Interpreting the Dynamics of Embankment Dams through a Time-Series Analysis of Piezometer Data Using a Non-Parametric Spectral Estimation Method

In-Soo Jung[1], Mario Berges[1], James H. Garrett Jr[1], Christopher J. Kelly[2]

[1]Carnegie Mellon University, Department of Civil and Environmental Engineering; 5000 Forbes Avenue, Pittsburgh, PA 15213;
e-mail: {ijung, mberges, garrett}@andrew.cmu.edu
[2]US Army Corps of Engineers, Risk Management Center; 1000 Liberty Avenue, Suite 1107, Pittsburgh, PA 15222; email: christopher.j.kelly2@usace.army.mil

ABSTRACT

A common approach used by engineers to monitor and analyze data collected from piezometers installed in embankment dams is to generate time history plots and visually identify any spikes or anomalies in them. However, such practice has several limitations when capturing complicated relationships among a number of factors that affect piezometric readings. This is especially true when periodic or dominant variations that exist in time-series data are of concern, given that environmental and process noise can sometimes mask these variations. In this paper, we propose applying Moving Principal Component Analysis (MPCA) and Robust Regression Analysis (RRA), which have been shown to be successful in other applications, to extract relevant components and detect anomalies in piezometer measurements, which are one of the most important data to be monitored when evaluating the performance of embankment dams. The proposed anomaly detection method provides a more efficient way of understanding and detecting changes in piezometer data.

INTRODUCTION

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. For example, according to the 2009 ASCE report card for the United States (U.S.) infrastructure, dams received an average grade of D, the lowest grade on that scale (ASCE 2009). Thus, more thorough inspections and immediate efforts are required to assess the condition of the dams and avoid any catastrophic consequences due to dam failures. In this study, we have focused specifically on one type of dams, namely embankment dams. These dams are the most common type of dam in use today (ASDSO and FEMA 2012). The most common aging scenario for embankment dams is internal erosion, which is mainly caused by seepage, and it is usually detected by periodic visual inspection and seepage measurement (USSD 2010). However, since it develops from the inside of dams, it is hard to be detected until it is too late.

When engineers monitor the performance of dams, they review instrumentation data and use additional sources, 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. Piezometers measure the static water pressure at different points along the embankment, and these measurements are generally collected manually on a monthly or daily frequency, or automatically every few hours. When analyzing such data, time-history plots and/or correlation plots are often generated. The relationship between hydrostatic level and reservoir level can be understood using a piezometer hysteresis plot (ASCE 2000). 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.

An example correlation plot is shown in Figure 1. The projection line is generated based on historic high pools, in which their corresponding maximum piezometer readings are included 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.

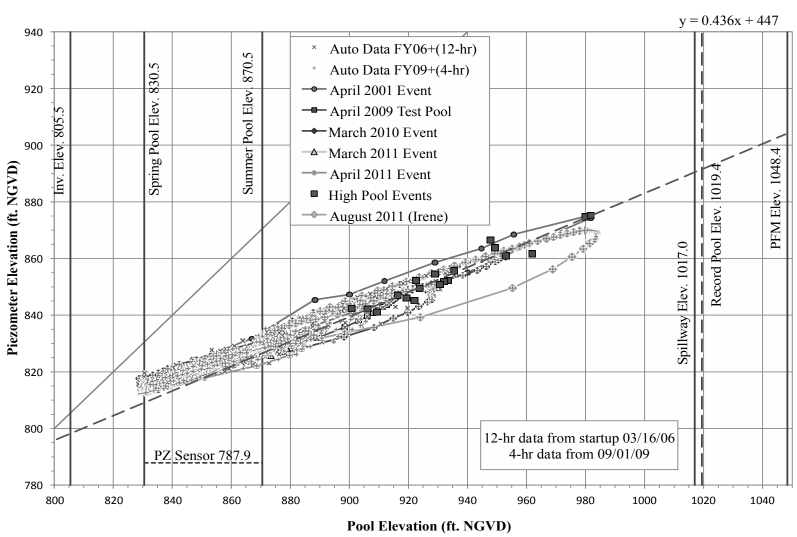

Figure 1. Correlation plot between piezometer and pool elevations

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 (internal erosion), 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.

There have been a number of data-driven anomaly detection methods considered in structural health monitoring (SHM) (e.g., Betti et al., 2006; Worden and Manson, 2007; Ying, 2012). Given the high dimensionality of many SHM datasets, as well as the complex relationships between the variables in the dataset, principal component analysis (PCA) has been widely applied and proved as a promising data analysis tool (e.g., Yan and Golinval, 2006; Yu et al. 2010; Tibaduiza et al., 2012). The main idea of PCA is computing eigenvectors associated with high variability of the data. Since the first few eigenvectors represent the directions of maximum variance or the variance of each independent component, the most dominant variation patterns can be captured. Due to this feature, observing changes in eigenvector structures over time can be adapted as an anomaly detection technique. For example, changes in these eigenvectors can be observed to detect if there have been any structural problems over time.

Many authors have applied this PCA-based method in combination with other statistical approaches for anomaly detection. Wang and Ong (2010), for instance, combined one of the control chart techniques (MEWMA control chart) with PCA to detect structural damage using vibrational response. Loh et al. (2011) have used Singular Spectrum Analysis (SSA), a technique with a similar mathematical basis as PCA, with an autoregressive model to extract the response feature from continuous monitoring of the static deformation of a dam. More recently, Laory et al. (2013) applied Moving PCA (MPCA) in combination with four regression analysis methods for damage detection in bridges. They compared these combined methods with stand-alone methods to see which ones provided highest levels of damage detectability as well as earlier detections.

In this paper, we present the application of 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 is 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. In the next section, we present a brief theoretical explanation of the proposed detection method, followed by a description of the case study dam as well as the application and results.

MOVING PRINCIPAL COMPONENT ANALYSIS (MPCA) AND ROBUST REGRESSION ANALYSIS (RRA)

By 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 $i = 1, \dots, M$ and $t = 1, \dots, N$, 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 $R_i(k)$ at each time value k, where $k = 1, \dots, (N - L)$. Then, a singular value decomposition (SVD) is performed on each one of the covariance matrices, $C = \frac{1}{N} R_i(k)^T \times R_i(k)$. During SVD, the matrix, T 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_l and the diagonal elements of S corresponds to the eigenvalues, e_l, of the matrix (where $l = 1, \dots, L$).

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

Now that the direction of most variability is known for each time step, the next phase is to determine whether this eigenvector changes over time, which would signal the presence of an anomaly. Robust Regression Analysis is known as a good regression technique to use in the presence of outliers. Among many types of robust regression models, we employed the method that uses iteratively reweighted least squares with a bisquare weighting function. So at the end of the proposed anomaly detection, 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.

CASE STUDY

To test the efficacy of this anomaly detection algorithm, we chose a case study dam located in the eastern part of the U.S., for which we had access to instrumentation measurements. 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.

Data validation and preparation. According to currently practiced validation criteria, any sensor measurements that fall out of the range between the top and tip elevations of the piezometers get eliminated. In addition, if there are missing data due to satellite transmission problems, these voids are filled by interpolation between the previous and proceeding readings. In addition, whenever there is no change in values among three consecutive readings, or within 12 hour-period, those readings get validated by engineers. In the application, we used the validated data that have passed the aforementioned criteria. Before September 2009, the piezometer data have been collected every 12 hours, instead of every 4 hours. To unify these two different frequencies, the time-series sampled at 4-hour intervals was down-sampled to match the 12-hour sampling period of the other time-series by selecting every fourth measurement from the 4-hour data.

APPLICATIONS AND RESULTS

The piezometer is the most common instrument used to measure water level on dams (Crum 2011). Besides the reservoir pool, piezometers are influenced by many factors such as precipitation, tail water, pressure, temperature, etc. However, since the main influencing variable to the collected piezometer data is the pool level, we applied MPCA 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 is 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 our approach to detect anomalies as well as to reduce any bias of where the anomalies are introduced, we decided to simulate anomalies to 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. We chose this approach, 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 we have not characterized real anomalies, 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, we only tested on the reordered datasets that have lower correlations than the original one.

The time series data shown in Figure 2 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 2 indicate non-normalized data.

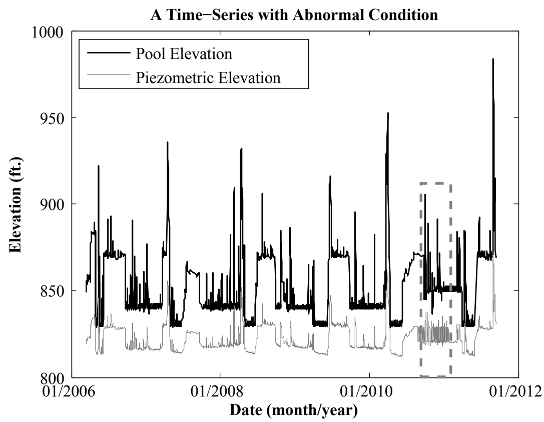

Figure 2. Time-series of the pool and the piezometric elevations with anomalies (dotted rectangle)

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, we observed the changes in the first eigenvectors only. 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. The Receiver Operating Characteristic (ROC) curve of the 13 standard deviations is shown in Figure 3. When the distance between the best possible detection point at (0,1) and each data point in Figure 3 was computed, the standard deviation of 6 had the shortest distance. The true positive rate (TPR) over the six experiments when ± 6 standard deviations were used was 0.82 while the false positive rate (FPR) was 0.12. 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. The contingency tables for all of the 6 experiments using the ± 6 standard deviations are also included in Table 1. As a background check, a dataset of no anomaly was tested also using ± 6 standard deviations, and 1238 false positives and 1957 true negatives were found.

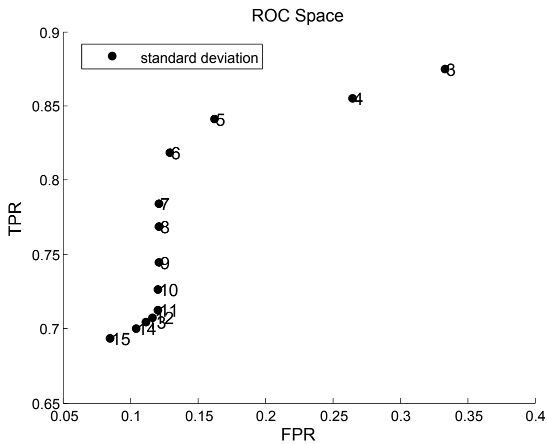

Figure 3. The ROC space from the 13 standard deviations

Table 1. Contingency Tables for the 6 Experiments

Exp.1		Exp. 2		Exp. 3		Exp. 4		Exp. 5		Exp. 6	
TP= 939	FP= 304	860	304	338	304	901	304	934	304	839	195
FN= 41	TN= 1911	120	1911	642	1911	79	1863	46	1911	141	2020

CONCLUSION

In an effort to implement a quantitative and robust approach to monitor the performance of embankment dams based on the piezometer data, we have applied MPCA and RRA 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 has 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 is due to randomly generated anomalies that were introduced in each experiment. Thus, instead of shuffling the original data, it would be also interesting to simulate anomalies that are caused by other scenarios, such as change in trends, movements in materials due to hydraulic loadings, high pool events, etc. In addition, other than observing changes in the relationship between pool and piezometer data, applying MPCA to a group of piezometers that have similar characteristics (e.g., response times, soil layers, etc.), would be useful. If multiple piezometers can be analyzed at the same time, engineers would not need to evaluate every single piezometer with pool data.

REFERENCES

ASCE. (2000). *Guidelines for Instrumentation and Measurements for Monitoring Dam Performance*. ASCE Publications.

ASCE. (2009). *2009 Infrastructure Fact Sheet: Facts about DAMS*. 2009 Report Card for America's Infrastructure, 15–23.

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

Betti, R., Lus, H., Franco, G., Imbimbo, M., and Yu, J. (2006). "Identification of structural damage using dynamic input-output measurements." *Proc. SPIE 6174, Smart Structures and Materials 2006: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems, 61741K*, 61741K–61741K.

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

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.

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).

Loh, C.-H., Chen, C.-H., and Hsu, T.-Y. (2011). "Application of Advanced Statistical Methods for Extracting Long-Term Trends in Static Monitoring Data from an Arch Dam." *Structural Health Monitoring*.

Tibaduiza, D. A., Mujica, L. E., and Rodellar, J. (2012). "Damage classification in structural health monitoring using principal component analysis and self-organizing maps." *Structural Control and Health Monitoring*, n/a–n/a.

USSD. (2010). "The Aging of Embankment Dams." (The Aging of Embankment Dams), 1–11.

Wang, Z., and Ong, K. C. G. (2010). "Multivariate Statistical Approach to Structural Damage Detection." *Journal of Engineering Mechanics*, 136(1), 12–22.

Worden, K., and Manson, G. (2007). "The application of machine learning to structural health monitoring." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851), 515–537.

YAN, A.-M., and Golinval, J.-C. (2006). "An Enhanced Principal Component Analysis for Structural Health Monitoring." *Structural Health Monitoring 2006: Proceedings of the Third European Workshop*, 555–562.

Ying, Y. (2012). "A Data-Driven Framework for Ultrasonic Structural Health Monitoring of Pipes." *Dissertations*.

Yu, L., Zhu, J., and Chen, L. (2010). "Parametric study on PCA-based algorithm for structural health monitoring." *Prognostics and Health Management Conference, 2010. PHM '10.*, 1 –6.