Civil and Project Journal, 2025, 7(10), 54-65
https://doi.org/10.22034/cpj.2025.553244.1408

Civil and Project Journal
http://www.cpjournals.com/

Damage detection in steel and concrete bridges under environmental and operational effects

Mohammadreza Mahmoudkelayeh[1], Behnam Adhami[2]*, Behzad Saeedi Razavi[3]

1- Department of Civil Engineering, CT.C., Islamic Azad University, Tehran, Iran.

2- Department of Civil Engineering, CT.C., Islamic Azad University, Tehran, Iran.*

3- Department of Construction and Mineral Engineering, Technology and Engineering Research Center, Standard Research Institute (SRI), Karaj, Iran

Received: 13 September 2025; Revised: 21 September 2025; Accepted: 18 October 2025; Published: 22 December 2025

Abstract
Environmental and operational changes, make the continuous monitoring of the health of civil engineering structures inaccurate and unreliable. The local unsupervised learning method based on double data clustering can help to solve this challenge. The main purpose is to extract the most relevant information insensitive to environmental and operational variations. The Local Density Peak Clustering under Minimum Spanning Tree (LDPC-MST) divides all available data points into main clusters. Using the representative sub-clusters of all main clusters, a damage detection indicator based on the Mahalanobis-squared distance is defined to detect any abnormal change caused by damage. Then, a steel arch bridge and a concrete box-girder bridge under strong environmental variations are investigated. Several comparative analyses are also performed to indicate the superiority of this method. The main innovation of this research is to develop a novel locally unsupervised learning method by using the process of double clustering and LDPC-MST. Results show that the proposed method is highly able to minimize the environmental and operational effects and provide reliable results.

Keywords: *Damage detection. continuous structural health monitoring; unsupervised learning; data clustering; environmental and operational variability*

Cite this article as Mahmoudkelayeh M., Adhami B., Saeedi Razavi B. (2025). 'Damage detection in steel and concrete bridges under environmental and operational effects', Civil and Project, 7(10), e232385. doi: https://doi.org/10.22034/cpj.2025.553244.1408

ISSN: 2676-511X /

Corresponding author E-mail address *beh.adhami@iauctb.ac.ir*

1. Introduction

Apart from the engineering challenge in continuous structural health monitoring (SHM) concerning the environmental and operational variability, an important technical challenge is related to the utilization of unsupervised learning for damage detection. In general, it is feasible to implement an unsupervised SHM method in terms of global and local algorithms. A global approach incorporates the whole (unlabeled) training data for developing an unsupervised learning model. In this regard, Jin et al. proposed an extended Kalman filter-based artificial neural network for eliminating temperature fluctuations from modal properties and performing accurate damage detection (Jin et al., 2016). Mario Azzara et al. presented the principal component analysis (PCA) and its nonlinear version, i.e., kernel PCA, for minimizing the influences of environmental variations on natural frequencies of a historical bell tower (Mario Azzara et al., 2018). Sousa Tomé et al. utilized an cointegration analysis for removing the environmental and operational effects on time series features obtained from structural responses of a cable-stayed bridge (Sousa Tomé et al., 2020). Qiu et al. proposed an improved dynamic Gaussian mixture model for damage detection in aircraft structures under changing environmental and operational conditions (Qiu et al., 2019). Saeedi Razavi et al. developed an automated hybrid algorithm in the time and frequency domains under unpredictable data nature in terms of simultaneously stationary or non-stationary behavior (Saeedi Razavi et al., 2021). Mahmoudkelayeh et al. proposed a locally unsupervised learning method based on a novel double data clustering and the LDPC-MST algorithm (Mahmoudkelayeh et al., 2024). Despite promising applications of global unsupervised learning methods, those may be influenced by any variability condition and noise. In particular, as these techniques incorporate the whole unlabeled data, the effect of noise and data variations may be problematic. Apart from this issue, the other limitation of the global methods is their computational inefficiency due to considering the whole data.

The major novelty of this research is to propose a locally unsupervised learning method based on a novel double data clustering and the LDPC-MST algorithm. The great advantage of this method is to significantly decrease the influences of the environmental and operational variations in continuous SHM programs. Real natural frequencies of a concrete box-girder bridge and a steel arch bridge under severe environmental and operational variations are used to verify the accuracy and good performance of the proposed method with some comparisons. Results demonstrate that the operational variability conditions and provide reliable results of continuous SHM.

2. Local Density Peak Clustering under Minimum Spanning Tree (LDPC-MST)

In the LDPC-MST method, one attempts to select some data points as useful features so that those are able to retain the shape of clusters and exclude the useless points, which may be influenced by any source of variability such as environmental and operational conditions. The LDPC-MST firstly makes attempt to find data points with local maximum density values, called LDPs, among their neighbors and allocates the remaining points to the corresponding LDPs. Second, it uses a shared-neighbor-based distance to measure the distance between the LDPs and their neighbors. Accordingly, one can construct the MST, which is a subgraph of the LDPs. Third, the method divides these LDPs into clusters so that the number of clusters has been pre-determined. Finally, the assigned data points to their LDPs are accommodated

into their clusters (Cheng et al., 2021). It is worth remarking that due to constructing the MST only on the LDPs of unlabeled data and also considering the theory of nearest-neighbor-search, it is possible to minimize the influence of useless features or outliers and improve the efficiency of the clustering process.

In order to find the LDPs and construct the MST, it is initially necessary to define the function of local density of data samples. In essence, the sum of distances between a sample in a dense region and its nearest neighbors is often smaller than the corresponding sample in a sparse region. Given a data sample **x,** its local density is defined as follows (Cheng et al., 2021):

$$\rho(\boldsymbol{x}) = \frac{N_\rho}{\sum_{\boldsymbol{y}\in NN(\boldsymbol{x})} d(\boldsymbol{x},\boldsymbol{y})} \qquad (1)$$

where N_ρ is the number of (reverse) nearest neighbors of **x**; d(**x**,**y**) denotes the distance between the samples **x** and **y**, which can be gained by the well-known Euclidean distance function; NN(**x**) represents the set of nearest neighbors of the sample **x**. To better realize the variable N_ρ, one needs to describe the idea of the reverse nearest neighbors.

The next step is to find the LDPs based on the concept of local density and the nearest neighbors of each sample. For this purpose, it only suffices to find a data point with the maximum local density among the original sample (**x**) and its nearest neighbors. After determining the main LDPs of the unlabeled sampling data, it is required to calculate their distances via the shared-neighbor-based distance for constructing the MST. Supposing that **x** and **y** are two LDPs, the function of their shared-neighbor-based distance is expressed as follows (Cheng et al., 2021):

$$d_{SN} = \begin{cases} \dfrac{d(\boldsymbol{x},\boldsymbol{y})}{N_\delta \times \sum_{\boldsymbol{v}\in\delta(\boldsymbol{x},\boldsymbol{y})}\rho(\boldsymbol{v})}, & if\ |\delta(\boldsymbol{x},\boldsymbol{y})| \neq 0 \\ max(d)\times\big(1+d(\boldsymbol{x},\boldsymbol{y})\big), & if\ |\delta(\boldsymbol{x},\boldsymbol{y})| = 0 \end{cases} \qquad (2)$$

where N_δ denotes the number of the shared neighbors between two LDPs (i.e., **x** and **y**), which are the intersections of their neighbors and fall in the set δ(**x**,**y**); ρ(**v**) is the local density of the sample **v**, which is one of the shared neighbors of both LDPs **x** and **y**. In this equation, max(d) represents the maximum distance between all pairs of LDPs. Accordingly, the LDPs and shared-neighbor-based distance values are applied to construct the MST. Based on the number of clusters, the LDPs are divided into clusters and get cluster indices. Finally, the assigned data points to their LDPs take the same cluster indices.

Once the MST has been constructed, one needs to distribute the main LDPs to clusters for receiving cluster labels or indices. Accordingly, it is important to pre-determine the cluster number before running the LDPC-MST. Generally, the cluster number is the underlying unknown parameter (called hyperparameter) in unsupervised clustering algorithms LDPC-MST (Sarmadi, 2021). Because this parameter significantly affects the clustering overall performance, one should attempt to select a proper cluster number for unlabeled data (Yang and Shami, 2020).

3. Locally Unsupervised Learning Method Based on Double Data Clustering

This method consists of two local levels of data clustering based on the LDPC-MST algorithm. In the first level, this algorithm is applied to partition the training data into main clusters, for which the Gap statistic is considered to select the number of clusters. After that, the second level begins by implementing another data clustering via the same LDPC-MST algorithm alongside a multivariate distance measure, which is used find a sub-cluster of each

main cluster. In this case, this sub-cluster contains the most relevant features insensitive to variability sources. These features are then considered to provide the main parameters of the damage detection indicator.

3.1 Clustering I

The first level of the method based on double data clustering, called here "Clustering I", begins by finding the optimal number of clusters based on the Gap statistic. Assume that $\mathbf{X}=[\mathbf{x}_1,\ldots,\mathbf{x}_n]$ is the training data (matrix) consisting of n data points (feature vectors) of t variables; that is, $\mathbf{x}_i=[x_1,\ldots,x_t]^T$, where $i=1,\ldots,n$. Using a sample cluster (K_{sam}), the Gap statistic is applied to determine K_{sam} values of this statistic; that is, $G(1),\ldots,G(K_{sam})$. Among these values, the sample cluster with the largest Gap statistic along with the maximum one-standard-error is selected as the optimal cluster number K. Hence, the unlabeled training data is divided into K clusters $\mathbf{C}_1,\ldots,\mathbf{C}_K$ by using the algorithm of the LDPC-MST as discussed in Section 2.1. Notices that each main cluster is a matrix containing some data points (i.e., $N_1,\ldots,N_K$) with t variables.

3.2 Clustering II

The second level of the method based on double clustering algorithm, called here "Clustering II", starts by dividing each of the main clusters $\mathbf{C}_1,\ldots,\mathbf{C}_K$ into sub-clusters. For this purpose, the Gap statistic is again utilized to determine the number of sub-clusters. Let $\mathbf{C}_r$ denotes the r^{th} main cluster from the first level, where $r=1,\ldots,K$. This cluster is in fact a matrix of the same variable as the training data but N_r samples rather than n; that is, $\mathbf{C}_r\in\mathbb{R}^{t\times N_r}$. By implementing the Gap statistic based on a new sample cluster (k_{sam}), a new optimal number of sub-clusters (k) is obtained according to the same strategy as explained. In this regard, the main cluster $\mathbf{C}_r$ is partitioned into k sub-clusters $\mathbf{S}_1,\ldots,\mathbf{S}_k$, each of which is a matrix of the sizes of $L_1,\ldots,L_k$ as the same variable of t. More precisely, supposing that $\mathbf{S}_h$ is the h^{th} sub-cluster, where $h=1,\ldots,k$, this matrix can be expressed as $\mathbf{S}_h\in\mathbb{R}^{t\times L_h}$.

Subsequently, the final step of Clustering II is to find a sub-cluster of each main cluster, which contains the most relevant features insensitive to environmental and operational variations. In other words, one should select one sub-cluster from $\mathbf{S}_1,\ldots,\mathbf{S}_k$ regarding the main cluster $\mathbf{C}_r$. For this goal, as the sub-clusters are matrices, this article presents a multivariate distance measure that can yield a scalar value. Accordingly, the sum of MSD between $\mathbf{C}_r$ and each of the sub-clusters $\mathbf{S}_1,\ldots,\mathbf{S}_k$ is computed in the following form:

$$d_M(\boldsymbol{S}_h,\boldsymbol{C}_r)=\sum_{j=1}^{L_h}\left(\boldsymbol{s}_h^{(j)}-\bar{\boldsymbol{c}}_r\right)^T\boldsymbol{H}_r^{-1}\left(\boldsymbol{s}_h^{(j)}-\bar{\boldsymbol{c}}_r\right) \quad (3)$$

where $\mathbf{s}_h^{(j)}$ is the j^{th} (column) vector of $\mathbf{S}_h$ so that $j=1,\ldots,L_h$; $\bar{\mathbf{c}}_r$ and $\mathbf{H}_r$ represent the mean vector and covariance matrix of $\mathbf{C}_r$, respectively. Based on Eq. (3), one can determine k distance values. Thus, the representative sub-cluster $\hat{\mathbf{S}}_r$ is one of the sub-clusters that provides the minimum distance value.

$$\hat{\boldsymbol{S}}_r\rightarrow min\left(d_M(\boldsymbol{S}_1,\boldsymbol{C}_r),\ldots,d_M(\boldsymbol{S}_k,\boldsymbol{C}_r)\right) \quad (4)$$

Notice that $\hat{\mathbf{S}}_r$ is the representative sub-cluster of the r^{th} main cluster. Hence, this process should be performed for all main clusters to determine K representative sub-clusters $\hat{\mathbf{S}}_r,\ldots,\hat{\mathbf{S}}_K$.

3.3 Damage detection indicator

Once the representative sub-clusters $\hat{\mathbf{S}}_1,\ldots,\hat{\mathbf{S}}_K$ of the main clusters $\mathbf{C}_1,\ldots,\mathbf{C}_K$ have been extracted, these are considered to provide the main parameters of the damage detection indicator, which is the well-known MSD. Accordingly, its main parameters are the mean

vectors and covariance matrices of the representative sub-clusters. Regarding the MSD, it should be pointed out that the research by Sarmadi demonstrated that if the concern about the environmental and operational variations is dealt with or their levels are low, this statistical distance measure can yield reliable and accurate results of damage detection (Sarmadi, 2021). As the main objective of this article is to minimize the influences such variability conditions by proposing the double clustering method, one can ensure that the MSD can succeed in damage detection.

Let $\mathbf{X}=[\mathbf{x}_1,\ldots,\mathbf{x}_n]$ and $\mathbf{Y}=[\mathbf{y}_1,\ldots,\mathbf{y}_m]$ denote the training and test datasets regarding the undamaged and unknown structural conditions, respectively. Using the training data, the optimal sub-clusters $\hat{\mathbf{S}}_1,\ldots,\hat{\mathbf{S}}_K$ are extracted to develop the MSD-based damage indicator by estimating their mean vectors $\bar{\mathbf{s}}_1,\ldots,\bar{\mathbf{s}}_K$ and covariance matrices $\mathbf{R}_1,\ldots,\mathbf{R}_K$. As such, the damage indicator of each training point is calculated as follows:

$$DI(\boldsymbol{x}_i) = min\{(\boldsymbol{x}_i - \bar{\boldsymbol{s}}_r)^T \boldsymbol{R}_r^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{s}}_r)\} \quad (5)$$

where $DI(\mathbf{x}_i)$ stands for the damage indicator of the i^{th} training point, which is the minimum distance value among all K distances of the sub-clusters. In a similar way, it is feasible to calculate the damage indicator of each test point in the following form.

$$\boldsymbol{DI}(\boldsymbol{y}_j) = \boldsymbol{min}\left\{(\boldsymbol{y}_j - \bar{\boldsymbol{s}}_r)^{\boldsymbol{T}} \boldsymbol{R}_r^{-1}(\boldsymbol{y}_j - \bar{\boldsymbol{s}}_r)\right\} \quad (6)$$

where $DI(\mathbf{y}_j)$ is the damage indicator of the j^{th} test sample so that $j=1,\ldots,m$. By considering all damage indicators of the training samples $DI(\mathbf{x}_1),\ldots,DI(\mathbf{x}_n)$, an upper-bound of threshold limit can be obtained by using the standard confidence interval. As such, each of the damage indicators associated with the training and test points is compared with this threshold. It is expected that $DI(\mathbf{x}_1),\ldots,DI(\mathbf{x}_n)$ fall under the threshold emphasizing the undamaged condition of the structure. If any damage indicator of the unknown state related to the test data is over the threshold, this means that the structure sustains an abnormal change attributable to damage and the structural current status is damaged.

4. Results and discussions

4.1 A steel arch bridge

This structure is a steel single-span tied arch railway bridge, called the KW51 bridge, in Leuven, Belgium (Maes et al., 2022). Fig. 1 shows a real image of this structure and its elevation view. The total length of this bridge is about 115 m and the width of 12.4 m. The deck contains two main girders, which are stiffened by thirty-three transverse beams. Additionally, the girders include three parts connected with steel plates. Due to a construction error associated with the bolted connections of the braces with the deck and the arch, it was decided to retrofit the KW51 bridge between 16 May 2019 until 15 September 2019. During this period, scaffolding was installed on the bridge, which led to increases in the structural mass and stiffness. All bolted connections between the braces and the arch as well as the braces and the deck were strengthened by some welded steel plates (Maes et al., 2022).

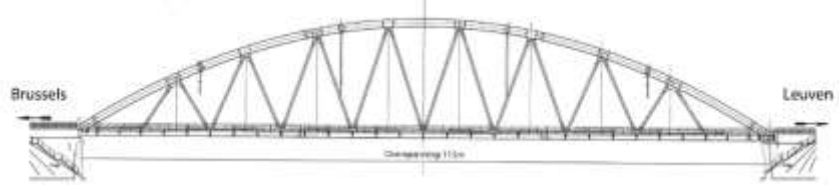

Fig. 1. The KW51 Bridge: the elevation view (Maes et al., 2022)

For SHM applications, the KW51 bridge was equipped with some sensors for measuring the environmental and dynamic data on 2 October 2018 and 15 May 2019 (before the retrofit),

16 May 2019 and 26 September 2019 (during the retrofit), and 27 September 2019 and 1 January 2020 (after the retrofit). Because the retrofitting process altered the inherent physical properties (i.e. the mass and stiffness) of the structure, it is possible to simulate the retrofitted structure as a simulated damaged condition. The underlying assumption is that such a structural change can alter dynamic characteristics similar to realistic damage cases (Maes et al., 2022). Therefore, the structural states before and after retrofit are considered the undamaged and simulated damaged conditions (Sarmadi et al., 2022).

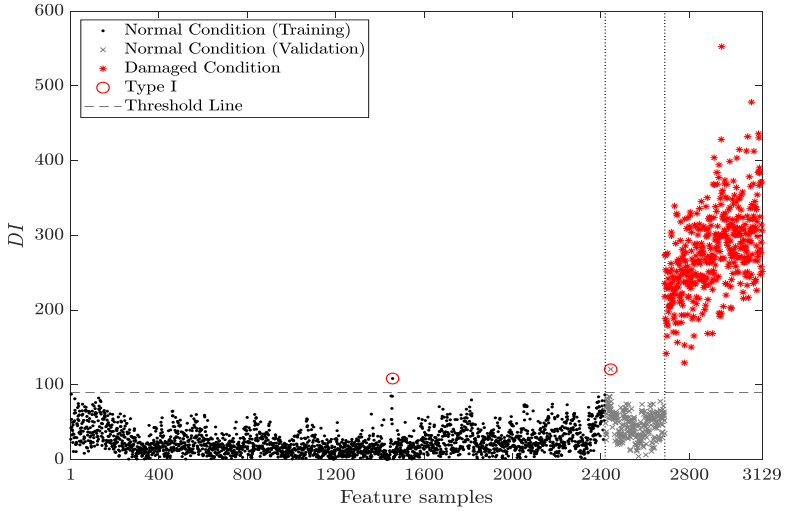

Fig. 2. Detection of damage to the KW51 bridge by the proposed unsupervised learning method under the double clustering process

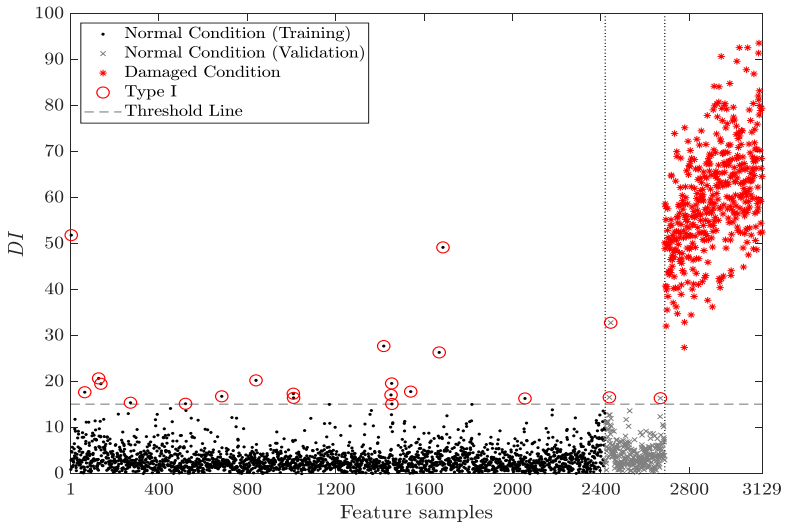

Fig. 3. Detection of damage to the KW51 bridge without the double clustering algorithm

The result of damage detection in the KW51 bridge based on the proposed method is presented in Fig. 2, where the horizontal line is the threshold obtained from the standard confidence interval under the 0.001 significance level. This figure clearly demonstrates that the environmental variability on the natural frequencies of the undamaged state has been mitigated due to the positive influence of the proposed unsupervised learning technique and its double data clustering. More precisely, with exception of the only two points, almost all the damage indicators of the training and validation data points fall below the threshold line. Moreover, all damage indicators of the simulated damaged state exceed the threshold of

interest. These conclusions verify the effectiveness of the proposed method for damage detection under the strong environmental variations.

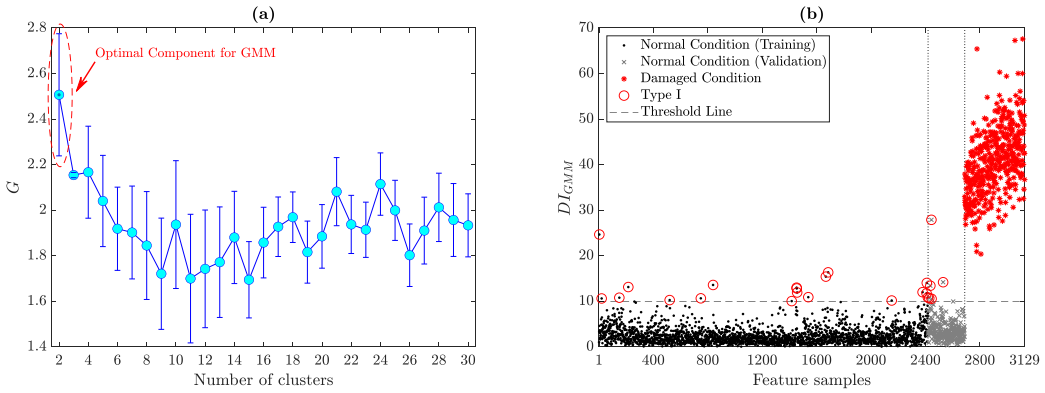

Fig. 4. Detection of damage to the KW51 bridge by the GMM: (a) selection of the optimal component number, (b) damage detection indicators based on the MSD

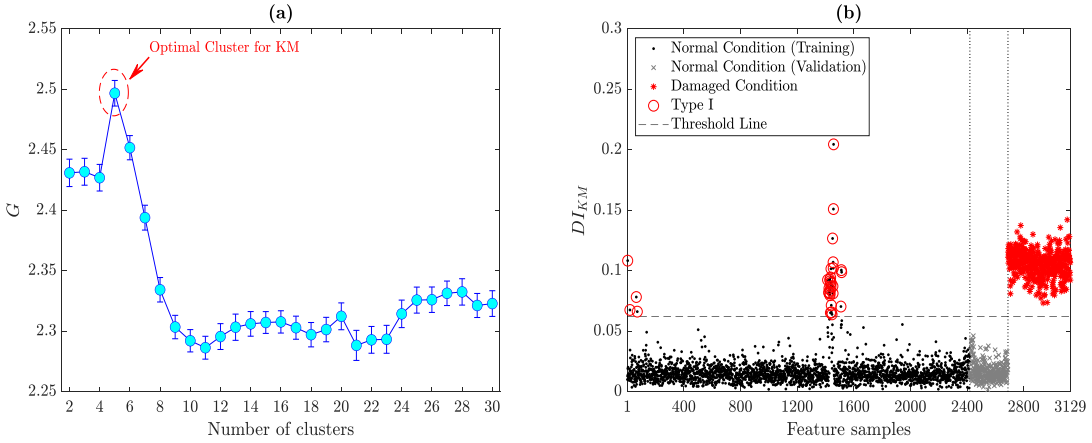

Fig. 5. Detection of damage to the KW51 bridge by the KM: (a) selection of the optimal cluster number, (b) damage detection indicators based on the ESD

To further demonstrate the performance and superiority of the proposed method, Fig. 3 shows the result of damage detection without the double clustering algorithm. Using the local information of the first level of clustering by the LDPC-MST, the two mean vectors and two covariance matrices are obtained from the main clusters $\mathbf{C}_1$ and $\mathbf{C}_2$ to calculate new damage indicators based on the MSD measure. From Fig. 3, there are relatively many false positives (Type I errors) in the damage indicators of the undamaged condition. In particular, the environmental effects on the natural frequencies of the bridge can still be observed, some of which have larger damage indicators than the damaged condition.

Furthermore, the proposed method is compared with the GMM in Fig. 4 and KM in Fig. 5. According to the Gap statistic, Fig. 4 (a) and Fig. 5 (a) reveal that the GMM and KM require 2 and 5 clusters for partitioning the training samples. By considering the MSD for GMM and ESD for KM, Fig. 4 (b) and Fig. 5 (b) show the results of detection of damage to the KW51 bridge. As can be seen, the GMM provides more false positive errors than the proposed method. However, the worst performance pertains to the KM, where the environmental effects on the natural frequencies are exactly observed in its output. Therefore, one can conclude that the proposed method with the double data clustering strategy is highly superior to the classical GMM and KM techniques.

4.2 A concrete box-girder bridge

This structure was a post-tensioned concrete two-cell box girder bridge, called the Z24 bridge. The total length of this structure was equal to 58 m containing the main span of 30m and two sides of 14m. Fig. 6 shows the real image and the elevation view of this bridge. Regarding the main structure of the Z24 bridge, the abutments contained triple concrete columns connected with concrete hinges to the girder. Additionally, the intermediate supports were concrete piers clamped into the girder. In 1998, it was decided to demolish the Z24 bridge in order to build a new larger bridge structure. Before this procedure, a continuous monitoring program was implemented by SHM systems to measure environmental and dynamic responses such as ambient temperature and acceleration time histories.

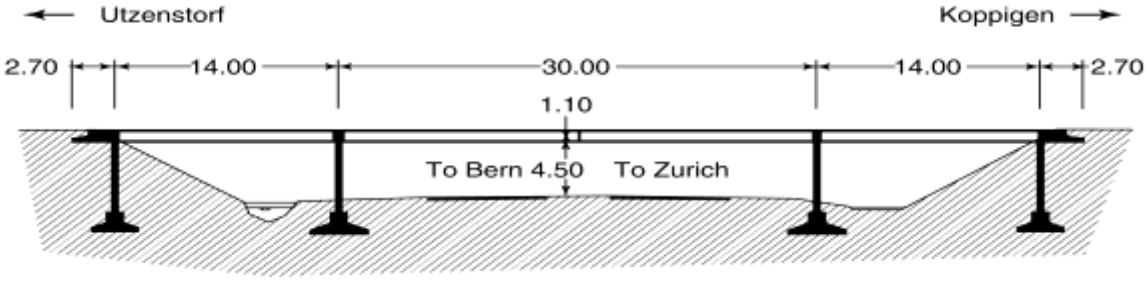

Fig. 6. The Z24 Bridge: the elevation view along with the main dimensions (in the unit of meters) (Reynders and De Roeck, 2009)

This program included two parts: (1) a one-year continuous monitoring test over the year before demolition for quantifying the environmental variations, in which case the bridge was supposed to be undamaged, (2) progressive damage tests over a month shortly before the complete demolition for simulating realistic damage patterns in a concrete civil structure. The damage patterns consisted of lowering of one pier, the tilt of the foundation, spalling of concrete, landslide at one abutment, failure of concrete hinge, failure of anchors, and rupture of the tendons. Due to the equipment of the Z24 bridge with various sensors (e.g., temperature sensors and accelerometers), some important environmental and dynamic parameters were measured during the SHM program.

Based on the estimated mean vectors and covariance matrices, the MSD measure is applied to compute the distance quantities of the 3128 training points $DI(\mathbf{x}_1)$,…,$DI(\mathbf{x}_{3128})$ and 804 test samples $DI(\mathbf{y}_1)$,…,$DI(\mathbf{y}_{804})$. The distance quantities of the training points are used to estimate a threshold based on the standard confidence interval under 0.001 significance level. The result of damage detection of the Z24 bridge gained by the proposed method is displayed in Fig. 7 As can be seen, the majority of the damage indicators of the training data are under the threshold line except for some points indicating an insubstantial false positive (Type I) error. The same conclusion is observable concerning the validation data. On the other hand, most of the damage indicators of the damaged state exceed the threshold line implying accurately detecting the occurrence of damage in the bridge. In this regard, a few points are detected at the below of the threshold line indicating an inconsiderable false negative (Type II) error.

One of the important notes in Fig. 7 is that the environmental effects, which are observable in the natural frequencies of the Z24 bridge, are no longer available in the damage indicators of the undamaged condition, particularly between the samples 400-1600, after using the proposed method. The other important note is that the damage indicators of the damaged state are suitably far away from the corresponding indicators concerning the undamaged condition

(i.e., the training and validation data points). This note conveys the good performance of the proposed method, particularly under the idea of the double data clustering.

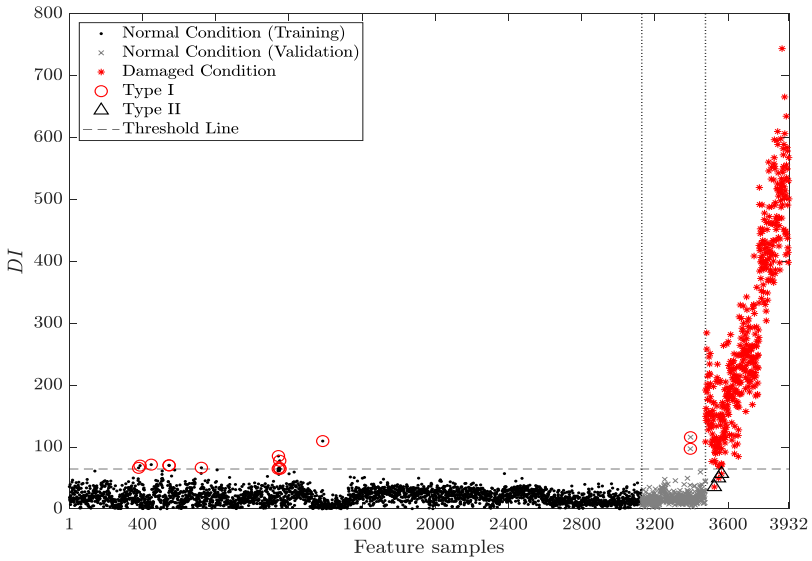

Fig. 7. Detection of damage to the Z24 bridge by the proposed unsupervised learning method under the double clustering process

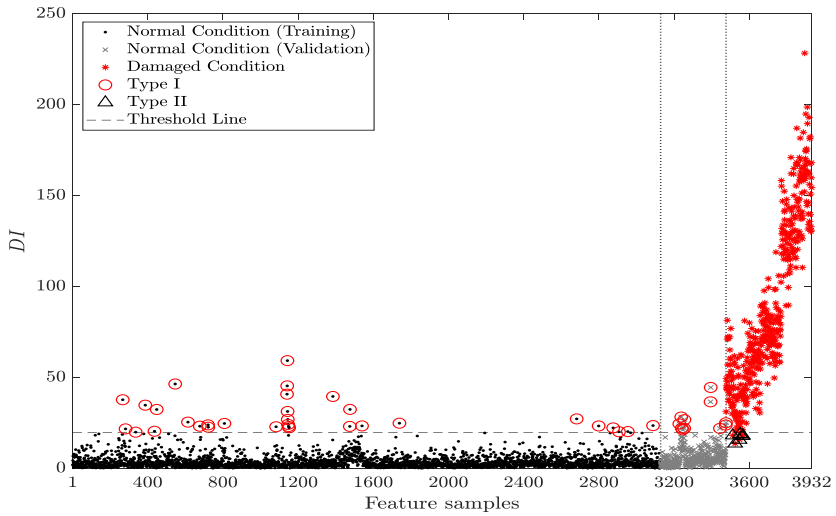

Fig. 8. Detection of damage to the Z24 bridge without the double clustering algorithm

To more indicate the correctness and reliability of the proposed method, it is compared with some unsupervised learning techniques. First, one attempts to investigate the effect of the proposed double data clustering algorithm. For this reason, the process of determining the damage detection indicators is carried out by the local information of the first level of the data clustering (i.e., the main clusters $\mathbf{C}_1,\ldots,\mathbf{C}_8$) and the MSD measure. Second, as the core of the proposed unsupervised learning method is based on a clustering algorithm, it is compared with some clustering-based SHM methods; that is, the *k*-means (KM) and Gaussian mixture model (GMM). The same cluster selection approach based on the Gap statistic is considered to determine the optimal cluster number regarding the KM and the optimal component number associated with the GMM. Once the original training data has been divided into clusters and components, their mean vectors and covariance matrices are obtained to compute new damage detection indicators for GMM-based SHM method through the MSD (Sarmadi,

2021). In relation to the KM method, after obtaining the center sets, the well-known Euclidean-squared distance (ESD) (Sarmadi et al, 2021) [11] is considered to determine new damage indicators.

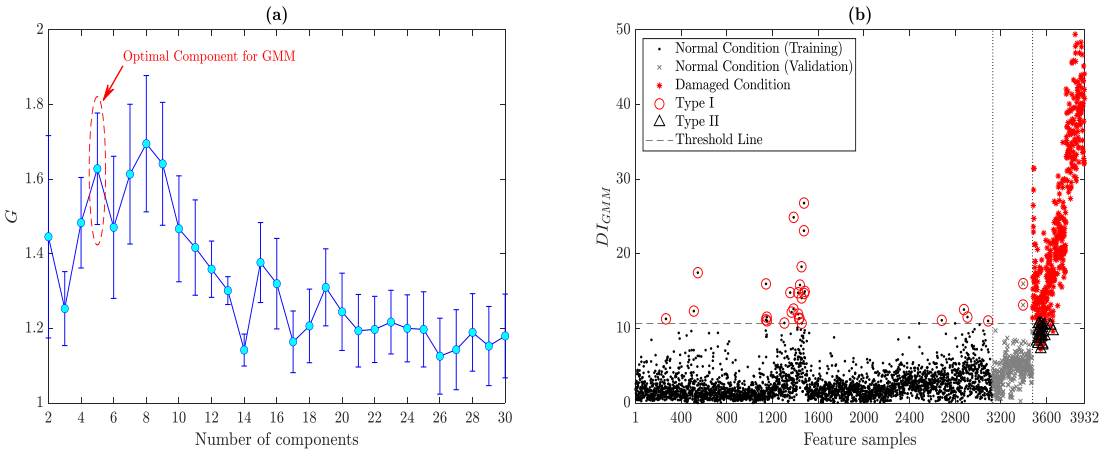

Fig. 9. Detection of damage to the Z24 bridge by the GMM: (a) selection of the optimal component number, (b) damage detection indicators based on the MSD

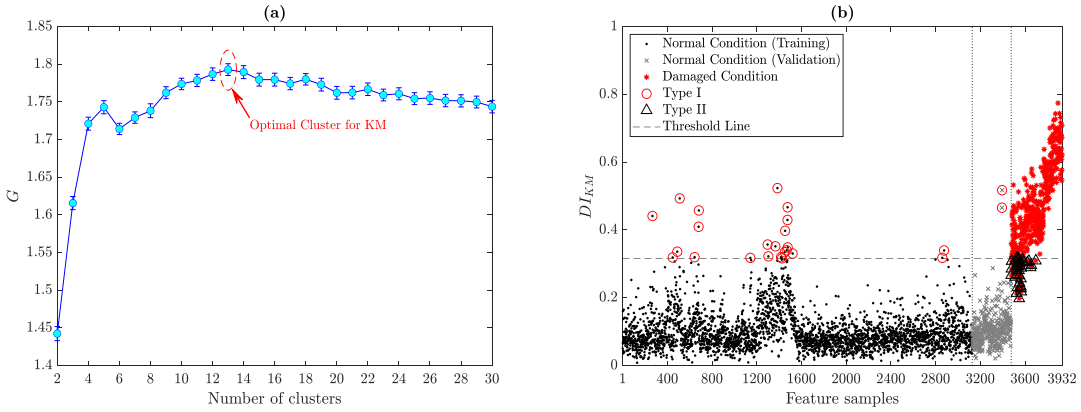

Fig. 10. Detection of damage to the Z24 bridge by the KM: (a) selection of the optimal cluster number, (b) damage detection indicators based on the ESD

For the first comparison, Fig. 8 indicates the result of damage detection without the double clustering algorithm. Although the use of local information of the first level of data clustering via the LDPC-MST can provide discriminative damage indicators, which are suitable for damage detection, some weaknesses can be observed. First, there are more false positive errors in the training and validation data points compared to the same result presented in Fig.10. The other important conclusion is that the environmental effects are somewhat observable in the damage indicators of the undamaged condition. On the other hand, further false negative errors occurred in Fig. 8. All these conclusions verify the positive effect of the idea of double data clustering by the LDPC-MST.

For the second comparison, Fig. 9 and Fig. 10 show the results of damage detection by the GMM and KM methods, respectively. From Fig. 9 (a) and Fig. 10 (a), where show the outputs of the Gap statistic approach, one can see that the optimal component for GMM and optimal cluster for KM are identical to 5 and 13, respectively. In Fig. 9 (b) and Fig. 10 (b), where illustrate the damage detection indicators of the GMM and KM methods, respectively, it is observed that many false positive errors happened in the features of the undamaged condition. In particular, the environmental effects in the natural frequencies are observable in the damage detection indicators. In relation to the damaged condition, it is seen that that both the GMM and KM techniques produce considerable false negative (Type II) errors. As a

comparison with the proposed method presented in Fig. 7, it can be proved that the proposed method outperformed the classical GMM and KM techniques in accurate damage detection and providing fewer errors.

5. Conclusion

For continuous SHM under environmental and operational variations, a damage detection indicator under the well-known MSD measure was applied to detect damage in structures. The continuous natural frequencies of the Z24 and KW51 bridges were considered to prove the proposed method with some comparative analyses.

Accordingly, all environmental effects in the Z24 and KW51 bridges were significantly removed leading to accurate damage detection with inconsiderable errors. One of the comparisons revealed that the performance of the proposed technique can be degraded by ignoring the implementation of the second level of the double data clustering. Indeed, one of the main reasons for the removal of the environmental effects on the natural frequencies of the Z24 and KW51 bridges is related to this level. Finally, the other comparison indicated that the proposed method outperformed the GMM and KM techniques in terms of mitigating the environmental and operational variations and providing smaller error rates.

Refrences

Cheng D, Zhu Q, Huang J, Wu Q, Yang L (2021) Clustering with Local Density Peaks-Based Minimum Spanning Tree. IEEE Transactions on Knowledge and Data Engineering 33 (2):374-387. doi:https://doi.org/10.1109/TKDE.2019.2930056

Jin C, Jang S, Sun X, Li J, Christenson R (2016) Damage detection of a highway bridge under severe temperature changes using extended Kalman filter trained neural network. J Civ Struct Health Monit 6 (3):545-560. doi:https://doi.org/10.1007/s13349-016-0173-8

Maes K, Van Meerbeeck L, Reynders EPB, Lombaert G (2022) Validation of vibration-based structural health monitoring on retrofitted railway bridge KW51. Mech Syst Sig Process 165:108380. doi:https://doi.org/10.1016/j.ymssp.2021.108380

Mahmoudkelayeh M.R, Adhami B, Saeedi Razavi B.S (2024) Continuous health assessment of bridges under sudden environmental variability by local unsupervised learning. J Perform Constr Facil 38 (5). doi: https://doi.org/10.1061/JPCFEV.CFENG-4323

Mario Azzara R, De Roeck G, Girardi M, Padovani C, Pellegrini D, Reynders E (2018) The influence of environmental parameters on the dynamic behaviour of the San Frediano bell tower in Lucca. Eng Struct 156:175-187. doi:https://doi.org/10.1016/j.engstruct.2017.10.045

Qiu L, Fang F, Yuan S, Boller C, Ren Y (2019) An enhanced dynamic Gaussian mixture model–based damage monitoring method of aircraft structures under environmental and operational conditions. Struct Health Monit 18 (2):524-545. doi:https://doi.org/10.1177/1475921718759344

Reynders E, De Roeck G (2009) Continuous Vibration Monitoring and Progressive Damage Testing on the Z24 Bridge. In: Encyclopedia of Structural Health Monitoring. John Wiley & Sons, Inc., Chichester, United Kingdom. doi:https://doi.org/10.1002/9780470061626.shm165

Saeedi Razavi B.S, Mahmoudkelayeh M.R, Saeedi Razavi S (2021) Damage identification under ambient vibration and unpredictable signal nature. J Civ Struct Health Monit. (11):1253–1273 doi:https://doi.org/10.1007/s13349-021-00503-x

Sarmadi H (2021) Investigation of machine learning methods for structural safety assessment under variability in data: Comparative studies and new approaches. J Perform Constr Facil 35 (6):04021090. doi:https://doi.org/10.1061/(ASCE)CF.1943-5509.0001664

Sarmadi H, Entezami A, Salar M, De Michele C (2021) Bridge health monitoring in

environmental variability by new clustering and threshold estimation methods. J Civ Struct Health Monit 11 (3):629-644. doi:https://doi.org/10.1007/s13349-021-00472-1

Sarmadi H, Entezami A, Behkamal B, De Michele C (2022) Partially online damage detection using long-term modal data under severe environmental effects by unsupervised feature selection and local metric learning. J Civ Struct Health Monit. doi:https://doi.org/10.1007/s13349-022-00596-y

Sousa Tomé E, Pimentel M, Figueiras J (2020) Damage detection under environmental and operational effects using cointegration analysis – Application to experimental data from a cable-stayed bridge. Mech Syst Sig Process 135:106386. doi:https://doi.org/10.1016/j.ymssp.2019.106386

Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. J Roy Stat Soc Ser B (Stat Method) 63 (2):411-423. doi:https://doi.org/10.1111/1467-9868.00293

Yang L, Shami A (2020) On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 415:295-316. doi:https://doi.org/10.1016/j.neucom.2020.07.061