Statistical distance analysis for seepage reconnaissance at Mactaquac dam

Emanuel de Gante
UNB ECE 4553
University of New Brunswick
Fredericton, Canada
emanuel.degante@unb.ca

I – INTRODUCTION

Mactaquac generating station is a hydroelectric facility located on the Saint John River 20 kilometers upstream of Fredericton and has been operating since 1968. Since the 1980s concrete portions of the hydro station have been affected by a chemical alkali aggregate reaction causing the concrete to swell and crack requiring substantial annual maintenance and repairs. NB Power is proposing a project to ensure the station can operate to its intended 100 year lifespan with a modified approach to maintenance [1].

All dams are designed and constructed to allow some seepage from the headpond to the downstream toe. Seepage becomes a concern when it leads to internal erosion, where statistical analyses [2] indicate that about 50% of embankment dam failures in the world before 1999 were due to internal erosion. The location of the internal erosion may occur within the embankment, in the foundation or from the embankment into the foundation.

Seepage is the infiltration of water from the upstream reservoir of the dam towards the downstream face. Water flow in porous media is governed y the hydraulic conductivity of the material and the hydraulic head gradient. Heat is transported through porous media primarily by conduction and convection, with convection being enhanced in regions with concentrated seepage flow. Thus heat can be used as a tracer and temperature analysis can be used as a method for seepage detection and monitoring [3]. Temperature monitoring can detect anomalous temperature variations triggered by increased seepage flow [4].

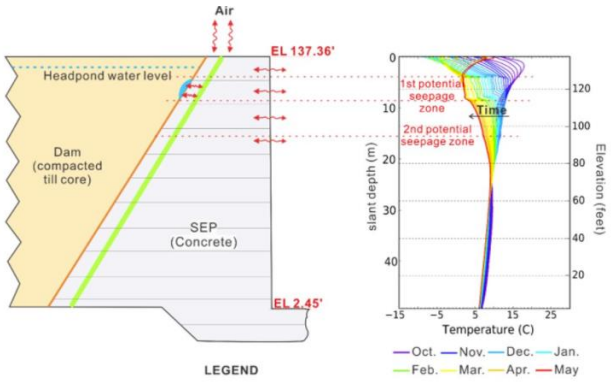


Figure (1) Fiber Optic installation at Mactaquac and temperature profiles.

Fibre Optic system have been installed in embankment around the world to monitor temperature and detect leaks or concentrated seepage along the embankment.

There is a seasonal anomaly detected between an elevation of 115’ and 125’. Visual inspections at the Mactaquac Generating Station in recent years also show some evidence of potential seepage near the dam/concrete interface.

For example, vegetation on the downstream slope of the embankment in the vicinity of the interface was observed in November 2014.

II – LITERATURE REVIEW

There are two main methods to use fiber optics for seepage recognition, the passive method and the active method. The passive method requires several months or years to adequately capture the seasonal temperature where reservoir water temperature oscillations do not propagate deep into the dam, and the temperature fluctuations within the dam should be relatively stable especially at distances far from the reservoir however when anomalous seepage occurs, temperature anomalies will be transported into the dam structure by means of convection and the normal temperature will be distorted. The magnitude and velocity of seepage flow may be estimated by means of the time lag and intensity of the temperature anomaly observing seasonal temperature variations [5][6]. In contrast the active method combines heating together with temperature measurement where temperature data are collected before and during the heating process. Leakage zones can then be detected by comparing the temperature measurements before heating and at the peak of heating because zones with higher seepage usually have higher heat loss leading to a lower peak temperature.

One method for temperature analysis using a passive method is based on the estimation of physical parameters associated with seepage is The Impulse Response Function Thermic Analysis [7], this method allows for seepage identification and its intensity estimation. The measured temperature is regarded as a superposition of the responses related to the reservoir water and the air temperatures, while other thermal contributions like geothermal and freezing processes, radiation and wind influence are neglected and the heat transport mechanisms of conduction and convection are analyzed.

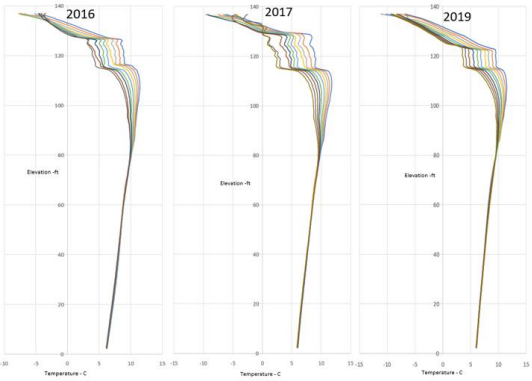
Usually, the data collected by the DTS/passive method in dams toe cannot be used directly for seepage detection. Many other environmental factors may influence the acquired temperature data, such as drainage pipes, ground heterogeneities, precipitation, and the temperature difference between seepage water and the ground in which the fibre cable is buried. Khan [8] introduced a source separation method to separate different thermal contributions or sources from the raw temperature data. Decomposition of the raw temperature data can be performed using singular value decomposition (SVD) and independent component analysis (ICA) based on the assumption that the sources are mutually independent. Both methods were validated on two different sites equipped with fibre optic temperature sensors along a dam, whereby anomalously increased seepage was detected.

III – METHODS

The proposed analysis have not being developed or published yet. One main difference in previous works is that most of the fibre optic cables are installed along the dam toe covering in some instances kilometers of installed cable, in contrast the DTS cable at Mactaquac is installed in the core of the embankment covering an elevation equivalent to 135 feet with the equivalent of 284 temperature measuring nodes distributed in the temperature system. There is data recorded within a timeframe comprehending between January 2014 to November 2021, but there have been major interruption during this time being the largest occurring in early 2018 where a major maintenance event occurred and the DTS cable was retrieved from the borehole. 48 temperature measurements are recorded each day with a sampling rate of every 30 minutes.

A. Subjects and Datasets.

The data is recorded in .ddf extension containing: date and time of the measurement event, temperature measurement every 0.5 ft along the total length of the cable, stokes and anti stokes from raman spectroscopy to calculate temperature and the calibrated temperature from the stokes and anti stokes. After this data processing step Data is transformed into a .h5 file containing cable length, sample time based on average measurement time for fata management purposes that can be divided into hourly, daily or weekly sampling time frame. The final data processing step is to convert into a .csv file containing the time stamp (it can be weekly, daily or hourly time frame), each measuring node aligned to the elevation of the dam and average temperature data for selected time frames.



Figure(2) Temperature vs elevation profiles at Mactaquac.

The main anomaly is observed during the early stages of winter from January to March. For consistency of the analysis presented in this paper temperature profiles of the year 2016, 2017, 2019 and 2021 have been selected for analysis where a weekly average time frame is selected from the beginning of January to the first week of March comprehending 10 weeks temperature analysis per year.

No dimensionality reduction technique is used because all the data presented is needed to perform the analysis, there is no course of dimensionality in this data set.

In order to perform the analysis of data will be normalized.

$$z = \frac{x-\mu}{\sigma} \qquad (1)$$

Where z is the score, μ is the mean of the data and σ is the standard deviation and x is the data to be converted.

Once normalized all the data, statistical distances will be used to measure how probability distribution differ one from another. And detect abrupt changes in divergences

Kullback-Liebler Divergence and Bhattacharyya distance will be calculated and will be compared to visualize changes in consecutive temperature probability distributions based on the elevation profile and temperature behavior in time.

B. Kullback-Liebler Divergence.

Kullback-Liebler Divergence also called relative entropy is used to measure how probability distribution is different one from another and is described by the following equation:

$$D_{KL}(p,q) = \sum_x p(x)\, log\, \frac{p(x)}{q(x)} \qquad (2)$$

Exploring the temperature measurements in each elevation node and having p(x) as the probability distribution of the first node, p(x) will be compared to q(x) which represent the next consecutive elevation node of the total elevation. This process will be iterated to get $D_{KL}(p,q)$ for all of the consecutive probability distribution of the elevation nodes. And a KL-Divergence vs elevation plot will be generated to analyze the relation of the differences.

C. Bhattacharyya Distance

Bhattacharyya Distance measures the similarity of two probability distributions (4). This is related to the Bhattacharyya Coefficient (3) that measures the amount of overlap between two statistical samples.

$$C_B(p,q) = \int \sqrt{p(x)q(x)} \qquad (3)$$

$$D_B(p,q) = -\ln\{C_B(p,q)\} \qquad (4)$$

Just as the KL-Divergence after the calculation of $C_B(p,q)$ and consequently $D_B(p,q)$, a plot of elevation vs $D_B(p,q)$ will be developed for all the consecutive elevation points for all years selected for this analysis

Once selected the zone of interest that can be characterized between two peaks of divergence or abrupt changes in statistical distances.

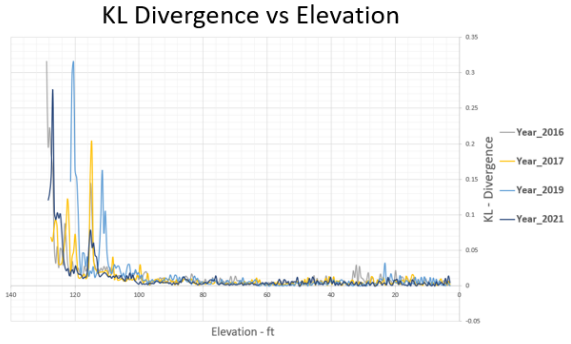
D. Construction of the Classifier

Once statistically detected the anomaly zones, the data set will be manipulated to label classes of Anomalous versus Non-anomalous temperature behavior based on the elevation of temperature profile. The classifiers will be the base to detect possible extensions of the temperature anomaly between the average range of 125' and 115' of elevation

Three classifier algorithms are tested to measure the accuracy: Support Vector Machines, K-Nearest Neighbor and Quadratic Discriminant Analyser. KNN and QDA are selected because they are generative models and SVM may provide a more accurate classification region for the anomalous zone compared to the other classifiers selected. In order to build these classifiers, the first week of the year will be used as the base feature to compare with the other features represented by the other 9 weeks selected where the anomaly is present with the purpose to provide more data density in each selected timeframe. To have a binary classification model the comparison of base temperature vs consecutive temperature the array will be concatenated at the bottom end of each array where each point will represent a measuring node within a specific comparison date, this way it will be possible to determine if the classifier accurately divide what is anomalous or what is normal and stablish models for regression purposes.

The three classifiers will use a 10 K-fold cross validation for training and testing purposes. And for the KNN, K=2 will be selected because it is desired to have only two clusters of data, the one that is labeled as Anomalous or "Y" and Normal or "N" is labeled as the elevation zone that is not considered anomalous.

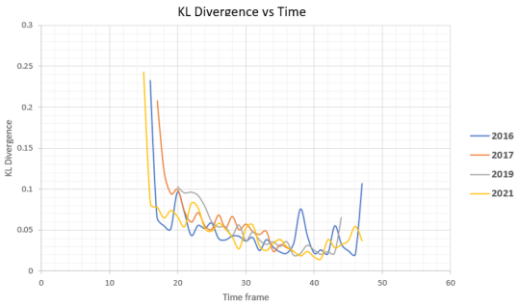
IV – RESULTS



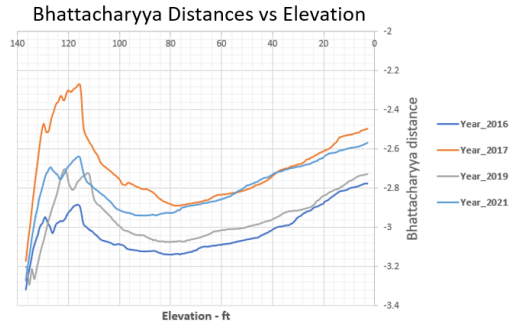
Figure(3) KL divergence vs elevation of probability distributions of temperatures sampled in each year.

	KL Divergence		
Year	Max	Min	Range
2016	123.2	115.18	8.02
2017	122.26	114.7	7.56
2019	120.37	111.14	9.23
2021	125.56	115.18	10.38

Table(1) Divergence peaks identified in each year for elevation in temperature probability distribution



Figure(4) KL divergence vs time of probability distributions of temperature sampled in each year.



Figure(5) Bhattacharyya Distance vs elevation of probability distributions of temperature sampled in each year.

	Bhattacharyya Distance		
Year	Max	Min	Range
2016	126.5	117.54	8.96
2017	128.6	116.56	12.04
2019	122.26	113.29	8.97
2021	127.92	117.54	10.38

Table(2) Distance peaks identified in each year for elevation in temperature probability distribution.

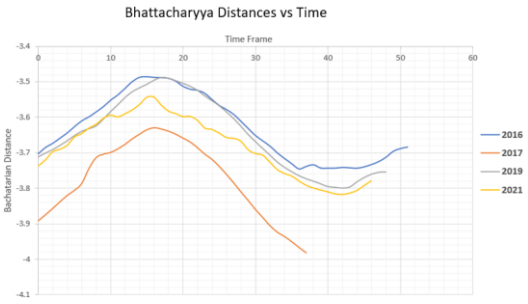


Figure (6) Bhattacharyya Distance vs time of probability distributions of temperature sampled in each year.

Once identified the ranges by KL divergences and Bhattacharyya distance where the region between the peaks is considered the anomalous zone, This zones are labeled as Y or anomalous and the rest are labeled as N for classification purposes. Once done this to label the data set the three classifier are evaluated. For year 2017 for the fourth week of February and the first week of march it is inspected by the temperature vs elevation profile that the anomaly was extended up to an elevation of 132 feet.

	Accuracy			
Classifier	**2016**	**2017**	**2017_Inc**	**2019**
QDA	0.9654	0.9611	0.9362	0.9724
SVM	0.9907	0.9864	0.9782	0.9973
KNN	0.9926	0.9856	0.9821	0.9977

Table(3) Accuracy results for Three different classifiers

KNN classifier was the classifier which accuracy more consistency presented .

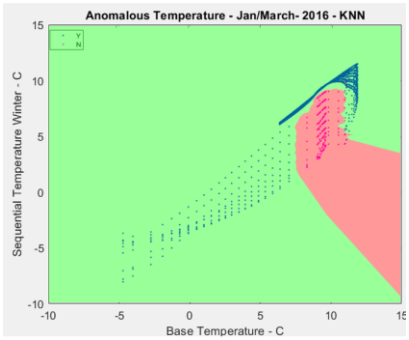


Figure (7) Scatter plot with KNN classifier applied to detect anomalous temperature classes and normal temperature classes for year 2016

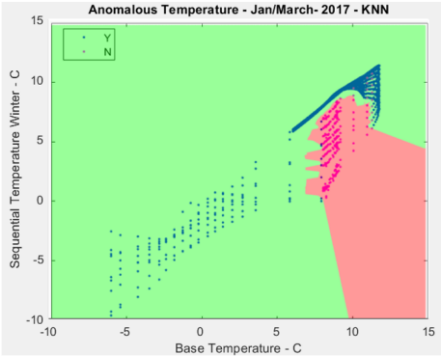


Figure (8) Scatter plot with KNN classifier applied to detect anomalous temperature classes and normal temperature classes for year 2017, with no anomaly extended in the last week of February and first week of March

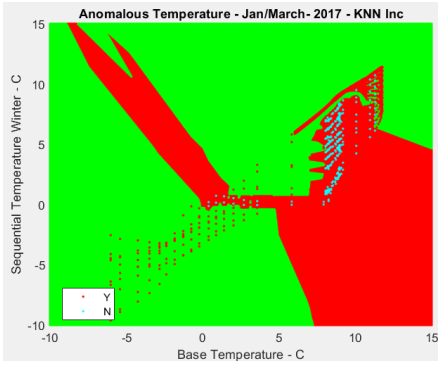


Figure (9) Scatter plot with KNN classifier applied to detect anomalous temperature classes and normal temperature classes for year 2017, with the anomaly extended in the last week of February and first week of March

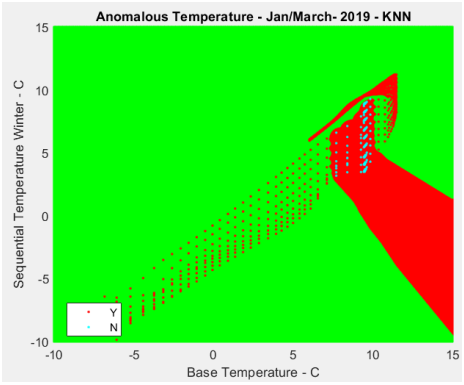


Figure (10) Scatter plot with KNN classifier applied to detect anomalous temperature classes and normal temperature classes for year 2019

V – DISCUSSION

Using the results from table 1 and table 2 to identify peaks in consecutive probability distributions vs elevation from KL divergence and Bhattacharyya, the results from KL divergence where used to label Normal “Y” and anomalous “N” classes to identify the zones where the temperature anomaly is present in this period of the year. After applying the three classifiers in 4 data sets comprehending year 2016, 2017, 2017 with extended anomaly and 2019. It is possible to appreciate that the KNN classifier had more consistency in the accuracy to detect the anomalous zone and draw a decision map where future possible anomalous zone can be identified compared to other classifiers. The anomalies where located between an average distance of 123’ of elevation and 115’ of elevation, but as presented in figure 2 the anomaly for year 2017 is extended up to an elevation of 132’ . KNN results from year 2016, 2017 and 2019 look very similar in shape in contrast of KNN where an extended anomaly was classified for late February and early march. The KNN model built for 2016 can be used for testing purposes for year 2017 without the anomaly extension and for year 2019 expecting a high accuracy. It is necessary to label the anomaly extension for late February and early March.

This type of DTS analysis on embankment dams are relatively new mainly for the data disposition to perform passive analysis. It is recommended to use time series analysis methods such as ARIMA to stablish confidence intervals and detect outliers. Due to the high density of data availability for this project. Also it will be recommendable to use the 2017 and 2019 on the 2016 model to measure accuracy and see if this classification method would work for anomaly detection for regression purposes and detect with false positives and false negatives if the anomaly grew or not.

VII - REFERENCES

[1] Mactaquac Life Achievement Project. https://www.nbpower.com/en/about-us/projects/mactaquac-project. Mactaquac Life Achievement project. Accessed November 2, 2021

[2] Foster, M., Fell, R., & Spannagle, M. (2000). The statistics of embankment dam failures and accidents. *Canadian Geothecnical Journal*, 37(5), 1000-1024

[3] Shija, N.P., MacQuarrie, K.T.B., 2014. Numerical Simulation of Active Heat Injection and Anomalous Seepage near an Earth Dam-Concrete Interface. *International Journal of Geomechanics. (ASCE 2018)*, **04014084**(1-11), doi: 10.1061(ASCE)GM.1943-5622.0000432

[4] Aufleger, M., Conrad, M., Goltz, M., Perzlmaier, S., & Porras, P. (2007). Innovative Dam Monitoring Tools Based on Distributed Temperature Measurement. Jordan Journal of Civil Engineering, 1(1), 29-37.

[5] Velásquez, J.P.P. (2007). Fibre optic temperature measurements: further development of the gradient method for leakage detection and localization in earthen structures. München: Technische Universität München.

[6] Aufleger, M., Goltz, M., Perzlmaier, S., & Dornstädter, J. (2008). Integral seepage monitoring on embankment dams by the DFOT Heat Pulse Method. Proceedings of the 1st International Conference on Long Time Effects and Seepage Behaviour of Dams.

[7] Radzicki, K., & Bonelli, S. (2011). Impulse Response Function Analysis model application to the thermical seepage monitoring in the earth dams. In 20ème Congrès Francais de Mècanique, Besancon, France.

[8] Khan, A.A., Vrabie, V., Mars, J., Girard, A., & D'Urso, G. (2010b). Automatic monitoring system for singularity detection in dikes by DTS data measurement. IEEE Transaction on Instrumentation and Measurement, 59(8), 2167-2175. doi:10.1109.

Appendix

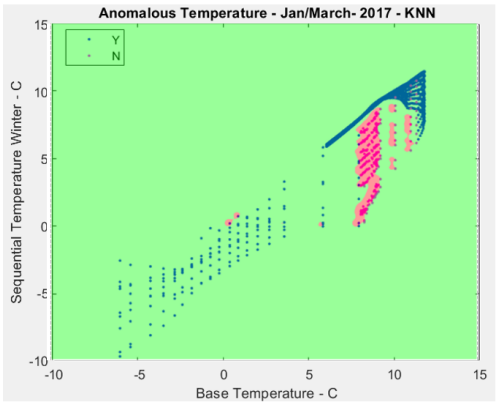


Figure (11) Scatter plot with SVM classifier applied to detect anomalous temperature classes and normal temperature classes for year 2017

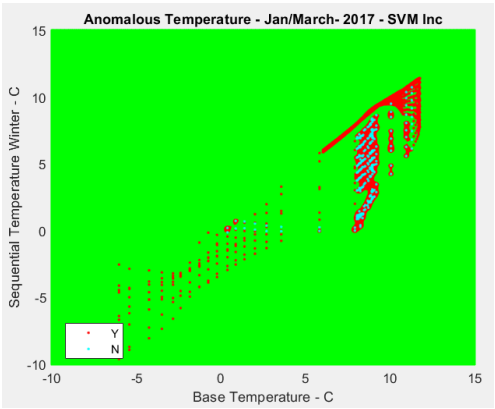


Figure (12) Scatter plot with SVM classifier applied to detect anomalous temperature classes and normal temperature classes for year 2017

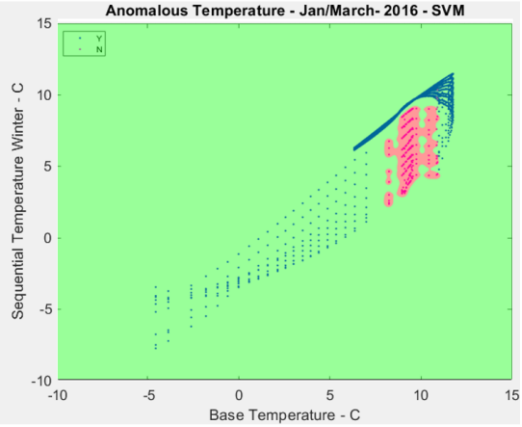


Figure (13) Scatter plot with SVM classifier applied to detect anomalous temperature classes and normal temperature classes for year 2016

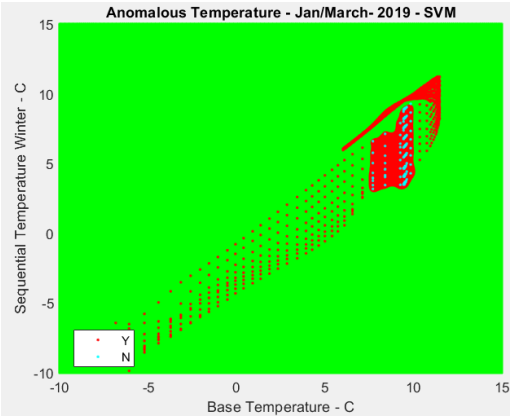


Figure (14) Scatter plot with SVM classifier applied to detect anomalous temperature classes and normal temperature classes for year 2016