Anomaly Detection on Piezometer Data Collected from Embankment Dams Using Physical Model-Based Simulation

In-Soo Jung[1], Mario Berges[2] and James H. Garrett Jr[3]

[1]Carnegie Mellon University, Department of Civil and Environmental Engineering; 5000 Forbes Avenue, Pittsburgh, PA 15213; ijung@andrew.cmu.edu
[2]Carnegie Mellon University, Department of Civil and Environmental Engineering; 5000 Forbes Avenue, Pittsburgh, PA 15213; marioberges.cmu.edu
[3]Carnegie Mellon University, Dean of Carnegie Institute of Technology; 5000 Forbes Avenue, Pittsburgh, PA 15213; garrett@cmu.edu

ABSTRACT

Embankment dams, like most other civil infrastructure systems, are exposed to harsh and largely unpredictable environments. However, unlike bridges, buildings and other structures, their design specifications and as-is properties are not generally known in the same level of detail due to, among other things, their age and the difficulties associated with assessing their internal structure. Hence, making sense of measurements collected from instruments used to monitor their behavior requires sound engineering judgment and analysis, as well as robust statistical analysis techniques to prevent misinterpretation. 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. In our previous work, we presented the effectiveness of applying statistical anomaly detection techniques such as Principal Component Analysis and Robust Regression Analysis when analyzing piezometer data collected from embankment dams. In this paper, we present how we could improve our work by testing with simulated anomalies that are indicative of internal erosion problems. In order to closely replicate more realistic anomalous scenarios, a physics-based model of an embankment dam was developed. By varying a hydraulic conductivity of a soil material in the model, corresponding detection accuracies and sensitivities of the statistical anomaly detection algorithm were evaluated. When we applied our proposed anomaly detection on more realistically simulated anomalous data using the numerical model, the detection accuracy came out to be 98.5%.

INTRODUCTION

There are 84,000 dams in the U.S., and their average age is more than 53 years old, which means they are in need of extensive rehabilitation given that their design life is typically 50 years (ASCE 2013). Moreover, dams have received a grade of D in the 2013 ASCE infrastructure report card (ASCE 2013). While dams provide drinking water, hydroelectric power, flood control, recreation and many other benefits, dams can pose significant risks to people living around them. Even though dam failures are low probability, they have high consequence (ASDSO and FEMA 2012). Thus, dam safety is one of the 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 dams in use today (ASDSO and FEMA 2012). 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 many different parts of the dams, and any changes in this behavior may signal problems. Thus, monitoring pore pressures as well as water levels of an embankment dam to observe potential seepage problems is important to prevent internal erosion and other structural failure modes that are common in these types of dams.

While embankment dams can fail due to overtopping, sliding, spillway/gate problems, sub-standard construction materials, poor maintenance, etc., one of the leading causes of their failures 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 inspections and seepage measurements (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. Thus, it is important to detect anomalies that are indicative of internal erosion problems in advance to prevent catastrophic consequences.

Engineers monitor instrument data regularly to ensure if 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, it indicates various potential problems, such as piping, cracks, malfunctions of the piezometers, etc. Embankment dams, like most other civil infrastructure systems, are exposed to unpredictable environments. However, 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 dams' behavior is not an easy task, requiring sound engineering judgment and analysis, as well as robust statistical analysis techniques to prevent misinterpretation. Because the current practice of analyzing the structural integrity of embankment dams relies primarily on manual *a posteriori* analysis of instrument data by engineers, it leaves much room for improvement through the application of automated data analysis techniques.

Piezometers are the most commonly used instruments in dams to monitor water levels, and they can be used to compute pore water pressures (Crum 2011). For embankment dams, piezometer levels and reservoir levels are directly related, so they are usually compared to monitor the seepage as well as to check the condition of the piezometers. Since dams often have slow responses, relatively small anomalies, which may be caused by initiations of any catastrophic events that are not obvious from engineers' views or those that are outside of the analysis period, may be easily overlooked. Thus, the traditional practice that subjectively detects deviations from the historical readings using time-series or correlation plots with raw data, which often contain much noise, may not be accurate enough to capture anomalous changes over time. Thus, in an effort to implement a quantitative and robust approach to monitor the performance of embankment dams based on piezometer data, we have previously applied Moving Principal Component Analysis and Robust Regression Analysis as the anomaly detection (Jung et al. 2013). To test anomaly detectability, several anomalies have been simulated by de-correlating the relationships between piezometer and reservoir readings over certain periods. However, we observed that such de-correlation approach might not simulate a realistic anomalous scenario. Thus, we improved our previous approach by collecting piezometer data using a numerical model to simulate anomalous scenarios that are more realistic.

APPROACH

Simulation for normal (baseline) and anomalous piezometer readings. Recently, numerical analyses as coded into computer programs have been the most widely used method 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 dams can be validated. In this study, an embankment dam was modeled with SEEP/W of the GeoStudio 2012 package (Geo-Slope International Ltd), which is often used to analyze seepage problems. It is a finite element software program that adopts an implicit numerical solution to solve Darcy's equation for saturated and unsaturated flow conditions over space and time (Krahn 2004). Using SEEP/W models, relevant parameters can be varied (e.g., reduction of hydraulic conductivity of soil layers and/or core), thus obtaining datasets that correspond to different conditions of a dam. In addition, steady-state and transient seepage analyses (during specified time sequences) are possible.

Before simulating anomalies, piezometer readings (assuming this piezometer is in the centerline of the dam, and its tip is located in the foundation) were collected based on the normal condition of the modeled dam. Here, a hydraulic boundary function was computed using the daily reservoir levels obtained from one case study dam during five years, i.e., Mar. 2006 to Sep. 2011. Figure 1 shows the corresponding time series of the piezometer readings and the reservoir levels. This 'normal' dataset was used as the baseline of this study. In this paper, we only present the result on one piezometer installed in one station of the modeled dam.

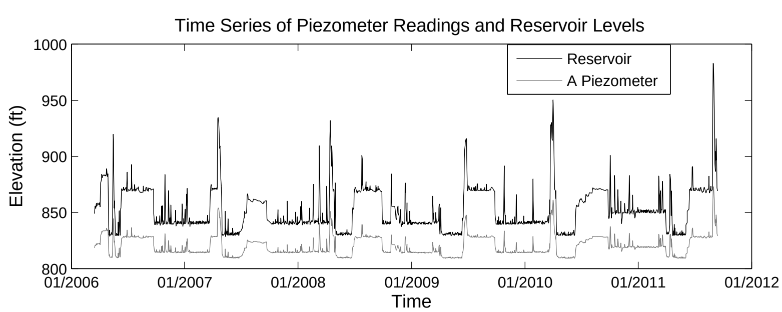

Figure 1. Time-series of the piezometer and the reservoir levels based on the normal dam condition

Table 1 and Figure 2 below show the soil properties used when modeling the embankment dam, and a cross section of the dam, respectively.

Table 1. Soil properties used when modeling an embankment dam

Material	k_{sat} (Saturated hydraulic conductivity in ft/s)	Anisotropy ($k_{vertical}/k_{horizontal}$)
US Rock Fill	1	1
US Gravel	0.001	0.2
Impervious	8e-006	0.2
Rock Fill	0.002	0.2
Till	0.0002	1
River Deposits	0.007	1
DS Gravel	0.0002	0.2

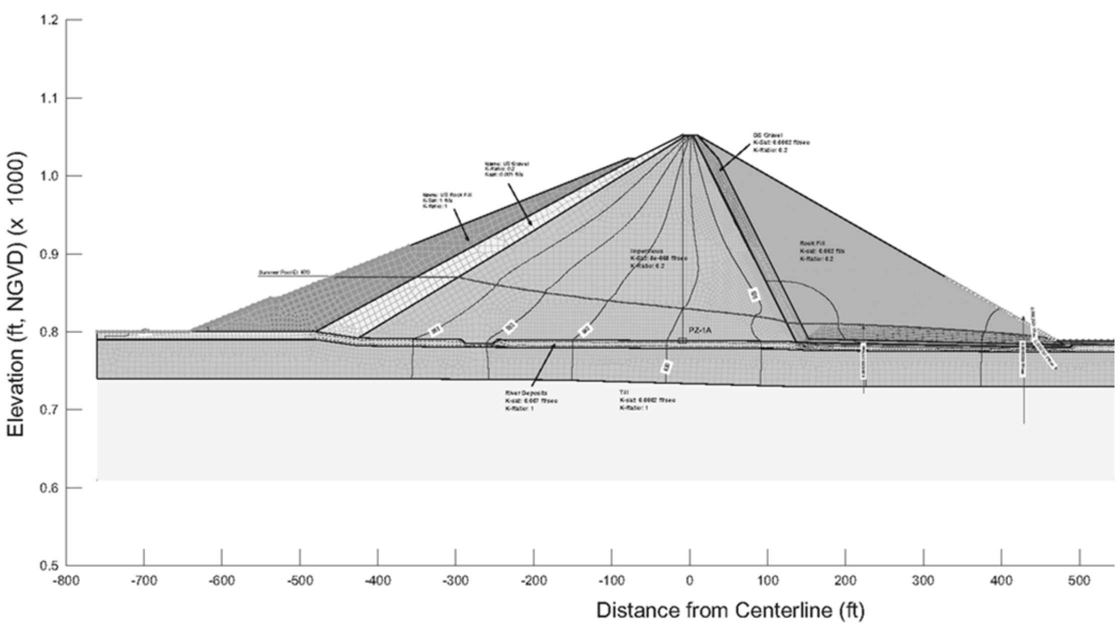

Figure 2. An embankment dam section modeled using SEEP/W

Anomaly detection based on Moving Principal Component Analysis (MPCA) and Robust Regression Analysis (RRA). In (Jung et al. 2013), we have applied an anomaly detection that is based on MPCA and RRA to see if this method can be also useful when analyzing piezometer data from embankment dams. To help the readers, we repeat here the basic theoretical explanation of the proposed detection method as discussed in (Jung et al. 2013).

In PCA, a data matrix is decomposed into a number of uncorrelated components where each of them represents a different degree of dominant variability embedded in the data. Suppose there is a data matrix, $Q \in \mathbb{R}^{N \times M}$, whose M columns are individual time-series of length N (e.g., measurements from M piezometers) that have been normalized with respect to each column. Each entry of this matrix can be denoted by $V_i(t)$, where $i = 1, \dots, M$ and $t = 1, \dots, N$, as shown in the equation below. $V_i(t)$ is the measurement of piezometer *i* at time *t*.

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

First, a singular value decomposition (SVD) is performed on the matrix, Q. During SVD, the matrix, Q first gets decomposed into matrices $U, S,$ and V, where $C = U * S * V^T$. 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, and the diagonal elements of S correspond to the square roots of the eigenvalues, e, of the covariance matrix. The eigenvectors represent the directions of the variance, or the variance of each component, and each of the corresponding eigenvalues indicates a degree of each component's proportional variance. Thus, the most dominant patterns can be captured by the first few sets of the eigenvectors after ordering the corresponding eigenvalues in a descending order.

While a common PCA approach is applied to the whole dataset, it can also be varied to analyze a subset, or a window, of the dataset. When analyzing time series data, for example, a window can slide from the beginning to the end of the dataset by performing PCA in each time step in order to detect any change in the main direction over time. Thus, this moving window approach is often called Moving PCA (MPCA). To apply MPCA on Q, first a sliding window of size L is applied to the matrix, to extract a sub-matrix, called $R(k) \in \mathbb{R}^{L \times M}$ at each time value k, where $k = 1, \dots, (N - L)$. Then, a singular value decomposition (SVD) is performed on each $R(k)$ in the same manner as PCA.

Once the direction of the most variability is computed for each time step, any changes in the eigenvectors over time, which would signal the presence of an anomaly, need to be detected. Robust Regression Analysis (RRA), which is known as a good regression technique in the presence of outliers, is performed to observe if any changes in the first few relevant eigenvectors from $R(k)$ have occurred over time. Among many types of robust regression models, we employ the method that uses iteratively reweighted least squares with a bisquare weighting function. Using only normal data, regression model is formed, and the threshold level is determined by computing certain degrees of 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 would be marked as anomalies.

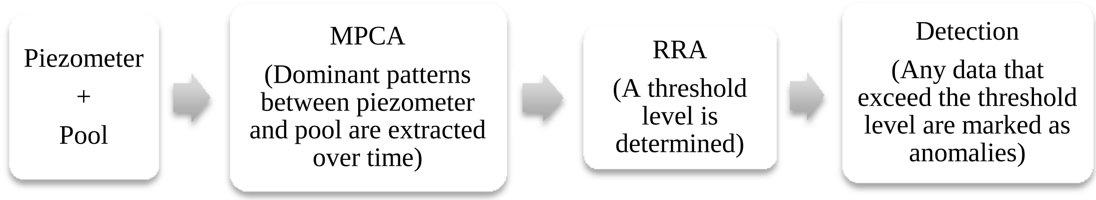

Figure 3. The proposed anomaly detection technique

APPLICATIONS AND RESULTS

As mentioned above, one of the main causes of embankment dam failures all over the world is internal erosion, or a piping event, which occurs due to the constant migration of soil particles towards free exits or into coarse openings (Flores-Berrones and Patricia 2011). One of the main factors affecting such erosion phenomenon is the hydraulic conductivity of the materials (Flores-Berrones and Patricia 2011). In this study, anomalies were simulated for 4 months starting from Jan. 2010 to Apr. 2010 by changing the hydraulic property of the downstream rock fill embankment shell, which is the rightmost material of the dam in Figure 2, given that most piping events would initiate from the toe area. The initial hydraulic conductivity of the rock fill was 0.002 ft/s. It was changed to a saturated-only material instead of an unsaturated/saturated material, and its hydraulic conductivity was increased to 0.2 ft/s to see how such changes would affect the seepage flow pathway, consequently piezometric responses.

When applying MPCA, a window size of 365, which corresponds to a year, was used to capture a periodic behavior of the dam. Since a piezometer responds to reservoir pool events, especially when it is located close to the upstream of a dam, and its tip is located in pervious soil layers, the proposed anomaly detection was performed using both of the piezometer readings and the reservoir levels. In our application, we observed the changes in the first eigenvectors only (to make the detection task not too sensitive to any minor changes), and the robust regression model (RRA) was developed based on the first year of the data. Any regression residuals computed from MPCA+RRA that exceed a threshold level, which was set as ±6 standard deviations of absolute values of the regression residuals were marked as anomalies. When the normal dataset was tested, 161 anomalies were detected. This corresponded to almost every April to summer season where the provided reservoir levels rapidly rise up and down due to precipitation and other seasonal effects. Such rapid filling and rapid drawdown can modify flow conditions inside a soil mass, thus causing uncontrolled saturation and seepage forces (Flores-Berrones and Patricia 2011).

When anomalous datasets were tested, 615 anomalies were detected, or 454 more anomalies than the normal dataset. The additional anomalies corresponded to Jan. 26, 2010 to Apr. 27, 2011, which do not overlap with the anomalies detected from the normal dataset (please see Figure 4 and 5). Since the window size of a year was used in our application, the anomalies were expected during one year before and one year after where the actual anomalies were introduced. Given this comparably big window size, it was a satisfactory result to see the detected anomalies during Jan. 2010 to Apr. 2011. The true positive rate, which is the fraction of true positives out of the total actual positives (true positives and false negatives), was 96.6% (=820 /(820+29)), and there were no false positives. The accuracy, which is obtained by taking the ratio of sum of true positives and negatives to the total, was 98.5% (=(820+1150)/(1999)).

In the application, we tested the proposed anomaly detection using the anomalous data that were generated by making the hydraulic conductivity of the rock fill 100 times more than the original. Due to the sensitivity of the proposed anomaly detection, it is possible that we may not obtain such high accuracy if we do not increase the conductivity as high as we have simulated in this application. However, our result still showed the potential of detecting realistically simulated anomalies that have been generated by the physical model.

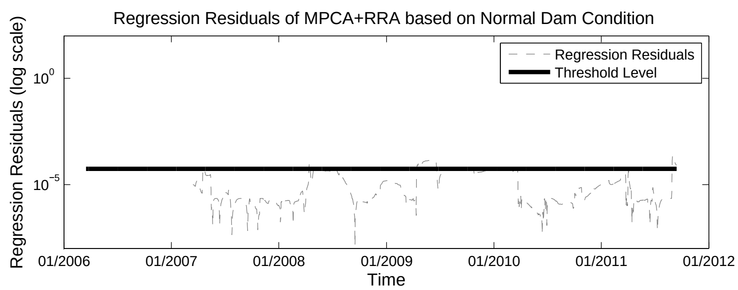

Figure 4. Result on the normal dam condition

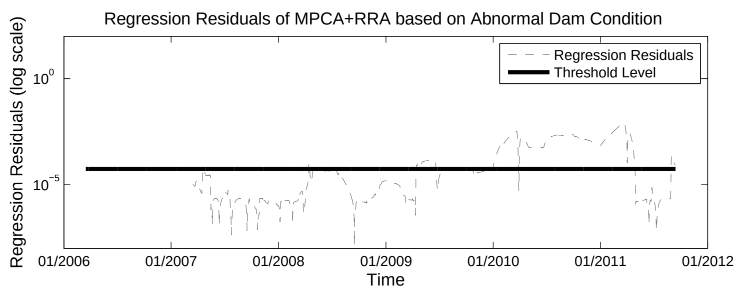

Figure 5. Result on the abnormal dam condition

CONCLUSION

In this paper, we applied the proposed anomaly detection (MPCA+RRA) and detected the simulated anomalies by observing how the relationships between the piezometer readings and the reservoir levels change over time. Since PCA can extract dominant patterns among the data masking any irrelevant patterns, the task of monitoring the performance of embankment dams could become more robust. In addition, such statistical approach reduced the subjective interpretation on the instrumentation data.

Since it is hard to obtain anomalous datasets given that catastrophic events are very rare, anomalies were simulated. In order to simulate a realistic anomalous scenario, we generated anomalous data using a physical seepage model, which would enhance the validation process of our previous work. This approach allowed us to understand how piezometric levels would be reflected by specific physical processes, which have been hard for engineers to recognize proactively especially given that problems often occur inside embankment dams without visual signs.

The most common failure mode of embankment dams is internal erosion, or a piping event, which often occurs from the toe area and develops backwards towards an embankment. Once a piping starts to initiate and eventually develop further, the hydraulic conductivity of affected soil materials would change. Thus, to simulate such anomalies, we changed the hydraulic property of the downstream rock fill. The hydraulic conductivity of the rock fill was increased from 0.002 ft/s to 0.2 ft/s. Then the corresponding piezometer readings were collected.

When the proposed anomaly detection (MPCA+RRA) was applied to the simulated datasets (the reservoir level and the anomalous piezometer readings), the simulated anomalies could be detected with a high accuracy of 98.5%. Thus, we can conclude that changes in hydraulic conductivity, consequently piezometer responses, can be successfully detected using the proposed anomaly detection. Thus, the proposed anomaly detection was validated that has the potential to aid in determining if a piping has occurred or not. In the future, it would be also beneficial to vary values of other parameters as well as degrees of anomalous severities in order to evaluate corresponding detection accuracies and sensitivities of the proposed anomaly detection.

REFERENCES

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

ASCE. (2013). “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.

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

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.

Jung, I., Berges, M., Garrett, Jr., J., and Kelly, C. (2013). “Interpreting the Dynamics of Embankment Dams through a Time-Series Analysis of Piezometer Data Using a Non-Parametric Spectral Estimation Method.” *Computing in Civil Engineering (2013)*, American Society of Civil Engineers, Los Angeles, California, 25–32.

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

Pelton, F. (2000). “Guidelines for Instrumentation and Measurements for Monitoring Dam Performance.”

Schurer, J., Wilkinson, E., Norfleet, J., Sciver, J. V., Huntington, C., Lee, C., and Rogers, A. (2002). “Dam Safety Manual.” State of Colorado.

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.