2017 16th IEEE International Conference on Machine Learning and Applications

Anomaly Detection in Earth Dam and Levee Passive Seismic Data using Multivariate Gaussian

Wendy Fisher, Blake Jackson, Tracy Camp
Department of Computer Science
Colorado School of Mines
Golden, CO USA
{wfisher, rbjackso, tcamp}@mines.edu

Valeria V. Krzhizhanovskaya
University of Amsterdam, The Netherlands
St. Petersburg Polytechnic University, Russia
ITMO University, Russia
V.Krzhizhanovskaya@uva.nl

***Abstract*—As earth dams and levees (EDLs) across the United States reach the end of their design lives, effectively monitoring their structural integrity is of critical importance. This paper investigates automatic detection of anomalous events in passive seismic data as a step towards continuous real-time monitoring of EDL health. We use a multivariate Gaussian machine-learning model to identify anomalies in experimental data from two different laboratory earth embankments. Additionally, we explore five wavelet transform methods for signal denoising; removing different signal components. The best performance is achieved with the Haar wavelets (removing the Level 3 component). We achieve up to 97.3% overall accuracy and less than 1.4% false negatives in anomaly detection. These promising approaches could eventually provide a means for identifying internal erosion events in aging EDLs earlier than is currently possible, thereby allowing more time to prevent or mitigate catastrophic failures.**

I. Introduction and Motivation

Earth Dams and Levees (EDLs) are essential for flood control, water storage, and irrigation. They are built using earthen materials such as rock, sand, and clay [1]. Many earth dams in the U.S. are over 60 years old and reaching the end of their design lives [2]. Thus, it is necessary to find ways to monitor the stability of these structures. The main causes of EDL failures are piping, slope instability, foundation issues, and overtopping [3] and internal erosion is actually the root cause of many of these failure mechanisms. Currently, failure detection methods rely on visual inspection of the surface of an EDL by human experts, which does not guarantee finding a problem early enough to prevent collapse. Figure 1 shows the result of a catastrophic failure of the Algodes I Dam in Brazil [4] during the Spring of 2009, which was previously deemed safe by engineers. A reported 50-meter hole opened in the dam, which submerged the city of Cocal da Estacão, with 30,000 inhabitants, under 20 meters of water. The results included more than seven fatalities, power outages, roads being completely washed away, and the destruction of more than 500 homes, which left more than 3,000 people homeless. Dam failure is a worldwide problem and early detection is critical. Our study focuses on using geophysical sensor technologies and machine learning methods to detect anomalous erosion events before they progress to failure.

We have investigated anomaly classification using unsupervised clustering algorithms [5] and the ability to separate normal from anomalous data observations using support vector machines [6]. Now, we take a step forward to an anomaly detection approach that could be used for the continuous real-time monitoring of earth dams and levees.

Fig. 1: The Algodes I Dam, Piaui, Brazil after it burst (top), the city of Cocal that was flooded (middle), and the aftermath including roads completely washed away (bottom) (figure adapted from [4]).

II. Background and Related Work

Anomaly detection aims to identify data observations that differ from the normal or expected pattern. Our goals necessitate the development of an anomaly detection scheme that can perform well despite the limited availability of ground-truth information and the large class imbalance present in many EDL datasets. With our multivariate Gaussian approach, we can train models on data that represent the normal state of the levee and subsequently detect deviations within a threshold; thus eliminating the need for levee specific labeled data.

Chandola et al. [7] separate the different anomaly detection techniques into: classification, clustering, nearest neighbor, statistical, information theoretic, and spectral. The various methods output either scores to select anomalies using an appropriate threshold or labels to identify anomalies versus the normal data observations. Anomaly detection has been used to investigate many real-world problems, such as fraud detection in finances [8] and insurance [9], intrusion detection in computer networks [10], image analysis in motion [11] and hyperspectral [12], and bioinformatics cancer gene analysis [13]. Our scheme uses a novel statistical-based anomaly detection workflow to differentiate normal seismic signals from anomalous internal erosion events in EDL passiv

0-7695-6321-X/17/31.00 ©2017 IEEE
DOI 10.1109/ICMLA.2017.00-81

Researchers from the University of Amsterdam have detected anomalies in earth levees from sensors installed inside the structures (e.g., temperature, pore water pressure, relative inclination) using a neural-cloud approach [14]. Mississippi State University researchers experimented with unsupervised and supervised methods to detect anomalies [15] and classify levee slides in synthetic aperture radar data [16], [17] along the Mississippi River using support vector machines. However, to our knowledge, our research is the first that uses machine learning for the automatic detection of anomalous internal erosion events in passive seismic EDL data. Our approach investigates detecting anomalous erosion or crack events using geophysical data collected from small sensors installed on the surface of the levee. By using surface sensors rather than sensors inside of the dam, we can implement an early warning system without damaging the integrity of the EDL structure.

III. Experimental Data

Two different passive seismic data sets from experimental earth embankments built at the United States Bureau of Reclamation (USBR) [1] were used during our experiments. The two data sets included different types of internal erosion events (e.g., cracking and piping).

A. Crack-Box Testbed

The crack-box data set is described in detail in previous work [5]. The data was collected from an experimental laboratory earth embankment that was built at the USBR and brought to failure to study internal erosion and cracking [18]. A hinged joint at the bottom centerline of the structure was used to induce the cracking in 0.5 cm and 1.0 cm increments. Figure 2 shows the initial state of the structure and the result of a 2.5 cm crack. The embankment was equipped with various geophysical instrumentation including a vertical array of 12 geophones that collected data at 500Hz for several days.

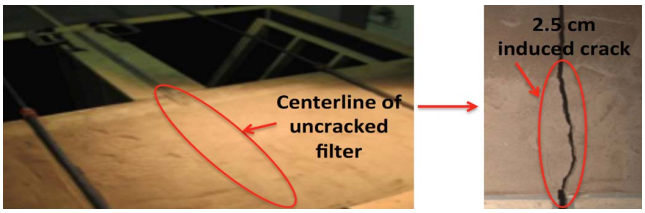

Fig. 2: Images from laboratory earth embankment structure built at the USBR (left) and result of a 2.5 cm induced crack (right).

We were provided with 4,140 seconds of passive seismic data for our experimentation that includes many events of crack, pump, and flow (30% normal or baseline activity and 70% anomalous events). We use the first third of our data for training our model since it is more representative of real-world conditions, with 80% normal data and 20% anomalous data.

B. Piping Experiment

The second data set is from an internal erosion and piping experiment also conducted at the USBR and described in detail in [19]. The structure was equipped with geophysical instrumentation and built with a long rebar embedded on the lower left side that was later removed to induce the piping event. Figure 3 shows a cross-sectional view of the experimental levee and piping activity after the rebar was removed and subsequent result of erosion activity.

Fig. 3: Experimental earth embankment built at the USBR showing cross-section result of a piping induced by the removal of the rebar (left) and result of significant erosion of materials from piping (right).

The data provided was sampled at 500Hz and included several baseline testing and noise events (e.g., setup of sensors, adjustment of equipment, hammer strikes, nearby machinery, and pumping of water into the reservoir). We focus on the events after the reservoir was filled and use the section of data that closely mimics a real-scenario of the initiation of a pipe, major water flow, and subsequent internal erosion events (80% normal or baseline data and 20% anomalous data).

There are several crack events clearly visible from the raw signal data. Despite this, the naive approach of simply flagging any large spike in amplitude as anomalous cannot be sufficient in this application for several reasons. First, we must keep in mind that this data was collected in artificial conditions, and, while there is some noise in the signals, we expect even more noise in real conditions (as the dams are exposed to tidal changes, weather, foot and vehicular traffic, and other irrelevant sources of vibration). Such circumstances are unavoidable, and our early warning system should view them as normal. Second, we seek to detect not only violent crack events, but also internal erosion events like piping (as seen in the second data set). These do not necessarily generate amplitude spikes in the signal data, especially during their incipient stages, but we still aim to detect them as early as possible. For these reasons, we consider features other than just amplitude in producing our model (see Section IV-C).

IV. Preprocessing

There are preprocessing steps needed before we can use the data as input for our machine learning approach. We first reduce the amount of noise in the data, then we divide the time-series data into smaller sections or frames, extract our features, standardize the data, and perform dimensionality reduction.

A. Noise Reduction

Our data sets include baseline or normal activity, crack or piping events, and many spurious noise events commonly found in passive seismic data. We explore wavelet denoising to help make a distinction between noise in the data and true anomalous events. Since the 1990s, wavelets have provided a powerful tool for denoising a variety of signals [20]. Decomposing a signal using a discrete wavelet transform yields

a set of wavelet coefficients that correspond to high frequency subband details within the signal. Small wavelet coefficients pertain to relatively minor details within the signal that are often created by noise. By setting all wavelet coefficients below some threshold to zero and then reversing the wavelet transform, we eliminate noise in our data [21].

We use a maximal overlap discrete wavelet transform (MODWT), which differs from the discrete wavelet transform in that it is highly redundant and nonorthogonal. This means that, among other advantages, the MODWT is well defined for all sample sizes, wavelet coefficients are not influenced by circular shifting of the input signal, and wavelet coefficients at each level align with the original signal [22]. We test various transform levels of the MODWT denoising process with Haar, Daubechies 2 (db2), Symlets 6 (sym6), Coiflets 4 (coif4), and Fejer-Korovkin 8 (fk8) wavelets and evaluate the efficacy of each in our results section.

B. Spectral Framing

We use the entire 4,140 seconds of crack-box time-series data (2,070,000 samples) and 2,460 seconds (1,230,000 samples) of the piping data set, both from a single sensor collected at 500Hz. We divide the data into frames (or segments) using spectral frame decomposition with MATLAB and the open-source MIRToolbox [23]. We have experimented with various frame-sizes and found similar performance for these data sets using 1, 2, 3, 5, and 10-second frames. Figure 4 illustrates a zoomed in portion of the segmented crack-box data using 3-second frames (the frame size used throughout this paper).

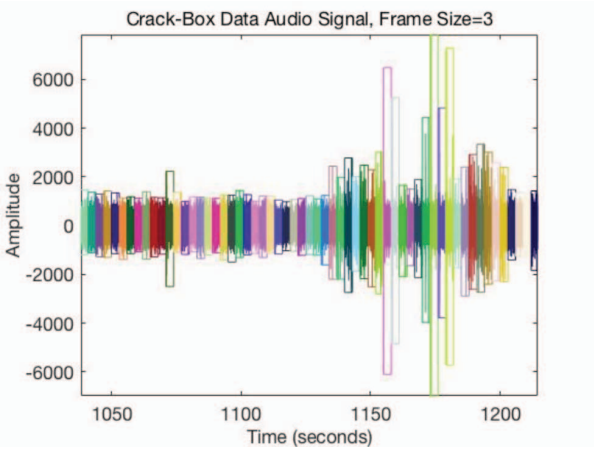

Fig. 4: A zoomed-in portion of the crack-box time series data showing the segments of a 3-second spectral frame decomposition.

C. Feature Extraction

The features selected to best represent our passive seismic data events are those commonly found in audio signal processing. We leverage our previous work to reduce the choices and extract the nine temporal, statistical, and spectral features described in [6] from each of the 3-second frames of data. These features are zerocross, centroid, spread, root mean square, rolloff, flatness, kurtosis, irregularity, and skewness.

D. Standardization

In our experiments, we used Equation 1 to standardize the feature values. Data standardization is common practice for transforming each of the feature vectors to have a zero mean (μ) and unit standard deviation (σ).

$$X' = \frac{X - \mu}{\sigma} \tag{1}$$

E. Dimensionality Reduction

In previous research, we used all nine extracted features with unsupervised clustering [5] and two-class support vector machines [6]. We further investigated using the top selected features from the ReliefF algorithm [24] with one-class support vector machines [6]. In this experiment, we applied the ReliefF method after the wavelet denoising and found that the top resultant features for these data sets are zerocross, spread, and the root mean square of the energy (described in Table I).

TABLE I: Top features of the nine spectral features extracted from each 3-second frame used in our anomaly detection algorithm.

Feature	Description
Zerocross (ZC)	A temporal feature which is a count of the number of times the spectral signal crosses the zero axis (changes sign).
Spread (SP)	A spectral feature that represents the standard deviation of the distribution.
Root Mean Square (RMS)	A temporal feature that represents the global energy (loudness) of the signal; root average of the square of the amplitude.

The top resultant features are not surprising given that the seismic anomalies that we seek to detect (e.g. cracking) are relatively violent in comparison to normal circumstances. Therefore, as the sensors are being jarred and generally experiencing more movement, we intuitively expect the signal to change sign more rapidly and be more peaked than usual. Furthermore, the fact that these events correspond to changes in the earthen structure that are not subsequently reversed suggests that there might be less symmetry in the anomalous portions of our signal data.

V. Anomaly Detection

One advantage of using a multivariate Gaussian approach is the ability to take all of the feature values into account at once instead of considering them one at a time and then combining the results [25]. The multivariate Gaussian may catch anomalies that a single system won't, especially if the features have some correlation with each other (i.e., the feature values have some dependencies and, thus, we should model $p(x)$ using the entire feature set). The multivariate Gaussian approach also works well when the number of normal data observations is not terribly skewed, which is true in our application.

Once the features are selected and preprocessed, we use the first 30% of the data to train our model (X_{trn}) and the remaining 70% for validation (X_{val}). We calculate the parameters of the Gaussian distribution of the training data set (the mean (μ) and the variance (σ^2) for X_{trn}). We then

compute the probability density function, $p(x)$, using the μ and σ^2 calculated for X_{trn}. Then, $p(x)$ is used to predict the probability that the next data observation is normal (as opposed to anomalous). Using a threshold value for ϵ found during the cross validation step, we test new observations to determine their probability of belonging to the normal class. In other words, using Equation 2, if

$$p(x_{new}) < \epsilon, \tag{2}$$

then the new observation is a potential anomaly.

VI. Results and Analysis

Using the multivariate Gaussian approach, Haar wavelet denoising, and 3-second frames, we achieved up to a 97.3% overall accuracy with the top features selected by the ReliefF algorithm for both of our data sets. After the preprocessing and denoising of the data, the model was trained and validated using the passive seismic data as described in Section III.

A. Crack-Box Data Results

The contour plot in Figure 5 shows the results with the data observations as black dots and the anomalies encircled in red. The lines of the contour represent the estimated probability that an observation at that level is normal.

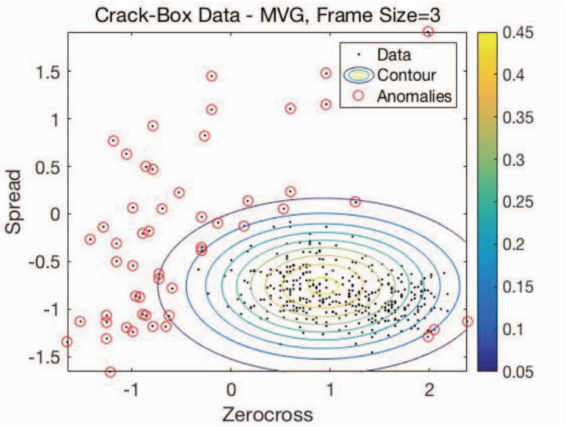

Fig. 5: Contour of the potential anomalies in our training data. Higher values represent normal data and values that fall below the threshold are considered anomalous. The anomalies selected with threshold $\epsilon = 0.0979$ are located outside the dark blue line (second to outermost).

Figure 6 is a plot of the resultant probabilities (using $\epsilon = 0.0979$) where values below ϵ are considered anomalous. Using known ground truth labels, we analyze the results and provide statistics to measure the performance of our anomaly detection approach. The results from our experiment with a 3-second frame size achieved a 97.3% overall accuracy and had relatively few misses (1.3%) and false alarms (1.3%).

We repeat our experiments to confirm that the multivariate Gaussian technique is effective with various training sizes. Our results are in Table II. The overall accuracy remains above 96% while maintaining a relatively low number of false negatives ($< 1.6\%$). Notice that the performance begins to degrade as we move past using the first 30% to train due to the presence of more anomalous observations in the training data.

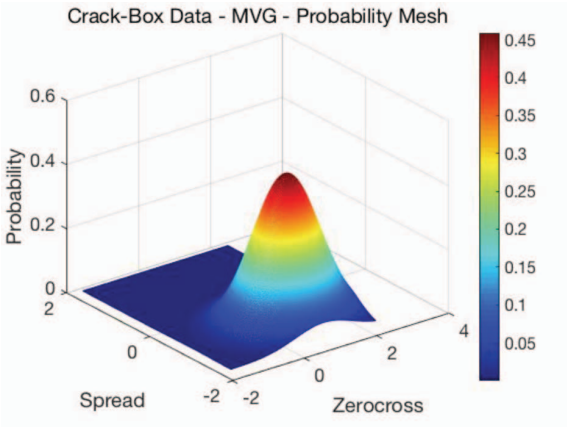

Fig. 6: Probabilities calculated using training data show values below $\epsilon = 0.0979$ are potential anomalies in the crack-box data.

TABLE II: Results of repeated runs of our data-driven workflow using a varied training percentage. The training percentages represent the first % of the time-series data.

Training %	Accuracy	TP	TN	FP	FN
20%	97.1	10.7	86.4	1.4	1.4
25%	97.2	4.6	92.6	1.2	1.6
30%	**97.3**	**2.7**	**94.6**	**1.3**	**1.3**
35%	97.0	3.1	93.9	1.8	1.2
40%	96.0	3.9	92.2	3.1	0.8

B. Piping Data Results

To further validate our approach, we applied the multivariate Gaussian technique to a second data set with a different type of internal erosion event (piping). The contour plot in Figure 7 again shows the results with the data observations as black dots and the anomalies encircled in red. The lines of the contour represent the estimated probability that an observation at that level is normal.

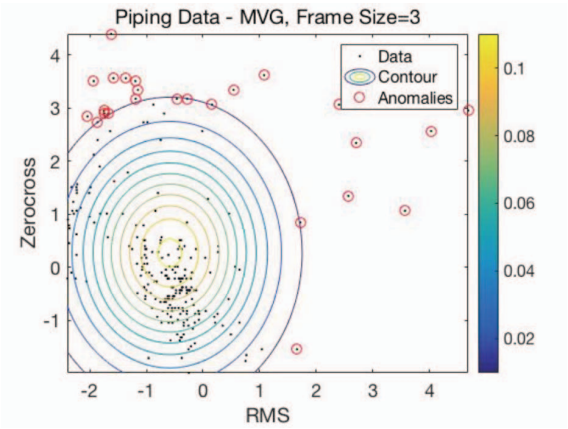

Fig. 7: Contour of the potential anomalies in our training data. Higher values represent normal data and values that fall below the threshold are considered anomalous. The anomalies selected using $\epsilon = 0.0113$ are located on or outside the dark blue line (outermost).

Figure 8 is a plot of the resultant probabilities where values below $\epsilon = 0.0113$ are considered anomalous. The results from our piping experiments using a 3-second frame size reveal an overall accuracy of 97.2%, a low number of misses (0.3%), and few false alarms (2.4%).

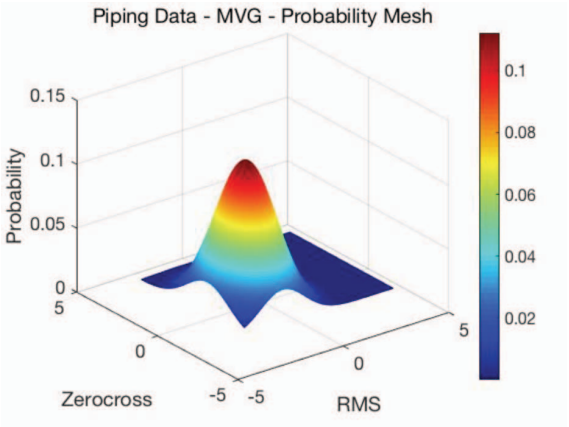

Fig. 8: Resultant probabilities calculated using training data show values below $\epsilon = 0.0113$ are potential anomalies in the piping data.

C. Validation Metrics

We continue our validation using several metrics commonly used with classification experiments, which are percentages representing ratios between the statistics found in the confusion matrix. These values are given and described in Table III.

TABLE III: Evaluation metrics typically used for the validation of machine-learning classifiers with results from our experiments with a 3-second frame size and Level 3 Haar wavelet denoising.

Metric	Description	Crack Box	Piping
Accuracy	Correctly classified events.	97.3%	97.2%
Specificity	Proportion of anomalous events correctly identified.	98.6%	84.1%
Precision	Proportion of events classified as normal that were truly normal.	67.5%	97.2%
Recall	Proportion of normal events correctly identified.	67.5%	99.6%
F1-score	Harmonic mean of precision and recall.	67.5%	98.4%

Both data sets achieved a high overall accuracy of over 97% with the piping data set producing a high F1-score of 98.4%. We hypothesize the increase in precision and recall between the crack-box data and the piping data could be due to the large amount of anomalous events present in the validation set of the crack-box data. The specificity of 98.6% for the crack-box data indicates that the model correctly identifies anomalous circumstances in the overwhelming majority of cases. Furthermore, a comparatively low recall, in this application, is better than decreasing specificity because it is safer to have the system trigger a false alarm than to accidentally ignore an internal erosion event. This trade-off between recall and specificity is elucidated in the ROC curves, which are discussed in detail in the next section. The ROC curve shows the relationship between the false positive rate on the x-axis (1 - specificity) versus true positive rate on the y-axis (sensitivity or recall), with values in the upper left hand corner (0,1) denoting a perfect classification. Unlike other common statistics that show results using only one threshold, the ROC curve shows performance using all possible thresholds.

D. Wavelet Analysis

We now provide results of our wavelet analysis using the crack-box data set; we observed similar results using the piping data sets. As shown in Figure 9 and Table IV, we tested the MODWT denoising process with several analyzing wavelets. Previous work showed we could separate normal from anomalous data observations with 86.5% accuracy using a one-class support vector machine and 3-second frames [6]. With our new method, the accuracy increased by over 9.8%. Adding the Haar wavelet denoising, we see an additional increased accuracy from 96.3% to 97.3% and an improvement for both true and false classification rates. Denoising using the other wavelets decreased the accuracy, presumably because some meaningful signal data was discarded with the noise.

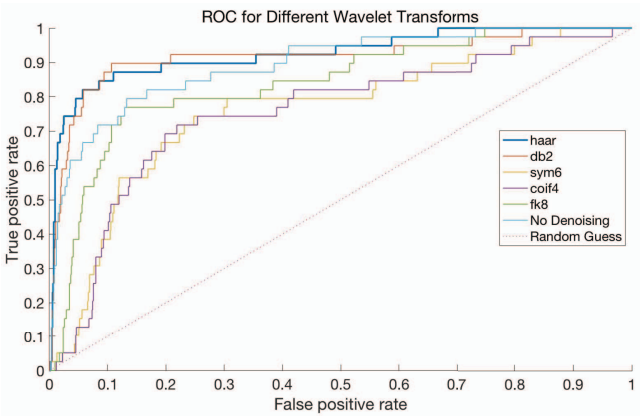

Fig. 9: ROC curves for the best performing level for a given wavelet (reference Table III). The Haar wavelet produces the highest overall result and is selected for our anomaly detection scheme.

TABLE IV: Performance impact of wavelet denoising.

Wavelet	Best Level	Accuracy	TP	TN	FP	FN
Haar	**3**	**97.3**	**2.7**	**94.6**	**1.3**	**1.3**
db2	3	95.6	2.9	92.7	1.1	3.3
sym6	1	86.7	2.3	84.4	1.8	11.6
coif4	1	87.7	2.0	85.7	2.1	10.2
fk8	1	92.5	2.2	90.3	1.9	5.7
None	-	96.3	2.0	94.3	2.1	1.7

We also tested various levels of denoising with each wavelet. These results are shown for the Haar wavelet in Table V. We see that level 3, which corresponds to a MODWT of a signal down to scale 2^3, yields the greatest performance increase. At levels 4-7, we again hypothesize that meaningful signal data was erroneously discarded as noise. These results

suggest that there are potential benefits to wavelet transform denoising in this application, and that one must be careful in selecting the level of denoising.

TABLE V: Performance impact of various levels of Haar wavelet transform denoising.

Level	Accuracy	TP	TN	FP	FN
1	93.9	3.1	90.8	0.9	5.2
2	96.8	2.5	94.3	1.6	1.7
3	**97.3**	**2.7**	**94.6**	**1.3**	**1.3**
4	54.5	2.9	51.6	1.1	44.4
5	16.7	4.0	12.7	0.0	83.2
6	67.5	3.1	64.4	0.9	31.5
7	32.4	3.1	29.3	0.9	66.7

VII. Conclusions and Future Work

We present a step towards continuous real-time EDL health monitoring using an anomaly detection approach. Our experiments with passive seismic data, collected from two different laboratory earth embankments, show promising results. We applied wavelet transform denoising techniques, extracted spectral features to represent our data, and used a multivariate Gaussian anomaly detection technique. Results indicate the ability to differentiate between normal and anomalous data observations with up to 97.3% accuracy. We plan to continue our research towards a generalized workflow that can be used to detect anomalous activity with any type of EDL. To fully evaluate our approach, we will experiment with additional test and real-world data. We have passive seismic data from the IJkdijk full-scale test embankment located in the Netherlands that was constructed to study seepage and internal erosion [18] [26]. We have also been provided seepage and erosion data from the real-world Colijnsplaat levee in the Netherlands [27] collected by a team of geoscientists. It is our goal to develop a system for the identification of internal erosion events in aging EDLs early enough to prevent or mitigate catastrophic failures.

Acknowledgment

This work is supported in part by the National Science Foundation Grant OISE-1243539. We also thank Justin Rittgers and Minal Parekh, former Ph.D. students at the Colorado School of Mines, for their assistance in collecting the data used in this study.

References

[1] "US Bureau of Reclamation," www.usbr.gov, accessed: 2016-10-1.

[2] "Aging water resource infrastructure in the United States," www.usbr.gov/newsroom/testimony/detail. cfm?RecordID=2441, accessed: 2016-10-1.

[3] M. Foster, R. Fell, and M. Spannagle, "The statistics of embankment dam failures and accidents," *Canadian Geotechnical Journal*, vol. 37, no. 5, pp. 1000–1024, 2000.

[4] "Algodes I Dam, Piau, Brazil," www.internationalrivers.org, accessed: 2016-10-1.

[5] W. Belcher, T. Camp, and V. V. Krzhizhanovskaya, "Detecting erosion events in earth dam and levee passive seismic data with clustering," *Proceedings of the 14th International Conference on Machine Learning and Applications (ICMLA)*, pp. 903–910, 2015.

[6] W. Fisher, T. Camp, and V. Krzhizhanovskaya, "Crack detection in earth dam and levee passive seismic data using support vector machines," *Procedia Computer Science*, vol. 80, pp. 577–586, 2016.

[7] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys (CSUR)*, vol. 41, no. 3, pp. 1–72, 2009.

[8] E. Ngai, Y. Hu, Y. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decision Support Systems*, vol. 50, no. 3, pp. 559–569, 2011.

[9] T. Ormerod, N. Morley, L. Ball, C. Langley, and C. Spenser, "Using ethnography to design a mass detection tool (MDT) for the early discovery of insurance fraud," *Extended Abstracts on Human Factors in Computing Systems (CHI)*, pp. 650–651, 2003.

[10] C.-H. Lin, J.-C. Liu, and C.-H. Ho, "Anomaly detection using LibSVM training tools," *International Conference on Information Security and Assurance (ISA)*, pp. 166–171, 2008.

[11] Z. Fu, W. Hu, and T. Tan, "Similarity based vehicle trajectory clustering and anomaly detection," *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, vol. 2, pp. II–602, 2005.

[12] C.-I. Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 40, no. 6, pp. 1314–1325, 2002.

[13] J. W. Macdonald and D. Ghosh, "COPA - cancer outlier profile analysis," *Bioinformatics*, vol. 22, no. 23, pp. 2950–2951, 2006.

[14] A. L. Pyayt, A. P. Kozionov, I. I. Mokhov, B. Lang, R. J. Meijer, V. V. Krzhizhanovskaya, and P. M. Sloot, "Time-frequency methods for structural health monitoring," *Sensors*, vol. 14, no. 3, pp. 5147–5173, 2014.

[15] L. Dabbiru, J. V. Aanstoos, M. Mahrooghy, W. Li, A. Shanker, and N. H. Younan, "Levee anomaly detection using polarimetric synthetic aperture radar data," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 5113–5116, 2012.

[16] L. Dabbiru, J. V. Aanstoos, and N. H. Younan, "Earthen levee slide detection via automated analysis of synthetic aperture radar imagery," *Landslides*, pp. 1–10, 2015.

[17] D. Han, Q. Du, J. V. Aanstoos, and N. Younan, "Classification of levee slides from airborne synthetic aperture radar images with efficient spatial feature extraction," *Journal of Applied Remote Sensing*, vol. 9, no. 1, pp. 097 294–1–097 294–10, 2015.

[18] M. A. Mooney, M. L. Parekh, B. Lowry, J. Rittgers, J. Grasmick, A. R. Koelewijn, A. Revil, and W. Zhou, "Design and implementation of geophysical monitoring and remote sensing during a full scale embankment internal erosion test," *Proceedings of the GeoCongress*, 2014.

[19] W. D. Fisher, T. K. Camp, and V. V. Krzhizhanovskaya, "Anomaly detection in earth dam and levee passive seismic data using support vector machines and automatic feature selection," *Journal of Computational Science*, vol. 20, pp. 143–153, 2017.

[20] R. Cohen, "Signal denoising using wavelets," *Project Report, Department of Electrical Engineering Technion, Israel Institute of Technology, Haifa*, 2012.

[21] "Signal denoising," eeweb.poly.edu/iselesni/Double Software, accessed: 2016-10-1.

[22] F. Vallianatos and G. Hloupis, "HVSR technique improvement using redundant wavelet transform," pp. 117–139, 2007.

[23] O. Lartillot and P. Toiviainen, "A MATLAB toolbox for musical feature extraction from audio," *Proceedings of the 10th International Conference on Digital Audio Effects*, 2007.

[24] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine Learning*, vol. 53, no. 1-2, pp. 23–69, 2003.

[25] "Coursera: Machine learning," www.coursera.org, accessed: 2016-10-1.

[26] J. Rittgers, A. Revil, T. Planes, M. Mooney, and A. Koelewijn, "4-D imaging of seepage in earthen embankments with time-lapse inversion of self-potential data constrained by acoustic emissions localization," *Geophysical Journal International*, vol. 200, no. 2, pp. 756–770, 2015.

[27] "Smart levee guideline," www.smartlevee.nl/projects-and-cases/totaallijst/colijnsplaat, accessed: 2016-10-1.