

RESEARCH ARTICLE

Initial service life data towards structural health monitoring of a concrete arch dam

Guru Prakash¹ | Ayan Sadhu²  | Sriram Narasimhan³  | Jean-Marie Brehe⁴

¹Department of Civil & Environmental Engineering, University of Waterloo, Waterloo, Canada

²Department of Civil Engineering, Lakehead University, Thunder Bay, Canada

³Department of Civil & Environmental Engineering and Mechanical & Mechatronics Engineering, University of Waterloo, Waterloo, Canada

⁴GKM Consultants, Montreal, Quebec, Canada

Correspondence

Sriram Narasimhan, Department of Civil & Environmental Engineering and Mechanical & Mechatronics Engineering, University of Waterloo, Waterloo, Canada.
Email: snarasim@uwaterloo.ca

Funding information

Natural Sciences Engineering research Council of Canada

Summary

This paper presents a statistical framework to monitor the performance of an operational concrete arch dam using sensory data acquired during its initial service life. One of the major challenges in dealing with a newly constructed dam is to predict its long-term behaviour by forecasting appropriate thresholds using limited data exhibiting nonstationarity. In this paper, a hybrid model is implemented to predict dam responses using environmental—hydrostatic, seasonal, and temperature—as well as age-related variables. The data from multiple sensors are first analyzed using principal component analysis to incorporate overall dam behaviour into a prediction model. The proposed prediction framework is then employed to estimate the residuals and control limits required to calculate thresholds under nonstationary operating conditions during its initial service life. The dam performance is then monitored using statistical control charts and anomalies are detected by comparing the test statistics, square prediction error, and Hotelling T-squared, calculated from the residuals with the preset control limits. The issue of limited data is addressed by updating the model parameters and thresholds periodically, which is aimed at minimizing the false alarm rate. The proposed method is demonstrated using a 130-m-high double-arch concrete dam located in Bulgaria.

KEYWORDS

control limits, principal component analysis, statistical analysis, structural health monitoring

1 | INTRODUCTION

Concrete dams play an important role in the socioeconomic growth of a nation by providing and facilitating irrigation, hydroelectric power, flood control, navigation, and tourism. They can be exposed to excessive operational and environmental loads due to floods, earthquakes, or seasonal variations in temperature during their service life. Their performance could also suffer due to age-related deterioration, for example, creep and shrinkage. Multiple catastrophic failures of concrete dams have been recorded in the past such as the Austin dam in Pennsylvania,^[1] Canyon Lake dam in Texas,^[2] Taum Sauk dam in Reynolds County,^[3] Lake Delhi dam in Iowa,^[4] and Vishnu-Prayag dam in India.^[5] Robust and reliable data-driven health monitoring strategies have since been regarded as crucial in order to prognose failures from anomalous behaviour during their service life. It is equally crucial to put in place such strategies during the initial service life of the dam, as failures (catastrophic or otherwise) could potentially occur at any time following construction.

With the availability of reliable low-cost sensors to an advancement in their miniaturization, data storage, low-power consumption, and communication technologies, there has been a marked adoption of data-driven prediction and automation methods for structural health monitoring (SHM) of dams.^[6–9] In these approaches, data collected from in situ sensors are utilized to

develop prediction models, data analytics to prognose failures, and to schedule dam safety programs. Dams are increasingly being regarded as critical structures, which call for mandatory application of SHM to improve their performance under natural hazards and to track their structural integrity over their lifetime.^[10] These structures have disproportionately large consequences of failure, which underscores the need to implement SHM during construction and also throughout their service life.

SHM tasks for dams are typically carried out by calibrating a prediction model between input variables such as environmental and age-related parameters, and the monitored dam responses, as illustrated conceptually in Figure 1. Existing prediction models can be broadly classified into two categories: deterministic and statistical models.^[11, 12] Deterministic models utilize numerical techniques such as finite element method to predict the dam responses. The advantage of deterministic models is that they can be applied during the first filling of the reservoir, which is generally considered to be the most critical period in the service life of a dam. Andonov et al.^[13] used such a model to demonstrate the safety of Tsankov Kamak dam (TK dam) located in Bulgaria under earthquake-induced natural hazards. However, such models require significant modeling effort, model calibration (which is often cumbersome), and are inadequate to represent nonlinear processes such as damage and leakage.^[14, 15]

On the other hand, statistical models operate directly on measured data (i.e., they fit models to the data directly) without utilizing physics-based models as predictors. These employ data-driven techniques such as neural networks,^[8] support vector regression,^[16, 17] and multivariate linear regression^[6, 7, 9, 18–20] to correlate environmental factors (water level, ambient temperature, and creep in concrete) with measured dam responses (e.g., displacement, crack, concrete stress, and strain). Ahmadi^[21] proposed a hydrostatic-seasonal-time (HST) model to model the response of a dam and applied it to displacement and strain data obtained from Idduki dam in India and to Daniel Johnson dam located in Canada. HST model was also used to model displacement measurements of a rock-fill embankment dam.^[7] A variant of HST model was proposed by Mata et al.^[9] in which seasonal functions are replaced by the recorded temperature to better represent temperature effects. The authors proposed a hydrostatic-thermal-time (HTT) statistical model to interpret the displacement responses of Alto Lindoso dam in Portugal. Loh et al.^[22] proposed a method to establish thresholds for an early warning system and applied it to the Fei-Tsui dam in Taiwan. The displacement data measured for a period of 22 years were analyzed using singular spectrum analysis with neural networks. The authors showed that the proposed method successfully captures periodic fluctuations resulting from seasonal and daily variations in temperature as well as trends due to creep.

Nearly all previous studies utilize long periods of measured data to model steady-state operating conditions. In contrast, new builds show nonstationarities with respect to measured parameters, which makes data-driven modeling a challenging task. Another important issue is data reduction from hundreds of monitored instruments and the identification of critical parameters of dam responses. Previous studies have utilized principal component analysis,^[6] artificial neural networks,^[23, 24] blind source separation,^[25] artificial immune algorithm,^[26] independent component regression,^[27] cointegration theory,^[28] and time-varying Bayesian approach^[29] to model and predict the dam responses. Time-frequency analysis (short-time Fourier transform) was employed by Mata et al.^[30] to identify the effect of daily variations of air temperature on the structural response of a concrete dam. In their study, short-time Fourier transform was used to identify the influence of the daily variation of air temperature on the horizontal displacement of concrete dams and to track its variation as a function of time, which was then used to predict the future dam performance. However, most of these data-driven and pattern recognition techniques are yet to be successfully applied where only the initial service life data (i.e., limited baseline information) is available. In a recent study, Cheng et al.^[19] proposed the use of control charts for dam safety monitoring. However, the aforementioned study is limited to univariate control charts that can be used to monitor one variable, the dam displacement. Usually dam responses are correlated and monitoring each component separately may result in an increase in the false alarm rate.

In this paper, a hybrid prediction model including hydrostatic, seasonal, temperature, and time (HSTT) is developed to better predict dam performance during initial service life. The thresholds are set based on residuals between the measured and predicted responses, in order to alert the owner of any abnormal behaviour in the dam responses, and these thresholds are updated as more

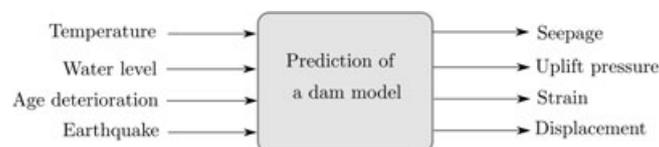


FIGURE 1 Structural health monitoring of a dam

data becomes available. Univariate as well as multivariate control charts are integrated to monitor multiple dam responses. The proposed method is then applied to the strain and displacement data of TK dam located in Bulgaria.

2 | DAM INSTRUMENTATION

This study utilizes measurements obtained from the TK dam (Figure 2), a double curvature concrete arch dam located on the Vacha river in south-western Bulgaria (see Table 1 for dam details). The stability of a dam is based on the load bearing capacity of the rock, which acts as the abutment. The main forces acting on the dam are the hydrostatic pressure due to the reservoir behind it, uplift water pressure beneath the foundation, and self-weight. Other forces that affect the dam performance are temperature, chemical reactions, settling, silt accumulation, and earthquakes. Concrete and air temperatures, reservoir water level, displacements in the superstructure, and in its foundation, rotations, joint movements, strains, and stresses in the concrete, pressures, and discharges in the foundation all contain vital information in order to predict the short- and long-term behaviour of the dam and are good candidates for measurements.

As shown in Table 2, several types of instruments were installed, and data were collected using an extensive data collection system. Among these, dam displacements and strains inside concrete are selected to be investigated in this paper.



FIGURE 2 Tsankov Kamak dam in Bulgaria (a) upstream and (b) downstream view

TABLE 1 Tsankov Kamak double-arch concrete dam details

Total crest length	480.0 m	Maximum height	130.5 m
Crest length (curved part)	340.0 m	Maximum base width	27.6 m
Maximum crest width	8.8 m	Wall volume	$\approx 4 \times 10^5 \text{ m}^3$
Reservoir capacity	$1.1 \times 10^8 \text{ m}^3$	Power capacity	2×40 MW
Total concrete poured	$3.1 \times 10^5 \text{ m}^3$	Power generation	188 GWh/a

TABLE 2 Instruments used in Tsankov Kamak dam

Instrumentation	Parameters measured
Strain gauge	Strain inside concrete block
Total pressure cell	Stress variation inside concrete block
Thermometer (telethermometers)	Concrete temperature
Joint meter	Effectiveness of the joint grouting
Extensometer	Rock deformation under the dam
Piezometer (inside the dam)	Uplift pressure
Piezometer (outside the dam)	Ground water pressure
Telependulums	Horizontal deformation of the dam
Weather station	Air temperature, water level, etc.

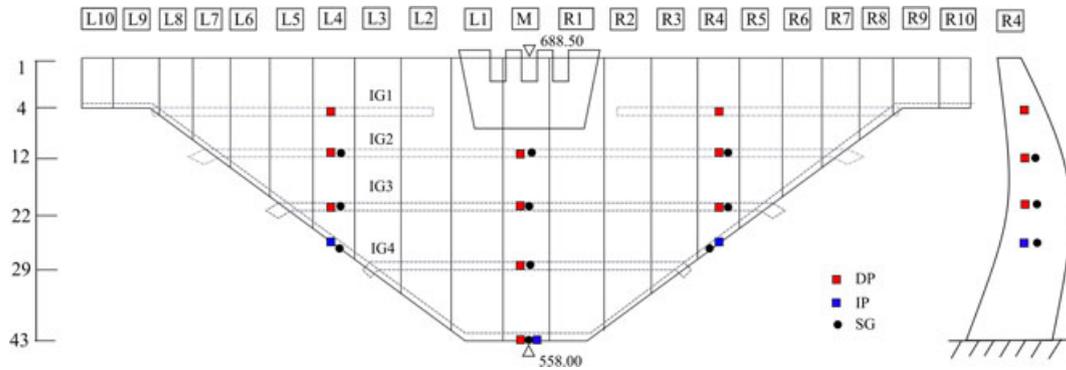


FIGURE 3 Dam instrumentation: strain gauges (SGs), direct (DP), and inverted pendulum (IP). There are 21 blocks and four internal horizontal galleries across the dam

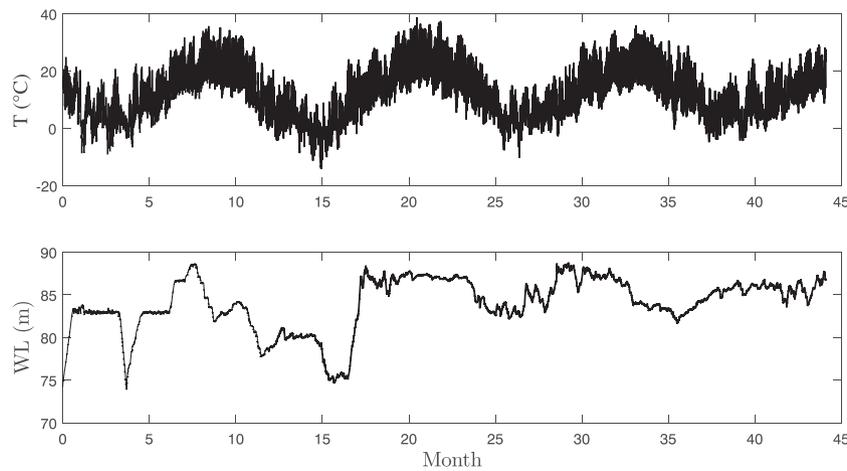


FIGURE 4 Variation of temperature and water level

Figure 3 shows direct pendulum (DP), inverted pendulum (IP), and strain gauges (SGs) installed across the TK dam. As shown in Figure 3, the instrumentation layout was divided into three major cross sections with left (L), middle (M), and right (R) bank profiles, which are stored in a central data acquisition system. A typical instrument is referred in the database based on its location, that is, the block-number and the row number.

For example, SG labeled SG-L4-31 is situated in L4 block, 31st row (each row is approximately 3-m high, with 43 rows) from the top. A scale of row numbers with dam height is also shown Figure 3 (left part). Rows 4, 12, 22, and 29 correspond to the inspection galleries IG1, IG2, IG3, and IG4, respectively. The data analyzed in this paper corresponds to a period between October 2010 and June 2014, resulting in more than 30,000 observations per variable. For illustration purposes, Figure 4 shows the time evolution of the reservoir water level and air temperature.

2.1 | Displacement measurements

DP and IP are used to measure dam displacements along its height. Pendulum method of displacement measurement is based on the relative position of a steel wire through the vertical line of the dam. As shown in Figure 5, the upper end of the wire of the DP is anchored to wall of the inspection gallery, whereas for an IP, the lower end is anchored in the deep zone of the foundation, which is assumed to not be affected by the dam displacement.

In this way, an IP measures the absolute dam displacement, whereas a DP measures the relative displacements^[9] between the levels it is installed.

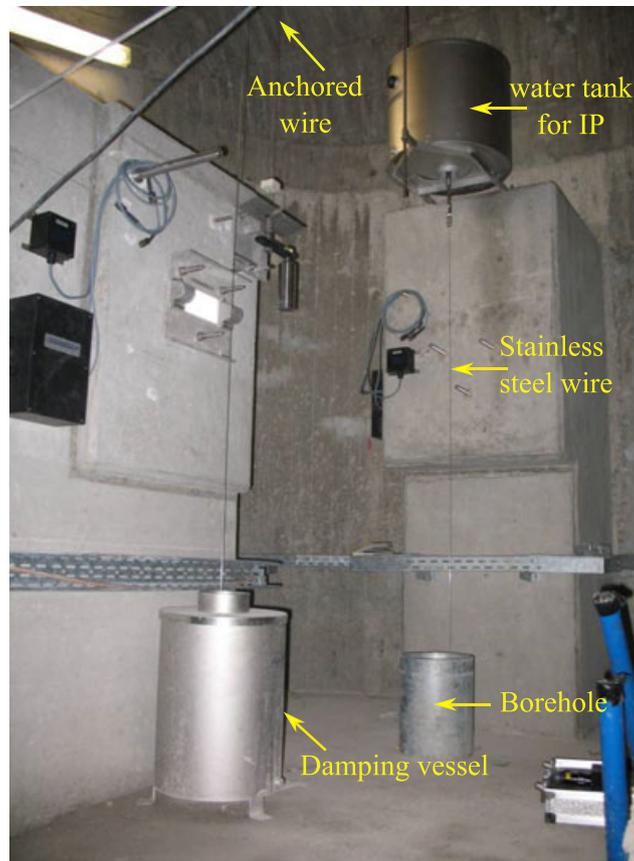


FIGURE 5 Direct pendulums and inverted pendulums (IPs) installation details

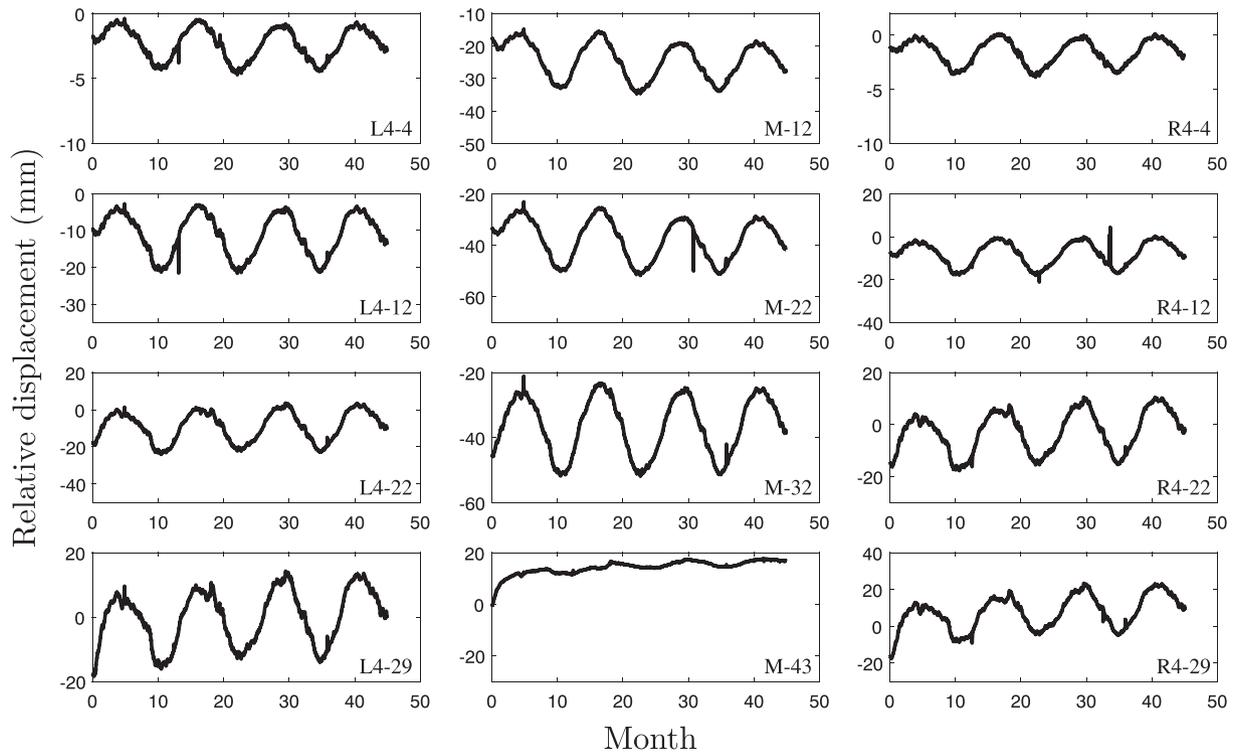


FIGURE 6 Relative displacement (downstream radial direction is positive) profile at different levels of left (L), middle (M), and right (R) blocks

The total displacement at any level can be found by summing relative DP readings across levels. The relative displacement in the radial direction measured at different levels for L-4, M, and R-4 block are shown in Figure 6. Note that in L-4 and R-4 block pendulums are installed in rows 4, 12, 22, and 29, whereas in the middle block it is in row number 12, 22, 32, and 43. Similar to strain measurements, most of the pendulum readings also show seasonal behavior clearly.

2.2 | Strain measurements

Figure 7 shows a typical SG installed in the dam superstructure. The SGs were mounted in a 6-gauge rosette fixture, for three-dimensional strain orientation measurements (45° in each plane), and each recess ($130 \times 110 \times 50$ cm) also includes a no-stress strain (dummy gauge) gauge assembly. The purpose of the dummy SG is to differentiate between measured strains due to hydration of the concrete and strains resulting from true loading of concrete.^[31]

The six strain components (after subtracting the zero-stress strain measurements) corresponding to SG-L4-31 and SG-R4-31 are shown in Figure 8, which can be used to calculate the principal strains and their orientations. It may be noted that the upstream strain does not show seasonal trends compared to the downstream gauge which shows clear seasonal trends. As well,

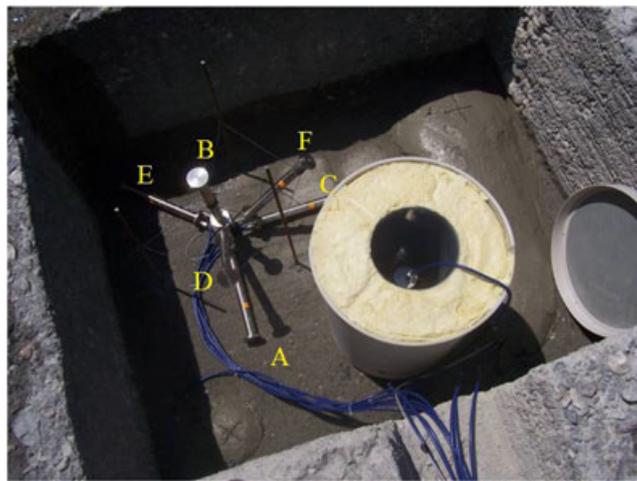


FIGURE 7 Strain gauge installed in Tsankov Kamak dam

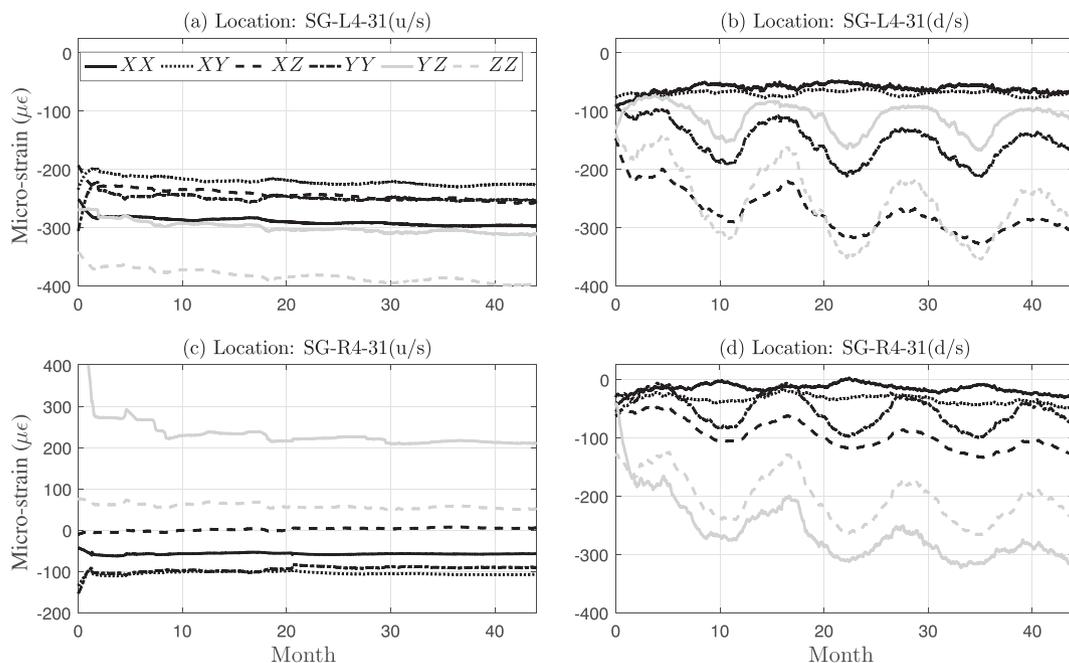


FIGURE 8 Variation of strains with time in the up-stream and down-stream faces

the downstream strain component ϵ_{zz} for SG-L4-31 shows a decreasing trend with time which can be considered as irreversible [8, 28, 29] due to residual deformations during the initial adjustment period of the dam and foundation during reservoir filling, dissipation of heat of hydration, and chemical reactions in concrete or linear bias in the SGs used. In order to investigate this further, zero-stress strains are shown in Figure 9. For all the four locations, these readings seem to be fairly constant, with an absence of significant trend. Moreover, it can be concluded that the decreasing trend in strain is more prominent for the first few months during the first filling of the dam, which appears to stabilize over time. The seasonal effect is also clearly evident

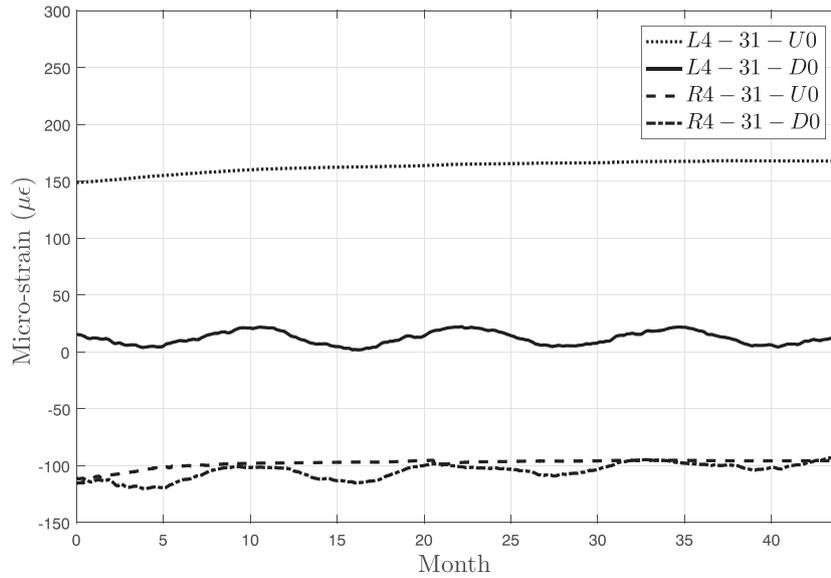


FIGURE 9 Dummy strain gauge readings in L4 and R4 block at level 31

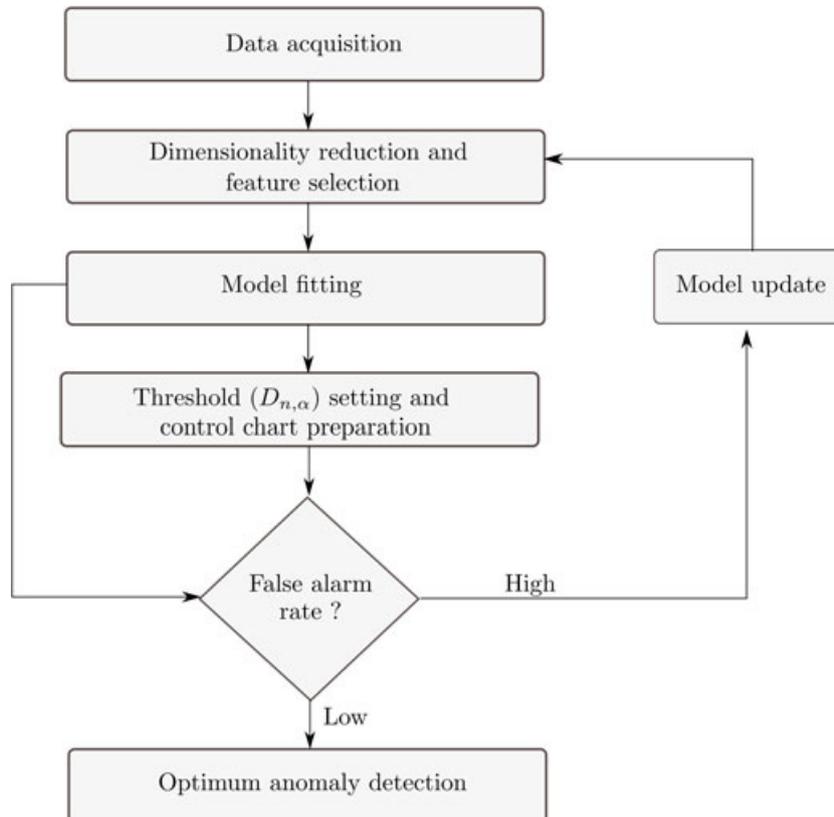


FIGURE 10 Flowchart of the proposed approach

in the downstream zero-stress gauges (temperature effect for the vibrating wire gauges themselves) compared to the upstream gauges where the temperature is rendered more or less constant due the presence of water.

3 | METHODOLOGY

A three-step procedure is proposed for the SHM of the dam under study: (a) dimensionality reduction and feature selection from dam monitoring data, (b) development of a statistical model, and (c) setting up thresholds by minimizing the false alarm rate and comparing the dam response against preset thresholds. A flowchart showing the key steps is presented in Figure 10. Each of these three steps is explained in detail, next.

3.1 | Dimensionality reduction

The first step is to extract the trending parameters (features) from the monitored data (strain and displacement). This can be done locally where data from a particular sensor is used or globally by aggregating similar sensors installed across a block of the dam.

The dimensionality reduction for the latter is achieved using principal component analysis.^[32] For example, crest displacement is a critical parameter in dam monitoring and is monitored locally. However, several thermometers are installed across the dam to measure the temperature. Figure 11 shows the variation of temperature upstream, downstream, and inside the central block at different heights. Note that, the thermometers are installed in row numbers 6, 14, 24, and 34, respectively, which are two rows ($\approx 6\text{m}$) below the pendulum locations.

It can be seen that the temperature is relatively constant on the upstream face, which gradually increases in the middle and to the downstream face of dam. Also, the temperature becomes relatively constant with increasing depth in the upstream face, but the seasonal trends are clearly evident in the downstream face at all depths. In general, the temperature across a block fluctuates; hence, using a particular sensor might result in inaccurate model predictions.

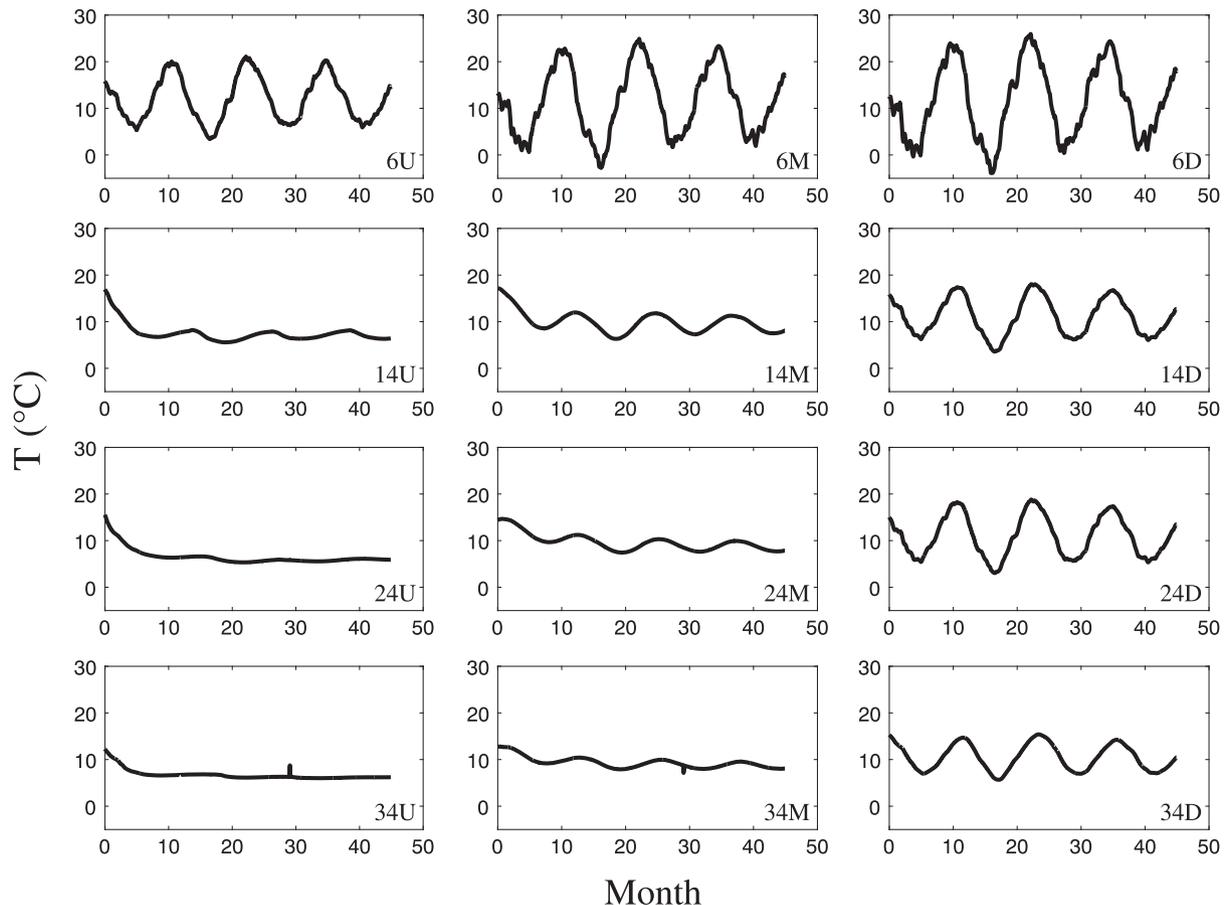


FIGURE 11 Temperature measured at different levels (6, 14, 24, and 34) from the central block, upstream (U), centre (M), and downstream (D)

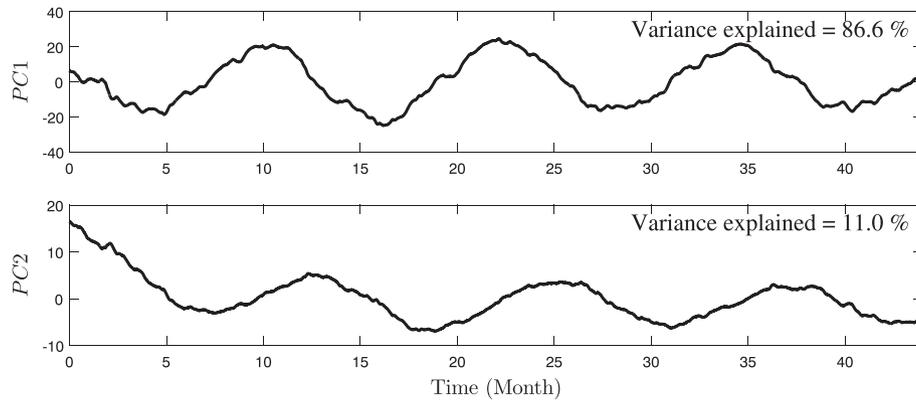


FIGURE 12 The first two principal components of temperature for the central block

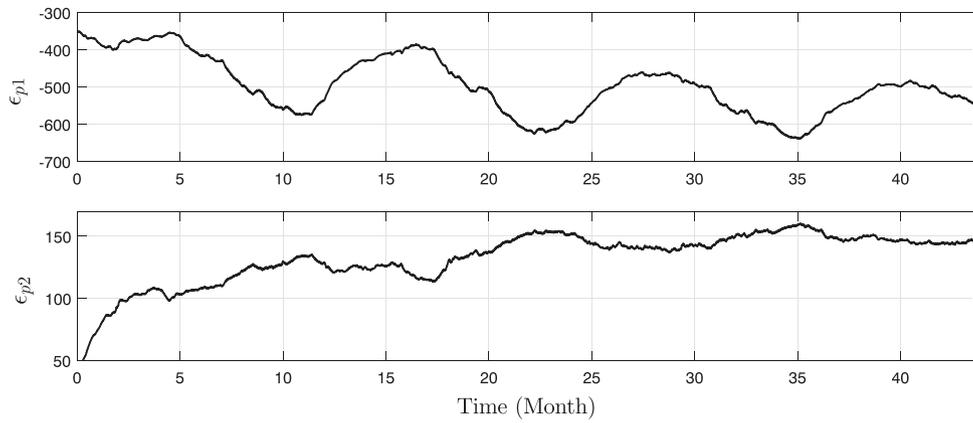


FIGURE 13 Principal strains at SG-L4-31 down-stream

When principal component analysis is performed on the temperature data acquired from all locations (see Figure 12), the first two components cumulatively explain 98% of total variation in the dam temperature and can be used as a global temperature indicator in the ensuing model.

Furthermore, dimensionality reduction can also be employed locally where a sensor measures multiple observations such as from an SG. The SG used in this study measures six strain components (in xx, yy, zz, xy, yz, and xz directions) and the corresponding strain matrix is represented by

$$\epsilon = \begin{bmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{xy} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{xz} & \epsilon_{zy} & \epsilon_{zz} \end{bmatrix}. \tag{1}$$

The principal strains (i.e., ϵ_{p1} , ϵ_{p2} , and ϵ_{p3}) are found through an eigenvalue decomposition of the strain matrix:

$$\epsilon = \Lambda P \Lambda^T, \tag{2}$$

where Λ is an orthogonal matrix and P is a diagonal matrix whose entities (ϵ_{p1} , ϵ_{p2} , and ϵ_{p3}) are the eigenvalues of ϵ .

Figure 8 shows strain components measured in L4 block in row 31. For the downstream data at this location, principal strain components are extracted using Equation 2. The major and minor strain components ϵ_{p1} and ϵ_{p2} are shown in Figure 13.

3.2 | Model fitting

Once the features are extracted, they are used to calibrate the model. The proposed statistical HSTT model is written as

$$y = y_H + y_s + y_T + y_t + \epsilon, \tag{3}$$

where y is the dam response and y_H , y_s , y_T , and y_t are the contributions due to hydrostatic load, seasonal effect, variation in temperature and age related deterioration. The error term ϵ is assumed to be independent and identically distributed. Unlike

existing models,^[14, 33, 34] in the literature, which include the terms y_H , y_s , and y_T only, the proposed method also accounts for age-related deterioration by introducing the y_t term. The following functional form for y_H , y_s , and y_t is considered:

$$\begin{aligned} y_H(H) &= \sum_{i=1}^N a_i H^i; & y_T(T) &= \sum_{i=1}^M b_i T_i^i; & y_t(t) &= c_1 t + c_2 \exp(-t); \\ y_s(\theta) &= d_1 \sin(\theta) + d_2 \cos(\theta) + d_3 \sin(\theta) \cos(\theta); & \text{with } \theta &= \frac{2\pi t}{24 \times 365}; \end{aligned} \quad (4)$$

where $a_i, b_i, c_1, c_2, d_1, d_2$ and d_3 are constants, H is the reservoir water level, T is the temperature, t is the time in hour from the beginning of dam operation, and N, M are the polynomial order with respect to H and T , respectively. In contrast to existing models, y_H is modified by ignoring the higher orders terms of H due to small fluctuations in reservoir level during the monitoring period. Moreover, a linear and quadratic variation of hydrostatic pressure and temperature related effect are considered respectively. Age effect is considered as a sum of linear and exponential terms.

The parameters of the above model can be estimated using the least squares method by minimizing the sum of squared residual errors. Let \mathbf{Y} and \mathbf{X} denote the dam response and transformed variables, respectively, then the estimate of coefficient Θ can be given by

$$\begin{aligned} \hat{\Theta} &= (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Y}) \\ \mathbf{X} &= [\mathbf{H}, \mathbf{H}^2, \mathbf{T}, \mathbf{t}, \exp(-\mathbf{t}), \mathbf{sin}(\theta), \mathbf{cos}(\theta), \mathbf{sin}(\theta)\mathbf{cos}(\theta)] \\ \mathbf{Y} &= [y_1, y_2, \dots \dots y_t]; \quad \Theta = [a_1, a_2, b_1, c_1, c_2, d_1, d_2, d_3]. \end{aligned} \quad (5)$$

The estimated model is used to check the future dam responses and setting up the thresholds.

As explained earlier, due to the initial adjustments of the foundation during the filling of the reservoir, there is a significant temporal effect causing nonstationarity in the dam responses as shown in Figure 13. However, the proposed model as shown above in Equation 3 itself is a stationary model. To overcome this limitation, the regression coefficients are tracked over time using recursive estimation. In order to formulate the recursive model, let us consider Equation 3 in the following form:

$$y(t) = \boldsymbol{\phi}^T(t) \Theta + \eta(t) \quad (6)$$

where $\Theta = \{a_1, a_2, b_1, c_1, c_2, d_1, d_2, d_3\}$ is the vector of unknown regression coefficients. Now, a weighted recursive least square estimation requires finding Θ by minimizing the weighted cumulative squared error:

$$J_k(\hat{\Theta}) = \frac{1}{k} \sum_{t=1}^k \lambda^{k-t} \left[y(t) - \boldsymbol{\phi}^T \hat{\Theta}(t-1) \right]^2, \quad (7)$$

which gives the following recursive estimation of regression coefficients³⁵:

$$\begin{aligned} \hat{\Theta}(t) &= \hat{\Theta}(t-1) + \mathbf{K}(t) \left[y(t) - \boldsymbol{\phi}^T \hat{\Theta}(t-1) \right] \\ \mathbf{K}(t) &= \frac{\mathbf{P}(t-1) \boldsymbol{\phi}(t)}{\lambda + \boldsymbol{\phi}^T(t) \mathbf{P}(t-1) \boldsymbol{\phi}(t)} \\ \mathbf{P}(t) &= \frac{1}{\lambda} \left[\mathbf{P}(t-1) - \mathbf{K}(t) \boldsymbol{\phi}^T(t) \mathbf{P}(t-1) \right]. \end{aligned} \quad (8)$$

The forgetting factor, λ in Equation 8 is a weighting parameter ($0 < \lambda \leq 1$) that may be used to weight recent data more heavily. $\mathbf{K}(t)$ is a gain on the the residual, whereas $\mathbf{P}(t)$ is the covariance matrix of unknown parameters. In this way, the nonstationary nature of the parameters can be tracked during reservoir filling and these parameters can be adjusted periodically to achieve steady-state conditions.

3.3 | Setting thresholds

One of the main goals of a dam monitoring system is the timely detection of any anomalous behaviour, which is indicative of impending failure or potential faults. Anomaly detection for dam safety is generally done by setting up an alarm limit based on the analysis of each individual instrument and plotting the control charts^[36] in real time. With limited data for a long life system such as a dam, selection of an appropriate threshold is a challenging task. Moreover, due to noise in measurements and malfunctioning of instruments, false alarms in control charts are unavoidable.

In this paper, univariate and multivariate control charts are used^[36, 37] for outlier detection in the dam behaviour utilizing control limits based on residuals. The squared prediction error (SPE) norm is defined as

$$SPE = \|y - \hat{y}\|^2, \tag{9}$$

where y is the measured dam response and \hat{y} is the model prediction, both of which are used for univariate control chart preparation. The selection of upper control limit (UCL) is based on fitting a distribution to SPE and choosing values corresponding to α_1 (UCL1) and α_2 (UCL2) percent of cumulative distribution function. A more general nonparametric approach of distribution fitting, namely, kernel density estimation is used here, which relaxes the distributional assumption of SPE.^[38] Given n values of SPE statistics ($SPE_1, SPE_2, \dots, SPE_n$) computed from in-control observations, the distribution of the SPE statistics can be estimated by the following kernel function:

$$\hat{f}_h(t) = \frac{1}{n} \sum_{i=1}^n K \left[\frac{(t - SPE_i^2)}{h} \right], \tag{10}$$

where K and h are the kernel function and smoothing parameter, respectively. The control limits can be determined by a percentile of the estimated kernel distribution. The UCL associated with $100.(1-\alpha)$ th percentile can be calculated by

$$UCL = \hat{f}_h(t)^{-1}(1 - \alpha). \tag{11}$$

For example, Figure 14 shows control limits UCL1 and UCL2 corresponding to $\alpha_1 = 0.95$ and $\alpha_2 = 0.99$. The SPE value corresponding to dotted and solid line is assigned to UCL1 (i.e., warning limit) and UCL2 (i.e., alarm limit), respectively.

For some instruments, the measured dam responses are correlated and simultaneous monitoring is required. For example, Figure 15 shows the correlation plot for principal strain residuals for L4 block in 31st row. Considering the dependency between the two variables, both responses can be monitored using a combined chart.

This is achieved by plotting Hotelling T^2 multivariate control chart. Similar to SPE, another test statistic called T^2 is calculated from the residuals ($\eta = y - \hat{y}$), and given as

$$\mathbf{T}^2 = (\eta - \bar{\eta})^T S^{-1} (\eta - \bar{\eta}), \tag{12}$$

where η is the residual matrix (row is the observation and column is the process variable), $\bar{\eta}$ is the sample mean vector, and S is the sample covariance matrix estimated from in-control process.^[38] T^2 statistics follow the F distribution with p and $(n - p)$ degrees of freedom and control limit can be determined by Equation 13:

$$UCL = \frac{p(n + 1)(n - 1)}{n^2 - np} F_{\alpha,p,n-p}, \tag{13}$$

where n and p are the number of observations and process variables, respectively. In other words, the $100.\alpha$ upper percentile of an F distribution is used as the control limit, where α is the type I error rate (i.e., false alarm rate).

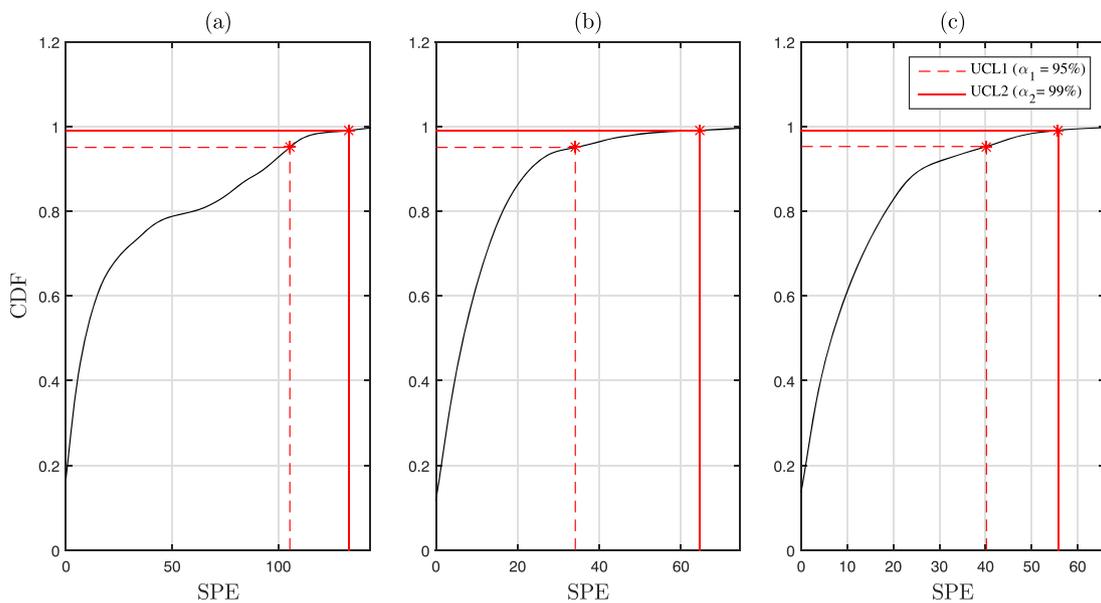


FIGURE 14 Two upper control limits for squared prediction error (SPE) estimated using Kernel density estimation method: (a) 1, (b) 2, and (c) 3 years of data used. UCL = upper control limit

4 | DEMONSTRATION OF THE PROPOSED METHODOLOGY

4.1 | Application to displacement data

First, the proposed method is illustrated for the displacement response in the radial direction. The deformation of dam occurs mainly due to water level and temperature. An increasing water level pushes the structure downstream, whereas a higher temperature causes the structure to shift in the upstream direction. Figure 16 shows the absolute displacements $D1$, $D2$, $D3$, and $D4$ for central block measured in 12th, 22nd, 32nd, and 43rd row, respectively, relative to anchored point above that level. The absolute displacement at a particular level is obtained by summing the relative displacements at lower elevations. For example, the crest displacement $D1$ is given by

$$D1 = \ell_1 + \ell_2 + \ell_3 + \ell_4 + \ell_{IP}, \quad (14)$$

where ℓ_i 's are the relative displacements and ℓ_{IP} is the indirect pendulum reading.

Next, the proposed HSTT model is employed on absolute displacements of central block, where the effect of water level,

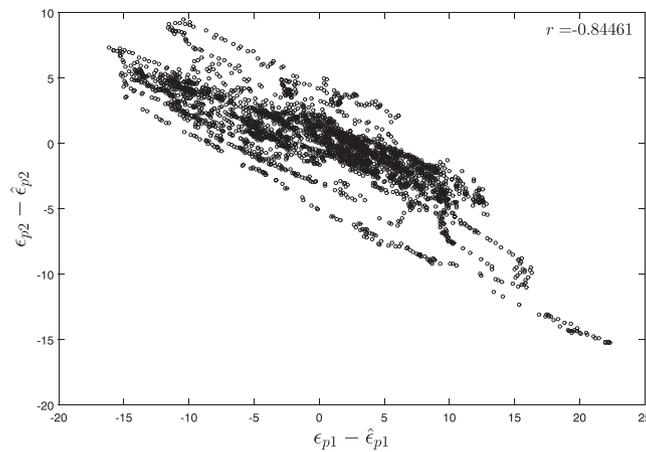


FIGURE 15 Correlation plot between principal tensile and compressive strain residuals

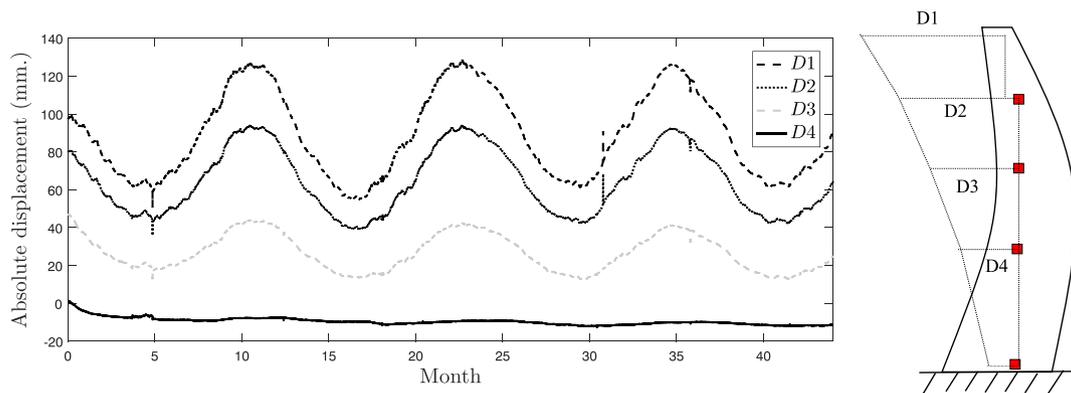


FIGURE 16 Absolute displacements obtained using direct and inverted pendulums

TABLE 3 HSTT model parameters estimated from the dam displacements in middle block

HSTT	H	H^2 (10^{-3})	$\sin\theta$	$\cos\theta$	$\sin\theta$ $\times \cos\theta$	T	t (10^{-3})	$\exp(-t)$	R^2	σ_ϵ (mm)
D1	-1.84	8.4	-21.8	23.5	4.98	0.06	1.14	6.76	0.98	2.88
D2	-1.34	6.5	-17.5	17.3	3.74	0.04	0.42	-0.49	0.99	2.09
D3	-0.52	2.6	-10.4	9.1	2.10	0.01	-0.20	-7.90	0.98	1.24
D4	0.24	-1.2	-1.1	0.1	0.20	-0.01	-0.26	-7.15	0.93	0.53

Note. HSTT = hydrostatic, seasonal, temperature, and time.

temperature, and seasonality are taken as $y_H = a_1H + a_2H^2$, $Y_T = b_1T$, and $y_s = d_1\sin(\theta) + d_2\cos(\theta) + d_3\sin(\theta)\cos(\theta)$, respectively. The time effect is negligible in the period being analyzed and included in the model for long-time monitoring. Table 3 shows the estimated model parameters, R^2 value, and standard deviation (σ_ϵ) of error. The high R^2 value (>0.98 for D_1, D_2 , and D_3) and low σ_ϵ (<3 mm) show excellent model performance under displacement modeling.

In contrast to the HSTT model, in HST and $HT_{pca}T$ models, the water level is normalized as $h = \frac{H - H_{min}}{H_{max} - H_{min}}$. In a previous work,^[9] the authors demonstrated that the fourth power of h is highly correlated with the dam deformation, and the displacements due to seasonal variations (y_s) can be captured either by a sum of two sinusoids (as in HST) or can be expressed in terms of principal components of temperature (as in $HT_{pca}T$). Under these assumptions the HST and $HT_{pca}T$ models can be given as

$$\begin{aligned} HST : \quad y &= ah^4 + d_1\sin\theta + d_2\cos\theta + k, \\ HT_{pca}T : \quad y &= ah^4 + b_1PC_1 + b_2PC_2 + k. \end{aligned} \tag{15}$$

The performance of $HSTT$ model can be compared with respect to these models. The principal components used in $HT_{pca}T$ model are discussed in Section 3 and shown in Figure 12. The estimated parameters, R^2 , and standard deviation of residuals are given in Table 4. In general, the R^2 value is maximum and the standard deviation of residuals (σ_ϵ) is minimum for the HSTT model at all levels of displacements. It is also important to mention that the HST model showed to be more effective than $HT_{pca}T$ at higher elevations (for $D1$ and $D2$). Once the model is calibrated, the model predictions can be used to set the thresholds for anomaly detection using control charts.

The thresholds for dam displacement in this paper are based on square prediction errors as discussed in Section 3. For a structure, where the first few years of operation are critical and at the same time monitoring data is limited during this period,

TABLE 4 Parameters estimated from the dam displacements in the middle block

	HST						HT _{pca} T					
	k	h ⁴	sinθ	cosθ	R ²	σ _ε	k	h ⁴	PC ₁	PC ₂	R ²	σ _ε
D1	-89.9	-3.24	23.5	-21.6	0.97	3.61	-91.9	10.2	-1.64	-0.56	0.95	5.1
D2	-66.1	-0.11	17.3	-17.2	0.98	2.5	-67.0	8.31	-1.23	-0.72	0.95	3.7
D3	-28.8	3.13	9.08	-9.87	0.95	2.1	-28.4	4.88	-0.66	-0.72	0.96	1.9
D4	8.47	3.22	0.16	-0.75	0.34	1.6	9.38	0.55	-0.02	-0.33	0.70	1.1

Note. HST = hydrostatic-seasonal-time; PCA = principal component analysis.

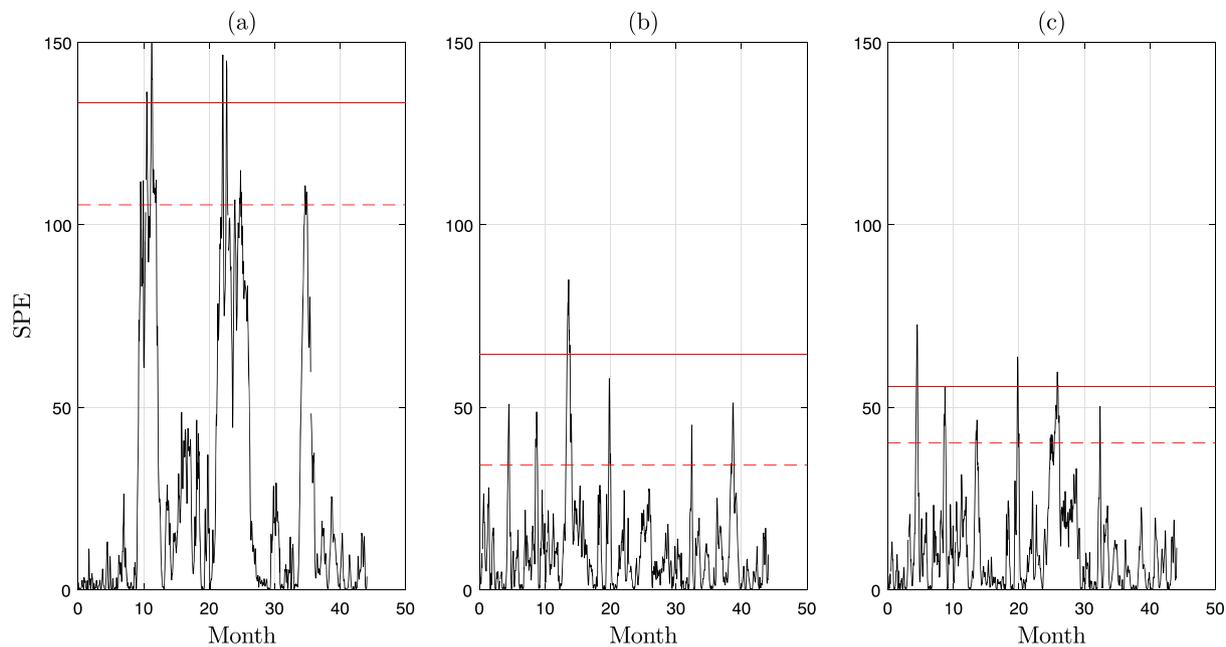


FIGURE 17 Threshold and squared prediction error control chart for crest displacement data when (a) 6, (b) 12, and (c) 18 months of data were used for model fitting

a proper selection of thresholds is a challenging task. One solution is that the model parameters can be re-estimated over time and the thresholds updated periodically as data becomes available.

Figure 17 shows threshold superimposed over SPE control chart for crest displacement (i.e., $D1$). The model parameters and control limits are estimated using 6, 12, and 18 months of data, respectively, and the performance validated over the entire available data (≈ 45 months). It can be seen that the SPE and the threshold values (see Figure 17) decrease as more data becomes available. Moreover, the false alarm rate FA-1 and FA-2 w.r.t. UCL1 and UCL2, respectively, reduce over time as shown in Table 5.

TABLE 5 False alarm rate of displacement and strain measurements

Data used (month)	Crest displacement		L4 block strain	
	FA-1 (%)	FA-2 (%)	FA-1 (%)	FA-2 (%)
6	5.5	1.1	6.7	0.62
12	5.1	1.0	1.9	0.86
18	4.7	0.9	1.5	1.1

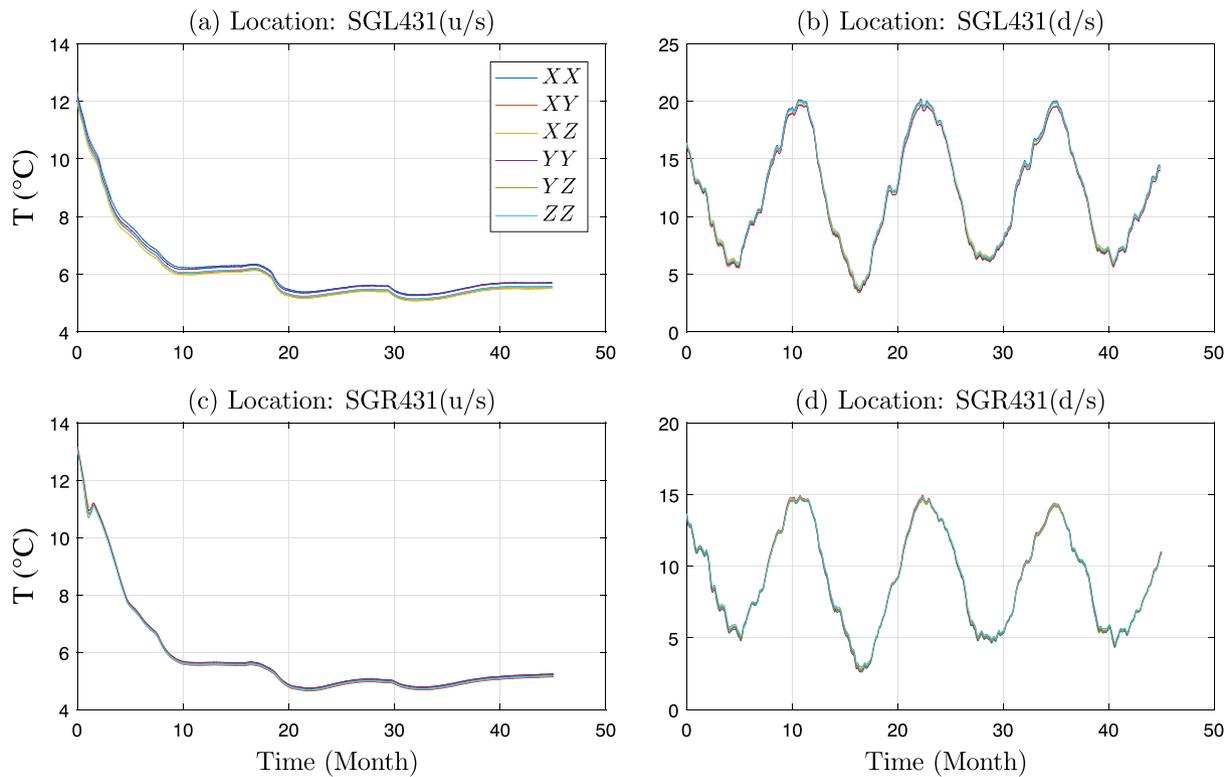


FIGURE 18 Temperature variation at strain gauge location in L4 block

TABLE 6 HSTT model parameters estimated using the strain data in L4 block

HSTT	h	h^2 (10^{-3})	$\cos\theta$	$\sin\theta$	$\sin\theta \times \cos\theta$	T	t	$\exp(-t)$ (10^2)	R^2	σ_ϵ
ϵ_{p1}	-7.5	4.5	-1.14	-32.4	-10.8	-13.3	-13.4	1.1	0.99	6.0
ϵ_{p2}	2.7	-1.6	1.56	11.8	6.2	3.0	1.6	-0.5	0.97	3.2

Note. HSTT = hydrostatic, seasonal, temperature, and time.

4.2 | Application to strain data

In this section, the proposed statistical modeling framework is applied to the SG data. In HSTT model of concrete strain, the temperature measured inside the concrete block is considered, because SGs are embedded with temperature sensors, unlike in the case of pendulums. Figure 18 shows the variation of temperature in all six directions for the L4 block, which shows similar magnitudes.

Moreover, fluctuations of downstream temperature are larger compared to the upstream data. This may be due to the fact that the downstream face is exposed directly to the sun, whereas the upstream face is submerged in water.

Table 6 shows the estimated parameters of the multivariate HSTT regression model. The R^2 value for compressive as well as tensile strains is greater than 0.97 and the standard deviation of the prediction error is reasonably low.

A plot of model predictions superimposed on the actual data along with the respective residual errors are shown in Figure 19.

In order to see the importance of different variables in the model, a stepwise regression is also performed. Stepwise regression is a systematic method for adding and removing terms from a multilinear model based on their statistical significance (e.g., t -statistics) in a regression. It can be seen from Table 7 that approximately 90% of variation in the strain is explained by three predictor variables, namely, h , $\cos\theta$, and t .

Figure 20 shows the recursive estimation (see Equation 8) of regression coefficients for principal strains ϵ_{p1} using the first 44 months of data. Initial fluctuations in the parameters can be observed due to the choice of an arbitrary covariance matrix. However, it may be noted that the regression coefficients stabilize over time and attain their steady-state values.

Because the data is relatively short in length (just $3\frac{1}{2}$ years of data, which is relatively small compared to the expected life of the dam) and the parameters are slowly changing, the effect of λ is observed to be nominal. However, it can be significant and its

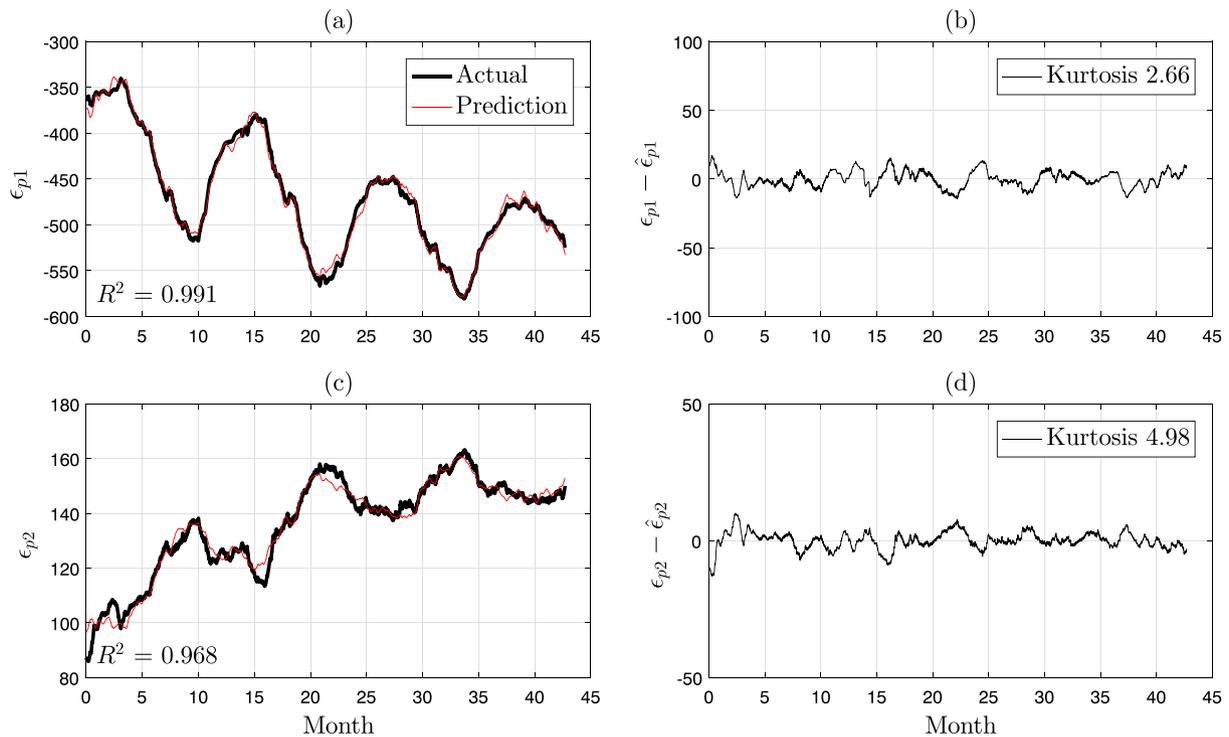


FIGURE 19 HSTT model and residual for (a) tensile and (b) compressive strain in L4 block

TABLE 7 Stepwise regression summary of principal strain

Tensile strain (ϵ_{p1})					Compressive strain (ϵ_{p2})				
Var.	R^2	RMSE	F (10^3)	t -stat	Var.	R^2	RMSE	F (10^3)	t -stat
h	0.15	56	5.5	-74	h	0.16	16	5.7	75
$\cos\theta$	0.64	36	28	207	$\cos\theta$	0.35	14	8.3	-95
t	0.94	14	182	-416	t	0.88	6	77	373
h^2	0.95	13	157	-65	h^2	0.9	5.4	69	406
T	0.96	12	132	-128	e^{-t}	0.94	4	10	-157

Note. RMSE = root mean square error.

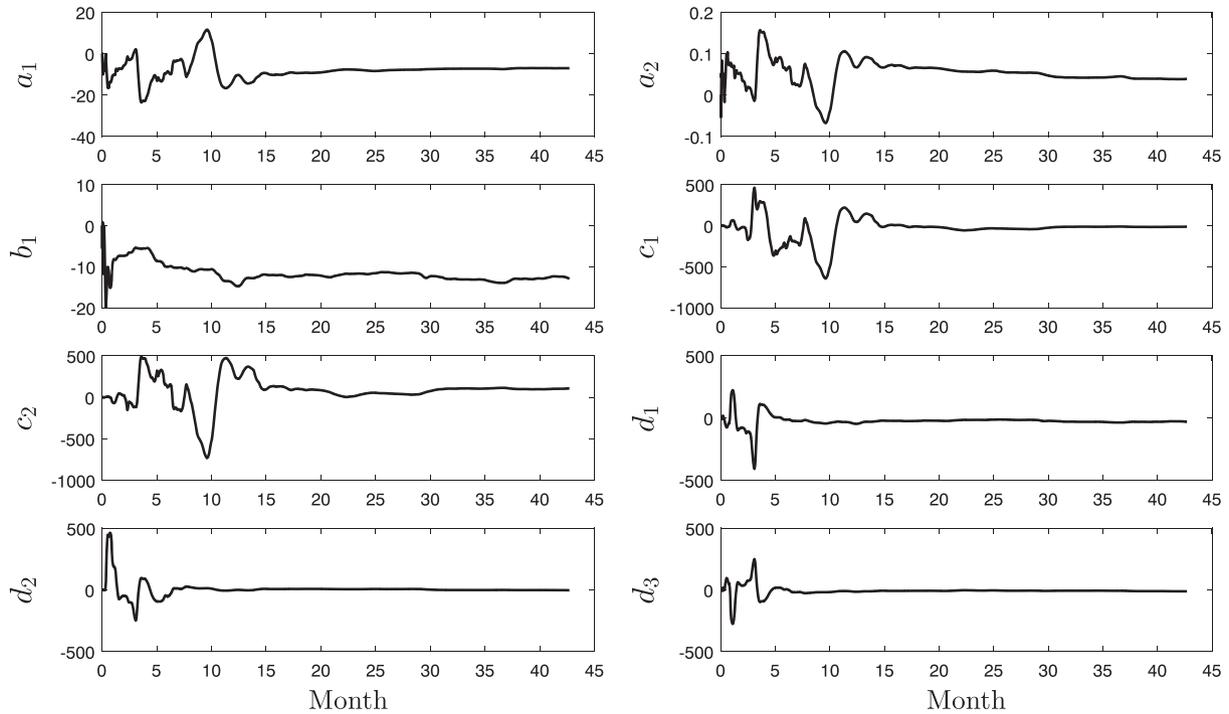


FIGURE 20 Recursive estimation of regression coefficients for ϵ_{p1} in L4 block

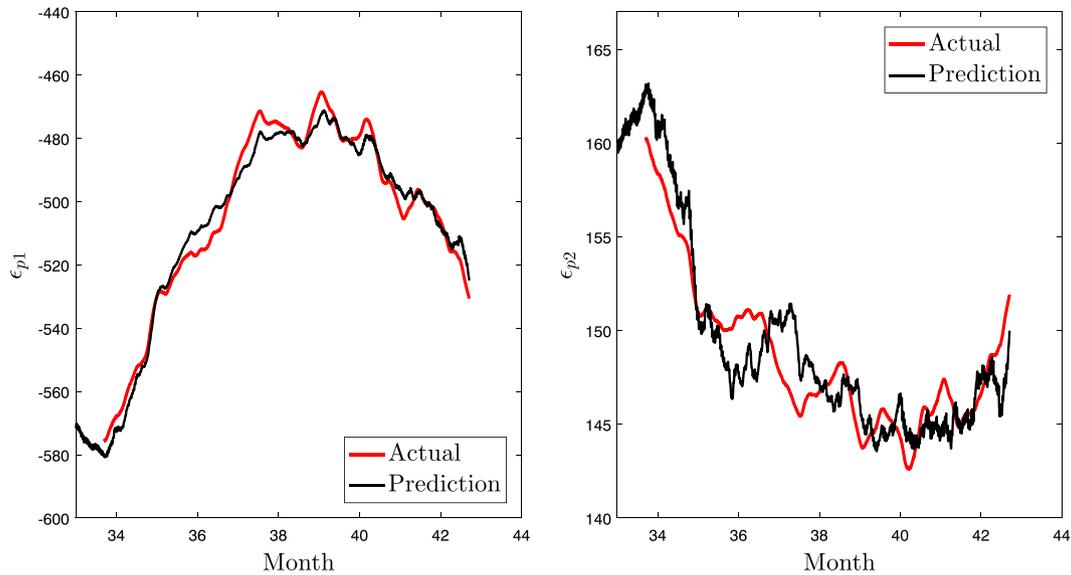


FIGURE 21 Predicted strains ϵ_{p1} and ϵ_{p2} using recursive regression model of in L4 block

effectiveness can be evaluated once more data comes available. Figure 21 gives the prediction of strain data for last 10 months (September 2013 to June 2014) using the proposed method based on the first 33 months of data with a R^2 value of 0.98.

Finally, the model performance with respect to HST and HTT models are compared. The inclusion of time variable is verified by stepwise regression (see Table 7), which can also be verified by a prominent trend in the principal strains, as in Figure 19. Another variation with respect to the displacement model is due to the temperature term. Because SGs have temperature sensors, absolute temperature is included in the model, not the principal components. Hence the model is HTT not $HT_{pca}T$ and given by

$$\begin{aligned}
 HTT : \quad y &= ah^4 + b_1T + c_1t + k, \\
 HST : \quad y &= ah^4 + d_1\sin\theta + d_2\cos\theta + c_1t + k.
 \end{aligned}
 \tag{16}$$

TABLE 8 Parameters estimated from HST and HTT model of principal strains in L4 block downstream

	HST							HTT					
	k	h^4	$\sin\theta$	$\cos\theta$	t	R^2	σ_ϵ	k	h^4	T	t	R^2	σ_ϵ
ϵ_{p1}	-402	-24	64	5.6	-37.4	0.95	13.0	-288	-15.80	-9.26	-37.3	0.97	11.1
ϵ_{p2}	111	9	-11	0.8	13.3	0.89	5.7	92	4.30	1.60	13.5	0.88	5.9

Note. HST = hydrostatic-seasonal-time; HTT = hydrostatic-thermal-time.

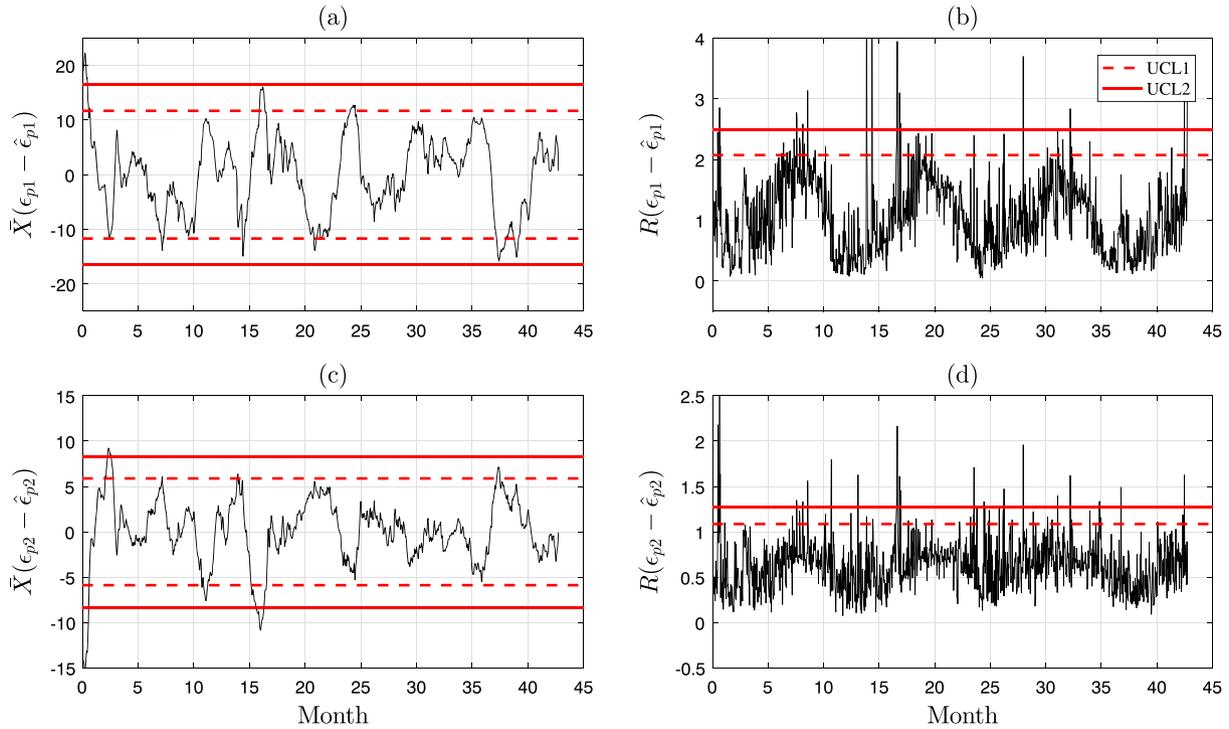


FIGURE 22 X bar and range chart for principal strains. UCL = upper control limit

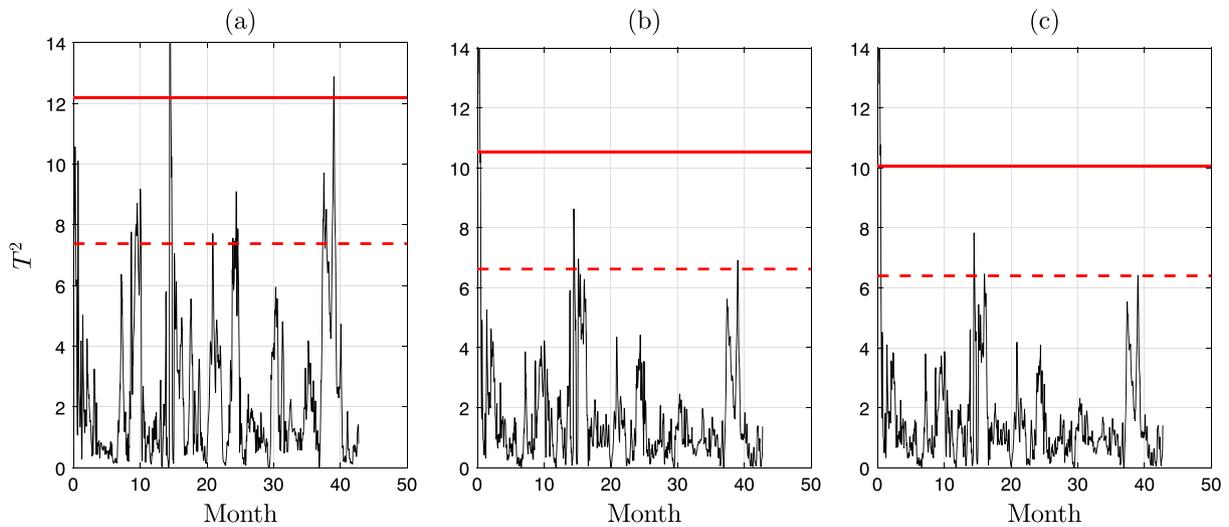


FIGURE 23 T^2 control chart for principal strains when (a) 6, (b) 12, and (c) 18 months of data were used for modal fitting

Table 8 presents the estimated HST and HTT parameters for principal tensile and compressive strain. A comparison of R^2 (see Table 6) shows better performance of HSTT with respect to these two models. Also standard deviation of residuals is lower for the proposed model indicating a smaller confidence interval.

The residuals obtained from the previous step contain important information for threshold setting and anomaly detection. The univariate mean (to detect the mean shift) and range (to detect the variability) control chart for principal strain residuals is shown in Figure 22. The threshold is set based on normality assumption of residuals.

The mean chart shows that the dam is functioning normally. However, the false alarm rate in the range chart is high for both cases, which is one of the difficulties encountered in individual monitoring using univariate control charts. This can be alleviated by a simultaneous monitoring of strains using multivariate control charts. Figure 23 shows the multivariate control chart for principal strains using T^2 test statistics as discussed in Section 3, Equation 12. The HSTT model and control limits were updated at 6-, 12-, and 18-month intervals.

As expected with more data T^2 value decreases as the model parameter becomes more accurate. Table 5 shows the false alarm rate (type I error) corresponding to 95% and 99% control limits.

5 | CONCLUSIONS

This paper presents a general approach to statistically model the dam responses and monitor its performance in real time using statistical control charts, when only the data during initial service life is available. Unlike most of the previous studies, where hydrostatic, seasonal, and time (HST) or hydrostatic, temperature, and time (HTT) are undertaken as predictor variables, the proposed HSTT model considers all the four variables to model the dam responses using limited sensor measurements. In this way, the model is more robust and can be applied to multiple dam responses such as displacement, stress, or strain. In order to check the overall dam performance, measurements from multiple sensors are included in the model where principal component analysis is utilized as a data reduction tool. The real-time monitoring of dam responses is undertaken using univariate (for displacement) and multivariate (for strain) control charts, where the control limits are assigned based on residuals. Test statistics such as SPE and T^2 value are used for anomaly detection in a real-time setting. Finally, to avoid noisy or unreliable measurements, the model parameters and control limits are updated periodically resulting in lower false alarms. The main advantage of using HSTT model compared to other statistical models (such as HST and HTT) is its better performance (in terms of model fitting), with lower residual standard deviation (i.e., narrower confidence intervals).

ACKNOWLEDGMENT

Financial support provided by the Natural Sciences Engineering research Council of Canada through their Engage program is gratefully acknowledged.

REFERENCES

- [1] T. P. Rich, *Int. J. Eng. Educ.* **2006**, 22(6), 1287.
- [2] W. J. Graham, *A Procedure for Estimating Loss of Life Caused by Dam Failure*, US Department of the Interior, Bureau of Reclamation, Dam Safety Office, Denver, Colorado **1999**.
- [3] J. D. Rogers, C. M. Watkins, J.-W. Chung, *Environ. Eng. Geosci.* **2010**, 16(3), 257.
- [4] L. McDaniel, J. Garton, W. Fiedler, W. King, N. Schwanz, Lake Delhi dam breach—two perspectives, in ASDSO Annual Conference, Washington D.C. **2011**, 2.
- [5] G. B. Baecher, *Georisk: Assess. Manage. Risk Eng. Syst. Geohazards* **2016**, 10(2), 92.
- [6] L. E. Chouinard, D. W. Bennett, N. Feknous, *J. Perform. Constr. Facil.* **1995**, 9(4), 286.
- [7] S. Gamse, M. Oberguggenberger, *Struct. Control Health Monit.* **2017**, 24(1), e1859.
- [8] J. Mata, *Eng. Struct.* **2011**, 33(3), 903.
- [9] J. Mata, A. Tavares de Castro, J. Sá da Costa, *Struct. Control Health Monit.* **2014**, 21(3), 423.
- [10] J. Brownjohn, *Philos. Trans. Royal Soc. London A: Math. Phys. Eng. Sci.* **2007**, 365(1851), 589.
- [11] M. Bianchi, R. Bremen, *Int. J. Restor. Build Monuments* **2001**, 7(3/4), 271.
- [12] F. Salazar, M. Toledo, E. Oñate, R. Morán, *Struct. Saf.* **2015**, 56, 9.
- [13] A. G. Andonov, A. Iliev, K. S. Apostolov, *Struct. Eng. Int.* **2013**, 23(2), 132.
- [14] L. Chouinard, V. Roy, Performance of statistical models for dam monitoring data, in Joint International Conference on Computing and Decision Making in Civil and Building Engineering, Montreal **2006**, 14.
- [15] B. Wang, Z. He, *J. Sound Vib.* **2007**, 302(4), 1037.
- [16] V. Ranković, N. Grujović, D. Divac, N. Milivojević, *Struct. Saf.* **2014**, 48, 33.
- [17] H. Su, Z. Chen, Z. Wen, *Struct. Control Health Monit.* **2016**, 23(2), 252.

- [18] R. Ardito, G. Cocchetti, *Eng. Struct.* **2006**, 28(14), 2036.
- [19] L. Cheng, D. Zheng, *Adv. Eng. Software* **2013**, 57, 48.
- [20] J.-S. Lew, C.-H. Loh, *J. Civil Struct. Health Monit.* **2014**, 4(4), 245.
- [21] B. Ahmadi-Nedushan. **2002**, Multivariate Statistical Analysis of Monitoring Data for Concrete Dams, Phd Thesis, McGill University, Montreal, Canada.
- [22] C.-H. Loh, C.-H. Chen, T.-Y. Hsu, *Struct. Health Monit.* **2011**, 10(6), 587.
- [23] C.-Y. Kao, C.-H. Loh, *Struct. Control Health Monit.* **2013**, 20(3), 282.
- [24] I. Karimi, N. Khaji, M. Ahmadi, M. Mirzayee, *Eng. Struct.* **2010**, 32(11), 3583.
- [25] T. D. Popescu, *Int. J. Innov. Comput. Inform. Control* **2011**, 7(7), 3811.
- [26] G. Xi, J. Yue, B. Zhou, P. Tang, *Comput. Math. Appl.* **2011**, 62(10), 3980.
- [27] W. Dai, L. Bin, X. Ding, D. Huang, *Trans. Nonferrous Metals Soc. China* **2013**, 23(7), 2194.
- [28] F. Li, Z. Wang, G. Liu, *Struct. Saf.* **2013**, 43, 12.
- [29] H. Su, Z. Wen, X. Sun, M. Yang, *Struct. Saf.* **2015**, 57, 1.
- [30] J. Mata, A. T. de Castro, J. S. da Costa, *Eng. Struct.* **2013**, 48, 658.
- [31] Committee AT, *ASCE* **2000**.
- [32] I. Jolliffe, *Principal Component Analysis*, Wiley Online Library **2002**.
- [33] A. De Sortis, P. Paoliani, *Eng. Struct.* **2007**, 29(1), 110.
- [34] P. Léger, M. Leclerc, *J. Eng. Mech.* **2007**, 133(3), 267.
- [35] L. Ljung, *System Identification: Theory for the User*, Englewood Cliffs, New Jersey **1987**.
- [36] S. Bersimis, S. Psarakis, J. Panaretos, *Qual. Reliab. Eng. Int.* **2007**, 23(5), 517.
- [37] D. C. Montgomery, *Introduction to Statistical Process Control*, John Wiley & Sons, New York, NY **2001**.
- [38] P. Phaladiganon, S. B. Kim, V. C. Chen, J.-G. Baek, S.-K. Park, *Commun Stat Simul Comput*[®] **2011**, 40(5), 645.

How to cite this article: Prakash G, Sadhu A, Narasimhan S, Brehe JM. Initial service life data towards structural health monitoring of a concrete arch dam. *Struct Control Health Monit.* 2017;e2036. <https://doi.org/10.1002/stc.2036>