Smart Structures and Systems, *Vol. 24, No. 2 (2019) 243-256*
DOI: https://doi.org/10.12989/sss.2019.24.2.243

Gaussian mixture model for automated tracking of modal parameters of long-span bridge

Jian-Xiao Mao[1,2a], Hao Wang*[1] and Billie F. Spencer Jr.[2b]

[1]*Key Laboratory of C&PC Structures of Ministry of Education, Southeast University, Nanjing 211189, China*
[2]*Nathan M. and Anne M. Newmark Endowed Chair of Civil Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, U.S.A.*

(Received December 6, 2018, Revised May 17, 2019, Accepted June 13, 2019)

Abstract. Determination of the most meaningful structural modes and gaining insight into how these modes evolve are important issues for long-term structural health monitoring of the long-span bridges. To address this issue, modal parameters identified throughout the life of the bridge need to be compared and linked with each other, which is the process of mode tracking. The modal frequencies for a long-span bridge are typically closely-spaced, sensitive to the environment (e.g., temperature, wind, traffic, etc.), which makes the automated tracking of modal parameters a difficult process, often requiring human intervention. Machine learning methods are well-suited for uncovering complex underlying relationships between processes and thus have the potential to realize accurate and automated modal tracking. In this study, Gaussian mixture model (GMM), a popular unsupervised machine learning method, is employed to automatically determine and update baseline modal properties from the identified unlabeled modal parameters. On this foundation, a new mode tracking method is proposed for automated mode tracking for long-span bridges. Firstly, a numerical example for a three-degree-of-freedom system is employed to validate the feasibility of using GMM to automatically determine the baseline modal properties. Subsequently, the field monitoring data of a long-span bridge are utilized to illustrate the practical usage of GMM for automated determination of the baseline list. Finally, the continuously monitoring bridge acceleration data during strong typhoon events are employed to validate the reliability of proposed method in tracking the changing modal parameters. Results show that the proposed method can automatically track the modal parameters in disastrous scenarios and provide valuable references for condition assessment of the bridge structure.

Keywords: Gaussian Mixture Model (GMM); baseline modal properties; automated mode tracking; long-span bridge; structural health monitoring (SHM)

1. Introduction

Structural modal parameters are important for structural health monitoring (SHM) of the civil infrastructure. Estimating and tracking modal parameters throughout the structure's life are important for effective operating management (Brownjohn *et al*. 2010, Ko and Ni 2005, Soyoz and Feng 2009, Spencer and Nagarajaiah 2003) and can provide a valuable reference for design and construction of new bridges (Annamdas *et al.* 2016, Peeters and De Roeck 2001). However, due to the unavoidable uncertainty of excitation and environments (Ou and Li 2010), human interventions are required, limiting the acquisition of modal parameters throughout the long-term structural monitoring.

Automated modal analysis (Reynders *et al*. 2012), i.e., identifying the structural modal parameters and tracking their evolution over time without human intervention, is one of the essential techniques for long-term structural health monitoring (SHM). Generally, automated modal analysis includes two steps, modal parameter estimation and mode tracking. The former consists of estimation of modal parameters from a single measured raw dataset. Numerous studies have been devoted to developing automated modal identification methods. Among the previously proposed methods, clustering-based methods are the most popular selections. Based on the stabilization diagram, Reynders *et al.* (2012) proposed a clustering-based method to autonomously filter the spurious modes and extract meaningful modes. Demarie and Sabia (2018) proposed an approach for the automated modal identification which relies on existing algorithms from machine learning and data mining fields including the autoregressive model, multiple linear regression method and the k-nearest neighbor classifier. Cardoso *et al.* (2017) proposed an enhanced automated approach for modal parameter estimation based on a two-step clustering analysis; numerical and experimental tests were utilized to validate the validity of the proposed methodology. The automated modal identification methods are popular for long-term SHM of bridge structures. However, limited efforts have been devoted to the second step, mode tracking.

*Corresponding author, Professor
E-mail: wanghao1980@seu.edu.cn
[a] Ph.D. Student
E-mail: jx1990@seu.edu.cn
[b] Professor
E-mail: bfs@illinois.edu

ISSN: 1738-1584 (Print), 1738-1991 (Online)

Mode tracking begins with determining the baseline list, i.e., a list of identifiable modal parameters including modal frequencies and mode shapes. Subsequently, modal parameters from new datasets are matched with the baseline list to track their evolution over long-term operation. Selection of the baseline list and associated bounds is the critical first step of mode tracking, being based on appropriate engineering data analysis, and is generally assumed to remain unchanged during the system operation. Previous studies are generally based on such predefined baseline modes based on engineering experience (Cabboi *et al.* 2017, El-Kafafy *et al*. 2017, Verboven *et al.* 2002). Recently, researchers claimed that variation of modal properties may be large due to structural damage (Feng and Bahng 1999, Reynders *et al.* 2014) and the variation property of environmental conditions (Asadollahi and Li 2017, Mao *et al.* 2018a, Zhou *et al.* 2017, Zhou *et al*. 2015). Besides, the identifiable modes may not be the same in different time periods due to change of excitation sources, i.e., modal identifiability. The above-mentioned features make the tracking process manipulation-required, posing significant challenges for long-term SHM of bridge structures. Therefore, an effective approach to automatically track the modal parameters of long-span bridge structures during long-term operation is required.

Due to their advantages in uncovering complex underlying relationships between processes, the emerging machine learning methods are proving to be good alternatives to deal with problems in structural engineering, resulting in significant human time saving and making the decision faster with less error rates (Salehi and Burgueño 2018). Therefore, in this study, Gaussian mixture model (GMM), a popular unsupervised machine learning method, is employed to automatically determine and update baseline modal properties from the identified unlabeled modal parameters. On this foundation, a novel mode tracking method is proposed for automated modal tracking for long-span bridges. Compared with the previous methods, the proposed one seeks to automatically track the evolution of modal parameters without any human participation, such that, one can obtain the timely information of the structural operating condition, which is a very important issue for long-term SHM.

This paper is structured as follows. In the following section, detailed description of proposed automated mode tracking method will be presented. Then, the key theories behind the proposed method, GMM and Bayesian Information Criteria (BIC), will be briefly introduced. A numerical case for a three-degree-of-freedom (3DOF) simply supported beam model and a field-monitoring case of a long-span cable-stayed bridge, Sutong Cable-stayed Bridge (SCB), are utilized to validate the feasibility of GMM for determining and updating the baseline modes in an automated way. SHM data of SCB during strong typhoon events are applied to validate the proposed automated mode tracking methods. Finally, conclusions and suggestions are reached to provide references for operational management and maintenance of similar bridges.

2. Methodology

2.1 Automated modal parameter identification

The parametric system identification method, stochastics subspace identification (SSI) method (Peeters and De Roeck 2001), is applied to estimate the structural modal parameters from raw monitoring data in this study. After that, the stabilization diagram are generated by changing the system order. The autonomous modal parameter identification method (AMPIM) (Mao *et al*. 2019) seeks to obtain modal parameters from the generated stabilization diagram without extra experience-based mode selection, i.e., human manipulation. The adopted AMPIM is briefly summarized using the following 6 steps as shown in Fig. 1.

Step 1: Modal validation criteria (including frequency variation (FV), modal assurance criteria (MAC), modal phase collinearity, etc.) are calculated to describe distances of one mode from the closest mode/pole in the stabilization diagram, including distance metric, mode shape complexity, energy contribution ratio, and other indicators.

Step 2: Modes with modal validation criteria larger than pre-defined thresholds are assumed as spurious ones and will be removed from the stabilization diagram. Observations of remaining modes can be assembled into a multi-dimensional matrix.

Step 3: Then principal component analysis (PCA) (Wold et al. 1987) is applied to condense the observed multi-dimensional matrix into one dimensional vector (the first principal component, PC1).

Step 4: Classify the remaining modes into 2 clusters using the k-means clustering (k=2) (Hartigan and Wong 1979) based on the squared Euclidian distance metric. Two clusters, as well as their centroid coordinates, could be obtained, and the cluster with a smaller centroid coordinate value is assumed as the possible physical one.

Step 5: Determine the optimal number of clusters to trim the hierarchical tree to achieve the best separation of close modes and clear the remaining spurious modes.

Step 6: After the optimal number is determined, the hierarchical clustering is performed to classify the modes into different clusters. Subsequently, the counted the modes of each cluster are utilized as the inputs for k-means clustering (k=2). The clusters with smaller counts are spurious modes. Ultimately, one can obtain the average and standard deviation values of the modal parameters belong to different orders based on the final sets of physical modes.

The AMPIM is the critical first step for automated modal analysis, which helps to identify modal parameters without any user intervention and guarantees the primary requirement for long-term SHM.

2.2 Autonomous tracking of modal parameters

Knowledge of the evolution of modal parameters is the final target of the modal analysis for long-term SHM. To do this, the mode tracking method, that is to compare and link the current identified modal parameters to those in the baseline list, is required. Generally, mode tracking includes

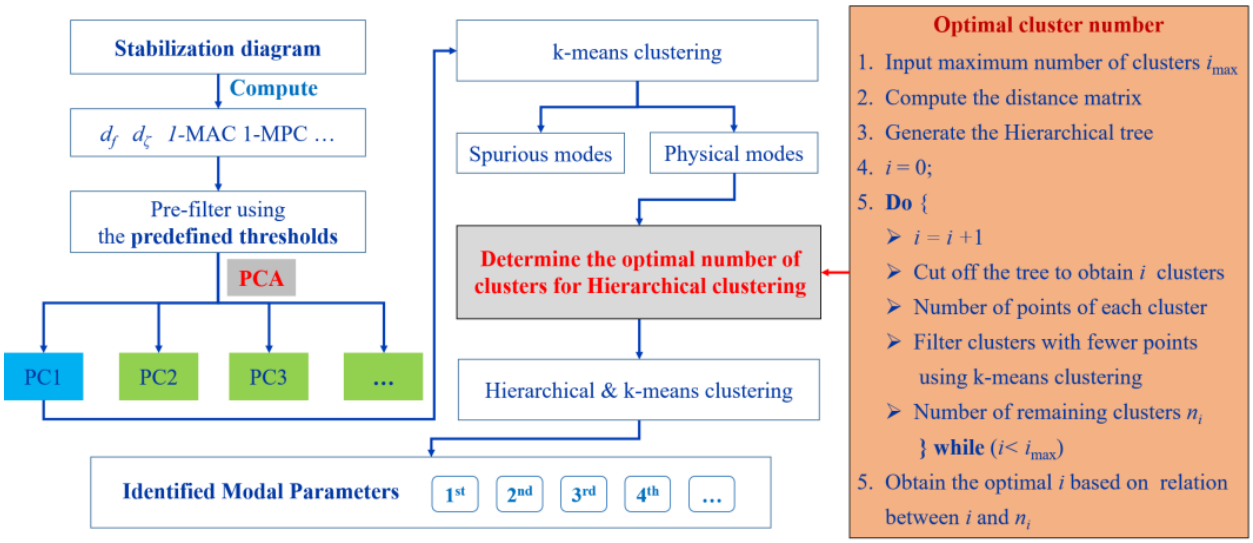


Fig. 1 Workflow of automated modal parameter identification method (AMPIM)

display the tracked modes.

two steps: 1) Determine the baseline list using the identified modal parameters for some time. 2) Compare and link the continuously identified modal parameters to those in the baseline list using similarity indexes, i.e., MAC and FV. The pre-defined baseline modal properties provide prior information for the matching/tracking process. However, due to the modal variation and identifiability properties, the acquisition of such prior information requires human intervention which depresses the feasibility for long-term SHM.

Therefore, a novel method for automated mode tracking is proposed in this study. The proposed tracking method seeks to keep automated track of modal parameters without any user-interaction based on the following three steps. (Ⅰ) Autonomously identify the modal parameters and save the identified results in the database. (Ⅱ) Generate the candidates for baseline modes using GMM and update the baseline list in an automated way. (Ⅲ) Track the currently identified modes using distance metrics, MAC and FV, with the baseline modes. In the following, each single step of the proposed method is described in detail as shown in Fig. 2.

Set the initial value for the time instant t=0, the time interval ∇t, and the minimum system operation time T_o.

(Ⅰ) Update the time instant $t = t+\nabla t$, load the vibration data recorded during (t-∇t,t], identify the structural modal parameters (frequency, damping ratios, and vibration shape) using AMPIM, and save the identified K_t modes to the data base. Repeat this step until $t \geq T_o$.

(Ⅱ) Call the identified modes from the database during the last T_o time, classify the modes into clusters in an automatic way using GMM which are assumed to be the candidate baseline modes, and update the baseline modes by comparing the candidate with the previous one.

(Ⅲ) For all the identified K_t modes during (t-∇t,t], calculate their distances with all the baseline modes using the sum of FV and MAC as shown in Eq. (1). Group each mode to the closest modes in the baseline list. Save and

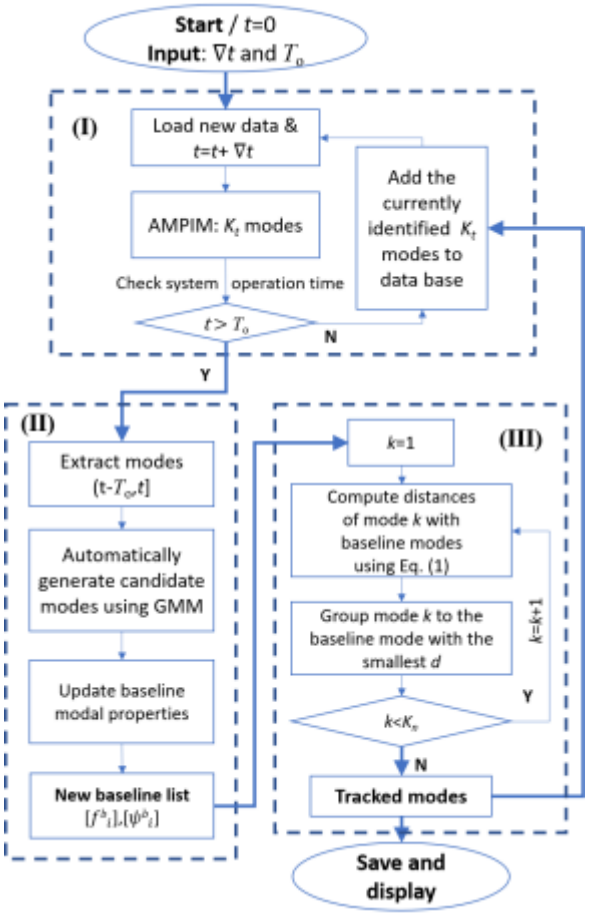


Fig. 2 Flowchart of the automated mode tracking method

Repeat step (Ⅰ) to (Ⅲ) until sensor failure or system maintenance for tracking the structural modes over a long period of time.

$$d_{ck,bi} = \left(d_{ck,bi}^{f}\right) + \left(d_{ck,bi}^{MAC}\right) = \left(\frac{\left|f_{ck} - f_{bi}\right|}{f_{bi}}\right) + \left(1 - \frac{\left|\varphi_{ck}^{*}\varphi_{bi}\right|^{2}}{\left(\varphi_{ck}^{*}\varphi_{ck}\right)\left(\varphi_{bi}^{*}\varphi_{bi}\right)}\right) \quad (1)$$

where f and φ denote the frequency and mode shape, respectively. The $d_{ck,bi}^{f}$ represents FV between the k^{th} frequency in the currently identified modal list and the i^{th} frequency in the baseline list, and $d_{ck,bi}^{MAC}$ represents the MAC index between the k^{th} mode shape in the currently identified modal list and the i^{th} mode shape in the baseline list.

To the best of the authors' knowledge, few studies on automatically determining and updating baseline list for mode tracking have been carried out. The proposed method in this study utilizes GMM and BIC to automatically determine and update the baseline list. GMM helps to classify identified modal parameters into different groups and determine the candidates for baseline list using statistical results of each group. Besides, the fitted GMM with smallest BIC value is regarded as the optimal one. In the next section, theories of GMM and BIC will be explained in detail.

3. Gaussian mixture model and bayesian information criteria

3.1 Gaussian mixture model

To obtain statistical properties of baseline modes, one should classify identified modal parameters into K groups, in each group only one kind of modes exist. GMM, one kind of unsupervised machine learning approach, is applied to estimate the probability density of the dataset and classify different modes into clusters automatically(Huang *et al.* 2005).

Suppose that an observation x follows a distribution $p(\mathrm{x}|\mathbf{z})$ conditional on a latent variable $\mathbf{z}$ for cluster assignment. This means that the data set could be modelled using a joint distribution, i.e., $p(\mathrm{x},\mathbf{z})=p(\mathrm{x}|\mathbf{z})p(\mathbf{z})$. Here, $\mathbf{z}$ is a K-dimensional binary random variable with a 1-of-K representation, i.e., an element z_k equals to 1 and all other elements equal to 0. $p(\mathbf{z})$ is the prior distribution over $\mathbf{z}$ and can be expressed as

$$p\left(z_k = 1\right) = \pi_k \quad (2)$$

where parameters $\{\ \pi_k\ \}$ satisfy $0 \le \pi_k \le 1$ and $\sum_{k=1}^{K} \pi_k = 1$.

Due to the 1-of-K representation of $\mathbf{z}$, Eq. (2) can also be expressed as follows.

$$p\left(\mathbf{z}\right) = \prod_{k=1}^{K} \pi_k^{z_k} \quad (3)$$

Suppose that x conditional on $\mathbf{z}$ follows a Gaussian distribution as

$$p\left(\mathrm{x} \mid z_k = 1\right) = \mathrm{N}\left(\mathrm{x} \mid \mu_k, \Sigma_k\right) \quad (4)$$

where μ_k and Σ_k are respectively the mean and covariance of x conditional on $z_k = 1$. The marginal distribution of x, denoted as $p(\mathrm{x})$, could then be obtained by summing the joint distribution $p(\mathbf{z})p(\mathrm{x}\mid\mathbf{z})$ over all possible $\mathbf{z}$ values

$$p(\mathrm{x}) = \sum_{k=1}^{K} \pi_k \mathrm{N}\left(\mathrm{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) \quad (5)$$

where $p(\mathrm{x})$ can be expressed using a linear superposition of Gaussians, that is a Gaussian mixture (Reynolds *et al.* 2000).

To estimate model parameters, including $\boldsymbol{\pi}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$, the most common approach is to find the parameters with maximum likelihood according to the observed data $\mathrm{D}_n = \{\mathrm{x}_1, \ldots, \mathrm{x}_N\}$. The likelihood function Λ for Δ_n sampled independently from the identical mixture distribution can be expressed as follows

$$\mathrm{L}\left(\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Sigma}\right) = \prod_{i=1}^{N} p(\mathrm{x}_i) = \prod_{i=1}^{N} \sum_{k=1}^{K} \pi_k \mathrm{N}\left(\mathrm{x}_i \mid \mu_k, \Sigma\right) \quad (6)$$

The objective function Λ is usually transformed into the log form ϑ.

$$\mathrm{J}\left(\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Sigma}\right) = \sum_{n=1}^{N} \log\left\{\sum_{k=1}^{K} \pi_k \mathrm{N}\left(\mathrm{x}_n \mid \mu_k, \Sigma_k\right)\right\} \quad (7)$$

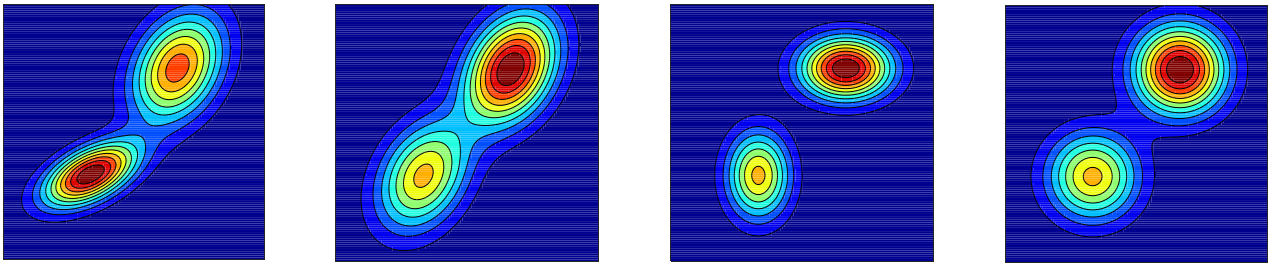
Due to the singular problem of Eq. (7), no close form solution can be found to maximize the objective function problem by setting its derivative equal to zero (Bishop 2006). The expectation-maximization algorithm (EM algorithm) has been proved to be an effective method to find the maximum likelihood for models with latent variables (Moon 1996). After the initial values are selected, EM algorithm alternates between two updates namely as expectation step (E step) and maximization step (M step). Each update for parameters is guaranteed to increase the log likelihood function. Detailed steps of EM algorithm is summarized as follows (Bishop 2006).

1) Set initial values for model parameters $\boldsymbol{\pi}_0$, $\boldsymbol{\mu}_0$, and $\boldsymbol{\Sigma}_0$, and evaluate the corresponding log likelihood according to Eq. (7).
2) **E step**. Evaluate the responsibilities using current parameters.

$$\gamma(z_{nk}) = \frac{\pi_k \mathrm{N}\left(\mathrm{x}_n \mid \mu_k, \Sigma_k\right)}{\sum_{j=1}^{K} \pi_j \mathrm{N}\left(\mathrm{x}_n \mid \mu_j, \Sigma_j\right)} \quad (8)$$

3) **M step**: Calculate the parameters using the updated responsibilities.

$$\mu_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathrm{x}_n \quad (9)$$



(a) Full (b) Tied (c) Diagonal (d) Spherical

Fig. 3 Exhibition of four two-component GMMs with different covariance matrices

$$\Sigma_k^{\text{new}} = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})(\mathrm{x}_n - \mu_k^{\text{new}})(\mathrm{x}_n - \mu_k^{\text{new}})^{\mathrm{T}} \tag{10}$$

$$\pi_k^{\text{new}} = \frac{N_k}{N} \text{ and } N_k = \sum_{n=1}^{N}\gamma(z_{nk}) \tag{11}$$

4) Estimate the current log likelihood value and check for convergence. If the convergence criterion is not satisfied return to step 2.

The initialization of the parameters is critically important for the convergence speed of EM algorithm. In practice, the clustering results using k-means algorithm are usually taken as the initializations (Kanungo *et al.* 2002).

A single Gaussian distribution is completely controlled by the mean vector μ_k, and covariance matrix Σ_k, where μ_k controls the central location and Σ_k determines the dispersion/shape. GMM is a linear combination of Gaussians, hence the covariance matrices for different mixture components may be different. Traditionally, four different types of covariance matrices, i.e., full, tied, diagonal, and spherical, are utilized to fit GMM. Full means the covariance matrices of different components are independently with each other. Tied means the covariance matrices of different components are the same, but the shape may be random. Diagonal means the covariance matrices of different components are all diagonal and may vary between components. Spherical indicates that the covariance matrix is diagonal with the same, but the matrix value may be different for different components. Four two-component GMMs with different covariance matrices are exhibited in Fig. 3 to illustrate the influence of the covariance matrices on the shape of GMM.

3.2 Bayesian information criteria

With the increase of mixture components, larger likelihood could be obtained, however, overfitting problems may take place (Cao and Wang 2014). Therefore, the users are required to estimate the suitable component number which makes the fitted GMM balanced between maximum likelihood and overfitting risk. In this study, BIC (Jeffreys 1998) is utilized to estimate posterior probability of the generated models conditional on observed data Δ_n and the fitted model with the smallest BIC is regarded as the optimal one.

Suppose there are K options, i.e., K parametric candidate models (M_1,..., M_K). The subscript k represents the component number of the candidate model. In Bayesian inference, the prior probability of the occurrence of model M_j is denoted as $\alpha_j = \mathrm{P}(\mathrm{M}_j)$. According to the Bayes' rule, the posterior probability of Model M_j given data Δ_n can be expressed as

$$\begin{aligned}\mathrm{P}(\mathrm{M}_j|\mathrm{D}_n) &= \frac{p(\mathrm{D}_n|\mathrm{M}_j)\mathrm{P}(\mathrm{M}_j)}{p(\mathrm{D}_n)} \\ &= \frac{p(\mathrm{D}_n|\mathrm{M}_j)\alpha_j}{\sum_{k=1}^{K}p(\mathrm{D}_n|\mathrm{M}_k)\alpha_k}\end{aligned} \tag{12}$$

where $p(\mathrm{D}_n)$ is the marginal distribution of $_n$ and could be represented as a constant, uncorrelated with the selected model. Besides, a uniform prior over the models is adopted: $\alpha_1 = \alpha_2 = \cdots = \alpha_K = 1/K$. Therefore, $p(\mathrm{D}_n|\mathrm{M}_j)$, the marginal likelihood function (MLF) for model M_j, can be regarded as indicators to determine the optimal the optimal model. According to the definition of the MLF, $p(\mathrm{D}_n|\mathrm{M}_j)$ can be estimated by applying the integration over the whole parameter space as follows

$$\begin{aligned}p(D_n|\mathrm{M}_j) &= \int p(D_n|\mathrm{M}_j,\theta_j)p_j(\theta_j|\mathrm{M}_j)d\theta_j \\ &= \int \mathrm{L}_n(\theta_j)p_j(\theta_j|\mathrm{M}_j)d\theta_j\end{aligned} \tag{13}$$

where $\mathrm{L}_n(\theta_j) = p(D_n|\mathrm{M}_j,\theta_j)$ is the likelihood function for model j, and $p_j(\boldsymbol{\theta}_j|\mathrm{M}_j)$, represents the prior probabilities of parameters $\boldsymbol{\theta}_j$ for model j.

Let $\hat{\theta}_j$ be the maximum a posterior (MAP) estimator under model M_j as shown in Eq. (14).

$$\frac{\partial}{\partial\theta_j}\mathrm{L}_n(\theta_j)\bigg|_{\theta_j=\hat{\theta}_j} = 0 \tag{14}$$

Apply the Taylor expansion at $\hat{\theta}_j$, and the estimation of $L_n(\theta_j)$ can be obtained

$$\mathrm{L}_n(\theta_j) \approx \mathrm{L}_n(\hat{\theta}_j)\exp\left(-\frac{1}{2}(\theta_j-\hat{\theta}_j)^T I_n(\hat{\theta}_j)(\theta_j-\hat{\theta}_j)\right) \tag{15}$$

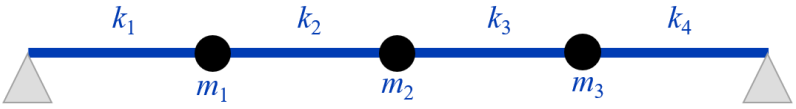


Fig. 4 General view of the 3DOF simply supported girder model

where $I_n(\theta_j) = -\frac{\partial^2 \log p(\mathrm{D}_n | \mathrm{M}_j, \theta_j)}{\partial \theta_j \partial \theta_j^T}$ is empirical Fisher information matrices for the dataset Δ_n.

Substitute the Eq. (15) into Eq. (13), and an approximation of $p(\mathrm{D}_n | \mathrm{M}_j)$ can be expressed as follows

$$p(\mathrm{D}_n | \mathrm{M}_j) \approx \mathrm{L}_n(\hat{\theta}_j) p_j(\theta_j | \mathrm{M}_j) \frac{(2\pi)^{d_j/2}}{n^{d_j/2} \left| I_1(\hat{\theta}_j) \right|^{1/2}} \quad (16)$$

where d is the dimensionality of model M_j and the obtained results can also be represented in the log format

$$\begin{cases} \log p(\mathrm{D}_n | \mathrm{M}_j) \approx \log \mathrm{L}_n(\hat{\theta}_j) - \frac{1}{2} d_j \log n + \mathrm{C} \\ \mathrm{C} = -\frac{1}{2} \log \left| I_1(\hat{\theta}_j) \right| + \frac{1}{2} d_j \log(2\pi) - \frac{1}{2} \log p(\hat{\theta}_j | \mathrm{M}_j) \end{cases} \quad (17)$$

It should be noted that C is of smaller order than $\log \mathrm{L}_n(\hat{\theta}_j) - \frac{1}{2} d_j \log n$, and we can approximate the $\log p(\mathrm{D}_n | \mathrm{M}_j)$ with Bayesian information criteria defined as follows

$$\mathrm{BIC}(\mathrm{M}_j) = \log \mathrm{L}_n(\hat{\theta}_j) - \frac{1}{2} d_j \log n \quad (18)$$

The BIC score provides a large sample approximation for the log posterior probability of the fitted model. The candidate model with the maximum BIC can be regarded as the optimal one, since it corresponds to the highest Bayesian posterior probability. It is also validated that BIC is model selection consistent, i.e., if the true model is within the candidate pool, the probability that BIC selects the true model goes to 1 as n goes to infinity (Wasserman 2000).

4. Validation of the automated baseline determination method

4.1 Numerical case

Determination of the baseline modal properties is one of the most important part for mode tracking. A numerical case of a simply supported beam is presented in this section to illustrate this issue. According to the lumped mass theory, the simply supported beam is simulated using a 3DOF system as exhibited in Fig. 4. The lumped mass values of the 3DOF system are m_1, m_2, and m_3, respectively. The vertical shear stiffness values between different components are denoted as k_1, k_2, k_3, and k_4, respectively. The damping ratios are not considered in this numerical case. Suppose that all the structural parameters are random variables generated from Gaussian distributions as listed in Table 1. The mass parameters, m_1, m_2, and m_3, are assumed to follow the same distribution, and so do the stiffness parameters k_1, k_2, k_3, and k_4.

Suppose that the structural parameters are generated, independent identical distribution, according to the distribution listed in Table 1. For each structural parameter, 100,000 samples are generated, hence totally 100,000 models with different structural parameters can be obtained.

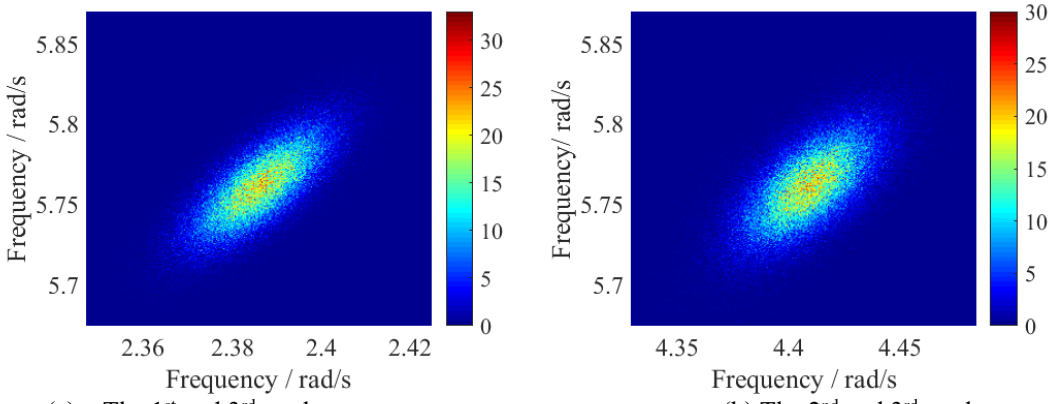
The theoretical modal frequencies of the generated 100,000 models are calculated and exhibited in Fig. 5.

Suppose that all the identified modal parameters are mixed in a group without any labels. Then, GMM is utilized to assign the modal parameters into different groups, where modal parameters belonging to the same mode are classified in one group. Determination of mixture component is one of the most difficult issue for GMM, and still relies on engineering experience. The BIC, indicating the posterior distribution of the fitted model conditional on observed data, is then applied to find optimal one from candidate models. The fitted model with the smallest BIC is assumed to be the optimal one. In the numerical case, GMMs with different mixture component numbers, i.e., from 1 to 9, are fitted using EM algorithm according to dataset. The BIC values corresponding to the fitted models are shown in Fig. 6.

As shown in Fig. 6, the BIC values of fitted GMMs with different covariance matrices show similar variation trend. BIC values decrease sharply with the increase of number of components when it is smaller than 3 and keep relative stable when the components number is larger than 3. From the zoom up diagram of Fig. 6, BIC slowly increases with the component number when it is larger than 3. Therefore, the GMM with 3 mixture components and full covariance matrix are automatically set as the optimal one. The classified modal frequencies in each group are shown in Fig. 7. One can see that GMM can successfully classify the modal frequencies of each order into the same group, based on which the candidate for the baseline modes can be determined.

Table 1 Distribution parameters for the 3DOF simply supported girder model

Notation	Variables	Gaussian distribution		Unit
		μ	σ	
Mass	m_1, m_2, and m_3	1, 800	18	kg
Stiffness	k_1, k_2, k_3, and k_4	17, 500	140	N/m



(a) The 1st and 3rd modes (b) The 2nd and 3rd modes

Fig. 5 Histogram of natural frequencies of the 3DOF simply supported girder model

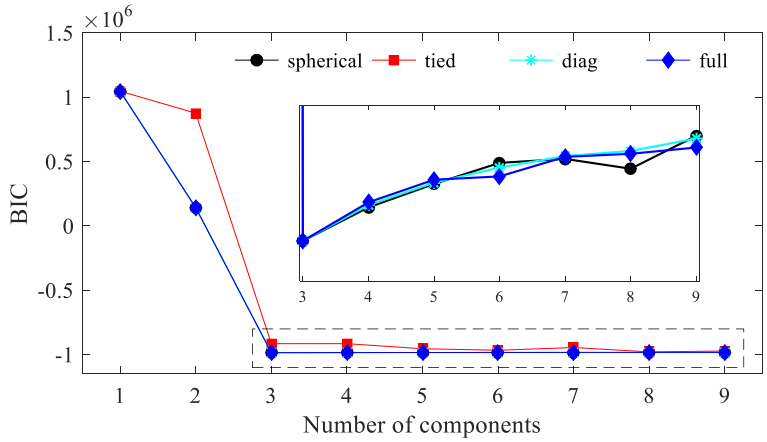


Fig. 6 BIC values for the fitted GMMs with different component number

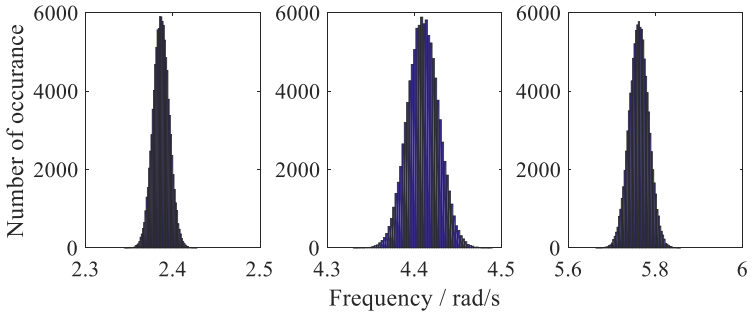


Fig. 7 Classified modal frequencies in 3 groups using GMM

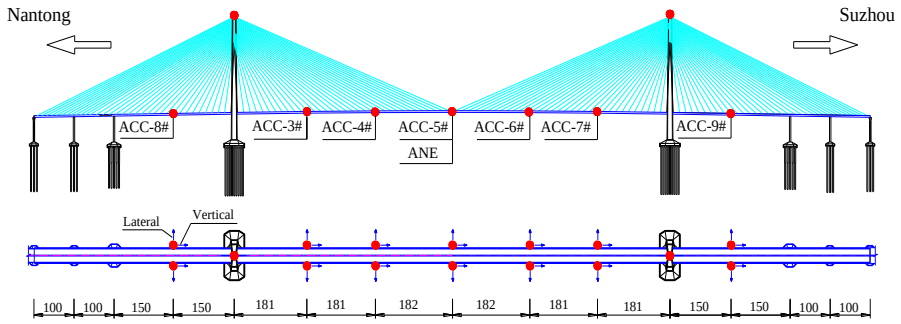


Fig. 8 General view of SCB and the installed accelerometers (Unit: m)

Table 2 The number of modes in each cluster of the dataset (T_o =7 days)

Mean /Hz	0.1854	0.2189	0.2386	0.3361	0.3742	0.3926	0.4246
Number	168	137	1	166	2	139	2
Notes	Physical	Physical	Outlier	Physical	Outlier	Physical	Outlier
Mean /Hz	0.4520	0.5330	0.5774	0.5847	0.6259	0.6719	
Number	156	167	149	8	113	58	
Notes	Physical	Physical	Physical	Outlier	Physical	Physical	

4.2 Field monitoring case

In this section, the field monitoring data of SCB, are applied to validate the potential usage of the proposed automated mode tracking method for long-term monitoring of the real civil infrastructure. SCB, with a main span of 1,088 m, is the longest cable-stayed bridge in China(Mao *et al.* 2018b). Comprehensive SHM system has been installed on this bridge since its completion in 2008. Among the massive sensors, 7 pairs of dual-axis accelerometers are installed at the bridge deck, and two dual-axis accelerometers are installed at the top of the two towers. The general view of SCB and the installed accelerometers is shown in Fig. 8. For more detailed information about the whole SHM system of SCB, one can refer to Wang *et al.* (2016).

Based on AMPIM illustrated in Fig. 1, the recorded vertical acceleration responses (sampling frequency is 20 Hz) of the bridge deck for 20 days (Jan. 1st - Jan. 20th, 2010) are utilized to identify the modal parameters. The time interval $\bigtriangledown t$ is set as 1 hour, and the total number of time segments for the selected 20 days is 480. The identified modal frequencies, shown in Fig. 9, are utilized to validate the feasibility of GMM for field monitoring application.

In real application, the minimum system operation time T_o is critical important for determining the baseline list. If T_o is too small, indicating the selected time segments are insufficient, some important modes may not be included in baseline list. However, the computation burden will decrease the efficiency of the tracking method when T_o is large. Therefore, different T_o values are applied to investigate their influence on the obtained baseline list using GMM. For that purpose, 10 datasets are generated with the corresponding T_o values defined as 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 7 days, 8 days, 9 days, and 10 days, respectively. These datasets consist of continuously identified modal parameters sampled from the total 480 segments (20 days) as shown in Fig. 9. Based on the GMM fitted results, BIC values are calculated using Eq. (18) and those for different datasets are shown in Fig. 10.

As shown in Fig. 10, the BIC values of the selected 10 datasets of SCB show similar variation trend with those of the numerical case (shown in Fig. 6), i.e., sharply decreasing at the beginning and then slowly increasing to a relatively stable value. Results show that the selected optimal parameters based on BIC values may not be the same for different datasets. For example, the optimal parameters for the dataset (T_o = 1 day) are 9 components with tied covariance matrices, while those for the dataset (T_o = 7 days) are 13 components with spherical covariance matrices. Some observed outliers during long-term monitoring process (as shown in Fig. 9) changed the data feature and led to the variation of obtained group number. The number of elements in each classified group of the dataset (T_o =7 days) as tabulated in Table 2.

As listed in Table 2, one can observe that 4 groups, whose mean values are respectively 0.2386 Hz, 0.3742 Hz, 0.4246 Hz, and 0.5847 Hz, include obviously fewer modes than those in other 9 groups and are assumed as spurious ones. The spurious modes can easily be filtered out using a predefined threshold, i.e., minimum number of elements, which is usually set as one fifth of the number of time segments, i.e., $T_o/(5\bigtriangledown t)$. In this scenario, the threshold is equal to 33.6 (168 divided by 5), and 4 groups with elements fewer than 33.6 are assumed as spurious modes.

In addition, the minimum system operation time T_o are required to define the minimum size of the dataset to generate representative candidates for baseline updating using GMM. This problem can easily be solved by setting T_o as 1 day because that we have found that GMM is able to classify the same modes into one group even when T_o is 1 day. For field monitoring, T_o can be set as 2 days, as a result, more representative statistical results of the candidate modal parameters can be obtained for determining and updating the baseline list.

Therefore, with the combination of GMM and BIC, one can automatically obtain the candidates (mean and standard deviation) for baseline list from datasets with different segments.

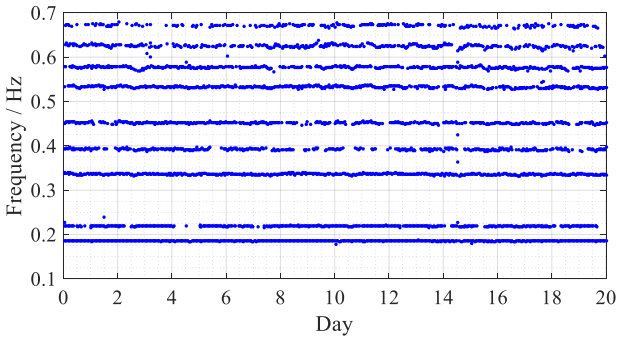


Fig. 9 Identified modal frequencies of SCB (Jan. 1- Jan. 20, 2010)

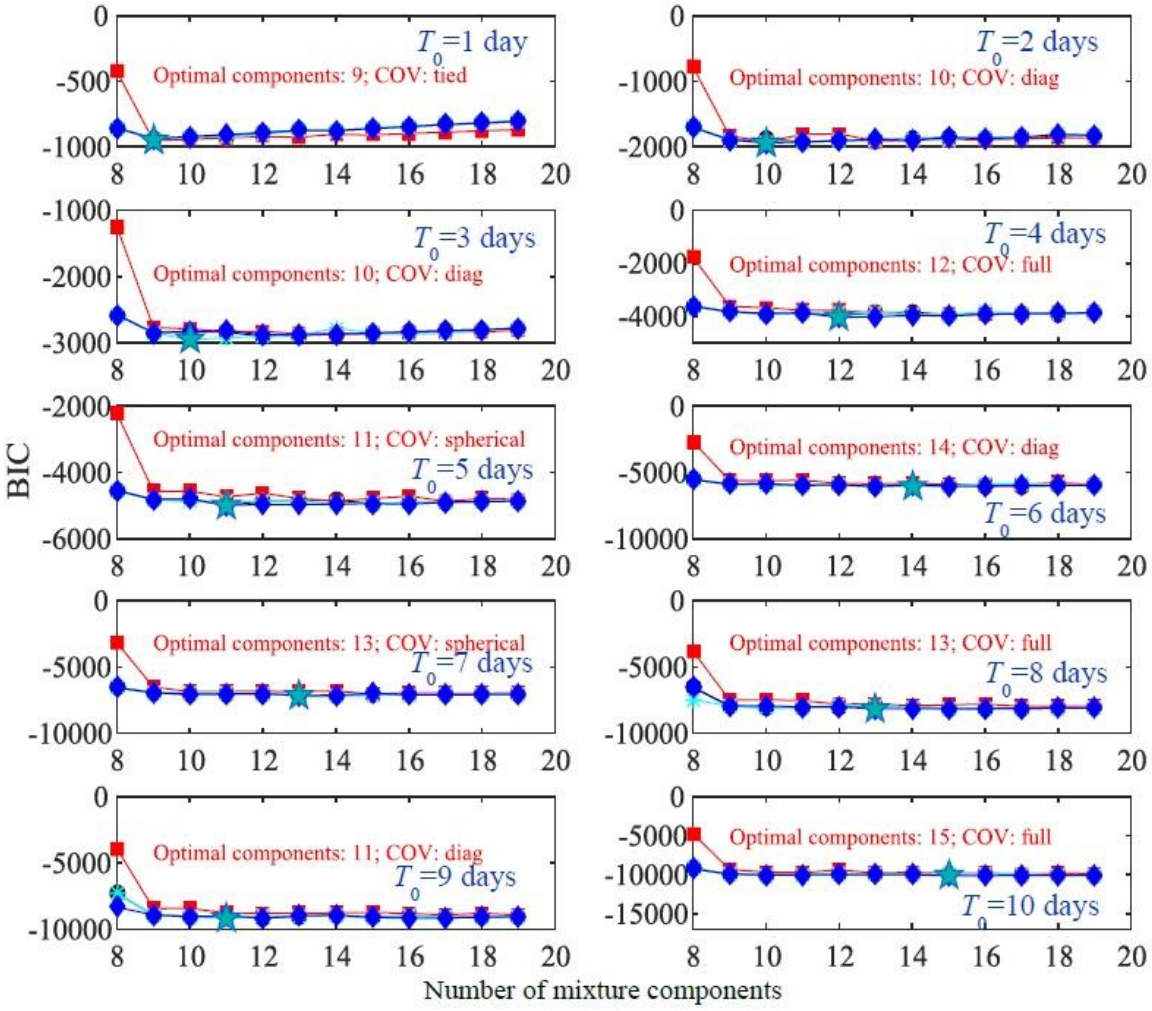


Fig. 10 GMM-based BIC values using GMM for different T_o values

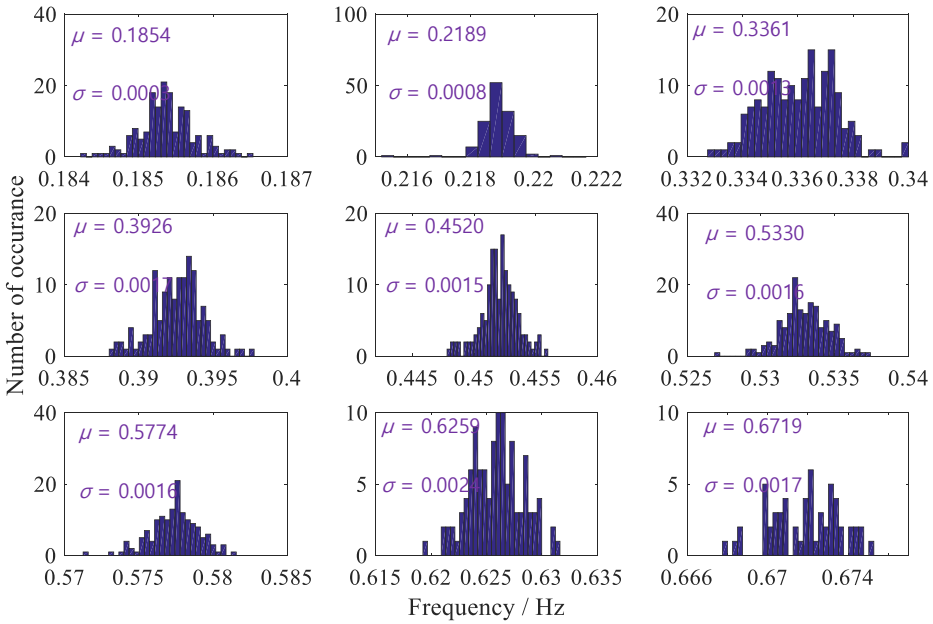


Fig. 11 The histograms of the obtained 9 groups of modal frequencies

Table 3 Baseline list of modal properties determined by GMM

Order	1	2	3	4	5	6	7	8	9
Frequency / (Hz)	0.1854	0.2189	0.3361	0.3926	0.4520	0.5330	0.5774	0.6259	0.6719
Damping ratio / (%)	0.76	0.74	2.30	1.95	1.23	1.40	1.11	1.43	1.60
Direction	V	V	V	V	V	V	T	T	V
shape	S	Anti-S	S	Anti-S	S	S	-	-	S
Notation	VS1	VAS1	VS2	VAS2	VS3	VS4	T1	T2	VS5

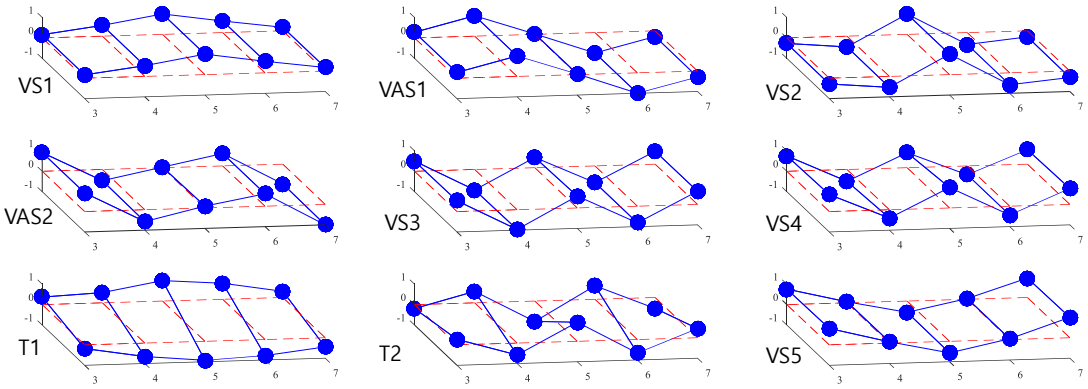


Fig. 12 The baseline mode shapes for automated modal tracking of SCB

The histograms of the remaining 9 groups of modal frequencies are shown in Fig. 11. The baseline list of the modal properties, for the mode tracking process, can be determined according to the statistics results of modes in each group and is tabulated in Table 3. Besides, the baseline mode shapes for automated modal tracking of SCB are shown in Fig. 12.

5. Application of the proposed method during typhoon events

5.1 Description of typhoon events

SCB is situated at the east coast of the China where strong monsoons and typhoons take place frequently. Timely acquisition of structural modal parameters is one of the most important aspects for SHM of the large-scale cable-stayed bridge. Therefore, the proposed automated mode tracking method is required to works effectively during such windy scenarios. In this section, the recorded acceleration responses during August 2012 are applied to validate the efficiency of the proposed method. Three typhoon events, namely as, Damrey, Haikui, and Bolaven, went through the east coast of China and had considerable influence on SCB during that time (Wang *et al.* 2019). The moving paths of three typhoons are exhibited in Fig. 13.

As shown in Fig. 13, typhoon Damrey was born at the northeast sea of Territory of Guam on July 27, 2012 (UTC+8) and decreased to tropical depression at Bohai sea on Aug. 4, 2012 (UTC+8). Typhoon Haikui was born at southeast sea of Iwo Jima on Aug. 1, 2012 (UTC+8) and downgraded to tropical storm at Chizhou City, Anhui Province by 12:00 on Aug. 9, 2012 (UTC+8). Typhoon Bolaven emerged at southwest sea of Territory of Guam on Aug. 17, 2012 (UTC+8) and decreased to tropical storm on Aug. 29, 2012 (UTC+8). The closest distances between the typhoon eye and SCB were respectively 255 km (Damrey), 210 km (Haikui), and 360 km (Bolaven), respectively.

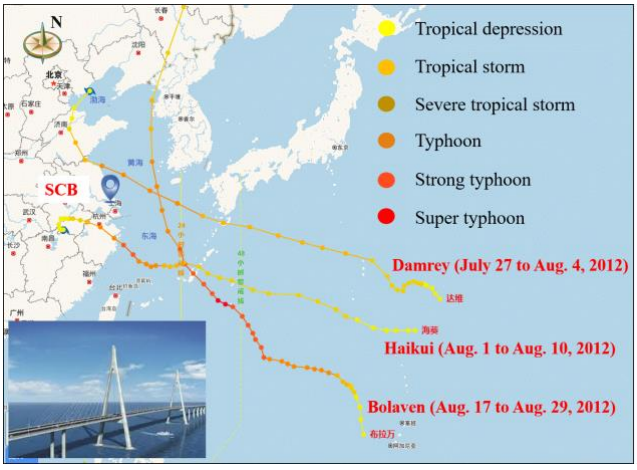


Fig. 13 Moving paths of typhoon Damrey, Haikui, and Bolaven

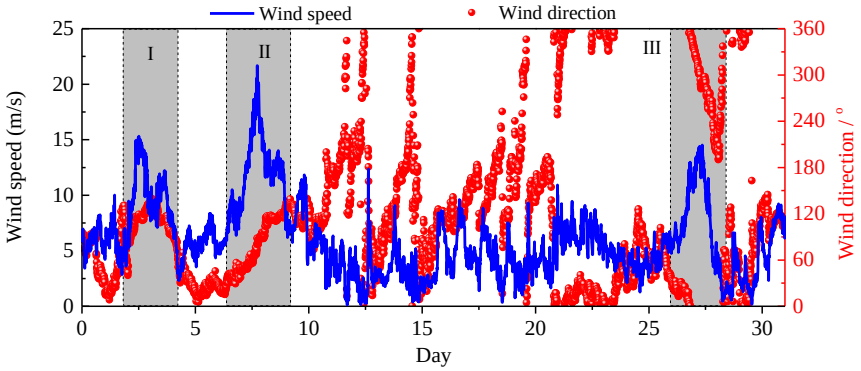


Fig. 14 10-minute average wind speed at the mid span of SCB

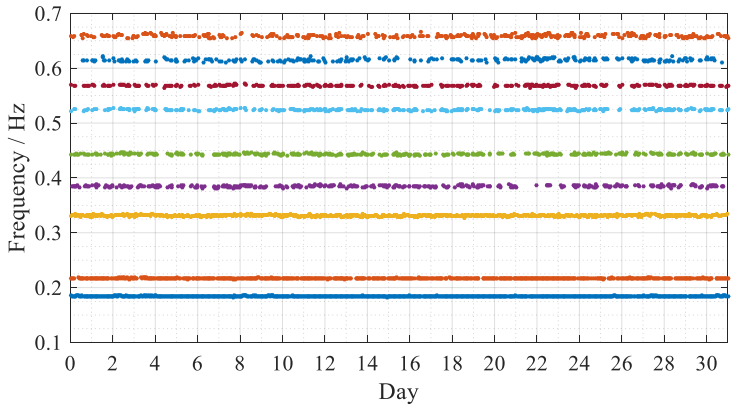


Fig. 15 Tracked modal frequencies of SCB during Aug. 2012

Fig. 14 exhibits the 10-minute average wind speed (mean wind speed occurring over a 10-minute period) at the mid-span of SCB during August 2012 (31 days in total). One can clearly observe that SCB was within the influence scope of the strong typhoons since maximum 10-minute average wind speed for the three typhoons are respectively 15.5 m/s (Damrey), 24.3 m/s (Haikui), 14.8 m/s (Bolaven). In such windy disasters, the bridge owners need to obtain the timely information of the bridge condition for risk evaluation and decision making, which is the motivation to develop automated modal identification and mode tracking methods.

5.2 Tracked modal parameters

Based on the recorded acceleration responses of the bridge deck during August 2012, the structural modal parameters are continuously identified using AMPIM by setting ∇t equal to 1 hour. The minimum system operation time T_o is set as 2 days. Based on the identified modal parameters during the first 2 days, the initial baseline list of modal properties is determined using GMM and BIC. The evolution of modal parameters during the first 2 days was tracked using the initial baseline list. After that, for every new time step, identified modal parameters are firstly included to update the baseline list and then paired to the updated list (mode tracking). The tracked modal frequencies and damping ratios of the deck of SCB during August are shown in Figs. 15 and 16, respectively. Results show that, the modal parameters cannot always be identified during the long-term monitoring process. To be specific, the identifiable segments for high frequency modes (i.e., T2 is 376 and VS5 is 485) are much smaller than the total time segments (744). These deficient modes cannot be identified during the periods with low level signal-to-noise ratio of wind excitation in high frequency range (Ni *et al.* 2015). Whatever, the proposed method can automatically track the evolution of each mode with a good reliability, and all the identified outliers can be filtered out successfully.

As shown in Fig. 15, the modal frequencies of SCB keep relatively stable, although strong typhoons had significantly influence on this bridge during these 31 days (Wang *et al.* 2018). Besides, similarly with the results by Hu (2011), the variation of the high frequency modes is obviously larger than that of the low frequency modes.

The damping ratios, featured as more scattered than natural frequencies, exhibit different variation features. As shown in Fig. 16, the damping ratios fluctuated in a fixed range for most of normal operation time. However, during the typhoon events, the damping ratios increased to a much higher level. Particularly, three peaks of the damping ratios

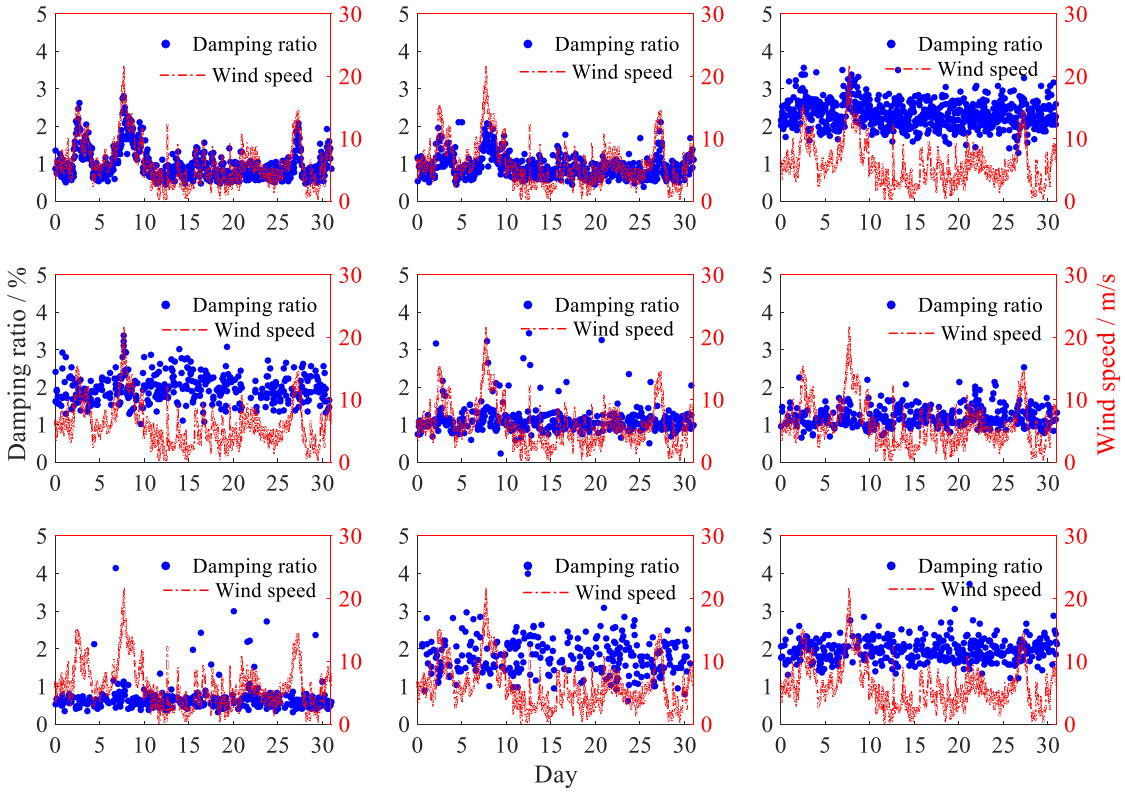
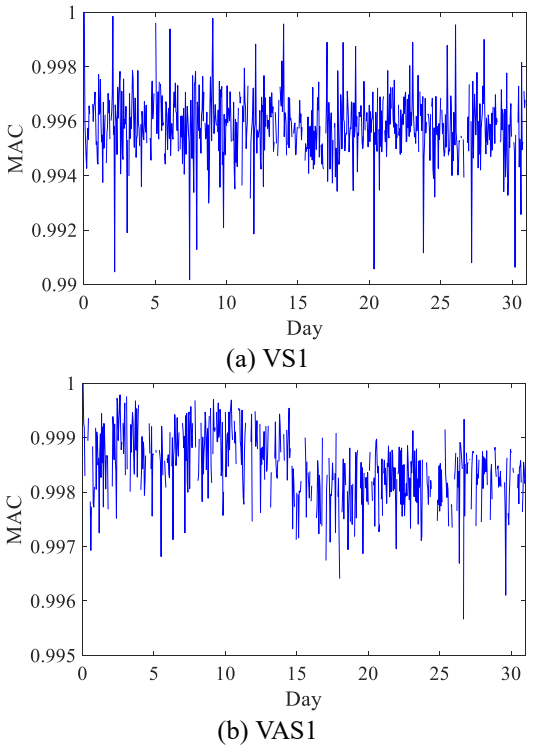


Fig. 16 Tracked damping ratios of SCB during Aug. 2012

can be observed clearly during typhoon Damrey, Haikui, and Bolaven, for the first two modes (VS1 and VAS1). For example, the damping ratios of VS1 varied between 0.60% and 1.30% during regular operating time, however, the maximum values increased to 2.81%, 2.92%, and 2.05% due to typhoon events. Structural damping generally includes contributions from two parts: energy dissipation inside and energy dissipation outside (Tamura and Suganuma 1996). For a long-span cable-stayed bridge (SCB in this case), the structural damping mainly stems from the inside energy dissipation including internal friction, vibration-control dampers, and plastic deformation of structural components, since the outside energy dissipation including aerodynamic damping at the experienced wind velocities is relatively insignificant (Wang *et al.* 2018). With the augment of vibration level, the dissipative energy from the towers, cables, dampers, and girders increases significantly. This might explain why the damping ratios of the first two modes increases with the augment of vibrations. On the other hand, the damping ratios of other modes don't show similar significant increasement for large vibration. The low-frequency modes contribute to most of the vibration responses instead of the high frequency modes due to low-frequency nature of wind excitation. Hence, this might explain that the increase amplitudes of the damping ratios for lower modes are much larger than those for high modes.



(a) VS1

(b) VAS1

Fig. 17 MAC values of the tracked mode shapes of SCB during Aug. 2012

In addition, the MAC values between the tracked and initial modes are calculated to show the evolution of mode shapes of the girder of SCB. All the MAC values are centered at in a small range between 0.99 and 1.0; the MAC values of VS1 and VAS1 are plotted in Fig. 17 to show their variation features. The tracked mode shapes keep relatively stable and show a much weaker relationship with the environmental factors than the frequencies and damping ratios.

6. Conclusions

In this study, an automated mode tracking method was proposed to automatically track the identified modal parameters of the long-span bridge; the proposed method can provide a practical tool for long-term structural health monitoring. Gaussian mixture model (GMM), a popular unsupervised machine learning method, was embedded in the proposed method to automatically determine and update baseline list according to the newly identified modal parameters in the dataset. On that basis, the automated mode tracking methodology is established with two major functions, i.e., updating the baseline modes and tracking the evolutional modal parameters.

With the combination of GMM and Bayesian information criteria (BIC), the candidates for the baseline modal parameters can be automatically determined and updated. However, two parameters, i.e., the minimum system operation time T_o and the minimum number of elements for the remaining clusters, are required to be defined. The minimum system operation time is applied to measure the minimum size of the dataset to generate enough representative candidates for baseline updating using GMM. As usual, T_o can be set as 2 days, as a result, ensuring an adequate quantity of examples in the dataset. The minimum number of elements is a threshold utilized to check the number of elements in the GMM-based clusters; these clusters, including fewer components than the threshold, are assumed to be spurious. The threshold is usually set as one-fifth of the number of time segments, i.e., $T_o/(5\bigtriangledown t)$.

The numerical and field monitoring tests were carried out to validate the validity of the proposed method. Results show that the spurious modes identified in the previous step can be successfully filtered out, such that, the proposed method can automatically track the evolution of modal parameters with a good reliability, even under strong wind events.

It is believed that the automated mode tracking method could be regarded as useful tool to provide timely information of the bridge structure during long-term operation. Insight into the modal parameters during long-term monitoring process will be conducted in future research to provide references for damage detection and condition assessment of the in-service bridge.

Acknowledgments

The authors would like to gratefully acknowledge the support from the National Natural Science Foundation of China (No. 51722804), the National Basic Research Program of China (973 Program) (2015CB060000), the Jiangsu Provincial Key Research and Development Program (BE2018120), the Project of Science and Technology Research and Development Program of China Railway Corporation (2017G002-K), the Scientific Research Foundation of Graduate School of Southeast University (YBJJ1761), and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX17_0127). The first author would also like to acknowledge the support of National Project funded by the China Scholarship Council (201706090073).

References

Annamdas, V.G.M., Bhalla, S. and Soh, C.K. (2016), "Applications of structural health monitoring technology in Asia", *Struct. Health Monit*., **16**(3), 324-346. DOI:10.1177/1475921716653278.

Asadollahi, P. and Li, J. (2017), "Statistical analysis of modal properties of a cable-stayed bridge through long-term wireless structural health monitoring", *J. Bridge Eng*., **22**(9), 04017051. DOI:10.1061/(ASCE)BE.1943-5592.0001093.

Bishop, C.M. (2006), *Pattern recognition and machine learning*. Singapore, Springer Science.

Brownjohn, J., Magalhaes, F., Caetano, E. and Cunha, A. (2010), "Ambient vibration re-testing and operational modal analysis of the Humber Bridge", *Eng. Struct.*, **32**(8), 2003-2018. DOI:10.1016/j.engstruct.2010.02.034.

Cabboi, A., Magalhães, F., Gentile, C. and Cunha, Á. (2017), "Automated modal identification and tracking: Application to an iron arch bridge", *Struct. Control Health Monit*., **24**(1), e1854. DOI:10.1002/stc.1854.

Cao, Z. and Wang, Y. (2014), "Bayesian model comparison and selection of spatial correlation functions for soil parameters", *Struct. Saf.*, **49**, 10-17. DOI:10.1016/j.strusafe.2013.06.003.

Cardoso, R., Cury, A. and Barbosa, F. (2017), "A robust methodology for modal parameters estimation applied to SHM", *Mech. Syst. Signal Pr*., **95**, 24-41. DOI:10.1016/j.ymssp.2017.03.021.

Demarie, G.V. and Sabia, D. (2018), "A machine learning approach for the automatic long-term structural health monitoring", *Struct. Health Monit*., 147592171877919. DOI:10.1177/1475921718779193.

El-Kafafy, M., Devriendt, C., Guillaume, P. and Helsen, J. (2017),. "Automatic tracking of the modal parameters of an offshore wind turbine drivetrain system", *Energies,* **10**(4), 574. DOI:10.3390/en10040574.

Feng, M.Q. and Bahng, E.Y. (1999), "Damage assessment of jacketed RC columns using vibration tests", *J. Struct. Eng*., **125**(3), 265-271. DOI:10.1061/(ASCE)0733-9445(1999)125:3(265).

Hartigan, J.A. and Wong, M.A. (1979), "Algorithm AS 136: A k-means clustering algorithm", *J. Royal Statistical Society. Series C (Applied Statistics)*, **28**(1), 100-108. DOI:10.2307/2346830

Hu, W.H. (2011), *Operational modal analysis and continuous dynamic monitoring of footbridges.* (Ph. D.), University of Porto, Porto.

Huang, Y., Englehart, K.B., Hudgins, B. and Chan, A.D. (2005), "A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses", *IEEE T. Bio-Med. Eng.*, **52**(11), 1801-1811. DOI:10.1109/TBME.2005.856295.

Jeffreys, H. (1998), *The theory of probability*. Oxford: Oxford University Press.

Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R. and Wu, A.Y. (2002), "An efficient k-means clustering algorithm: Analysis and implementation", *IEEE T. Pattern Anal. Machine Intell.*, (7), 881-892. DOI:10.1109/TPAMI.2002.1017616.

Ko, J.M. and Ni, Y.Q. (2005), "Technology developments in structural health monitoring of large-scale bridges", *Eng. Struct.*, **27**(12), 1715-1725. DOI:10.1016/j.engstruct.2005.02.021.

Mao, J.X., Wang, H., Feng, D.M., Tao, T.Y. and Zheng, W.Z. (2018a), "Investigation of dynamic properties of long-span cable-stayed bridges based on one-year monitoring data under normal operating condition", *Struct. Control Health Monit.*, **25**(5), e2146. DOI:10.1002/stc.2146.

Mao, J.X., Wang, H., Fu, Y.G. and Spencer, Jr., B.F. (2019), "Automated modal identification using principal component and cluster analysis: Application to a long-span cable-stayed bridge", *Struct. Control Health Monit.*, e2430. DOI:10.1002/stc.2430.

Mao, J.X., Wang, H. and Li, J. (2018b). "Fatigue reliability assessment of a long-span cable-stayed bridge based on one-year monitoring strain data", *J. Bridge Eng.*, **24**(1), 05018015. DOI:10.1061/(ASCE)BE.1943-5592.0001337.

Moon, T.K. (1996), "The expectation-maximization algorithm", *IEEE Signal Proc. Mag.*, **13**(6), 47-60. DOI:10.1109/79.543975

Ni, Y., Wang, Y. and Xia, Y. (2015), "Investigation of mode identifiability of a cable-stayed bridge: comparison from ambient vibration responses and from typhoon-induced dynamic responses", *Smart Struct. Syst.*, **15**(2), 447-468. DOI:10.12989/sss.2015.15.2.447.

Ou, J.P. and Li, H. (2010), "Structural health monitoring in mainland China: review and future trends", *Struct. Health Monit.*, **9**(3), 219-231. DOI:10.1177/1475921710365269.

Peeters, B. and De Roeck, G. (2001), "Stochastic system identification for operational modal analysis: A review", *J. Dynamic Systems, Measurement, and Control*, **123**(4), 659. DOI:10.1115/1.1410370.

Reynders, E., Houbrechts, J. and De Roeck, G. (2012). "Fully automated (operational) modal analysis", *Mech. Syst. Signal Pr.*, **29**(2012), 228-250. DOI:10.1016/j.ymssp.2012.01.007.

Reynders, E., Wursten, G. and De Roeck, G. (2014), "Output-only structural health monitoring in changing environmental conditions by means of nonlinear system identification", *Struct. Health Monit.*, **13**(1), 82-93. DOI:10.1177/1475921713502836.

Reynolds, D.A., Quatieri, T.F. and Dunn, R.B. (2000), "Speaker verification using adapted Gaussian mixture models", *Digital Signal Process.*, **10**(1-3), 19-41. DOI:10.1006/dspr.1999.0361.

Salehi, H. and Burgueño, R. (2018), "Emerging artificial intelligence methods in structural engineering", *Eng. Struct.*, **171**, 170-189. DOI:10.1016/j.engstruct.2018.05.084.

Soyoz, S. and Feng, M.Q. (2009), "Long-term monitoring and identification of bridge structural parameters", *Comput.-Aided Civil Infrastruct. Engi.*, **24**(2), 82-92. DOI:10.1111/j.1467-8667.2008.00572.x.

Spencer, Jr., B.F. and Nagarajaiah, S. (2003), "State of the art of structural control", *J. Struct. Eng.*, **129**(7), 845-856.

Tamura, Y. and Suganuma, S.Y. (1996), "Evaluation of amplitude-dependent damping and natural frequency of buildings during strong winds", *J. Wind Eng. Ind. Aerod.*, **59**(2-3), 115-130. DOI:10.1016/0167-6105(96)00003-7.

Verboven, P., Parloo, E., Guillaume, P. and Van Overmeire, M. (2002), "Autonomous structural health monitoring-part I: modal parameter estimation and tracking", *Mech. Syst. Signal Pr.*, **16**(4), 637-657. DOI:10.1006/mssp.2002.1492.

Wang, H., Mao, J.X. and Spencer, Jr., B.F. (2019), "A monitoring-based approach for evaluating dynamic responses of riding vehicle on long-span bridge under strong winds", *Eng. Struct.*, **189**, 35-47. DOI:10.1016/j.engstruct.2019.03.075.

Wang, H., Tao, T., Gao, Y. and Xu, F. (2018), "Measurement of wind effects on a kilometer-level cable-stayed bridge during typhoon Haikui", *J. Struct. Eng.,* **144**(9), 04018142. DOI:10.1061/(ASCE)ST.1943-541X.0002138.

Wang, H., Tao, T., Li, A. and Zhang, Y. (2016), "Structural health monitoring system for Sutong cable-stayed bridge", *Smart Struct. Syst.*, **18**(2), 317-334. DOI:10.12989/sss.2016.18.2.317.

Wasserman, L. (2000), "Bayesian model selection and model averaging", *J. Math. Psychol.*, **44**(1), 92-107. DOI:10.1006/jmps.1999.1278.

Wold, S., Esbensen, K. and Geladi, P. (1987), "Principal component analysis", *Chemometrics and intelligent laboratory Systems*, **2**(1-3), 37-52.

Zhou, G.D., Yi, T.H., Xie, M.X. and Li, H.N. (2017), "Wireless sensor placement for structural monitoring using information-fusing firefly algorithm", *Smart Mater. Struct.*, **26**(10), 104002.

Zhou, G.D., Yi, T.H., Zhang, H. and Li, H.N. (2015), "Energy-aware wireless sensor placement in structural health monitoring using hybrid discrete firefly algorithm", *Struct. Control Health Monit.*, **22**(4), 648-666. DOI:10.1002/stc.1707.

HJ