

Sensor fault diagnosis for bridge monitoring system using similarity of symmetric responses

Xiang Xu¹, Qiao Huang^{*1}, Yuan Ren¹, Dan-Yang Zhao¹ and Juan Yang²

¹School of Transportation, Southeast University, Sipailou, Xuanwu District, Nanjing 210-096, People's Republic of China

²Department of Engineering, Nanjing No.3 Yangtze River Bridge Ltd, Tianfu Road, Pukou District, Nanjing 211-808, People's Republic of China

(Received June 11, 2018, Revised January 29, 2019, Accepted January 30, 2019)

Abstract. To ensure high quality data being used for data mining or feature extraction in the bridge structural health monitoring (SHM) system, a practical sensor fault diagnosis methodology has been developed based on the similarity of symmetric structure responses. First, the similarity of symmetric response is discussed using field monitoring data from different sensor types. All the sensors are initially paired and sensor faults are then detected pair by pair to achieve the multi-fault diagnosis of sensor systems. To resolve the coupling response issue between structural damage and sensor fault, the similarity for the target zone (where the studied sensor pair is located) is assessed to determine whether the localized structural damage or sensor fault results in the dissimilarity of the studied sensor pair. If the suspected sensor pair is detected with at least one sensor being faulty, field test could be implemented to support the regression analysis based on the monitoring and field test data for sensor fault isolation and reconstruction. Finally, a case study is adopted to demonstrate the effectiveness of the proposed methodology. As a result, Dasarathy's information fusion model is adopted for multi-sensor information fusion. Euclidean distance is selected as the index to assess the similarity. In conclusion, the proposed method is practical for actual engineering which ensures the reliability of further analysis based on monitoring data.

Keywords: structural health monitoring; sensor fault diagnosis; similarity; symmetric structure responses; multi-sensor information fusion; evidential reasoning

1. Introduction

SHM techniques have been increasingly applied in bridge engineering to ensure the safe and sustainable operation of bridges (Ko and Ni 2005). Lots of effort has been made to investigate the condition assessment or damage identification based on mining of data from the SHM systems (Li *et al.* 2012, Yi *et al.* 2016); however, studies on the precision or creditability of monitoring data have been rarely conducted (Hernandez-Garcia and Masri 2013). In reality, sensor faults have become more frequent issues than structural damages owing to inefficient manufacture, harsh operation environment and performance degradation (Chang *et al.* 2017, Huang *et al.* 2015). Therefore, it is essential to study the diagnosis of sensor faults to provide high quality data for further data mining.

The process of fault diagnosis usually includes fault detection, fault isolation and fault reconstruction (Yi *et al.* 2017): (i) fault detection addresses determination of the presence of faults in a sensor system; (ii) fault isolation tends to locate the faulty component and identify the type of fault; and (iii) fault reconstruction aims to correct distorted signals of faulty sensors. Methods for sensor fault diagnosis generally take root in the redundancy of information and

can be categorized into model-based and data-driven types according to the sources of redundant information (Reppa *et al.* 2014). The former methods utilize the redundancy of information provided by mathematical models such as finite element models; while the latter relies on the monitoring data from remaining or extra sensors. The performance of the model-based methods is completely dependent on the accuracy of the established mathematical model. Unfortunately, the mathematical model can hardly capture the realistic behavior of the system in view of the effects from uncertainties of model parameters, complexity of system dynamics and modeling of damages occurring in the system (Reppa *et al.* 2014). This study focuses on the data-driven methods for fault diagnosis of sensors.

Data-driven methods then can be grouped into the physical redundancy-based and multivariate statistical methods (Yi *et al.* 2017, Reppa *et al.* 2014). The former requires a field test to be conducted using temporary co-located sensors for measuring the same response quantity as done by the existing sensor, while the latter utilizes the valuable redundant information provided by the existing sensor network. The physical redundancy-based method was not widely used before due to the high installation cost, however, the evolution of micro-technology in recent decades has reduced the size and fabrication cost of sensors, thereby making the method become more promising in engineering practice. Considering a large number of sensors installed in bridge structures, it is still unacceptable to inspect all the sensors blindly using the physical

*Corresponding author, Professor
E-mail: qhuanghit@126.com

redundancy-based method.

The multivariate statistical approach relies existing multiple heterogeneous sensors (which are not necessarily co-located) on collecting measurements having redundant information. Among these methods, the principle component analysis (PCA) is one of the most popular techniques for sensor fault diagnosis (Kerschen *et al.* 2004), including two mostly proposed indices: the D-statistic and Q-statistic. The D-statistic is a Mahalanobis-like measure of variations of measured variables in the latent-variable subspace, whereas Q-statistic measures the discrepancy between the measured variables and the latent-variable model (Alcala and Qin 2011). The PCA approach is effective in SHM field since measured structural responses from sensors follow a multivariate process and statistical correlations among these measurements can be easily modeled by PCA. Whereas, previous PCA models did not quantify the fault sensitivity to each principal direction when only the normal monitoring data are obtained, which may weaken the ability of fault detection. To overcome the limitations, the weighted principle component analysis was developed for statistical monitoring with improved fault-detection capability using fault-sensitive index (Huang *et al.* 2017). In addition, other methods, including the parity equation-based method, minimum square error estimation-based method and independent component analysis-based method, were also sought for sensor fault detection (Huang *et al.* 2016, Yi *et al.* 2017).

These aforementioned methods were proven to be effective in detecting single-fault sensors; however, multi-fault diagnosis models (which is a common issue in practice) are rarely investigated. A model based on a combination of PCA and neural networks was developed for multi-fault diagnosis of sensors (Zhu *et al.* 2009). Unfortunately, such method requires a large amount of data in training the model and is difficult to be implemented in practical applications. Moreover, the structure is usually assumed being damage-free in developing the scheme for sensor fault diagnosis, however, in reality there could be a scenario that both structural damages and sensor faults coincide and the response measurements by sensor failure are coupled with that resulting from the damages of structure. The sensor network is usually modelled as a Gauss process to distinguish between sensor fault and structural damage based on the fact that impact of sensor faults on structural responses is local while that by structural damage is global (Kullaa 2011). Nevertheless, such method has the constraint (and potential issues in practical application) that the number of sensors must be greater than the number of active modes of the structure. Thus, it is challenging to address the multi-fault diagnosis issue and distinguish the abnormality from structural damage and sensor fault.

This paper proposes a practical methodology for sensor fault diagnosis based on the fact that structural responses at symmetric locations should be quite similar with each other due to the symmetry of structure geometry and operational loadings for bridge structures. First, the similarity of symmetric structure responses is discussed using the real-time monitoring data from the studied bridge. For one SHM

system installed on a bridge, all the sensors are initially paired and each pair is usually composed of transverse symmetric sensors. The sensor faults will be detected pair by pair, by which multi-fault diagnosis of sensor systems could be addressed. Next, Dasarathy's multi-sensor information fusion model is used for data fusion, feature fusion and decision fusion. The data fusion deals with preprocessing of the monitoring data in terms of data continuity and trend. The feature fusion is dedicated to evaluating the similarity of the monitoring data of the sensor pair with the similarity index determined by sensitivity analysis. To resolve the coupling response between structural damage and sensor fault, decision fusion is performed to distinguish the response abnormalities resulting from structural damage and sensor fault. If the sensor pair is determined to be the suspected pair, the physical redundancy-based approach can be applied to the sensor pair for isolating and reconstructing the faulty sensor. Finally, a case study is adopted to demonstrate the effectiveness of the proposed methodology by using the cable force and deflection monitoring data from a cable-stayed bridge.

2. Similarity analysis of symmetric responses

Sensor fault diagnosis methods generally rely on the foundation of the redundancy of information. For bridge structures, the structure responses at the transverse symmetric locations tend to have the strongest redundant information due to the symmetry of both structure geometry and major loadings on the bridge, scilicet, the measurements from symmetric sensor pair are expected to be similar or highly correlated. However, some unbalance (asymmetric) loadings such as the traffic and environmental loadings (temperature, wind etc.) and measurement noise may impact the similarity between measurements collected from the symmetric sensor pair. Hence, it is critical to assess the similarity of symmetric responses prior to introducing the proposed methodology for sensor fault diagnosis.

2.1 Influencing factors on similarity

The monitoring data from sensors installed at the transverse symmetric locations are not identical due to various influencing factors: (i) the similarity will be affected by measurement noise which is unavoidable; (ii) the environmental loadings such as the transverse temperature difference across the girder, as well as the unbalance traffic loadings, will certainly result in the dissimilarities between symmetric responses, although the dead loading for a straight type bridge is usually symmetrical; and (iii) the symmetric responses could have a bias due to certain construction factors, for instance, cable force adjustment during the cable-stayed bridge construction period may lead to the differences in monitoring cable forces.

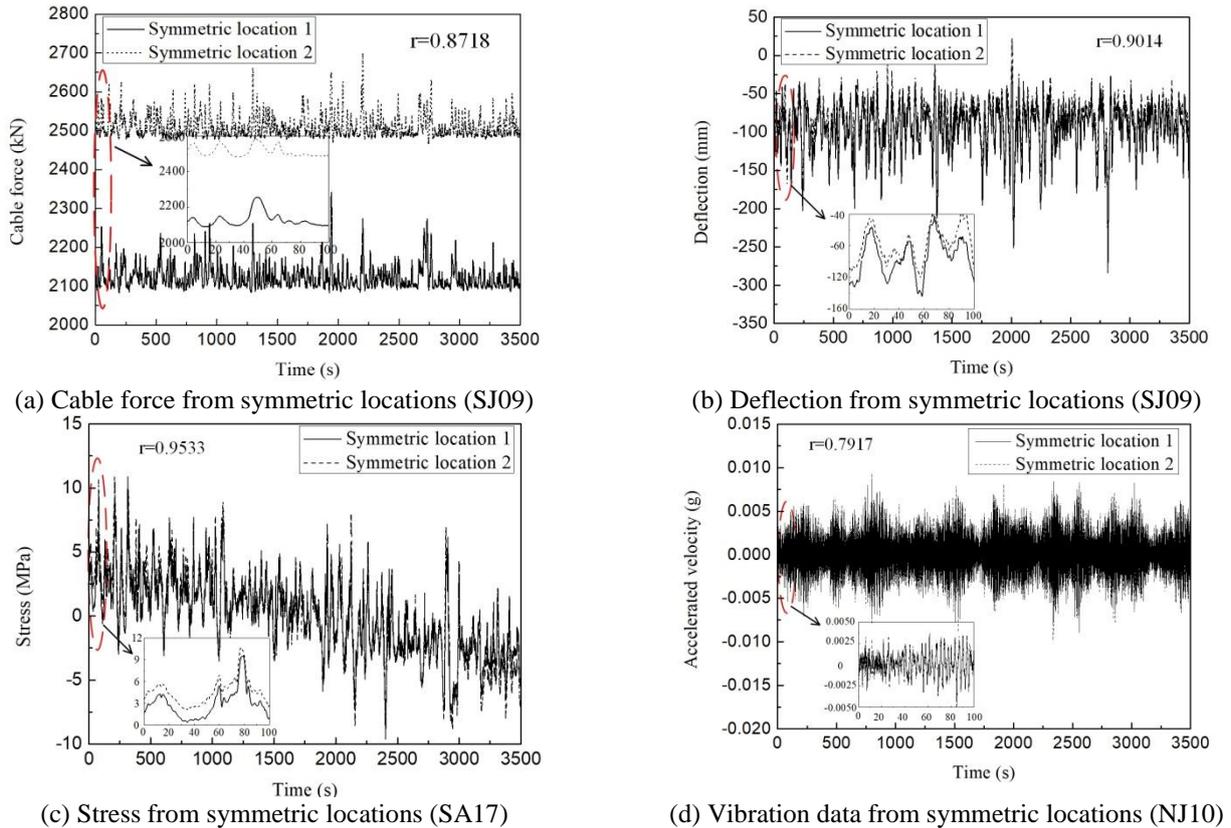


Fig. 1 Correlation analysis for symmetric monitoring data from four typical sensor types

2.2 Similarity evaluation using field monitoring data

To evaluate impact of the influencing factors (as discussed in Section 2.1) on the similarity of symmetric responses, the monitoring data from typical sensors (i.e., the anchor load cell for cable force, connected pipe for deflection, strain gauge for stress and acceleration for vibration) installed on the studied bridge are investigated. The detailed information on the studied bridge and its monitoring system will be illustrated in Section 4.

The monitoring data during an earlier period (8:00 a.m. ~ 9:00 a.m., January 03, 2007) are used for the correlation analysis since the sensors are in good condition at that time. Typical measurements in one-hour window are presented in Fig. 1 for four sensor types, in which the anchor load cell and connected pipe at location SJ09, strain gauge measuring girder stress at location SA17 (sensor No.1 and No.10 on Section B-B of Fig. 8(b)), and acceleration at location SJ10 are selected. For the real-time monitoring absolute values, tiny difference survives in the deflection, stress and acceleration measurements owing to the above mentioned measurement noise, environmental and unbalance traffic loadings, whereas, an apparent bias is in the measurements of cable force as shown in Fig. 1(a) resulting from the cable force adjustment in the construction period. What's more, for the variation trend of each variable that is without the effect of dead loadings and mainly dominated by vehicle loadings, similarity is remarkable for all the four types of sensors which can be visually seen from Fig. 1. The linear

correlation coefficient (r) is used to assess the similarity between the symmetric monitoring data from different sensor types. As a result, 91.9% of the calculated linear correlation coefficients are greater than 0.8000 while the minimum is 0.7523.

Following the above assessment, it can be concluded that the correlation between symmetric monitoring data from these four sensor types is strong. Considering limited types of sensors used in the studied bridge, other types of sensors were not discussed. When the proposed method is applied to other types of sensors, the similarity between the symmetric monitoring data from those sensors should be first assessed.

3. The proposed methodology

Based on the fact that measurements of most sensor types from symmetric locations have remarkable similarity features as being discussed in Section 2, the similarity of symmetric structure responses can be established as an alternative index for sensor fault diagnosis. A systematic sensitivity analysis has been conducted to assess the performance of the possible indices using the simulated sensor fault signals. Unsatisfactory similarity for measurements of sensors at transverse symmetric locations can be attributed to either structural damages or sensor faults. Among the entire sensor zones of the sensing network on the bridge, a sub-zone (termed as target zone,

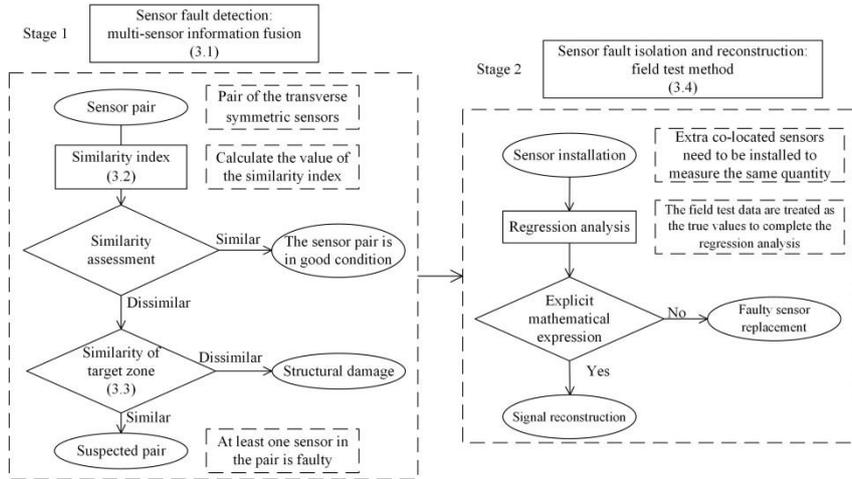


Fig. 2 Two stages of the proposed methodology for diagnosis of sensor fault

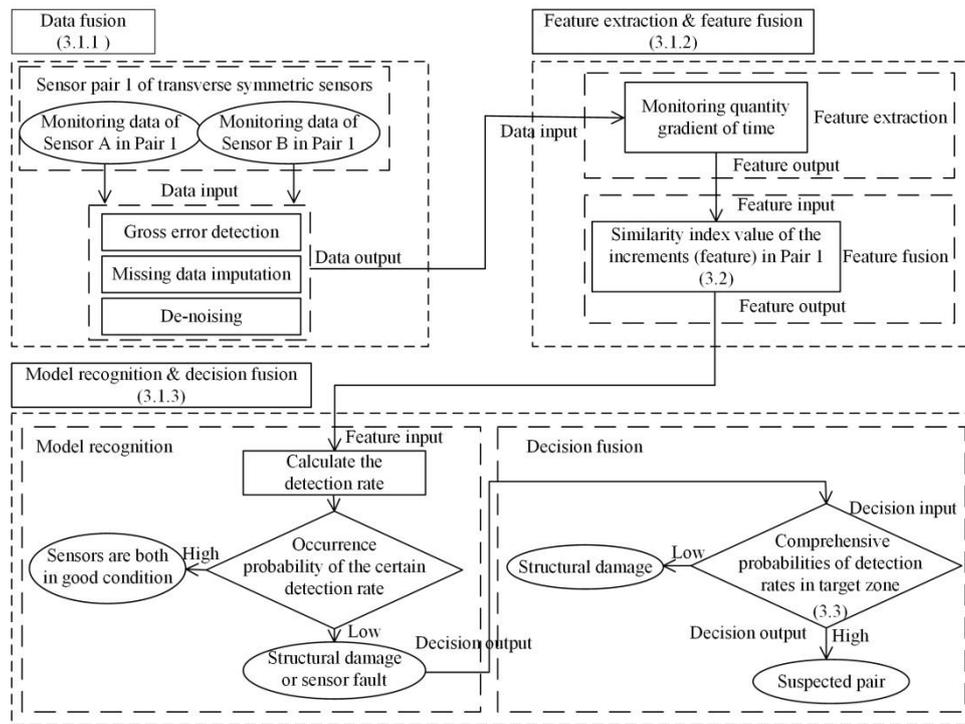


Fig. 3 The Architecture of Dasarthy's input/output information fusion model

with an assembly of sensor pairs to be studied) will be determined and selected to study similarity between responses of sensor pair. Based on the rationale and assumption that structural damages will completely impact the similarity of all sensor responses within the target zone and the probability of a simultaneous failure of all the sensors is very low (Kullaa 2011), the similarity between responses of sensor pairs within the target zone can be evaluated for the purpose of sensor fault diagnosis. If the similarity level of the studied sensor pair is low while the similarity within the target zone is good, it means that at least one sensor in the pair becomes faulty. On the other hand, if the similarity levels for the studied pair and the target zone are both low, it is concluded that there are

structural damages on the bridge. When the suspected sensor pair is detected, field test method may be required (depending on the necessity) for additional verification and confirmation of the faulty sensors. Fig. 2 presents the process of the proposed methodology for sensor fault diagnosis.

3.1 Sensor fault detection: multi-sensor information fusion

Multi-sensor information fusion is used for combining sensor data into a common representational format with three different processing levels according to the stage at which the fusion takes place: signal level, feature level, and

decision level. Dasarathy’s model is one of the most widely used information fusion models (Dasarathy 1994), and it is determined to be more appreciate within the proposed methodology. Based on the Dasarathy’s fusion model, the architecture for sensor fault detection has been developed and presented in Fig. 3.

3.1.1 Data fusion

There are many factors that can affect data collection in long-term monitoring system, including the power system failure, noise, communication malfunctions etc. (Amiri and Jensen 2016). Specifically, the preprocessing techniques include gross error detection, missing data imputation and de-noising. Initially, the gross errors may occur suddenly in the measurement signals at a particular time. A variety of serial elimination strategies have been developed, such as the generalized likelihood ratio test, Bayesian test, and principal component test etc. (Narasimhan and Mah 1987, Tamhane *et al.* 1988, Tong and Crowe 1995). Three-sigma rule is adopted in this study to detect the gross errors owing to its effectiveness and practicality for processing of the real-time mass measurement data. The presence of missing data could dramatically degrade the interpretation results as drawn from the datasets. The methods for missing value imputation can be categorized in two different ways, by either the mathematical or statistical methods or machine learning methods (Amiri and Jensen 2016). Due to the simplicity of the algorithms, the cubic spline imputation technique has been used in this study. It is based on the fitting of cubic polynomials for a series of observed data. Actually, the measured structure responses through the SHM system always accompany with noise. Among many state-of-art de-noising methods such as the moving average and moving differential methods (Qin *et al.* 2012), the wavelet method is used in removing noise.

3.1.2 Feature extraction and feature fusion

The recorded data at transverse symmetric locations are always affected, and maybe greatly different, by various factors such as unbalance loadings, construction factors etc. However, the increments of the measurement responses over time will behave a better trend in describing the similarity of responses than that by the direct measurements.

Therefore, the increment ($k_i = \frac{x_{i+1} - x_i}{t_{i+1} - t_i}$, where x_i is the i^{th}

monitoring datum in the series and t_i denotes the sampling instant.) is determined as the feature for sensor fault detection.

It is essential to develop an effective feature, which should be sensitive to the possible faults, to reveal the measurement response similarity. Typical sensor faults have been summarized and the faults include the bias, drift, precision degradation, gain, and complete failure (Yi *et al.* 2017). Many indicators are able to represent the similarity, including the Euclidean distance, Pearson correlation, slope relational degree and cosine similarity. In this paper, the Euclidean distance is determined as the similarity index based on the sensitivity study in the following Section 3.2.

3.1.3 Model recognition and decision fusion

The value of the similarity index, calculated from the step of feature fusion, should be compared with the threshold value to evaluate the degree of similarity. This paper employs a statistical approach, i.e., the three-sigma rule, to define the threshold. Based on the three-sigma rule, the occurrence probability of an index value outside the threshold is 0.13% for the normal distribution. Aiming to

avoid contingency, the detection rate $\eta_0 = \frac{N_{be}}{N_0}$ (where N_{be}

is the number of similarity index values outside the threshold and N_0 represents the total number of similarity index values.) is proposed to assess the similarity. Assuming that both the structure and sensors are in good condition, the exceedance probability for a given detection rate η_0 can be derived as

$$P(\eta \geq \eta_0) = \sum_{i=0}^{N_{be}} \left[C_{N_0}^i \times (0.13\%)^i \times (1-0.13\%)^{N_0-i} \right] \quad (1)$$

where C is a combination symbol. Based on Eq. (1), the probability density function of the detection rate is formulated. Once the detection rate η_0 is determined, the cumulative distribution model $P(\eta \geq \eta_0)$ is used to describe the probability of the case that the measurements of the symmetric locations are normal.

If the calculated probability value of $P(\eta \geq \eta_0)$ is very low, it means that the structure and sensor system must have some issues, and additionally, it is worthy of determining the cause of the issue, i.e., by the sensor fault or localized structural damage. Based on the rationale and assumption that structural damages will completely impact the similarity of all the sensor responses within the target zone and the probability of a simultaneous failure of all the sensors is very low (Kullaa 2011), the similarity between indices of probability of detection rate derived from the sensor pair (within the target zone) can be evaluated using for sensor fault diagnosis. Specifically, sensor pairs in the target zone are selected and the probability of detection rate for each pair is calculated. The evidential reasoning method is then used to combine each of the probability of detection rate into a comprehensive one to represent the similarity within the target zone (as introduced in Section 3.3). Finally, if the similarity of the studied pair is significantly different from that of the target zone, the sensor pair is judged to be the suspected pair with at least one faulty sensor. If the similarities of the studied sensor pair and target zone are both unsatisfactory, the structure is judged to be with some damages.

3.2 Similarity index

Assuming that the feature vectors of a sensor pair are $\mathbf{K}^A = (k_1^A, k_2^A, \dots, k_n^A)$ and $\mathbf{K}^B = (k_1^B, k_2^B, \dots, k_n^B)$, respectively. A number of existing indicators can be used to evaluate the similarity between these two feature vectors, namely,

Table 1 Mathematical measurement models for seven typical sensor faults

Sensor fault type	Mathematical expression	Description
Bias	$x_i = x_i^* + w_i + C$	Outputs differ from the measurements of the healthy sensor with a constant difference
Drift	$x_i = x_i^* + w_i + C + D \cdot i$	Differences between the outputs and measurements of the healthy sensor vary linearly with time
Gain	$x_i = G(x_i^* + w_i)$	Outputs are proportional to the measurements of the healthy sensor
Precision degradation	$x_i = x_i^* + w_i + s_i$	Outputs are added with a random noise on top of the measurements of the healthy sensor
Complete failure 1: constant	$x_i = C$	Outputs remain constant with time
Complete failure 2: constant with noise	$x_i = C + s_i$	Outputs are combination of the constant and noise
Complete failure 3: bottom noise	$x_i = s_i$	Outputs are noise only

(i) Euclidean distance:

$$d_{Eu}(\mathbf{K}^A, \mathbf{K}^B) = \left[(k_1^A - k_1^B)^2 + (k_2^A - k_2^B)^2 + \dots + (k_n^A - k_n^B)^2 \right]^{0.5}$$

(ii) Pearson correlation:

$$r_{Pe}(\mathbf{K}^A, \mathbf{K}^B) = \frac{n \sum k_i^A k_i^B - \sum k_i^A \sum k_i^B}{\sqrt{n \sum (k_i^A)^2 - (\sum k_i^A)^2} \sqrt{n \sum (k_i^B)^2 - (\sum k_i^B)^2}}$$

(iii) Slope relational degree:

$$r_{AB} = \frac{1}{n-1} \sum_{k=1}^{n-1} \left[1 + \left| \frac{x_k^A}{x_{k+1}^A} - \frac{x_k^B}{x_{k+1}^B} \right| \right]^{-1}$$

(iv) Cosine similarity:

$$\cos(\mathbf{K}^A, \mathbf{K}^B) = \frac{\sum k_i^A k_i^B}{\sqrt{\sum (k_i^A)^2} \sqrt{\sum (k_i^B)^2}}$$

For the above four similarity indexes, if two vectors are extremely similar, values of Euclidean distance tend to approach null while values for Pearson correlation, slope relational degree and cosine similarity are close to unity.

A total of seven typical sensor faults have been summarized, which include the bias, drift, gain, precision degradation, constant, constant with noise, and bottom noise. Let x_i^* represent the true value of the measured quantity and w_i be the measurement noise, the measurement datum by a healthy sensor can be expressed as $x_i = x_i^* + w_i$. Then, the mathematical models for the seven typical sensor faults can be derived and listed in Table 1, where C , D and G are parameters controlling the magnitude of sensor faults and s_i is the parameter for Gaussian random process with zero mean and unknown standard deviation.

Since the real-time monitoring data covering all the sensor fault types are usually not available from the SHM system, simulated signals are constructed and used for sensitivity analysis of similarity. However, the baseline signal $\mathbf{X}^0 = (x_1^0, x_2^0, \dots, x_n^0)$, which is the real-time cable

force monitoring data from a fault-free sensor in the studied bridge as shown in Fig. 4, is used in the sensitivity analysis. Additional signals from faulty sensors are simulated on top of the baseline signal \mathbf{X}^0 based on the mathematical fault models in Table 1. \mathbf{X}^0 has a total of 1000 samples with a frequency of 10Hz and the reliability of \mathbf{X}^0 in representing the outputs of fault-free sensor has been validated. For instance, the simulated signal for sensor drift fault, as shown in Fig. 4, is generated based on the aforementioned baseline signal and its mathematical expression with parameters of $C = 200$ and $D = 0.20$.

Among the seven types of sensor faults, the feature is the same for the measurements from fault-free sensors and that from the sensors with bias based on the mathematical model of bias. Hence, the sensor fault of bias cannot be detected by the proposed feature. However, sensors with the fault of bias are still able to precisely measure the increments of the response quantity, which sometimes are more important than the absolute values.

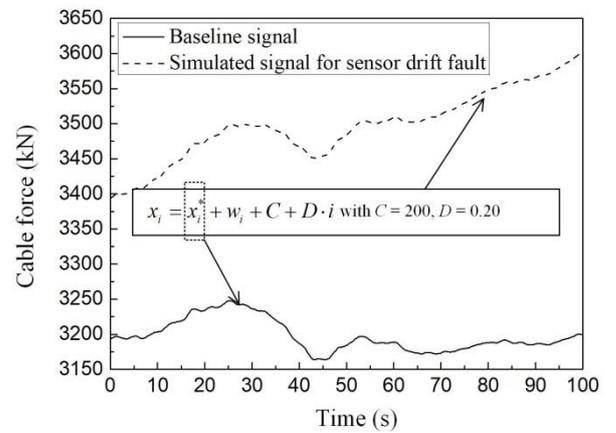


Fig. 4 Signal construction for sensitivity analysis of similarity indexes

Table 2 Different simulated fault magnitudes for the four sensor faults

Fault type	Parameter	Fault magnitude						
Drift	D	0.00	0.20	0.50	1.00	2.00	5.00	
Gain	G	1.00	1.01	1.05	1.10	1.20	1.30	
Precision degradation	σ	1.00	5.00	15.00	20.00	25.00	30.00	
Complete failure 2: constant + noise	σ	1.00	5.00	15.00	20.00	25.00	30.00	

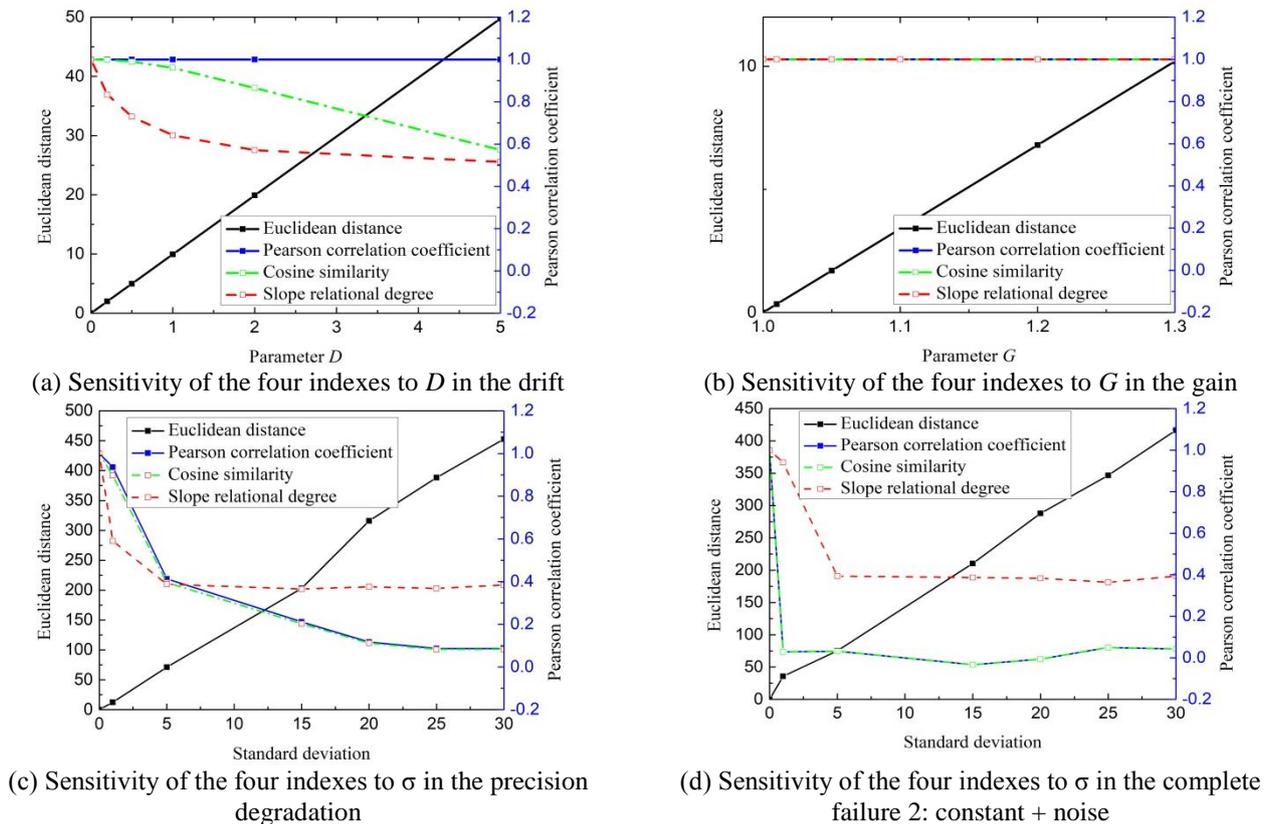


Fig. 5 Sensitivity analysis results for the similarity indexes under various sensor faults

Furthermore, the bias usually does not come alone in bridge health monitoring system but follows the drift fault. In addition, there are three types of the complete failure fault, namely the constant, constant + noise, and noise. Only the fault of constant + noise is detected in this study since faults of the constant and noise can be visually detected in a straightforward way. As a result, a total of four types of sensor faults, including the drift, gain, precision degradation and constant + noise, are included in the following sensitivity studies. These sensor faults are simulated with six different magnitude levels as listed in Table 2, in which parameter σ represents the standard deviation of the noise. Based on the baseline signal of Fig. 4 and parameter values in Table 2, a total of 24 groups of simulated signals are generated for sensitivity analysis of similarity indexes.

To save computation time, all the signals are resampled by a new frequency of 1Hz (the original sampling frequency is 10Hz) before calculating the similarity index. The values of similarity indexes are calculated based on the

fault-free sensor data and the data with specific sensor fault and fault magnitude. The sensitivity analysis results for the similarity indexes under various sensor faults are shown in Fig. 5 (Cosine similarity and slope relational degree refer to the right “Pearson correlation coefficient” coordinate axis).

According to Fig. 5, the indexes of Pearson correlation coefficient, cosine similarity and slope relational degree (especially the Pearson correlation coefficient and cosine similarity) are sensitive to those sensor faults with uncertainties such as precision degradation and complete failure 2. The values of these indexes decrease significantly with a small value of the standard deviation, then remain in a relatively stable level. However, they have a poor sensitivity to those sensor faults such as drift and gain. On the other hand, the index of Euclidean distance has a strong stable sensitivity to all the types of sensor faults, and therefore was used as the similarity index in this proposed methodology.

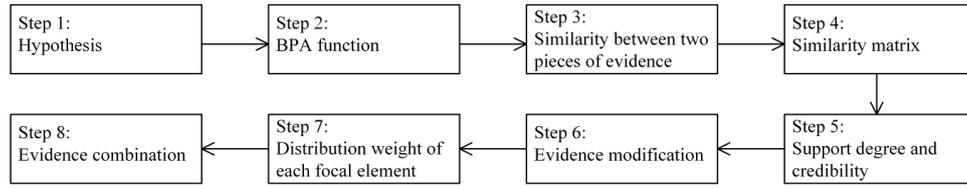


Fig. 6 Steps of the evidence similarity-based method for evidential reasoning in decision level

3.3 Decision fusion level: evidential reasoning

As discussed in Section 3.1.3, the comprehensive similarity index for the target zone is needed to distinguish the dissimilarity owing to sensor fault or localized structural damage. The comprehensive similarity index should be derived from the similarity indices from sensor pairs within the zone. The occurrence probability of detection rate of each sensor pair can be treated as a piece of evidence that supports the similarity of the target zone. These pieces of evidence can be used to derive the comprehensive similarity index, while it is common that conflicts also exist among these pieces of evidence. For instance, the similarities of some pairs in the target zone could be unsatisfactory due to the sensor faults and it is possible that the damage does not impact the similarity of a certain sensor pair in the target zone. It is thus difficult for decision makers to make a correct decisions relying on several conflicting pieces of the evidence. To address the conflict problem, evidential reasoning method is adopted to combine these pieces of evidence into a comprehensive similarity index.

3.3.1 Fundamental concept of evidential reasoning

Evidential reasoning theory, as introduced by Dempster and extended later by Shafer, is concerned with the question of belief in a proposition and systems of propositions (Yang and Xu 2002). Let Θ be a finite non-null set of mutually exclusive alternatives and Θ is defined as the identification framework

$$\Theta = \{F_1, F_2, \dots, F_n\} \tag{2}$$

where F_i is called the single subset of Θ . The basic probability assignment (BPA), which is critical in evidential reasoning, reflects the degree of belief in a hypothesis and satisfies the following relation

$$\begin{aligned} m: 2^\Theta &\rightarrow [0,1] \\ \sum_{A \in \Theta} m(A) &= 1 \\ m(\emptyset) &= 0. \end{aligned} \tag{3}$$

in which BPA reflects the degree of evidence supports for the proposition of A , namely $m(A)$ and \emptyset represents an empty set. The rule of combination (also referred as the orthogonal sum of evidence) is used to aggregate multiple sources information. The typical Dempster's rule of combination is defined as

$$\begin{aligned} m(A) &= (m_1 \oplus m_2)(A) = \frac{1}{1-k} \sum_{B \cap C = A} m_1(B)m_2(C) \\ k &= \sum_{B \cap C = \emptyset} m_1(B)m_2(C). \end{aligned} \tag{4}$$

where k is a normalization constant and termed as the conflict coefficient for measuring the degree of conflict between m_1 and m_2 .

3.3.2 Evidential reasoning in decision level

Considering that the Dempster-Shafer evidence theory cannot handle the data conflicts effectively, an evidence combination method based on the evidence similarity is used to handle evidence with strong conflicts (Bi *et al.* 2017). In particular, if a body of evidence is supported by other bodies of evidence greatly, its credibility degree is higher and the subjective evidence has more impact on the final combination results. On the contrary, if a piece of evidence is always conflicting with the other evidence with high degree, its credibility degree is lower and this evidence should have less effect on the final combination results. The steps of evidence similarity-based method for evidential reasoning in decision level are presented in Fig. 6.

Step 1: Define the hypothesis

The framework of identification in this case is set as

$$\Theta = \{F_1, F_2\} \tag{5}$$

where F_1 is a hypothesis that the similarity of the target zone is satisfactory while F_2 denotes as a hypothesis that the similarity of the target zone is not satisfactory.

Step 2: Deliver the BPA function

The BPA function $m(A)$ describes the proportion of all relevant and available evidence in support of the claim that an element of Θ belongs to the set A . In this case, A can be the hypothesis F_1 indicating the similarity of the target zone. $m(A)$ is the occurrence probability of the detection rate and it describes the similarities of sensor pairs distributed in the target zone. Assuming that there are a total of n adjacent sensor pairs selected into the analysis, there exists n pieces of evidence for the similarity of the target zone, namely

$$[E_1 : m_1(A)], [E_2 : m_2(A)], \dots, [E_n : m_n(A)] \tag{6}$$

where E_i is the evidence i and $m_i(A)$ is the occurrence probability of its detection rate.

Step 3: Calculate the similarity $sim(m_i, m_j)$

The existing algorithms were reviewed in measuring the degree of conflict between two pieces of evidence and proposed a new algorithm for evidence conflict measurement (Qin *et al.* 2012). The similarity between two pieces of evidence is expressed as

$$sim(m_1, m_2) = \frac{\sum_{i=1}^m m_1(F_i) \cdot m_2(F_i)}{\sum_{i=1}^m m_1(F_i)^2 + \sum_{i=1}^m m_2(F_i)^2 - \sum_{i=1}^m m_1(F_i) \cdot m_2(F_i)} \quad (7)$$

The value of evidence similarity of Eq. (7) is between 0 and 1. The bigger this value is, the higher the similarity is.

Step 4: Build a similarity matrix

After obtaining the similarities between each two pieces of evidence, the $n \times n$ similarity matrix SIM can be established

$$SIM = \begin{bmatrix} 1 & sim(m_1, m_2) & \dots & sim(m_1, m_n) \\ sim(m_2, m_1) & 1 & \dots & sim(m_2, m_n) \\ \vdots & \vdots & \ddots & \vdots \\ sim(m_n, m_1) & sim(m_n, m_2) & \dots & 1 \end{bmatrix} \quad (8)$$

Step 5: Determine the degree of support and credibility

The degree of support $Sup(m_i)$ that all the other evidence gives to the evidence i can be expressed as

$$Sup(m_i) = \sum_{\substack{j=1 \\ j \neq i}}^n Sim(m_i, m_j) \quad (9)$$

The credibility of evidence can be obtained after normalization of the degree of support. The credibility of the evidence i , $Cred_i$, can be expressed as

$$Cred_i = \frac{Sup(m_i)}{\max_{1 \leq i \leq n} [Sup(m_i)]} \quad (10)$$

Step 6: Modify the evidence

The credibility parameter of $Cred_i$ is used as the weight of evidence for a weighted correction to the BPA function (i.e., the occurrence probability of the detection rate in this paper). The modified occurrence probability of the detection rate $\tilde{m}_i(A)$ can be rewritten as

$$\tilde{m}_i(A_i) = Cred_i \cdot m_i(A_i) \quad (11)$$

Step 7: Determine the distribution weight of each focal element

The relative credibility $Ccred_i$ for the evidence i is

$$Ccred_i = \frac{Sup(m_i)}{\sum_{i=1}^n Sup(m_i)} \quad (12)$$

The distribution weight $\delta(A, \tilde{m})$ of the focal element A is

$$\delta(A, \tilde{m}) = \sum_{i=1}^n Ccred_i \cdot \tilde{m}_i(A) \quad (13)$$

Step 8: Evidence combination

The comprehensive index for the similarity of the target zone can be derived by combining all the evidence through the following Eq. (13)

$$m(A) = \sum_{\cap A_i = A, 1 \leq j \leq n} \Pi \tilde{m}_j(A_i) + K \cdot \delta(A, \tilde{m}) \quad (14)$$

$$K = \sum_{\cap A_i = \emptyset, 1 \leq j \leq n} \Pi \tilde{m}_j(A_i)$$

3.4 Field test approach

Once the suspected sensor pair, which has at least one faulty sensor, is located, the field test can follow to isolate and reconstruct sensor fault if needed. The aforementioned seven types of faults can be grouped into two categories: the first includes with clear math expressions such as the bias, drift and gain, and the second is affected by uncertainties such as the precision degradation, complete failure 2 etc. For the former, fault-free data are able to be reconstructed if the mathematical expressions are known. Whereas for the latter, it is impossible to obtain fault-free signals through reconstruction since the true values are mixed with the noise or even lost. As a result, sensor replacement is the only option for sensors with the second type of faults.

In the field test, additional sensors need to be installed on the co-located locations of the faulty sensors to measure the same response quantity at the same time instants. The measurements from the additional sensors can be treated as the fault-free outputs. According to the mathematical expressions for the sensor faults of Table 1, a comprehensive expression can be formed as $y = C + G \cdot x + D \cdot i$, where y is the monitoring data, x the field test data, and i the corresponding number of time series. Based on the reliability of parameters C , D and G , the specific sensor fault is able to be isolated. Moreover, once these parameters are available, the fault-free outputs can be reconstructed as

$$x = \frac{y - C - D \cdot i}{G}$$

Specifically, when there is only one reliable parameter of $G = 1$, it indicates that the sensor is fault-free. When reliable parameters are all unavailable after regression analysis, it means that the monitoring data are strongly influenced by uncertainties or unstable values of parameters (C , D and G), preventing sensor fault reconstruction. In this case, sensor replacement measure is supposed to be

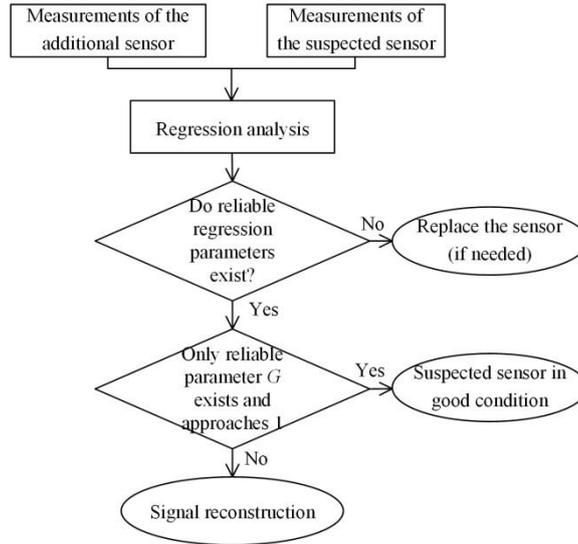


Fig. 7 Flow chart of the field test method

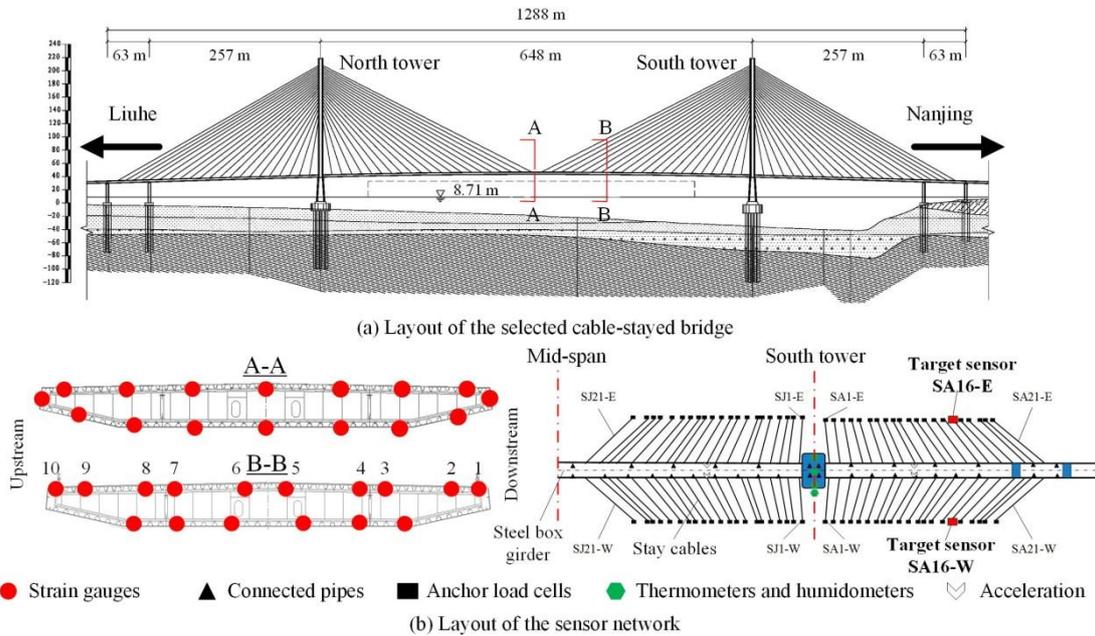


Fig. 8 Layout of the studied bridge and sensor network

implemented if needed. The flow chart of the field test method is shown in Fig. 7.

4. Case study

The 3rd Nanjing Yangtze River Bridge in China is employed as an example to demonstrate the proposed methodology for diagnosis of sensor faults. The cable-stayed bridge is a vital transportation link crossing the middle and lower Yangtze River and connecting Nanjing City and its Liuhe District. It is a cable-stayed bridge with a main span of 648m and two steel towers, as shown in Fig. 8(a). The superstructure deck has a depth of 3.2 m and the orthotropic

steel box girder has a width of 37.5 m to accommodate three traffic lanes in each direction. The deck is supported by a total of 168 stay cables and each cable consists of 109 to 241 wires having a diameter of 7 mm.

A sophisticated long-term monitoring system was devised and implemented to monitor the structural condition in the second year after completion of construction. The goal of the structural health monitoring system was to monitor the structural behavior in conditions of extreme traffic, high temperature, humidity or wind. In order to achieve this goal, the following sensor applications were selected: anchor load cells for cable forces, connected pipes for vertical deflections, strain gages, thermometers etc. The detailed sensor layout is presented in Fig. 8(b). The

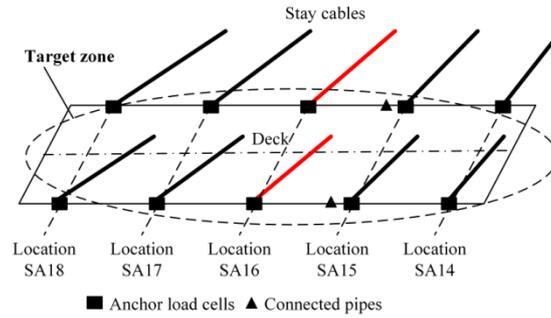


Fig. 9 The sensor pairs in the target zone

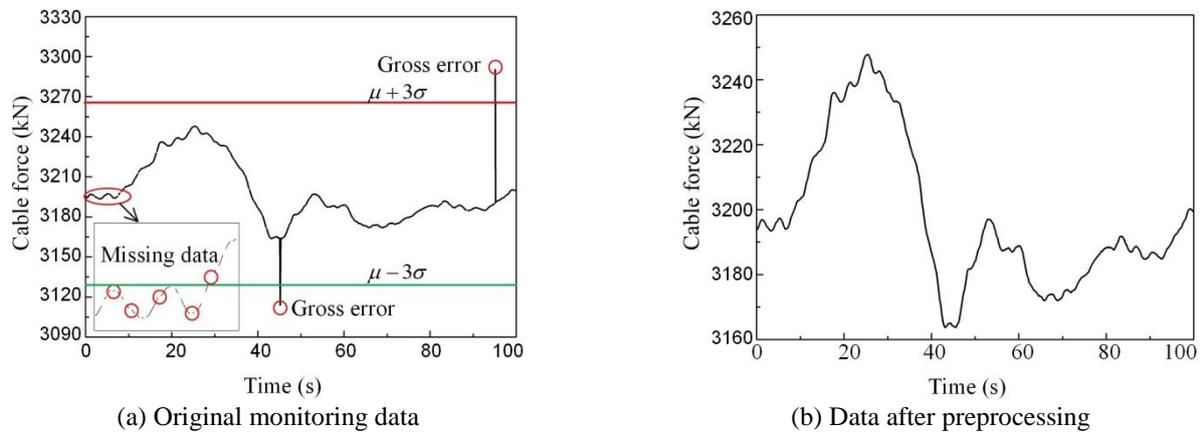


Fig. 10 Monitoring data preprocessing procedure

load cells are installed in the anchors of stay cables, which measure cable force directly. The total 168 stay cables are all equipped with anchor load cells with a relative error of $\pm 1\%$. The sample frequency of the anchor load cell is 10 Hz. The pair of the transverse symmetric sensors highlighted in Fig. 8(b) (which are the anchor load cells installed in cables SA16-E and SA16-W), was selected as the target pair in demonstrating the proposed sensor fault diagnosis approach. Aiming to determine whether the sensor faults or structural damages impact the similarity, sensor pairs located near the studied sensor pair are taken into consideration (which are anchor load cells in locations of SA18, SA17, SA15, SA14, and connected pipes near location SA15 as shown in Fig. 9).

4.1 Sensor fault detection

The recorded data (e.g., the recorded cable force data at 8:00 a.m. April 30th, 2007 from the target sensor pair are shown in Fig. 10(a)) are affected by the gross error, missing data as well as noise. Based on the three-sigma rule, the threshold values (i.e., $\mu - 3\sigma$ and $\mu + 3\sigma$) are used to detect gross errors. Then, cubic spline imputation method is implemented to achieve a continuous data source (Junninen *et al.* 2004). Finally, the dB8 wavelet basis with six decomposition levels is selected as the de-noising method for data from bridge monitoring system (Xing *et al.* 2012). After the above three preprocessing steps, the quality of data is improved and shown in Fig. 10(b).

The preprocessing one-hour monitoring data from the target sensor pair during May 13th, 2016 are shown in Fig. 11. Bias between these two groups of data may result from the sensor fault, cable force adjustment, structural damage or unbalance loadings. It is difficult to visually judge whether some sensor faults occurred. Due to the unbalance loading and noise, the similarity of the sensor pair features could be different from each other even for the healthy sensor and structure. Therefore, it is vital to determine the threshold values of the Euclidean distance of the feature for healthy sensors. It is assumed herein that all the sensors in the network as well as the structure are in the healthy state at the earlier stage of bridge operation. Thus, the early year monitoring data between January and March, 2007 of the sensor pair located in SA16 are taken as the testing data, aiming to develop the threshold of Euclidean distance. A total of 6888 values of Euclidean distance are calculated from these monitoring data of the sensor pair and are used for the distribution analysis. A Gauss distribution with a mean of $\mu = 8.86$ and standard deviation of $\sigma = 2.44$ is employed for the data fitting and is shown in Fig. 12. According to the three-sigma rule, the threshold value is calculated as 16.18.

The detection rate of the target sensor pair is calculated as 83.700% and given in Fig. 13(c). Based on the detection rate of 83.700%, we have almost 100% confidence to conclude that there are either structural damages or faults in the sensor pair located in SA16.

Table 3 Evaluation process of comprehensive evidence for similarity of target zone

Items	η_0 (%)	$P(\eta \geq \eta_0)$ (%)	$Cred_i$	$\delta(A, \tilde{m})$	K	Comprehensive evidence
SA14 cable force	33.533	0.00	0.0000			
SA15 cable force	0.067	100.00	1.0000			
SA17 cable force	0.033	100.00	1.0000	0.8202	1	82.02%
SA18 cable force	0.000	100.00	1.0000			
SA15 deflection	0.133	46.71	0.3627			

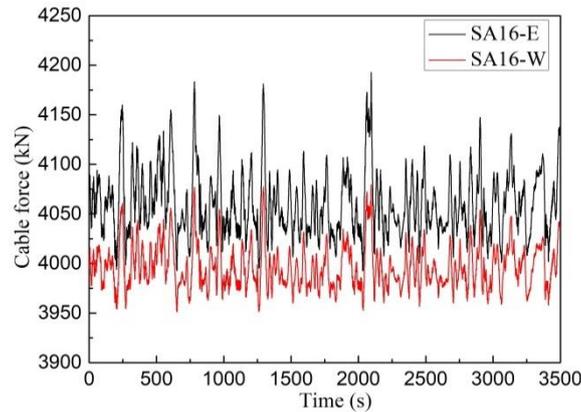


Fig. 11 Monitoring data of symmetric sensors

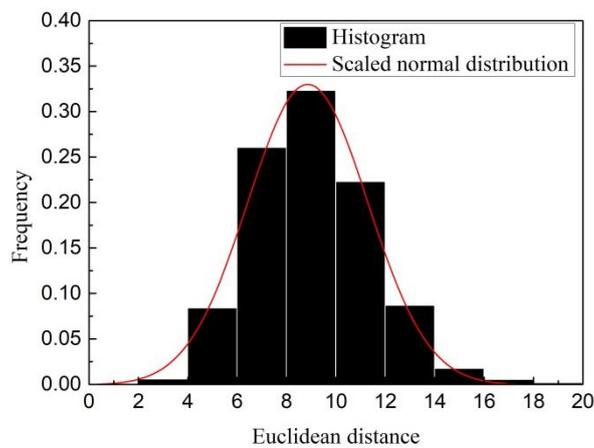


Fig. 12 Gauss distribution of Euclidean distance

Similar to the data processing of sensor at location SA16, the early year monitoring data between January and March, 2007 of the other six sensor pairs as shown in Fig. 9 are taken as the testing data for threshold determination.

Similar to the sensor pair in location SA16, the detection rates for all the other six sensor pairs are obtained and shown in Fig. 13. It can be seen from Fig. 13 that the anchor load cell pairs in the locations SA14 and SA 16 have a high detection rates of 33.533% and 83.700%, respectively, while the detection rates for the other sensor pairs are quite similar. Based on the values of the detection rates, the corresponding occurrence probabilities are calculated and listed in Table 3. For instance, the probability of

$P(\eta \geq \eta_0) = 46.71\%$ for deflection of SA15 means that we have 46.71% confidence to conclude that the symmetric response measurements are normal. According to the method developed in Section 3.3, the comprehensive similarity of the target zone is evaluated and listed in Table 3.

Based on the above analysis, we have a confidence of 82.02% to conclude that the similarity of the target zone is satisfactory. However, the studied anchor load cell pair in location SA16 in the target zone is determined to be faulty with a 100% confidence.

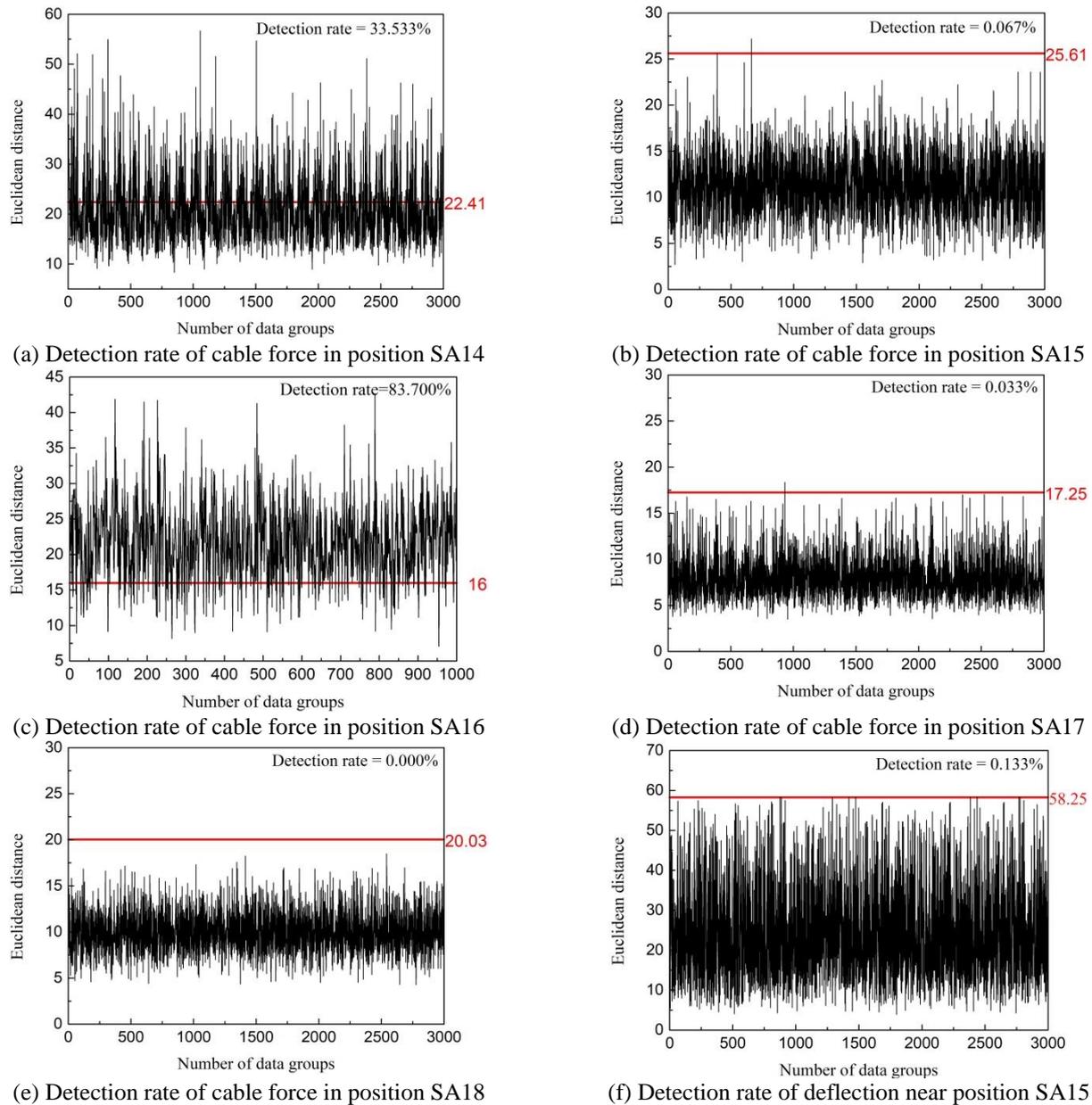


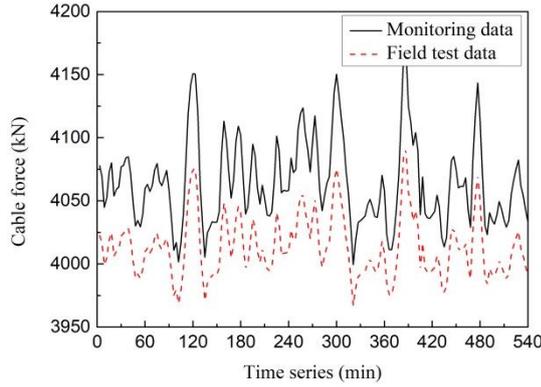
Fig. 13 Detection rates for each pair of the symmetric sensors

4.2 Sensor fault isolation and reconstruction

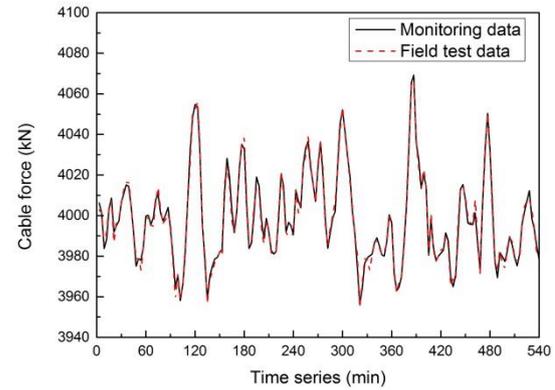
Aiming to isolate and reconstruct the sensor faults, field test has been implemented to measure the force of the studied stay cables. A wireless acceleration transducer (HCF 400) is used for cable force field test, which has a resolution of 0.25 mg. The ambient vibration method is employed to estimate the cable force in this study. It is known that the estimation error of the method bound of 3% generally due to the uncertainty of the two parameter free length of vibration and bending stiffness. Chen *et al.* (2013) developed a new concept of combining the modal frequencies and mode shape ratios to obtain more accurate values of length of vibration and bending stiffness. Moreover, in order to control the vibration of stay cables, damping devices are usually attached to the cable, which

will influence the vibration behavior of the stay cable. Sun *et al.* (2014) presented a five-parameter fractional derivative model to portray a general linear viscoelastic damper for a taut cable. Zhou *et al.* (2014) derived a frequency equation of the cable-damper-spring system for free vibration of taut cable. In practical applications, effective vibration length is updated to model the influence of the damper to cable force. To improve the accuracy of the ambient vibration method, some actions are taken as

1) In view of the sensor fixed close to the anchor construction of the cable in terms of technical and economic reasons (accessibility, installation time, etc.), the signal collected by the sensor is much more affected by the higher-order frequency than fundamental frequency. Thus, the average value of high-order frequencies is employed for the calculation of cable force.



(a) Monitoring and test data of stay cable SA16-E



(b) Monitoring and test data of stay cable SA16-W

Fig. 14 Comparison of the monitoring data and the field test data of the sensor pair

2) Aiming to improve the accuracy of the estimation, bending stiffness is taken into account by using

$$T = 4mL^2 \left(\frac{f_n}{n} \right)^2 - \frac{n^2 \pi^2}{L^2} EI, \text{ where } T, m, L, f_n, \text{ and } EI$$

denote cable tension, mass density, free length of vibration, n^{th} natural frequency, and bending stiffness respectively.

3) The theoretical derivation is based on the assumption that the cable is supported by hinges. The realistic dynamic behavior of a stay cable is between the borderline cases of an infinitely flexible string supported by hinges and a stiff cable with clamped boundary conditions. The free length of vibration is used to model the boundary condition.

4) Optimization is used to determine the optimal values of the uncertain variables (i.e., free length of vibration and bending stiffness). The target function is the error of the estimated cable force and the measured cable force by the anchor load cell. It is assumed that the anchor load cells work well at their early age (during the stage near the instant of completion of monitoring system). Therefore, the monitoring data at the early stage are used for optimization to determine the uncertain parameters (i.e., free length of vibration and bending stiffness).

By conducting the aforementioned strategies, accuracy in the range of 0.3% may be achieved. Forces of stay cables in the location SA16-E and SA16-W are measured by the acceleration transducer. The 3-min monitoring and field test data of the sensor pair in SA16 are shown in Fig. 14.

The monitoring and field test data of the cable force in SA16-W are almost identical, while there is large bias between the monitoring and field test data for SA16-E. The sensor located in SA16-E is determined to be affected by faults, while the one located in SA16-W is determined to be healthy.

Regression analysis of both the monitoring and field test data and the estimated parameters for the two regression models are listed in Table 4, in which models 1 and 2 are established by using the monitoring and field test data of cables SA16-W and SA16-E, respectively, and the p value is an index to evaluate the reliability of the regression model. If $p \leq 0.05$, the regression model is judged to be reliable.

Table 4 Estimated values of the parameters

Model	C	D	G	p
Model 1	11.8546	3.5000×10^{-3}	0.9969	0
Model 2	-1537.4	-4.8331×10^{-4}	1.3957	0

In model 1, parameter of $C = 11.8546$ is relatively smaller compared to the magnitude of the cable force in Fig. 14. The parameter of $D = 3.5000 \times 10^{-3}$ is almost zero and parameter of G is 0.9969. These estimated parameters all indicate that the sensor in SA16-W is in good condition. Moreover, the p value of model 1 is 0 indicating a reliable regression model 1. In model 2, the estimated value of parameter D in the drift expression approaches to zero, meaning that drift is not the fault for the target sensor. The estimated values of parameter C and G are -1537.4 and 1.3957, respectively, which indicates that both the bias and gain faults accompany with the sensor simultaneously. The monitoring signals of the sensor in SA16-E can have a satisfactory quality by reconstruction using the equation of $x = \frac{y + 1537.4}{1.3957}$.

5. Conclusions

A data-driven methodology based on the similarity of symmetric structure responses has been proposed and applied to sensor fault diagnosis for the bridge SHM system. The multi-sensor information fusion approach is used for sensor fault detection. The suspected sensors are then isolated and reconstructed by means of field test approach. The following conclusions can be drawn from this research:

(i) Based on the rationale that structure responses at the transverse symmetric locations tend to be quite similar with each other due to the symmetry of both structure geometry and loadings, a data-driven sensor fault detection method was proposed. Dasarathy's information fusion model is adopted for multi-sensor information fusion, including data fusion, feature extraction, feature fusion,

model recognition, and decision fusion.

(ii) According to the seven typical sensor faults and the sensitivity comparison, Euclidean distance is selected as the index to assess the similarity among the other similarity indexes such as Pearson correlation, slope relational degree, and cosine similarity.

(iii) Based on the rationale and assumption that structural damages will completely impact the similarity of all sensor responses within the target zone and the probability of a simultaneous failure of all the sensors is very low, a method was presented to evaluate the similarity of the target zone by using the evidential reasoning theory. Case study was carried out to validate and demonstrate the effectiveness of the proposed sensor fault diagnosis approach in practical engineering application. Sensors installed in a cable-stayed bridge in China are employed as the study example and sensor fault in the studied sensor pair of SA16 was detected, isolated, and reconstructed by the proposed approach. The study proposed a practical technique for sensor fault diagnosis of SHM system using the cable force and deflection measurements; however, when the proposed method applies to other sensor types or structures, similarity between symmetric responses should be assessed first.

Acknowledgements

The research reported in this paper was supported in part by the National Natural Science Foundation of China under Grant No. 51208096, the Key Science and Technology Project of Jiangsu Transport Department, China under Grant No. 2014Y02, the Scientific Research Foundation of Graduate School of Southeast University under Grant No. YBJJ1845 and China Scholarship Council.

References

- Alcala, C.F. and Qin, S.J. (2011), "Analysis and generalization of fault diagnosis methods for process monitoring", *J. Process Contr.*, **21**(03), 322-330.
- Amiri, M. and Jensen, R. (2016), "Missing data imputation using fuzzy-rough methods", *Neurocomputing*, **205**(C), 152-164.
- Bi, W., Zhang, A. and Yuan, Y. (2017), "Combination method of conflict evidences based on evidence similarity", *J. Syst. Eng. Electron.*, **28**(3), 503-513.
- Chang, C.M., Chou, J.Y., Tan, P. and Wang, L. (2017), "A sensor fault detection strategy for structural health monitoring systems", *Smart Struct. Syst.*, **20**(1), 43-52.
- Chen, C.C., Wu, W.H., Huang, C.H. and Lai G. (2013), "Determination of stay cable force based on effective vibration length accurately estimated from multiple measurements", *Smart Struct. Syst.*, **11**(4), 411-433.
- Dasarathy, B.V. (1994), *Decision Fusion*, IEEE Computer Society Press, Amsterdam, Netherlands.
- Hernandez-Garcia, M.R. and Masri, S.F. (2013), "Application of statistical monitoring using latent-variable techniques for detection of faults in sensor networks", *J. Intel. Mat. Syst. Str.*, **25**(2), 121-136.
- Huang, H.B., Yi, T.H. and Li H.N. (2015), "Sensor fault diagnosis for structural health monitoring based on statistical hypothesis test and missing variable approach", *J. Aerospace Eng.*, **30**(2), B4015003.
- Huang, H.B., Yi, T.H. and Li, H.N. (2016), "Canonical correlation analysis based fault diagnosis method for structural monitoring sensor networks", *Smart Struct. Syst.*, **17**(6), 1031-1053.
- Huang, H.B., Yi, T.H. and Li, H.N. (2017), "Bayesian combination of weighted principal-component analysis for diagnosing sensor faults in structural monitoring systems", *J. Eng. Mech.*, **143**(09), 04017088.
- Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J. and Kolehmainen, M. (2004), "Methods for imputation of missing values in air quality data sets", *Atmos. Environ.*, **38**(18), 2895-2907.
- Kerschen, G., De Boe, P., Golinval, J.C. and Worden, K. (2004), "Sensor validation using principal component analysis", *Smart Mater. Struct.*, **14**(01), 36-42.
- Ko, J.M. and Ni, Y.Q. (2005), "Technology developments in structural health monitoring of large-scale bridges", *Eng. Struct.*, **27**(12), 1715-1725.
- Kullaa, J. (2011), "Distinguishing between sensor fault, structural damage, and environmental or operational effects in structural health monitoring", *Mech. Syst. Signal Pr.*, **25**(8), 2976-2989.
- Li S., Zhu, S., Xu, Y.L., Chen, Z.W. and Li, H. (2012), "Long-term condition assessment of suspenders under traffic loads based on structural monitoring system: application to the Tsing Ma Bridge", *Struct. Control Health.*, **19**(1), 82-101.
- Narasimhan, S. and Mah, R.S.H. (1987), "Generalized likelihood ratio method for gross error identification", *AICHE J.*, **33**(09), 1514-1521.
- Qin, Z., Chen, L. and Bao, X. (2012), "Wavelet denoising method for improving detection performance of distributed vibration sensor", *IEEE Photonic. Tech. L.*, **24**(05), 542-544.
- Reppa, V., Polycarpou, M.M. and Panayiotou, C.G. (2014), "Adaptive approximation for multiple sensor fault detection and isolation of nonlinear uncertain systems", *IEEE T. Neural Netw. Learn. Syst.*, **25**(1), 137-153.
- Sun, L. and Chen, L. (2015), "Free vibrations of a taut cable with a general viscoelastic damper modeled by fractional derivatives", *J. Sound Vib.*, **335**, 19-33.
- Tamhane, A.C., Iordache, C. and Mah, R.S. (1988), "A Bayesian approach to gross error detection in chemical process data: Part I: Model development", *Chemometr. Intell. Lab.*, **4**(1), 33-45.
- Tong, H. and Crowe, C.M. (1995), "Detection of gross errors in data reconciliation by principal component analysis", *AICHE J.*, **41**(7), 1712-1722.
- Xing, D., Zhang, L. and Duan, Y. (2012), "Research on long-span bridge health monitoring data pre-processing based on wavelet analysis", *Highway Eng.*, **37**(2), 33-36.
- Yang, J.B. and Xu, D.L. (2002), "On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty", *IEEE T. Syst. Man Cy.*, **32**(3), 289-304.
- Yi, T.H., Huang, H.B. and Li, H.N. (2017), "Development of sensor validation methodologies for structural health monitoring: a comprehensive review", *Measurement*, **109**, 200-214.
- Yi, T.H., Li, H.N., Song, G. and Guo, Q. (2016), "Detection of shifts in GPS measurements for a long-span bridge using CUSUM chart", *Int. J. Struct. Stab. D.*, **16**(4), 1640024.
- Zhou, H., Sun, L. and Xing, F. (2014), "Free vibration of taut cable with a damper and a spring", *Struct. Control Health.*, **21**(6), 996-1014.
- Zhu, D., Bai, J. and Yang, S.X. (2009), "A multi-fault diagnosis method for sensor systems based on principle component analysis", *Sensors*, **10**(1), 241-253.