

Online anomaly detection for long-term structural health monitoring of caisson quay walls

Taemin Lee ^{a,b}, Seung-Seop Jin ^c, Sung Tae Kim ^a, Jiyoung Min ^{a,*}

^a Department of Structural Engineering Research, Korea Institute of Civil Engineering and Building Technology, Goyang 10233, South Korea

^b School of Architecture, Soongsil University, Seoul 06978, South Korea

^c Department of Civil and Environmental Engineering, Sejong University, Seoul 05006, South Korea

ARTICLE INFO

Keywords:

Structural health monitoring
Anomaly detection
Online learning
Quay wall
Port structures

ABSTRACT

To assess the current state and develop maintenance strategies for proactive management of infrastructure, research on Structural Health Monitoring (SHM) has been actively performed. For port facilities, the need for sensor-based monitoring is increasing to analyze the effects of various factors such as aging, ship activities, backfill earth pressure, and waves. However, few researchers have conducted long-term monitoring of caisson quay walls. In this study, an SHM system was developed with different types of sensors installed on two caisson quay walls and monitored over one year. A new online adaptive anomaly detection approach was proposed to identify the anomalous status of each caisson in real-time by analyzing multiple variables. The method was validated with seven simulated anomaly scenarios, demonstrating high accuracy in anomaly detection despite significant environmental variations, outperforming other approaches. These results highlight the potential to provide timely and accurate alerts when anomalous states occur in port structures.

1. Introduction

The maintenance of deteriorating concrete infrastructure poses a significant challenge, imposing a substantial economic burden worldwide. Over the past decade, billions of dollars have been spent globally on the repair and restoration of aging structures [1–3]. Unfortunately, many of these projects were initiated without adequate consideration for their service life, leading to premature deterioration. The absence of a systematic, service-life-oriented approach puts civil infrastructure at significant safety risks, including the potential for rapid, unforeseen degradation or, in extreme cases, sudden collapse. Consequently, investigating cost-effective methods to extend the service life of civil engineering structures is crucial, thereby reducing the necessity for major repairs.

Port structures are critical social infrastructures that manage the cargo volume of imports and exports for each country, playing a vital role in logistics. Among port facilities, mooring and berthing facilities such as quays and docks offer an important function and space for ships to load and unload cargo, directly correlating with the safety of workers and operational performance. For these, the Port of Rotterdam in the Netherlands, a representative smart port in the world, has implemented various advanced technologies to maintain berthing and mooring

facilities, including embedded sensor-based monitoring [4]. Especially, timely inspection and maintenance technology for port facilities have become one of the major issues in recent years due to the increased frequency of typhoons, heavy rainfall, and high waves surpassing design levels, attributed to climate change.

In monitoring technologies for the safety and usability of port facilities, it is key to define the types of damage that may occur and detect damage timely and correctly. Displacement-related damage in port facilities such as sliding, inclination, and settlement exceeding the criteria are critical failures significantly affecting the stability of the structure [5–7]. Various types of sensors should be installed to monitor the displacement-related damage, such as inclinometer and displacement gauges, and also these multiple measurements should be evaluated individually and comprehensively. In this stage, it is important from the standpoint of a non-professional manager that the entire signal analysis process is fully automated by eliminating the heuristic parameters.

Various studies have been reported on anomaly detection techniques for road facilities such as bridges, incorporating deep learning, machine learning, etc. Ni et al. proposed a framework using a one-dimensional convolutional neural network, capable of directly extracting features from input signals [8]. To assess its effectiveness, they utilized acceleration data measured from a 1 km long span bridge in China. The results

* Corresponding author.

E-mail address: amote83@kict.re.kr (J. Min).

indicated that their technique could detect anomalies with a precision of 97.93 %, a recall of 97.13 %, and an overall accuracy rate of 99.15 %, leading to a final F-measure of 97.53 %. Li et al. proposed an end-to-end framework for recognizing anomalous states of bridges using monitoring data [9]. They validated their algorithm against traditional methods such as support vector machine, k-nearest neighbors, and linear regression. Their framework exhibited superior performance, achieving identification precision and recall rates of 0.962 and 0.967, respectively. Turrisi et al. presented a cointegration-based damage detection technique for automatically assessing the health of structures [10]. It produced cointegration residuals that mitigate the impact of environmental and operational variables. It was applied to the monitoring data measured from the roof of Milan's G.Meazza Stadium. The principal component analysis (PCA) has been widely applied for anomaly detection based on the information loss derived from reconstruction [11,12]. Zhu et al. studied that temperature effects could skew the outcomes of structural anomaly detection and introduced a temperature-driven moving principal component analysis method [13]. Pan et al. developed a transfer learning-based approach for anomaly detection in SHM systems. This method used convolutional neural networks to identify anomalies in various sensor data, such as acceleration, strain, displacement, humidity, and temperature. By applying transfer learning, the model adapted knowledge from one bridge's monitoring data to another, reducing the need for extensive retraining [14]. Glashier et al. applied moving PCA and cointegration to SHM data from the MX3D Bridge. Using the TB-MI method, they significantly increased the anomaly detection rate from 17 % to 67 %, showcasing improved identification of structural damage and providing robust detection across various sensor clusters [15]. These methods isolated data fluctuations attributable to temperature to facilitate structural anomaly detection.

The anomaly detection techniques vary depending on the unique characteristics of target structures exposed to various deterioration environments. So, it should be performed based on long-term monitoring data. However, there is limited research on long-term monitor port structures, especially mooring and berthing facilities [16]. The only research has been reported from the project at the Port of Rotterdam, while primarily aims to monitor long-term behavioral characteristics of the docks. Inventec Corp. has installed fiber optic sensors to monitor the anchorage force, deformation, and relaxation of grout anchors, assess the compression on pile heads due to ships, measure back soil pressure on quay walls, and examine the correlation between groundwater level and port water level using fiber optic piezometers, all authorized by the Port of Rotterdam Authority [17–19]. Grosso et al. conducted the long-term monitoring of strain and displacement in port facilities at the San Giorgio pier using the Surveillance d'Ouvrages par Fibres Optiques system, achieving a minimum measurement precision of 2 μm within a temperature range of -10 to 40°C [20,21]. The observed strain and displacement showed a strong correlation with temperature and a weak correlation with factors such as ship docking and crane positions. Recent initiatives aimed at creating a smart quay wall by integrating monitoring systems with inclinometers, strain gauges, groundwater level meters, fiber optic sensors, satellites, AI, and IoT sensors, thus enhancing the allowable quay load by approximately 20 % over conventional methods [18,19]. Pengel et al. developed the UrbanFlood Early Warning System (EWS) for dikes and levees, utilizing real-time data from a sensor network and AI to predict failures and simulate flood scenarios. This system enhances flood prevention by providing decision support tools for emergency management, integrating sensor data with weather information [22]. Bolourani et al. developed a SHM system for harbor caissons, utilizing support vector machines and principal component analysis to detect damage under dynamic loads such as earthquakes and vessel impacts. The study used finite element modeling to simulate the caisson structure and identified damage-sensitive features, achieving accurate classification of undamaged and damaged states with an F1 score above 90 % for most cases [23]. Despite these advances, research

on real-time damage detection from a structural perspective remains unreported.

Therefore, this study represents the first step toward this objective. A new approach is proposed to automatically monitor and evaluate damage events occurring on the caisson quay wall by combining multiple measurement features on each caisson unit. It is important to note that caisson-type facilities are generally managed on a per-unit basis. First, sensors were installed on a caisson quay wall in Incheon, South Korea, to measure slope, settlement, concrete spacing, and temperature, and were monitored over one year to examine structural performance and environmental effects. Next, several scenarios of anomalous states were artificially designed, varying in difficulty level, considering the types of failures in mooring facilities. Then, the proposed algorithm was applied and validated to determine whether the occurrence of damage could be automatically detected immediately by comprehensively utilizing sensor information.

2. Proposed anomaly detection method for a caisson quay wall

2.1. Procedures of the proposed method

The procedure of the proposed online anomaly detection technique was summarized as follows with Fig. 1. Details on determining the threshold for autonomous anomaly detection and updating the initial threshold adaptively in long-term SHM are explained in the next section.

The key distinction of the proposed method compared to existing approaches lies in its ability to overcome the nonlinear correlation with temperature while performing anomaly detection. This is achieved by incorporating a GMM (Gaussian Mixture Model) clustering technique. Unlike traditional methods that may struggle to address the complexity of temperature-related variations, the use of GMM allows for more accurate feature extraction and clustering by accounting for the nonlinear relationships between temperature and structural responses. This adaptive approach enhances the reliability of anomaly detection in environments where temperature fluctuations play a significant role, ensuring that false alarms are minimized and true structural anomalies are detected with greater precision.

Step 1. Initial principle component analysis (PCA) Setting: Perform PCA on the initial training data (*multivariable data set each caisson*), determine the number of principal components, derive residual error through Q-statistics, and set threshold values.

Step 2. Online Adaptive-PCA Fitting: Receive new data (\mathbf{X}_{new}) and update the current PCA model, using PCA components obtained from Step 1 to calculate residual error. Check if the Q value of \mathbf{X}_{new} is below the threshold and, if normal, update the PCA based on Gaussian Mixture Model (GMM). If an anomaly is detected, trigger, alert, and isolate the data without adding it to the learning dataset.

Step 3. GMM Clustering Based on Similarity: Use the newly added normal data to fit the GMM. Calculate the Bayesian Information Criterion (BIC) for each GMM model and select the optimal GMM with the lowest BIC.

Step 4. Recursive Baseline Update: Add normal state data not present in the initial training dataset and continuously learn about normal states, ensuring the baseline model and thresholds exhibit time-dependent characteristics. This adaptive thresholding allows for effective anomaly detection over time.

2.2. Determination of threshold for autonomous anomaly detection

PCA is a widely used technique for reducing dimensions by identifying principal components (PCs) that capture the most significant data variance based on correlations among variables [24–26]. This method projects the original data onto a hyperplane that maximizes variance while minimizing information loss. A small number of PCs can account for most of the total variability, with PC_1 typically representing the greatest variability and explaining the key factors of change in the

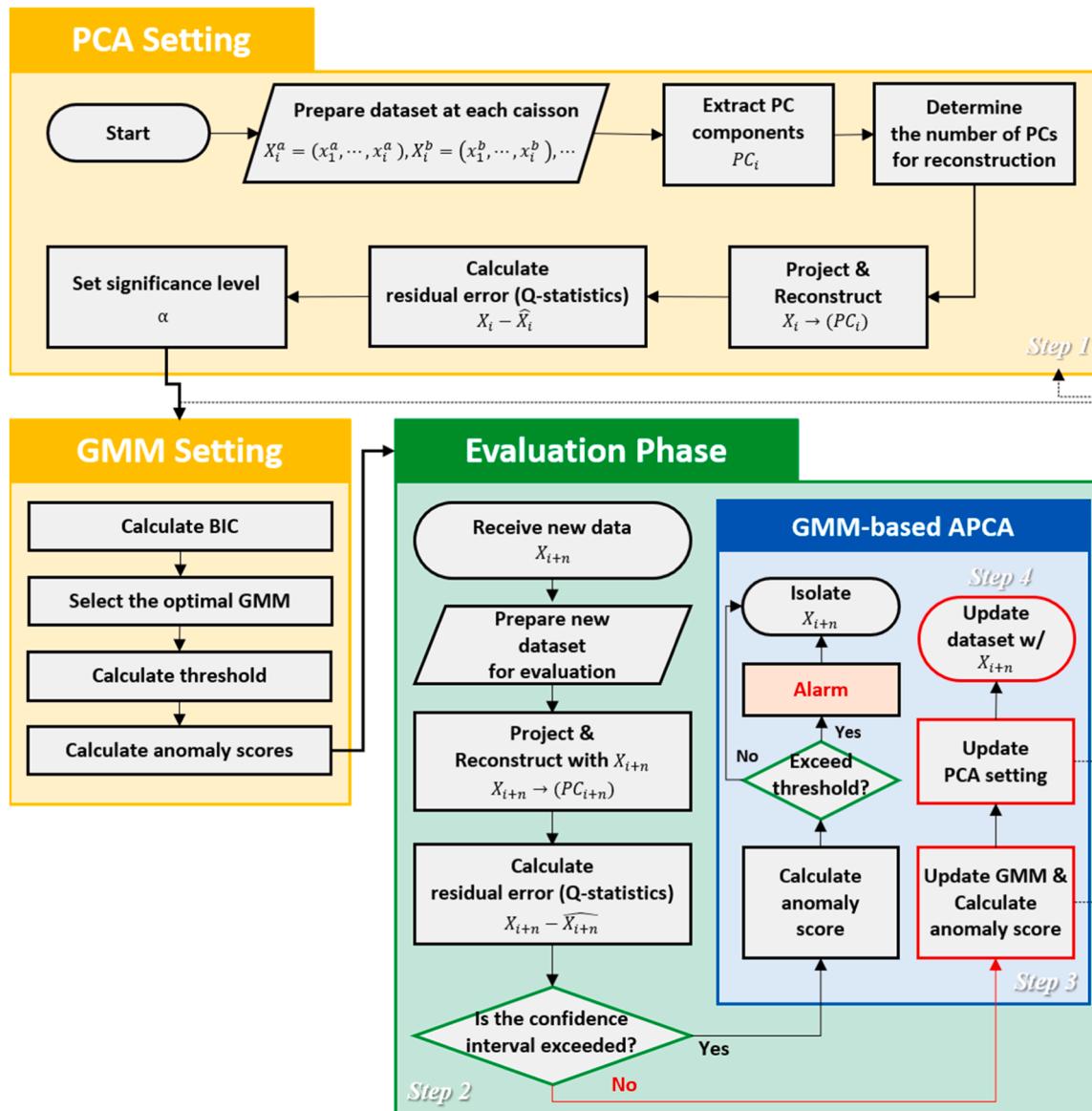


Fig. 1. Flowchart of proposed anomaly detection method.

dataset.

Mathematically, PCA begins by computing the covariance matrix C of the original data. The covariance matrix summarizes the relationships between different features (variables) in the dataset. It is defined as follows:

$$C = \frac{1}{n}(\mathbf{X} - \bar{\mathbf{X}})^T(\mathbf{X} - \bar{\mathbf{X}}) \quad (1)$$

where, \mathbf{X} is the data matrix, with each row representing a data point and each column representing a feature, $\bar{\mathbf{X}}$ is the mean vector of the data, computed across each feature, and n is the number of data points.

PCA then proceeds to find the eigenvectors (\mathbf{V}_i) and corresponding eigenvalues (λ_i) of the covariance matrix C . The eigenvectors represent the directions of maximum variance, while the eigenvalues quantify the amount of variance explained by each eigenvector. The eigenvectors and eigenvalues are computed such that:

$$\mathbf{C}\mathbf{V}_i = \lambda_i \mathbf{V}_i \quad (2)$$

The eigenvectors are typically ordered by their corresponding eigenvalues in decreasing order, such that the eigenvector with the largest

eigenvalue captures the most variance in the data. By selecting the top k eigenvectors (where k is the desired dimensionality of the reduced space), PCA constructs a projection matrix \mathbf{W} consisting of these eigenvectors. This projection matrix is used to transform the original data onto a lower-dimensional subspace spanned by the selected eigenvectors.

In many SHM studies, PCA is employed to compensate for temperature variations, as these fluctuations primarily influence the behavior of the signals [13,26–32]. The eigenvalues and eigenvectors derived from PCA represent the principal components that capture the dominant patterns and variances in the data, including those related to temperature changes.

The anomaly scores can be computed using a widely adopted statistical approach known as Q-Statistics (SPE, squared prediction error) [29–33], which assesses reconstruction errors. The reconstruction error in Q-Statistics represents the deviation observed when projecting data back into the original space and is computed as the Euclidean distance between the original and reconstructed data. The squared error for each data point, measured as SPE, reflects the discrepancy between the point and the original dataset. Structural anomalies are detected by evaluating deviations that arise during data reduction to principal components and

subsequent reconstruction. The methodology for computing SPE is outlined as follows:

$$\mathbf{Q}(\tilde{\mathbf{X}}, \mathbf{P}_{1:r}) = \tilde{\mathbf{X}}(\mathbf{I} - \mathbf{P}_{1:r}\mathbf{P}_{1:r}^T)\tilde{\mathbf{X}}^T = \|\tilde{\mathbf{X}} - \hat{\mathbf{X}}\|^2 \quad (3)$$

where $\tilde{\mathbf{X}}$ represents the scaled data, $\mathbf{P}_{1:r}$ represents the PCs extracted through the eigen-gap technique, and $\hat{\mathbf{X}}$ denotes the reconstructed data using the first r PCs [34]. Based on the derived reconstruction error, a threshold for anomaly detection can be calculated as follows:

$$Q_{lim,\alpha} = \frac{\theta_2}{2\theta_1} \chi_{\alpha}^2(\mathbf{h}) \quad (4)$$

where θ_1 is the sample mean, θ_2 is the sample variance, and $\chi_{\alpha}^2(\mathbf{h})$ is the chi-squared distribution with \mathbf{h} degrees of freedom and a significance level of α .

2.3. Adaptive online updating of threshold for long-term monitoring

The limitations of PCA arise from its reliance on static datasets [35]. Typically, PCA is beneficial for understanding the main structures of data and reducing dimensions based on static datasets. However, data is dynamic and can change over time in real-world scenarios. When applying PCA to dynamic datasets where data evolves, the model may only consider past data for learning, potentially missing out on capturing evolving data patterns. This is particularly evident in datasets like time series data, where data changes over time, making it challenging for PCA to adapt flexibly to these changes.

Adaptive PCA (APCA) addresses these issues by incorporating the concept of online learning, moving windowing [36–38], and recursive adaptation techniques [39,40]. The moving window technique adjusts the variations in normal states as data is incrementally added. Essentially, it involves continuously updating the moving window to accommodate changes in the normal state by discarding the oldest data and incorporating the latest data. This allows the system to adapt to changing normal state variations. On the other hand, the recursive adaptation technique employs recursive formulas to mitigate the influence of outdated data when updating the baseline. Using recursive calculations alleviates the impact of older data on baseline updates.

However, these techniques are based on linear PCA, which may not effectively capture the nonlinearity of long-term data, leading to an increased false alarm rate in anomaly detection when nonlinearity is present. Jin and Jung (2018) addressed this issue by proposing online learning with a K-means clustering technique that segments long-term monitoring data exhibiting nonlinearity into bi-linear sections [29]. However, it is noted that K-means clustering is limited by its assumptions of spherical and equally sized clusters, its sensitivity to initial centroid placement, and its poor performance in the presence of outliers, noise, and non-linear data structures [41].

To address these limitations, the algorithm proposed in this paper employs a Gaussian Mixture Model (GMM). GMM models the data as a collection of multiple Gaussian distributions, allowing it to handle clusters of varying shapes and sizes, better accommodate outliers and noise, and effectively capture the underlying patterns in non-linear data structures. By assuming that the data consists of several clusters, each following a Gaussian distribution, GMM provides a more flexible and robust clustering approach than K-means [42–44]. GMM is composed of K Gaussian components, and its density function is derived according to the formula as follows:

$$f(\mathbf{x}) = \sum_{k=1}^K \pi_k N(\mathbf{x} | \boldsymbol{\mu}_k, \Sigma_k) \quad (5)$$

where, $\boldsymbol{\mu}_k$ represents the mean of the k -th component, Σ_k represents the covariance matrix for the k -th component, π_k represents the mixing coefficient corresponding to the k -th component, and $N(\mathbf{x} | \boldsymbol{\mu}_k, \Sigma_k)$ is the probability distribution of the k -th component with respect to the

variable \mathbf{x} .

One of the key parameters of the GMM is the number of clusters, determined by using the Bayesian Information Criterion (BIC). BIC evaluates the performance of a model by considering both the complexity of the model and the fit of the data. It is derived according to the formula below:

$$BIC = p \cdot \log(n) - 2 \cdot \log(\hat{L}) \quad (6)$$

where, p represents the number of parameters in the model, n is the number of sample data points, and \hat{L} denotes the maximum likelihood of the model. Fig. 2 is an illustration showing the baseline being adaptively updated using GMM Clustering.

With APCA, the model is updated whenever new data arrives, allowing it to adapt in real-time to changes in the dataset. By dynamically updating the model based on incoming data, APCA offers a flexible adaptation to evolving data patterns. This online learning approach continuously discerns abnormal states from continuously measured data. When a state is classified as normal, it is added to the learning database to encompass various normal states. Conversely, when measured data is deemed abnormal, it is isolated without being added to the learning data. This continual model learning method enables the addition of normal state data that was not present in the initial learning data, allowing for continuous learning of normal states. Additionally, the baseline model and thresholds exhibit time-dependent characteristics, adapting to changes over time. As a result, unlike traditional learning methods, the thresholds are adaptively determined over time, enabling the detection of abnormal states. Through the online learning approach, effective resolution of errors in learning data composition can be achieved, and the onset of abnormal states can be promptly detected.

3. Long-term SHM in a caisson quay wall

3.1. Target structure

Caisson-type quay walls account for 18 % of mooring and berthing facilities and 24 % of breakwater facilities in South Korea. To validate the proposed method, a quay wall in the Port of Incheon, constructed in 2018 and managed by the Incheon Port Authority, was selected as a case study. This quay wall spans 850 m, with two berths for 50,000 DWT vessels and one berth for 30,000 DWT vessels. The cap concrete is approximately 20 m wide, and the caissons are about 40 m in width, with two cap concretes mounted on each caisson. Despite being less than 10 years old, the facility requires long-term monitoring due to various operational events. As this is a security facility, detailed descriptions are restricted. Fig. 3 shows the foreground of the target structure.

3.2. Sensors and their installation

Previous studies have shown that each caisson in caisson-type structures can behave differently, exhibiting movements like sliding and slipping. Since caissons are positioned independently, they may respond individually to external loads such as waves and backfill pressure [45]. In this study, sensors were installed on two cap concretes (approximately 5.5 m thick), each influenced by different caissons, to monitor their behavior.

Fig. 4 describes an overview of the monitoring system. The sensors were placed on two caissons, labeled C.A. #2 and C.A. #4. Three parameters were monitored for each: sliding, spacing between cap concretes, and settlement. Measuring sliding was particularly challenging due to safety concerns. Linear displacements like settlement and spacing were recorded using a crack meter (Tokyo Measuring Instruments Lab; KG-5A) mounted at the joints between the cap concretes. Inclinometers (ZIS I&C; STM-5) were installed on the seaside of the structure, with two per location. The inclinometer positioned at the bottom was influenced by tides. All sensors were bolted to the structure for stability and housed

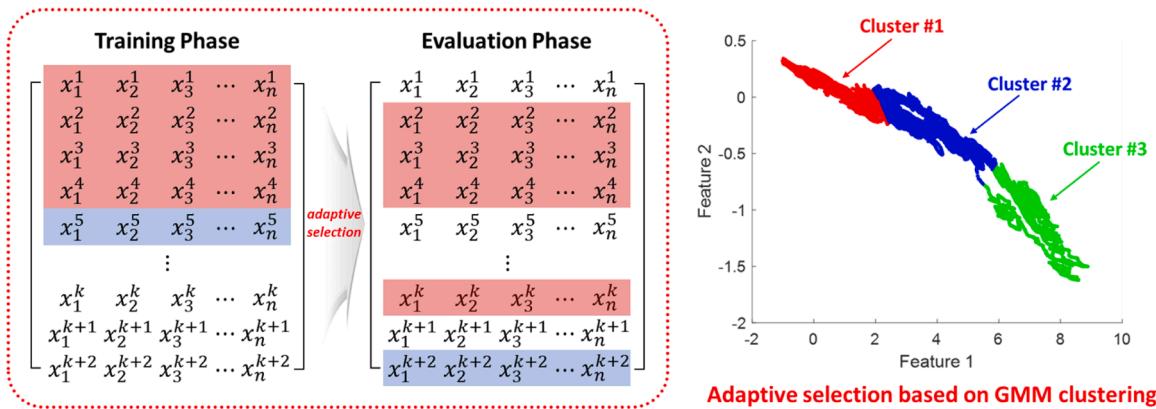


Fig. 2. Illustration of adaptive baseline update with GMM.

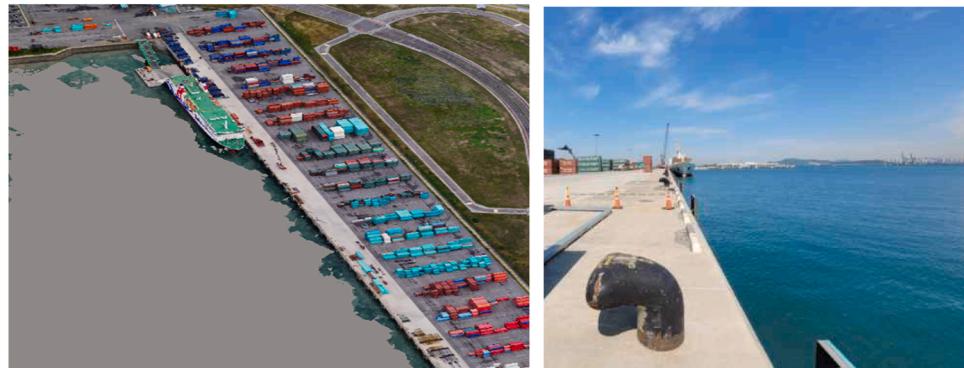


Fig. 3. The target structure.

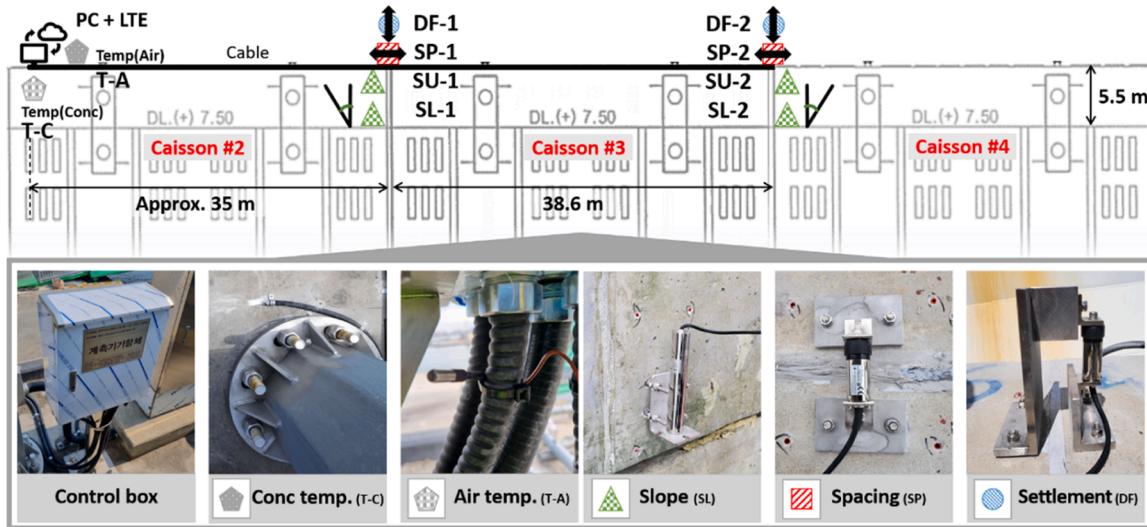


Fig. 4. SHM system for a target structure.

in waterproof and dustproof aluminum casings. The spacing and settlement meters had a precision of 0.001 mm, while the inclinometers had a precision of 0.0025°, operating within a temperature range of -20 to 60 °C. The data acquisition system (DAQ) and PC were located 40 m from the sensors to ensure a stable power supply. Data was stored on the PC and transmitted to the administrator via LTE communication. Two thermocouple T-type thermometers were also installed to monitor the temperature of the concrete and ambient air.

Due to the nature of gravity-type quay walls, which are not suited for high-rate dynamic measurements, the sampling rate was set to 10 min. Tidal data, collected at 1-minute intervals from a public website (<http://www.khoa.go.kr/oceangrid>), was used to assess the correlation between the measured data and tidal levels. Tidal data and sensor measurements were synchronized for analysis.

3.3. Measured signals

The target structure was monitored over 13 months from October 1, 2022, to October 31, 2023. The data are presented in Fig. 5. Fig. 5 shows: (a) presents the spacing between the cap concrete, (b) relative settlement, (c) upper side slope, (d) lower side slope, (e) member and air temperature, and (f) tidal level. Sensor maintenance was conducted from February 1 to 6, 2023, during which no measurements were conducted. As shown in the overall graphs, over a year of measurements confirmed that seasonal effects are clearly evident in this structure, as with other SOC facilities. In particular, given that port structures are massive structures primarily made of concrete, the effects of concrete expansion and contraction were significantly reflected in the spacing data. Fig. 6 shows examples of one-week data, illustrating variability based on temperature during clear days. Short-term variability due to daily temperature changes was observed, but the effect by the tide level was not clearly observed. Table 1 presents the statistical values of each sensor measurement data.

The spacing between the cap concretes at C.A. #2 and C.A. #4 increased to a maximum of 12.028 mm. Throughout the measurement period, the spacing range for both locations was similar, spanning from -2.150 mm to 10.206 mm. Additionally, a difference in settlements between the two sites was observed during winter: the settlement at C.A. #2 decreased to -1.617 mm (DF-1), while at C.A. #4, it increased by approximately 0.343 mm (DF-2). The slope data measured at the top varied from -0.135° to 0.092°, showing consistency across both sensors. In comparison to linear displacements such as spacing and settlement, the slope exhibited a relatively lower seasonal effect. This variation is likely due to differences in sensor sensitivity, leading to the conclusion that rotational displacements (slope) are less affected than linear displacements, given the concrete's drying shrinkage characteristics with temperature changes. However, for inclinometers located near sea level (SL-1, SL-2), there was noticeable fluctuation in values.

3.4. Long-term behavior of caisson

South Korea experiences distinct seasons, with temperature differences between summer and winter reaching 50 °C and tidal ranges on the west coast around 8–10 m. These factors significantly influence SHM data over a year. Fig. 7 shows the correlation between measured signals and temperature, with cap concrete spacing showing the highest correlation and slope the lowest. The correlation shifted around an air temperature of 20 °C. Linear regression analysis between SP-1 and SP-2 and temperature yielded correlation coefficients R^2 of 0.91 and 0.82, respectively, while for DF-1 and DF-2, they were 0.82 and 0.0009, and for SU-1 and SU-2, 0.73 and 0.20. This suggests that relative settlement differences were not solely due to temperature. For DF-1, a clear bilinear relationship with temperature was observed, but deviations were noted for DF-2. These results indicate that anomaly detection from the long-term monitoring data should be carefully performed considering the different boundary conditions, superimposed load conditions, and operational conditions of caissons.

Despite the large tidal range, the impact of tides on the structure was minimal, with correlation coefficients ranging from 1.25E-5 to 6.9E-3. Since the quay wall is mass concrete with a safety factor above 1.2, tidal effects are insignificant unless wave heights exceed design limits. More accurate analysis would require time-synchronized tidal data. The SHM data revealed that temperature was the most significant influence, but each caisson had different effects on the variables measured. Detecting anomalies from multiple physical quantities requires setting adaptive threshold values that account for complex variability. Although deep learning and machine learning models have been developed for anomaly detection using single measurement sensors, caisson-type structures require safety assessment based on multiple variables under varying conditions.

When establishing an SHM system for SOC and building facilities, either a three-sigma criterion (about 99.73 %) is commonly applied or anomaly detection is performed based on a fixed threshold value set by

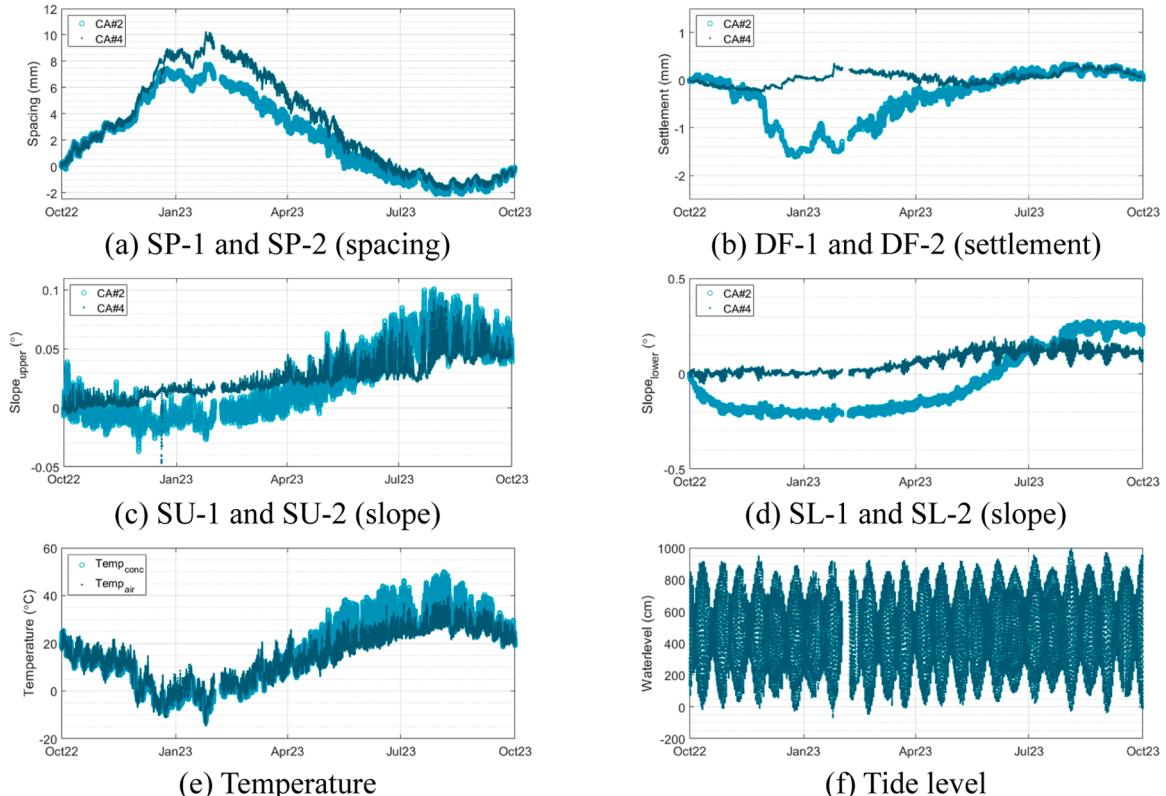


Fig. 5. Measured signals.

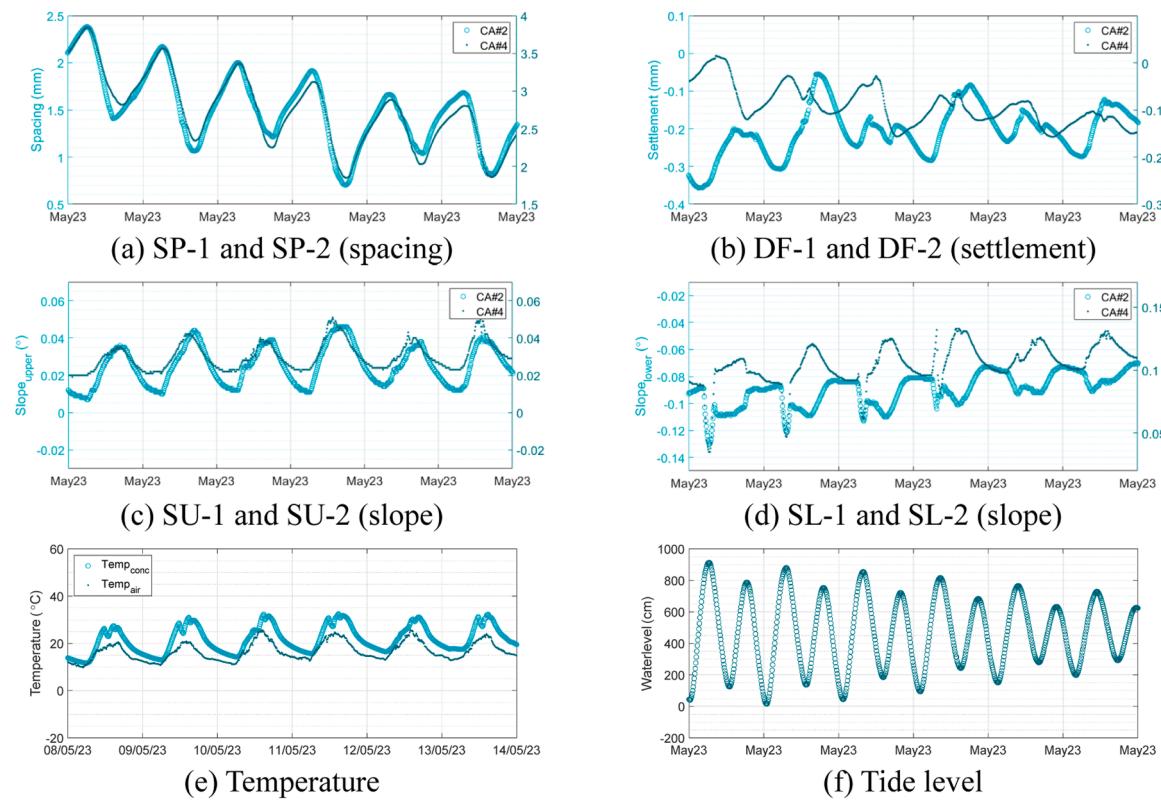


Fig. 6. Examples of 1-week measured data.

Table 1

Statistics for each sensor over one year.

Sensor ID		Unit	Max. value	Min. value	Difference	μ	σ
SP	SP-1	mm	7.763	-2.150	9.913	2.254	2.9565
	SP-2	mm	10.206	-1.822	12.028	3.195	3.562
DF	DF-1	mm	0.348	-1.617	1.965	-0.353	0.550
	DF-2	mm	0.343	-0.248	0.591	0.0399	0.134
SU	SU-1	degree	0.101	-0.037	0.138	0.020	0.029
	SU-2	degree	0.092	-0.135	0.227	0.025	0.016
SL	SL-1	degree	0.277	-0.245	0.522	-0.045	0.171
	SL-2	degree	0.191	-0.053	0.244	0.064	0.0547
Temp.	Concrete	°C	49.940	-13.130	63.07	15.545	13.229
	Air	°C	39.500	-14.540	54.04	14.484	10.676
Tide level		cm	990	-65	1055	468.349	222.340

an initial training dataset using a one-shot learning static reference framework [46]. The three-sigma threshold provides a broad safety margin, making it less sensitive to minor changes or noise, which reduces the probability of false alarms in anomaly detection. However, even with the use of a three-sigma criterion, if baseline data that sufficiently reflects the seasonal effects is not secured, false alarms due to temperature variation can occur. For port structures, determining whether the displacement is progressive is crucial for applying appropriate management threshold values to assess safety. Therefore, during long-term monitoring, it is essential to minimize the effects of temperature to accurately evaluate the progressiveness of displacement [47]. This means that the threshold should be adaptively determined in consideration of environmental variability, with high sensitivity to anomalies in the field.

4. Validation of the proposed anomaly detection method

4.1. Anomalous state scenarios for validations

To validate the performance of the proposed algorithm, scenarios simulating anomalous states of varying difficulty levels were designed. The types of anomalies on the quay wall were described in previous studies [5,6]. When sliding or slope changes occur, the signals exhibit an abrupt drift rather than a smooth transition. This study introduced different ranges of artificial drift within the temperature fluctuations applied to the monitoring dataset. The dataset consisted of measurements from multiple sensors installed on each caisson unit, including temperature, slope, spacing, and settlement. It was designed based on the understanding that each caisson exhibits similar structural behaviors and is influenced by comparable environmental conditions.

The scenarios were classified into three types: normal (healthy), single-damaged, and multi-damaged. It is important to note that slope data has relatively low variability and high noise, spacing data is highly dependent on temperature, and settlement data shows irregular,

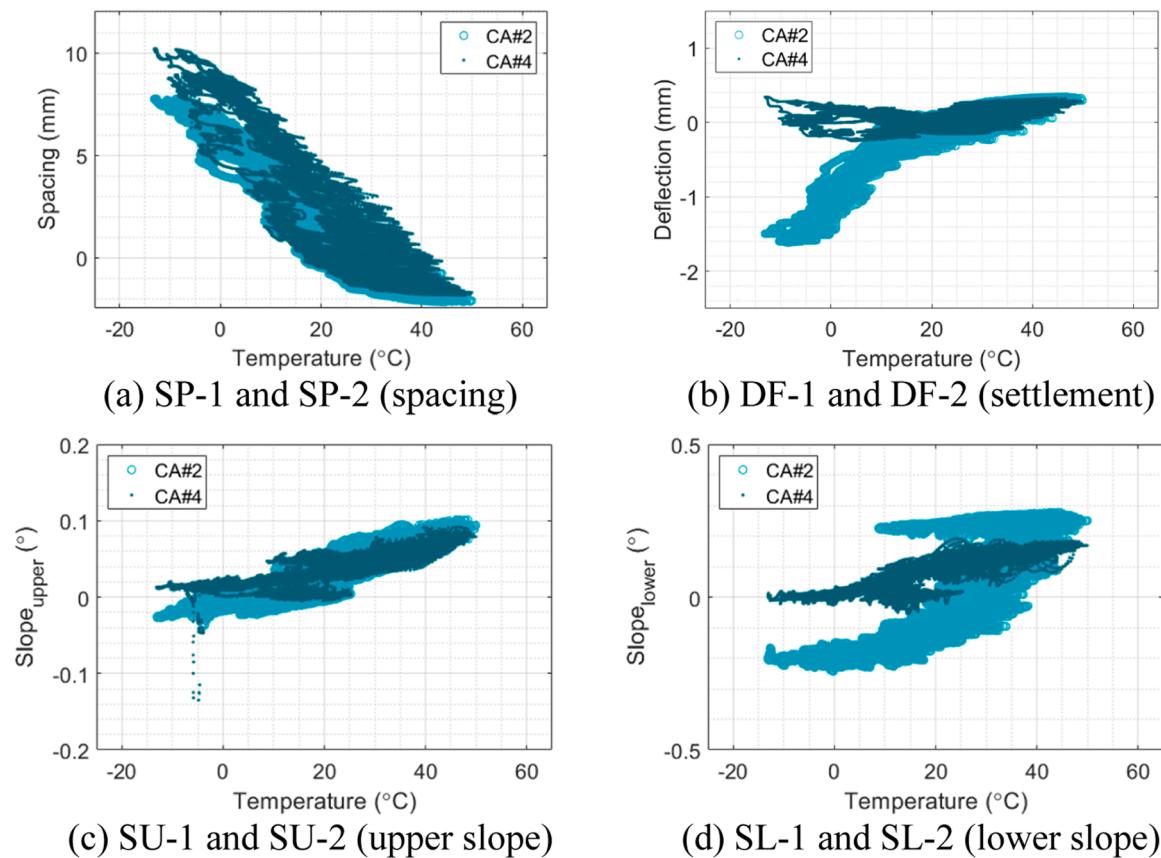


Fig. 7. Correlation between measurement signals and concrete temperature.

complex variations. Scenarios #1 and #2 are normal cases. Scenarios #3 and #4 involve a 1-sigma anomaly in spacing and slope data, respectively. Scenario #5 includes a 0.3-sigma anomaly in slope data, which is hard to detect from the raw signal. Scenario #6 involves both linear displacements (spacing and settlement) at the same caisson, while Scenario #7 adds rotational displacement to Scenario #6. These scenarios were designed to reflect the complex anomalies likely to occur in real structures. For example, settlement due to caisson foundation issues can cause simultaneous spacing and settlement changes, and settlement can also lead to slope changes. The seven scenarios are outlined in Table 2, and Figs. 8 and 9 show examples of the anomalous conditions.

4.2. Effect of the initial training dataset

In long-term monitoring, the selection of the initial training dataset is critical because it can affect the accuracy and sensitivity of the anomaly detection algorithm. Therefore, the performance of the proposed algorithm was first examined with different lengths of the initial training dataset.

Anomalies at a level of 1 sigma of the initial training data were

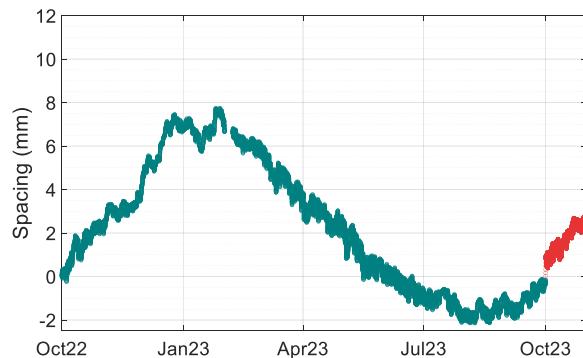


Fig. 8. A sample of 0.3σ anomalies in SP-1.

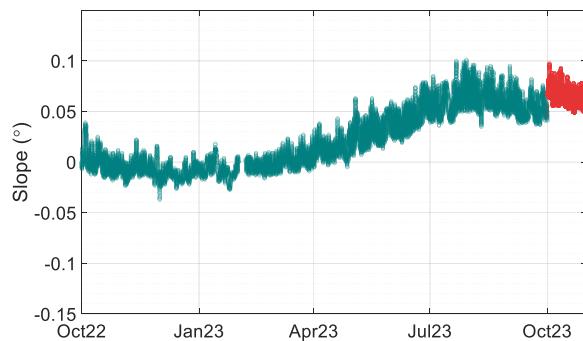


Fig. 9. A sample of 1.0σ anomalies in SU-1.

Table 2
Scenarios of various anomalous states.

No.	Descriptions	Class
1	Normal condition (C.A.#2)	Normal
2	Normal condition (C.A.#4)	(healthy)
3	Anomalous condition of the 1-signal level in SP-1	Sigle-damaged
4	Anomalous condition of the 1-signal level in SU-1	damaged
5	Anomalous condition of the 0.3-signal level in SP-1	
6	Anomalous condition of the 0.3-signal level in SP-1&DF-1	Multidamaged
7	Anomalous condition of the 0.3-signal level in SP-1&DF-1&SU-1	

applied based on 1 year, 6 months, 3 months, and 1 month of SP-1 signal. Fig. 10 shows the anomaly score calculated from the proposed method depending on the initial training dataset. The blue dot means the training dataset, the red dot means the result determined to be abnormal, and the black dot means the result determined to be healthy. Since the spacing data exhibited significant variations due to concrete expansion in the summer, the detection performance varied depending on whether this data was included in the initial training data. When only 1 month of data was used for training, behaviors due to temperature were detected as anomalies; whereas, including a part of the temperature-related behavior significantly reduced the number of false alarms (Fig. 10(b) to (d)). In this figure, the false alarm rate decreased to 0.6 % when using 6 months of data and to 0.005 % when using 12 months of data.

The proposed algorithm confirmed that anomaly detection could be realized despite using data for more than 3-months as a baseline. The online learning method allows for the base model and its threshold to adaptively change as initial training data is added and expanded. As a result, once a certain length of initial training data is secured, the occurrence of false positives significantly decreases, and anomalies are accurately detected in real time.

4.3. Comparative analysis of previously proposed methods

To validate the performance of the proposed technique in this paper, we compared it with static linear PCA (SPCA), incremental linear PCA (IPCA), and online learning with K-means clustering as proposed by Jin and Jung (2018) [29]. Here, SPCA refers to the method where the threshold is determined solely based on the initial training dataset without any online updating, while IPCA refers to the method that performs online updating. Jin and Jung (2018) proposed the adaptive statistical process monitoring using an online learning algorithm with the variable moving window strategy (ASPM).

Anomaly detection was conducted under the 0.3 sigma anomalous condition (Scenario #5) from Table 2, using initial training datasets of 6 and 12 months of data. Fig. 11 presents the anomaly detection results using four different methods. With 6 months of data, the false alarm rates were 37.77 % for SPCA, 18.15 % for IPCA, 14.36 % for ASPM, and 5.78 % for the proposed method. The detection accuracy of the proposed method improved by 8.58 %, compared to ASPM. Using 12 months of data, the false alarm rates were 7.39 % for SPCA and 7.95 % for IPCA, indicating a significant improvement in accuracy compared to the 6-month data, as the entire year's fluctuation data was used for training. For ASPM, the false alarm rate was 0.245 %, while the proposed method

achieved 0.23 %. Comparing Figs. 11(g) and 11(h), there was a performance difference in the timing of anomaly detection between ASPM and the proposed method. The 0.015 % difference in detection accuracy between the two methods was due to their ability to detect anomalies at the moment they occur, confirming the superior accuracy of the proposed method.

4.4. Anomaly detection results for various scenarios

The proposed algorithm was applied to set an adaptive threshold with a confidence interval set at 99 %. The anomaly score was calculated for each caisson, same as the actual port inspection, and it allows for the management of anomalies occurring in each caisson. The results of anomaly detection were presented in Fig. 12.

For Scenario #1, the proposed algorithm results revealed that the anomaly score for all data points was below the threshold, thus indicating that they were determined as normal conditions. Even if the newly measured raw data point was located out of the range of the initial training dataset, the principal axis of the new dataset remained in the range of normal conditions and so this was updated with the training dataset. In Scenario #2, except for detecting a short anomalous state that occurred temporarily in December 2022 in SU-2, all were judged as normal.

From Figs. 12(c) and 12(e), the detection performance according to the sigma level can be compared. Scenario #5, with 0.3 sigma applied, exhibited a less distinction between normal and anomalous data in terms of the anomaly score when compared with Scenario #3, but accurately detected the anomaly immediately after the anomaly score surpassed the threshold. The results indicate that the proposed algorithm is capable of detecting small anomalies at the caisson unit at a level of 0.3 sigma, even in signals containing various variabilities. Scenario #4 involved a case of 1-sigma rotational displacement in SU-1 data, which had a relatively small measurement range and high noise level when compared with other sensors. The proposed algorithm detected anomalies with several false alarms before and after the abnormal point, but was able to detect anomalies with high accuracy.

Scenarios #6 and #7 are cases wherein anomalies were applied to multiple sensor data. When anomalies occur in multiple sensors, the principal axes may simultaneously change and it decreases the detection performance. However, as shown in Figs. 12(f) and 12(g), the proposed algorithm showed robust performance in anomaly detection despite multiple complex anomalous states.

Considering that port structures including mooring and berthing facilities span hundreds of meters to kilometers, many sensors should be

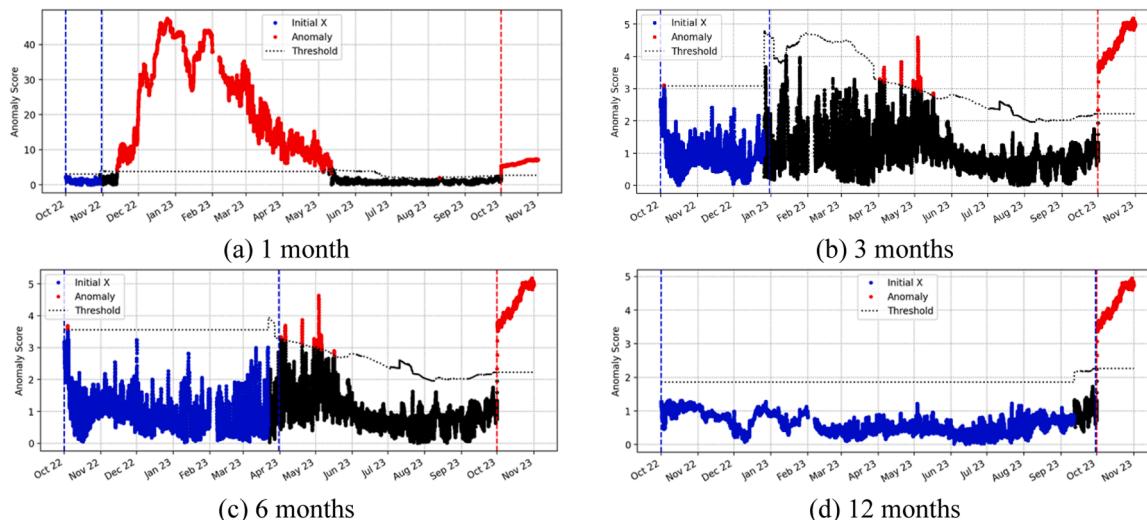


Fig. 10. Effect of the initial training dataset for the proposed method.

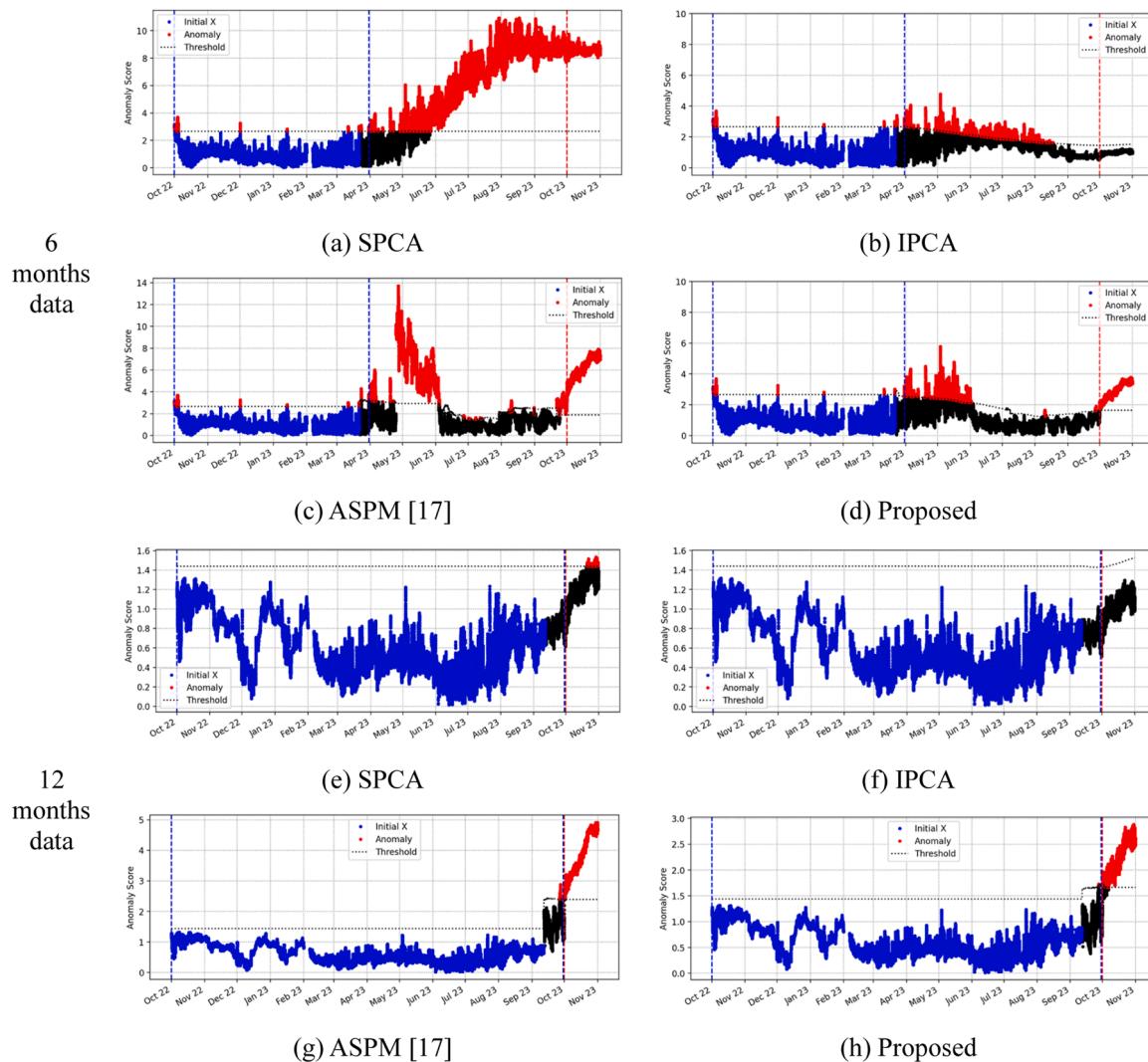


Fig. 11. Comparison of results depending on the initial training dataset.

installed and the volume of monitoring data is significantly large; thus, the automation of anomaly detection algorithms plays a critical role. Furthermore, due to the waves, temperature, tides, ships, cranes, and heavy vehicles on port structures, there are limitations to managing anomalous states with fixed thresholds and identifying signals individually. The online learning-based anomaly detection technique proposed in this study could adaptively and autonomously decide the thresholds and accurately detect anomalies with an average accuracy of 99.90 %, thereby reducing the occurrence of false alarms and enabling stable safety management of the facility.

5. Conclusions

This study conducted long-term monitoring of caisson-type quay walls for slope, settlement, spacing, and air/concrete temperature over a year in South Korea. It represents the first SHM system for caisson-type quay walls. The multiple monitored data showed significant variations influenced by environmental factors such as temperature and boundary conditions among caissons. To address these variations and set robust thresholds autonomously, this study proposed a GMM-based adaptive PCA algorithm applicable to each caisson equipped with multiple sensors. The effectiveness of the proposed algorithm was evaluated by inducing various artificial anomalies, reflecting the types of anomalies observed in caisson-type facilities. The results showed the following:

- 1) Long-term monitoring was conducted on caisson quay walls for over a year. Slope, settlement, concrete block spacing, and temperature were measured on two different caissons using multiple sensors.
- 2) Correlation analysis showed that temperature had a significant effect on the measured parameters, while tidal levels had little correlation with the structural behaviors. A new anomaly detection method was needed for the SHM of caisson-type structures to address data variability due to environmental factors, including temperature variations, and to ensure high sensitivity to anomalies while reducing false alarms.
- 3) In long-term SHM, the length of the initial training dataset is crucial for the efficiency of anomaly maintenance. The proposed method effectively detected anomalies with data collected over a period of more than 3-months and demonstrated 100 % accuracy with a 12-month dataset.
- 4) The proposed GMM-based adaptive PCA method demonstrated high effectiveness, adaptively setting thresholds and detecting anomalies with 99.90 % accuracy across various scenarios. This approach significantly reduced false alarms and offered more stable and efficient anomaly detection compared to SPCA, IPCA, and the online learning method using K-means clustering proposed by Jin and Jung (2018).

For port facilities constructed on reclaimed land, monitoring

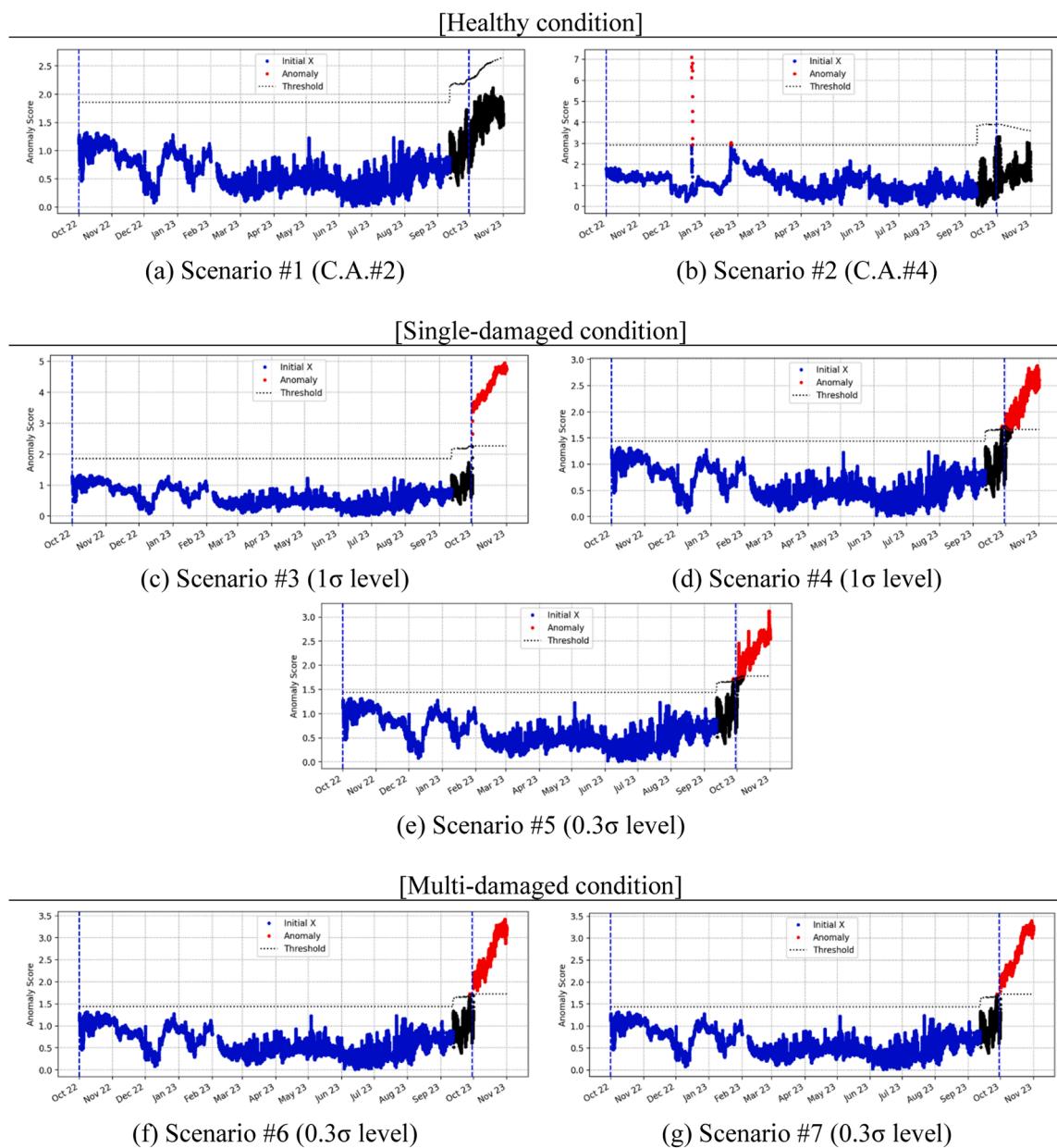


Fig. 12. Results of the proposed method according to scenarios.

displacements such as slope, sliding, and settlement is crucial for ensuring the safety and usability of the structure. In particular, when displacement progresses over time, the facility requires special management. In long-term SHM, displacements may not always be clearly reflected in the measured signals due to surrounding environmental factors, which is why the proposed anomaly detection method is necessary. In the future, analytical studies should be conducted to further explore the close relationship between the safety of the structure and the measured signals.

CRediT authorship contribution statement

Taemin Lee: Writing – original draft, Methodology, Conceptualization. **Seung-Seop Jin:** Validation, Software, Methodology. **Sung Tae Kim:** Resources, Investigation, Data curation. **Jiyoung Min:** Writing – original draft, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jiyoung Min reports financial support was provided by Korea Ministry of Oceans and Fisheries. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the Korea Institute of Marine Science & Technology Promotion [grant number 20210659] funded by the Ministry of Oceans and Fisheries.

Data availability

The data that has been used is confidential.

References

- [1] Nolan S, Rossini M, Knight C, Nanni A. New directions for reinforced concrete coastal structures. *J Infrastruct Preserv Resil* 2021;2:1–12.
- [2] Hou B, Li X, Ma X, Du C, Zhang D, Zheng M, et al. The cost of corrosion in China. *NPJ Mater Degrad* 2017;1(1):4.
- [3] Koch GH, Brongers MP, Thompson NG, Virmani YP, Payer JH. Corrosion cost and preventive strategies in the United States (No. FHWA-RD-01-156, R315-01). United States: Federal Highway Administration; 2002.
- [4] van der Wel, T. (2018). Reliability based Assessment of Quay Walls.
- [5] Jin S, Min J, Kim Y, Kim R. Development of Online-learning based Adaptive Anomaly Detection Algorithm for Monitoring Data Analysis on Caisson Type Breakwater. *J Coast Disaster Prev* 2023;10(1):1–12.
- [6] USACE Publications. (2006). Coastal Engineering Manual.
- [7] Ministry of Land. (2018). Infrastructure and Transport. Detailed guidelines for safety and maintenance of facilities (performance evaluation).
- [8] Ni F, Zhang J, Noori MN. Deep learning for data anomaly detection and data compression of a long-span suspension bridge. *Comput Civ Infrastruct Eng* 2020;35(7):685–700.
- [9] Li J, He H, He H, Li L, Xiang Y. An end-to-end framework with multisource monitoring data for bridge health anomaly identification. *IEEE Trans Instrum Meas* 2020;70:1–9.
- [10] Turrisi S, Cigada A, Zappa E. A cointegration-based approach for automatic anomalies detection in large-scale structures. *Mech Syst Signal Process* 2022;166:108483.
- [11] Kromanis R, Kripakaran P. Support vector regression for anomaly detection from measurement histories. *Adv Eng Inform* 2013;27(4):486–95.
- [12] Wang H, Barone G, Smith A. A novel multi-level data fusion and anomaly detection approach for infrastructure damage identification and localisation. *Eng Struct* 2023;292:116473.
- [13] Zhu Y, Ni YQ, Jin H, Inaudi D, Laory I. A temperature-driven MPCA method for structural anomaly detection. *Eng Struct* 2019;190:447–58.
- [14] Pan Q, Bao Y, Li H. Transfer learning-based data anomaly detection for structural health monitoring. *Struct Health Monit* 2023;22(5):3077–91.
- [15] Glashier T, Kromanis R, Buchanan C. Temperature-based measurement interpretation of the MX3D Bridge. *Eng Struct* 2024;305:116736.
- [16] Negi P, Kromanis R, Dorée AG, Wijnberg KM. Structural health monitoring of inland navigation structures and ports: a review on developments and challenges. *Struct Health Monit* 2024;23(1):605–45.
- [17] Broos, E.J. (2010). Design & construct contract for a deepsea quay wall in the Port of Rotterdam: Case study Brammen terminal.
- [18] den Adel, N. (2018). Load testing of a quay wall: Evaluating the use of load testing by application of Bayesian updating.
- [19] Schouten, O. (2020). Optimising the functionality of smart quay walls using measurement data obtained during the construction process: A case study in the port of Rotterdam: HHHT-quay.
- [20] Del Grosso, A., Inaudi, D., & Lanata, F. (2000, July). Strain and displacement monitoring of a quay wall in the Port of Genoa by means of fibre optic sensors. In 2d European Conference on Structural Control, ENPC, Paris.
- [21] Del Grosso A, Lanata F, Brunetti G, Pieracci A. Structural health monitoring of harbour piers. *SHMII* 2007;3:2007.
- [22] Pengel BE, Krzhizhanovskaya VV, Melnikova NB, Shirshov GS, Koelewijn AR, Pyat AL, et al. Flood early warning system: sensors and internet. *IAHS Red Book* 2013;357:445–53.
- [23] Bolourani A, Bitaraf M, Tak AN. Structural health monitoring of harbor caissons using support vector machine and principal component analysis. *Structures*, 33. Elsevier; 2021. p. 4501–13.
- [24] Maćkiewicz A, Ratajczak W. Principal components analysis (PCA). *Comput Geosci* 1993;19(3):303–42.
- [25] Bro R, Smilde AK. Principal component analysis. *Anal Methods* 2014;6(9):2812–31.
- [26] Granato D, Santos JS, Escher GB, Ferreira BL, Maggio RM. Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. *Trends Food Sci Technol* 2018;72:83–90.
- [27] Bisheh HB, Amiri GG. Structural damage detection based on variational mode decomposition and kernel PCA-based support vector machine. *Eng Struct* 2023;278:115565.
- [28] Ma Z, Luo Y, Yun CB, Wan HP, Shen Y. An MPPCA-based approach for anomaly detection of structures under multiple operational conditions and missing data. *Struct Health Monit* 2023;22(2):1069–89.
- [29] Jin SS, Jung HJ. Vibration-based damage detection using online learning algorithm for output-only structural health monitoring. *Struct Health Monit* 2018;17(4):727–46.
- [30] Li W, Peng M, Wang Q. False alarm reducing in PCA method for sensor fault detection in a nuclear power plant. *Ann Nucl Energy* 2018;118:131–9.
- [31] Camacho J, Pérez-Villegas A, García-Teodoro P, Maciá-Fernández G. PCA-based multivariate statistical network monitoring for anomaly detection. *Comput Secur* 2016;59:118–37.
- [32] Penha, R., & Hines, J.W. (2001, May). Using principal component analysis modelling to monitor temperature sensors in a nuclear research reactor. In Maintenance and reliability conference (MARCON 2001).
- [33] Nomikos P, MacGregor JF. Multivariate SPC charts for monitoring batch processes. *Technometrics* 1995;37(1):41–59.
- [34] Davis C, Kahan WM. The rotation of eigenvectors by a perturbation. *III SIAM J Numer Anal* 1970;7(1):1–46.
- [35] Malthouse EC. Limitations of nonlinear PCA as performed with generic neural networks. *IEEE Trans Neural Netw* 1998;9(1):165–73.
- [36] Wang X, Kruger U, Irwin GW. Process monitoring approach using fast moving window PCA. *Ind Eng Chem Res* 2005;44(15):5691–702.
- [37] Rafferty M, Liu X, Laverty DM, McLoone S. Real-time multiple event detection and classification using moving window PCA. *IEEE Trans Smart Grid* 2016;7(5):2537–48.
- [38] Liu X, Kruger U, Littler T, Xie L, Wang S. Moving window kernel PCA for adaptive monitoring of nonlinear processes. *Chemom Intell Lab Syst* 2009;96(2):132–43.
- [39] Li W, Yue HH, Valle-Cervantes S, Qin SJ. Recursive PCA for adaptive process monitoring. *J Process Control* 2000;10(5):471–86.
- [40] Jeng JC. Adaptive process monitoring using efficient recursive PCA and moving window PCA algorithms. *J Taiwan Inst Chem Eng* 2010;41(4):475–81.
- [41] G. Celeux Data Clustering: Theory Algorithms and Applications by Gan, G., Chaoyun, M. A., and Wu, J *Biometrics* 64 2 2008 656 657.
- [42] Reynolds DA. Gaussian mixture models. *Encycl Biom* 2009;741:659–63.
- [43] Rasmussen C. The infinite Gaussian mixture model. *Adv Neural Inf Process Syst* 1999;12.
- [44] Chaleshtori AE, Aghaei A. A novel bearing fault diagnosis approach using the Gaussian mixture model and the weighted principal component analysis. *Reliab Eng Syst Saf* 2024;242:109720.
- [45] Korea Institute of Ocean Science and Technology. (2014). Performance Evaluation of Interlocking Caisson-Type Breakwaters against Abnormal High Waves. Report No. PE99251.
- [46] Li W, Peng M, Wang Q. Improved PCA method for sensor fault detection and isolation in a nuclear power plant. *Nucl Eng Technol* 2019;51(1):146–54.
- [47] Ministry of Land, Infrastructure and Transport. Detailed Guidelines for the Safety and Maintenance of Facilities (Part of Performance based Evaluation). Jinju: Korea Authority of Land & Infrastructure Safety; 2021.