Carnegie Mellon University

CARNEGIE INSTITUTE OF TECHNOLOGY

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF Doctor of Philosophy

TITLE Anomaly Detection of Piezometer Data

Collected from Embankment Dams

PRESENTED BY In-Soo Jung

ACCEPTED BY THE DEPARTMENTS OF

Civil and Environmental Engineering

Mario Berges	February 18, 2015
ADVISOR, MAJOR PROFESSOR	DATE
James H. Garrett	February 18, 2015
ADVISOR, MAJOR PROFESSOR	DATE
David A. Dzombak	February 19, 2015
DEPARTMENT HEAD	DATE

APPROVED BY THE COLLEGE COUNCIL

Vijayakumar Bhagavatula	February 19, 2015
DEAN	DATE

Anomaly Detection of Piezometer Data Collected from Embankment Dams

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Department of Civil and Environmental Engineering

In-Soo Jung

B.S., Civil Engineering, Carnegie Mellon University
M.S., Civil Engineering, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

February, 2015

Acknowledgments

First, I would like to acknowledge my advisor Professor Garrett, who has been my mentor since I was an undergraduate student in Carnegie Mellon University. Prof. Garrett has been the greatest support for me and guided me to transform from an undergraduate student to a PhD. His passion for work, care for his students and punctuality (and many others that I cannot list everything here) have motivated me to work harder and to reflect myself as a student and as a person.

Throughout my PhD journey, he helped me through tough times, provided me valuable advices and opportunities to grow as a better researcher and a stronger person. Without him, I would not have been able to start this journey.

I also thank my advisor Mario Berges, who has assisted me to research in various perspectives. I respect his enthusiasm for various research topics and willing to collaborate and share his research experience with his advisees. I also thank him for providing his students the opportunities to share their research experiences and give feedback to each other.

Next, I thank my thesis committee for agreeing to be a part of this work. Prof. Barnabas Poczos was instrumental in the technical aspects of the anomaly detection techniques employed in the research papers. His knowledge and expertise in this field have expedited my research. I also thank Prof. Matteo Pozzi for his insights and feedbacks that helped me refine my research approach and discussion.

I would like to acknowledge the financial support from US Army Corps of Engineers (USACE). The research in this thesis would not have been possible without the support from the USACE engineers, and I have had the pleasure of working with the remarkable people. I would like to thank Chris Kelly, who arranged important meetings and provided all of the relevant documents for me (his dedication and support were crucial in this research), and Andy Harkness, Paula Boren, Meghann Wygonik, and Cathy Bazán-Arias for helping me understand dam problems and for providing valuable feedback and discussion.

I cannot imagine my PhD without my colleagues. In particular, I thank Inferlab colleagues, Matineh Ebs, Suman Giri, Leneve Ong, Emre Can Kara, and Saurabh Taneja. Without them, I would not have been able to manage my stress and enjoy my work. It has been a pleasure going through my Ph.D life with such a great group.

Last but not least, I thank God for the grace he has given me. Without his grace, I would not be where I am today. And of course, I thank my family. I would not have been able to go through this journey without their encouragement and sacrifices. I would like to also thank my fiancé, Jongho Lee, who has given me endless support and made me smile even during tough times. I look forward to our lifelong journey.

Abstract

There are more than 85,000 dams in the U.S., the majority of which were built decades ago. It is not surprising then that the number of deficient dams, as qualified by different evaluation methods, has increased in recent years. Dams can pose significant risks to people living around them, and since they are exposed to harsh and largely unpredictable environments, it is important to maintain and inspect the conditions of dams effectively. In the United States (US), the current practice of analyzing the structural integrity of embankment dams relies primarily on manual *a posteriori* analysis of instrument data by engineers, leaving much room for improvement through the application of automated data analysis techniques. Accurately evaluating measurements collected from instruments used to monitor dam behavior is not an easy task, requiring sound engineering judgment and analysis, as well as robust statistical analysis techniques to prevent misinterpretation. In the research presented in this thesis, different types of anomaly detection techniques are investigated in an effort to i) propose which data analytics are appropriate for various anomalous scenarios as well as piezometer locations, and ii) to test if the widely-held assumptions on piezometer data, i.e., linearity between piezometer data and pool levels, as well as normally distributed piezometer data, are necessary in the anomaly detection task. This thesis specifically focuses on anomaly detection techniques that are applicable in analyzing piezometer data (collected from embankment dams) and anomalous scenarios that may lead to internal erosion. In order to validate how well the anomaly detection techniques perform, various anomalous scenarios and piezometer data of a case study dam are simulated using a numerical model. For each technique, various parameter settings are also examined.

In the real world, piezometers (and other instruments) in an embankment dam may not always be located at the optimum places, and their behaviors may not continue consistently over years. Thus, it is important to recognize how the data from a set of piezometers have changed over time as a group rather than evaluating only a single significant piezometer at a time. Since anomalies occurring inside dams may initiate with very small deformations without any visual signs, it is difficult to know exactly when and where the anomalies have initiated and would be located. However, by observing deviations among multiple (or grouped) piezometer readings over time, or the piezometers around a specific location, there is potential for obtaining better interpretations. Thus, this thesis also presents the research work on analyzing multiple piezometers together. The research examines if analyzing multiple piezometers together can improve interpretation and detection of piezometric anomalies, and shows the impacts of analyzing multiple piezometers compared to individual piezometers.

CONTENTS

List of Tables

List of Figures

CHAPTER 1. INTRODUCTION

1.1 Motivation

There are 84,000 dams in the U.S., and their average age is more than 53 years old. Since the design life for most of these dams was 50 years, this means they are in need of extensive rehabilitation (ASCE 2013a). Moreover, dams have received a grade of D in the 2013 ASCE infrastructure report card (ASCE 2013b). While dams provide drinking water, hydroelectric power, flood control, recreation and many other benefits, dams can pose significant risks to people living around them. According to ASDSO and FEMA (2012), even though dam failure is low probability, it causes high consequence; it typically happens somewhere in the U.S. every year. Thus, dam safety is one of the most important issues among the U.S. infrastructure systems that needs to be improved. To efficiently expedite the improvement, systematic inspections as well as more advanced monitoring systems are necessary.

Embankment dams are the most common type of dam in use today (ASDSO and FEMA 2012). For more information about embankment dams, please refer to Appendix A. In general, embankment dams are constructed of natural materials of the earth, commonly soil and rock (Schurer et al. 2002). In embankment dams, water seeps through different soil layers of the dam, and any changes in this behavior may signal problems. One of the leading causes of failure of embankment dams has been internal erosion, which can occur due to normal operations that may pose higher risks to a dam than remote loading conditions like floods and earthquakes (URBR 2010). Internal erosion problems are usually detected by periodic visual inspection and seepage

measurement (USSD Committee on Materials for Embankment Dams 2010). However, since they mostly occur without any visual signs, it is often too late by the time problems are actually identified. Any anomalous changes in pore pressure within a dam signal potential seepage problem. Thus, monitoring pore pressure as well as water levels of an embankment dam to observe potential seepage problems is important; anomalies that are indicative of internal erosion problems should be detected in advance to prevent catastrophic consequences.

Engineers monitor a variety of data collected from different instruments regularly to ensure that a dam is performing properly and as expected (Pelton 2000). If there is any abnormal change in flow rates or volumes of seepage over time, for example, this can be an indication of various problems, such as piping (pathways of water flow through the soil), cracks, malfunctions of the piezometers, etc. However, interpreting such movements from piezometer measurements is not always straightforward. Embankment dams, like most other civil infrastructure systems, are exposed to unpredictable environments. Moreover, their design specifications and as-is properties are not generally known due to, among other things, their age and the difficulties associated with assessing their internal structure. Hence, accurately evaluating measurements collected from instruments used to monitor dam behavior is not an easy task, requiring sound engineering judgment and analysis, as well as robust statistical analysis techniques to prevent misinterpretation. In the U.S., the current practice of analyzing the structural integrity of embankment dams relies primarily on manual *a posteriori* analysis of instrument data by engineers, leaving much room for improvement.

Piezometers are the most commonly used instruments to monitor the water level in dams, which can often be used to compute pore water pressure (Crum 2011). Besides the reservoir pool, piezometers are influenced by many factors such as precipitation, tailwater present in downstream, soil permeability, temperature, etc. For embankment dams, piezometer levels are highly influenced by reservoir levels, so they are usually compared together to monitor the seepage as well as to check the condition of the piezometers. In addition to regularly collected piezometric and reservoir readings, the data collected during past historical events, such as flood events, high pool operation, etc., are also considered so that the dam behavior during such extreme events can be understood. Engineers look for any anomalies by observing correlation plots that graph unexpected deviations from the established norm (or the trend line) that reflects the relationship between the two variables (See Figure 1). By observing how much the collected data are tracing or scattered around the trend line, it is possible to quickly verify if the piezometers are behaving as expected or not. However, it is often difficult to quantitatively determine if identified deviations depicted in the correlation plot actually represent problematic behaviors. In addition, besides yearly checkups, more intense analysis of the collected instrument data, including piezometer data, is normally done in 5+-year intervals in the U.S. Since dams often have slow responses, relatively small anomalies, which may be caused by initiations of any catastrophic events that are not obvious by just looking at correlation plots or those that are outside of the analysis period, may be easily overlooked. The traditional practice that subjectively detects deviations from the historical readings using time-series or correlation plots of raw data may not be accurate enough to capture anomalous changes over time. Thus, a

more advanced way of analyzing collected instrument data is necessary to improve interpretation of the data and thereby enhance dam safety.

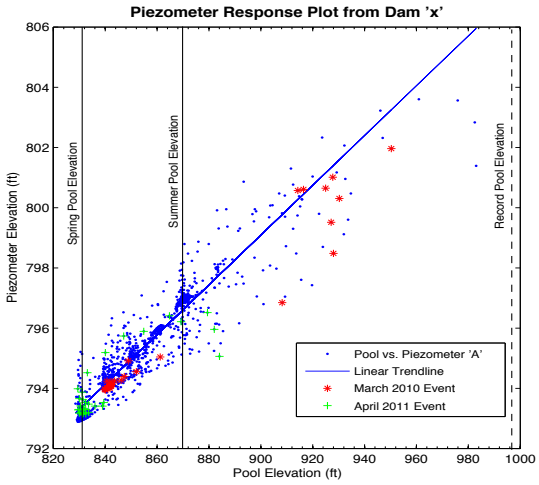

Figure 1. Traditional correlation plot (Pool vs. Piezometer)

In the dam safety domain, there is an increased interest on improving instrumentation systems by employing advanced sensors such as fiber optics, vibration wire piezometers, etc. (Zhu et al. 2011; Duarte 2013). However, there have been fewer efforts on improving the traditional ways of analyzing instrumentation data. There is opportunity to improve rational data analysis tools for the existing instruments, especially given that it is unlikely that new instrumentation systems will be deployed given the constraints of federal budgets (in the U.S.).

Thus, this thesis focuses on the development of advanced data analytics that are especially applicable to piezometer data (given that piezometers are the most commonly installed instruments in dams) that are obtained from embankment dams, so that anomalies that are indicative of the internal erosion problems can be detected in a more systematic way and robust

enough to be not affected by noise. Since it is important to not only detect anomalies as soon as possible, but also not missing any anomalies, anomaly detection techniques that can satisfy both criteria are studied and evaluated.

1.2 Research Questions

Having described the current limitations associated with analyzing piezometer data obtained from embankment dams, three main research questions are formed as follows:

- RQ.1. What are the anomaly detection techniques that are applicable to detect piezometric anomalies (in this research, anomalies are referred to any anomalous indicators that may possibly lead to internal erosion)?

Any unexplained changes (increases or decreases) in piezometric levels with time, which have been corrected for changing reservoir levels, indicate changes in the seepage behavior within a dam or foundation (USBR 2011). Such fluctuations in pressures indicate an episodic internal erosion process, so it is important to closely evaluate any change in piezometric levels from what has been observed in the past (USBR 2011). However, due to low inspection frequencies and subjective evaluation, oftentimes either important anomalous behavior may be overlooked or engineers may be misled by false alarms. Thus, applying more advanced anomaly detection techniques that are data-driven would improve the performance of the piezometer data analytics. The most challenging part of detecting internal erosion problems is that the initiation of the internal erosion process does not provide visible signs. For example, the amount of transported

sand does not change initially. In addition, there may be little warning for the failure until the process of the problem starts to evolve very quickly (Fell et al. 2003). During the full-scale experiments performed by (van Beek et al. 2010), the authors observed that water pressure decreases as a pipe forms, but starts to increase as the pipe enlarges. Thus, a local increase and/or decrease of water pressure is a good indicator of internal erosion; analyzing piezometer data has potential of indicating internal erosion problems. The research presented in this thesis studies various mechanisms associated with internal erosion as well as different types of statistical techniques that are appropriate in detecting anomalies among piezometer data.

RQ.2. What are advantages and disadvantages of the anomaly detection techniques that are applied to the piezometer data in dam safety?

In this research, different categories of anomaly detection techniques are examined that have been widely applied and validated from various domains by refining them to be applicable to detect anomalous piezometer data collected from embankment dams. Different categories of anomaly detection techniques are applied, either one at a time or combinations of them by considering critical conditions, advantages, disadvantages, and carefully understanding the nature of piezometer data. In this way, it becomes possible to analyze which of the techniques are more appropriate with respect to various anomalous scenarios. The importance of spatial locations of piezometers is also considered with respect to various anomalous scenarios.

RQ.3. How can analyzing multiple piezometers simultaneously improve the anomaly detection?

Since past behavior of piezometers may not continue indefinitely into the future, it is important to recognize how piezometers have progressed over time as a group (or interacting responses among them), rather than evaluating only a single significant piezometer at a time. Internal erosion has chaotic tendencies, so it is difficult to know exactly when and where it will occur. However, by observing deviations among the grouped piezometers over time, or the piezometers around a specific one, there is potential that better interpretations about anomalies can be obtained. In order to achieve this, various ways to group piezometers are investigated, e.g., by geotechnical characteristics, as well as understanding their expected behaviors given the characteristics of the groups. Based on which group of piezometers has been analyzed together, it would be possible to infer more information about the detected anomalies within the group, such as locations of where the anomalies might have been initiated, what other piezometers may be possibly affected shortly, etc.

The initiation of piping is often missed until it is physically observed, because the instruments do not always happen to be in the critical locations along every seepage pathway. However, if correlational structures can be understood by analyzing multiple piezometers together, it would be more intuitive to predict potential problems and their causes. In addition, by analyzing the groups' behaviors, significant patterns that may have overlooked from analyzing a single piezometer may be recognized. Thus, if the applications of anomaly detection techniques to multiple piezometers can detect anomalies with similar or better performance than analyzing individual piezometers, there will be much improved contribution to piezometer data analysis.

1.3 Roadmap

Chapter 2 of this thesis provides background that would be useful to understand the research problems. In Section 2.1, internal erosion is described. Sections 2.2 includes literature reviews on structural health monitoring (SHM) and SHM on dam safety. Section 2.3 introduces anomaly detection in general. We then show the potential of our research by presenting the applications of advanced statistical techniques on real piezometer data collected from two case study dams in Section 2.4 and describe the observed improvement in Section 2.5.

In Chapter 3, we explore and evaluate three different anomaly detection techniques that are selected to detect anomalies among piezometer data, and we extend our work to multiple piezometer analyses as presented in Chapter 4. Chapter 5 includes in-depth discussion of the research, and lastly, this thesis concludes with a summary, contributions, and future directions.

CHAPTER 2. BACKGROUND

2.1 Internal Erosion

According to Torres (2009), internal erosion is defined as "a process that develops concentration of seepage flows and in time may develop large cavities that may produce embankment failure with a catastrophic and uncontrolled release of the reservoir." There are very complicated mechanisms of internal erosion, and various flaws and physical factors can lead to internal erosion. Classical piping, progressive erosion, blowout, scour and suffusion are the well-known mechanisms of internal erosion. Internal erosion can happen due to, but not limited to, improper filters, when pressure relief measures are not used at seepage discharge points, when soils are not properly compacted adjacent to outlet pipes/sharp changes in foundation grades or slopes, when water flows along cracks or other defects in the soil, pressurized water flow into the embankment, etc. In most cases, soil particles are initially removed at the discharge point of the seepage, and pipes tend to enlarge and progress upstream with increasing flow quantity (MDEQ 2007).

According to Bureau of Reclamation (2012), internal erosion can be categorized based on loadings, i.e., 1) static/normal operation, 2) hydrologic and 3) seismic, and based on physical locations, i.e., internal erosion 1) in the embankment, 2) in the foundation, 3) from the embankment into the foundation, 4) along or into embedded structures, 5) into drains, etc. Thus, it is important to closely observe a number of areas during operation for signs of fracturing or unusual settlement that can lead to internal erosion problem (MDEQ 2007).

There are 4 main phases of internal erosion: initiation, continuation, progression and breach as shown in Figure 2. This figure is showing a specific case of when internal erosion in embankment is initiated by concentrated leak (Fell et al. 2003).

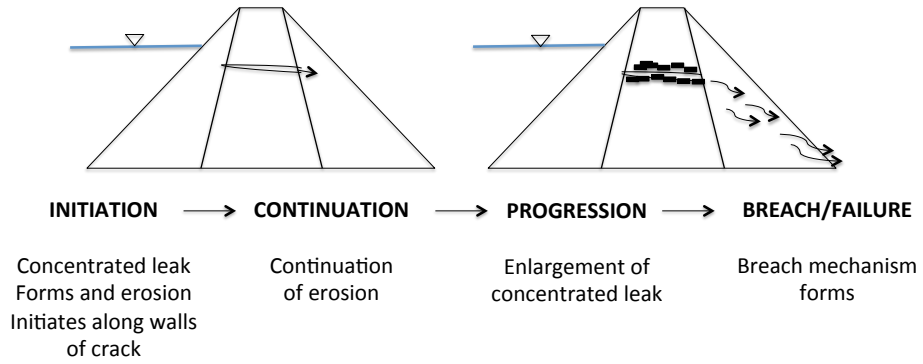

Figure 2. Main phases of Internal Erosion (Source adopted from (Fell et al. 2003)).

There are various important factors to be examined for each phase of internal erosion. For example, during the initial stage, engineers should look for the potential for concentrated seepage paths, the possibilities of defects, the hydraulic gradients along the path, and soil erodibility (Fell et al. 2003). In addition, earthquakes and defects in embankment dams, such as cracks, pervious layers, and many others should be carefully considered, because they might increase the likelihood of initiation of internal erosion (Bureau of Reclamation 2012).

As explained above, internal erosion is a very complicated process, and site-specific conditions should be also considered to better understand its likelihood to develop further. In addition, internal erosion problem cannot be completely analyzed using numerical models; additional

information on dam and soil behavior need to be assessed to determine the risks associated with internal erosion (Bureau of Reclamation 2012).

Any abnormalities, such as heave, cracks, removal of materials, etc. in a dam may or may not continue to develop actual internal erosion. However, such observations should be carefully examined due to the possibility of being further developed to initiate internal erosion. Thus, it is difficult to state that detected anomalies are actual internal erosion; rather, they should be considered as potential *indicators* of internal erosion. Thus, only secondary consequences of internal erosion are considered, meaning that we do not judge whether detected anomalies will always lead to internal erosion problems, but rather, we try to detect any types of *potential* anomalous responses that may possibly continue to actual internal erosion.

2.2 Structural Health Monitoring

Due to deterioration and extreme events, civil infrastructures, including dams, should be carefully monitored to maintain their performance and safety. Thus, structural health monitoring (SHM) has been emerging and studied for the last three decades (Chang et al. 2003). SHM techniques have been developed to assess conditions of different types of structures. During SHM process, it is important to detect damage, thus preventing further catastrophic failure. Since damage can be caused by various reasons, including varying operational and environmental conditions, the process of assessing the nature of damage in a structure is a complicated task especially when the damage is not substantial. SHM techniques can be largely divided into model-based and data-driven techniques (Ying et al. 2013; Barthorpe 2010).

The model-based approaches require understanding of complicated relationships among relevant parameters, and there are often uncertainties related to conditions that may affect the behavior of the structure. Thus, data-driven approaches are often preferred especially when there is no prior knowledge about the structure.

Data-driven approaches analyze the data collected from the structure of interest, so the main advantage of such approaches is that they require only sufficient measured field data. Since the behavior of the structure is monitored over time, it is important to understand the influence of 'time' on the data to detect any behavioral changes of the structure. A large number of studies have been performed to improve anomaly detection in time series (Huang et al. 1998; Tsay et al. 2000; Bianco et al. 2001; Ma et al. 2003; Basu and Meckesheimer 2007).

In SHM domain, for example, Nair et al. (2006) used the vibration signals obtained from sensors and modeled autoregressive moving average (ARMA) time series. They used a hypothesis test using the t-test and localization indices defined in the Autoregressive (AR) coefficient space to identify and localize minor to severe damage in the benchmark structure, which is a four story, steel-braced frame. Omenzetter and Brownjohn (2006) analyzed the strain time series obtained from a major bridge structure by applying a vector seasonal autoregressive integrated moving average (ARIMA) model, and they concluded that the proposed method could detect the structural damage by observing various changes in the ARIMA model coefficients.

Another common technique that has been used to model the time series is subspace identification approach, which uses statistical decomposition techniques like Principal Component Analysis (PCA) (Yan and Golinval, 2006; Yu et al. 2010; Wang and Ong 2010; Tibaduiza et al. 2012). The idea is to observe the changes in the significant correlation structure of the time series. (Yan et al. 2008) applied PCA to extract essential features from the recorded traffic-induced bridge-deck strain time histores, and those features were used as input to further classify classes of vehicles using Neural Network.

Posenato et al. (2010) applied MPCA, which is Moving PCA that accomodates time-varying process behaviors by moving a window that contains data of a defined time period, with combination of other statistical methods and concluded that these techniques could discover anomalous behavior in civil engineeirng structures. Recently, Zhang et al. (2013) applied PCA to tire excited acoustic signal measured by a moving vehicle equipped with a directional microphone to estimate the MTD (Mean Textrue Depth) value of pavement, which is used to quantify the pavement surface condition. With PCA, they could successfully eliminate the noise that is not associated with the road surface and extract important information about the road surface. Other data-based SHM approaches can be found in (Farrar and Worden 2012), which introduces various statistical models that use machine learning algorithms, including Neural Network, Decision Trees, Maximum Likelihood, Factor Analysis, etc.

SHM on Dam Safety

In the dam engineering community, SHM is equivalent to surveillance (Brownjohn 2007). Typical dam safety surveillance consists of visual inspections and monitoring instrumentation data. Recently, the use of non-intrusive geophysical techniques has been largely investigated to facilitate early detection of anomalous phenomenon, e.g., anomalous seepage, piping and internal erosion (Lum and Sheffer 2005). To assist in monitoring for dam safety and detect internal erosion problems, there have been a number of efforts by the dam safety community. In the past, mathematical models and empirical relations were developed to prevent piping (Bligh 1910; Sellmeijer 1988). Then further research has been performed to validate such pioneering models as well as to monitor seepage problems.

French and Tschantz (1986) performed a correlation analysis as wells as regression analysis between piezometric head and tailwater in an effort to identify changes in the seepage field in time. A multiple linear regression model based on the modified Bernoulli's equation, Darcy's Law and Laplace equation was compared with a hydrostatic model and an overall modest improvement in performance of the multiple linear models over simple linear models was observed. The authors, however, questioned the applicability of such models due to errors resulted from simplifications of the modified diffusion equations. More refinement of the regression models was also necessary to better describe headwater/tailwater time functions for comparison with piezometer data. In addition, since headwater/tailwater fluctuations would not be immediately transmitted to piezometers at depth, the authors suggested that other predictor variables should be also considered to improve model performance.

The rate of development of internal erosion and piping was studied in (Fell et al. 2003) by accounting for the nature of the soils and materials of the dam and its foundation, and the mechanisms of internal erosion. Then the authors discussed the ability of monitoring methods to detect internal erosion by estimating time for development of internal erosion in embankment dams. However, the methods they proposed for assessing the rate of development of internal erosion are very approximate and should be only used as a general guide to performance.

In order to predict seepage path through the body of an earthfill dam, both physical-based model and a data-driven model were developed and compared in (Tayfur et al. 2005). Even though such models showed potential to detect the anomalies in seepage path, their main objective of developing such models was to calibrate and verify the performance of dams, thus lack systematic ways to identify anomalies using quantitative measures.

More recent effort was on investigating the failure process of dams via a full-scale experiment (van Beek et al. 2010). Four phases of the failure process were identified and discussed in detail. In addition, the relationship between time and hydraulic head with respect to the failure process was presented. Even though this work would be useful to understand different phases of piping in detail, they also did not address ways to quantitatively assess dam condition. In addition, in order to understand failure process in other dams, developing a full-scale experiment is very unlikely in real life, and significant factors of determining the potential piping problems such as soil plasticity and grain size are expected to vary.

In an effort to address such gaps in analyzing performance of dams, we investigate systematic anomaly detection techniques that can detect anomalies using piezometer data, which is the most common instrument in embankment dams. We considered the techniques that can detect anomalies by reducing subjectivity involved with engineering judgment as well as that are robust from noise present around dam sites.

2.3 Anomaly Detection

Anomaly detection refers to "the problem of finding patterns in data that do not conform to expected behavior." (Chandola et al. 2009b). Since 19th century, detecting anomalies in data has been studied in the statistics community, and there have been a variety of anomaly detection techniques developed for certain application domains and for more generic problems (Xu 2010).

Anomaly detection techniques/algorithms should be selected by considering if they are suitable for the data sets in terms of the correct distribution model, the correct attribute types, the scalability, the speed, etc. (Hodge and Austin 2004). In addition, selecting appropriate approaches would depend on the data type, the ground truth of the labeling, how one wishes to detect anomalies and to handle them (Hodge and Austin 2004). While there are a number of anomaly detection techniques that have been validated from various domains, applying a technique developed in one domain to another is not an easy task, especially because how an anomaly is defined varies for different applications (Chandola et al. 2009a). In most times, it is necessary to modify some parts of the algorithms to meet the objectives of an application. Broad and general examples of different anomaly detection categories would be classification-based,

nearest-neighbor based, clustering-based, statistics-based, and spectral-based techniques (Hodge and Austin 2004; Anscombe and Guttman 1960).

Each category of the anomaly detection techniques has different advantages and disadvantages. First, classification-based techniques use a set of labeled instances to learn a classifier and test if a test instance is normal or not using the trained classifier (Weekley et al. 2010). Example algorithms are Neural Networks, Bayesian Networks, Support Vector Machines, Rule-based, etc. Even though the testing phase of classification-based techniques is comparatively fast since the classifier has been already computed by the time, one of the disadvantages of this type of techniques is that their performance can be heavily affected by the accuracy of available data labels, which are often rare in anomaly detection (Chandola et al. 2009a).

Nearest-neighbor based techniques assume that normal instances occur close together while anomalies occur far from their neighbors (Zhao and Saligrama 2009). There can be an infinite number of distance or density measures, including Euclidean distance, Mahalanobis distance, etc. to compute the closeness or similarity among data. Using such metrics, anomaly scores are assigned to each instance to distinguish anomalies from normal instances. Example algorithms would be kNN (k-Nearest Neighbor), Density-based techniques, etc. One of the advantages of this type of techniques is that they are data-driven approaches, so no assumption on the distribution of the data is required. On the other hand, problems may arise if normal instances are not enough compared to anomalies since anomalies may have more numbers of close neighbors. The computation time would be another problem since the distance of each test instance should

be computed. Lastly, appropriate distance measure should be carefully selected since its performance will be heavily affected by it.

Clustering-based techniques are very similar to nearest-neighbor based techniques, because a distance or density is computed to detect anomalies, e.g., distances between instances to centroids of clusters or densities of identified clusters (Chandola et al. 2009a). In clustering-based techniques, anomalies are defined as those that do not belong to any clusters or those that form comparably small clusters. Example algorithms are K-means Clustering, Expectation Maximization (EM), etc. One of the biggest problems of clustering can arise when anomalies are forced to belong to a large cluster, thus considered as normal (Chandola et al. 2009a).

In statistical anomaly detection techniques, a stochastic model is assumed, and any observations that are partially or wholly irrelevant to the model are considered as anomalies (Anscombe and Guttman 1960). There are two techniques, parametric and non-parametric to learn a statistical model. Parametric techniques assume the underlying distribution and estimate the parameters from the given data (Eskin 2000). A statistical hypothesis test, for example, can be associated with a test statistic to compute probabilistic anomaly score for test instances (Chandola et al. 2009a). Here, anomalous instances are assumed to occur in the low probability regions of the stochastic model. Gaussian model-based (e.g., Grubb's test, box plot), and regression model-based, which use residual values to determine anomaly scores (e.g., Autoregressive Moving Average) are the representative examples of parametric techniques. Examples of nonparametric techniques, which do not require any *a priori* information, are histogram (counting)-based,

kernel function-based, etc. In statistical anomaly detection techniques, the assumption of data distribution is very important, which can make the application either advantageous or disadvantageous.

Lastly, spectral anomaly detection techniques try to find anomalies by determining subspaces that capture the variability among the data. Any deviations from the correlated structure or low variance will get a larger anomaly scores (Parra et al. 1996). Principal Component Analysis (PCA) and variants of PCA, such as Singular Spectrum Analysis are the examples of spectral techniques. Since this type of techniques can reduce dimensionality, they are very useful when analyzing high dimensional data sets; however, computational complexity can be often high.

The next section presents the initial studies that were conducted to demonstrate potential of applications of statistical techniques (that have been widely implemented in other domains) in improving piezometer data analytics.

2.4 Motivating Initial Study: Application of MPCA on piezometer data

When engineers monitor the performance of dams, they review instrumentation data and use additional sources (Pelton 2000), e.g., past inspection reports, construction photos, historical events, etc., as references. Especially when engineers are concerned about possible seepage problems, piezometer readings are closely analyzed. If the piezometer level is directly influenced

by the reservoir level with no other significant stimulus, the correlation plot will show a straight line (Gall 2007). Based on where the piezometers are located, i.e., the soil layers, distance between the piezometer and the reservoir, etc., the slopes of the correlation plots will vary due to different response times.

As mentioned above, the projection line of the correlation plot is generated to analyze piezometer data, and engineers include historic high pools as well as their corresponding maximum piezometer readings so that they can be compared with the previous piezometer responses. The projections are used as monitoring/action limits during future events to quickly verify that the piezometers are reacting in a predictable manner or whether additional monitoring is required or not. As the piezometer is influenced by other factors and lags behind the reservoir level, the data points will scatter along a sloped line, forming an elliptical envelope.

The major problem of analyzing such plots is that it is difficult to set the threshold values, or levels that need to be set around the straight line (or the envelope) to distinguish unexpected readings, or anomalies from normal readings. Spikes in the piezometer data are rare, but since they are theoretically indicative of filter failure, piping, hydraulic fracturing, or other phenomena, it is difficult for engineers to dismiss them as bad data (Crum 2011). Thus, instead of visually identifying the data that deviate from the established norm, there should be a more quantitative and robust approach of detecting anomalies, not only to ensure dam safety, but also to reduce any subjectivity and efforts required by engineers. Initially, we applied statistical anomaly detection techniques that have been widely used and validated from various domains

and observe if such techniques can be also useful when analyzing piezometer data from embankment dams.

As mentioned in Section 2.3, PCA-based methods in combination with other statistical approaches have been widely used for anomaly detection. A recent work applied Moving PCA (MPCA) in combination with four regression analysis methods for damage detection in bridges (Laory et al. 2013). They compared these combined methods with stand-alone methods to see which ones provided highest levels of damage detectability as well as earlier detections. Thus, we also applied MPCA, which has been tested by Laory et al. (2013), to see if this method can be also useful when analyzing piezometer data from embankment dams. The main motivation was to improve piezometer data analysis by implementing statistical anomaly detection, thus reducing the subjective quality of the analyses that are commonly carried out by engineers today and the errors that come from this practice. Before presenting the applications and results, brief theoretical explanations of MPCA and one of the regression analyses, Robust Regression Analysis (RRA) are presented in the following paragraphs.

MPCA

By linearly decomposing a data matrix into a number of independent components, PCA can identify periodic variations that are dominant in the data. While PCA is often applied to the whole dataset, it can also be applied to a subset, or a window, of the dataset. MPCA performs PCA by sliding this window, so that any change in the first several principal components over time can be detected.

More formally, consider a data matrix, $T \in \mathbb{R}^{N \times M}$, whose M columns are individual time-series of length N (e.g., measurements from individual piezometers) that have been normalized with respect to each column. Each entry of this matrix can be denoted by $V_i(t)$, where $i = 1, \ldots, M$ and $t = 1, \ldots, N$, as shown in the equation below. $V_i(t)$ would indicate the measurement of piezometer *i* at time *t*.

$$T = \begin{bmatrix} V_1(1) & V_2(1) & \cdots & V_M(1) \\ V_1(2) & V_2(2) & \cdots & V_M(2) \\ \vdots & \vdots & \cdots & \vdots \\ V_1(N) & V_2(N) & \cdots & V_M(N) \end{bmatrix}$$

To apply MPCA on T, first a sliding window of size L is applied to the matrix, to extract a sub-matrix, called $R(k)$ at each time value k, where $k = 1, \ldots, (N - L)$. Then, a singular value decomposition (SVD) is performed on each one of the covariance matrices, $C = \frac{1}{N} R(k)^T \times R(k)$. During SVD, the matrix, T gets decomposed into matrices $U, S,$ and V, where $C = USV^T$. The columns of U are the left singular vectors while those of V are the right singular vectors. S is a diagonal matrix with singular values along the diagonal. Since *C* is symmetric, the right singular vectors correspond to the eigenvectors, E_l, and the diagonal elements of *S* correspond to the square roots of the eigenvalues, e_l, of the matrix (where $l = 1, \ldots, L$).

The eigenvectors of the covariance matrix represent the directions of maximum variance, or the variance of each independent component, and the corresponding eigenvalue indicates a degree of each component's proportional variance. Thus, the most dominant patterns between reservoir

and piezometric readings can be captured by the first few sets of the eigenvectors after ordering the corresponding eigenvalues in a descending order.

RRA

After the direction of most variability is computed in each time step, the next phase is to determine whether this eigenvector changes over time, which would signal the presence of an anomaly. RRA is known as a good regression technique to use in the presence of outliers. Among many types of robust regression models, the method that uses iteratively reweighted least squares with a bisquare weighting function was employed. So after applying MPCA, RRA is performed to observe if any changes in the first few relevant eigenvectors from $R_i(k)$ have occurred over time. The number of eigenvectors to be monitored can be determined based on how sensitive the anomaly detection needs to be. The regression model is formed based on the normal data, and the threshold level is determined by computing the standard deviation of absolute values of the regression residuals (a difference between actual and predicted values) in the normal data. Any regression residuals that exceed this threshold are marked as anomalies.

2.4.1 Case Study and Discussion

To test the efficacy of this anomaly detection algorithm, a case study dam located in the eastern part of the U.S. was chosen, for which instrumentation measurements were accessible. The dam is an earth and rock fill structure and is composed of a central core of impervious rolled fill with the upstream side slope protected by rock fill on gravel bedding. On the downstream of the core, there is a large zone of rock fill. There are a total of 26 piezometers installed, and currently,

measurements are collected automatically every 4 hours. In the application, the time-series data sampled at 12-hour intervals were used.

Applications and Results

Since the main influencing variable to the collected piezometer data is the pool level, MPCA was applied to this pair of highly correlated variables (i.e., pool and one of the piezometers installed in the case study dam). The Pearson's correlation coefficient between the two variables was 0.983.

Other than minor seepage, which is common in embankment dams, there have been neither major structural problems nor serious turbid water effects in the dam. Thus, to test the ability of the selected approach to detect anomalies as well as to reduce any bias of where the anomalies are introduced, anomalies were simulated in a number of time periods with a constant interval. By setting the first two years of the data as a normal condition of the dam, 4 months of anomalous data were introduced for each subsequent time period in each experiment. Given the length of the piezometer data, we could experiment with 6 unique abnormal time periods without overlaps. To simulate the anomalies, the original piezometer data during the chosen "abnormal" periods were randomly reordered in time. This approach was chosen, instead of artificially generating new data, so that the resulting data were kept within the piezometer ranges of the original periods and the primary effect would be a de-correlation between the pool level and this particular piezometer during that period. Even though real anomalies were not characterized, this simulated anomaly would represent a problem when the piezometers are not responding to the pool levels, which may occur due to serious seepage problems and/or when the piezometers

themselves are malfunctioning. To make sure the randomly reordered datasets are de-correlated compared to the original dataset, only the reordered datasets that have lower correlations than the original one were tested.

The time series data shown in Figure 3 corresponds to one of the 6 experiments we have tested on. The piezometric elevation corresponds to the readings from one of the piezometers installed. A dotted rectangle indicates the period where the anomalies were added to the piezometer data. While the normal piezometer data remain highly correlated with the pool data, the abnormal piezometer data no longer respond well to the pool levels. Both readings shown in Figure 3 indicate non-normalized data.

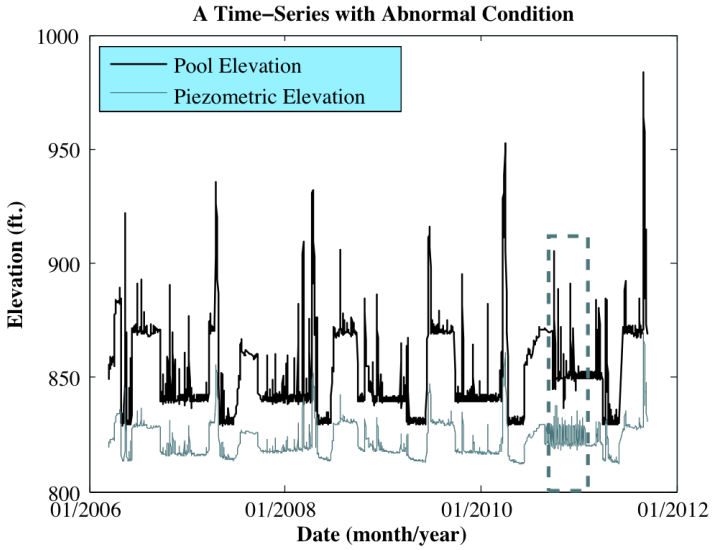

Figure 3. Time-series of the pool and the piezometric elevations with anomalies (dotted rectangle)

When applying MPCA, a window size of 730, which corresponds to a year, was used to capture a periodic behavior of the dam. In this experiment, the changes only in the first eigenvectors were observed. After computing eigenvectors through MPCA, Robust Regression Analysis was

performed. The regression residuals were calculated based on the normal data, and 13 different standard deviations (from 3 to 15) of the regression residuals were cross-validated over the 6 experiments to see which standard deviation can detect the anomalies most accurately. When the sensitivity analysis was performed, the standard deviation of 6 had the best true positive rate (TPR) and the false positive rate (FPR) over the six experiments, which were 0.82 and 0.12, respectively. This high TPR and the low FPR validate that the tested method can detect anomalies in the piezometer data successfully with an average accuracy of 0.86, which was obtained by taking the ratio of sum of true positives and negatives to the total. Figure 4 shows the result of the MPCA+RRA statistical detection technique applied to one of the piezometers installed in the case study dam after simulating anomalies over 4 months. As shown in the figure, the residuals exceed the threshold line when there are simulated anomalies, showing the potential of this method to analyze piezometer measurements.

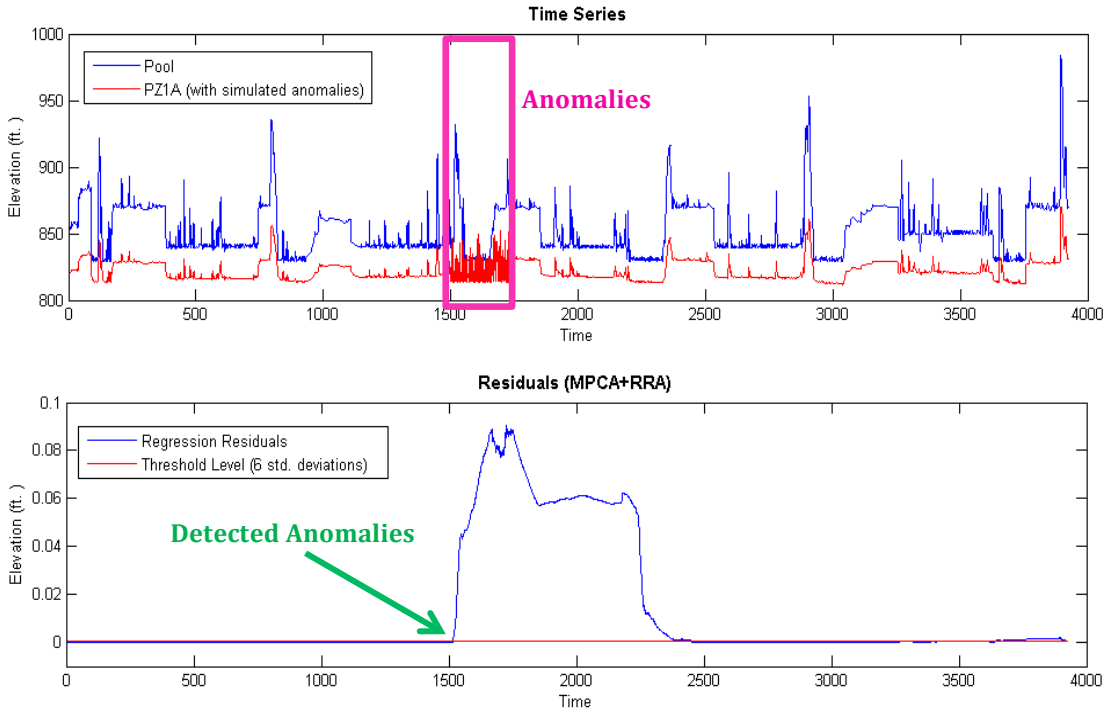

Figure 4. The result of MPCA+RRA (above: time series with simulated anomalies. below: regression residuals) applied to a piezometer of the case study dam

Discussion

In this initial study, MPCA and RRA were applied as anomaly detection. To test anomaly detectability, 6 different anomalous datasets were introduced subsequently after the period of the normal data. Then, several standard deviations of the regression residuals were cross-validated to find a proper threshold level. The results of the 6 experiments presented a high true positive rate. Thus, the proposed anomaly detection method showed the potential as a promising data-driven method to analyze piezometer measurements.

When a contingency table for each experiment was generated, variations among the numbers of true positives, true negatives, false positives and false negatives could be observed. This result was due to randomly generated anomalies that were introduced in each experiment. Thus, instead of shuffling the original data, we decided to simulate anomalies that are caused by more realistic scenarios. The next chapter presents our new simulation approach using a numerical model.

2.4.2 Case Study 2 and Discussion

In addition to the first case study dam, we assessed piezometer data collected from another case study dam (CSD2), which had known anomalous behaviors in the past. Thus, instead of having to simulate anomalous datasets, we applied the anomaly detection technique, MPCA+RRA to CSD 2 data directly. In this dam, the piezometer data have been collected daily since 1987, so 26 years of data were tested.

Figure 5 shows the result obtained from piezometer 'A', which is one of the piezometers installed in CSD 2. The piezometer A is installed on the upstream side of the crest of the dam. The year 2007 was the first time to raise its conservation level, which means new baseline was established. As shown in the time series (the upper plot of Figure 5), there has been no piezometric response to pool until 2007, but it started to respond more to the pool. The result of the proposed anomaly detection shows that such change in 2007 could be detected. Since the water levels started to rise sharply during the conservation pool and drop gradually, the anomalies might have been caused by unusual lags in the piezometric response. If such anomalies continue, it is recommended that engineers investigate if any of the internal properties

of the dam have been affected by this extreme event. According to Fell et al. (2003), in the vast majority of cases of piping through the embankment, the reservoir was either at a historic high level, or within 1m of the historic high level when progression of erosion occurred to form a pipe. Thus, considering unusual changes in reservoir level would be an important task.

Figure 6 shows the result obtained from another piezometer in CSD 2. According to the engineers who manage CSD 2, a single row grout curtain was installed in upstream side of the right abutment in 2001-2002. This installation might have caused the anomalies detected in 2001. The anomalies occurred in 2003 could not be explained with any solid evidence. The anomalies of 2009 might have been caused by turbid water observed at one of the tunnel drain in 2009.

As shown in Figure 5 and 6, the MPCA+RRA could provide additional anomalous behaviors (previously unknown) to the engineers (though this approach might have been too sensitive) as well as validate the previously known anomalous events. Since the engineers from CSD2 used to analyze their piezometer data every 5 year, they might not have detected changes that have occurred before the 5-year period, possibly overlooking any long-term changes. However, since the tested statistical technique has a capability to detect deviations over long periods of time, any chance of missing important changes in the data could be reduced.

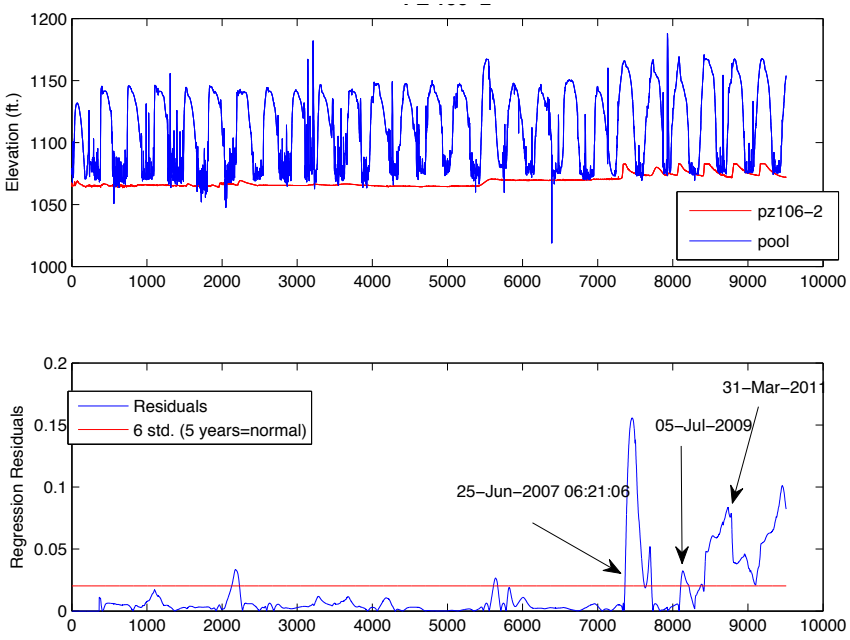

Figure 5. The result of MPCA+RRA (above: time series with simulated anomalies. below: regression residuals) applied to a piezometer of CSD2.

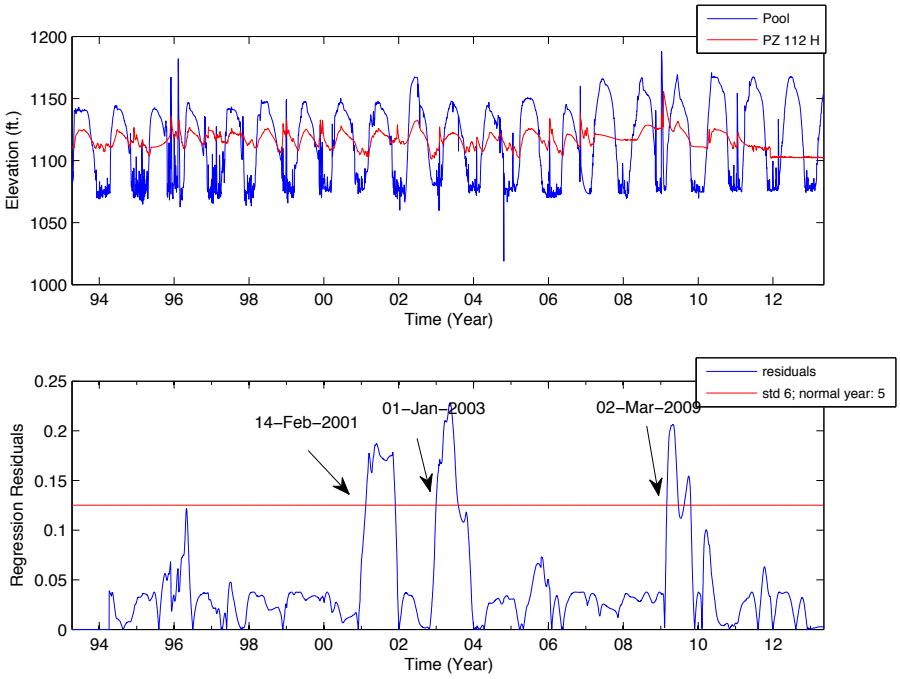

Figure 6. The result of MPCA+RRA (above: time series with simulated anomalies. below: regression residuals) applied to a piezometer of CSD 2.

2.4.3 Autoregressive (AR) Model

AR model, which will be presented more in detail in Section 3.1.3, is another statistical approach we have applied, which has been widely used in various domains to describe time-varying processes. The AR model is based on a linear regression method, in which the current value of the time series is estimated based on linear combination of its previous values of the series. The number of the prior values used is defined as the order of the AR model. Equation (1) below shows an AR model of order *n*. The x_t represents the value of *x* at time, *t*. The equation shows that x_t can be estimated using a linear relationship among *n* previous *x* values and their corresponding coefficients, *a*'s. ε_t is the Gaussian white noise term. There are different ways of estimating such coefficients, e.g., least-squares, Yule-Walker approach (Kay 1999), Forward-backward approach (Marple 1986; Armour and Morgera 1991), etc.

$$\boldsymbol{x_t} = \left(\boldsymbol{a_1 x_{(t-1)}} + \boldsymbol{a_2 x_{(t-2)}} + \boldsymbol{a_3 x_{(t-3)}} + \cdots + \boldsymbol{a_n x_{(t-n)}}\right) + \boldsymbol{\varepsilon_t} \quad (1)$$

Applications and Results

In the AR application, a dataset from CSD2 was used. The standard least squares technique was used to determine AR coefficients and an AR model with 5^{th} order was used. There are many different procedures to choose the best order or AR processes, e.g., FZC (First Zero Crossing of the autocorrelation function), the FPE (Final Prediction Error), AIC (Akaike's Information Criterion), etc. (Schlindwein and Evans 1990). However, due to computation time, we first considered the order that is less than 10 with a trial-and-error method. As a result, the order of 5 gave us the best result (higher orders either did not improve or decreased accuracy) with minimum false alarms and false negatives.

First, the AR coefficients were estimated from the normal time series data, x, and residual errors, ε, were calculated. Then, residual errors of 'damaged' data, (x') was calculated using the AR model developed from the 'normal' data as shown in Equation (2).

$$\varepsilon'_{(t)} = x'_{(t)} - \left(a_1 x'_{(t-1)} + a_2 x'_{(t-2)} + a_3 x'_{(t-3)} + a_4 x'_{(t-4)} + a_5 x'_{(t-5)}\right) \quad (2)$$

Given the total length of the data, the first 5 years of the data were assumed to be normal. Then, AR coefficients (AR (5)) obtained from the normal data were fitted to the next data points that may contain anomalies.

The result shown in Figure 7 is a control chart based on the normalized residual error time series in which its upper and lower control limits were determined by +/- 3 standard deviations and the centerline was determined by its mean. Any points that exceed the control limits were flagged. The resulting AR model was too sensitive to any deviations contained in the data; this is because only piezometer data were analyzed instead of analyzing how the relationship between the pool and piezometer readings change over time. In this way, unusual behaviors while they are *not* abnormal (such as unexpectedly high pool levels or maintenance periods) may not be reflected, thus possibly causing many false alarms. Possible improvements would come from consideration of seasonality or annual periodicity when selecting the model order.

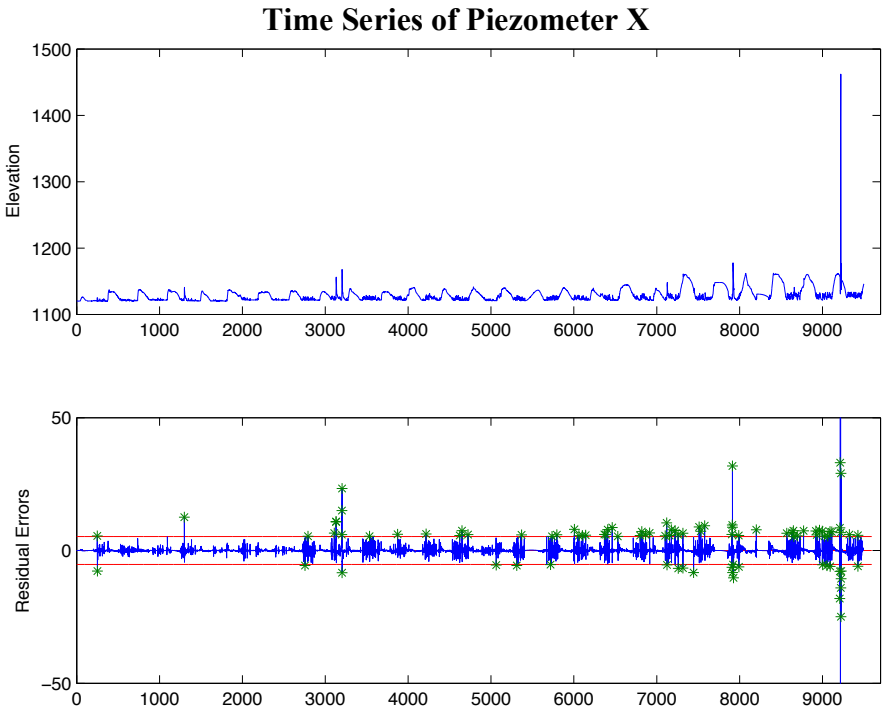

Figure 7. Above: time series of a piezometer from another case study dam; below: the result of AR application (green '*'s are the detected anomalies from AR while red lines indicate the threshold range)

2.5 Improvements Observed from Applications of Statistical Anomaly Detection Techniques

While the traditional practice from the engineers generates correlation plots directly from the instrumentation data and qualitatively detect any deviations from such plots, one of the statistical anomaly detection approaches we applied, MPCA detects abnormal behaviors from the *dominant patterns*, or the major variability between the two variables, i.e., pool and piezometer readings, of the instrumentation data over time.

Figure 8 shows a correlation plot as most engineers would generate with the piezometer data. When a simple regression method is applied directly on the data, many anomalies (red dots) especially those that are embedded in the hysteresis cannot be detected. However, when dominant patterns obtained from MPCA are used, instead of the data directly, most of the anomalies even inside the hysteresis can be detected as shown in Figure 9. Thus, we can conclude that anomalies that are embedded in the hysteresis, which are often hard to be visually detected, can be better detected when MPCA is applied before performing the regression analysis.

As shown in this section, the applications of the statistical anomaly detection technique had advantages over the current data analytics. Thus, it has shown the potentials of applying statistical anomaly detection techniques to detect piezometric anomalies. We can infer from the different results obtained from MPCA+RRA and AR that distinct statistical techniques would bring different results as well as interpretations, meaning there maybe well-suited techniques for different failure scenarios.

The next chapter presents exploration and evaluations of different categories of anomaly detection techniques by considering critical conditions, advantages, disadvantages, and carefully understanding the nature of piezometer data. Such various approaches resulted in different anomaly detection performance, providing more insights to understand problems.

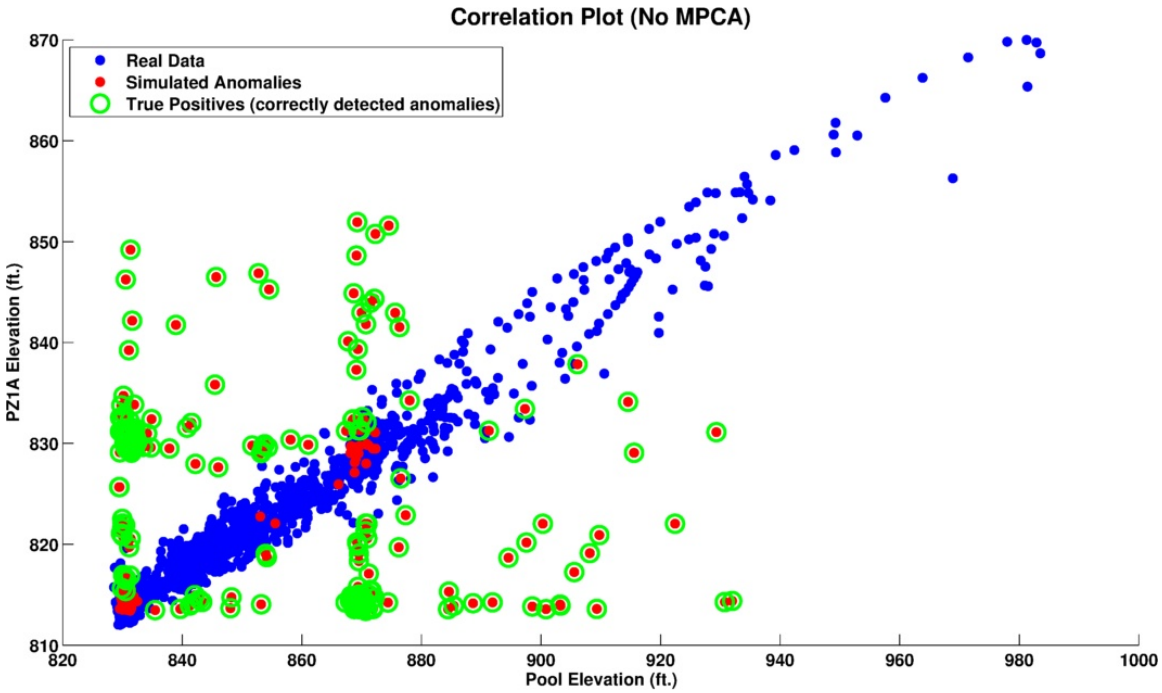

Figure 8. Correlation plot without the proposed statistical technique

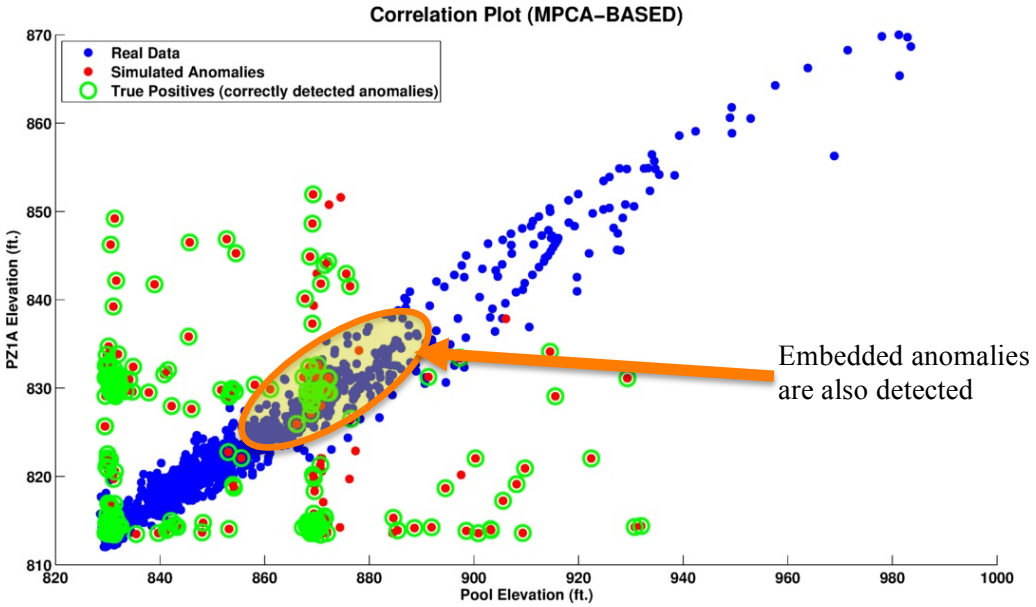

Figure 9. Correlation plot generated based on the statistical technique (MPCA+RRA) applied. Even those red dots, which are the simulated anomalies, inside the hysteresis are detected.

Chapter 3. APPLICATIONS OF AR, MPCA, AND KL ANOMALY DETECTION TECHNIQUES TO EMBANKMENT DAM PIEZOMETER DATA

Given that different parts of a dam behave differently and undergo complicated processes over time, it is possible to improve the traditional anomaly detection approach using more advanced data analysis techniques. Thus, we investigated different types of anomaly detection techniques in an effort to propose/recommend data analytics that are more appropriate for various anomalous scenarios as well as piezometer locations. In this section, we present how we tested Autoregressive (AR) model, Moving Principal Component Analysis (MPCA) and Kullback-Leibler divergence, which entail different characteristics.

In the next section, we first describe the details of the simulations performed for the various scenarios considered in this work. Then, we introduce the three anomaly detection techniques that were evaluated as part of this research; we describe the principles and characteristics of those techniques as well as why they are appropriate for the anomaly detection task given our data. The following sections include the experiments we conducted and the corresponding results obtained from these experiments. Lastly, we conclude with the discussion of these results.

3.1 Approach

3.1.1 Anomalous Data Collection

Since the anomalous problems on which we focus in this work often occur inside embankment dams without visual signs until they become severe enough to be observed, it is hard to know exactly when such problems are initiated. Thus, a simulation process is necessary in order to accurately evaluate the ability of the anomaly detection techniques to detect anomalies long before they become severe enough to be visually observed. Moreover, such simulations will feed additional information, such as how piezometric levels would be affected by the specific physical processes we simulate, which have been hard for engineers to understand based only on observations of and data collected from real embankment dams. The datasets collected form these embankment dam simulations will not only benefit the validation of anomaly detection approaches, but also contribute to the understanding by dam engineers of the behavior of piezometers in embankment dams under different conditions. In order to simulate realistic anomalous embankment dam scenarios, a case study dam we introduced in 2.4.1 was used in which its reservoir and instrumentation measurements were accessible to the research team.

Recently, numerical analyses as coded into computer programs have been the most widely used method to analyze seepage issues. Specifically, 2-dimensional (2D) or 3D numerical models are often developed to analyze seepage issues (USBR 2011). Such models are often used for simulation of infiltration, seepage and seepage path, etc., so that the performance of the dam can

be validated. In our work, we used SEEP/W[1] to model a section of the case study dam in order to simulate various anomalous scenarios, and consequently, to collect corresponding piezometer data. It is a finite element software program that adopts an implicit numerical solution to solve Darcy's equation[2] for saturated and unsaturated flow conditions over space and time (Krahn 2004). Using SEEP/W, it is possible to vary relevant parameters, such as reducing the hydraulic conductivity of soil layers and/or the core, thus obtaining datasets that correspond to different phases of internal erosion. In addition, steady-state and transient seepage analyses (during specified time sequences) are possible.

The existing 26 piezometers already installed in the case study dam are clustered in the core area, so we selected some of the existing piezometers and several new locations to collect piezometer data that have various response characteristics to the pool levels, and are spatially spread along the dam section we model. Using the daily reservoir levels obtained from the case study dam for five and a half years, specifically Mar. 2006 to Sep. 2011, corresponding daily piezometer data were collected from the 10 specific locations in the model as shown in Figure 10. Table 1 provides more details about each piezometer.

[1] SEEP/W is a finite element software program for analyzing ground water and excess pore-water pressure dissipation problems in a porous media. Users can compute flow quantities, uplift pressures and many others by selected interest locations in the model (Source: http://www.geo-

[2] Darcy's law states that $q=ki$, where q is the specific discharge, k is the hydraulic conductivity and i is the gradient of total hydraulic head (for more theory of SEEP/W, please refer to Appendix B

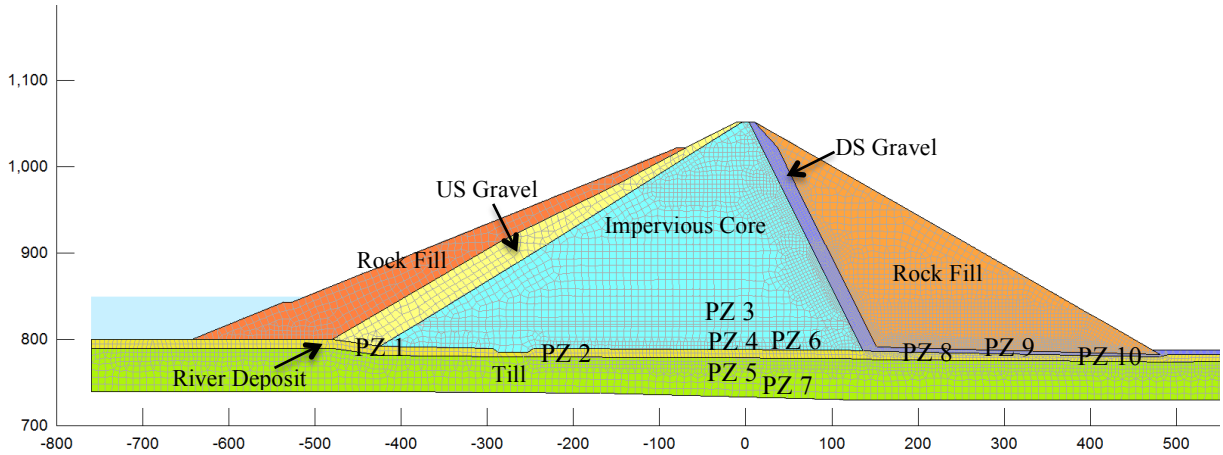

Figure 10. Screenshot of SEEP/W model with piezometer locations

Table 1. Properties of the Piezometers

Piezometer	X-coordinate (ft.)	Y-coordinate (ft.)	Material/Region	Hydraulic Conductivity (ft./sec)
PZ 1	-434.908	788.4537	River Deposit	0.007
PZ 2	-198.799	784.7021	River Deposit	0.007
PZ 3	-8.8836	818.8395	Impervious Core	8e-008
PZ 4	-8.8836	800.683	Impervious Core	8e-008
PZ 5	-8.9629	785.4531	River Deposit	0.007
PZ 6	71.1705	801.7607	Impervious Core	8e-008
PZ 7	63.24691	788.2472	River Deposit	0.007
PZ 8	200.6962	787.4987	DS Gravel	0.0002
PZ 9	300.8705	795.7294	Rockfill	0.002
PZ 10	415.5292	775.7503	River Deposit	0.007

3.1.2 Anomalous Scenarios

Out of many different internal erosion mechanisms possible, we only considered two that can be simulated using SEEP/W and possibly occur given the properties of the case study dam.

Case 1: Foundation Erosion

In most cases, internal erosion occurs from the toe area[3] and develops backwards towards an embankment. This is often caused by the constant migration of soil particles towards free exits or into coarse openings (Flores-Berrones and Patricia 2011). Once piping initiates and eventually develops further, the hydraulic conductivity of affected soil materials change due to such soil erosion. Such backward erosion piping can occur from either the foundation or embankment. In our work, we introduced anomalies in the downstream foundation of the case study dam, whose soil property is designated as river deposit. In order to simulate various scenarios given this mechanism, we first selected a thin layer within the river deposit and modeled 5 segments in the layer, in which various anomalies could be introduced. Severities and durations of the anomalous scenarios were varied as follows:

- Severity Low: original hydraulic conductivity times 2, 4, 6, 8, 10
- Severity High: original hydraulic conductivity times 10, 25, 50, 75, 100
- Duration Slow: three months for each change
- Duration Fast: one month for each change
- Uniformity Graded: 5 segments in the anomalous layer have different hydraulic conductivities
- Uniformity Uniform: 5 segments in the anomalous layer have a uniform hydraulic conductivity

[3] The junction of the downstream face of a dam with the natural ground surface (New Hampshire Department of Environmental Services 2011)

For example, if the anomalous scenario was based on 'Severity: High, Duration: Fast, and Uniformity: Graded', we first changed the hydraulic conductivity of the right most segment of the selected layer by 2 times the original hydraulic conductivity of the river deposit for three months. Then in the next three months, we increased the hydraulic conductivity of the second right most segment of the layer by 2 times, and the right most segment by 4 times, and so on (See Figure 11). Thus, this particular anomalous scenario spanned a total 15 months (3 months times 5 segments). In the uniform case, the hydraulic conductivities of all five segments were uniformly increased. That is, the hydraulic conductivities of all five segments were the same for three months or one month depending on Duration Slow or Fast, respectively. Thus, as summarized in Table 2, there were a total of 8 anomalous scenarios generated and simulated based on this particular mechanism.

Duration (months)					
1~3	original	original	original	original	2*original
4~6	original	original	original	2*original	4*original
7~9	original	original	2*original	4*original	6*original
9~12	original	2*original	4*original	6*original	8*original
12~15	2*original	4*original	6*original	8*original	10*original

Figure 11. Simulation method based on Severity: Low, Duration: Slow and Uniformity: Graded

Case 2: Poorly Compacted Layer (in Core)

Another anomalous scenario we simulated was internal erosion caused by poorly compacted zones (or high permeability zones) within an embankment core. When constructing an embankment dam, each soil layer is compacted on top of the previously compacted layers.

However, due to occasional freezing on seasonal shutdown, poor compaction, and many other reasons, highly permeable zones/ layers can be possibly formed within a core, foundation or even at abutment contacts. To simulate a similar problem, we created another thin layer along the core. Since the core had been compacted to be impervious, we increased its hydraulic conductivity more severely, i.e., on the order of 1, 2, 3 and 4 times the original magnitude (Severity: extreme-High), and applied only the 'Uniform' scenario to simulate a *pervious* layer. Table 2 also includes two anomalous scenarios generated from this type of mechanism.

Table 2. Anomalous Scenarios

Scenario	Location	Severity	Duration	Uniformity
1	River Deposit	Low	Slow	Graded
2	River Deposit	Low	Slow	Uniform
3	River Deposit	High	Slow	Graded
4	River Deposit	High	Slow	Uniform
5	River Deposit	Low	Fast	Graded
6	River Deposit	Low	Fast	Uniform
7	River Deposit	High	Fast	Graded
8	River Deposit	High	Fast	Uniform
9	Core	e-High	Slow	Uniform
10	Core	e-High	Fast	Uniform

Figure 12 shows the time series of one of the piezometers (PZ 4) based on various scenarios. The figure shows the data from Jan. 2009 to Apr. 2010 only to highlight the nature of the anomalous data in each scenario. We can see that piezometer PZ4 responded differently in the different scenarios. Figure 13 shows the time series of three different piezometers that were generated from Scenario 1. Even though the data were generated from the same anomalous scenario, it is clear that different piezometers behaved differently. Some anomalous data seem to involve only amplitude changes while some also involve shape changes, probably due to time lags caused by changes in seepage rates.

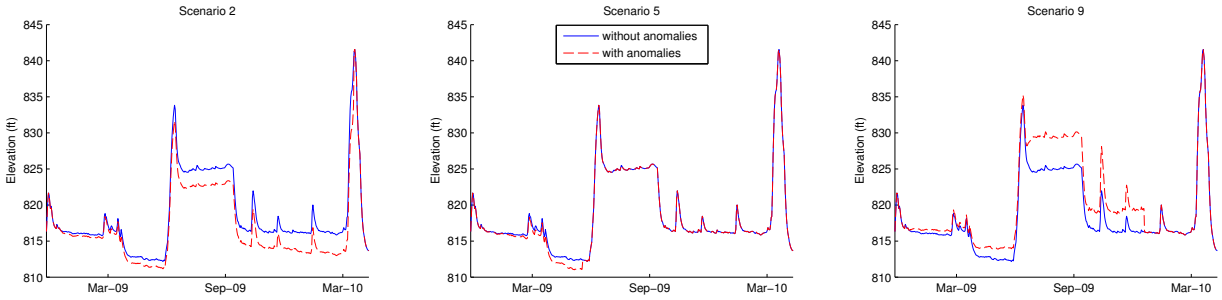

Figure 12. Parts of the simulated PZ 4 time series based on various scenarios

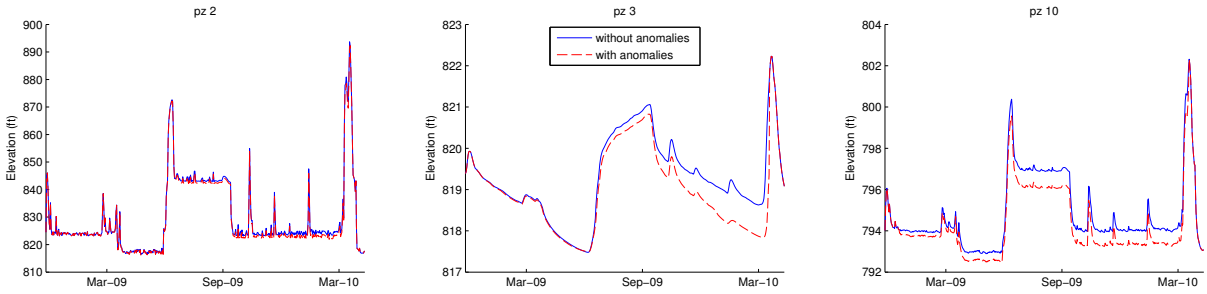

Figure 13. Parts of the simulated anomalous time series based on Scenario 1

3.1.3 Anomaly Detection Techniques

Our objective for the research is to examine several data analysis techniques to detect anomalous patterns in piezometer data that do not follow its previous historical behavior or do not behave as expected. Since the 19th century, anomaly detection has been widely studied by the statistics community, and there have been a variety of anomaly detection techniques developed for certain application domains and for more generic problems (Xu 2010). Anomaly detection techniques should be selected by considering if they are suitable for the data sets in terms of the correct distribution model, the correct attribute types, the scalability, the speed, etc. (Hodge and Austin 2004). In addition, selecting appropriate approaches depends on the data type, availability of data labels, how one wishes to detect anomalies and to handle them (Hodge and Austin 2004). In this

research, we carefully investigate how an anomaly should be defined in piezometer data analysis as well as the sensitivity required and the validation methods that should be used.

Piezometer data are collected at a regular frequency (e.g., every 4 hours, 12 hours, etc.), so temporal information should not be ignored but perceived as an important factor. Moreover, since piezometer data are highly affected by pool levels, seasonality present in data should be also considered. As mentioned above, anomalies are defined as data or patterns that deviate from the normal data. Assuming that the first few years of data are normal, we detect anomalies by looking for any unexpected changes in that data over time.

By considering the nature of the piezometer data, three types of anomaly detection techniques were selected to be studied as part of our research. The three techniques we evaluate are: Autoregression (AR), Moving Principal Component Analysis (MPCA), and Kullback-Leibler (KL) Divergence, which will be explained in details below. These three techniques can handle multi-dimensional data (pool and piezometer data) and be adjusted to retain temporal information. In addition, since they entail different characteristics, it would be possible to provide relevant insights to various anomalous scenarios as well as different piezometers. Even though AR is one of the most widely used time series techniques, there is a limitation in such a predictive technique, because accurately modeling time series with some noise present is challenging, especially given that AR only considers the linear dependence in a time series (Chandola et al. 2009b). Thus, we also tested techniques that initially extract features from the data (e.g., MPCA and KL), which are more robust to noise present in the data.

Given that piezometer data fluctuate based on pool levels, it is easy to assume that piezometer data are highly correlated to pool, thus they have linear relationships/trends in general. Moreover, especially when looking for any deviations from normal data, engineers tend to assume a normal (Gaussian) distribution. However, based on where piezometers are located along the dam, their measurements can be either linear/nonlinear or highly correlated/uncorrelated in response to pool levels as shown in Figure 14. This means that detecting any deviations with a Gaussian approach may not be appropriate. Thus, we tested both parametric and nonparametric techniques, MPCA and KL divergence, respectively, to see if the existing assumptions on linearity as well as Gaussian distribution are appropriate when analyzing piezometer data. In the next subsections, we introduce each model and technique in more detail.

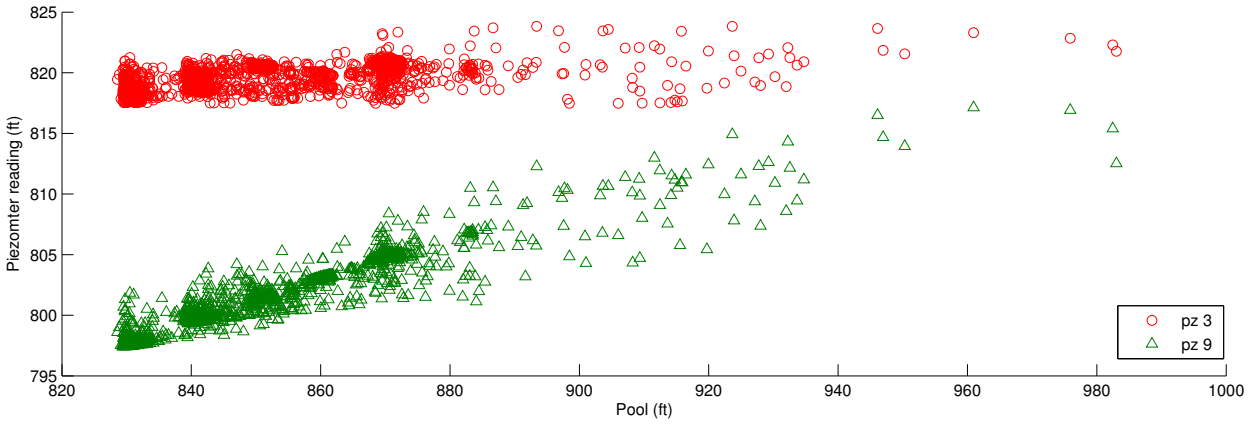

Figure 14. Pool vs. Piezometer

1. Autoregressive (AR) Model

One of the most widely used stochastic time series models is Autoregressive Integrated Moving Average (ARIMA) (Zhang 2003; Adhikari and Agrawal 2013). While ARIMA models have been mainly adopted in economic and statistics domains, they also have been popular for

structural health monitoring or damage detection (Nair et al. 2006; Omenzetter and Brownjohn 2006; Gul and Catbas 2011). ARIMA assumes that a time series dataset follows a normal distribution and has a linear form meaning that future values of the time series are estimated based on linear combination of previous values of the series. An ARIMA model can be divided into subclasses, such as Autoregressive (AR), Moving Average (MA), and Autoregressive Moving Average (ARMA) models (Adhikari and Agrawal 2013). Particularly, the AR model is expressed in terms of past values, MA model is expressed in term of past errors, and ARMA is a combination of AR and MA models.

In order to determine the most appropriate model given time series data, autocorrelation (ACF) and partial autocorrelation functions (PACF) should be computed. Autocorrelation measures the linear dependence of a variable at different lags while partial autocorrelation measures the linear dependence between different lags after removing the mutual autocorrelation in-between (controls intermediate autocorrelations). Equation (3) and (4) represent the ACF and PACF of a time series, $x_{t,}$ at lag j, respectively (Shumway and Stoffer 2011).

$$\rho_j = corr\left(x_{t,}x_{t-j}\right) = \frac{cov\,(x_t,x_{t-j})}{\sqrt{var(x_t)var(x_{t-j})}} \tag{3}$$

$$\phi_{jj} = corr(x_{t+j} - \hat{x}_{t+j}, x_t - \hat{x}_t),\ j \geq 2 \tag{4}$$

where $\hat{x}_{t+j}$ denotes the regression of x_{t+j} on $\{x_{t+j-1,}x_{t+j-2,\cdots,}x_{t+1}\}$

After examining these functions, either AR or MA is selected to appropriately model time series. For example, if a PACF plot of time series displays a sharp cutoff at a certain lag while ACF decays more slowly, the time series are said to have an AR signature given that AR considers

cumulative effects of all previous values. In this case, the order number is determined based on at which lag PACF cuts off. The opposite scenario applies for a MA model. In MA models, since only errors at certain lags (let's say, *p*) affect the current value, ACF will cutoff after lag *p* while PACF will decay more slowly. All of our piezometers have AR signatures; each ACF decays slowly while PACF cuts of sharply at a certain lag. So, we focus on developing AR models in this work.

Vector Autoregressive (VAR) Model

The Equation (1) shown in Section 2.4.3 was based on a univariate time series. However, since piezometer readings are highly affected by the pool levels, we should consider the relationship between the pool and piezometer readings together. Vector Auto Regressive model (VAR) is a multivariate time series model that can track many series at once instead of a single time series at a time. Since VAR jointly considers dynamics among multiple time series, it is possible to recognize how much one variable is influenced by the other variables. Equation (5) shows a VAR (*n*) model equation:

$$\begin{pmatrix} x_{1,t} \\ \vdots \\ x_{d,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_d \end{pmatrix} + \begin{pmatrix} a_{1,1}^1 & \cdots & a_{1,n}^1 \\ \vdots & \ddots & \vdots \\ a_{d,1}^1 & \cdots & a_{d,n}^1 \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ \vdots \\ x_{d,t-1} \end{pmatrix} + \cdots + \begin{pmatrix} a_{1,1}^n & \cdots & a_{1,n}^n \\ \vdots & \ddots & \vdots \\ a_{d,1}^n & \cdots & a_{d,n}^n \end{pmatrix} \begin{pmatrix} x_{1,t-n} \\ \vdots \\ x_{d,t-n} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{d,t} \end{pmatrix} \quad (5)$$

where $x_{d,t}$ denotes d^{th} time series at a time *t*, a^n's are AR coefficients that correspond to n^{th} lag, and ε_t is a vector of uncorrelated innovations.

2. Moving Principal Component Analysis (MPCA)

Please refer to Section 2.4, which include the theoretical explanation of the MPCA.

3. KL Divergence

A nonparametric technique that we tested is based on divergence measurement, which is the distance between two probability densities (Szabo 2014). One of the most well known divergence estimators is the Kullback-Leibler (KL) divergence (Kullback and Leibler 1951; Bishop 2007; Polani 2013). The KL divergence between *p(x)* and *q(x)* is expressed in Equation (6).

$$D(p,q) = \int_{\mathbb{R}^d} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx \tag{6}$$

When fitting parameterized models to the samples being tested, various principles can be used to estimate KL divergence (Wang et al. 2009; Basseville 2013), e.g., k-nearest neighbors, Maximum Likelihood Method, maximum spacing estimators, etc. In this work, we adapt the k-Nearest Neighbors (kNN) method (Wang et al. 2009) to estimate the KL divergence, or distance between the samples from two distributions, p and q.

First, the distance between each point in the d-dimensional samples, $\{X_{i=1,\ldots,n}\}$ of p to its k[th] nearest neighbor in $\{X_j\}_{j\neq i}$ are computed. Then, the k[th] nearest neighbor distance of each point from $\{X_i\}$ to the d-dimensional samples, $\{Y_{l=1,\ldots,m}\}$ of q are also computed. Those distances are denoted as $XX_k(i)$ and $XY_k(i)$, respectively. The divergence estimator is then expressed as

$$\widehat{D}(p,q) = \frac{d}{n}\sum_{i=1}^{n} \log\left(\frac{XY_k(i)}{XX_k(i)}\right) + \log\frac{m}{n-1} \tag{7}$$

In the next section, we will explain how we applied the three statistical techniques in order to improve anomaly detection of piezometer data.

3.2 Applications

As mentioned in Section 3.1.2, the locations of piezometers (including soil layers in which their tips are installed) as well as different anomalous scenarios influence piezometers responses, and consequently, anomaly detection performance. In this section, we present two experiments designed to evaluate three different anomaly detection techniques for the selected piezometers of the case study dam (Table 1), and for various anomalous scenarios (Table 2).

For every time series dataset, $X \in \mathbb{R}^{T \times D}$, where T is the length of the dataset, and D is the dimension of the data, we assumed that the first two years of the data were normal (i.e., training data), and the rest of the data were treated as testing data. In every technique, since piezometers respond to reservoir pool events, we analyze both the piezometer readings and the reservoir levels together. Thus, given 10 different anomalous scenarios and 10 piezometers, we test 100 2-dimensional datasets (each dataset is composed of two dimensional times series of pool and piezometer) total.

The metric we used to evaluate the anomaly detection techniques is the F_2 score, which is a combined measure of precision and recall, but emphasizes recall twice as much as precision (See Equation (8)). We selected this evaluation metric, because in the dam safety domain, missing anomalies is worse than raising false alarms. For each technique, we tested various parameters, i.e., order number of AR models, window sizes, threshold levels, etc. Once the optimized parameter settings were reviewed, we also examined a specific set of parameters that may be intuitive when analyzing piezometer data.

$$F_2 = (1 + 2^2)\frac{(precision * recall)}{(2^2 * precision) + (recall)} \quad (8)$$

We constructed a control chart for each technique, where threshold levels are established based on the features extracted from the normal data. Any deviations that go beyond certain sigma levels (standard deviations) of the threshold level are then marked as anomalies. 1 to 3 standard deviations are tested for each technique.

3.2.1 Vector Auto-Regressive Model (VAR)

In the VAR application, we used the standard least squares technique to determine VAR coefficients and AIC (Akaike 1998) to select the order of VAR models. For each dataset, VAR coefficients were estimated from the scaled training data, and residual errors of the testing data (which contain both normal and anomalous data) were calculated. Here, residual errors are the difference between actual and predicted values that were computed from the VAR model developed using the training data. Since reservoir data were given (or we do not need to predict), we computed the residual errors of the piezometer data only (the VAR model considers both of the previous reservoir and piezometer data to predict new piezometer data.)

3.2.2 Moving Principal Component Analysis (MPCA)

It is necessary to select a window size, *w* that is large enough to capture periodicity of the data and not be influenced by measurement noise (Laory et al. 2013; Posenato et al. 2008). Given our daily data, we tested window sizes that range from 365 to 400. We did not test shorter windows

in order to capture the entire seasonality (1 year) of the data, and did not test larger windows to make sure the numbers of anomalous windows did not exceed those of normal windows.

Once a w was selected, principal components of the data contained in the window, $W \in \mathbb{R}^{w \times D}$ were computed (in our case we obtained 2 eigenvectors given the 2 dimensional data). As W_i (a window that contains data from time i to $i + w$ -1) moved towards the end of the time series, we observed any changes in the most dominant pattern only, i.e., first principal component. Since our data is 2-dimensional data (i.e., D=2), there were two elements in each principal component. Thus, when observing how the first principal components changed over time, we looked for changes in the angles formed by the two elements.

When moving a window, we moved in 30 day increments instead of one day increments because i) the anomalous problems we are focusing on are those that are difficult to be visually detected, which means they are mostly slowly moving processes, not requiring anomaly detection of daily changes, and ii) any anomalies found in one day data were likely caused by noise, and the eigenvectors would not significantly change due to one additional day data. The threshold level, σ of the control chart was established using the mean +/- 1 to 3 standard deviations of the training data.

3.2.3 Kullback-Leibler Divergence (KL)

We applied the kNN method-based KL divergence using the Information Theoretical Estimators (ITE) Toolbox (Szabo 2014). We used a k value of 3. First, the divergence, d between every

window in a two-year training period, $W_i, i \in t$, where t =1, … 730-*w*, and the windows in the testing period, $W_j', j \in t'$, where t'=730-*w*+1, ... *T-w,* were computed. We then analyzed the median divergences from the testing windows to the training windows, $\tilde{d}\ (W_i \parallel W_j')$ to see if there were any significant deviations. We only considered median divergences to observe central tendency. For example, let's say there were 20 training windows*, W* and 50 testing windows, *W'*. Then, for each of the 50 *W'*, there would be a total 20 d's computed against 20 *W*. Then only the median of those 20 *d*'s was saved from each *W'*. As the window moves toward the end, a total 50 $\tilde{d}$'s were examined in order to detect if there were any significant increases. As in the MPCA experiment, we also skipped by increments of 30 days (i.e., the duration between the two successive moving windows is 30 days).

We set the threshold levels by using the mean of those median divergences computed from the training windows only, and the standard deviations among them. Any KL divergences that went beyond certain sigma levels (i.e., standard deviations) from the mean value were marked as anomalies. Given that it is more important to detect every anomaly rather than having false alarms, we used the metric of F_2 score to evaluate the performance. F_2 score is a combined measure of precision and recall, but weighs recall twice as much as precision, meaning it penalizes false negatives more than false positives. We performed the parametric analyses with various window sizes, and threshold levels. After reviewing the optimized parameter settings, we also examined a specific set of parameters that might be intuitive given the sensitivity required when analyzing piezometer data in real life.

In order to establish a threshold level, σ of the control chart, we first computed every possible distance among training windows. We used the mean of those entire distances being computed, and the standard deviations among the mean distances from each training window to the rest of the training windows. Various threshold levels were obtained from the mean plus three different sigma levels of the standard deviations.

3.3 Results

Table 3 presents the results obtained from VAR based on the best order numbers as discussed in Section 5 (please also see Figure 15, which is a heat map generated for this result). The threshold levels, σ and the model orders, M shown in the table correspond to the combination that brings the best F_2 scores. Most piezometers perform badly while Piezometer 2, 5 and 7 perform comparatively better in extremely severe scenarios (i.e., Scenario 3, 4, 7, 8). The performances on the scenarios that simulate pervious core layer are poor.

Table 4 shows the results of MPCA based on the optimal parameter settings (of w and σ) that bring the best F_2 scores. Table 5 summarizes the results of KL, also based on the optimal parameter combinations. Figure 16 and Figure 17 are the corresponding heat maps.

In general, the performance of MPCA applied on the graded scenarios is not as good as that applied on the uniform scenarios; the performance of each piezometer improves in Scenario 4 and 8 except Piezometer 10. The results of KL are better compared to MPCA. KL has consistent

performance among different piezometers. KL performs better on the slow scenarios (i.e., Scenario 1, 2, 3, 4, 9).

Table 3. Results of VAR based on Optimal Parameters

	Scenario 1			**Scenario 2**			**Scenario 3**			**Scenario 4**			**Scenario 5**		
PZ	M	σ	F_2	M	σ	F_2	M	σ	F_2	M	σ	F_2	M	σ	F_2
1	4	1	9%	4	1	12%	4	1	19%	4	1	63%	4	1	6%
2	5	1	33%	5	1	69%	5	1	70%	5	2	97%	5	1	26%
3	8	1	1%	8	1	2%	8	1	2%	8	1	2%	8	1	-
4	4	1	1%	4	2	1%	4	2	1%	4	1	2%	4	2	1%
5	5	1	29%	5	1	66%	5	1	68%	5	2	97%	5	1	22%
6	5	1	2%	5	1	2%	5	1	3%	5	1	5%	5	2	2%
7	5	1	27%	5	1	65%	5	1	66%	5	1	97%	5	1	17%
8	5	1	7%	5	1	11%	5	1	64%	5	1	82%	5	1	4%
9	5	1	4%	5	1	4%	5	1	48%	5	1	59%	5	1	2%
10	4	1	3%	4	1	3%	4	1	6%	4	1	6%	4	1	-
	Scenario 6			**Scenario 7**			**Scenario 8**			**Scenario 9**			**Scenario 10**		
PZ	M	σ	F_2	M	σ	F_2	M	σ	F_2	M	σ	F_2	M	σ	F_2
1	4	1	6%	4	1	7%	4	1	31%	4	1	6%	4	1	6%
2	5	1	60%	5	1	62%	5	2	92%	5	1	7%	5	1	6%
3	8	1	-	8	1	-	8	1	-	8	1	38%	8	1	31%
4	4	1	2%	4	2	3%	4	3	3%	4	1	3%	4	1	4%
5	5	1	71%	5	1	76%	5	2	91%	5	1	6%	5	1	4%
6	5	1	3%	5	1	4%	5	1	5%	5	1	6%	5	1	6%
7	5	1	55%	5	1	57%	5	2	92%	5	1	6%	5	1	3%
8	5	1	8%	5	1	49%	5	1	77%	5	1	3%	5	1	2%
9	5	1	2%	5	1	11%	5	1	18%	5	1	1%	5	1	-
10	4	1	2%	4	1	2%	4	1	4%	4	1	2%	4	1	-

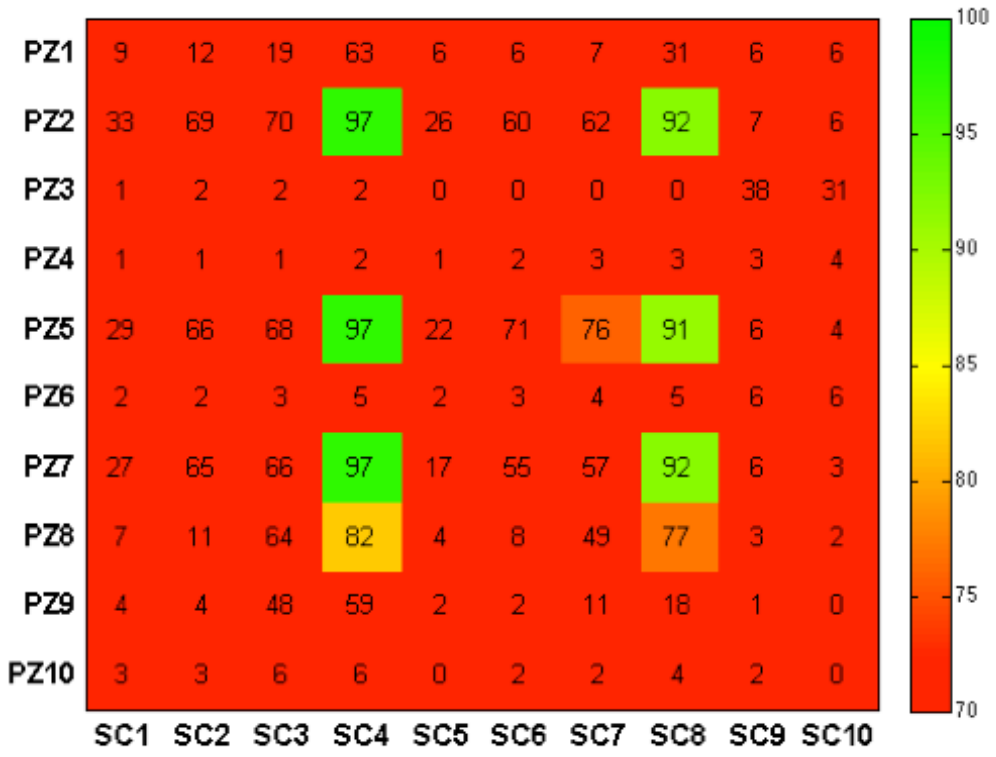

Figure 15. Heat map generated from the VAR results shown in Table 3

Table 4. Results of MPCA based on Optimal Parameters

	Scenario 1			**Scenario 2**			**Scenario 3**			**Scenario 4**			**Scenario 5**		
PZ	*w*	*σ*	F_2	*w*	*σ*	F_2	*w*	*σ*	F_2	*w*	*σ*	F_2	*w*	*σ*	F_2
1	365	1	72%	365	1	85%	365	1	90%	390	1	91%	365	1	84%
2	395	1	79%	365	1	85%	365	1	85%	380	3	94%	370	1	82%
3	365	1	72%	365	1	76%	395	1	76%	380	1	82%	390	1	57%
4	375	1	49%	395	1	69%	365	1	73%	365	1	84%	375	1	79%
5	365	1	76%	370	1	85%	365	1	88%	385	1	97%	370	1	82%
6	365	1	52%	365	1	77%	375	1	80%	400	1	82%	365	1	77%
7	365	1	76%	370	1	85%	365	1	88%	385	1	97%	370	1	82%
8	385	1	76%	380	1	82%	365	1	88%	380	3	94%	370	1	82%
9	385	1	73%	395	1	76%	370	1	87%	385	1	91%	370	1	82%
10	365	1	82%	395	1	54%	385	1	88%	380	1	63%	370	1	74%
	Scenario 6			**Scenario 7**			**Scenario 8**			**Scenario 9**			**Scenario 10**		
PZ	*w*	*σ*	F_2	*w*	*σ*	F_2	*w*	*σ*	F_2	*w*	*σ*	F_2	*w*	*σ*	F_2
1	365	1	84%	365	1	84%	375	3	91%	365	1	54%	365	1	74%
2	385	2	83%	385	2	83%	380	3	91%	380	1	71%	385	1	75%
3	380	1	52%	380	1	52%	365	1	67%	380	1	92%	370	1	92%
4	395	1	78%	395	1	83%	375	1	83%	380	1	84%	390	1	52%
5	390	3	91%	380	3	96%	390	2	91%	380	1	71%	380	1	74%
6	380	1	78%	375	1	83%	370	1	84%	375	1	85%	400	1	83%
7	370	1	82%	390	3	86%	390	2	91%	380	1	71%	380	1	74%
8	380	1	80%	370	1	82%	380	3	91%	365	1	73%	380	1	74%
9	370	1	78%	390	3	86%	380	2	87%	380	1	71%	380	1	74%
10	375	1	76%	370	1	82%	375	1	76%	380	1	71%	380	1	74%

Table 5. Results of KL based on Optimal Parameters

	Scenario 1			**Scenario 2**			**Scenario 3**			**Scenario 4**			**Scenario 5**		
PZ	*w*	*σ*	F_2	*w*	*σ*	F_2	*w*	*σ*	F_2	*w*	*σ*	F_2	*w*	*σ*	F_2
1	395	2	97%	395	2	97%	395	2	97%	395	2	97%	380	2	92%
2	400	2	97%	400	2	97%	400	2	97%	400	2	97%	370	2	92%
3	395	2	96%	385	3	97%	395	2	96%	395	3	97%	390	2	89%
4	395	1	96%	395	1	96%	395	1	96%	395	1	96%	395	1	90%
5	370	2	97%	385	3	97%	385	3	97%	385	3	97%	370	2	92%
6	395	1	96%	395	1	96%	395	1	96%	385	3	97%	370	2	90%
7	370	2	97%	385	3	97%	385	3	97%	385	3	97%	370	2	92%
8	370	2	97%	390	3	97%	390	3	97%	385	3	97%	370	2	92%
9	370	2	97%	370	2	97%	370	2	97%	390	3	97%	370	2	92%
10	395	2	97%	395	2	97%	390	3	97%	395	2	97%	370	2	92%
	Scenario 6			**Scenario 7**			**Scenario 8**			**Scenario 9**			**Scenario 10**		
PZ	*w*	*σ*	F_2	*w*	*σ*	F_2	*w*	*σ*	F_2	*w*	*σ*	F_2	*w*	*σ*	F_2
1	380	2	92%	380	2	92%	380	2	92%	395	1	75%	395	2	93%
2	370	2	92%	370	2	92%	370	2	92%	400	1	78%	400	2	93%
3	390	2	89%	390	2	89%	390	2	89%	380	3	94%	395	3	93%
4	395	1	90%	395	1	90%	395	1	90%	395	1	94%	395	2	91%
5	370	2	92%	370	2	92%	370	2	92%	395	1	81%	395	2	91%
6	370	2	90%	370	2	90%	370	2	90%	395	1	94%	395	2	91%
7	370	2	92%	370	2	92%	370	2	92%	395	1	81%	395	2	91%
8	370	2	92%	370	2	92%	370	2	92%	395	1	81%	395	2	91%
9	370	2	92%	370	2	92%	370	2	92%	400	1	81%	395	2	91%
10	370	2	92%	370	2	92%	370	2	92%	400	1	81%	395	2	92%

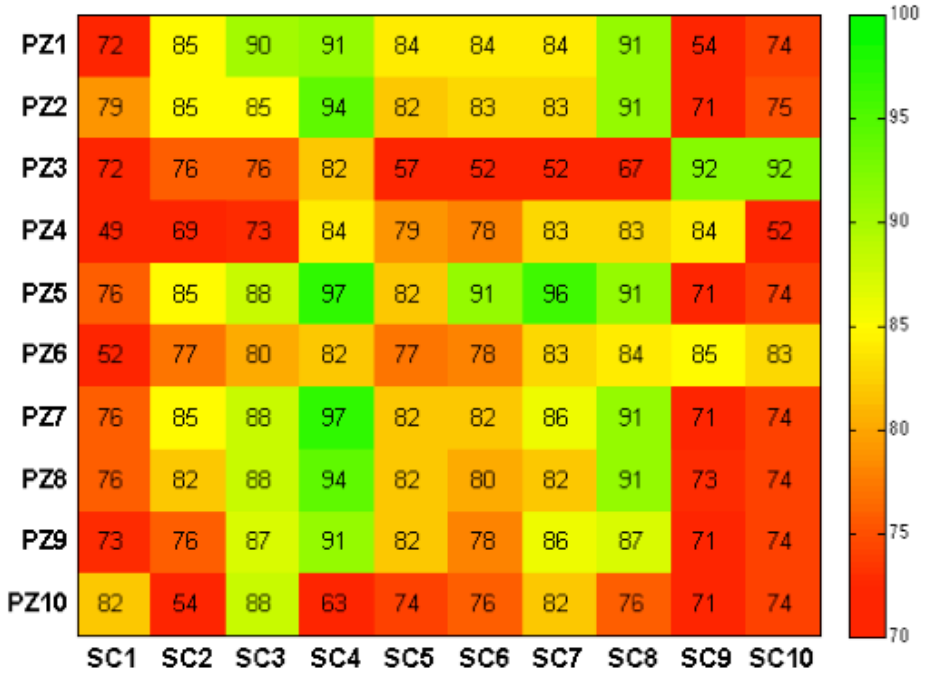

Figure 16. Heat map generated from the MPCA results shown in Table 4

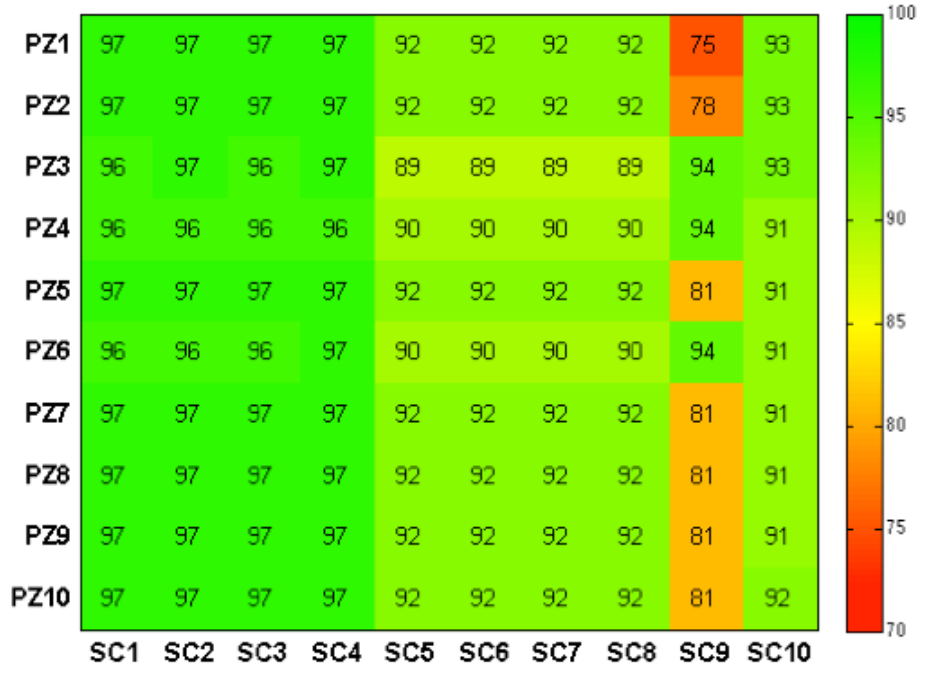

Figure 17. Heat map generated from the KL results shown in Table 5

In real life, it is not possible to determine the optimal parameter settings for anomaly detection especially given that various types of anomalies are not known in advance. Thus, we also tested parameters that may be intuitively selected by those who analyze the data to see how MPCA and KL perform. Given that it is fundamental to inspect annual condition changes, a w of one year and σ of 2 were chosen. However, sigma levels should be appropriately adjusted based on how sensitive your anomaly detection system should be. Table 6 (Figure 18) clearly shows that KL computed from each piezometer performs better than MPCA on every scenario.

Table 6. Results of MPCA and KL using a window size of 365 and a sigma level of 2

	Scenario 1		**Scenario 2**		**Scenario 3**		**Scenario 4**		**Scenario 5**	
PZ	MPCA	KL	MPCA	KL	MPCA	KL	MPCA	KL	MPCA	KL
1	36%	96%	74%	96%	77%	96%	86%	96%	74%	91%
2	60%	96%	79%	96%	79%	96%	91%	96%	72%	91%
3	32%	95%	50%	95%	50%	95%	67%	95%	25%	88%
4	8%	95%	32%	95%	50%	95%	69%	95%	62%	89%
5	63%	95%	76%	95%	76%	95%	94%	95%	72%	90%
6	28%	95%	60%	95%	67%	95%	79%	95%	63%	89%
7	63%	95%	79%	95%	76%	95%	94%	95%	72%	90%
8	57%	95%	73%	95%	82%	95%	88%	95%	71%	89%
9	53%	95%	63%	95%	85%	95%	88%	95%	67%	89%
10	53%	96%	24%	96%	76%	96%	28%	96%	66%	91%
	Scenario 6		**Scenario 7**		**Scenario 8**		**Scenario 9**		**Scenario 10**	
PZ	MPCA	KL	MPCA	KL	MPCA	KL	MPCA	KL	MPCA	KL
1	79%	91%	79%	91%	83%	91%	44%	71%	62%	89%
2	77%	91%	77%	91%	86%	91%	47%	71%	60%	89%
3	25%	88%	25%	88%	19%	88%	84%	92%	85%	87%
4	62%	89%	67%	89%	76%	89%	55%	76%	-	88%
5	86%	90%	90%	90%	86%	90%	47%	70%	60%	89%
6	63%	89%	67%	89%	77%	89%	66%	80%	55%	88%
7	77%	90%	77%	90%	86%	90%	47%	70%	60%	89%
8	76%	89%	76%	89%	81%	89%	47%	70%	60%	88%
9	72%	89%	77%	89%	82%	89%	47%	70%	60%	88%
10	61%	91%	75%	91%	71%	91%	47%	71%	64%	89%

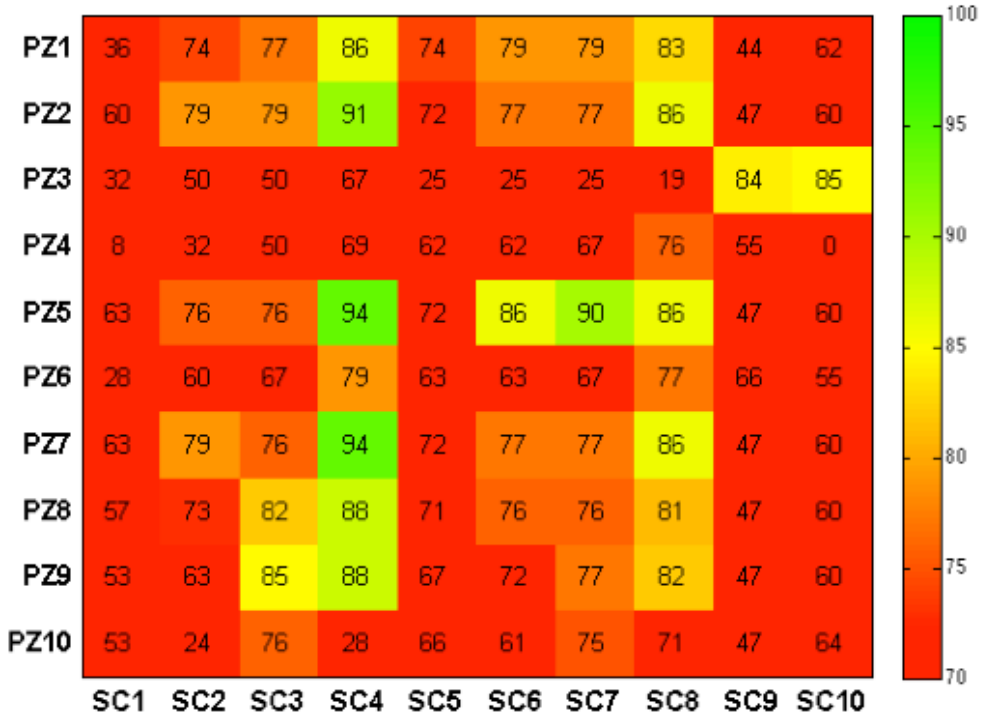

Figure 18. Heat map generated from the MPCA results shown in Table 6

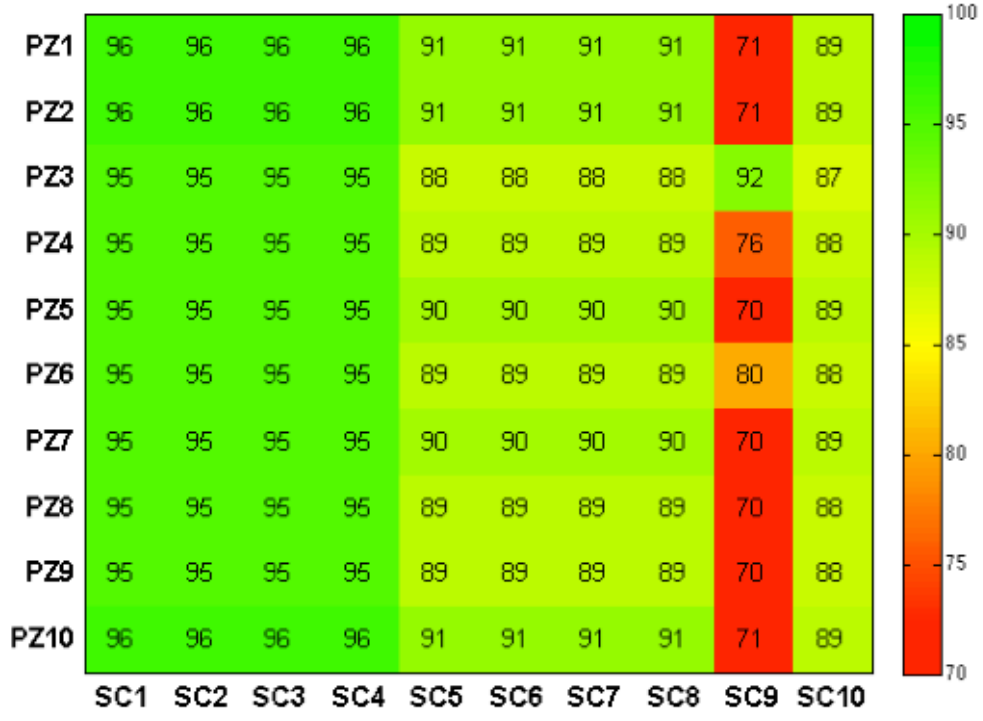

Figure 19. Heat map generated from the KL results shown in Table 6

3.4 Discussion

3.4.1 VAR

As summarized in Table 3 (and Figure 15), VAR performs very poorly in most piezometers and scenarios. It is worth noting that in Scenarios 1, 2, 3, 4 and 9, which are based on Duration: Slow, the anomalies are introduced from Jan. 2009 to Mar. 2010 while in Scenarios 5, 6, 7, 8, 10, which are based on Duration: Fast, the anomalies lie between Jan. 2009 and May 2009.

When VAR does perform well, the scenario is based on severe anomalies that are uniformly introduced (i.e., Scenario 4, 8) along the entire layer within the downstream deposit (not the gradual changes introduced). Figure 20 shows the entire time series (both training and testing data) of PZ 2 and the results of AR based on Scenario 4. Here, the regression residuals start to exceed the threshold line early enough to capture the anomalies introduced in the first 3 months (Jan. 2009 to Mar. 2009).

However, even based on the same scenario (Scenario 4), when the piezometric response changes are not clearly reflected enough (e.g., PZ 9), the result is not satisfactory as shown in Figure 21; the anomalies are detected late, thus leaving many false negatives.

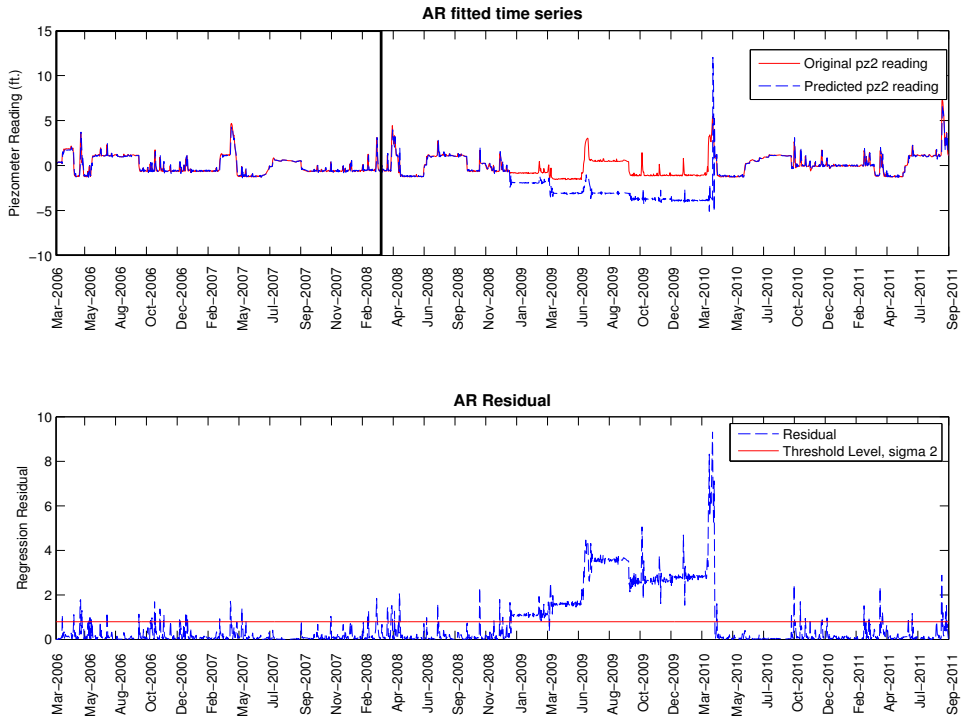

Figure 20. Result of AR based on Scenario 4 and PZ 2

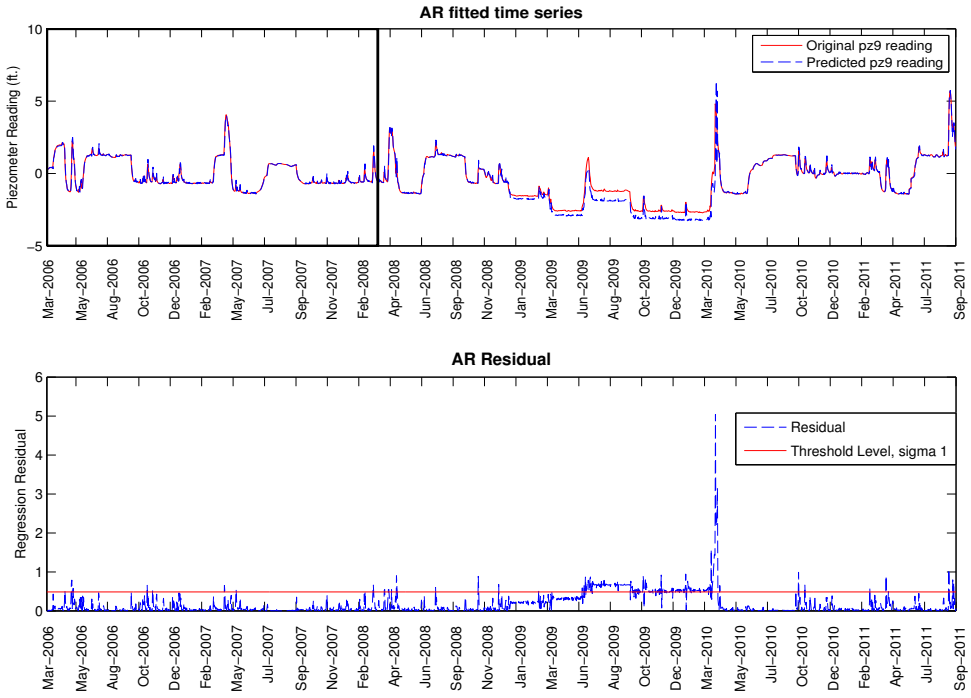

Figure 21. Result of AR based on Scenario 4 and PZ 9

In most cases, the poor performance of VAR is due to significant noise being present in the training data. For example, based on the PZ 1 data from Scenario 3 (Figure 22), we can visually see slight increases in the regression residuals during the anomalous period. However, due to a number of fluctuations in the residuals computed from the training data, the threshold level gets inadequately established, thus the anomalies cannot be accurately detected.

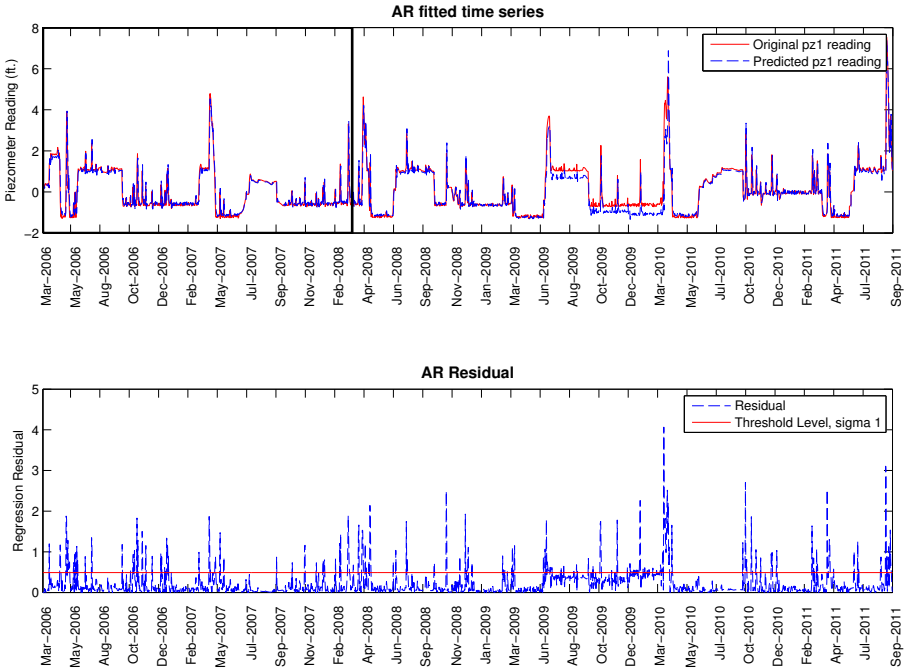

Figure 22. Result of AR based on Scenario 3 and PZ 1

3.4.2 MPCA vs. KL

Under the optimal parameter settings, MPCA performs better than VAR (Table 4 and Figure 16). There are differences in the performance over different scenarios and piezometers. Relatively, the piezometers installed in the impervious core area seem to perform poorly when anomalies are introduced in the downstream river deposit. However, their performance improves (even

outperforms compared to other piezometers) under the core anomalous scenarios (i.e., Scenario 9 and 10). MPCA seems incapable to capture less severe anomalies compared to severe ones as shown in the results based on Scenario 1 vs. Scenario 3 and Scenario 2 vs. Scenario 4. Accordingly, MPCA works better when the anomalies are uniformly distributed throughout the thin layer, which takes less time to affect further dam areas.

Compared to MPCA, KL performs better (Table 5 and Figure 17). It is clear that all 10 piezometers perform similarly in each scenario and over different piezometers. In general, the performances of KL on Scenarios 1- 4, which are based on Duration: Slow, are better than those on Scenario 5-8, which are based on Duration: Fast. When the results are closely analyzed, the lower F_2 scores on Scenario 5-8 are caused by more false positives (not necessarily more false negatives) compared to Scenario 1-4. Table 6 shows that under the realistic parameter setting, KL also performs better than MPCA.

In Figure 23, the top plot is a scatter plot of PZ 7 readings vs. pool levels. The second plot is the time series of PZ 7. In the time series, the piezometer data that lie in the anomalous period are plotted with plus (+) signs. The third and fourth plots correspond to the results of KL and MPCA, respectively, based on both normal scenario and Scenario 1 (dashed line). Here, the normal scenario means there are no anomalies introduced throughout the testing period. We show the results based on this normal condition in order to emphasize the difference in the results of the normal and anomalous scenarios (please note that these normal condition-based results are not the same as the threshold levels we used to test the performance of each anomaly detection technique). The KL divergence and the angle computed from MPCA using the data

contained in a particular window are indicated with asterisks. The range of this window is also indicated with side arrows in the second plot. The downward arrows in the third and fourth plots indicate the range of the anomalies where we expect to see certain increases in the KL divergence and MPCA angle. In this figure, we can clearly see that KL divergence starts to increase during the anomalous period while MPCA seems to fluctuate within the normal range.

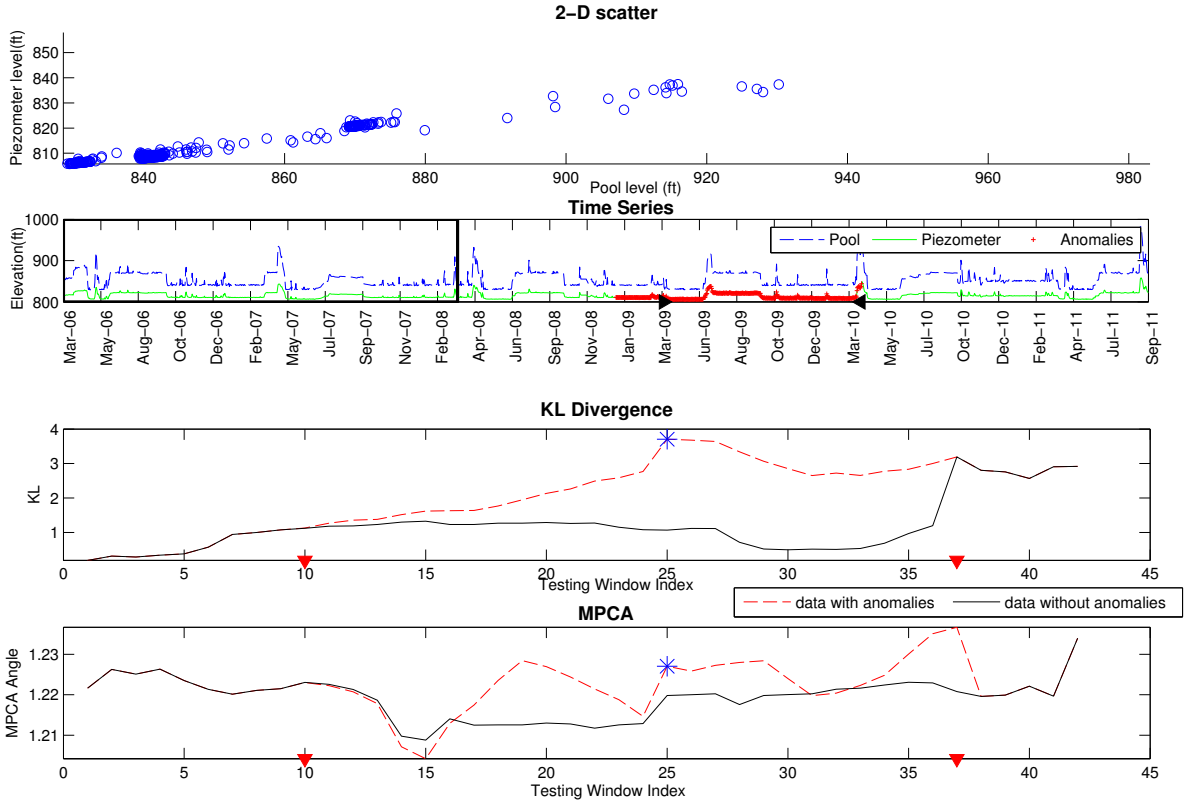

Figure 23. Results of KL and MPCA (window size 365, Scenario 1, PZ 7)

Figure 24 shows when both KL and MPCA perform well (Scenario 4). Both techniques show increasing trends within the anomalous period. The asterisks shown in the third and fourth plot correspond to the KL divergence and MPCA angles, respectively when the window contains

anomalous data only. As the window slides toward the end, it starts to include the normal data, and both KL and MPCA decrease slowly.

In each KL divergence plot, we see a sudden increase toward the end, even though such increase is not expected from the dataset without any anomalies. After closely looking at the data during that particular time period, we postulate that this increase may be caused by the unusually high pool level (pointed with a green arrow) as shown in the left plot of Figure 25. If the pool level suddenly becomes very high, there possibly can be some seepage-induced lags in the piezometer responses. Since KL picks up this change while MPCA does not, it indicates that KL is more sensitive in the task of piezometer anomaly detection.

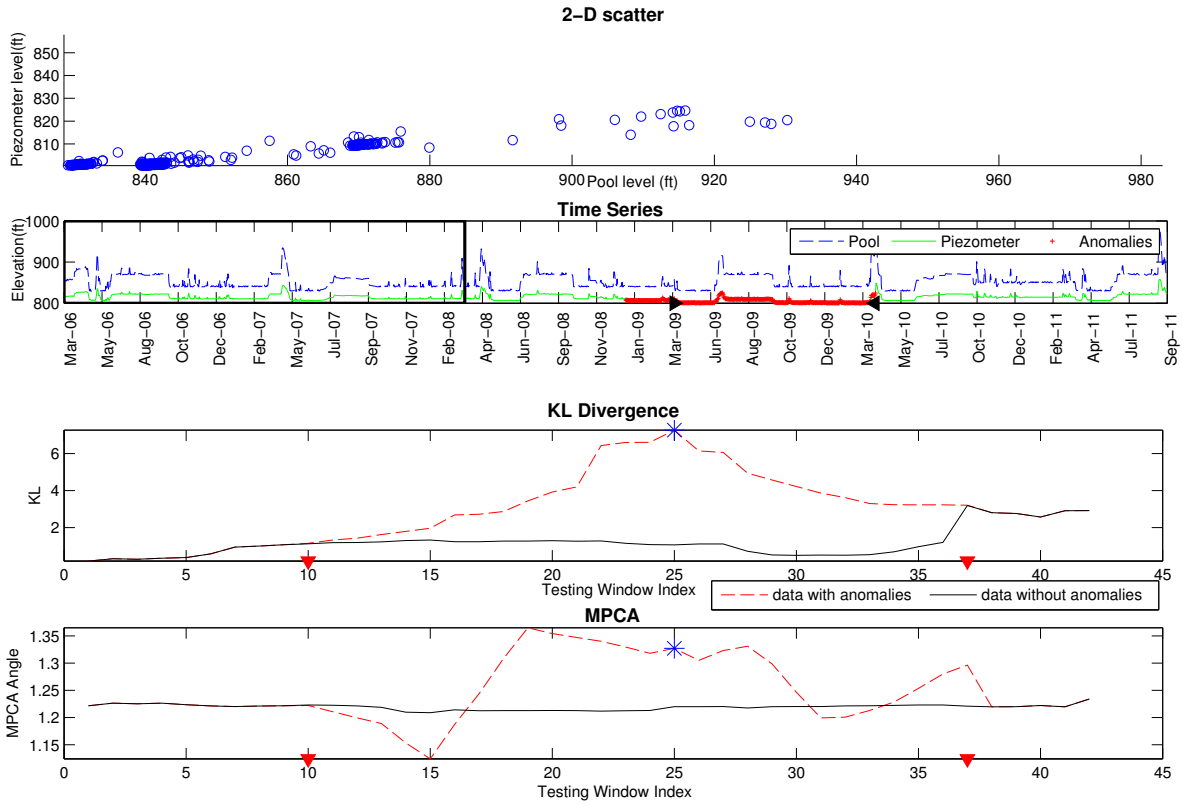

Figure 24. Results of KL and MPCA (window size 365, Scenario 4, PZ 7)

While MPCA sometimes show poor performance when applied to the upstream piezometers (i.e., PZ 1, 2, 3), we observe that KL performs well in most cases. As explained before, MPCA looks for changes in the angles formed from the eigenvectors, so in the case of when the anomalous piezometer data only reflect changes in the amplitudes, the angles are not much affected, thus such changes are hard to be detected.

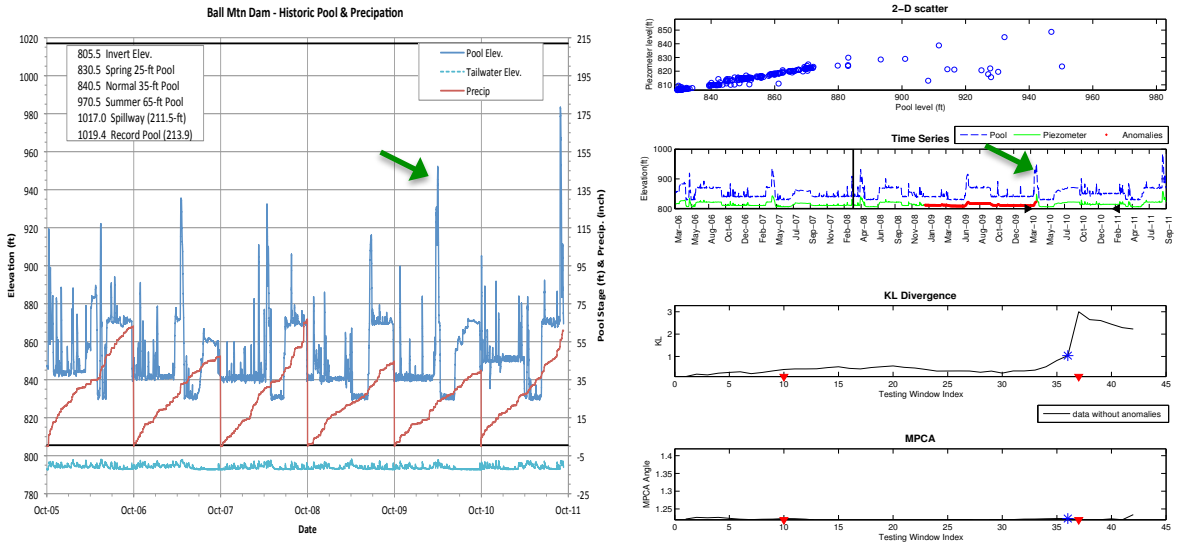

Figure 25. Pool information (left); Results of KL and MPCA (right)

Given that KL shows consistently strong results among every piezometer and anomalous scenarios, we can conclude that KL is more sensitive to various anomalies and is capable of handling a more wide range of piezometer locations. As mentioned above, KL is a nonparametric technique that does not make any assumptions on data distribution and linearity. Thus, from our experiment, we can see that such assumptions that may be easily made during piezometer data analysis are not always appropriate, and anomaly detection performance can be improved by not making such assumptions about the data.

CHAPTER 4. IMPACT OF ANOMALY DETECTION ON MULTIPLE PIEZOMETER DATA COLLECTED FROM EMBANKMENT DAMS

As shown in the last chapter, the detection performances of KL were more consistent with respect to various anomalous scenarios as well as different piezometers we analyzed. However, in the real world, piezometers in an embankment dam may not always be located at the optimum places, and their behaviors may not continue consistently over years. Thus, it is important to recognize how the data from a set of piezometers have changed over time as a group rather than evaluating only a single significant piezometer at a time. Since anomalies occurring inside dams may initiate with very small deformations without any visual signs, it is difficult to know exactly when and where the anomalies have initiated and would be located. However, by observing deviations among multiple (or grouped) piezometer readings over time, or the piezometers around a specific location, there is potential for obtaining better interpretations. Thus, in this chapter, the analyses of multiple piezometers are presented. The objective of the research described in this chapter is to determine if analyzing multiple piezometers together can improve interpretation and detection of piezometric anomalies. If the application of anomaly detection techniques using multiple piezometers can detect anomalies with better performance than analyzing individual piezometers, this will be a significant contribution to piezometer data analysis.

Even though our Kullback-Leibler (KL)-based method resulted in satisfying anomaly detection performance, we observed that the performance might often be highly affected by threshold levels, which need to be determined by users. Since patterns among piezometric responses would vary based on where they are located, it would often become uncertain to identify the impacts of various piezometer combinations on anomaly detection tasks using specified threshold levels. Thus, this chapter presents the new approach that combines KL with Support Vector Machine (SVM), which do not require prior knowledge about proper sensitivity levels.

The remainder of this chapter is organized as follows. In Section 4.1, we present our initial approach of analyzing multiple piezometers using the KL divergence technique and explain some of the weaknesses with this approach we have observed. Section 4.1.1 then describes our new approach, in which we combine KL and SVM in order to enhance anomaly detection. The applications and the experimental results are presented in Section 4.2 and 4.3, respectively. The chapter concludes with a discussion of the findings.

4.1 Approach

As an initial approach, we tested every pair of piezometers on the 10 anomalous scenarios. So, instead of a 2-dimensional dataset (i.e., pool and single piezometer reading), 3-dimensional data were analyzed (i.e., pool and two piezometer readings). In the previous research (Chapter 3), the final results were based on a window size, w of 365 days and a threshold level, σ of 2 standard deviations. Thus, we also used the same parameter setting for this experiment. The table below summarizes the results of the experiments in which the KL of the piezometer pairs improved

compared to when each of the pairs was analyzed by itself. The table only shows the cases where the performance from *both* of the piezometers improved. For example, in the case of (Scenario 5, Piezometer 2, Piezometer 5), the KL performance using Piezometer 2 and 5 together improved by 1 % and 2% compared to analyzing only Piezometer 2 and 5, respectively. As shown in Table 7, there are only 18 cases (out of 450 possible cases) where the piezometer pairs resulted in better performance; there are no dramatic improvements either.

Table 7. F_2 scores of KL based on the piezometer pairs (PZ A and PZ B) that improved the detection performance

Scenario	PZ A	PZ B	KL		
			PZ A	PZ B	PZ A+PZ B
5	2	5	91%	90%	92%
5	5	7	90%	90%	91%
5	8	9	89%	89%	90%
6	2	5	91%	90%	92%
6	5	7	90%	90%	91%
6	8	9	89%	89%	90%
7	2	5	91%	90%	92%
7	5	7	90%	90%	91%
7	8	9	89%	89%	90%
8	2	5	91%	90%	92%
8	5	7	90%	90%	91%
8	8	9	89%	89%	90%
9	2	5	71%	70%	71%
9	5	7	70%	70%	71%
9	8	9	70%	70%	70%
10	2	5	89%	89%	90%
10	5	7	89%	89%	89%
10	8	9	88%	88%	89%

We also tested various piezometer pairs by using a different σ (i.e., standard deviation of 3). This time, there were more cases (119 cases) where the piezometers pairs improved the performance

although only 7 cases improved by more than 5%. We observed that the performances were highly affected by the threshold levels, thus the impacts of additional piezometers could not be properly acknowledged. In addition to different levels of sensitivity, the results might have also been heavily influenced by the training dataset, which in turn was used for establishing threshold levels. Moreover, as expressed in Equation (6) and (7), KL measures the divergence between two distributions; however, with additional dimensions (i.e., additional piezometers), significant piezometric responses may often be distracted by noisy features.

In order to make our performance more reliable, we applied one popular classification technique called Support Vector Machine (SVM). SVM is known to perform well even when applied to data outside the training period. SVM tries to solve classification problems that can be generalized in unseen testing data. SVM is based on the principle of Structural Risk Minimization (SRM) instead of the traditional principle of Empirical Risk Minimization (ERM), which minimizes the errors on the training data, thus often suffer from overfitting. SRM, on the other hand, minimizes an upper bound on the expected risk (El-Naqa et al. 2002). It focuses on the training data that are most difficult to classify. In the next section, we further provide principles of SVM and introduce our new anomaly detection approach, which is based on KL and SVM.

4.1.1 New Anomaly Detection Approach: KL+SVM

The Support Vector Machine approach is a binary classifier; it finds the best hyperplane (or one with the largest margin between the two classes) that separates two different classes (Figure 26).

While there can be an infinite number of hyperplanes that can separate the two classes, a hyperplane that is very close to the training examples can be sensitive to noise, thus not generalize well in unseen testing data. However, with the hyperplane that maximizes the margin between the classes, or the hyperplane that is farthest from all training examples, SVM can generalize better.

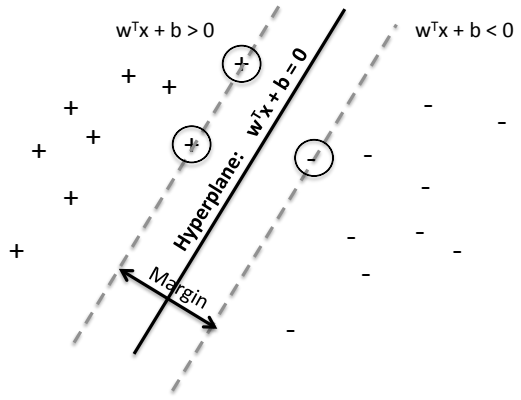

Figure 26. SVM classification showing a hyperplane that separates '+' class from '-' class with maximum separating margin.

In the case where data cannot be completely separated, SVM uses a soft margin that separates many data, but not all, meaning it allows a minimal number of errors (Cortes and Vapnik 1995). Given *n* training examples, the objective function of such SVM is (Vapnik 1995):

$$\min_{w,b,\xi} \frac{1}{2} \| w \|^2 + C \sum_{i=1} \xi_i$$

subject to (9)

$$y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

where $i = 1, 2, \ldots, n$, *w* is a normal vector that is perpendicular to the hyperplane, C is a penalty variable and ξ_i are positive slack variables, or training errors.

Sometimes, high dimensional data cannot be separated with a simple linear hyperplane. In such cases, kernel functions can be used to transform the input space into a feature space, so that SVM can produce nonlinear boundaries by constructing a linear boundary in the transformed feature space (James et al. 2013).

$$K(x, y) =< \phi(x), \phi(y) > \tag{10}$$

In our experiments, we use a Gaussian radial basis function to train the SVM classifier:

$$K(x, y) = \exp(\frac{\|x-y\|^2}{2\sigma^2}) \tag{11}$$

In the next section, we present how we applied SVM on the KL divergence using various combinations of piezometers.

4.2 Applications

KL

We adopt the same KL-based application approach as presented in Chapter 3.2.3. By using a window size, *w* of 365, we moved the window by increments of 10 days instead of 30 days in order to obtain a sufficient number of training and testing data for the classification task later. We applied kNN-based KL divergence using the Information Theoretical Estimators (ITE) Toolbox (Szabo 2014). We used a value of k of 3. We first computed the KL divergence, *d* among the windows that lie within the 2-year normal period and saved only the median divergence of each window, i.e., $\tilde{d}\,(W_i \parallel W_i)$. Then, *d* between each of the subsequent

windows was computed against every training window, and the corresponding medians were saved, i.e., $\tilde{d}\ (W_i \parallel W_j')$.

SVM

After computing the KL divergence of each piezometer based on each scenario, we used the KL divergence to train SVM classifiers (classify either 0 (normal) or 1 (abnormal)). In order to test/classify each scenario, we performed leave-one-out cross validation. For example, to test Scenario 1, we trained on the 9 other scenarios (Scenario 2-10) and to test Scenario 2, we trained on Scenario1 and Scenario 3-10, and so on. In order to minimize possible over-fitting, we also included more training datasets: i) KL divergence computed from the normal scenarios (where we did not introduce any anomalies throughout the testing periods), and ii) the simulated datasets (from all of the scenarios) with white Gaussian noise added. For these additional training datasets, we did not leave any one scenario out but used all of them to train, because they are different from the originally simulated datasets.

We used individual piezometers as well as multiple piezometers to train an SVM classifier for each scenario. Throughout the research described in this paper, this approach means either a one-dimensional feature (i.e., KL divergence) vector or higher dimensional feature vector was used to train an SVM classifier for each scenario. That is, the training data, x_i in Equation (6) are now either i) a one-dimensional KL divergence, *D* based on a single piezometer or ii) a multivariate *D,* corresponding to one of the multiple piezometers. For the latter case, we considered various feature vectors (each feature consists of the KL divergence computed from a single piezometer)

to see if they can improve the classification performance compared to when using individual piezometers only. We took three approaches to train the SVM classifier that is based on multiple piezometers (for each scenario):

1) Start by training on a two-dimensional dataset using a pair of the piezometers (e.g., piezometer 1 and of 2) and increase the dimension of the training dataset by adding each of the other piezometers at a time until it consists of all of the 10 piezometers (i.e., a 10-dimensional dataset).

2) Remove one piezometer at a time and observe how the performances change accordingly.

3) Train on distinct groups of piezometers, which are grouped by various characteristics (e.g., piezometer locations, elevations, soil materials, x-coordinates, etc.).

The results of these three steps are presented in the following section.

4.3 Results

We trained SVM classifiers using a Gaussian kernel (Equation (11)). Table 8 shows the F_2 scores when the SVM classifier for each scenario was trained by individual piezometers. Figure 27 shows a heat map generated from this result. Every piezometer has distinctive results on every scenario without clear patterns. For the *fast* anomalous scenarios that are based on the foundation erosion mechanisms occurring at the downstream river deposit (i.e., Scenario 5 - 8), most

piezometers except Piezometer 3 worked consistently while their results varied in the other scenarios. For the anomalous scenarios caused by the poorly compacted layer within the core (i.e., Scenario 9-10), Piezometers 3, 4 and 6 performed comparatively better, but every piezometer performed poorly in general.

Table 8. F_2 scores of KL+SVM based on individual piezometers

	PZ 1	PZ 2	PZ 3	PZ 4	PZ 5	PZ 6	PZ 7	PZ 8	PZ 9	PZ 10
SC1	78%	73%	44%	83%	58%	82%	58%	67%	58%	77%
SC2	93%	67%	75%	63%	90%	66%	93%	89%	66%	93%
SC3	82%	88%	65%	54%	93%	78%	92%	93%	93%	76%
SC4	63%	87%	61%	81%	87%	90%	88%	87%	91%	84%
SC5	69%	81%	46%	70%	82%	71%	82%	85%	75%	79%
SC6	80%	85%	71%	82%	88%	82%	87%	87%	75%	71%
SC7	85%	85%	71%	82%	88%	82%	88%	87%	76%	83%
SC8	87%	86%	80%	84%	88%	85%	88%	87%	78%	79%
SC9	7%	7%	51%	65%	15%	55%	10%	16%	-	5%
SC10	14%	-	82%	62%	-	66%	-	-	-	-

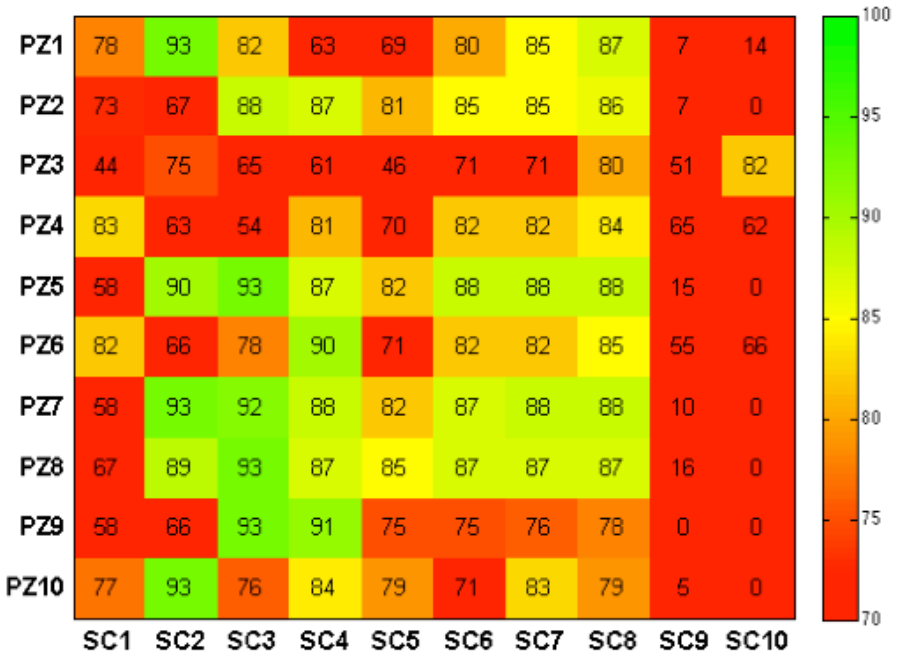

Figure 27. Heat map generated from the results shown in Table 8

Next, we trained the SVM classifiers using multiple piezometers. For this particular experiment, we increased the dimension of the training data by including one additional piezometer at a time. Table 9 summarizes the results. The first row shows which piezometers were used together to train a SVM classifier for each scenario. For example, 1-3 means the training dataset was comprised of the KL divergences of Piezometer 1, 2 and 3. Table 9 also includes average scores for the two anomalous cases, i.e., foundation erosion in the downstream (i.e., Scenario 1-8) and the pervious core layer (Scenario 9-10).

Table 9. F_2 scores of KL+SVM using multiple piezometers (starting from Piezometer1)

Piezometer / Scenario	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9	1-10
1	94%	92%	93%	95%	95%	95%	96%	96%	96%
2	96%	95%	97%	97%	98%	98%	97%	97%	97%
3	96%	95%	97%	98%	98%	97%	97%	97%	97%
4	96%	93%	93%	93%	93%	93%	93%	94%	94%
5	87%	82%	86%	89%	89%	89%	89%	89%	89%
6	91%	87%	92%	92%	92%	92%	92%	91%	89%
7	91%	87%	91%	91%	92%	92%	92%	91%	91%
8	89%	86%	86%	84%	84%	84%	84%	84%	86%
9	18%	85%	85%	85%	85%	85%	85%	84%	84%
10	5%	81%	83%	81%	81%	81%	81%	81%	81%
Ave. (1-8)	**92%**	**90%**	**92%**	**92%**	**93%**	**93%**	**93%**	**92%**	**92%**
Ave. (9-10)	**12%**	**83%**	**84%**	**83%**	**83%**	**83%**	**83%**	**83%**	**83%**

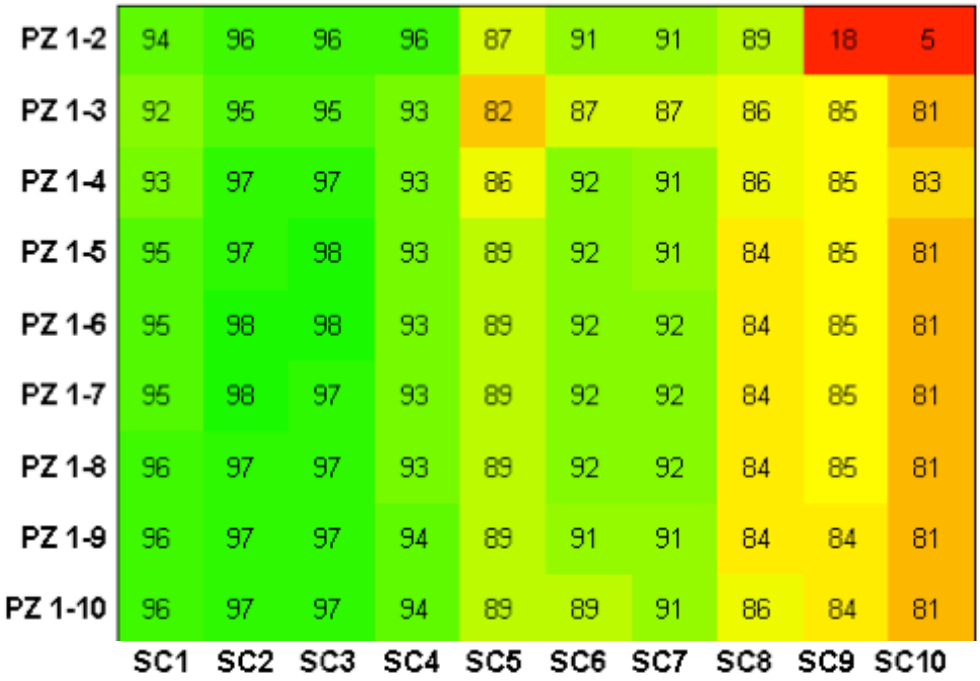

Figure 28. Heat map generated from the results shown in Table 9

As shown in Table 9 (and Figure 28), the performance on the foundation erosion scenarios improved with the average F_2 scores ranging from 90% to 93%. Even for the pervious core scenarios, we can clearly see the improvements throughout the various piezometer combinations.

Next, in order to recognize the contribution from each piezometer, we continued the experiments by removing one piezometer at a time and observed how the performances changed accordingly. Table 10 (see also Figure 29) summarizes the results, but this time, starting from Piezometer 2, instead of Piezometer 1. In general, the average scores of the foundation erosion scenarios did not change much, but the average scores for the pervious core scenarios got lower. Any of the combinations did not reach 80%.

Table 10. F_2 scores of KL+SVM using multiple piezometers (starting from Piezometer 2)

Piezometer / Scenario	2-3	2-4	2-5	2-6	2-7	2-8	2-9	2-10
1	86%	89%	93%	93%	95%	96%	96%	96%
2	86%	91%	96%	96%	96%	96%	96%	96%
3	95%	96%	96%	96%	96%	96%	96%	96%
4	95%	95%	96%	96%	96%	96%	96%	96%
5	80%	82%	84%	84%	87%	87%	87%	87%
6	84%	89%	89%	89%	89%	89%	89%	89%
7	85%	89%	89%	89%	89%	89%	89%	89%
8	85%	89%	87%	87%	87%	87%	87%	89%
9	60%	67%	67%	65%	72%	73%	76%	77%
10	75%	70%	76%	66%	68%	72%	70%	78%
Ave. (1-8)	**87%**	**90%**	**91%**	**91%**	**92%**	**92%**	**92%**	**92%**
Ave. (9-10)	**68%**	**68%**	**71%**	**66%**	**70%**	**73%**	**73%**	**77%**

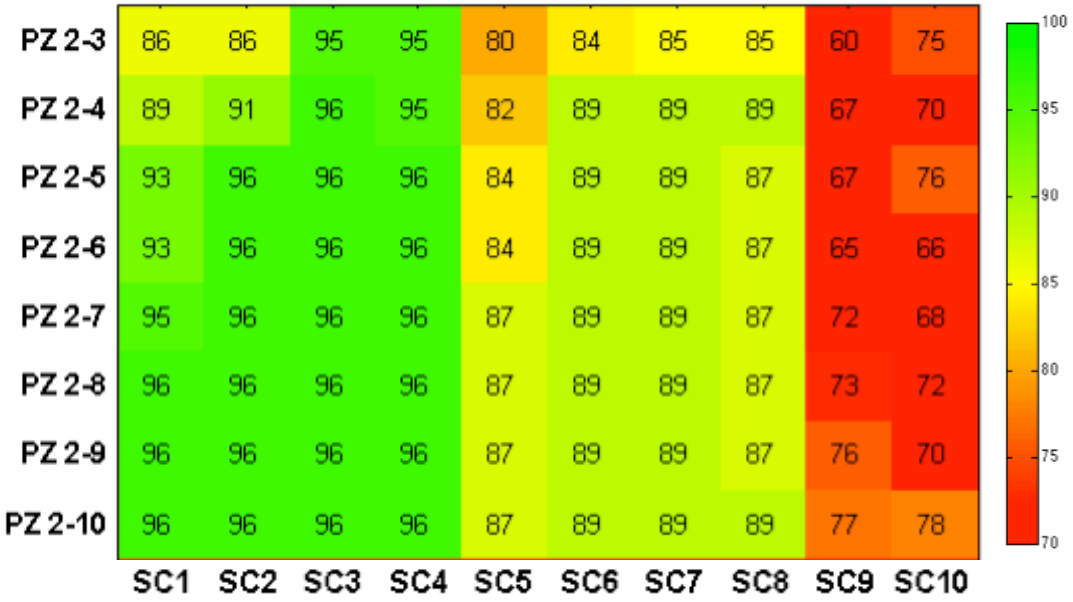

Figure 29. Heat map generated from the results shown in Table 10

In Table 11, in order to highlight how certain piezometers contribute to the classification performance, we only show the average F_2 scores over the two anomalous cases: the foundation

erosion (first 8 scenarios) and the pervious core layer (last two scenarios). Compared to the average scores shown in the last two tables, which include the first two piezometers, Table 11 shows that the average scores does not improve further, especially for the last two scenarios. The average F_2 scores on the first eight scenarios also does not reach 90% once the first four piezometers are removed.

Table 11. Average F_2 scores of KL+SVM using multiple piezometers (starting from Piezometer 3)

	3-4	**3-5**	**3-6**	**3-7**	**3-8**	**3-9**	**3-10**
Ave. (1-8)	78%	91%	91%	92%	92%	92%	92%
Ave. (9-10)	64%	62%	64%	64%	66%	66%	71%
		4-5	**4-6**	**4-7**	**4-8**	**4-9**	**4-10**
Ave. (1-8)		88%	88%	89%	89%	89%	90%
Ave. (9-10)		29%	49%	52%	55%	57%	64%
			5-6	**5-7**	**5-8**	**5-9**	**5-10**
Ave. (1-8)			88%	88%	88%	88%	89%
Ave. (9-10)			35%	29%	41%	42%	54%
				6-7	**6-8**	**6-9**	**6-10**
Ave. (1-8)				87%	88%	88%	89%
Ave. (9-10)				44%	30%	41%	53%
					7-8	**7-9**	**7-10**
Ave. (1-8)					84%	87%	90%
Ave. (9-10)					7%	7%	8%
						8-9	**8-10**
Ave. (1-8)						87%	90%
Ave. (9-10)						7%	10%
							9-10
Ave. (1-8)							89%
Ave. (9-10)							7%

Past behavior of piezometers may not continue indefinitely into the future, so it is important to recognize how piezometer readings have changed over time as a group (or interacting responses among them). Internal erosion has chaotic tendencies, so it is difficult to know exactly when and

where it will occur. However, by observing deviations among the grouped piezometers over time, or the piezometers around a specific piezometer, there is potential that better interpretations about anomalies can be obtained. Thus, we first created several groups of piezometers by considering various geotechnical characteristics, i.e., tip elevations, soil materials, and lateral locations, given that they may share common piezometric responses. In this way, any deviations observed within a specific group maybe useful to infer potential locations of anomalies. Table 12 and Table 13 below show the piezometer groups we selected and their corresponding results, respectively. The first row of Table 12 indicates the characteristics we considered when grouping the piezometers. For example Piezometers in G1 are installed in the foundation while Piezometers in G2 are installed in the core.

The F_2 scores of Table 13 show that some of the average scores are higher than the maximum scores (i.e., 93%) that resulted from the various piezometer combinations presented in Table 9-11. The scores on the pervious core scenarios are also higher in general; especially, G5 scored 90% and 87%, which are the best scores among the other defined groups as well as previous combinations we have tested.

Table 12. Defined groups of piezometers

Soil material	**Tip elevation (Based on mean value: 792 ft.)**	**Upstream vs. Downstream**	**X-coordinate**	
G1: Foundation (1,2,5,7,10)	G3: lower than 792 ft. (1,2,5,7,8,10)	G5: U/S (1,2,3,4,5)	G7: (1,2)	G9: (6,7)
G2: Core (3,4,6)	G4: higher than 792 ft. (3,4,6,9)	G6: D/S (6,7,8,9,10)	G8: (3,4,5)	G10: (8,9,10)

Table 13. F_2 scores based on the defined piezometer groups

Piezometer / Scenario	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
1	96%	90%	97%	97%	97%	96%	96%	96%	92%	90%
2	98%	95%	98%	98%	98%	96%	98%	98%	96%	98%
3	98%	94%	98%	98%	99%	96%	98%	98%	96%	95%
4	98%	96%	98%	98%	96%	97%	98%	98%	97%	98%
5	93%	76%	93%	89%	93%	87%	92%	88%	84%	91%
6	94%	86%	94%	90%	95%	88%	94%	91%	86%	91%
7	93%	85%	93%	89%	94%	90%	94%	92%	86%	91%
8	93%	88%	93%	90%	89%	91%	93%	91%	89%	92%
9	32%	68%	28%	71%	90%	67%	26%	69%	56%	22%
10	8%	76%	8%	75%	87%	61%	8%	75%	54%	8%
Ave. (1-8)	**95%**	**89%**	**95%**	**93%**	**95%**	**93%**	**95%**	**94%**	**91%**	**93%**
Ave. (9-10)	**20%**	**72%**	**18%**	**73%**	**89%**	**64%**	**17%**	**72%**	**55%**	**15%**

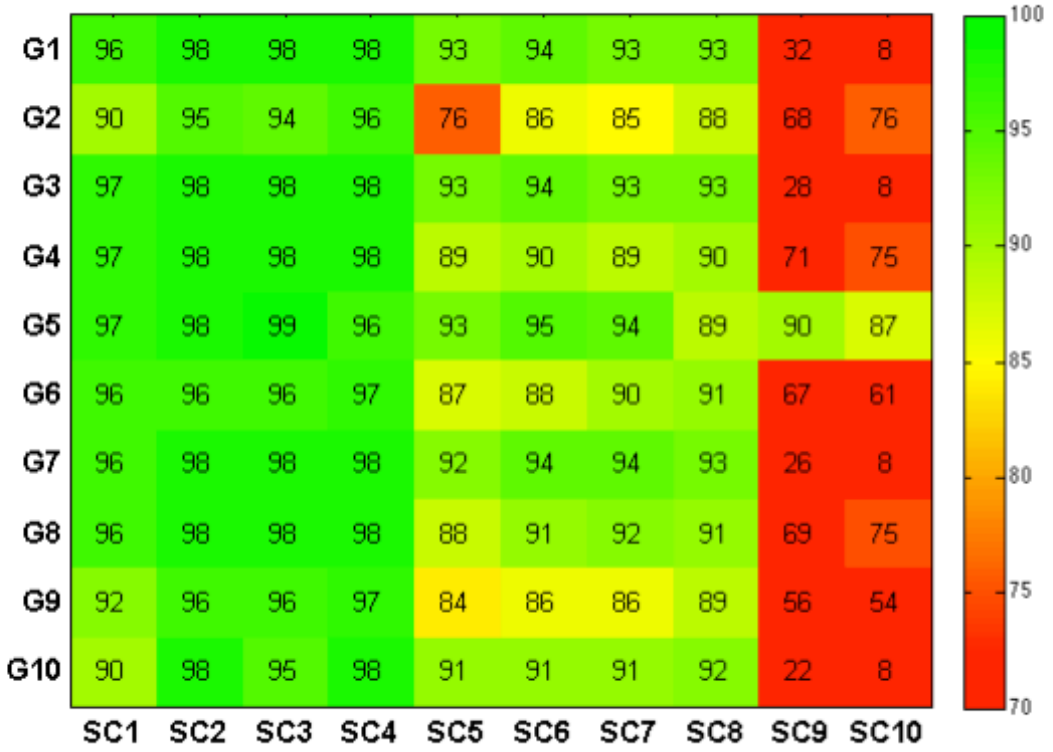

Figure 30. Heat map generated from the results shown in Table 13

4.4 Discussion

When we analyzed individual piezometers separately (Table 8), Piezometers 5, 7, and 8 performed the best with the average F_2 scores of 84-85% on the foundation erosion scenarios. This result was expected given that these particular piezometers are located near the anomalous foundation layer (Piezometer 6 is also located close to the anomalous foundation layer, but is installed within the core area with higher tip elevation). For the pervious core scenarios, however, Piezometers 5, 7 and 8 do not seem to be enough by themselves to classify anomalies. Instead, Piezometers 3, 4, and 6 work comparatively better, probably due to their locations within the core.

In Table 9 (Figure 28), it is interesting to see that the overall classification performance improved with additional piezometers. However, when Piezometer 3 (installed in the core) was added, the F_2 scores of the foundation erosion scenarios became lower, which means it did not provide any useful information, but probably introduced disturbing features. Piezometer 3, however, helped classifying the pervious core anomalies; the F_2 scores improved by at least 67%. Other core piezometers, i.e., Piezometer 4 and 6 also seemed to be more useful compared to the piezometers installed in the foundation (please see Table 14 below).

Table 14. KL+SVM on Piezometer 1 and 2 with core or foundation piezometers

	PZ 1,2 + Core Piezometer			**PZ 1,2 + Foundation Piezometer**	
	1,2,3	**1,2,4**	**1,2,6**	**1,2,5**	**1,2,7**
Sc. 9	85%	67%	60%	37%	31%
Sc. 10	81%	55%	59%	3%	3%

The F_2 scores in Table 9 and 10 show that by analyzing multiple piezometers together, the classification performances become more consistent among the 10 scenarios, meaning more reliable to cover various anomalies occurring at different locations. The first combination (i.e., Piezometer 2-3) in Table 10 does not perform better without Piezometer 1; however, once certain number of piezometers is added, the performance starts to stabilize and improve as with Piezometer 1.

Based on our experiment, it seems the combination of Piezometers 1 and 2 is not significant in classifying the pervious core scenarios. It is probably due to their locations with respect to the anomalous area (Piezometer 1 and 2 are located in the upstream); the seepage path may not be affected much until it meets or passes the core area. On the other hand, although each of Piezometer 3 and 4 performs better in classifying the pervious core scenarios as shown in Table 8, when they are combined together, the performance does not improve. Interestingly, when Piezometer 1 and 2 are combined with Piezometer 3 and 4 together, the performance improves. Here, we can observe that the piezometers with good performance do not always promise improvement as a group, but rather, poorly performing piezometers can often provide additional information regardless of their locations, thus improving the classification performance.

Table 13 also supports this observation. It is easy to assume that the core area piezometers (G2) or the downstream piezometers (G6, G9, G10) would perform better in the two respective anomalous cases due to their locations being close to the anomalous layers. However, their average performances on the two anomalous cases were not as good as the other groups. For

example, G2 is expected to perform the best, because G2 includes all of the core piezometers. However, the performance was much better when they were analyzed with the upstream piezometers as shown in the result of G5.

Based on these results, it is likely that the upstream piezometers (especially, Piezometer 1 and 2) provide useful information when detecting various anomalous responses, including very hard-to-detect responses. Table 10 and Table 11 also show that excluding the upstream piezometers do not bring better results compared to when including the upstream piezometers. Even though the erosion mechanism was simulated in the downstream foundation area, such changes may eventually affect the seepage path, and correspondingly the responses of upstream piezometers. The anomalous changes in the upstream piezometers are often not as dramatic as the downstream piezometers; however, since upstream piezometers, in most times, have steady relationships with pool levels due to their close locations with respect to pool, comparatively even slight piezometric changes may assist in signaling anomalous responses.

CHAPTER 5. IN-DEPTH DISCUSSION

As presented in the previous chapters, we learned that the widely held assumptions on piezometer data, i.e., linearity between piezometer data and pool levels, as well as normally distributed piezometer data, are not always appropriate, and the anomaly detection performance can be much improved by adopting non-parametric statistical techniques. In addition, by analyzing multiple piezometers together, the anomaly detection performance can be improved.

However, it is still difficult to state that certain techniques will always detect anomalies better than the others. The performance of different anomaly detection techniques will vary based on the data type, the data collection frequency, the sensitivity, the presence of noise, the types of anomalies, etc. For example, even though the overall performance of KL was better than that of MPCA, there will be certain conditions (though rare in real life) in which MPCA will work better. Figure 31, for instance, illustrates the a case where the pool level drastically increases in a short period, but the linear relationships between the piezometer readings and the reservoir levels still stay the same. In this case, MPCA will be able to correctly label the data as normal (due to no changes observed in the principal components) while KL may signal alarms due to the deviated data distribution. Thus, rather than relying on one particular technique, it is also important to consider in which conditions the selected techniques would be more appropriate.

In order to further understand the implications of the multiple piezometer analysis, the piezometer data were evaluated in more detail. We particularly investigated which piezometers

are correlated/uncorrelated, and if such correlational characteristics have any relevance in the anomaly detection task.

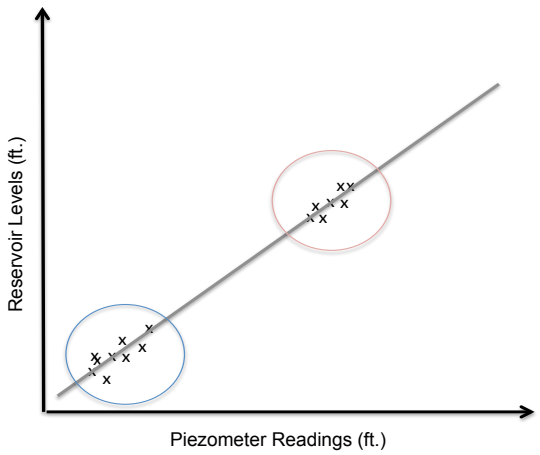

Figure 31. Piezometer vs. Reservoir

Figure 32 corresponds to the images of Pearson product-moment correlation coefficients[4] among 10 piezometers based on Scenario 9 (left) and 10 (right). The figures do not illustrate the correlations over the entire analysis period. Rather, they correspond to the correlation coefficients that are calculated based on the piezometer data collected during the anomalous periods only, which means they represent the correlational relationships when the anomalies are introduced.

[4] Pearson product-moment correlation coefficient measures the linear correlation between two variables. The coefficient of 1 indicates perfect positive correlation (-1 is total negative correlation) while 0 indicates no correlation.

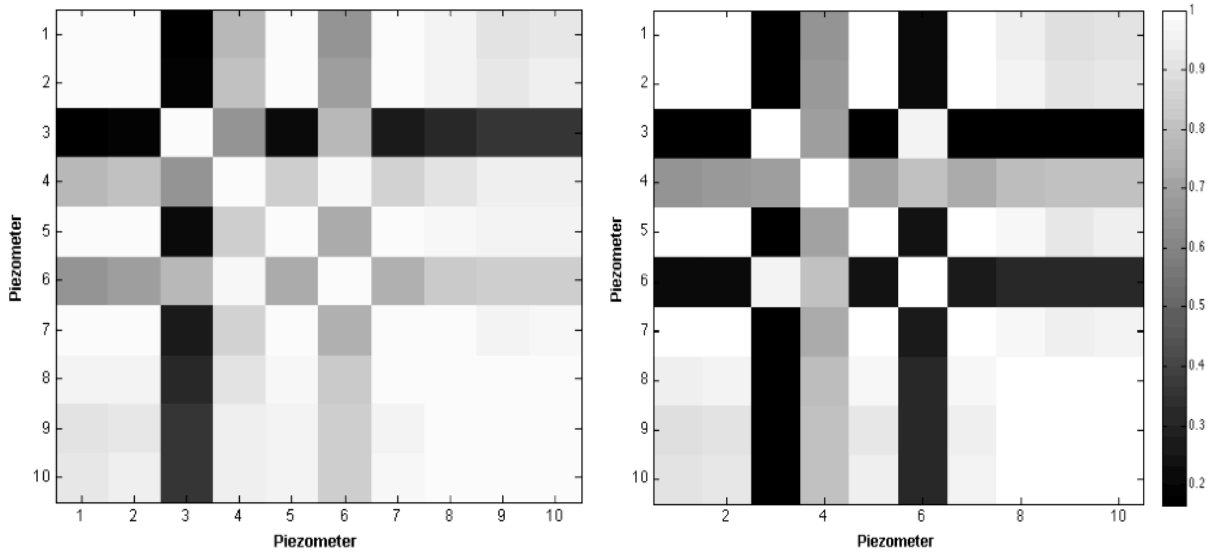

Figure 32. The Images of Pearson Product-Moment Correlation Coefficients among 10 piezometers based on Scenario 9 (left) and 10 (right)

In Table 9, we have observed that analyzing only PZ 1 and 2 (which are upstream piezometers) does not work well on the scenarios that simulate the poorly compacted core layers (i.e., Scenario 9 and 10). However, once PZ 3 was analyzed together, the average F_2 score (between SC 9 and 10) increased from 12% to 83%. As shown in the two image figures above, while the responses of PZ 1 and 2 are correlated to the corresponding anomalous scenarios, PZ 3 has lower correlation coefficients than the other piezometers, meaning it behaves differently under the anomalous scenarios.

In addition, Table 11 showed that the average F_2 scores of Scenario 9 and 10 drastically decreased when only PZ 7, 8, 9 and 10 were analyzed, which have at least 95% correlations (see the dark red areas on the lower right corner of Figure 23). On the other hand, the biggest improvement observed in Table 11 was between PZ 4-5 and PZ 4-6; the average F_2 score of PZ 4-5 was 29% and with additional PZ 6, it increased to 49%. Similar to the previous observation,

such improvement was observed once the piezometer with the low correlation coefficient was analyzed together, i.e., PZ 5 and 6 are not correlated much, especially in Scenario 10.

In Chapter 4, we observed that the anomaly detection performance was improved when the upstream piezometers were analyzed with the core piezometers (i.e., the piezometer group, G5, which consists of PZ 1,2,3,4 and 5). This observation also supports that the piezometers with low correlation coefficients (in this case, PZ 3 and 4) help improve anomaly detection performance. For G1, G3, G7, and G10 (See Table 12), the performances were also poor, in which all of the piezometers included in each group were highly correlated to each other.

The correlation coefficient was one of the metrics that was tested in order to see how the piezometers that respond similarly (high coefficients in this case) or dissimilarly would affect the classification performance. Based on this simple experiment, we observed that uncorrelated piezometers might provide additional information that has not been available from closely correlated piezometers, thus possibly improving the classification performance. However, it is important to note that the magnitude of correlation coefficients does not necessarily have a direct relationship to the degree of classification performance improvement. For example, the first columns of the two image figures, which correspond to the correlation coefficients between PZ 1 and the other piezometers, show that PZ 3, 4 and 6 are comparatively less correlated to PZ 1 than the other piezometers. In both Scenarios 9 and 10, PZ 3 has the lowest correlation coefficients, while PZ 4 has the highest correlation coefficients among the three piezometers. However, when PZ 4 was analyzed together with PZ 1 and 2 (see Table 14), the performance on Scenario 9 was

67%. The performance was higher than the analysis with PZ 6 (i.e., 60%) although PZ 6 is less correlated to PZ 1 and 2 compared to PZ 4.

The inference drawn here should not be perceived as if correlated piezometers (or those that respond similarly under certain anomalous scenarios) are not useful when detecting anomalies. The correlation coefficients we used do not capture all the possible relationships between piezometers that may be useful to improve the anomaly detection performance. Even those piezometers with high correlations may provide additional useful information. It is though noteworthy that by performing the analysis *with* uncorrelated piezometers, there is a high chance that anomalies that may be easily overlooked can be detected better.

CHAPTER 6. CONCLUSION

6.1 Summary of the work

In this thesis, various anomaly detection techniques were explored and evaluated in an effort to improve anomaly detection of embankment dams. We focused on analyzing piezometer data collected from embankment dams, especially to detect internal erosion problem, or potential *indicators* of internal erosion.

In order to see the potential of applying advanced statistical techniques compared to the traditional ways of analyzing piezometer data, we first examined the anomaly detection techniques that are based on Moving Principal Component Analysis (MPCA) as well as Autoregressive (AR) model using the real field data collected from the two case study dams. We observed that those anomalies that would not be detected using the existing data analytics could be detected better. Thus, it has shown the potentials of applying statistical anomaly detection techniques to detect piezometric anomalies.

We then simulated various anomalous scenarios and collected corresponding piezometer data from 10 different locations in order to validate how well the selected anomaly detection techniques can detect anomalies. The piezometer data were simulated using a numerical model called SEEP/W. In our research, two anomalous mechanisms that may possibly be caused by internal erosion were explored.

After the data simulation, we first analyzed the performance of VAR models. Given that most predictive models perform poorly when there is noise present in data, we also examined more robust techniques that extract relevant features from the data in addition to temporal information. Most piezometer readings and pool levels are highly correlated, so practicing engineers typically assume that pool and piezometer measurements are linearly related, and are drawn from a normal distribution. However, our analysis shows that depending on various characteristics of piezometers, such assumptions may not always be the best decision in analyzing the data. Thus, we investigated both parametric and nonparametric techniques (MPCA and KL, respectively) in order to compare how they perform in various piezometers and anomalous scenarios. We tested both various parameter settings (of window sizes and threshold levels) that optimized the performance and a specific parameter combination that may be intuitively adopted in real life. Each of the performance was evaluated using an F_2 score.

As expected, VAR models were not robust to noise present in the training data, so they performed poorly in most piezometers and anomalous scenarios. Both MPCA (parametric), and KL (nonparametric) worked well under the optimal parameters and the specific setting (please refer to Chapter 3). Under the specified setting, we observed that the results of MPCA varied among different piezometers while those of KL were consistent for every piezometer. This implies that KL is sensitive enough to detect even less severe anomalies and applicable to be used in various piezometer locations. Even though we cannot state that KL is better than MPCA, we learned that piezometer data do not always follow linearity and normal distribution, thus the techniques that are based on such assumptions may not always be the best option when analyzing

piezometer data. By using nonparametric techniques, the performance of anomaly detection can be improved.

Chapter 4 presented the follow-up research that was conducted to improve anomaly detection by analyzing multiple piezometers together. We proposed the refined anomaly detection technique that is based on Kullback-Leibler (KL) divergence and Support Vector Machine (SVM). We first computed the KL divergence of each piezometer over time and trained SVM classifiers on various KL divergences of either individual piezometers or multiple piezometers. By considering additional features from multiple piezometers, the anomaly detection performance improved for both of the anomalous cases we have simulated (i.e., foundation erosion and pervious core layer). The piezometers that could not detect anomalies well by themselves could improve in their performance with the additional information provided by the other piezometers, which did not have to be located close or share common characteristics. Based on the experiments, we recognized the value of analyzing multiple piezometer data together, which can possibly provide significant information to detect a wide range of anomalies.

6.2 Contributions

The research presented in this thesis has three main contributions: i) Application and evaluation of different types of anomaly detection techniques that are applicable to piezometer data to systematically detect anomalies that may lead to internal erosion, ii) Improvement of piezometer data analysis by combining more than one piezometer data/features, iii) Exploration and identification of the relationships between the physical anomalous processes being simulated and

the corresponding responses of the differently located piezometers**,** iv) Formalization of an approach as a data analysis process.

i) Application and evaluation of different types of anomaly detection techniques that are applicable to piezometer data to systematically detect anomalies that may lead to internal erosion

Given that there were still many gaps to be filled in the SHM process of dams, we investigated various types of data-driven anomaly detection techniques (which are often preferred when there is no prior knowledge about the structure) that are applicable to piezometer data collected from embankment dams. We investigated both parametric and non-parametric statistical techniques to see if engineers' common assumptions on piezometer data, i.e., linearity and normal distribution, are justifiable. The developed anomaly detection approaches could detect anomalies that might have been overlooked with the existing method, and showed the potential to forecast potential problems by understanding the results of the proposed approach.

Specifically, we showed that s*tatistical anomaly detection techniques to piezometer data from embankment dams detect anomalies that are indicative of internal erosion, more accurately and earlier than qualitative examinations performed by engineers.* We initially observed the weaknesses involved with the current data analytics, and presented the potential of applying statistical anomaly detection techniques to further improve the data analysis approach (Section 2.5). The statistical techniques examined in this research reduced subjectivity involved with engineering interpretation due to robust anomaly detection techniques that are mostly driven by

data themselves and not affected by minor noise.

When regression models (i.e., VAR) were examined, the detection performances on 100 anomalous cases (corresponding to 10 piezometers and 10 scenarios) were poor; more than 80% of the cases had F_2 scores below 60%. When MPCA, which extracts dominant patterns embedded in data, was applied, the overall performance improved; 56 of the 100 simulated cases scored higher than 80%. However, the performances were still poor in some piezometer locations and anomalous scenarios, especially when the piezometric responses were not clearly affected and during the last two anomalous scenarios (i.e., pervious core layer).

While the anomaly detection performances of those two parametric techniques (i.e., VAR and MPCA) varied based on different piezometer locations and anomalous scenarios, those of the nonparametric technique (i.e., KL) were constantly better with the average F_2 score of higher than 90% in most anomalous scenarios we simulated. Compared to the average F_2 scores of VAR and MPCA, especially on the anomalous scenarios of the pervious core layer (which were 7% and 54%, respectively), the average F_2 score of KL, which was 81% showed the clear improvement.

Through in-depth evaluations, we also showed that the prior assumptions on data distribution and linearity, which may be easily made during piezometer data analysis, are not always appropriate, and anomaly detection performance can be improved by not making such assumptions about the data.

ii) Improvement of piezometer data analysis by combining more than one piezometer data/features

In Chapter 4, we tested if *analyzing anomalies in the behavior of multiple piezometers allows for the detection of potential internal erosion problems more accurately than analyzing anomalies from a single piezometer*. We first enhanced anomaly detection by combining the KL-based anomaly detection with a classification technique, Support Vector Machine (SVM). We tested various combinations of piezometers, and analyzed which groups or combinations of piezometers perform better for predicting the various anomalous scenarios. By considering additional features from multiple piezometers, the anomaly detection performance improved for the anomalous cases we have simulated.

The piezometers that could not detect anomalies well by themselves could improve in their performance with the additional information provided by the other piezometers, which did not have to be located close or share common characteristics. Compared to when using a single piezometer (whose average performance over the 10 scenario ranged from 65% to 78%), once the dimensions of the data increased (with additional piezometer at a time), the average F2 scores became much higher ranging from 76% to 91%. In addition, the average performance for the last two anomalous scenarios (i.e., the pervious core layer) based on the best piezometer group we formed was at least 22% higher than that of the best performing single piezometer.

Based on the experiments, we recognized the value of analyzing multiple piezometer data together, which can possibly provide significant information to detect a wide range of anomalies.

Furthermore, based on which group of piezometers has been analyzed together, engineers would be able to infer more information about the detected anomalies within the group, such as locations of where the anomalies might have been occurring and which piezometers provide useful information.

iii) Exploration and identification of the relationships between the physical anomalous processes being simulated and the corresponding responses of the differently located piezometers

The simulation approach presented in this thesis to generate anomalous data using a physical seepage model enhanced the validation process by providing various anomalous scenarios, which are often insufficient and not easily identifiable in real life.

In order to simulate realistic scenarios, possible mechanisms of internal erosion problems that are specific to the case study dam (i.e., after reviewing the historical assessment of the case study dam and potential failure modes) were studied: a) foundation erosion, and b) poorly compacted zone. In most cases, internal erosion often occurs from the toe area and develops backwards towards an embankment. By understanding that such process is often caused by the constant migration of soil particles towards free exits or into coarse openings, we varied the hydraulic conductivity of the foundation soil layer to simulate the foundation erosion. For the poorly compacted zone, thus highly permeable zone, we varied the hydraulic conductivity of the core layer, whose soil property is different from the foundation. In order to examine various anomalies associated with those anomalous mechanisms, we also varied severity and duration of

the anomalous scenario. After identifying potential anomalous mechanisms, we collected the corresponding piezometer data and observed how they are affected by the various anomalous processes.

This simulation approach provided additional information of how piezometric levels are reflected by the specific physical processes, which have been hard for engineers to recognize proactively especially given that problems often occur inside embankment dams without visual signs. Thus, such simulated datasets not only benefit the validation processes but also contribute to dam engineers who have been having hard times understanding how instrumentation data would be reflected based on different dam conditions.

iv) Formalization of an approach to be able to replicate data analysis process

Even though our experiments were based on the case study dam, the research approaches presented in this thesis can be replicated to other dam projects.

- *Simulating various anomalous mechanisms to investigate behavior of instrument data*

Since different dam projects behave differently (based on their locations, construction materials, environmental factors, etc.), their potential failure modes/mechanisms will be also different. Even though this thesis presented only two possible anomalous mechanisms that are possible from the case study dam, the proposed simulation approach with the numerical model can be also used to simulate and obtain other types of anomalous mechanisms (at various locations) that are possible from different dam projects.

- *Selecting appropriate anomaly detection techniques, e.g., parametric vs. non-parametric*

techniques

The application results presented in this thesis showed that the non-parametric technique performed better, thus not to rely on the prior assumptions on piezometer data which do not always provide the best performance for anomaly prediction. As mentioned before, different dams behave differently, meaning there would be certain cases where one type of technique will work better than the other. As we explored and evaluated different types of statistical techniques, which entail different characteristics, appropriate data analysis technique should be selected by carefully understanding various parametric settings, the nature of the data, critical conditions, sensitivity, and many other relevant factors, in order to obtain better performance and insights to understand problems.

- *Investigating various characteristics of different groups of piezometers*

In Chapter 4 and 5, we tried to formalize the piezometer grouping approach by evaluating the detection performance of various piezometer groups and by correlating the anomalous behaviors of different piezometers on various anomalous scenarios. Even though this research did not cover all of the other potential significant characteristics/features (other than the correlation feature), such interpreting approach presented in this thesis is a good practice to explore different features of data that may contribute to improve anomaly detection performance.

Broader Impacts: Out of 83,000 dams listed in the National Inventory of Dams (NID), the number of deficient dams has risen by 85% since 1998. This significant increase in the number of high-hazard potential dams nationwide whose failure would cause loss of life is a big concern (FEMA 2009).

Our research on improving poor dam conditions by advancing data analysis tools have potential to assess such high numbers of deteriorating dams more efficiently and accurately. The research approaches and results have been discussed with dam operators and engineers from government agency such as USACE and US Bureau of Reclamation, and other private sectors whose goal is to improve managing infrastructure system, as validation purposes and to get feedbacks on the approaches taken. The investigated advanced techniques will enhance dam safety as well as benefit the practitioners, too. In addition, since there is no specific procedure to assess internal erosion with a wide variety of engineering judgment, being able to systematically analyze internal erosion using only piezometer data will be a significant contribution to the dam safety domain.

6.3 Future Directions

This thesis focused on detecting anomalies among piezometer data only. Given that there are various instruments installed in dams, such as strain gages, inclinometers, resistance thermometers, accelerometers, etc., the research presented in this thesis can be broadened to analyze various instrument data. Since data collected from different types of instruments may contain distinctive information that has not been available from the piezometer data, there is potential that anomaly detection performance can be much improved.

We only presented the analysis on the daily piezometer data. However, given that different dam projects have different frequencies of collecting instrument data (due to their required sensitivities, environmental conditions, etc.), it would be also interesting to analyze instrument

data with various sampling rates (which will affect the appropriate parameter settings) and observe if there are any improvements in detectability of the anomaly detection techniques.

In Chapter 3, the finite element software, SEEP/W was introduced which was used to simulate the anomalous piezometer data. The problems that can be solved in SEEP/W are limited to be defined in two dimensions, so users can only analyze a single cross-section of a dam at a time. Given that seepage flows are not limited to 2-dimensional, and anomalies can occur across more than one cross-section, using a 3-dimensional numerical model will deepen the research work.

In order to further improve the piezometer analysis, the approaches proposed in this research can be tested in other dam projects, in which historically dense instrumentation data are available, so that the inferences made on various piezometer groups (in multiple piezometer analysis) can be generalizable. In addition, if the most appropriate anomaly detection techniques can be automatically and interactively selected (e.g., using an user interface) for different dam projects by understanding historical behaviors, frequencies of instrumentation data, sensitivities required, and may other relevant factors, the data analytics will be much enriched. Moreover, even though we examined both parametric and nonparametric techniques in order to see if the common assumptions on piezometer data are appropriate or not, we could not explore all of the statistical techniques available. Thus, anomaly detection performance can be further improved if other data-driven statistical techniques can be also investigated. Applications of different techniques will provide better insights to understand various problems.

The research presented in this thesis analyzed two anomalous cases (foundation erosion and poorly compacted foundation layer) that may lead to internal erosion. However, it would be also interesting to explore additional anomalous cases such as sliding, liquefaction during an earthquake, slope-instability, etc. in order to understand broad ranges of anomalous scenarios and corresponding instrument data.

In Chapter 5, we examined the linear correlational relationships among various anomalous piezometer data in an effort to formalize the piezometer-grouping approach. This analysis can be further extended to explore other characteristics of the piezometer data within various groups, which resulted in better anomaly detection performance.

Lastly, this research can be also extended to other domains such as medical informatics and fraud detection, whose objective is to detect unexpected behaviors, or deviations from normal behavior. However, given that the performances of anomaly detection techniques vary based various factors (e.g., data type, anomaly type, sensitivity, etc.), further studies will be required in order to appropriately apply them to other domains.

APPENDIX A

Embankment Dams

Embankment dams are primarily constructed of the natural materials of the earth, i.e., soil and rock. Compared to concrete dams, embankment dams are more economical to build and various foundations and topography are allowed (Veesaert, Chris J. et al. n.d.). It is, however, important to allow sufficient height and spillway capacity to keep an embankment dam safe. The principal types of embankment dams are 1) earth dams and 2) rock-fill dams.

Earth dams are composed of suitable soils. They are spread and compacted in layers by mechanical means (URBR 2010). Rock-fill dams are mainly composed of fragmented rocks with an impervious earth core where the core is separated from the shells by various transition zones that are built of properly graded materials (URBR 2010). Figure 24 illustrates major features of an embankment dam.

The upstream and the downstream slopes are the inclined surfaces of the dam in contact with the reservoir or away from the reservoir, respectively. In order to protect them from erosion, riprap (a layer of large stones, broken rock, or precast blocks) or some other protection material should be placed. Usually, grass or rock is used to protect the downstream slope (Veesaert, Chris J. et al. n.d.).

Toe of an embankment dam refers to the junction of the downstream slope with the ground surface. It is important to control seepage by inspecting any changes in the seepage rates at the

downstream toe. Usually, relief wells are installed to control and safely discharge seepage from under the dam. The contacts between the downstream slope and the abutments are prone to seepage, because they are less dense and more pervious compared to other parts of the embankment (Veesaert, Chris J. et al. n.d.).

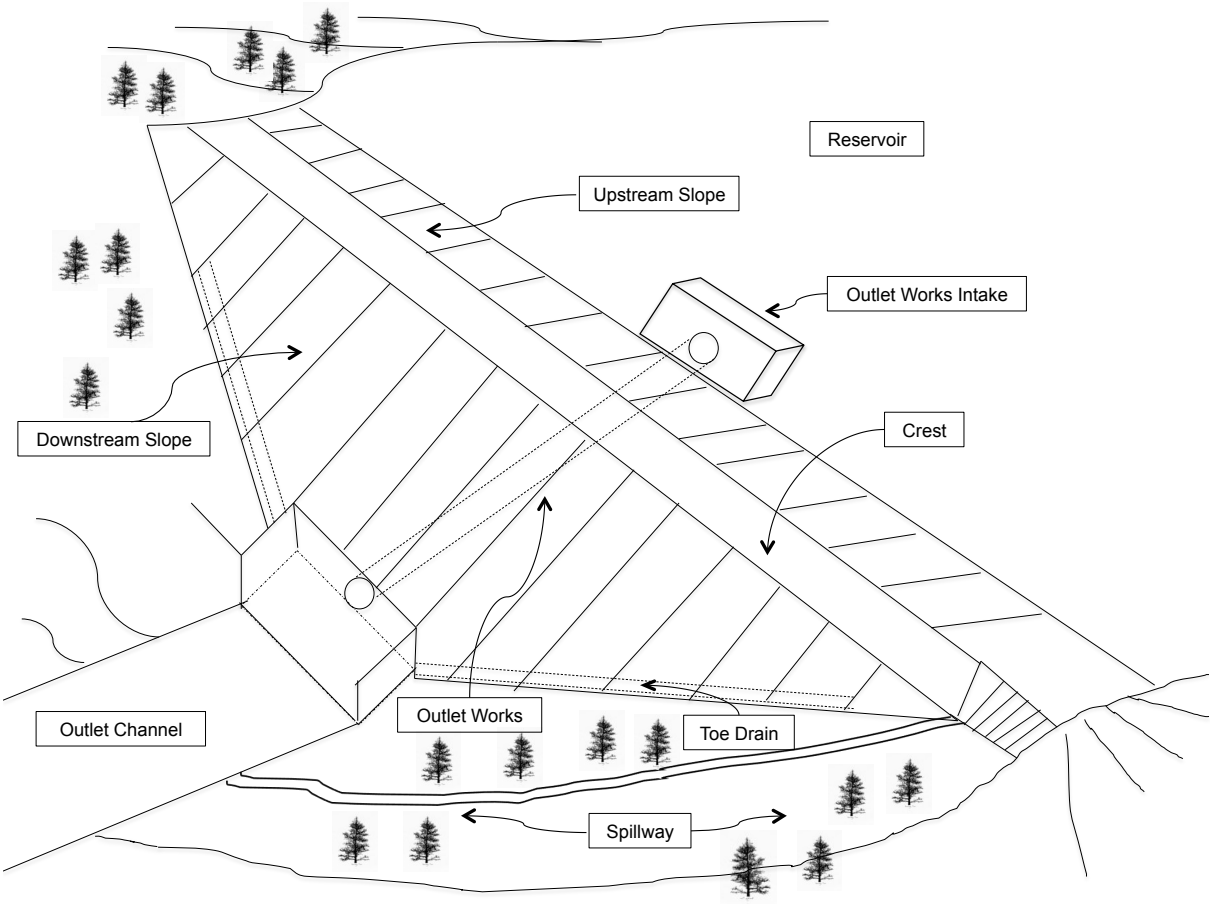

Figure 33. Typical Embankment Dam

A conduit is a closed channel that conveys the discharge through or under a dam while outlet is an opening through which water can be discharged for a particular purpose from a reservoir

(New Hampshire Department of Environmental Services 2011). Spillway discharges flood flows. In addition to seepage control, cracking and sliding are also important deficiencies to be inspected to prevent dam failures.

APPENDIX B

SEEP/W Seepage Analyses (USBR 2011; GEO-SLOPE International 2012)

As a part of the GeoStudio software package, SEEP/W is a finite element numerical method computer program that is used to model fluid flow and pore-water pressure distribution within porous materials. SEEP/W can model conditions of a saturated-only flow or both saturated and unsaturated flow. In addition, seepage can be analyzed as a function of time (i.e., transient analysis). When performing a transient analysis, users need to provide initial conditions (e.g., reservoir levels, loadings, etc.) or perform steady-state analysis first to reach a steady condition, which in turn becomes the initial condition of the transient analysis.

The flow of water through both saturated and unsaturated soil computed in SEEP/W follow Darcy's law. Darcy's law states that

$$q = ki$$

where q is the specific discharge, k, the hydraulic conductivity, and i the gradient of total hydraulic head.

Partial differential water flow equations for two-dimensional seepage implemented in SEEP/W are expressed as:

$$\frac{\partial}{\partial x}\left(k_x \frac{\partial H}{\partial x}\right) + \frac{\partial}{\partial y}\left(k_y \frac{\partial H}{\partial y}\right) + Q = \frac{\partial \theta}{\partial t}$$

where H is the total head, k, the hydraulic conductivity in either x or y direction, Q, the applied boundary flux, θ, the volumetric water content, and t, the time.

According to the equation, the difference between the flow entering and leaving an element at a point in time is equal to the change in storage of the soil systems. Under steady-state conditions, the right side of the equation will be zero, meaning the flow (flux) entering and leaving is the same at all times.

The finite element equation of SEEP/W with temporal integration is expressed as:

$$(\Delta t[K] + [M])\{H_1\} = \Delta t(\{Q_1\}) + [M]\{H_0\}$$

where t is the time increment, K is the element characteristic matrix, M is the element mass matrix, H_1 is the head at end of time increment, H_0 is the head at start of time increment, and Q_1 is the nodal flux at end of time increment. In order to know the head at the end of the time increment, the initial conditions must be known.

REFERENCES

Adhikari, R., and Agrawal, R. K. (2013). *An Introductory Study on Time Series Modeling and Forecasting*. LAP LAMBERT Academic Publishing, Saarbrücken.

Akaike, H. (1998). "Information Theory and an Extension of the Maximum Likelihood Principle." *Selected Papers of Hirotugu Akaike*, Springer Series in Statistics, E. Parzen, K. Tanabe, and G. Kitagawa, eds., Springer New York, 199–213.

Anscombe, F. J., and Guttman, I. (1960). "Rejection of Outliers." *Technometrics*, 2(2), 123–147.

Armour, B., and Morgera, S. D. (1991). "An exact forward-backward maximum likelihood autoregressive parameter estimation method." *IEEE Transactions on Signal Processing*, 39(9), 1985–1993.

ASCE. (2013a). "ASCE | 2013 Report Card for America's Infrastructure | Grade Sheet: Gpa." <http://www.infrastructurereportcard.org/a/#p/grade-sheet/gpa> (Sep. 4, 2013).

ASCE. (2013b). "ASCE | 2013 Report Card for America's Infrastructure | Dams: Conditions & Capacity." <http://www.infrastructurereportcard.org/a/#p/dams/conditions-and-capacity> (Sep. 4, 2013).

ASDSO, and FEMA. (2012). "Living With Dams: Know Your Risks." NDSP-National Dam Safety Program.

Barthorpe, R. J. (2010). "On Model- and Data-based Approaches to Structural Health Monitoring." Thesis, University of Sheffield.

Basseville, M. (2013). "Review: Divergence Measures for Statistical Data processing-An Annotated Bibliography." *Signal Process.*, 93(4), 621–633.

Bishop, C. M. (2007). *Pattern Recognition and Machine Learning*. Springer, New York.

Bligh, W. G. (1910). "Dams Barrages and weirs on porous foundations." *Engineering News*, 64 (26).

Brownjohn, J. M. W. (2007). "Structural health monitoring of civil infrastructure." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851), 589–622.

Bureau of Reclamation. (2012). "Best Practices (26. Internal Erosion Risk)." U.S. Department of the Interior.

Chandola, V., Banerjee Arindam, and Kumar, V. (2009a). "Anomaly Detection - A Survey." *ACM Computing Surveys*, 41(3).

Chandola, V., Cheboli, D., and Kumar, V. (2009b). *Detecting Anomalies in a Time Series Database*. CS Technical Report, Technical Report, Computer Science Department, University of Minnesota.

Chang, P. C., Flatau, A., and Liu, S. C. (2003). "Review Paper: Health Monitoring of Civil Infrastructure." *Structural Health Monitoring*, 2(3), 257–267.

"Combining Model Based and Data Based Techniques in a Robust Bridge Health Monitoring Algorithm." (n.d.). <http://cait.rutgers.edu/cait/research/combining-model-based-and-data-based-techniques-robust-bridge-health-monitoring-algori> (Nov. 3, 2013).

Cortes, C., and Vapnik, V. (1995). "Support-vector networks." *Machine Learning*, 20(3), 273–297.

Crum, D. (2011). "Rogue Piezometers." *31st Annual USSD Conference*, 21st Century Dam Design - Advances and Adaptations, 1591–1602.

El-Naqa, I., Yang, Y., Wernick, M. N., Galatsanos, N. P., and Nishikawa, R. (2002). "A support vector machine approach for detection of microcalcifications in mammograms." *2002 International Conference on Image Processing. 2002. Proceedings*, II–953–II–956 vol.2.

Eskin, E. (2000). "Anomaly Detection over Noisy Data Using Learned Probability Distributions."

Fell, R., Wan, C., Cyganiewicz, J., and Foster, M. (2003). "Time for Development of Internal Erosion and Piping in Embankment Dams." *Journal of Geotechnical and Geoenvironmental Engineering*, 129(4), 307–314.

Flores-Berrones, R., and Patricia, N. (2011). "Internal Erosion Due to Water Flow Through Earth Dams and Earth Structures." *Soil Erosion Studies*, D. Godone, ed., InTech.

Gall, E. (2007). "Correlation Plot-A Modern Tool in Dam Safety Monitoring." *27th Annual USSD Conference*, Modernization and Optimization of Existing Dams and Reservoirs, 537–551.

GEO-SLOPE International. (2012). *Seepage Modeling with SEEP/W*. GEO-SLOPE International Ltd., Calgary, Alberta., Canada.

Gul, M., and Catbas, F. (2011). "Damage Assessment with Ambient Vibration Data Using a Novel Time Series Analysis Methodology." *Journal of Structural Engineering*, 137(12), 1518–1526.

Hodge, V. J., and Austin, J. (2004). "A Survey of Outlier Detection Methodologies." *Artificial Intelligence Review*, 22(2), 85–126.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer, New York.

Kay, S. (1999). *Modern Spectral Estimation: Theory and Application*. Prentice Hall, Englewood Cliffs, N.J.

Krahn, J. (2004). "Seepage modeling with SEEP/W: An engineering methodology." GEO-SLOPE International Ltd., Calgary, Alberta., Canada.

Kullback, S., and Leibler, R. A. (1951). "On Information and Sufficiency." *The Annals of Mathematical Statistics*, 22(1), 79–86.

Laory, I., Trinh, T., Posenato, D., and Smith, I. (2013). "Combined Model-Free Data-Interpretation Methodologies for Damage Detection During Continuous Monitoring of Structures." *Journal of Computing in Civil Engineering*, (ja).

Lum, K. Y., and Sheffer, M. R. (2005). "Dam Safety: Review of Geophysical Methods to Detect Seepage and Internal Erosion in Embankment Dams." <http://www.hydroworld.com/articles/hr/print/volume-29/issue-2/articles/dam-safety-review.html> (Sep. 3, 2013).

Marple, S. L. (1986). *Digital Spectral Analysis: With Applications*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

MDEQ. (2007). "Evaluation of Seepage Conditions." Mississippi Department of Environmental Quality.

Nair, K. K., Kiremidjian, A. S., and Law, K. H. (2006). "Time series-based damage detection and localization algorithm with application to the ASCE benchmark structure." *Journal of Sound and Vibration*, 291(1–2), 349–368.

New Hampshire Department of Environmental Services. (2011). "Environmental Fact Sheet." DES Dam Bureau.

Omenzetter, P., and Brownjohn, J. M. W. (2006). "Application of time series analysis for bridge monitoring." *Smart Materials and Structures*, 15(1), 129.
Parra, L., Deco, G., and Miesbach, S. (1996). "Statistical independence and novelty detection with information preserving nonlinear maps." *Neural Comput.*, 8(2), 260–269.
Pelton, F. (2000). "Guidelines for Instrumentation and Measurements for Monitoring Dam Performance."
Polani, D. D. (2013). "Kullback-Leibler Divergence." *Encyclopedia of Systems Biology*, W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota, eds., Springer New York, 1087–1088.
Posenato, D., Lanata, F., Inaudi, D., and Smith, I. F. C. (2008). "Model-free data interpretation for continuous monitoring of complex structures." *Advanced Engineering Informatics*, Intelligent computing in engineering and architecture, 22(1), 135–144.
Schlindwein, F. S., and Evans, D. H. (1990). "Selection of the order of autoregressive models for spectral analysis of Doppler ultrasound signals." *Ultrasound in medicine & biology*, 16(1), 81–91.
Schurer, J., Wilkinson, E., Norfleet, J., Sciver, J. V., Huntington, C., Lee, C., and Rogers, A. (2002). "Dam Safety Manual." State of Colorado.
Sellmeijer, J. B. (1988). "On the mechanism of piping under impervious structures." *Delft University of Technology, Delft*.
Shumway, R. H., and Stoffer, D. S. (2011). *Time Series Analysis and Its Applications - With R Examples*. Springer.
Szabo, Z. (2014). "Information Theoretical Estimators Toolbox." *arXiv:1405.2106 [cs, math, stat]*.
Torres, R. (2009). "Considerations for Detection of Internal Erosion in Embankment Dams." *GEO-Velopment*, American Society of Civil Engineers, 82–98.
URBR. (2010). "Internal Erosion, Dam Safety Risk Analysis Best Practices." US Department of Interior, Bureau of Reclamation.
USBR. (2011). "Embankment Dams- Chapter 8: Seepage." US Department of Interior, Bureau of Reclamation.
USSD Committee on Materials for Embankment Dams. (2010). *The Aging of Embankment Dams*. United States Society on Dams.
Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., New York, NY, USA.
Veesaert, Chris J. et al. (n.d.). "Training Aids for Dam Safety: Inspection of Embankment Dams."
Wang, Q., Kulkarni, S. R., and Verdu, S. (2009). "Divergence Estimation for Multidimensional Densities Via -Nearest-Neighbor Distances." *IEEE Transactions on Information Theory*, 55(5), 2392–2405.
Weekley, R. A., Goodrich, R. K., and Cornman, L. B. (2010). "An Algorithm for Classification and Outlier Detection of Time-Series Data." *Journal of Atmospheric and Oceanic Technology*, 27(1), 94–107.
Xu, X. (2010). "Sequential anomaly detection based on temporal-difference learning: Principles, models and case studies." *Applied Soft Computing*, 10(3), 859–867.

Ying, Y., Garrett, J., Oppenheim, I., Soibelman, L., Harley, J., Shi, J., and Jin, Y. (2013). "Toward Data-Driven Structural Health Monitoring: Application of Machine Learning and Signal Processing to Damage Detection." *Journal of Computing in Civil Engineering*, 27(6), 667–680.

Zhang, G. P. (2003). "Time series forecasting using a hybrid ARIMA and neural network model." *Neurocomputing*, 50, 159–175.

Zhao, M., and Saligrama, V. (2009). *Anomaly Detection with Score functions based on Nearest Neighbor Graphs*. arXiv e-print.