Journal of Zhejiang University-SCIENCE A 2025 26(11):1052-1069
www.jzus.zju.edu.cn; www.springer.com/journal/11582
E-mail: jzus_a@zju.edu.cn

Research Article
https://doi.org/10.1631/jzus.A2400538

A two-stage framework for automated operational modal identification using OPTICS-KNN-based clustering

Yi CHEN[1,2], Wenwei FU[3,4]✉, Yaozhi LUO[1], Yanbin SHEN[1,4], Hui YANG[2], Shiying WANG[5]

[1]*College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China*
[2]*Zhejiang Communications Construction Group Co., LTD., Hangzhou 310051, China*
[3]*School of Civil Engineering, Suzhou University of Science and Technology, Suzhou 215011, China*
[4]*Future City Laboratory, Innovation Center of Yangtze River Delta, Zhejiang University, Jiaxing 314102, China*
[5]*Hangzhou Institute of Communications Planning Design & Research, Hangzhou 310030, China*

Abstract: Modal analysis, which provides modal parameters including frequencies, damping ratios, and mode shapes, is essential for assessing structural safety in structural health monitoring. Automated operational modal analysis (AOMA) offers a promising alternative to traditional methods that depend heavily on human intervention and engineering judgment. However, estimating structural dynamic properties and managing spurious modes remain challenging due to uncertainties in practical application conditions. To address this issue, we propose an automated modal identification approach comprising three key aspects: (1) identification of modal parameters using covariance-driven stochastic subspace identification; (2) automated interpretation of the stabilization diagram; (3) an improved self-adaptive algorithm for grouping physical modes based on ordering points to identify the clustering structure (OPTICS) combined with *k*-nearest neighbors (KNN). The proposed approach can play a crucial role in enabling real-time structural health monitoring without human intervention. A simulated 10-story shear frame was used to verify the methodology. Identification results from a cable-stayed bridge demonstrate the practicality of the proposed method for conducting AOMA in engineering practice. The proposed approach can automatically identify modal parameters with high accuracy, making it suitable for a real-time structural health monitoring framework.

Key words: Structural health monitoring; Covariance-driven stochastic subspace identification; Automated operational modal analysis (AOMA); Ordering points to identify the clustering structure (OPTICS); *k*-nearest neighbors (KNN)

1 Introduction

Civil infrastructure is continuously subjected to material degradation and time-varying environmental degradation, resulting in gradual or sudden reductions in structural safety and reliability (Farrar and Worden, 2007). It is essential to identify the key structural parameters of civil infrastructure in real-time using structural health monitoring (SHM) technology. SHM technology offers an effective way to update finite element models (Pan et al., 2020; Ereiz et al., 2022), track structural status (Yun et al., 2014; Ni et al., 2020; Wu et al., 2023), and detect damage (Garcia-Perez et al., 2013; Gu et al., 2017; Sun et al., 2020; Ran et al., 2023). In recent decades, many customized SHM systems have been implemented in various forms of civil infrastructure systems such as bridges (Jeong et al., 2019; Jin et al., 2021), high-rise buildings (Liu et al., 2013; Huang et al., 2020; Doroudi et al., 2022), dams (Kang and Li, 2020; Mostafaei et al., 2021), and other large-scale space structures (Luo et al., 2021, 2022). These systems provide raw data for identifying the modal parameters, including frequencies, damping ratios, and mode shapes, which are the key indicators of changes in the structural condition of infrastructure during long-term operation.

Most civil infrastructure is large-scale and can be excited only in operational excitations, which limits the application of experimental modal analysis (EMA). In contrast, operational modal analysis (OMA) has been proposed to extract modal parameters from continuous acceleration data of structures with minimal

✉ Wenwei FU, fww@usts.edu.cn
Wenwei FU, https://orcid.org/0009-0000-7117-1964

Received Nov. 18, 2024; Revision accepted Feb. 6, 2025;
Crosschecked Sept. 15, 2025; Online first Nov. 7, 2025

influence on their operation. Generally, OMA approaches are divided into three categories: frequency domain, time domain, and time-frequency analysis (Hou and Xia, 2021). These approaches have been shown to be effective in several high-profile projects. The Z24 Bridge (Civera et al., 2023), Beichuan River Bridge (Ren and Zong, 2004), Confederation Bridge (Desjardins and Lau, 2022), and Tsing Ma Bridge (Chen et al., 2004) are among notable examples of OMA applications in bridge engineering. The Z24 bridge, a three-span concrete continuous beam bridge, underwent continuous monitoring and modal testing, serving as a benchmark model that comprehensively validated numerous operational modal identification approaches. In a comparative study of current methods for operational modal identification for the Z24 bridge, Peeters et al. (2004) investigated the varied identification results obtained by different researchers. Reynders et al. (2008) introduced a time-domain hybrid subspace identification approach by taking the Z24 bridge as a case study. This study stands as the most comprehensive literature on modal parameter identification for the Z24 bridge, demonstrating strong identification accuracy and affirming the effectiveness of the time-domain method. The Beichuan River Bridge, located in China, has attracted significant attention in recent times. To identify its modal properties, extensive dynamic tests were carried out using the peak picking method and the stochastic subspace identification (SSI) method (Ren and Zong, 2004). The findings highlighted the efficacy of the stabilization diagram method in circumventing the inherent issues of randomness and ambiguity that can arise during modal identification. Consequently, the time-domain SSI method outperformed its frequency-domain counterpart (Peeters and de Roeck, 1999). A massive amount of acceleration data is continuously collected from structures under operational and environmental excitations, making it necessary to process data in real-time for modal parameter identification. In such cases, conventional modal analysis methods that rely on human manipulation become time-consuming due to the required engineering judgment. It is necessary to develop a robust and efficient automated operational modal analysis (AOMA) approach, which is useful for the application of SHM in engineering practice (Yaghoubi et al., 2018; Mugnaini et al., 2022).

Many researchers are actively engaged in automating the interpretation of stabilization diagrams and eliminating spurious modes to address these challenges within the field of AOMA. SSI-based AOMA is widely recognized for its robust mathematical foundation and effectiveness in identifying closely spaced and weakly excited modes (Cabboi et al., 2017; Wu et al., 2019). Consequently, SSI has gained significant popularity and has become one of the most widely used methodologies in AOMA. However, SSI-based AOMA faces a major problem in determining the optimal system order. Over-specifying the model order and setting it high enough to include all physical modes within the frequency of interest is a typical approach used to resolve this problem. Nonetheless, this approach unavoidably introduces numerous spurious modes, which can either be mathematically derived or induced by noise. To mitigate this issue, the use of a stabilization diagram that provides a visual representation of the model order plotted against frequency has become standard practice. It serves as a valuable tool for distinguishing physical modes from spurious modes. Physical modes of structure have a regular distribution in the stabilization diagram that allows them to be seen through vertical lines as the model order progressively increases. In contrast, spurious modes appear scattered. The main part of the implementation of SSI-based AOMA involves the automated interpretation and clearing of the stabilization diagram in situations where spurious modes are present. Significant efforts have been dedicated to automating the interpretation of the stabilization diagram to effectively distinguish physical modes from spurious modes. Spurious mode removal and physical mode clustering constitute the two primary components of the SSI-based AOMA methodology (Feng et al., 2024).

Many modal validation criteria have been used to pre-filter the stabilization diagram and eliminate the spurious modes. One of these criteria is the modal complexity index (MCI), which combines the modal phase deviation (MPD) and modal phase collinearity (MPC) (Rainieri and Fabbrocino, 2014; Greś et al., 2021). The modal transfer norm (MTN), based on the modal contribution ratio, is an additional standard for validating modal parameters (Reynders and de Roeck, 2008). While many real-world scenarios can be constrained by the specifications, these standards may be less useful when significant measurement noise and non-proportional damping are present. It is suggested that the modal uncertainty, which is currently measured by metrics like the coefficient of variation C_v, is

an effective instrument for the removal of spurious modes (Reynders et al., 2016). A lower C_v is typically observed in physical modes (Döhler and Mevel, 2013). The second step of the AOMA procedure is defining a collection of physically related modes using different techniques (He et al., 2022; Sadeqi et al., 2022). Diverse clustering techniques play a pivotal role in the mode clustering step, effectively grouping modes with similar characteristics (Saxena et al., 2017). Clustering algorithms are divided into two categories: hard and soft. Hard clustering algorithms classify each data sample precisely once, but soft clustering algorithms assign a probability to each data sample for belonging to each cluster rather than hard connecting it to only one cluster. Many hard clustering techniques have been proposed to automate the interpretation of stabilization diagrams, such as *k*-means clustering and its derivatives, hierarchical clustering, and density-based clustering (Civera et al., 2023). Many researchers have used hierarchical clustering extensively and consider it to be the most common strategy (He et al., 2021). The ability to pick physical clusters is a significant advantage of hierarchical clustering. The computational cost of hierarchical clustering is high due to the need for user-defined tree cutoff distances and manual intervention. Another popular clustering method involving partitioning is commonly known as *k*-means clustering (Neu et al., 2017). This method has the advantage of being computationally fast. However, the number of clusters must be predetermined and is sensitive to cluster seeds. These approaches do not take into account the uncertainty of modal parameters and the inaccuracy of mode shapes. The cluster can be described as a multivariate normal distribution function for data samples in a Gaussian mixture model (GMM), a type of soft clustering also known as probability model-based clustering (Zeng and Hu, 2022). The adoption of GMM, an unsupervised learning technique that automatically assigns each sample to a cluster with a probability, speeds up the construction of clusters based on these probabilities (McLachlan et al., 2019). It has been reported that the assumption of modal parameters following a normal distribution may not accurately reflect real-world conditions (Fan et al., 2019). Sun et al. (2017) proposed an AOMA methodology using multi-level clustering techniques, such as *k*-means and hierarchical clustering, which was applied to a single-tower cable-stayed bridge with a main span of 46 m. With the rapid advancements in artificial intelligence and big data, machine learning has the potential to refine supervised clustering approaches, thereby enhancing the accuracy of automatic modal parameter identification and reducing the influence of spurious modes (Mao et al., 2019; Liu et al., 2023). Civera et al. (2022) proposed a machine learning-based AOMA method applied to masonry arch bridges, combining the density-based spatial clustering of applications with noise and the SSI.

Researchers have focused on distance-based grouping algorithms, hierarchical clustering, and frequency-based overlapping histograms to identify collections of physically related modes. Distance-based grouping algorithms classify modes by considering the separation between elements and applying distance criteria progressively. To the best of our knowledge, a density clustering approach known as ordering points to identify the clustering structure (OPTICS) has rarely been proposed to efficiently assess and automate the discovery of stable columns. OPTICS uses the reachability plot and reachability distance concepts to determine the core distance that clusters a minimum set of objects for each component. All objects that can be allocated to physical modes and outliers are identified based on the reachability distance, and the findings are displayed on a reachability plot. In this paper, we present a novel AOMA framework based on OPTICS density-based clustering to address the drawbacks of earlier approaches. The method comprises two main stages: (1) two widely used modal validation criteria are used to preprocess the stabilization diagram, remove spurious modes, and reduce the computational burden for the clustering stage; (2) the optimal cluster number is determined using the *k*-nearest neighbors (KNN) method, and mode clustering is performed using the OPTICS process. The proposed method does not involve any manually updated indexes. Instead, it uses two widely known and verified criteria. The efficiency of the proposed AOMA methodology was investigated using simulated data obtained from a 10-story shear frame and monitoring data measured from the Taoyaomen Bridge, China. The identification results demonstrate that our proposed approach can successfully eliminate spurious modes and separate closely spaced modes.

The structure of this paper is as follows. The computational procedure for the proposed two-stage AOMA is described in Section 2. Section 3 demonstrates the

effectiveness of the proposed methodology through a numerical simulation of a 10-story shear frame, and applies the methodology to monitoring data from the Taoyaomen Bridge. Section 4 provides concluding remarks.

2 Two-stage automated operational modal identification

2.1 Spurious mode pre-filtering stage

The whole stabilization diagram derived from SSI seems complicated and frequently contains spurious modes. The basic theory of stochastic subspace identification is provided in Section S1 of the electronic supplementary materials (ESM). Various validation criteria are used in the pre-filtering stage to identify and filter out these spurious modes. This approach speeds up and simplifies the automated interpretation of the stabilization diagram.

2.1.1 Hard and soft validation criteria

Hard and soft validation criteria (HVC and SVC), also named conventional validation criteria, are defined as follows:

$$0\% < \xi_p \leqslant 20\%, \tag{1}$$

$$\mathrm{Re}(\lambda_p) \leqslant 0, \tag{2}$$

$$\mathrm{Im}(\lambda_p) \neq 0, \tag{3}$$

where ξ_p and λ_p denote the damping ratio and eigenvalue for the pth pole of an mth model order, respectively. Eq. (1) assumes that the damping ratio of the practical structure cannot have negative values and exceed a threshold of 20%. Eqs. (2) and (3) limit the scope of research to decaying oscillations. Furthermore, the complexity of the mode shape $\boldsymbol{\Phi}_p$ can be quantified by MPC. The definition $M_{\mathrm{pc}}(\boldsymbol{\Phi}_p)$ is as follows:

$$M_{\mathrm{pc}}(\boldsymbol{\Phi}_p) = \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^2, \tag{4}$$

$$\boldsymbol{S}_{\mathrm{cov}} = \begin{bmatrix} \mathrm{Re}(\boldsymbol{\Phi}_p)^{\mathrm{T}}\mathrm{Re}(\boldsymbol{\Phi}_p) & \mathrm{Re}(\boldsymbol{\Phi}_p)^{\mathrm{T}}\mathrm{Im}(\boldsymbol{\Phi}_p) \\ \mathrm{Re}(\boldsymbol{\Phi}_p)^{\mathrm{T}}\mathrm{Im}(\boldsymbol{\Phi}_p) & \mathrm{Im}(\boldsymbol{\Phi}_p)^{\mathrm{T}}\mathrm{Im}(\boldsymbol{\Phi}_p) \end{bmatrix}, \tag{5}$$

where λ_1 and λ_2 are the eigenvalues of matrix $\boldsymbol{S}_{\mathrm{cov}}$; $\boldsymbol{S}_{\mathrm{cov}}$ is the covariance matrix. Making a stabilization diagram is the first step in separating real poles from fake mathematical ones. The poles found for various model orders are shown in stabilization diagram as a function of frequency. Physical modes can be found by observing the alignments of stable poles with the increasing model orders, while erroneous mathematical poles often stay unstable and dispersed. The alignments of stable poles can start at a lower or higher model order based on the excitation level of the mode. To produce the stabilization diagram, the poles derived from a one-order lower model are compared to those associated with a particular model order. Only those poles that satisfy predetermined traditional validation criteria are considered stable. Inequalities are frequently used to express these stability criteria. Physical modes represent the underlying behavior of the structure and are expected to exhibit similar characteristics across different model orders. Conversely, the modal parameters of spurious modes may vary significantly between two successive model orders and may thus be successfully removed by comparing the modal parameter variations to tolerable levels. Consequently, spurious modes are dispersed arbitrarily on the frequency-model order plane, while physical modes can align vertically to form several stability axes on this plane. The relative deviations of frequency, damping ratio, and mode shape can be calculated as follows:

$$\frac{f^{(m)} - f^{(m-1)}}{\max(|f^{(m)}|, |f^{(m-1)}|)} \leqslant \delta_f, \tag{6}$$

$$\frac{\xi^{(m)} - \xi^{(m-1)}}{\max(|\xi^{(m)}|, |\xi^{(m-1)}|)} \leqslant \delta_\xi, \tag{7}$$

$$1 - M_{\mathrm{ac}}(\boldsymbol{\psi}^{(m)}, \boldsymbol{\psi}^{(m-1)}) \leqslant \delta_\psi, \tag{8}$$

where δ_f, δ_ξ, and δ_ψ denote the tolerances of frequency f, damping ratio ξ, and mode shape $\boldsymbol{\psi}$, respectively; $f^{(i)}$, $\xi^{(i)}$, and $\boldsymbol{\psi}^{(i)}$ (i=1, 2, ⋯, m) denote the ith frequency, damping ratio, and mode shape, respectively. Modal assurance criteria (MAC) between the mode shapes $M_{\mathrm{ac}}(\boldsymbol{\psi}^{(m)}, \boldsymbol{\psi}^{(m-1)})$ can be expressed as:

$$M_{\mathrm{ac}}(\boldsymbol{\psi}^{(m)}, \boldsymbol{\psi}^{(m-1)}) = \frac{\left|(\boldsymbol{\psi}^{(m)})^{\mathrm{H}}[\boldsymbol{\psi}^{(m-1)}]\right|}{(\boldsymbol{\psi}^{(m)})^{\mathrm{H}}[\boldsymbol{\psi}^{(m)}](\boldsymbol{\psi}^{(m-1)})^{\mathrm{H}}[\boldsymbol{\psi}^{(m-1)}]}, \tag{9}$$

where H denotes the conjugate transpose operator.

2.1.2 Uncertainty validation criterion

Spurious modes cannot be thoroughly eliminated using typical modal validation criteria, which may impede physical mode clustering. Uncertainty validation criteria (UVC) should be used to eliminate as many false modes as possible based on their greater efficiency. Uncertainty on modal parameters via SSI is caused mainly by six factors: an insufficient amount of monitoring data, unmeasured environmental excitation, operational noise represented as white noise, assumption of linear and stationary behavior, imperfect data filtering, and inappropriate model order selection. UVC can enhance the accuracy of the identified modal parameters. In reality, the uncertainty of physical modes is less than that of spurious modes. The standard deviations of modal parameters can be used to separate physical modes and spurious modes based on this phenomenon. Uncertainty computation has a significant benefit for removing spurious modes, which have more uncertainties than physical modes (Reynders et al., 2016). Modal uncertainty may be used to clean up the stabilization diagram by using uncertainty-based criteria. Modes with smaller C_{V} appear stably when the model order increases sequentially. These modes are visible through vertical alignments in the stabilization diagram, whereas the spurious modes with larger C_{V} appear scattered. Modes having a C_{V} greater than 1.5% in terms of frequency are considered spurious and should be removed. The threshold for eliminating the spurious modes in this study was set to 2.0% (Döhler and Mevel, 2013):

$$C_{\mathrm{V}} > 2.0\%. \tag{10}$$

2.2 Physical mode clustering stage

2.2.1 OPTICS density-based clustering

The second step of the AOMA procedure is defining a collection of physically related modes using OPTICS, which is more efficient and automated for detecting stable columns, especially in bigger datasets. To elucidate the construction of reachability plots, several definitions need to be introduced. S denotes a set of potential physical modes, and X_p represents an object that encapsulates the properties of the pth mode. In our case, each object X_p is characterized by its frequency f_p and damping ratio ξ_p:

$$X_p = \{f_p, \xi_p\},\ p \in S. \tag{11}$$

The distance between two objects X_p and X_q (an object corresponding to the qth mode) $D(X_p, X_q)$ is denoted as a function of f and $\boldsymbol{\psi}$, defined as follows:

$$D(X_p, X_q) = \frac{|f_p - f_q|}{\max(f_p, f_q)} + 1 - M_{\mathrm{ac}}(\boldsymbol{\psi}^{(p)}, \boldsymbol{\psi}^{(q)}), \tag{12}$$

where f_q is the frequency of X_q; $\boldsymbol{\psi}^{(p)}$ and $\boldsymbol{\psi}^{(q)}$ are the pth and qth mode shapes, respectively; $M_{\mathrm{ac}}(\boldsymbol{\psi}^{(p)}, \boldsymbol{\psi}^{(q)})$ is the MAC between the pth and qth mode shapes.

The damping ratio distance is disregarded due to potential identification inaccuracies and the likelihood of multiple modes with similar damping ratios, which can lead to misleading results. Similarly, the eigenvalue distance, which combines frequency and damping ratio, is not considered. Instead, a distance is used to account for the variable importance of differences in frequency and mode shape. Let $|N_\varepsilon(X)|$ be the number of objects at a distance ε from any project X:

$$|N_\varepsilon(X)| = C_{\mathrm{s}}\left\{X_p \middle| D(X_p, X_q) \leqslant \varepsilon\right\}, \tag{13}$$

where C_{s} refers to the cardinality of a set, indicating the number of elements it contains. The variable $O_{\min}$ refers to the minimum number of objects required to form a cluster. Additionally, the core distance C can be defined as the distance at which a minimum number of objects, represented by $O_{\min}$, are encompassed within the set. Note that if $O_{\min}$ exceeds the total number of elements in the set, the C becomes undefined, indicating that there are insufficient elements to form a cluster based on the given minimum requirement. The $C(X)$ is defined as follows:

$$C(X) = \begin{cases} \text{undefined}, & \text{if } |N_{\varepsilon\to\infty}(X)| < O_{\min}, \\ \min\limits_{\varepsilon}(|N_\varepsilon(X)|) \leqslant O_{\min}, & \text{if } |N_{\varepsilon\to\infty}(X)| \geqslant O_{\min}, \end{cases} \tag{14}$$

where $|N_{\varepsilon\to\infty}(X)|$ is the number of objects that are infinitely distant from any project X.

The reachability distance R of object X_p concerning object X_q is defined as the maximum value

between the C of X_q and the distance between the two objects. It is important to note that if the minimum number of objects required for cluster formation, $O_{\min}$, exceeds the total number of elements in the set, the reachability distance will be undefined.

$$R(X_p,X_q)=\begin{cases}\text{undefined}, & \text{if } |N_{\varepsilon\to\infty}(X_p)|<O_{\min},\\ \max\left(C(X_p),D(X_p,X_q)\right), & \text{if } |N_{\varepsilon\to\infty}(X_p)|\geqslant O_{\min}.\end{cases} \tag{15}$$

The reachability plot is constructed by sequentially evaluating the R between each object and all other objects in the set. In each iteration, the R between the selected object and the remaining objects is computed. This plot serves as a graphical representation of the ordering of objects based on the reachability distance metric, facilitating the cluster identification (Boroschek and Bilbao, 2019).

2.2.2 Determination of optimal parameters

The neighborhood size parameter ε must be determined for performing physical mode clustering. The KNN search is used to estimate ε. This search helps in estimating ε by considering the distance of any point P to its kth nearest neighbor. Let $D_k^{\mathrm{N}}(P)$ denote the neighborhoods surrounding point P, which includes its KNN as well. Notably, the $D_k^{\mathrm{N}}(P)$ comprises k+1 points including the point P itself. The outline of the estimation algorithm is summarized below:

(1) Determining all points in its $D_k^{\mathrm{N}}(P)$.

(2) Accumulating the distances of all points within their respective $D_k^{\mathrm{N}}(P)$ into a single vector and sorting this vector in ascending order based on the distances.

(3) Plotting the sorted k-distance graph, which displays the sorted distances against the corresponding point numbers.

(4) Locating the knee of the curve on the graph where the distance at that particular point serves as an estimate of ε.

The value of ε can be determined using a k-distance graph in the mode clustering stage, where the distances to the KNN are plotted in descending order (García-Pedrajas et al., 2017; Abu Alfeilat et al., 2019). This graph typically exhibits a multi-linear distribution with a pronounced elbow point. The elbow rule refers to the selection of the value at the elbow, which represents an optimal setting for the search distance. It is expected that the k-distances of core points and border points fall within a certain range, while noise points may exhibit much larger k-distances. However, in some cases, the knee point may not be visible or there may be multiple knee points, making it difficult to determine. The proposed approach entails calculating the average distance between each point and its KNN. The value of k, which corresponds to minimum samples (Minpts), is specified by the user. These k-distances are then shown in ascending order. The objective is to identify the knee point, which represents the optimal ε parameter. The knee point signifies a threshold where a significant change occurs in the k-distance curve. Finally, the determination of the modal parameters that accurately depict the remaining clusters is conducted based on the guidelines proposed by Liu et al. (2023), whereby the mean values ($\bar{f}_p$ and $\bar{\xi}_p$) of f_p and ξ_p are attributed to each respective cluster. The proposed method enables the estimation of modal participation factors using output-only data through three steps (mode shape extraction, influence vector assumption, and modal participation factor calculation), grounded in classical modal analysis theory (Ewins, 2003; Chopra, 2017).

2.3 Procedure of the two-stage automated operational modal identification

In this section, the procedure for two-stage automated operational modal identification is presented in Fig. 1. The flowchart of the proposed method is as follows:

(1) The original data Y measured from practical engineering should be preprocessed to eliminate the spikes and noise. The block Hankel matrix $\boldsymbol{H}$ is established from the processed data. The observability matrix $\boldsymbol{O}_i$ of the ith mode order is estimated based on the singular value decomposition (SVD). Finally, the frequency, the mode shape, and the damping ratio are identified through state matrix $\boldsymbol{A}$ and output matrix $\boldsymbol{C}$ deduced from the shifted observability matrix $\boldsymbol{O}_i^{\uparrow}$ and reversed controllability matrix $\boldsymbol{O}_i^{\downarrow}$.

(2) The modal validation criteria (such as HVC, SVC, and UVC) are applied in this step to eliminate spurious modes by Eqs. (1)–(3) and Eq. (10).

(3) The optimal ε is estimated by Eqs. (12)–(15) based on the KNN method. An enhanced adaptive clustering method is used to identify the physical

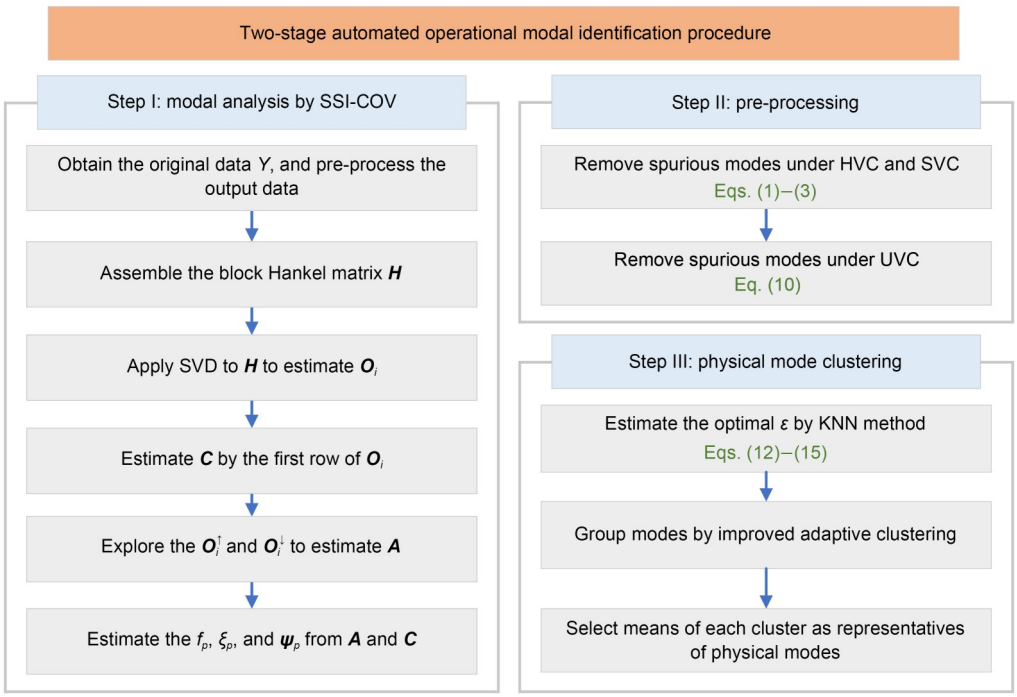

Fig. 1 Procedure of the two-stage automated operational modal identification. COV is the covariance

modes. The means of each cluster are selected as representatives of the physical modes.

3 Methodology verification and application

3.1 Numerical verification through a 10-story shear frame

3.1.1 Description of the 10-story shear frame

A linear time-invariant 10-story shear frame was used for methodology verification (Fig. 2). The model represents floors as masses m_α (α=1, 2, ⋯, 10) interconnected by springs k_α and dampers c_α. The matrices of physical parameters (including mass $\boldsymbol{M}_\alpha$, stiffness $\boldsymbol{K}_\alpha$, and damping $\boldsymbol{C}_\alpha$) can be calculated as follows:

$$\boldsymbol{M}_\alpha=\begin{bmatrix}10 & & & \\ & 10 & & \\ & & \ddots & \\ & & & 10\end{bmatrix},\ \boldsymbol{K}_\alpha=\begin{bmatrix}2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 1\end{bmatrix}\times 10^4,$$

$$\boldsymbol{C}_\alpha=\begin{bmatrix}20 & -10 & & & \\ -10 & 20 & -10 & & \\ & \ddots & \ddots & \ddots & \\ & & -10 & 10\end{bmatrix}. \tag{16}$$

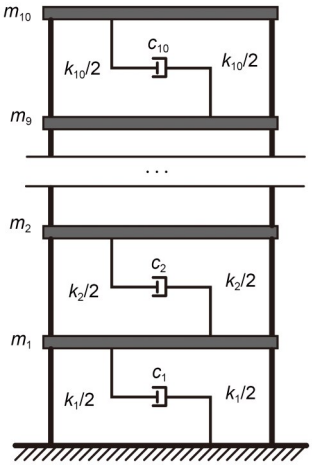

Fig. 2 Schematic diagram of the 10-story shear frame

Each floor is excited by a stationary Gaussian white noise with a mean of 0 to obtain simulated dynamic responses. The excitation lasts 500 s with a sampling frequency of 50 Hz for the requirement of the Nyquist-Shannon sampling theorem. In addition, it is assumed that each floor has an accelerometer arranged at the top of the column for collecting acceleration data. Fig. 3 shows that the signal components exhibit frequency concentration between 2.5 and 10.0 Hz in the time-frequency transform. The noise content becomes more similar to the physical modes in the

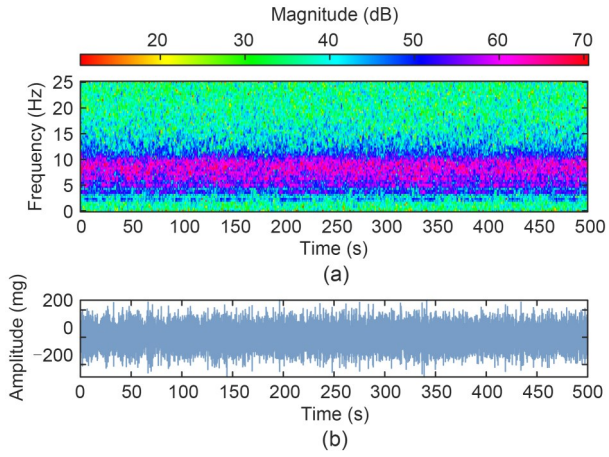

Fig. 3 Results of the time-frequency transform: (a) time-frequency graph; (b) time history

higher frequency band, making it harder to identify the true physical modes. Consequently, the frequency range of interest for this shear frame is 0–10 Hz. The precomputed parameters of the SSI approach are as follows: i=240, and model order n is between 12 and 120. Note that a significant challenge encountered by long-term monitoring systems is the possibility of erroneously interpreting these noise frequencies as physical modes.

3.1.2 Identification results

The proposed approach was used to analyze the measured data, as described in Section 2. All calculations were performed on a computer equipped with a 13th Gen Intel® Core™ i7-13700KF 3.40 GHz Processor and 32 GB of RAM. The same computational setup was used for all subsequent calculations in this study. The computation time for a single run of the proposed method was 156.5 s. The identified results of the spurious modes pre-filtering stage are presented in Fig. 4. Following the application of conventional modal validation criteria, such as the damping ratio check and modal complexity check, all the remaining possible physical modes in the stabilization diagram are shown in Fig. 4a. The scattered poles of the stabilization diagram are filtered out based on the uncertainty validation criterion in Fig. 4b. It is evident that, when relying solely on traditional validation criteria, the stabilization diagram still appears chaotic and contains numerous scattered poles. The uncertainty criterion can be used to eliminate spurious modes more efficiently than the usual validation criterion, hence

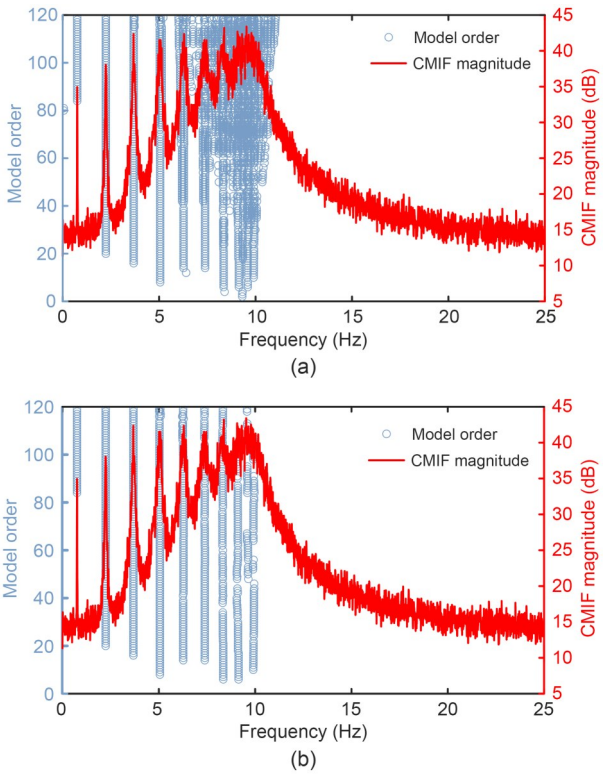

Fig. 4 Modal identification results: (a) after conventional validation criteria; (b) after proposed modal clustering in Step II. CMIF represents the complex mode indicator function

speeding up the automation process. The results of the spurious modes pre-filtering stage show that the uncertainty criterion outperforms the standard validation criterion.

The final step is the automatic detection of clusters from the reachability plot. The reachability plot serves as a graphical representation of the objects' order, obtained through the reachability distance metric. It is derived from the remaining poles in the stabilization diagram following the spurious modes pre-filtering stage. This plot aids in identifying clusters based on the obtained order. By observing the reachability plot, the clustering structure can be observed and derived. As shown in Fig. 5, regions with low reachability distance indicate objects belonging to the same cluster, while regions with high reachability distance are likely to represent outliers. Through a manual inspection, the parameter ε was selected as 0.05.

To increase automation, Minpts and ε should be set based on the framework of the improved self-adaptive algorithm discussed previously. Fig. 6 shows the estimation of an optimal ε considering the optimal

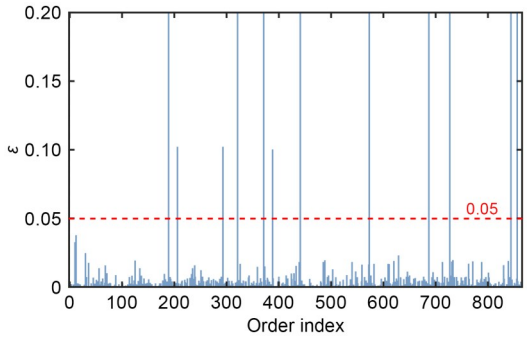

Fig. 5 Reachability plot

values of the KNN search. The line that joins the curve's initial and final points is defined using this method. The value of ε is defined by the ordinate of the point on the sorted k-distance graph that is perpendicular to and farthest from the line. The algorithm determines the average estimate of ε for each curve when a range of k is given. The distances, which range from 0.1 times the number of model orders to 12–120 times the number of model orders, are plotted against the point index k in Fig. 6. The knee of the time-averaged ε occurs at 0.046, which is close to the results of manual inspection. The points over ε are regarded as noise, whereas those below it are part of a cluster.

Fig. 7 presents the identified results obtained at the modal clustering stage after applying the self-adaptive algorithm. It shows the relationship between damping ratios and frequencies with 10 different colors denoting 10 distinct groups. The identified damping ratios and frequencies exhibit a close and consistent agreement, indicating the reliability of the results. Table 1 shows the identification results of modal parameters, including frequencies, damping ratios, and MAC, along with their corresponding theoretical results. The computation of the C_V is performed using uncertainty validation criterion analysis, as described in Section 2.1. Note that the proposed AOMA approach effectively captures closely spaced modes from identified frequencies. The closely spaced modes 9.60 and 9.92, are successfully separated. As shown in Fig. 4a, several spurious closely spaced modes are present in the frequency range from 6 to 10 Hz. The 7th to 10th modes are effectively separated, demonstrating the feasibility of the proposed method in identifying closely spaced modes. Fig. 8 illustrates the identified mode shapes of the structural model under ambient vibration. The identified and theoretical mode shapes are illustrated by red dotted lines and blue solid lines, respectively, while the measured positions are represented by dots. The identified mode shapes are close to the theoretical ones, indicating the effectiveness of the identification process. In conclusion, the proposed approach enables the automatic identification of modal parameters, while preserving the physical modes at specific frequencies. The method accurately identifies closely spaced modes without needing to specify the number of clusters or a threshold value for clustering.

Due to measurement errors and load uncertainties, noise inevitably contaminates the actual measured acceleration responses. To assess the robustness of the AOMA method, Gaussian white noise at a specified noise level was introduced to the structural acceleration response. The noise level R_n is defined as follows:

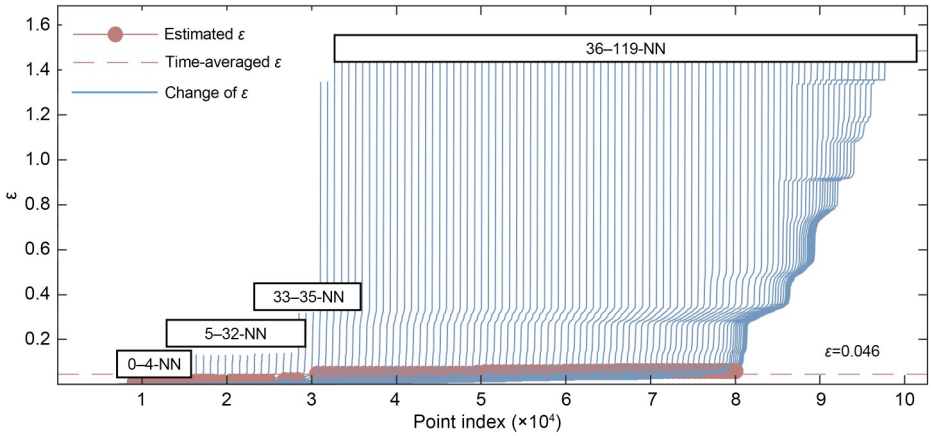

Fig. 6 Determination of Minpts and ε by KNN. In *a*–*b*-NN, *a*–*b* represents the range of order of estimated ε, and NN represents the nearest neighbor

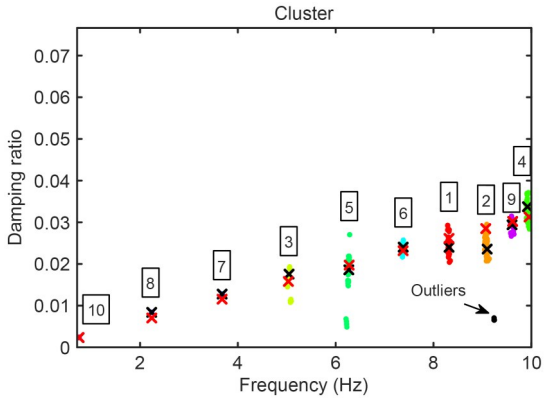

Fig. 7 Plot of clusters identified from the damping ratio and frequency. Black and red crosses represent the mean values of the identified and theoretical modal parameters, respectively. References to color refer to the online version of this figure

$$\boldsymbol{A}_{\text{noisy}} = \boldsymbol{A}_0 + R_{\text{n}} \boldsymbol{N} \sigma(\boldsymbol{A}_0), \tag{17}$$

where $\boldsymbol{A}_{\text{noisy}}$ denotes the structural acceleration response with added noise; $\boldsymbol{A}_0$ denotes the theoretical structural acceleration response; $\boldsymbol{N}$ denotes the Gaussian white noise sequence with a standard normal distribution of 0 mean and unit variance; $\sigma(\boldsymbol{A}_0)$ denotes the variance of the true structural acceleration response. Considering noise levels ranging from 0% to 5%, the relative deviations of parameter identification results for the frame were calculated. Fig. 9 shows the average relative deviations of the identification results at different noise levels. The curve indicates that the impact of measurement noise on each case is generally consistent. As the noise level increases, the accuracy of the

Table 1 Proposed AOMA identification results for the 10-story shear frame

Mode	Frequency			Damping ratio			MAC
	Identified (Hz)	C_{V} (%)	Theoretical (Hz)	Identified (%)	C_{V} (%)	Theoretical (%)	
1	0.7532	0.177	0.7522	0.23	9.483	0.24	0.9953
2	2.2382	0.251	2.2399	0.84	10.064	0.70	0.9997
3	3.6781	0.267	3.6775	1.28	8.851	1.16	0.9994
4	5.0463	0.144	5.0329	1.76	7.436	1.58	0.9995
5	6.2689	0.261	6.2760	1.86	20.850	1.97	0.9996
6	7.3808	0.102	7.3788	2.40	3.395	2.32	0.9989
7	8.3178	0.019	8.3168	2.40	0.744	2.61	0.9994
8	9.0970	0.001	9.0690	2.36	0.003	2.85	0.9995
9	9.6061	0.233	9.6186	2.94	5.168	3.02	0.9976
10	9.9213	0.003	9.9534	3.37	1.273	3.13	0.9938

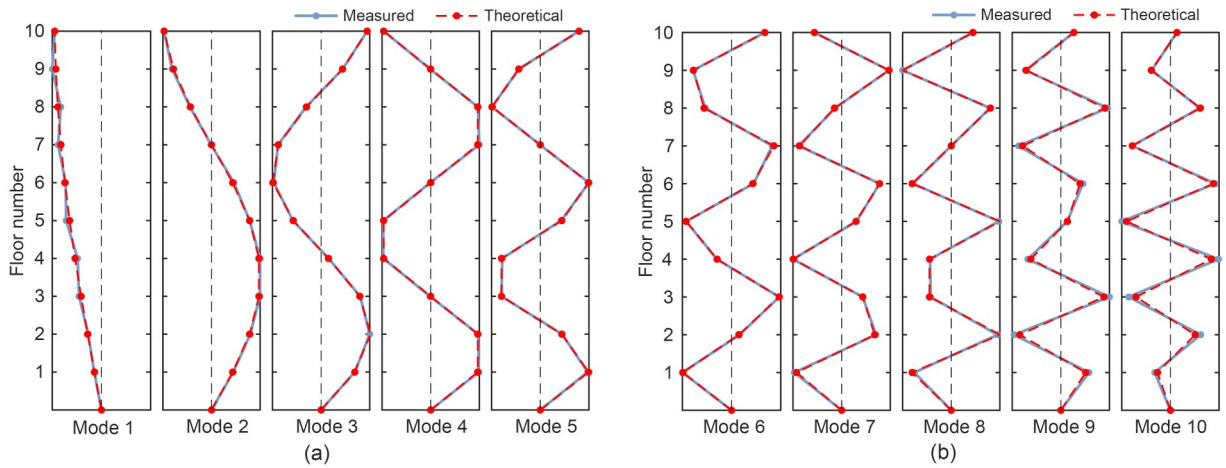

Fig. 8 Comparison between measured and theoretical mode shapes: (a) modes 1–5; (b) modes 6–10. References to color refer to the online version of this figure

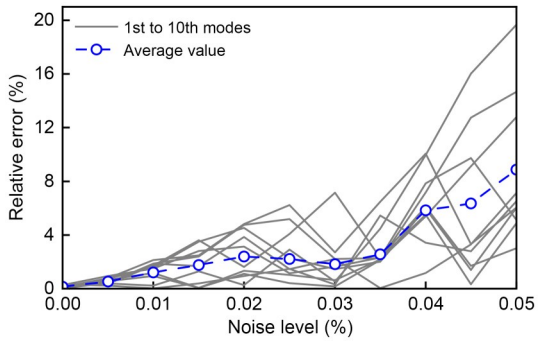

Fig. 9 Relative deviations of parameter identification results under different noise levels

identification results gradually decreases. For different modes, the proposed method provides accurate identification results for frequency when the measurement noise is less than 4%, with the average relative deviation for each case not exceeding 5%, demonstrating the robustness of the proposed method against environmental noise.

3.2 Application to field measurement of a large-scale bridge

3.2.1 Description of the Taoyaomen Bridge and its monitoring system

Monitoring data from the Taoyaomen Bridge were adopted to further demonstrate the performance of the presented AOMA approach. The Taoyaomen Bridge, a cable-stayed bridge, is located in Zhoushan City, Zhejiang Province, China (Fig. 10a). The bridge spans the Taoyaomen waterway and connects Fuchi Island and Cezi Island. It is the third sea-crossing bridge of the Zhoushan mainland–island connection project, with a bridge deck width of 27.6 m and four lanes for two-way traffic. The main girder is a steel structure stiffening girder, and the bridge foundation is drilled piles. The upper part of the bridge is a diamond-shaped tower, with a height of 151 m. Ten accelerometers were strategically installed on the deck for collecting vertical acceleration data under environmental and operational effects (Fig. 10b). The Taoyaomen Bridge is equipped with a structural health monitoring system that consists of a data acquisition transmission system, a data management system, and an evaluation system (Fig. 10c). The method proposed in this study can be integrated into the evaluation system to analyze real-time monitoring data transmitted over the 4G network, facilitating the automated identification of changes in the bridge's modal parameters. These accelerometers continuously recorded the vibrations at a sampling frequency of 50 Hz. The acceleration data used in this study were collected on July 7, 2022, at 3:00 pm. A dataset with a length of 15000 samples was selected for estimating the modal parameters. The parameters of the SSI-COV algorithm are set as follows: the model order ranged from 6 to 60, and i=120. The computation time for a single run of the proposed method was 126.5 s.

3.2.2 Identification results

The identification results obtained with traditional validation criteria are shown in Fig. 11a, where many spurious modes are present. The singular value spectrum is illustrated in the stabilization diagram. The conventional and uncertainty validation criteria were applied to improve the clarity of the diagram (Fig. 11b). The inset box in Fig. 11b shows the identification of nine vertical blue alignments, while the conventional stabilization diagram still looks busy. This scattering indicates the presence of a spurious mode at 3.5 Hz, indicating the significant influence of environmental changes on the high-order mode fluctuations. The identification results of Stage II show that the uncertainty criterion works better than conventional validation criteria. Accurate assessment of these fluctuations is essential for any AOMA approach, regardless of its particular application.

The evolution of the frequencies of the first six modes is plotted. The physical modes are clustered using OPTICS combined with the KNN algorithm. The physical mode clustering step then begins with an estimated clustering threshold. The estimated threshold in this case was 0.193 (Fig. 12). This threshold represents a stricter criterion that enables the removal of more spurious modes while preserving the physical modes. The first nine modes, labeled as modes 1 to 9 (a total of nine clusters), were used as a baseline in Fig. 13. Note that the forth mode and fifth mode have close spacing, which is a prevalent problem in AOMA. To address this issue, robust outlier identification was used to eliminate extremely variable frequencies and damping ratios. The relationship between damping ratios and frequencies is illustrated in Fig. 13. Each of the nine different colors represents a distinct cluster. The results are more reliable because of the strong agreement and consistency between the identified damping ratios and frequencies.

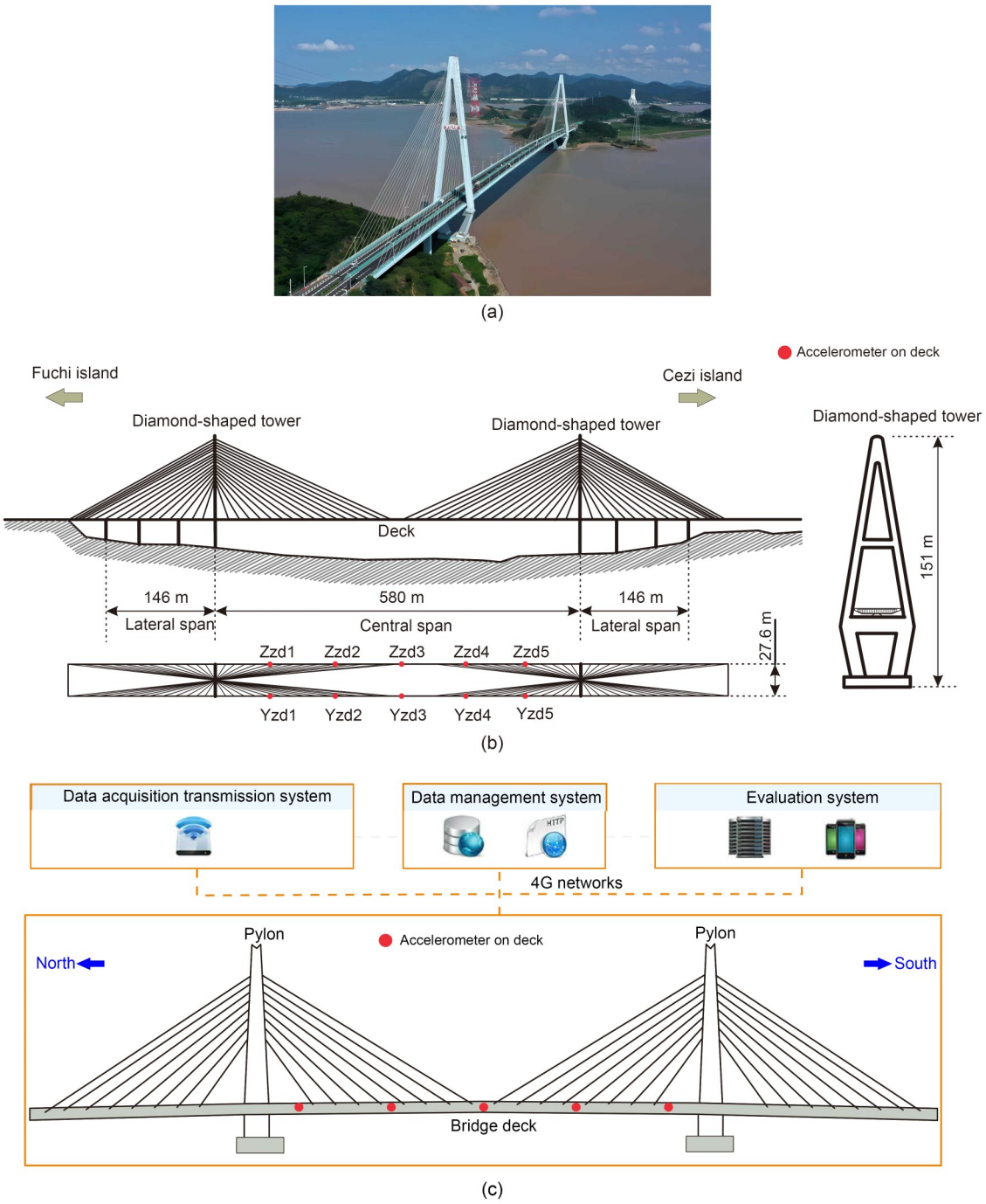

Fig. 10 Description of the Taoyaomen bridge: (a) photo; (b) geometry of the bridge and layout of the sensors; (c) monitoring system. Zzd*i* and Yzd*j* (*i*, *j*=1, 2, ⋯, 5) represent the measurement point number

The distribution of identified frequencies and damping ratios of the first six mode shapes for 96 datasets for two days is shown in Figs. 14 and 15. The yellow dashed line in the graph represents the mean of the normal distribution, while the red solid line depicts the probability density function of the corresponding distribution. All frequencies of the first six modes fit well with Gaussian normal distributions. On the other hand, the damping ratios show a heavily-tailed Weibull distribution. The identification results indicate that modal parameter identification is significantly affected by changes in the environment. In conclusion, the proposed two-stage AOMA technique performed remarkably well in modal parameter tracking

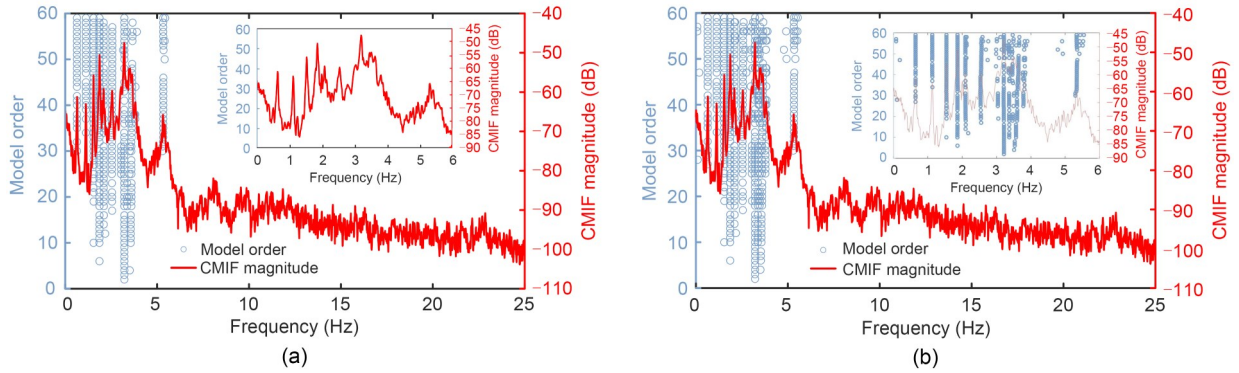

Fig. 11 Modal identification results of the Taoyaomen Bridge: (a) after conventional validation criteria; (b) after proposed modal clustering in Stage II

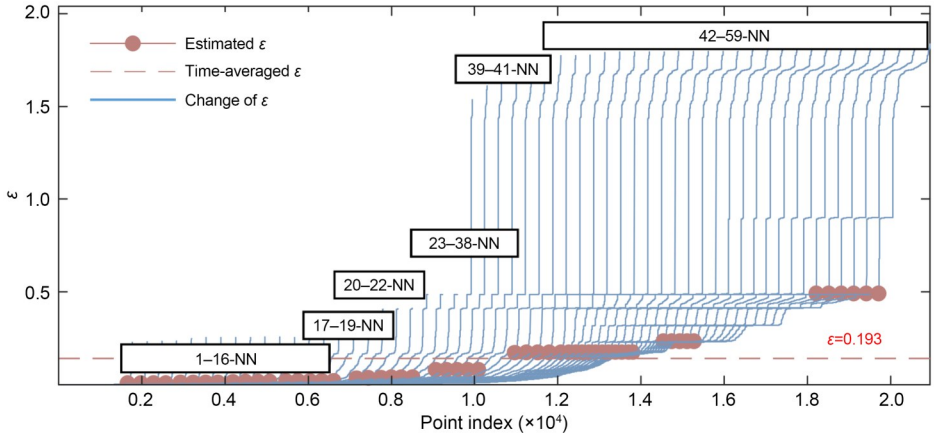

Fig. 12 Determination of Minpts and ε by KNN. In a–b-NN, a–b represents the range of order of estimated ε, and NN represents the nearest neighbor

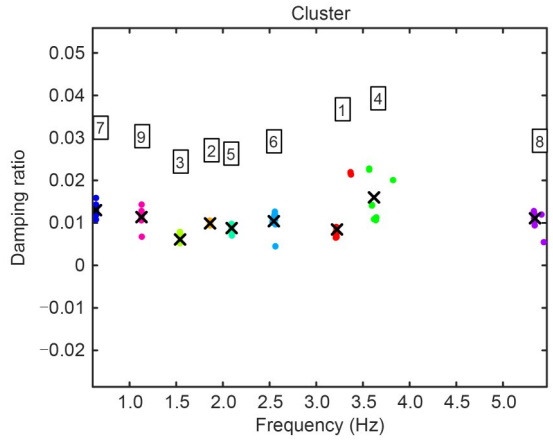

Fig. 13 Clusters identified from the damping ratio versus frequency plot. Black crosses indicate the mean damping ratio. References to color refer to the online version of this figure

over 96 observed datasets during two days without the need for human intervention. The results of this study show that the proposed AOMA approach is promising, robust, and applicable for online structural health monitoring with potential extension to different types of structures.

The mode shapes identified from measured data and finite element model (FEM) analysis were compared to assess the identification accuracy of the proposed AOMA approach. Fig. 16 shows the first six normalized mode shapes of the Taoyaomen Bridge. The blue lines are the nodes of the bridge deck, and the red dots are the positions of the accelerometers. The first, second, and fifth modes are the flexural modes in the vertical direction. The third, fourth, and sixth modes are the torsional modes. The accuracy of the identified modal parameters meets engineering requirements, as shown by the average MAC of the mode shapes in AMOA and FEM analysis being greater than 0.95. Because of the influence of environmental noise, the MAC of high-order mode shapes is slightly lower than that of low-order mode shapes.

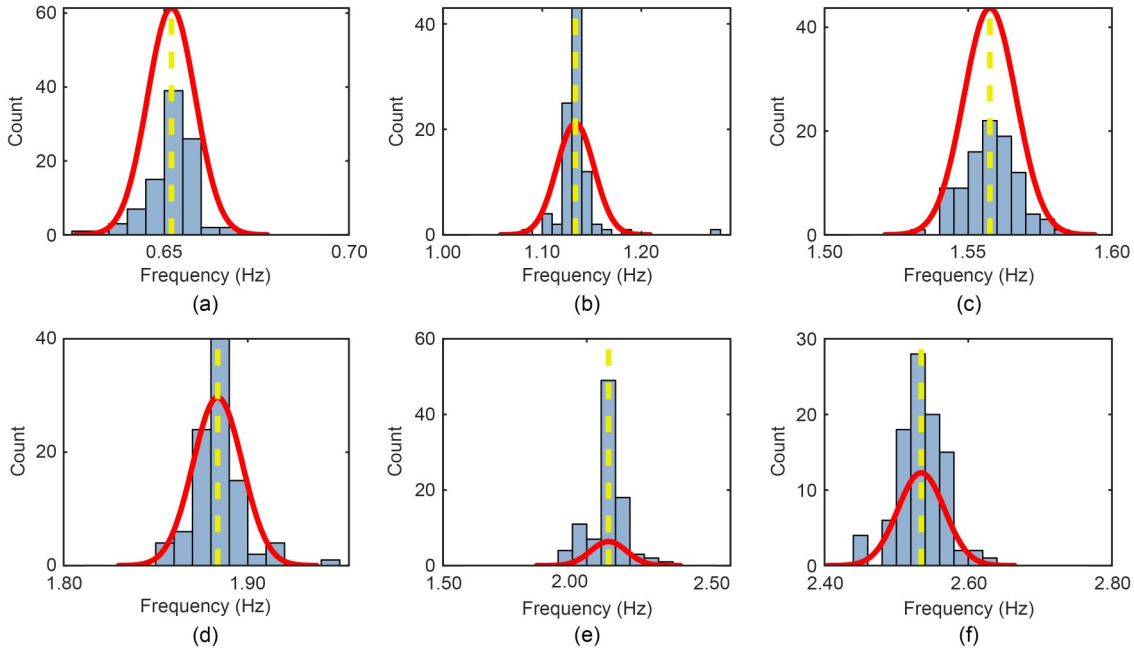

Fig. 14 Frequency distributions (normal distribution): (a) mode 1 (0.65 Hz); (b) mode 2 (1.13 Hz); (c) mode 3 (1.56 Hz); (d) mode 4 (1.88 Hz); (e) mode 5 (2.08 Hz); (f) mode 6 (2.53 Hz). References to color refer to the online version if this figure

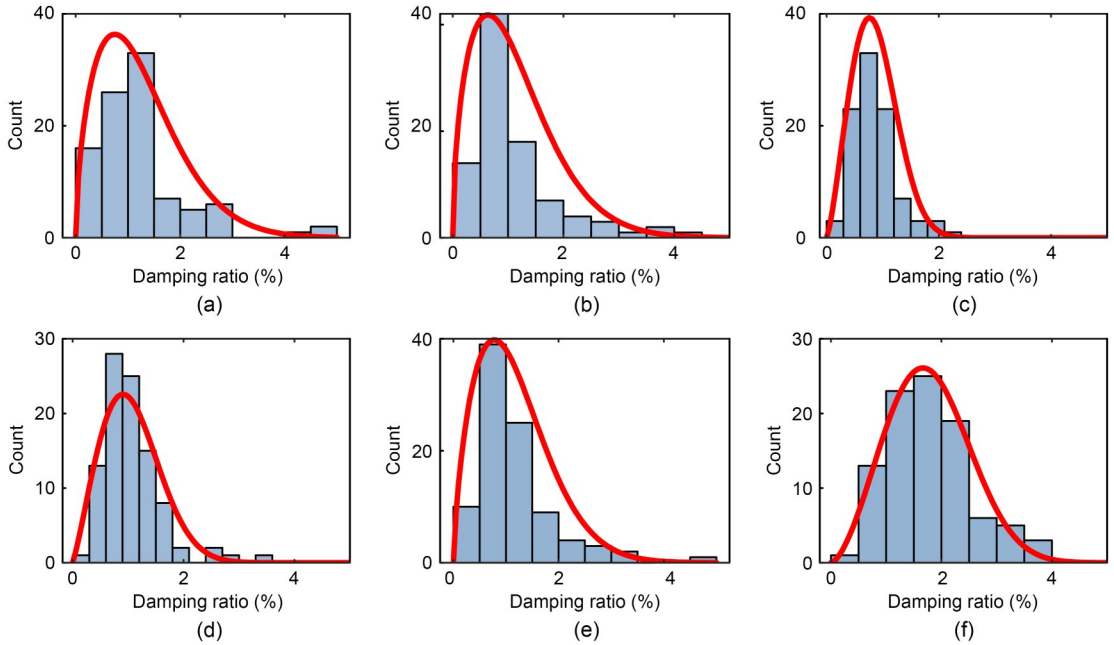

Fig. 15 Damping ratio distributions (Weibull distribution): (a) mode 1 (1.39%); (b) mode 2 (1.26%); (c) mode 3 (0.97%); (d) mode 4 (1.21%); (e) mode 5 (1.31%); (f) mode 6 (2.00%). References to color refer to the online version if this figure

The proposed AOMA approach was used to identify the modal parameters continually. The results of the frequencies identified for the first six mode shapes are shown in Fig. 17. The proposed approach had a high accuracy in the continuous automatic identification of modal parameters. The identified frequencies did not seem to be sensitive to the environmental effects. The damping ratio is another important modal parameter, which is important for vibration control of the large-span structures. The time histories of damping ratios are shown in Fig. 18. Although more spread than the frequencies, all of the measured damping

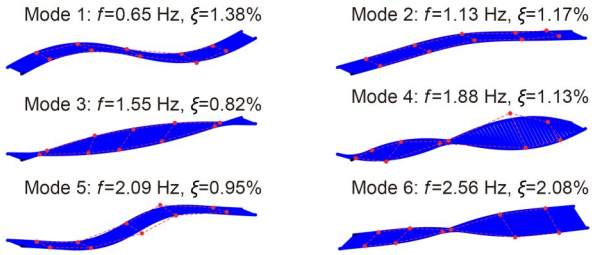

Fig. 16 First six normalized mode shapes of the Taoyaomen Bridge. References to color refer to the online version of this figure

ratios maintained a respectable fluctuation amplitude. Sudden peaks can be observed for the first, second, and fifth mode shapes. According to the recorded data from the anemometer at the bridge site, sudden wind gusts during this period may have induced a large amplitude spike in the damping ratios.

4 Conclusions

In this paper, we present a new two-stage framework for AOMA based on OPTICS. The distinctive feature of this framework is the incorporation of a fully automatic density-based clustering approach without requiring specification of the number of clusters in advance. The effectiveness and applicability of the proposed approach are demonstrated through numerical verification on a 10-story shear frame and practical application to the Taoyaomen Bridge. The main summaries and conclusions are as follows:

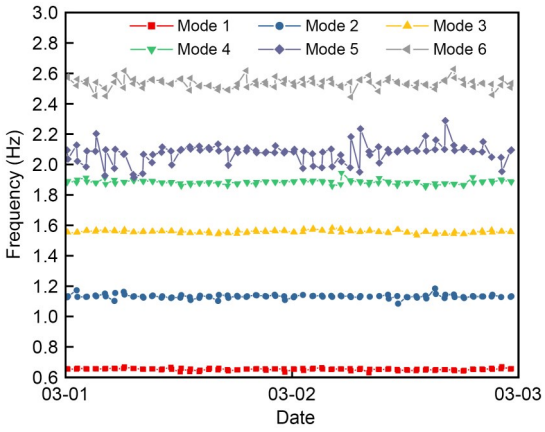

Fig. 17 Identified frequencies of the first six modes

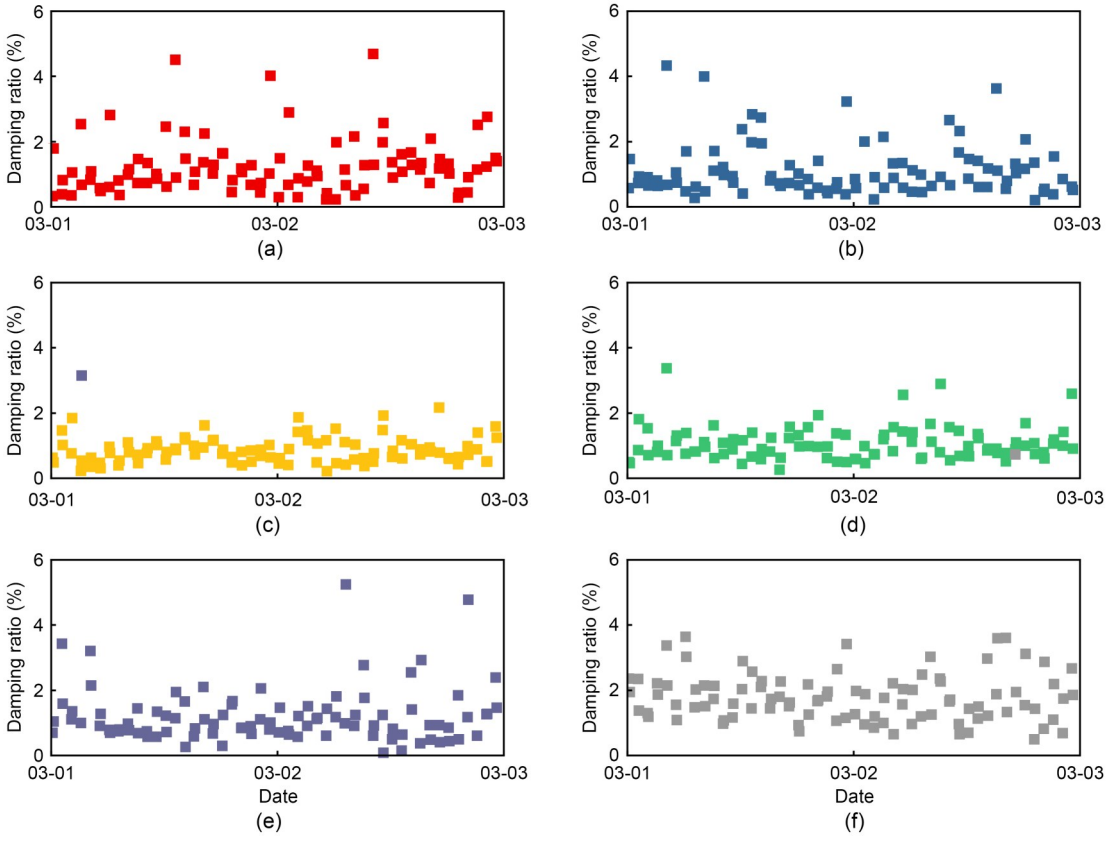

Fig. 18 Identified damping ratios of the first six modes: (a) mode 1; (b) mode 2; (c) mode 3; (d) mode 4; (e) mode 5; (f) mode 6

1. According to the results of the modal parameters identified from the 10-story frame and the Taoyaomen Bridge, the frequencies and damping ratios of each mode remain within a narrow range, validating that the proposed approach can effectively remove spurious modes and identify physical modes.

2. The proposed approach performs well in identifying closely spaced modes, as demonstrated by numerical verification on the shear frame. These modes have been successfully discovered without the need for specifying the number of clusters or a threshold value for clustering.

3. Six mode shapes of the Taoyaomen Bridge were identified, and the modal parameters were continuously tracked over several days. The analysis of extensive datasets collected from SHM systems confirmed the practicability of the approach.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 52408200), the Natural Science Foundation of Jiangsu Province (No. BK20240996), China, the Suzhou Science and Technology Plan (Basic Research) Project (No. SJC2023002), China, and the Natural Science Research Projects of Colleges and Universities in Jiangsu Province (No. 24KJB560022), China.

Author contributions

Yi CHEN and Wenwei FU designed the research and wrote the first draft of the manuscript. Yaozhi LUO and Yanbin SHEN revised and edited the final version. Hui YANG supervised and administrated the project. Shiying WANG helped to organize the manuscript.

Conflict of interest

Yi CHEN, Wenwei FU, Yaozhi LUO, Yanbin SHEN, Hui YANG, and Shiying WANG declare that they have no conflict of interest.

References

Abu Alfeilat HA, Hassanat ABA, Lasassmeh O, et al., 2019. Effects of distance measure choice on *K*-nearest neighbor classifier performance: a review. *Big Data*, 7(4):221-248. https://doi.org/10.1089/big.2018.0175

Boroschek RL, Bilbao JA, 2019. Interpretation of stabilization diagrams using density-based clustering algorithm. *Engineering Structures*, 178:245-257. https://doi.org/10.1016/j.engstruct.2018.09.091

Cabboi A, Magalhães F, Gentile C, et al., 2017. Automated modal identification and tracking: application to an iron arch bridge. *Structural Control and Health Monitoring*, 24(1):e1854. https://doi.org/10.1002/stc.1854

Chen J, Xu YL, Zhang RC, 2004. Modal parameter identification of Tsing Ma suspension bridge under Typhoon Victor: EMD-HT method. *Journal of Wind Engineering and Industrial Aerodynamics*, 92(10):805-827. https://doi.org/10.1016/j.jweia.2004.04.003

Chopra AK, 2017. Dynamics of Structures-Theory and Applications to Earthquake Engineering. Prentice Hall, Englewood Cliffs, USA.

Civera M, Mugnaini V, Fragonara LZ, 2022. Machine learning-based automatic operational modal analysis: a structural health monitoring application to masonry arch bridges. *Structural Control and Health Monitoring*, 29:e3028. https://doi.org/10.1002/stc.3028

Civera M, Sibille L, Fragonara LZ, et al., 2023. A DBSCAN-based automated operational modal analysis algorithm for bridge monitoring. *Measurement*, 208:112451. https://doi.org/10.1016/j.measurement.2023.112451

Desjardins S, Lau D, 2022. Advances in intelligent long-term vibration-based structural health-monitoring systems for bridges. *Advances in Structural Engineering*, 25(7):1413-1430. https://doi.org/10.1177/1369433222

Döhler M, Mevel L, 2013. Efficient multi-order uncertainty computation for stochastic subspace identification. *Mechanical Systems and Signal Processing*, 38(2):346-366. https://doi.org/10.1016/j.ymssp.2013.01.012

Doroudi R, Hosseini Lavassani SH, Shahrouzi M, et al., 2022. Identifying the dynamic characteristics of super tall buildings by multivariate empirical mode decomposition. *Structural Control and Health Monitoring*, 29(3):e3075. https://doi.org/10.1002/stc.3075

Ereiz S, Duvnjak I, Jiménez-Alonso JF, 2022. Review of finite element model updating methods for structural applications. *Structures*, 41:684-723. https://doi.org/10.1016/j.istruc.2022.05.041

Ewins DJ, 2003. Modal Testing: Theory, Practice and Application. Research Students Press LTD., Baldock, UK.

Fan G, Li J, Hao H, 2019. Improved automated operational modal identification of structures based on clustering. *Structural Control and Health Monitoring*, 26(12):1-23. https://doi.org/10.1002/stc.2450

Farrar CR, Worden K, 2007. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):303-315. https://doi.org/10.1098/rsta.2006.1928

Feng YH, Su YH, Zhao C, et al., 2024. A two-stage automated OMA framework for transmission towers based on clustering algorithms. *Structures*, 61:106023. https://doi.org/10.1016/j.istruc.2024.106023

García-Pedrajas N, del Castillo JAR, Cerruela-García G, 2017. A proposal for local *k* values for *k*-nearest neighbor rule. *IEEE Transactions on Neural Networks and Learning Systems*, 28(2):470-475. https://doi.org/10.1109/TNNLS.2015.2506821

Garcia-Perez A, Amezquita-Sanchez JP, Dominguez-Gonzalez A, et al., 2013. Fused empirical mode decomposition and wavelets for locating combined damage in a truss-type structure through vibration analysis. *Journal of Zhejiang University-SCIENCE A*, 14(9):615-630. https://doi.org/10.1631/jzus.A1300030

Greś S, Döhler M, Andersen P, et al., 2021. Uncertainty quantification for the modal phase collinearity of complex mode shapes. *Mechanical Systems and Signal Processing*, 152:107436. https://doi.org/10.1016/j.ymssp.2020.107436

Gu JF, Gul M, Wu XG, 2017. Damage detection under varying temperature using artificial neural networks. *Structural Control and Health Monitoring*, 24(11):e1998. https://doi.org/10.1002/stc.1998

He M, Liang P, Li J, et al., 2021. Fully automated precise operational modal identification. *Engineering Structures*, 234:111988. https://doi.org/10.1016/j.engstruct.2021.111988

He Y, Yang JP, Li YF, 2022. A three-stage automated modal identification framework for bridge parameters based on frequency uncertainty and density clustering. *Engineering Structures*, 255:113891. https://doi.org/10.1016/j.engstruct.2022.113891

Hou RR, Xia Y, 2021. Review on the new development of vibration-based damage identification for civil engineering structures: 2010-2019. *Journal of Sound and Vibration*, 491:115741. https://doi.org/10.1016/j.jsv.2020.115741

Huang CS, Le QT, Su WC, et al., 2020. Wavelet-based approach of time series model for modal identification of a bridge with incomplete input. *Computer-Aided Civil and Infrastructure Engineering*, 35(9):947-964. https://doi.org/10.1111/mice.12539

Jeong S, Ferguson M, Hou R, et al., 2019. Sensor data reconstruction using bidirectional recurrent neural network with application to bridge monitoring. *Advanced Engineering Informatics*, 42:100991. https://doi.org/10.1016/j.aei.2019.100991

Jin SS, Jeong S, Sim SH, et al., 2021. Fully automated peak-picking method for an autonomous stay-cable monitoring system in cable-stayed bridges. *Automation in Construction*, 126:103628. https://doi.org/10.1016/j.autcon.2021.103628

Kang F, Li JJ, 2020. Displacement model for concrete dam safety monitoring via Gaussian process regression considering extreme air temperature. *Journal of Structural Engineering*, 146(1):05019001. https://doi.org/10.1061/(asce)st.1943-541x.0002467

Liu DW, Bao YQ, Li H, 2023. Machine learning-based stochastic subspace identification method for structural modal parameters. *Engineering Structures*, 274:115178. https://doi.org/10.1016/j.engstruct.2022.115178

Liu YC, Loh CH, Ni YQ, 2013. Stochastic subspace identification for output-only modal analysis: application to super high-rise tower under abnormal loading condition. *Earthquake Engineering & Structural Dynamics*, 42(4):477-498. https://doi.org/10.1002/eqe.2223

Luo YZ, Chen Y, Wan HP, et al., 2021. Development of laser-based displacement monitoring system and its application to large-scale spatial structures. *Journal of Civil Structural Health Monitoring*, 11(2):381-395. https://doi.org/10.1007/s13349-020-00459-4

Luo YZ, Fu WW, Wan HP, et al., 2022. Load-effect separation of a large-span prestressed structure based on an enhanced EEMD-ICA methodology. *Journal of Structural Engineering*, 148(3):04021288. https://doi.org/10.1061/(ASCE)ST.1943-541X.0003263

Mao JX, Wang H, Fu YG, et al., 2019. Automated modal identification using principal component and cluster analysis: application to a long-span cable-stayed bridge. *Structural Control and Health Monitoring*, 26(10):e2430. https://doi.org/10.1002/stc.2430

McLachlan GJ, Lee SX, Rathnayake SI, 2019. Finite mixture models. *Annual Review of Statistics and Its Application*, 6:355-378. https://doi.org/10.1146/annurev-statistics-031017-100325

Mostafaei H, Ghamami M, Aghabozorgi P, 2021. Modal identification of concrete arch dam by fully automated operational modal identification. *Structures*, 32:228-236. https://doi.org/10.1016/j.istruc.2021.03.028

Mugnaini V, Fragonara LZ, Civera M, 2022. A machine learning approach for automatic operational modal analysis. *Mechanical Systems and Signal Processing*, 170:108813. https://doi.org/10.1016/j.ymss p.2022.108813

Neu E, Janser F, Khatibi AA, et al., 2017. Fully automated operational modal analysis using multi-stage clustering. *Mechanical Systems and Signal Processing*, 84:308-323. https://doi.org/10.1016/j.ymssp.2016.07.031

Ni YQ, Wang YW, Zhang C, 2020. A Bayesian approach for condition assessment and damage alarm of bridge expansion joints using long-term structural health monitoring data. *Engineering Structures*, 212:110520. https://doi.org/10.1016/j.engstruct.2020.110520

Pan YX, Ventura CE, Xiong HB, et al., 2020. Model updating and seismic response of a super tall building in Shanghai. *Computers & Structures*, 239:106285. https://doi.org/10.1016/j.compstruc.2020.106285

Peeters B, de Roeck G, 1999. Reference-based stochastic subspace identification for output-only modal analysis. *Mechanical Systems and Signal Processing*, 13(6):855-878. https://doi.org/10.1006/mssp.1999.1249

Peeters B, Lowet G, van der Auweraer H, et al., 2004. A new procedure for modal parameter estimation. *Sound and Vibration*, 38(1):24-29.

Rainieri C, Fabbrocino G, 2014. Operational Modal Analysis of Civil Engineering Structures. Springer, New York, USA. https://doi.org/10.1007/978-1-4939-0767-0

Ran L, Ding Y, Chen QZ, et al., 2023. Influence of adjacent shield tunneling construction on existing tunnel settlement: field monitoring and intelligent prediction. *Journal of Zhejiang University-SCIENCE A*, 24(12):1106-1119. https://doi.org/10.1631/jzus.A2200573

Ren WX, Zong ZH, 2004. Output-only modal parameter identification of civil engineering structures. *Structural Engineering and Mechanics*, 17(3-4):429-444.

https://doi.org/10.12989/SEM.2004.17.3_4.429

Reynders E, de Roeck G, 2008. Reference-based combined deterministic–stochastic subspace identification for experimental and operational modal analysis. *Mechanical Systems and Signal Processing*, 22(3):617-637. https://doi.org/10.1016/j.ymssp.2007.09.004

Reynders E, Pintelon R, de Roeck G, 2008. Uncertainty bounds on modal parameters obtained from stochastic subspace identification. *Mechanical Systems and Signal Processing*, 22(4):948-969. https://doi.org/10.1016/j.ymssp.2007.10.009

Reynders E, Maes K, Lombaert G, et al., 2016. Uncertainty quantification in operational modal analysis with stochastic subspace identification: validation and applications. *Mechanical Systems and Signal Processing*, 66-67:13-30. https://doi.org/10.1016/j.ymssp.2015.04.018

Sadeqi A, Esfandiari A, Sanayei M, et al., 2022. Automated operational modal analysis based on long-term records: a case study of Milad Tower structural health monitoring. *Structural Control and Health Monitoring*, 29(10):e3037. https://doi.org/10.1002/stc.3037

Saxena A, Prasad M, Gupta A, et al., 2017. A review of clustering techniques and developments. *Neurocomputing*, 267: 664-681. https://doi.org/10.1016/j.neucom.2017.06.053

Sun LM, Shang ZQ, Xia Y, et al., 2020. Review of bridge structural health monitoring aided by big data and artificial intelligence: from condition assessment to damage detection. *Journal of Structural Engineering*, 146(5):04020073. https://doi.org/10.1061/(ASCE)ST.1943-541X.0002535

Sun M, Alamdari MM, Kalhori H, 2017. Automated operational modal analysis of a cable-stayed bridge. *Journal of Bridge Engineering*, 22(12):05017012. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001141

Wu WH, Wang SW, Chen CC, et al., 2019. Modal parameter identification for closely spaced modes of civil structures based on an upgraded stochastic subspace methodology. *Structure and Infrastructure Engineering*, 15(3):296-313. https://doi.org/10.1080/15732479.2018.1547770

Wu Y, Fu HR, Bian XC, et al., 2023. Impact of extreme climate and train traffic loads on the performance of high-speed railway geotechnical infrastructures. *Journal of Zhejiang University-SCIENCE A*, 24(3):189-205. https://doi.org/10.1631/jzus.A2200341

Yaghoubi V, Vakilzadeh MK, Abrahamsson TJS, 2018. Automated modal parameter estimation using correlation analysis and bootstrap sampling. *Mechanical Systems and Signal Processing*, 100:289-310. https://doi.org/10.1016/j.ymssp.2017.07.004

Yun CB, Cho S, Park HJ, et al., 2014. Smart wireless sensing and assessment for civil infrastructure. *Structure and Infrastructure Engineering*, 10(4):534-550. https://doi.org/10.1080/15732479.2013.769011

Zeng JC, Hu Z, 2022. Automated operational modal analysis using variational Gaussian mixture model. *Engineering Structures*, 273:115139. https://doi.org/10.1016/j.engstruct.2022.115139

Electronic supplementary materials

Section S1