A Two-stage Framework for Automated Operational Modal Identification

Jice Zeng[a], Young Hoon Kim[b]
[a] *Ph.D. Candidate. Department of Civil and Environment Engineering. University of Louisville, Louisville, Kentucky, 40292, USA.*
[b] *Corresponding author. Associate Professor. Department of Civil and Environment Engineering. University of Louisville, Louisville, Kentucky, 40292, USA*
Email: young.kim@louisville.edu (Y. H. Kim)

A Two-stage Framework for Automated Operational Modal Identification

Abstract: Automated operational modal analysis (OMA) is attractive and has been extensively used to replace traditional OMA, which involves much empirical observation and engineers' judgment. However, the uncertainties on modal parameters and spurious modes are still challenging to estimate under the field conditions. For addressing this challenge, this research proposed an automated modal identification approach. The proposed approach consists of two steps: (1) modal analysis using covariance-driven stochastic subspace algorithm (SSI-cov/ref); (2) automated interpretation of the stabilization diagram. An additional uncertainty criterion is employed to initially remove as many spurious modes as possible. A novel threshold calculation for clustering is proposed with incorporating uncertainty of modal parameters and the weighting factor. An improved self-adaptive clustering with new distance calculation is used to group physical modes, followed by the final step of robust outlier detection to select outlying modes. The proposed automated approach requires minimum human intervention. Two field tests of the footbridge and a post-tensioned concrete bridge are used to verify the proposed approach. A modal tracking was used for continuously measured data for demonstrating the applicability of the approach. Results show the proposed approach has fairly good performance and be suitable for automated OMA and long-term health monitoring.
Keywords: Operational modal analysis; Automated interpretation; Uncertainty criterion; Threshold calculation; Clustering; Minimum human intervention; Long-term health monitoring

1. Introduction

Over the last few decades, long-term structural health monitoring (SHM) using vibration responses has been applied to civil infrastructures such as bridges and buildings (Chen & Ni 2018). It is widely recognized that continuous monitoring is useful in tracking changes in dynamic modal parameters: frequencies, mode shapes, and damping ratios. In particular, operational modal analysis (OMA) is paid attention to because it can be implemented efficiently, economically, and safely, avoiding interruption to the normal operation of observed structure and requiring no artificial loadings (Brownjohn *et al.* 2010). Therefore, only ambient excitation (i.e., wind and traffic) is used to identify dynamic modal parameters by measuring vibration responses. Moreover,

dynamic modal parameters are beneficial for: (1) finite element model updating (Zeng & Kim 2020); (2) vibration-based damage detection (Honfi *et al.* 2020); (3) real-time alarm systems for bridges and buildings (Chen *et al.* 2011). It can be incorporated with maintenance systems to observe variations of modal parameters over time and track abnormality to prevent disastrous failure at an early stage. However, it requires a huge amount of recorded data and data analysis in a short amount of time. Therefore, it is labor-intensive, even impractical, to process massive measured data and identify modal parameters by manual intervention and engineers' experience. Much user intervention on a large amount of vibration data can be an obstacle in a real application. For overcoming this, it is essential to have an automated evaluation of structural conditions in almost real-time.

In the last decades, there are many researches focusing on the development of automated OMA techniques. Generally, OMA techniques can be categories into two groups: direct methods and indirect methods. Direct methods directly measure vibration responses of observed structures to identify modal parameters, requiring networks of sensors installed on the target structures. Many system identification algorithms have been developed to be compatible with direct methods, such as stochastic subspace identification (SSI) (Cabboi *et al.* 2017), eigensystem realization algorithm (ERA) (Yang *et al.* 2019), poly-least squares complex frequency domain method (Poly-MAX) (Magalhães *et al.* 2009, May), fast Bayesian FFT method (Au 2011), etc. On the other hand, indirect methods are referred to as the primary use of mobile sensors installed inside in the passing-by vehicles to perform modal identification. No sensors are needed to be deployed on structures as the indirect methods. Recently, much progress has been made in terms of system identification using networks of mobile sensors. For example, mass-modeled vehicle scanning method for bridges (Yang *et al.* 2004, Yang & Sun 2020), hybrid sensors based method (Marulanda *et al.*

2017), extended structural identification using expectation optimization (STRIDEX) (Matarazzo & Pakzad 2018), matrix completion method (Sadeghi Eshkevari *et al.* 2020), crowdsensing based method (Matarazzo *et al.* 2020). A detailed literature review on direct and indirect methods can be found (Malekjafarian *et al.* 2015, Yang & Yang 2018).

The present work belongs to the class of direct methods. Vibration response is measured from sensors installed on investigated structures. Among various system identification algorithms in the direct methods, SSI has been widely applied to various structures to identify modal parameters. It offers accurate identification results and simple implementation, which are important attributes accounting for its popularity. In addition, due to SSI's explicit mathematical nature, SSI tends to be more suited for automated modal identification. However, the major challenge in SSI is spurious modes inevitably appear in outputs.

Commonly, spurious modes consist of pure mathematical (i.e., non-physical) and noise modes (Reynders *et al.* 2008). The most common strategy to deal with this challenge is to construct a stabilization diagram, a plot of model order vs. frequency for an extensive range of model order. In the stabilization diagram, physical modes are referred to as those poles that cross most of the model orders consistently. Therefore, physical modes should be graphically recognized and homogeneously distributed along vertical alignments in a stabilization diagram (Cabboi *et al.* 2017). On the contrary, spurious modes appear in the stabilization diagram in a scattered way. Spurious modes are eliminated in a manual analysis depending on empirical discovery and engineers' judgment, which is subjective, time-consuming, and leads to incorrect modal identification. For addressing this issue, most strategies are proposed in the literature to interpret stabilization diagram and remove spurious modes automatically. Among these methods, the process can be divided into three steps:

(1) Step 1: Apply the modal validation criteria to eliminate as many spurious modes as possible in the stabilization diagram

(2) Step 2: Group modes with similar characteristics, i.e., frequencies, damping ratios, and mode shapes by clustering strategies

(3) Step 3: Detect outliers in each cluster to improve the accuracy of modal parameters and select representative of each cluster

Over the last decades, several methods aiming at minimizing human involvement in the interpretation of the stabilization diagram have been developed. For example, in step one, many modal validation criteria are proposed to detect spurious modes in the stabilization diagram. These criteria can be divided into hard criteria, which yield a binary answer, and soft criteria, which yield a certain range of values. Reynders et al. (2012) tthoroughly reviewed and summarized hard and soft criteria. However, conventional modal validation criteria limitedly remove a certain number of spurious modes; many spurious modes remain in the stabilization diagram that affects parameter estimates' accuracy and imposes a computational burden to the following step (clustering process).

In step 2, various clustering strategies are widely employed in group modes with similar characteristics. For example, hierarchical clustering has been extensively applied by many researchers and is considered the most natural approach (Reynders *et al.* 2012, de Almeida Cardoso *et al.* 2017, Sarlo *et al.* 2018). Hierarchical clustering has a significant advantage of allowing a good selection of physical clusters. However, the main drawbacks include a user-defined tree cutoff distance, which can be considered a significant human intervention, resulting in computationally demanding. Furthermore, the hierarchical clustering is sensitive to outliers. Another popular clustering strategy is partitioning methods, often referred to as K-means clustering (Neu *et al.* 2017, Yang *et al.* 2020). K-means clustering has the benefit of being fast

processing. However, the number of clusters has to be predefined, and it is sensitive to cluster seeds (initial centroid). By merging the benefits of hierarchical clustering and K-means clustering and overcome some of their limitations, self-adaptive clustering is recently proposed (Cabboi *et al.* 2017, Fan *et al.* 2019). The self-adaptive clustering has outstanding features: 1) simple implementation; 2) fast computation; 3) No need for the number of clusters; 4) Clustering threshold is iteratively trained during the clustering process.

While it still starts with a user-defined clustering threshold, which requires some level of human intervention. Some methods are proposed to automatically calculate clustering threshold based on statistical properties, i.e., mean and standard derivation or median, of the distance between two closed poles in the stabilization diagram (Magalhães *et al.* 2009, May, Reynders *et al.* 2012, Sun *et al.* 2017, Yang *et al.* 2019). However, these methods do not consider uncertainty on modal parameters and inaccuracy of mode shapes. In practice, modal parameters' uncertainty is inevitable due to modeling error and measurement noise; it can be a more reasonable approach to consider uncertainty when calculating the clustering threshold. Also, measurement on mode shapes is less accurate than that on frequencies. Thus, a weighting factor can reduce the inaccuracy of mode shape difference on threshold calculation (Boroschek & Bilbao 2019).

In step 3, some outliers are undesirably involved in identified physical clusters; this phenomenon is pronounced in a damping ratio with a scattered nature. Most outlier detection techniques need to define limit bounds, such as box-plot outlier detection (Yang *et al.* 2019, Wu *et al.* 2020). A bound-free outlier detection method is needed to improve the accuracy of parameter estimations. In summary, challenges to the current automated interpretation of the stabilization diagram are listed as follows:

1. Conventional modal validation criteria are not efficient enough, causing computational cost in the clustering process.
2. The clustering threshold and distance calculation in the clustering process does not consider the uncertainty of parameters and the weighting factor.
3. Uncertainties on identified modal parameters and physical clusters are unavailable.
4. Outlier detection requires to define limit bounds.

This paper attempts to address the aforementioned challenges. A two-stage framework for automated operational modal identification is proposed. Figure 1 shows the flowchart of the proposed framework: (1) modal analysis using covariance-driven reference-based SSI (SSI-cov/ref); (2) two-stage automated interpretation of stabilization diagram. SSI-cov/ref is adopted to perform modal analysis and construct a stabilization diagram. Subsequently, a two-stage automated analysis for the stabilization diagram is carried out. At the pre-processing stage, besides applying conventional modal validation criteria, such as damping ratio check and modal complexity check, to eliminate spurious modes, a new supplementary criterion: uncertainty criterion, which is applied for further removal of spurious modes. At the clustering stage, a novel threshold calculation, which incorporates the uncertainty of modal parameters and weighting factor, is proposed. An improved self-adaptive clustering with new distance calculation is then employed to group modes with similar characteristics and identify physical clusters. Finally, robust outlier detection is implemented to exclude outliers. The average of each cluster's elements is chosen as representative frequency, damping ratio, and mode shape.

Two field tests were used to evaluate the proposed approach. The first field example is the Dowling Hall footbridge located at Turfs University, a two-span steel frame bridge with 144 ft (44 m) long and 12 ft (2.7 m) width having a reinforced concrete deck. The second field test example

is the Z24 bridge located in Switzerland, a post-tensioned concrete bridge with a main span of 100 ft (30 m) and two sides span of 46 ft (14 m), which is considered as a benchmark in the research community.

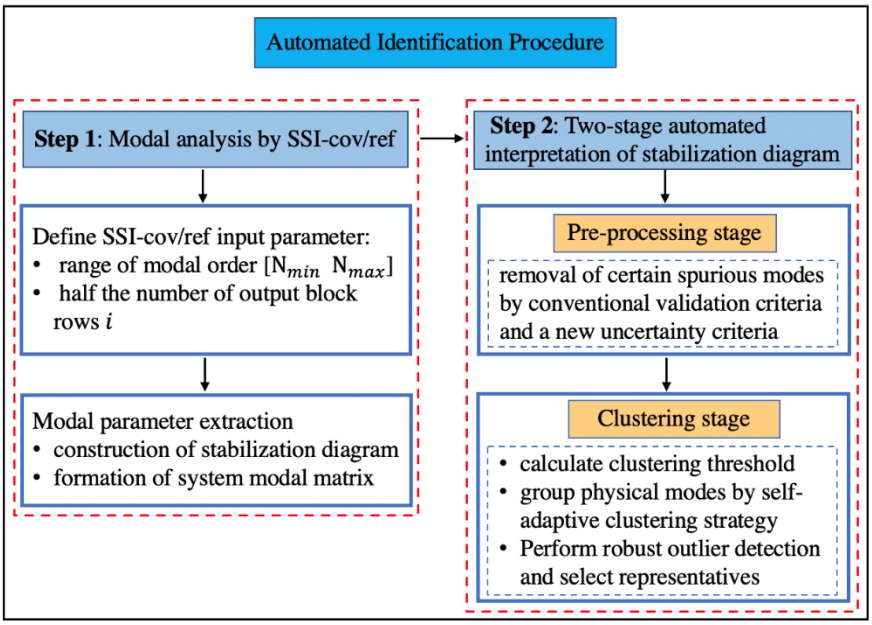


Figure 1. A framework of the proposed automated approach

This paper is organized as follows: In Section 2, the background of SSI-cov/ref is briefly introduced. In Section 3, a two-stage approach for proposed automated modal identification is presented. In Section 4, the capability of the proposed approach is validated by two field tests along with the modal tracking results. Finally, the conclusion is presented in Section 5.

2. Background of SSI

SSI has been extensively spread over the field of OMA during the past few decades accounting for its quick implementation and high accuracy (Peeters 2000). In this paper, covariance-driven reference-based SSI (SSI-cov/ref) is employed to reduce the dimensions of the output matrix and computational cost. Theoretical fundamentals in detail of SSI-cov/ref is fully described in the literature (Peeters & De Roeck 1999). Briefly, SSI-cov/ref is developed based on

assuming a linear and stationary N degree of freedoms (DOFs) system with a dynamic motion characterized by the discrete-time state-space equation:

$$\begin{aligned} x_{k+1} &= Ax_k + \omega_k \\ y_k &= Cx_k + v_k \end{aligned} \tag{1}$$

where subscript k denotes time step; $A \in \mathcal{R}^{n\times n}$ denotes system state matrix with $(n = 2N)$; $C \in \mathcal{R}^{l\times n}$ is an output matrix, l is defined as the number of measured signals; $x_k \in \mathcal{R}^{n\times 1}$ and $y_k \in \mathcal{R}^{l\times 1}$ are discrete-time state vector and measured response vector, respectively; $\omega_k \in \mathcal{R}^{n\times 1}$ is a process white noise vector. $v_k \in \mathcal{R}^{l\times 1}$ is the measurement of the white noise vector. It has been shown in (Peeters & De Roeck 1999) that modal parameters can be identified from system matrices A and C.

Two main SSI preparation parameters significantly affect the accuracy of identification results: (1) model order; (2) time lag, i. Unfortunately, the value of model order and i, which yield the best identification results are never known (Ubertini *et al.* 2013, Fan *et al.* 2019). In practice, it is necessary to over-specify model order to cover weakly-excited modes, but spurious modes increase with increasing model order. These spurious modes must be singled out in the subsequent procedure. On the other hand, the value of i determines the size of the response covariance function. The smaller i may fail to identify the fundamental mode, but the larger value of i yields many spurious modes and increases computational time. The value of i may be chosen at least estimated value as follows (Fan *et al.* 2019):

$$i \geq T_i/t \tag{2}$$

where T_i denotes fundamental period, (unit: second); t denotes sampling interval, (unit: second).

3. A two-stage automated modal identification

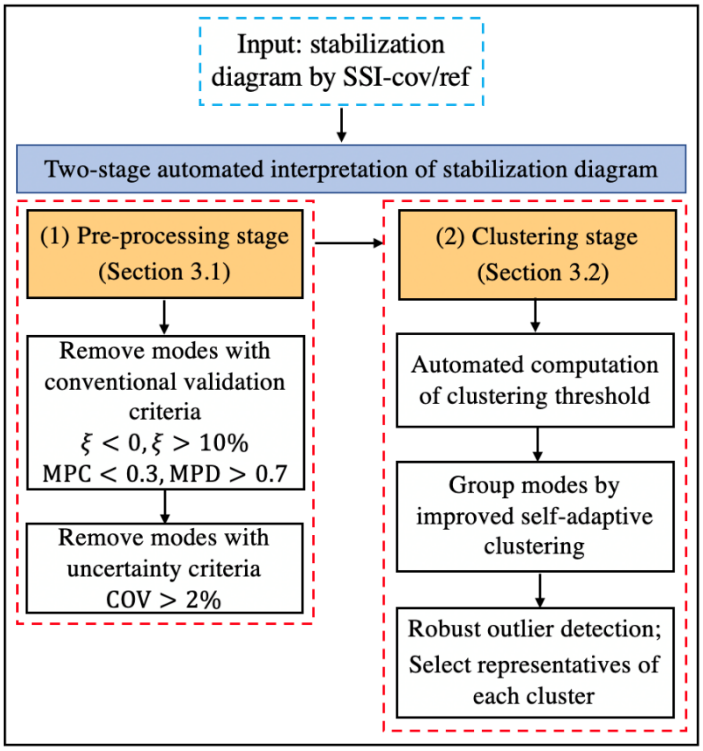


Figure 2. A flowchart of the proposed two-stage approach of automated modal identification

In this section, a two-stage framework for automated modal identification is proposed. Section 3.1 presented the pre-processing stage for automated identification, including conventional modal validation criteria (Section 3.1.1) and a new additional uncertainty criterion (Section 3.1.2). Subsequently, the clustering stage is introduced in Section 3.2. A newly proposed threshold calculation for clustering is provided in Section 3.2.1. An improved self-adaptive clustering is then employed to determine physical clusters in Section 3.2.2. Finally, robust outlier detection is performed to improve the accuracy of modal parameter estimates in Section 3.2.3. The pseudocode of the proposed automated framework is also provided in Appendix A. The flowchart of the entire automated process in detail is presented in Figure 2.

3.1. The pre-processing stage

A full stabilization diagram looks very complex and contains the spurious modes. At the pre-processing stage, different validation criteria are utilized to initially screen spurious modes, clarifying the stabilization diagram and consequently accelerating automated interpretation of the stabilization diagram.

3.1.1. Modal validation criteria

First of all, for civil engineering structures, a negative or high damping ratio hardly appears in practice. Therefore, a damping ratio with less than 0 and higher than 10% is discarded (Cabboi *et al.* 2017, de Almeida Cardoso *et al.* 2017, Fan *et al.* 2019). On the other hand, two popular modal validation criteria are used to measure the complexity of mode shape vectors, namely, modal phase collinearity (MPC) and mean phase deviation (MPD). These two indicators have been utilized by various researchers to distinguish physical modes from spurious modes (Reynders *et al.* 2012, de Almeida Cardoso *et al.* 2017, Neu *et al.* 2017, Sun *et al.* 2017). The real (Re) and imaginary (Im) part of mode shapes display a linear correlation, which can be assessed by the MPC indicator. The value of MPC for the tth mode shape, ϕ_t, is expressed as (Reynders *et al.* 2012):

$$\mathrm{MPC}(\phi_t) = \frac{\left\|\mathrm{Re}(\tilde{\phi}_t)\right\|^2 + \frac{1}{\alpha}\mathrm{Re}(\tilde{\phi}_t^T)\mathrm{Im}(\tilde{\phi}_t^T)(2(\alpha^2+1)\sin^2\gamma - 1)}{\left\|\mathrm{Re}(\tilde{\phi}_t)\right\|^2 + \left\|\mathrm{Im}(\tilde{\phi}_t)\right\|^2} \tag{3}$$

The k^{th}component of $\tilde{\phi}_t$ is given: $\tilde{\phi}_{t,k} = \phi_{t,k} - \frac{\Sigma_{k=1}^{L}\phi_{t,k}}{L}$, L is the number of components in ϕ_t.

$$\alpha = \frac{\left\|\mathrm{Im}(\tilde{\phi}_t)\right\|^2 - \left\|\mathrm{Re}(\tilde{\phi}_t)\right\|^2}{2\mathrm{Re}(\tilde{\phi}_t^T)\mathrm{Im}(\tilde{\phi}_t^T)} \tag{4}$$

$$\gamma = \arctan\left(|\alpha| + \mathrm{sign}(\alpha)\sqrt{1+\alpha^2}\right) \tag{5}$$

MPC values are dimensionless; they lie within the range of 0 and 1. MPC value closer to 1 indicates that mode shape, ϕ_t, is more collinear and 'monophase,' which is usually regarded as a physical mode.

With regard to MPD, it represents the phase degree of each identified mode shape vector. The value of MPD/90 lies between 0 and 1. A smaller quantity of MPD implies that mode shape vector is more likely to be physical. A detailed discussion can be found in (Reynders *et al.* 2012). For the tth mode shape, ϕ_t, the mean phase (MP) is defined as:

$$\mathrm{MP}(\phi_t) = \arg_\theta \min \left(\frac{\|\mathrm{Im}(\phi_t) - \tan\theta\, \mathrm{Re}(\phi_t)\|^2}{1 + \tan\theta}\right) \tag{6}$$

where θ is phase angle in degree, Equation (6) can be solved by the least square as:

$$\mathrm{MP}(\phi_t) = \arctan\left(\frac{-V_{12}}{V_{22}}\right), USV^T = [\mathrm{Re}(\phi_t)\ \mathrm{Im}(\phi_t)] \tag{7}$$

where USV^T is singular value decomposition, V_{12} and V_{22} denotes elements (1,2) and (2,2) of V matrix, respectively. Then, MPD can be determined as:

$$\mathrm{MPD}(\phi_t) = \frac{\Sigma_{k=1}^{L} \omega_k \arccos\left(\mathrm{Re}(\phi_{t,k}) V_{22} - \mathrm{Im}(\phi_{t,k}) V_{12}\right)}{\Sigma_{k=1}^{N} \omega_k} \tag{8}$$

where ω_k is a weighting factor that equals to the k^{th}component of the tth mode shape, ϕ_t.

The selection of threshold values of MPC and MPD depends on measurement conditions and dynamic vibration properties. For a structure with clear linear behavior and high signal-to-noise ratio, the threshold of MPC and MPD can be conservatively chosen as 0.7 and 0.3, which implies that modes whose MPC are less than 0.7 and MPD exceed 0.3 are regarded as spurious modes. Conversely, the threshold of MPC and MPD are chosen as 0.3 and 0.7 in the case of structures with complex behavior (Cabboi *et al.* 2017, Fan *et al.* 2019). Two representative field tests with complex measurement conditions are used to validate the methods. Thus, 0.3 and 0.7 are selected as a threshold for MPC and MPD, respectively, in this study.

3.1.2. Uncertainty criterion

Although conventional modal validation criteria remove certain spurious modes, many spurious modes remain in the stabilization diagram resulting in slowing down the following process (herein clustering process). More effective validation criteria should be adopted to delete as many spurious modes as possible. This study employed supplementary uncertainty criteria at the pre-processing stage to further eliminate spurious modes.

Uncertainty on modal parameters by SSI mainly arise from five sources: (1) finite number of data sample; (2) unmeasured excitation and measurement noise modeled as white noise; (3) assumption of linear and stationary behavior ; (4) imperfect filter of data; (5) incorrect choice of model order (Reynders *et al.* 2008). Reynder et al. (2008) initially developed the uncertainty computation based on the propagation of first-order perturbation from measured data to modal parameters. Also, some validation and application are summarized in Reynders et al. (2016). Later, Döhler et al. (2013) significantly improved the computational efficiency of uncertainty at multiple model orders, which has been applied in various structures (Döhler *et al.* 2013, Mellinger *et al.* 2016). Uncertainty quantification can provide information to measure the accuracy of identified modal parameters. It is only valid when the uncertainty of physical modes is smaller than those of spurious modes. Based on this information, coefficient of variation (COV) (standard derivation/mean) with respect to frequency may be used to distinguish physical modes from spurious modes (Döhler & Mevel 2013, Nord *et al.* 2019). Some research has introduced uncertainty features in the stabilization diagram (Döhler & Mevel 2013, Mellinger *et al.* 2016, Nord *et al.* 2019), but uncertainty criterion is not used or does not attribute to further automated modal procedure. General procedures of uncertainty computation are summarized as follows:

- Input parameters: the number of block rows in Hankel matrix, q; the amount of data blocks,n_b; the range of model order, (n_{min}, n_{max});
- Compute Hankel matrix, H, system state matrix, A and output matrix, C in Equation (1), as well as observability matrix, O, based on SSI-cov/ref, then compute transform matrix, T
- Compute covariance and sensitivity of subspace matrix from SSI-cov/ref, given by $\hat{\Sigma}_{H^{cov}}$ and $J_{O,H}$, respectively
- Compute sensitivity and covariance of system state matrix, A, and output matrix, C from SSI-cov/ref, given by $J_{A,O}, J_{C,O}$ and $\Sigma_{A,C}$, respectively
- For each mode i at successive modal order, compute sensitivity matrix: $J_{f_i,A}, J_{\xi_i,A}$ and $J_{\phi_i,A}$. Finally, compute covariance of modal parameters, frequency, f_i, damping ratio, ξ_i; mode shape, ϕ_i: $\mathrm{cov}\left(\begin{bmatrix} f_i \\ \xi_i \end{bmatrix}, \begin{bmatrix} f_j \\ \xi_j \end{bmatrix}\right)$ and $\mathrm{cov}\left(\begin{bmatrix} \mathrm{Re}(\phi_i) \\ \mathrm{Im}(\phi_i) \end{bmatrix}, \begin{bmatrix} \mathrm{Re}(\phi_j) \\ \mathrm{Im}(\phi_j) \end{bmatrix}\right)$

Comprehensive derivation for uncertainty estimation can be found in Döhler and Mervel (2013). When the COV in the frequency is chosen as a threshold (herein, 2%), modes with the COV of frequency larger than the threshold will be discarded.

3.2. Clustering stage

The clustering stage is sequentially performed to assemble modes based on similarities in modal parameters in this section. A novel method is proposed to calculate the clustering threshold; an improved self-adaptive clustering is then used to identify physical clusters. Finally, robust outlier detection is implemented, and each representative of modal parameters is determined.

3.2.1. Automated computation of clustering threshold

Typically, two kinds of thresholds are usually adopted for clustering: (1) static threshold; (2) automatically computed threshold. A static threshold relies on the engineers' judgment. Also, during long-term health monitoring, a well-defined static threshold may be suitable for some initial datasets; however, there is no guarantee that the static threshold will be keeping appropriate for all datasets. This is more challenging in the case of handling massive datasets. In this study, a novel method is proposed to calculate the clustering threshold based on possible physical modes at the pre-processing stage. First, the mutual distance between the two modes is defined as:

$$\text{distance} = \frac{\left|F_i - F_j\right|}{\max\left(F_i, F_j\right)} + \omega\left(1 - \frac{(\Phi_i^T \Phi_j)^2}{(\Phi_i^T \Phi_i)(\Phi_j^T \Phi_j)}\right) \tag{9}$$

where $F_i = f_i + 2\sigma_{f_i}$, $F_j = f_j + 2\sigma_{f_j}$, f_i and f_j are i^{th} and j^{th} identified frequency at a pre-processing stage, respectively; σ_{f_i} and σ_{f_j} are corresponding standard derivation, respectively; $\Phi_i = \phi_i + 2\sigma_{\phi_i}$, $\Phi_j = \phi_j + 2\sigma_{\phi_j}$, ϕ_i and ϕ_j are i^{th} and j^{th} mode shapes at a pre-processing stage, respectively; σ_{ϕ_i} and σ_{ϕ_j} are corresponding standard derivation, respectively. ω is a weighting factor of mode shape difference, $\omega = \frac{(\sigma_{\phi_i} + \sigma_{\phi_j})}{2}$.

Equation (9) does not consider the damping ratio difference because it is difficult to accurately measure the damping ratio in practice. In addition, there is a high probability of two different modes with a similar damping ratio. A weighting factor, ω, in Equation (9) represents different participation for frequency difference and mode shape difference. Generally, mode shape is measured with limited sensors, yielding missing components of mode shape; frequency is usually measured with an accurate level. Therefore, the use of ω can reduce the effect of measurement inaccuracy of mode shapes on distance calculation (Boroschek & Bilbao 2019). An uncertainty quantification using standard derivation is used to form a weighting matrix for Finite

Element Model Updating (Lam *et al.* 2014, Yang & Lam 2018). Similarly, this work adopted the average of the standard derivation of mode shapes to define ω. Furthermore, as uncertainty on modal parameters is inevitable in practice, it is more reasonable to incorporate uncertainty in distance calculation. Here, two standard derivations are considered in Equation (9).

At the next model order, the mutual distance between one mode and all other modes is computed by Equation (9), then the minimum distance is determined. Assuming n modes have been identified at the pre-processing stage, each mode has its minimum mutual distance with forming a minimum distance vector, $V = (d_{min}^1, d_{min}^2, d_{min}^3 \cdots d_{min}^n)$, ($n$ denotes the number of modes, d_{min} denotes the minimum distance between one mode and all other modes). Finally, the sum of mean and two standard derivations of V are used to compute the clustering threshold, $\bar{d}$ (Reynders *et al.* 2012, Sarlo *et al.* 2018):

$$\bar{d} = \bar{\mu} + 2\bar{\sigma} \tag{10}$$

Typically, modal features are usually assumed to follow Gaussian normal distribution, such as frequency, damping ratio and mode shape (MAC value) (Au 2011). In this study, Equation (9) defines modal distance which is the sum of frequency difference and mode shape difference between two modes. Therefore, modal distance turns out to be Gaussian normal distribution. Two standard derivations in Equation (10) guarantee the distance between two modes should be captured within a 95% confidence interval with the assumption of Gaussian distribution.

It is worth mentioning that conventional threshold calculation for clustering does not consider the uncertainty of modal parameters and weighting factor. A newly proposed threshold calculation in Equation (9) incorporates both uncertainty and weighted factor, which is considered a contribution of this study to the advancement of automated OMA.

3.2.2. Mode clustering

Mode clustering starts with a calculated threshold in Section 3.2.1. to group individual physical modes with similar modal characteristics. This study adopts self-adaptive clustering (Cabboi *et al.* 2017) to accomplish automated process. But different from original Cabboi et al.'s work, a weighted distance with an uncertainty of modal parameters is proposed. The ith weighed distance at model order, n, is defined as:

$$d_{n,i} = \left(\frac{\left| \tilde{F}_z - F_{n,i} \right|}{\tilde{F}_z} \right) + c\left(1 - MAC(\tilde{\Phi}_z, \Phi_{n,i})\right) \tag{11}$$

where $\tilde{F}_z = \tilde{f}_z + 2\sigma_{\tilde{f}_z}$, $\tilde{\Phi}_z = \tilde{\phi}_z + 2\sigma_{\tilde{\phi}_z}$, $F_{n,i} = f_{n,i} + 2\sigma_{f_{n,i}}$, $\Phi_{n,i} = \phi_{n,i} + 2\sigma_{\phi_{n,i}}$. $\tilde{f}_z$ and $\tilde{\phi}_z$ are mean frequency and mean mode shape at the zth cluster, respectively; $\sigma_{\tilde{f}_z}$ and $\sigma_{\tilde{\phi}_z}$ are corresponding mean standard derivation at the zth cluster. $f_{n,i}$ and $\phi_{n,i}$ are the ith frequency, and mode shape at model order, n, respectively; $\sigma_{f_{n,i}}$ and $\sigma_{\phi_{n,i}}$ are corresponding standard derivation, respectively. MAC represents the modal assurance criteria (Allemang & Brown 1982). c is a weighting factor to reduce the effect of inaccurate mode shape on distance calculation ($c = \frac{\sigma_{\tilde{\phi}_z} + \sigma_{\phi_{n,i}}}{2}$).

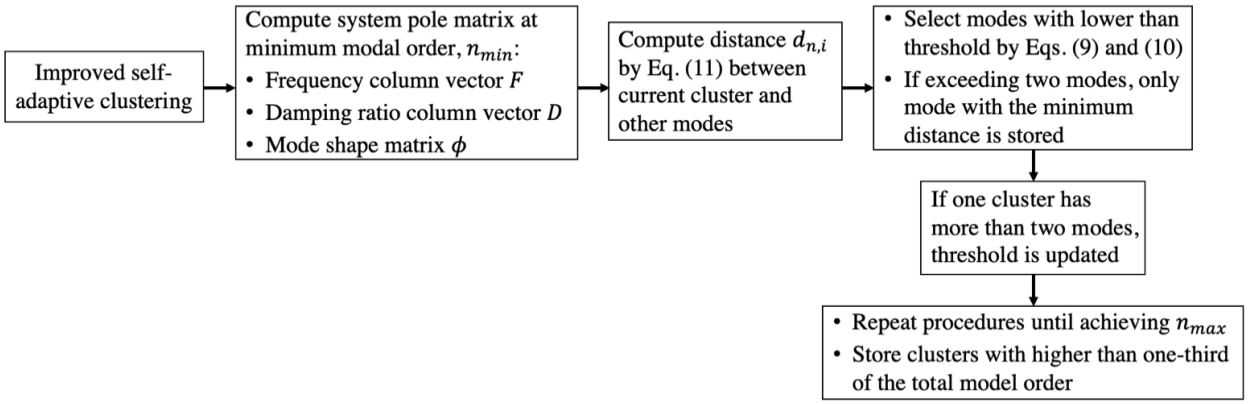


Figure 3. The flowchart of an improved self-adaptive clustering

Equations. (9) and (11) are similar; both consider the uncertainty of parameter estimates and the importance of mode shape difference. Figure 3 shows the flowchart of an improved clustering strategy. A more detailed introduction of self-adaptive clustering is referred to Cabboi et al. (2017).

3.2.3. Robust outlier detection

The number of physical poles in the stabilization diagram has trends with the increase of model order, exhibiting variability of modal estimates (Neu *et al.* 2017). The phenomenon more frequently appears in the damping ratio because the damping ratio has a high scattered nature. Outlier detection is applied to penalize undesirable modes in the final clusters for reducing identification variance from different measurements. In this study, robust outlier detection based on the minimum covariance determinant (MCD) is employed to identify outlying values from physical clusters as described in Hubert et al. (2017). A robust distance (RD) is defined as:

$$\mathrm{RD}(x) = \mathrm{d}(x, \hat{\mu}_{MCD}, \hat{\Sigma}_{MCD}) \tag{12}$$

where observation sample, x, is either frequency or damping ratio in a physical cluster in our case. $\hat{\mu}_{MCD}$ is the MCD estimates of location; $\hat{\Sigma}_{MCD}$ is the MCD covariance estimate. Explicit derivation and introduction can be found in Hubert et al. (2017).

A robust MCD estimator based on Equation (12) is very powerful to flag outliers, as RD in Equation (12). It is not sensitive to diagnostic tools' masking effect compared to statistical distance and Mahalanobis distance (Cerioli 2010). Also, MCD has a high resistance to outliers and are more robust and efficient (Hubert *et al.* 2017). Furthermore, MCD has the advantage of requiring no user-defined threshold, like a box-plot method that needs to define limit bounds (Sarlo & Tarazaga 2019). Robust outlier detection in this work can be done by the function ‘robustcov’ in MATLAB.

After outlier removal, the average frequency, damping ratio, and mode shape in each physical cluster are taken as a representative. For evaluating the quality of each identified cluster, uncertainty on the zth physical clusters is quantified by Euclidean norm of uncertainty on modal parameters:

$$\sigma_z = \sqrt{(\tilde{\sigma}_{f,z}^2 + \tilde{\sigma}_{\phi,z}^2 + \tilde{\sigma}_{\xi,z}^2)} \tag{13}$$

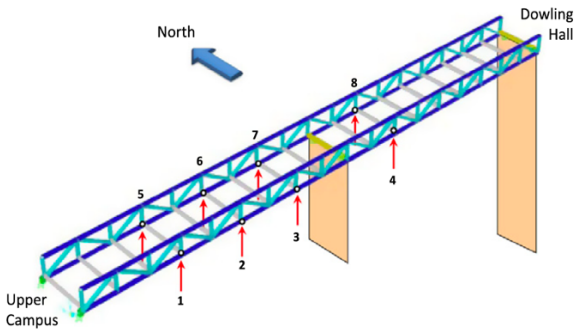
where σ_z is the standard derivation of the zth clusters; $\tilde{\sigma}_{f,z}$, $\tilde{\sigma}_{\phi,z}$, and $\tilde{\sigma}_{\xi,z}$ are the average values of standard derivations of all frequencies, damping ratios, and mode shapes in the zth clusters.

4. Application

In this section, the performance of the proposed approach is validated by two field tests on the bridge, namely, the Dowling Hall footbridge located at Turfs University in the U.S. and the Z24 bridge benchmark located in Switzerland. The data are open sources, and many researchers used these data to test the algorithms in the research community.

4.1. Application 1: Steel frame footbridge

Dowling Hall footbridge is located at Tufts University, as shown in Figure 4 (a). The bridge is a two-span steel frame bridge, 144 ft (44 m) long and 12 ft (2.7 m) wide, and has a reinforced concrete deck. A continuous health monitoring was designed and performed on Dowling Hall footbridge from January 2010 to May 2010. The layout of eight accelerometers is shown in Figure 4 (b). More details of Dowling Hall footbridge's information can be found in Moser and Moaveni (2013). In this study, the first six modal characteristics are used as baseline results that are obtained from the literature (Moser & Moaveni 2011) to evaluate the performance of the proposed approach.



(a) (b)

Figure 4. Description of Application 1: (a) Dowling Hall footbridge; (b) Sensor layout (Moser & Moaveni 2011)

The acceleration data used in this study are obtained from vertical measurement under ambient excitation collected in the first week at 1:00 P.M. on January 7th, 2010. The frequency range of interest is 0-30Hz. The sampling frequency is 128Hz. Preparation parameters for SSI-cov/ref in this application are: $i = 60$, model order $n = 40{\sim}150$, reference sensor $= (1,2,3,4,5,6,7,8)$.

4.1.1. Identification results

The proposed approach described in Section 3 is utilized to analyze measured data. Figure 5 (a)-(c) show modal identification results at the pre-processing stage. The singular value spectrum (appeared in the curves in Figure 5) is plotted below the stabilization diagram. The standard derivation ($\pm\sigma$) uncertainty bounds of the frequency are shown as horizontal bars. Figure 5 (a) displays all possible physical modes remaining in the stabilization diagram after applying conventional validation criteria, e.g., damping ratio check and modal complexity check. Figure 5 (b) shows the stabilization diagram filtered by a supplementary uncertainty criterion. It is observed only using conventional validation criteria, the stabilization diagram still looks busy, including some scattered poles, which are spurious modes. However, uncertainty criterion can eliminate as many spurious modes as possible compared to conventional validation criteria, which will speed

up later automated processes. The pre-processing stage's identification results demonstrate that the uncertainty criterion is more effective than conventional validation criteria.

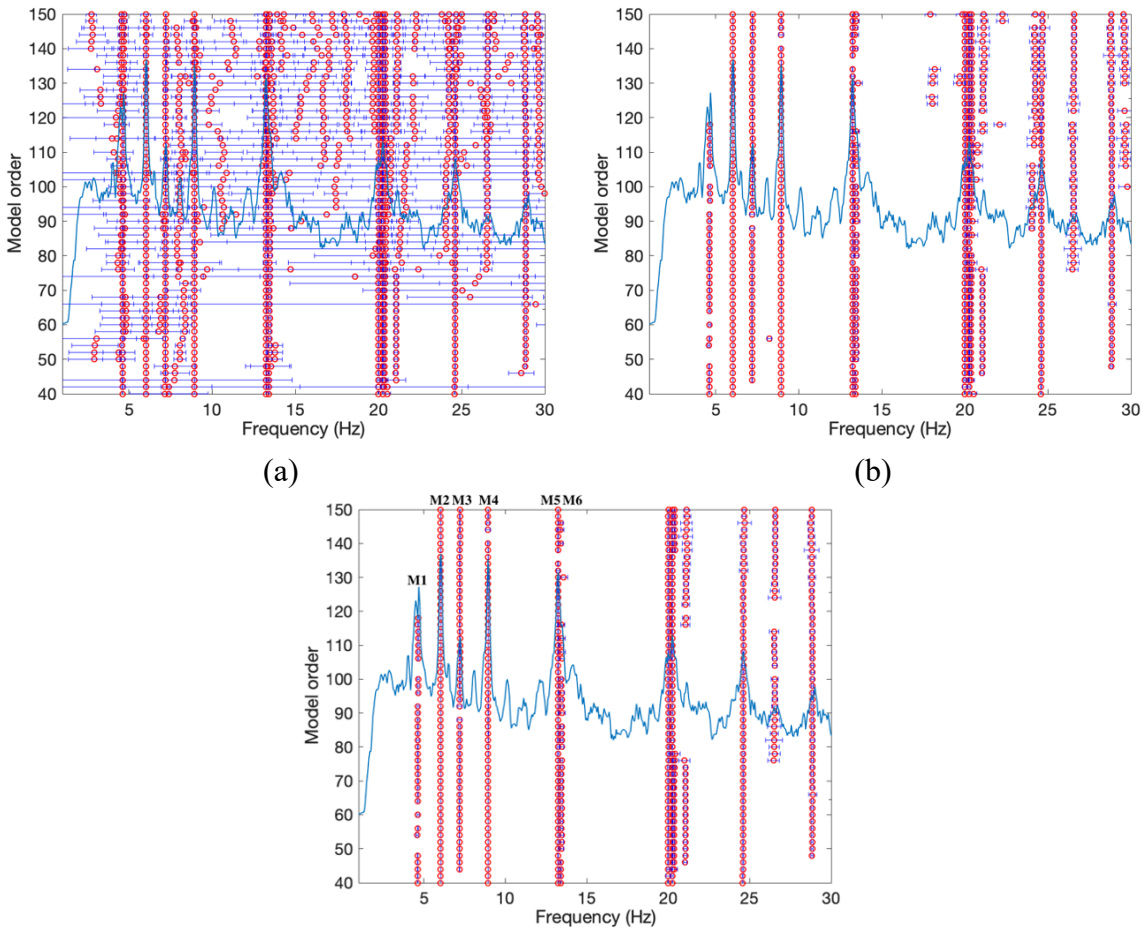
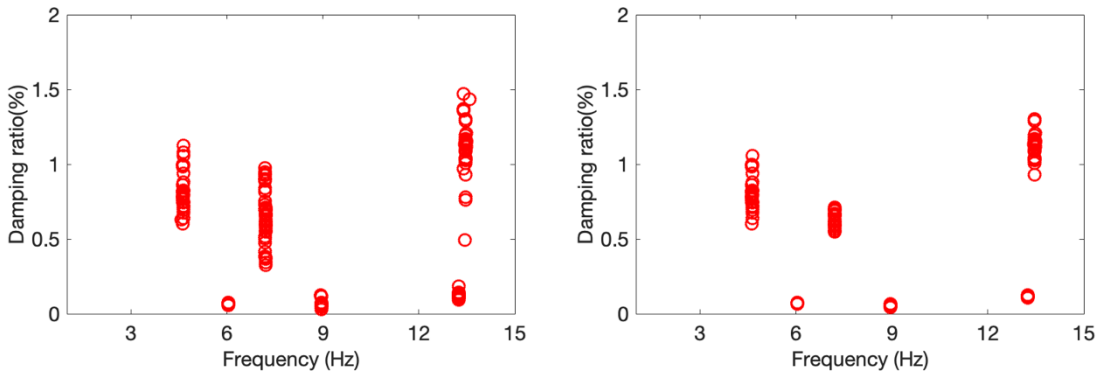


Figure 5. Identification results: (a) after conventional validation criteria; (b) after uncertainty criterion (c) after improved self-adaptive clustering

The clustering stage then starts with a calculated clustering threshold using Equations. (9) and (10) based on the remaining modes in Figure 5 (b). The proposed method's calculated threshold in this example is 0.022, while without the weighting factor, ω, it is 0.0488. It implies the stricter threshold by Equations. (9) and (10) that allows removing more spurious modes with keeping physical modes. Furthermore, the updated threshold by improved clustering with Equation (11) is 0. 0086. Still, the original work (distance calculation without weighting factor) gives the updated threshold as 0.0304, indicating that the weighted distance tends to give a smaller value of the

updated threshold. The identified modes are more consistent and stable. It may be attributed to the use of ω can improve the accuracy of measured mode shapes by distance calculation. Figure 5 (c) shows the modal identification results after performing the improved self-adaptive clustering. It is observed that clustering procedures remove spurious modes, the stabilization diagram is clarified with only remaining stable modes (vertical alignments). The first six modes in the reported work (Moaveni & Behmanesh 2012) are used as a baseline for comparison, marked as M_1 to M_6 (A total of six clusters) in Figure 6. It is noted that M_5 and M_6 are closely spaced modes, which are a common challenge in the OMA. The robust outlier detection is to remove outlying frequencies and damping ratios. As seen in Figure 6 (b), damping ratios are tighter and more consistent.



(a) (b)

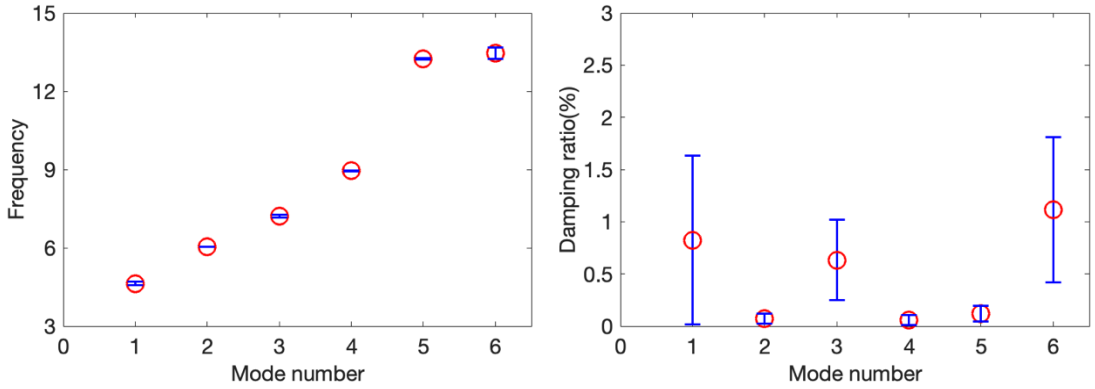
Figure 6. Damping ratio vs frequency: (a) before outlier detection; (b) after outlier detection

Table 1 presents identified frequencies and damping ratios along with the baseline data. Identified frequencies in this work agree well with those in the literature; the maximum relative difference (2.05%) is observed in the third mode. While larger variation is found in terms of damping ratio. It is because two tests were performed at a different time. Moaveni and Behmanesh (2012) reported baseline data, measured on April 4, 2009. In this study, measured data was collected on January 7, 2010. When considered the effect of environmental variables such as temperature, it is not surprising to have these differences. The frequency is less sensitive to environmental effects than the damping ratio.

Table 1. Identification results

Modes	Frequency (Hz)		Damping ratio (%)	
	Baseline*	The proposed approach	Baseline*	The proposed approach
1st (M_1)	4.63	4.63	1.0	0.8
2nd (M_2)	6.07	6.04	0.6	0.1
3rd (M_3)	7.07	7.21	0.7	0.6
4th (M_4)	8.90	8.95	0.3	0.1
5th (M_5)	13.13	13.24	0.8	0.1
6th (M_6)	13.56	13.46	1.1	1.1

Note: *: Moaveni & Behmanesh, 2012;



(a) (b)

Figure 7. error bar of frequency (left, $\pm 2\sigma$) and damping ratio (right, $\pm \sigma$)

Table 2. Uncertainty of physical clusters

No. cluster	1	2	3	4	5	6
STD (%)	0.184	0.001	0.430	0.010	0.119	2.712

Note: STD denotes standard derivation

The uncertainty on modal parameters and physical clusters are also investigated in this example. The frequency and damping ratio in each mode are plotted as an open circle overlapping the standard derivation (σ) error bar in Figure 7. And the uncertainty of modal frequencies is much smaller than those of damping ratios. It is often more difficult to accurately measure the damping ratio in practice. The uncertainty of identified physical clusters is also quantified by Equation (13) and shown in Table 2. And the uncertainty of the sixth cluster is much larger than those of others, suggesting it is more challenging to identify the sixth cluster because this cluster contains weakly-excited and closely spaced modes.

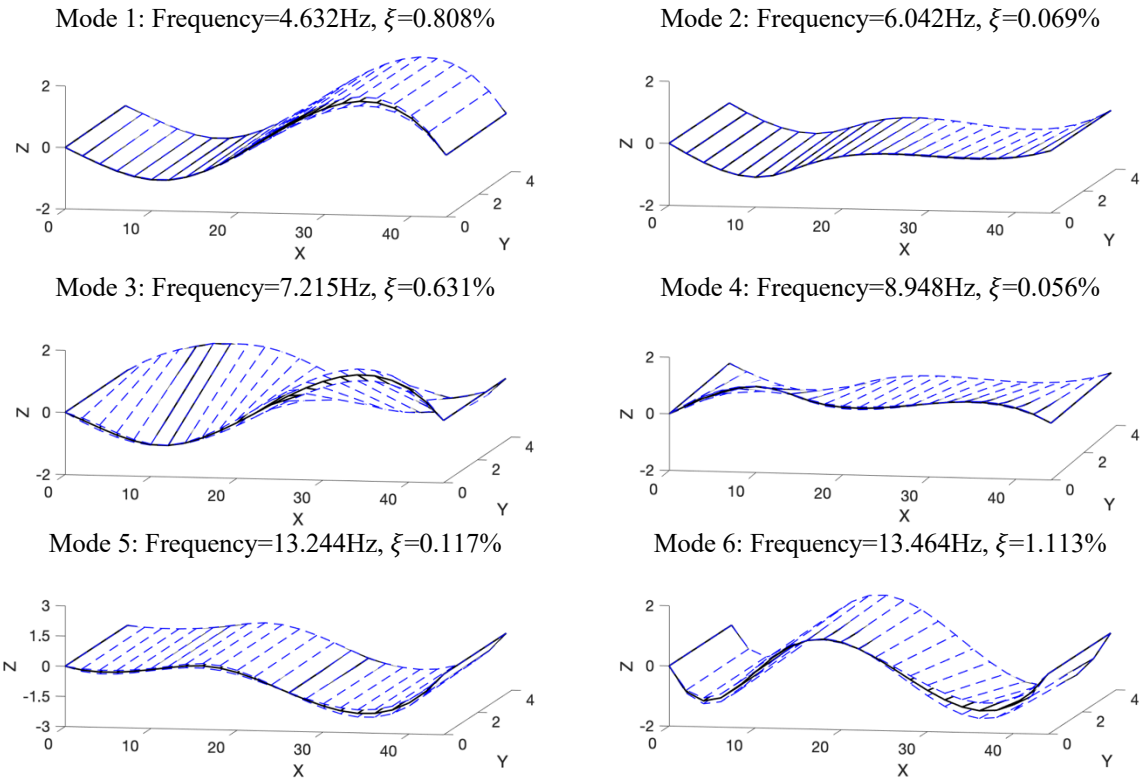


Figure 8. Identified mode shapes by the proposed approach and $\pm 2\sigma$ uncertainty bounds (blue dashed lines)

Overall, the proposed approach successfully identifies six modes under ambient vibration, as shown in Figure 8. The first six global mode shapes with corresponding uncertainties are presented; $\pm 2\sigma$ uncertainty bounds are plotted as blue dashed lines. Identified mode shapes have good agreement with those identified in the reported work (Moaveni & Behmanesh 2012). Modes 3 and 4 are bending-torsional mode with evident rotational motion, while only vertical deformation is found on other modes. In addition, uncertainty bounds for all modes are narrow, which concludes that the identification of mode shapes is accurate.

For continuous SHM, it is crucial to track the change of modal parameters over time. In this example, the proposed approach is applied to modal tracking with measured data collected at every 1:00 P.M. from January 5th to February 28th, in 2010 (total 55 datasets). The same procedures as the former data analysis are applied for modal tracking. As shown in Figure 9, solid

black lines indicate frequency estimates, and grey areas cover ± 2 standard derivations. All six modes are identified and tracked for all datasets by the proposed approach. It is not surprising that frequencies varied over time, mainly because of environmental change and ambient excitation's randomness. The frequency at mode 6 has a relatively larger variation for two months, as this mode is not excited well and unstable to environmental change. The results illustrate the proposed approach can analyze massive datasets with minimum human intervention.

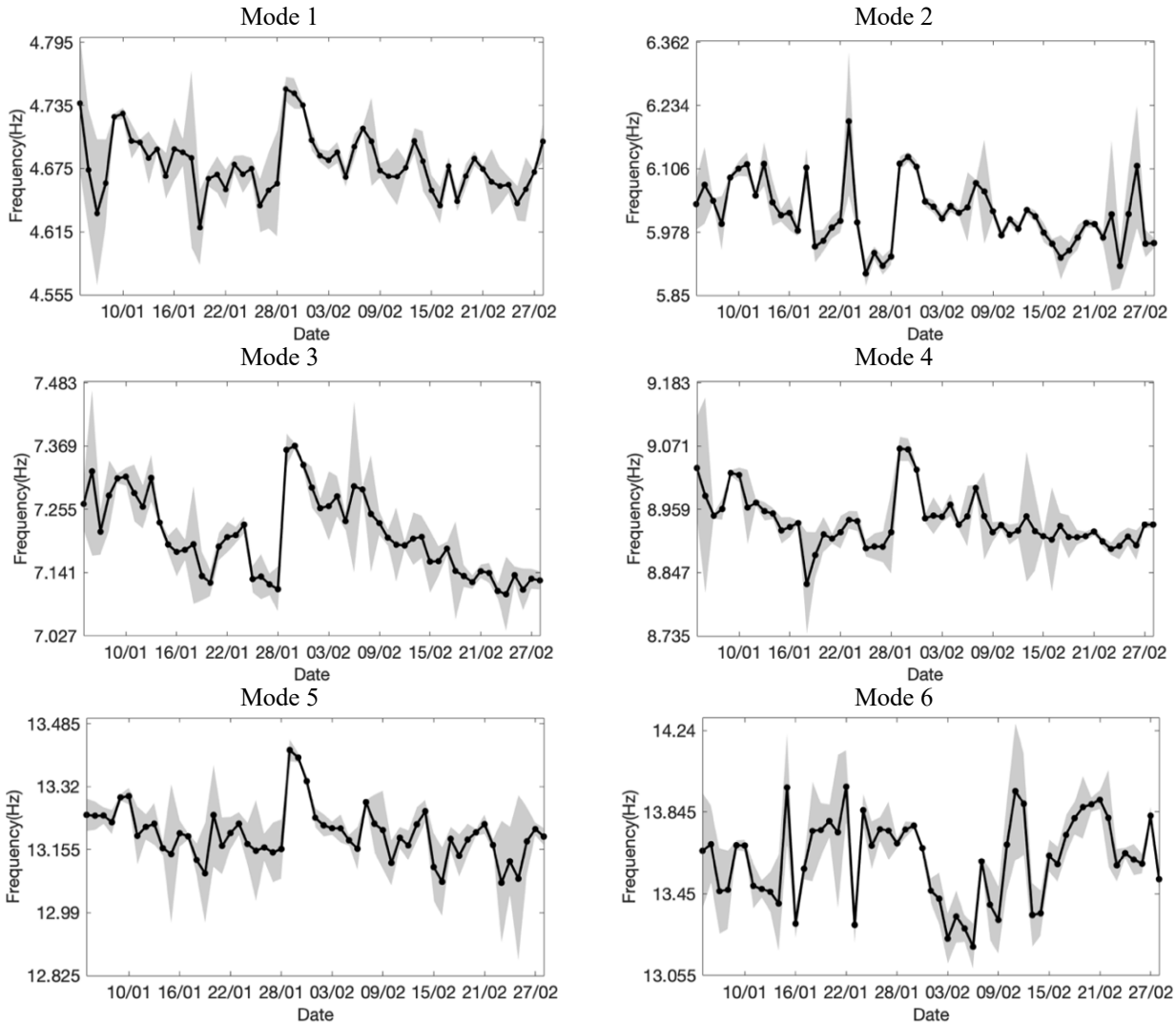
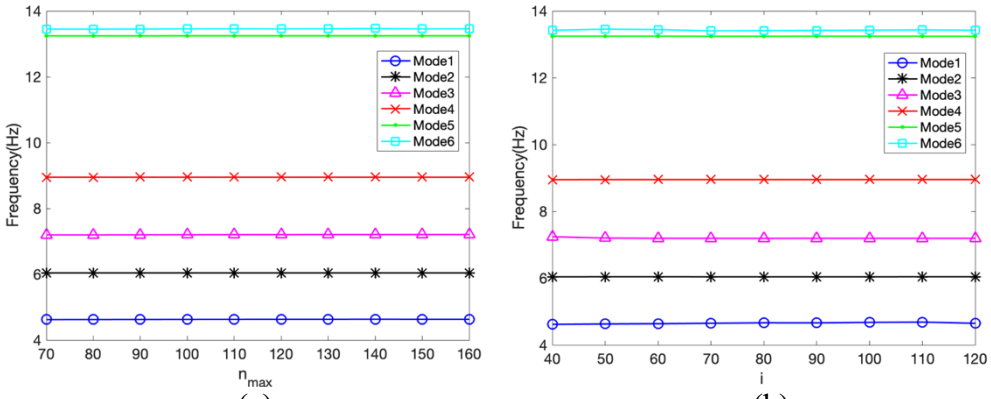


Figure 9. Identified frequency of modes 1-6 with two-month data. Black solid lines: frequency estimates; grey shaded areas: $\pm$two standard derivations

4.1.2. *Sensitivity analysis*



(a) (b)

Figure 10. Frequencies at different parameters: (a) model order range sensitivity (fixed $i = 60$); (b) time lag range sensitivity (fixed $n_{max} = 100$)

Two preparation parameters in SSI, e.g., maximum model order, n_{max}, and time lag, i, significantly affect identification results as described in Section 2. The influence of n_{max} and i is investigated to demonstrate the proposed approach is robust to their choice. n_{max} and i are varied from 70 to 160 and 30 to 120 in intervals of 10, respectively. As shown in Figure 10, identified frequencies are almost invariant to a different choice of n_{max} and i, suggesting the proposed approach is robust and not sensitive to these two preparation parameters. It is very difficult to identify the best n_{max} and i in practice (Ubertini *et al.* 2013, Neu *et al.* 2017). Thus, insensitivity to them allows to more conveniently perform automated OMA and continuous health monitoring.

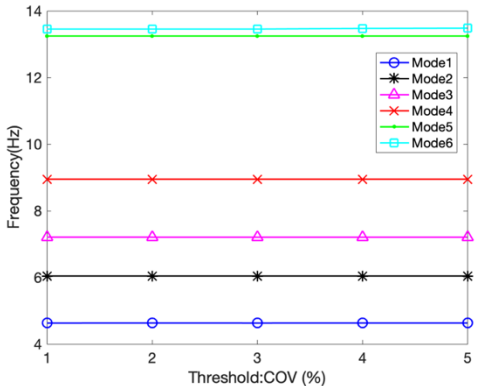
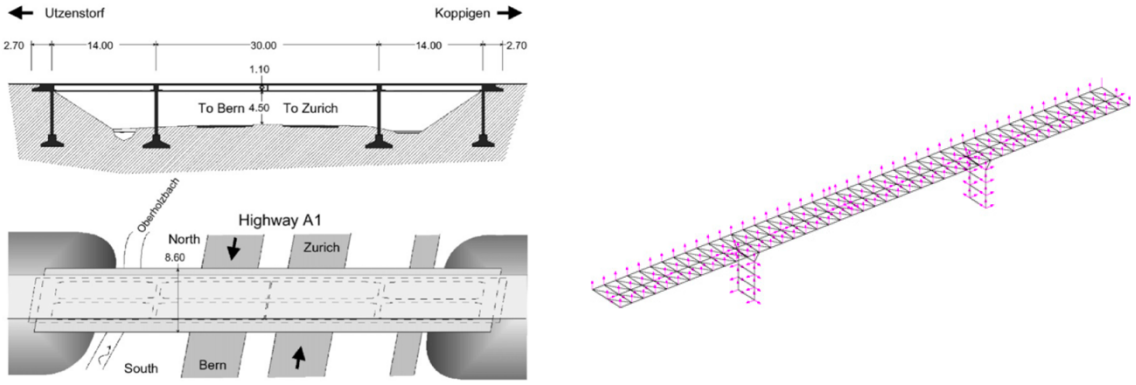


Figure 11. Frequencies at different COV threshold: ($i = 60$; $n_{max} = 100$)

On the other hand, a different choice of uncertainty threshold (COV of frequency) is utilized to evaluate its effect on identification results. The uncertainty threshold is varied from 1% to 5% in the interval of 1%. As shown in Figure 11, the proposed approach yields the same frequencies regardless of COV thresholds, indicating a COV threshold can be safely chosen in the range of $1\% \leq \mathrm{COV} \leq 5\%$.

4.2. Application 2: Concrete box girder bridge



(a) (b)

Figure 12. Description of the Z24 bridge: (a) Front and Top view (Maeck & De Roeck 2003); (b) sensor layout (Döhler *et al.* 2013)

The proposed approach is also applied to the Z24 bridge benchmark to validate its performance. The Z24 bridge was built in 1963 and located in Switzerland, serving to connect Koppigen with Utzenstorf and crossing over the A1 highway (See Figure 12 (a)). It is a post-tensioned concrete box-girder bridge with a main span of 100 ft (30 m) and two sides span of 46 ft (14 m). Detailed Introduction of the Z24 bridge can be found in Maeck and De Roeck (2003). The Z24 bridge was demolished at the end of 1998. Before the complete demolition, a short-term progressive damage test was implemented on the bridge to investigate the effect of simulated damage on the safety of the bridge. A total of 17 different damage scenarios were designed under full forced and ambient excitation (Reynders & Roeck 2009). In this work, acceleration response data from the scenario of No.8 for the new reference condition under ambient excitation is used to

assess the proposed approach. A total of 291 DOFs were measured (See Figure 12 (b)). Due to the limited number of sensors, only at most 33 DOFs were measured for each set-up. Therefore, nine measurement set-ups were recorded with most 33 sensors to have full location coverage of the whole bridge, containing five reference sensors that are common to each set-up and 28 moving sensors whose location changes with different set-ups. For the No.5 set-up, only 22 moving sensors were used. Samples of 65536 data were recorded at each set-up at a 100 Hz sampling rate.

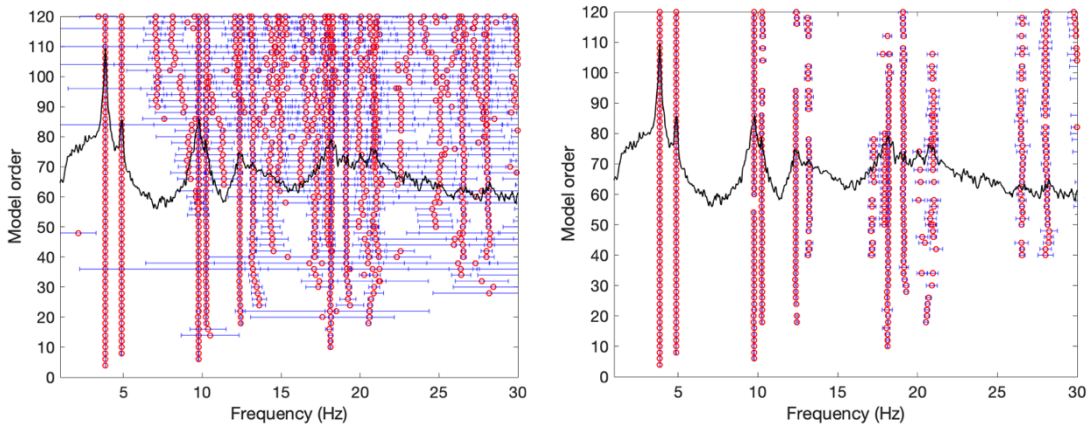
For each dataset, preparation parameters in SSI-cov/ref are defined as: time lag is $i = 50$, model order ranges from 2 to 120, to create stabilization diagrams. Reference sensors are selected as No. 29-33 (for set-up No.5, as No. 23-27). After the stabilization diagram is created, the proposed approach in Section 3 is applied to automatically interpret the stabilization diagram.

4.2.1. Identification results

Nine stabilization diagrams corresponding to each set-up are created; results of the fifth set-up are only presented in Figure 13 due to space limitation in this paper. The singular value spectrum (appeared in curves in Figure 13) is also plotted below the stabilization diagram. $\pm\sigma$ (standard derivation) uncertainty bounds of frequency are plotted as horizontal bars. Figure 13 (a) displays modal identification results using conventional validation criteria, many scattered poles which are definitely spurious modes, still retain in the stabilization diagram. However, the uncertainty criterion can remove most spurious modes, demonstrating that the uncertainty criterion is more effective than conventional validation criteria (See Figure 13 (b)).

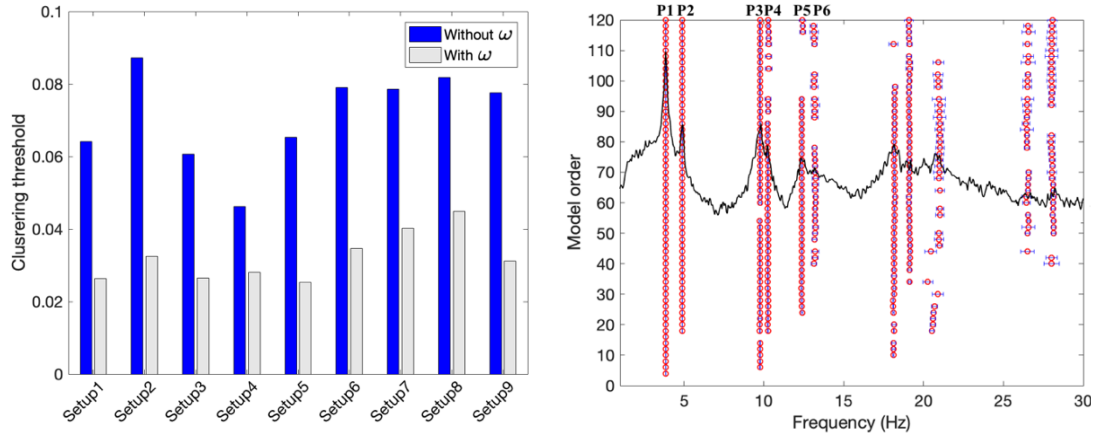
Based on the remaining poles in the stabilization diagram after the pre-processing stage, the clustering threshold for each measurement set-up is calculated using Equations. (9) and (10). As shown in Figure 14 (a), all threshold values are significantly reduced compared to those calculated without weighting factor, as weighting factor can offset the effect of mode shape

difference. Manual clustering thresholds in commercial OMA software are usually below 0.06 (Neu *et al.* 2017). The threshold derived from the newly proposed method is closer to the one from manual analysis, indicating proposed method's rationality and feasibility in practice. Mode clustering described in Section 3.2.2 is then implemented to group physical modes. The number of scattered poles is greatly removed by the proposed approach, only remaining obvious vertical alignments in the stabilization diagram in Figure 14 (b).



(a) (b)

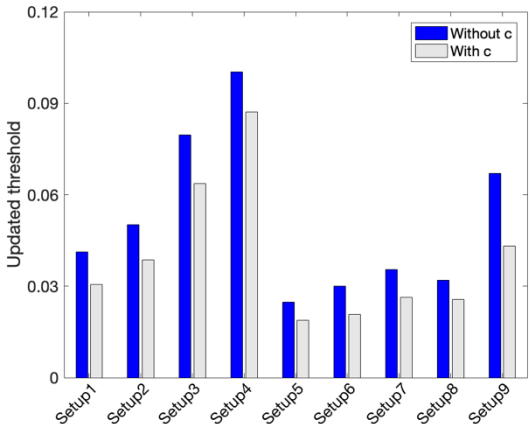
Figure 13. Pre-processing stage for No. 5 set-up: (a) after validation criteria; (b) after uncertainty criterion



(a) (b)

Note: ω is weighting factor in Equation (9)

Figure 14. The clustering stage: (a) calculated clustering threshold; (b) after improved self-adaptive clustering for No. 5 set-up



Note: c is the weighting factor in Equation (11)

Figure 15. Updated threshold

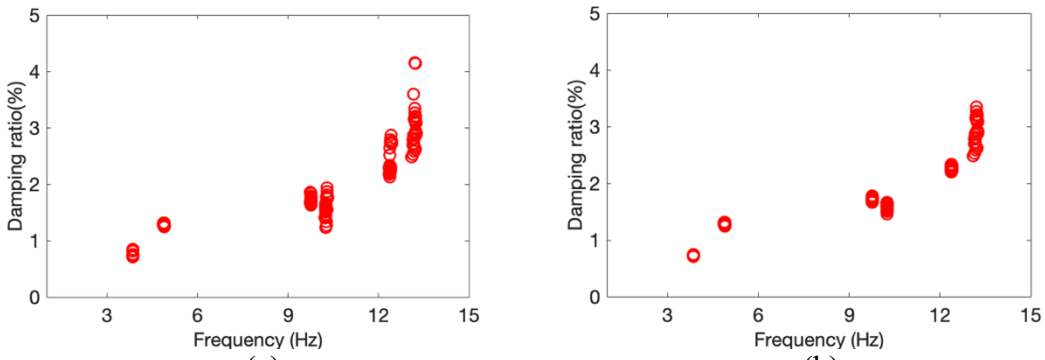


Figure 16. Damping ratio vs frequency: (a) before outlier detection; (b) after outlier detection

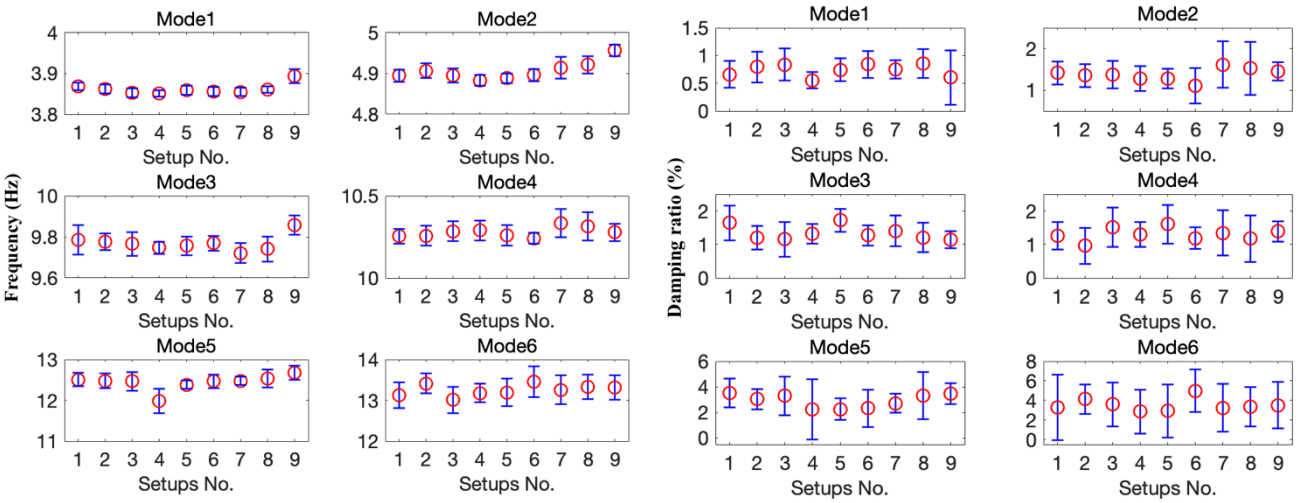
In addition, an improved self-adaptive clustering that considers the weighting factor of c in Equation (11) tends to give a smaller updated threshold, implying identified modal parameters are more stable and consistent with each other (See Figure 15). The use of the weighting factor can improve the performance of clustering. Because only the first six modes are present in baseline work (Reynders *et al.* 2012), the first six clusters are presented in Figure 14 (b), marked as P_1 to P_6. Robust outlier detection is used to identify outlying modes (See Figure 16). Finally, the average of modal parameters in each cluster is selected as representative. Table 3 shows the sample mean and sample standard derivation of frequency and damping ratio over nine measurement set-ups obtained from the proposed approach. The standard deviation in Table 3 represents the setup-to-setup sample statistics among all set-ups. The calculation of sample standard derivation in Table

3 only considers the environmental change among different set-ups rather than uncertainty sources summarized in Section 3.1.2. It is seen from Table 3 that the damping ratio has more significant variability than frequency, implying it is more challenging to identify damping ratio in practice, as the damping ratio is sensitive to environmental change. Overall, the proposed approach's identified frequencies and damping ratios are almost identical to those from baseline work, demonstrating low demand for human intervention.

Table 3. Identification results

Modes	Frequency (Hz)				Damping ratio (%)			
	Baseline*	STD	The proposed approach	STD	Baseline*	STD	The proposed approach	STD
1st (P_1)	3.86	0.01	3.86	0.01	0.8	0.1	0.74	0.11
2nd (P_2)	4.90	0.01	4.91	0.02	1.4	0.2	1.38	0.15
3rd (P_3)	9.76	0.02	9.77	0.04	1.4	0.2	1.34	0.21
4th (P_4)	10.3	0.09	10.28	0.03	1.3	0.2	1.30	0.19
5th (P_5)	12.41	0.19	12.44	0.19	2.8	0.4	2.91	0.53
6th (P_6)	13.22	0.15	13.25	0.14	3.4	1.1	3.54	0.66

Note: *: Reynders et al. (2012);



(a) (b)

Figure 17. $\pm 2\sigma$ standard derivation error bar ratio across nine setups: (a) frequency; (b) damping ratio

Uncertainties of modal parameters airing from assumptions made in SSI, such as linear, stationary structural behavior, white noise, etc., are also studied. Figure 17 shows the variability

of frequency and damping ratio from modes 1 to mode 6 across nine measurement set-ups, respectively, with open circles representing the parameter estimates and error bars covering $\pm 2\sigma$ standard derivations. Both frequencies and damping ratios change over time, while the damping ratios have larger uncertainties. The negative damping ratio is immaterial in Figure 17 (b), such as mode 5 at No. 4 set-up and mode 6 at No. 1 set-up, merely because of the Gaussian distribution approximation and the larger standard derivation. Mode 6 has relatively larger uncertainty since the mode is not excited well. Table 4 presents the average of standard derivation for each cluster over nine measurement set-ups using Equation (13). As expected, the sixth cluster has the highest uncertainty, implying it is relatively harder to identify this cluster, which is also reflected in Figure 14 (b) that the sixth vertical alignments from the left form at a very ambiguous peak. Generally speaking, quantities of identified frequency and damping ratio are consistent from one to another set-up numbers., suggesting robust and fair performance on modal analysis.

Table 4. Average of standard derivation (STD) for physical clusters among nine setups

No. cluster	1	2	3	4	5	6
STD (%)	0.56	1.37	3.06	5.25	10.19	17.61

The global mode shapes are directly assembled from a local one in a single dataset by multiplying by a scaling factor so that their common DOFs, at the location of reference sensor, agree well with each other through data fitting, namely, the method of least squares (Au 2017). For the sake of article spaces, detailed procedures for calculating scaling factors are referred to Au (2017). As seen in Figure 18, the entire six modes are successfully identified from vibration response in all the nine measurement set-ups, which are in good accordance with those in Reynders, Houbrechts and De Roeck (2012). Mode 1 is a typical bending mode with a symmetric shape that has the maximum deflection at midspan. Mode 2 is the first torsional mode with a slight rotational dynamic behavior along y-axis (transverse direction). Similar to mode 2, but more significant

rotation is observed on modes 3 and 4; they are another two torsional modes. Modes 5 and 6 are vertical modes with asymmetric shapes. Furthermore, five mode shapes at the only vertical direction (corresponding z-axis in Figure 18) and one mode shape at the only transverse direction (corresponding y-axis in Figure 18) are also presented in Figures. 19 and 20, $\pm 2\sigma$ uncertainty bounds are plotted as blue dashed lines. Figure 19 shows only mode 6 has relatively wider uncertainty intervals since it is weakly-excited, while others have narrow bounds, implying mode shapes are identified with an accurate level.

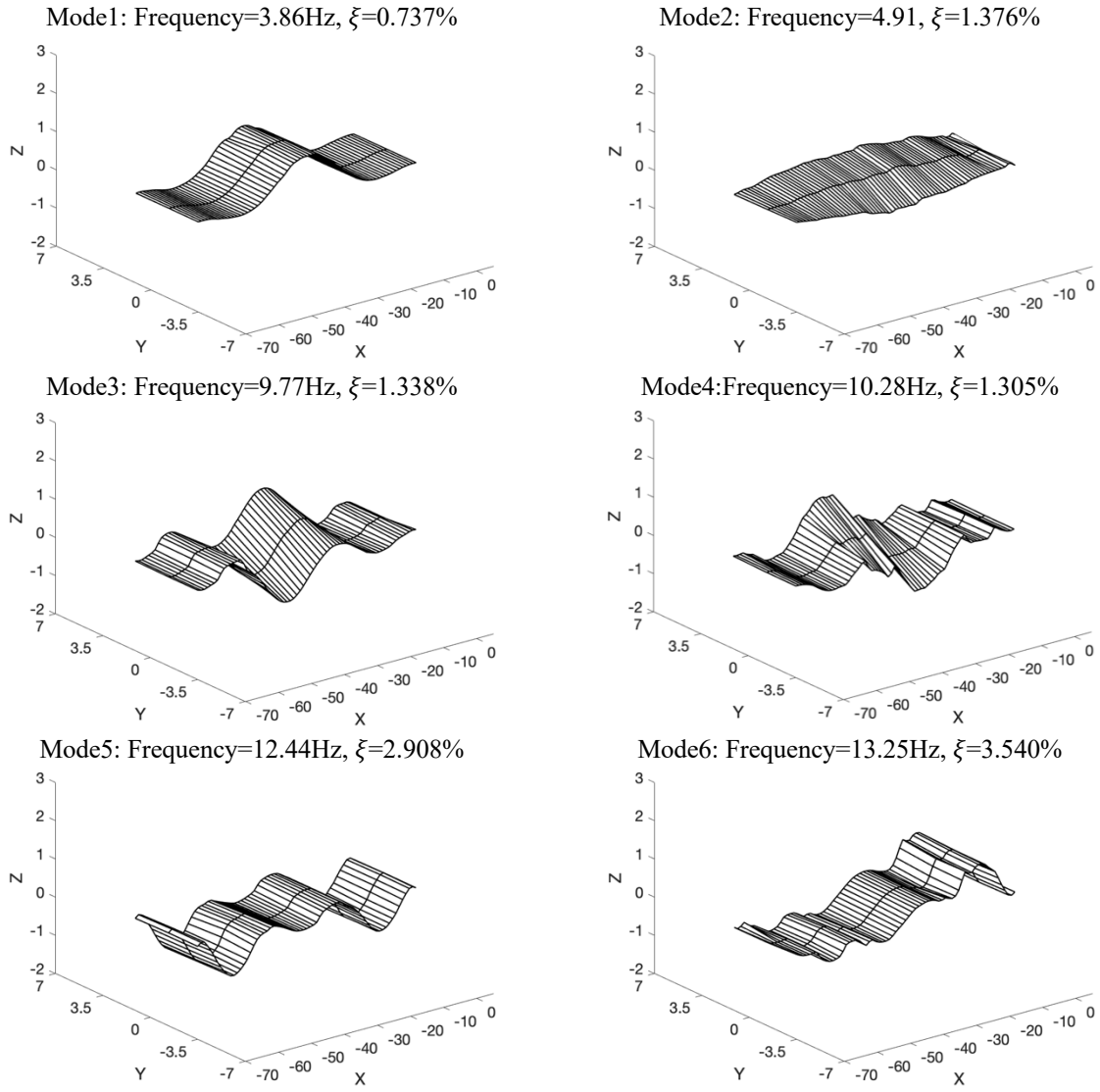


Figure 18. Mode shapes of the Z24 bridge

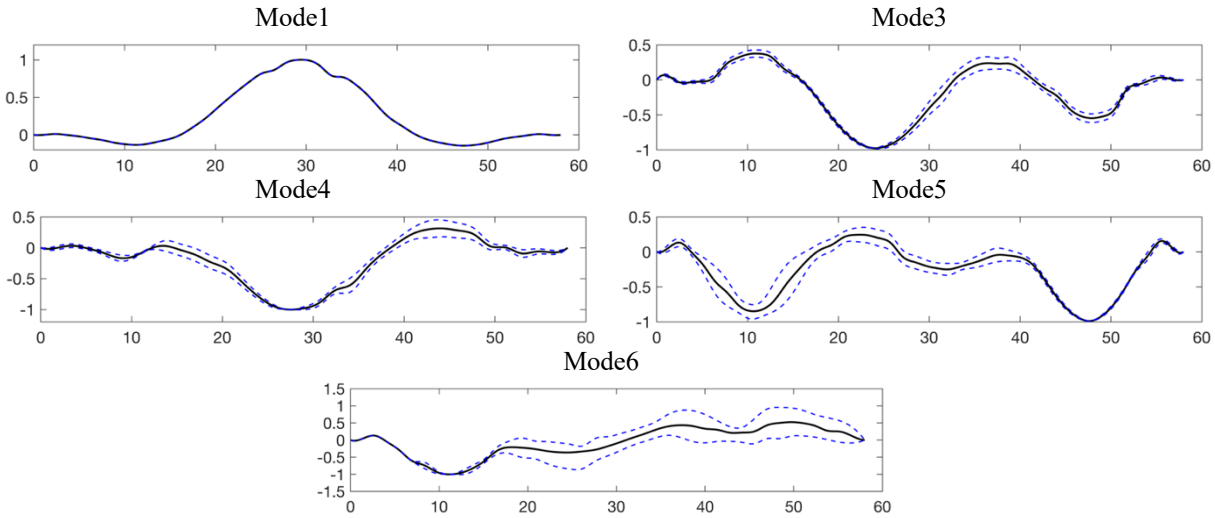


Figure 19. Mode shapes at X-Z plane with $\pm 2\sigma$ uncertainty bounds

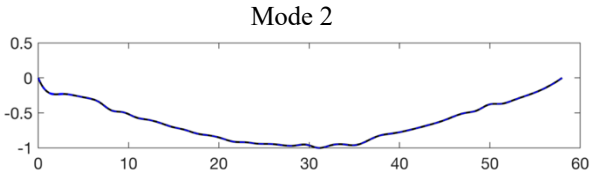


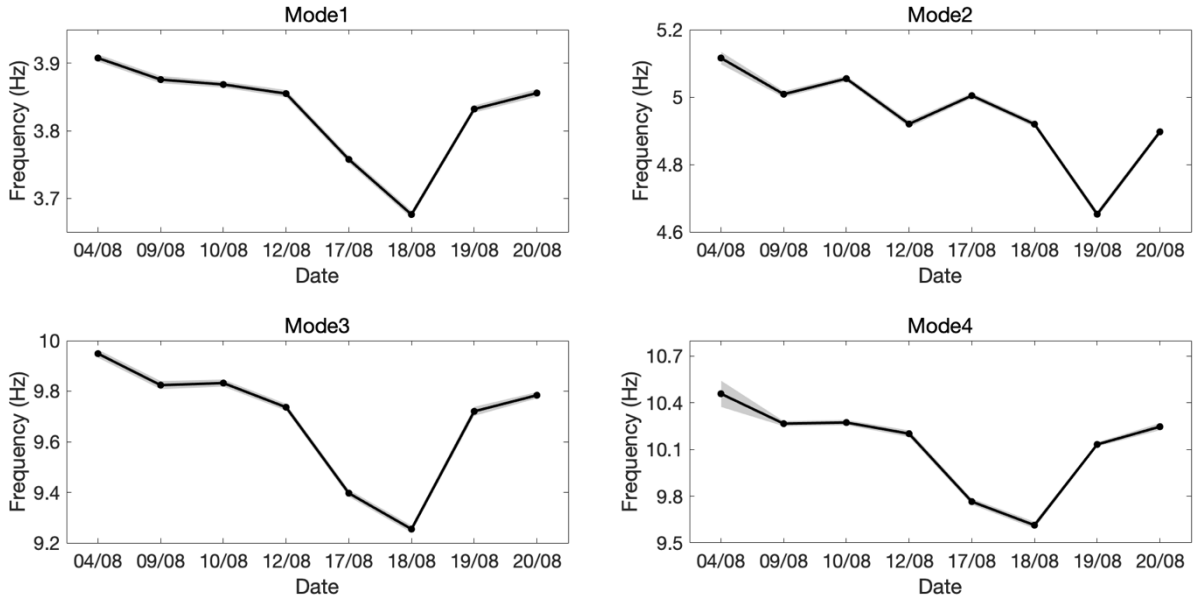
Figure 20. Mode shape at X-Y plane with $\pm 2\sigma$ uncertainty bounds (not visible)

Table 5. The first eight damage scenarios during the progressive damage test in 1998 (Maeck & De Roeck 2003)

Measurement No	Date	Scenario
1	04, August	First reference measurement
2	09, August	Second reference measurement
3	10, August	Settlement of pier, 20mm
4	12, August	Settlement of pier, 40mm
5	17, August	Settlement of pier, 80mm
6	18, August	Settlement of pier, 95mm
7	19, August	Tilt of foundation
8	20, August	Third reference measurement

To further investigate the performance of the proposed approach for continuous health monitoring. As seen in Table 5, the proposed approach is applied to eight different damage scenarios during the short-term progressive damage test. A total of 72 datasets consists of nine individual measurement setups for each damage scenario. The tracked frequencies, damping ratios, and associated uncertainty are plotted in Figures. 21 and 23, with sample mean (solid black lines)

and two averages of standard derivation (grey shaded areas) among all measurement set-ups. It is observed that the maximum frequency happened at scenario No. 1 corresponding to undamaged condition; the minimum frequency happened at modes 1, 3 ,4 and 5 in scenario No. 6 and modes 2 and 6 in scenario No. 7 for corresponding to 95mm settlement of pier and tilt foundation. As seen in Figure 22, damage scenarios in Table 5 have significant effect on frequency, especially when the pier is settled, and foundation is tilted. For example, frequency reduction at modes 1, 3, 4 and 5 reaches the maximum magnitude when the pier has the maximum settlement, 95mm, ranging from 5.93% to 8.08%. On the other hand, modes 2 and 6 have the maximum frequency reduction of 9.06% and 3.66%, respectively, due to foundation's tilt, respectively. In contrast, 95-mm pier settlement still impairs on frequency, suggesting pier and foundation may be paid more attention during SHM. Figure 23 shows the variability of damping ratio is smaller than of frequency, implying damping ratio is not sensitive to global damage scenarios in Table 5, but the damping ratio has much larger uncertainty. The results demonstrate potential benefits to handle a large amount of data with an acceptable level of performance while reducing human involvement. Therefore, the proposed approach is suitable for continuous health monitoring and modal tracking.



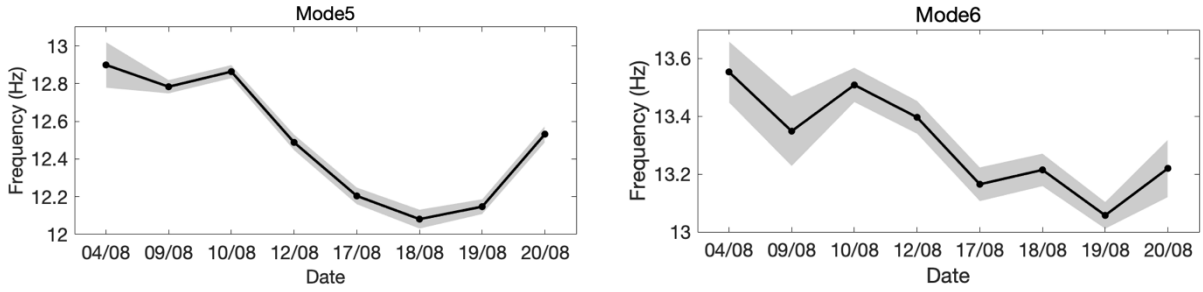


Figure 21. Identified frequencies for different damage scenarios

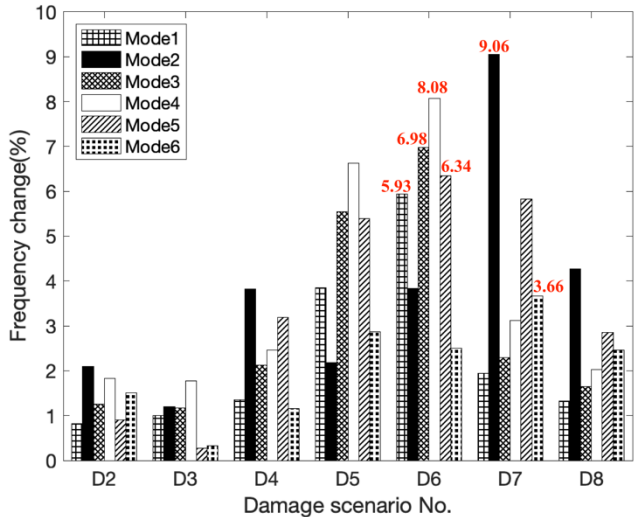
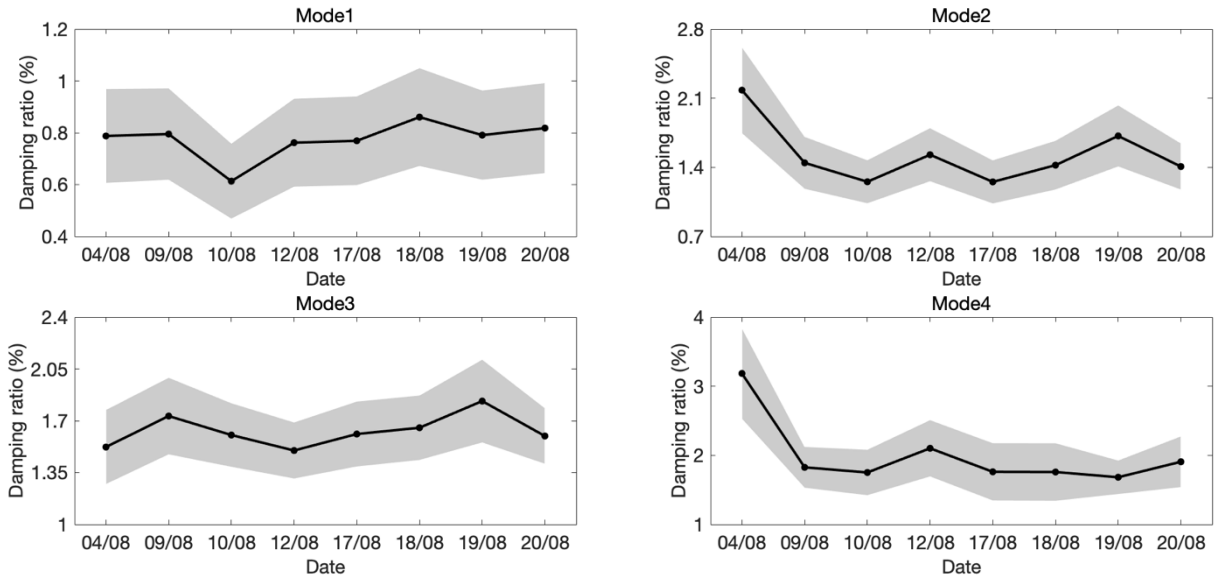


Figure 22. Frequency change due to damage



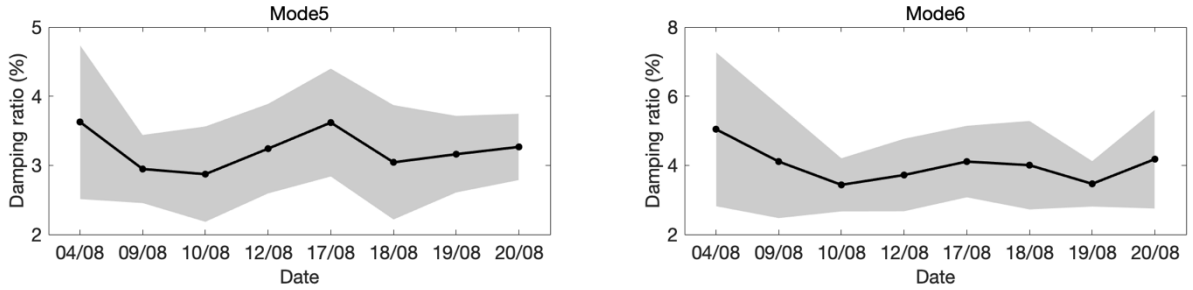
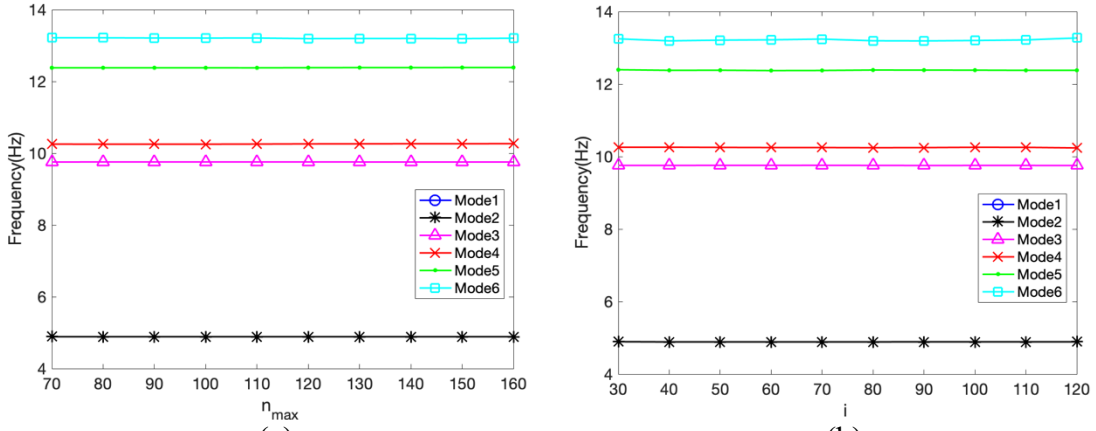


Figure 23. Identified damping ratio for different damage scenarios

4.2.2. Sensitivity analysis



(a) (b)

Figure 24. Frequencies at different parameters: (a) model order range sensitivity (fixed $i = 50$); (b) time lag range sensitivity (fixed $n_{max} = 90$)

To examine the performance of the proposed approach in case of a different combination of preparation parameters in SSI-cov/ref, e.g., the maximum mode order, n_{max}, and time lag, i, the sensitivity analysis is conducted for No. 5 measurement setup in this example. n_{max} and i range from 70 to 160 and from 30 to 120, respectively. As seen in Figure 24, the proposed approach has consistent behavior for identifying six modes using different SSI-cov/ref preparation parameters. Similar to Application 1, Figure 25 shows that any threshold between 1 and 5% yields the same outcomes. The sensitivity analysis demonstrates that the proposed approach is insensitive to two crucial parameters in SSI-cov/ref: model order and time lag. Generally, model order is over-estimated to identify weakly excited modes, yielding more spurious modes; a small value of time lag may fail to generate enough stable poles in the stabilization diagram. It is very difficult to

determine the best model order and time lag in real test. The proposed approach provides more flexibility for the selection of the two parameters, significantly facilitating automated modal identification in practice.

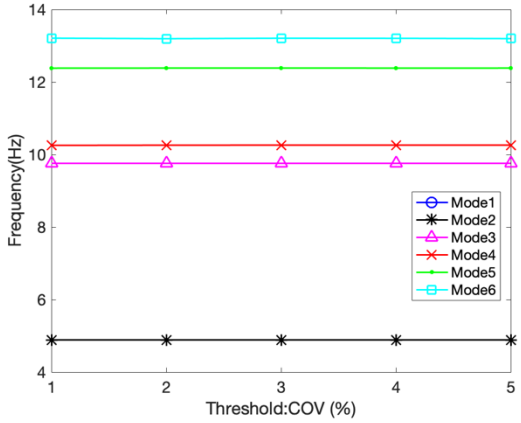


Figure 25. Frequencies at different COV threshold: ($i = 50$, $n_{max} = 120$)

4.3. Discussion and recommendation

The proposed approach accurately identifies the modes of interest and concurrently eliminates spurious modes. The weakly excited and closely spaced modes are identified on two bridges under ambient vibration. The procedures only require a few initial parameters setting, e.g., model order range, time lag, the threshold of MPC/MPD, and uncertainty criterion. In short, the proposed approach is insensitive to these parameters, especially, two crucial parameters: model order and time lag. Finally, recommendations are summarized in Table 6. Both n_{max} and i are over-defined, but the proposed approach will remove spurious modes. MPC and MPD can be regarded as standard values, with no need for adjustment. Uncertainty threshold, COV, can also be safely chosen in the range of 1-5%. In addition, two-month period data (55 datasets: Application 1) and short-term progressive damage test (72 datasets: Application 2) demonstrates the feasibility of automated OMA and modal tracking of massive data for continuous health monitoring.

Table 6 Recommendations on initial parameters for structures under complex test condition

Initial parameters	n_{max}	i	MPC	MPD	COV
Recommendation	100-160	2-3 times of Equation (2)	0.3	0.7	1%-5%

Note: n_{max} is the maximum model order; i is time lag; MPC is modal phase collinearity; MPD is mean phase deviation; COV is coefficient of variation of frequency.

5. Conclusion

A two-stage framework for automated operational modal identification based on SSI-cov/ref is proposed in this study. Firstly, a stabilization diagram is created by SSI-cov.ref. Two-stage framework, e.g., pre-processing stage and clustering stage, is then implemented to interpret the stabilization diagram with low demand of user intervention. Two field tests on the bridge are employed to validate the capability of the proposed approach. The main contributions are summarized as follows:

- This uncertainty criterion is efficient in eliminating many undesired modes at the processing stage, which speeds up the later automation process.
- A novel distance calculation with the uncertainty of modal parameters and weighting factor yields a reasonable threshold for clustering.
- An improved self-adaptive clustering is proposed based on weighted distance calculation with the uncertainty of modal parameters.
- The uncertainty on modal parameters and identified physical clusters are also quantified and providing additional information about quality of identified results.
- A robust outlier detection requiring no setting of threshold improves modal parameters' accuracy.

In short, the proposed framework has a minimal user's involvement to achieve sufficient accuracy. Therefore, the proposed work can be suitable for long-term health monitoring, e.g., modal tracking. Some limitations should be further investigated. For example, both steel frame pedestrian bridge

and highway concrete bridge have a relatively wider frequency range (e.g., 0-15Hz), including few weakly-excited modes and closed spaced modes. In contrast, long span or suspension bridges exhibit low frequency range (e.g., 0-1Hz) and multiple extremely closed-spaced modes (Zhang *et al.* 2016, Brownjohn *et al.* 2018). Therefore, the further verification as the future study is needed. Another is the optimized sensor location should be determined to improve computational efficiency.

Disclosure statement

No potential conflict of interest was reported by the authors.

Acknowledgement

The authors would like to acknowledge Tufts University and KU. Leuven for the data used in this paper. The data of Z24 bridge and Dawling Hall footbridge can be downloaded from https://bwk.kuleuven.be/bwm/z24 and https://engineering.tufts.edu/cee/shm/research.asp, respectively.

CRediT author statement

Jice Zeng: Methodology, Writing-Original draft preparation: Young Hoon Kim: Conceptualization, Supervision. All authors reviewed the results and approved the final version of the manuscript.

Appendix A. The pseudocode of an improved self-adaptive clustering

Generate system pole matrix, W, by SSI-cov/ref. W includes:

- Frequency and damping column vector: F and D
- Mode shape matrix: Φ

Stage 1 starts: initially eliminate spurious modes:

conventional validation criteria: • damping ratio check: remove ζ<0 and ζ>10%

• modal complexity check: remove MPC<0.3 and MPD>0.7

uncertainty criterion: • uncertainty check: remove COV of frequency > 2%

Stage 2 starts: compute clustering threshold, d_{max} by Eqs. (9) and (10)

Initialize the cluster number z as 1

For $i = 1: n_{max}$ (starting from pole matrix from the lowest to highest model order)

For $j = 1$: the last pole (scan poles from the first to the last pole at each model order)

If pole at W$[i, j]$ is not assembled

Create a new cluster C_z

Save pole W$[i, j]$ at the first component of this cluster

For $x = (i + 1): n_{max}$ (Consecutively scan model order)

For $y = 1$: the last element

If pole at W$[x, y]$ is not assembled

Compute the weighted distance $d_{n,i}$ in cluster C_z by Eq. (11)

End if

End for

If the weighted distance $d_{n,i}$ of pole m is lower than d_{max}

Collect m to cluster C_z. If no pole is detected, put 0

(if detect more than one pole, select pole with the minimum distance;)

If cluster C_z contains more than 2 poles

Update clustering threshold $d_{max} = \sqrt{median(|d_{n,i} - median(d_{n,i})|)}$

End if

End if

End for

Repeat procedures until achieving n_{max}

Delete cluster containing poles $< \frac{1}{3} n_{max}$

End if

End for

End for

Detect outlying modes by robust outlier detection: function 'robustcov' in MATLAB

Calculate the average of all elements in each cluster as representative

Calculate uncertainty of each physical cluster by Eq. (13)

Reference

Allemang, R. J. & Brown, D. L. (1985). A correlation coefficient for modal vector analysis. *Proceedings of the 1st international modal analysis conference,* SEM Orlando, FL, 110–116.

Au, S., (2017). *Chapter 14: Multi–setup problem. Operational modal analysis: Modeling, bayesian inference, uncertainty laws.* Singapore: Springer, pp, 415–450.

Au, S.-K. (2011). Fast bayesian FFT method for ambient modal identification with separated modes. *Journal of Engineering Mechanics,* 137(3), 214–226.

Boroschek, R. L. & Bilbao, J. A. (2019). Interpretation of stabilization diagrams using density-based clustering algorithm. *Engineering Structures,* 178, 245–257.

Brownjohn, J., Magalhaes, F., Caetano, E. & Cunha, A. (2010). Ambient vibration re-testing and operational modal analysis of the humber bridge. *Engineering Structures,* 32(8), 2003–2018.

Brownjohn, J. M. W., Au, S.-K., Zhu, Y., Sun, Z., Li, B., Bassitt, J., Hudson, E. & Sun, H., (2018). Bayesian operational modal analysis of Jiangyin Yangtze river bridge. Mechanical Systems and Signal Processing, 110, 210-230.

Cabboi, A., Magalhães, F., Gentile, C. & Cunha, Á. (2017). Automated modal identification and tracking: Application to an iron arch bridge. *Structural Control and Health Monitoring,* 24(1), e1854.

Cerioli, A. (2010). Multivariate outlier detection with high-breakdown estimators. *Journal of the American Statistical Association,* 105(489), 147–156.

Chen, H.-P. & Ni, Y.-Q. (2018). *Structural health monitoring of large civil engineering structures* Hoboken, NJ, Wiley Blackwell: Wiley Online Library.

Chen, W., Lu, Z., Lin, W., Chen, S., Ni, Y., Xia, Y. & Liao, W. (2011). Theoretical and experimental modal analysis of the guangzhou new tv tower. *Engineering Structures,* 33(12), 3628–3646.

De Almeida Cardoso, R., Cury, A. & Barbosa, F. (2017). A clustering-based strategy for automated structural modal identification. *Structural Health Monitoring,* 17(2), 201–217.

Döhler, M., Lam, X.-B. & Mevel, L. (2013). Uncertainty quantification for modal parameters from stochastic subspace identification on multi-setup measurements. *Mechanical Systems and Signal Processing,* 36(2), 562–581.

Döhler, M. & Mevel, L. (2013). Efficient multi-order uncertainty computation for stochastic subspace identification. *Mechanical Systems and Signal Processing,* 38(2), 346–366.

Fan, G., Li, J. & Hao, H. (2019). Improved automated operational modal identification of structures based on clustering. *Structural Control and Health Monitoring,* 26(12), e2450.

Honfi, D., Björnsson, I., Ivanov, O. L. & Leander, J. (2020). Informed successive condition assessments in bridge maintenance. *Journal of Civil Structural Health Monitoring,* 10(4), 729–737.

Hubert, M., Debruyne, M. & Rousseeuw, P. J. (2017). Minimum covariance determinant and extensions. *WIREs Computational Statistics,* 10(3), e1421.

Lam, H. F., Peng, H. Y. & Au, S. K. (2014). Development of a practical algorithm for bayesian model updating of a coupled slab system utilizing field test data. *Engineering Structures,* 79, 182–194.

Maeck, J. & De Roeck, G. (2003). Description of Z24 benchmark. *Mechanical Systems and Signal Processing,* 17(1), 127–131.

Magalhães, F., Reynders, E., Cunha, Á. & De Roeck, G. (2009). Online automatic identification of modal parameters of a bridge using the P-LSCF method. *Proceedings of the 3rd International Operational Modal Analysis Conference*, Portonovo, AN, Italy, 21–28.

Malekjafarian, A., Mcgetrick, P. J. & Obrien, E. J. (2015). A review of indirect bridge monitoring using passing vehicles. *Shock and Vibration,* 2015, 286139.

Marulanda, J., Caicedo, J. M. & Thomson, P. (2017). Modal identification using mobile sensors under ambient excitation. *Journal of Computing in Civil Engineering,* 31(2), 04016051.

Matarazzo, T. J., Kondor, D., Santi, P., Milardo, S., Eshkevari, S. S., Pakzad, S. N. & Ratti, C. (2020). Crowdsourcing bridge vital signs with smartphone vehicle trips. *arXiv preprint arXiv:2010.07026*.

Matarazzo, T. J. & Pakzad, S. N. (2018). Scalable structural modal identification using dynamic sensor network data with stridex. *Computer-Aided Civil and Infrastructure Engineering,* 33(1), 4–20.

Mellinger, P., Döhler, M. & Mevel, L. (2016). Variance estimation of modal parameters from output-only and input/output subspace-based system identification. *Journal of Sound and Vibration,* 379, 1–27.

Moaveni, B. & Behmanesh, I. (2012). Effects of changing ambient temperature on finite element model updating of the dowling hall footbridge. *Engineering Structures,* 43, 58–68.

Moser, P. & Moaveni, B. (2011). Environmental effects on the identified natural frequencies of the dowling hall footbridge. *Mechanical Systems and Signal Processing,* 25(7), 2336–2357.

Neu, E., Janser, F., Khatibi, A. A. & Orifici, A. C. (2017). Fully automated operational modal analysis using multi-stage clustering. *Mechanical Systems and Signal Processing,* 84, 308–323.

Nord, T. S., Petersen, O. W. & Hendrikse, H. (2019). Stochastic subspace identification of modal parameters during ice-structure interaction. *Philos Trans A Math Phys Eng Sci,* 377, 20190030.

Peeters, B. (2000). System identification and damage detection in civil engeneering [doctoral dissertation, katholieke universiteit leuven]. Faculteit toegepaste wetenschappen arenbergkasteel, b-3001 heverlee (Belgium). Http://www.Bwk.Kuleuven.Ac.Be/bwm/

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.

Reynders, E., Houbrechts, J. & De Roeck, G. (2012). Fully automated (operational) modal analysis. *Mechanical Systems and Signal Processing,* 29, 228–250.

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.

Reynders, E. & Roeck, G. D. (2009). Continuous vibration monitoring and progressive damage testing on the z 24 bridge. *Encyclopedia of structural health monitoring,* (2009).

Sadeghi Eshkevari, S., Pakzad, S. N., Takáč, M. & Matarazzo, T. J., (2020). Modal identification of bridges using mobile sensors with sparse vibration data. *Journal of Engineering Mechanics,* 146(4), 04020011.

Sarlo, R. & Tarazaga, P. A. (2019). Modal parameter uncertainty estimates as a tool for automated operational modal analysis: Applications to a smart building. *Dynamics of civil structures, volume 2.* 177–182.

Sarlo, R., Tarazaga, P. A. & Kasarda, M. E. (2018). High resolution operational modal analysis on a five-story smart building under wind and human induced excitation. *Engineering Structures,* 176, 279–292.

Sun, M., Alamdari, M. M. & Kalhori, H. (2017). Automated operational modal analysis of a cable-stayed bridge. *Journal of Bridge Engineering,* 22(12), 05017012.

Ubertini, F., Gentile, C. & Materazzi, A. L. (2013). Automated modal identification in operational conditions and its application to bridges. *Engineering Structures,* 46, 264–278.

Wu, G., He, M., Liang, P., Ye, C. & Xu, Y. (2020). Automated modal identification based on improved clustering method. *Mathematical Problems in Engineering,* 2020.

Yang, J.-H. & Lam, H.-F. (2018). An efficient adaptive sequential monte carlo method for bayesian model updating and damage detection. *Structural Control and Health Monitoring,* 25(12), e2260.

Yang, J. P. & Sun, J.-Y. (2020). Pitching effect of a three-mass vehicle model for analyzing vehicle-bridge interaction. *Engineering Structures,* 224, 111248.

Yang, K., Yu, K. & Wang, H. (2020). A hybrid method of multi-objective particle swarm optimization and k-means clustering and its application to modal parameter estimation in the time–frequency domain. *Journal of Vibration and Control,* 26(9-10), 769–778.

Yang, X.-M., Yi, T.-H., Qu, C.-X., Li, H.-N. & Liu, H. (2019). Automated eigensystem realization algorithm for operational modal identification of bridge structures. *Journal of Aerospace Engineering,* 32(2), 04018148.

Yang, Y. & Yang, J. P. (2018). State-of-the-art review on modal identification and damage detection of bridges by moving test vehicles. *International Journal of Structural Stability and Dynamics,* 18(02), 1850025.

Yang, Y.-B., Lin, C. & Yau, J. (2004). Extracting bridge frequencies from the dynamic response of a passing vehicle. *Journal of Sound and Vibration,* 272(3-5), 471–493.

Zeng, J. & Kim, Y. H. (2020). Identification of structural stiffness and mass using bayesian model updating approach with known added mass: Numerical investigation. *International Journal of Structural Stability and Dynamics,* 20(11), 2050123.

Zhang, F.-L., Ni, Y.-Q. & Ni, Y.-C., (2016). Mode identifiability of a cable-stayed bridge based on a Bayesian method. *Smart Structures and Systems*, 17(3), 471-489.