

Data Interpretation and Continuous Modal Parameter Identification of Cable Stayed Bridge

Inamullah Khan^{*}, Deshan Shan, Qiao Li, Jie He and Fei Long Nan

Bridge Engineering Department, Civil Engineering School, Southwest Jiaotong University, Chengdu, Sichuan Province, China

Abstract: For reliable identification of modal parameters, it is important to distinguish between abnormal data due to defects, malfunctioning, and anomalies in the sensors, from that of precise data. In case of long-term continuous monitoring data, it is imperative to identify any defects in the raw data very quickly and accurately to ensure that the identification is trustworthy. Exploratory Data Analysis (EDA) is employed for the purpose of quickly visualizing any defects and anomalies in the sensor's data. Outlier analysis is employed to make some data treatment followed by auto and cross correlation to further elucidate any defects and anomalies in the collected data. Finally, covariance driven stochastic subspace identification (CO-SSI) with some improvements is employed to carry out the continuous modal parameter identification. The Sutong Yangtze river bridge, a long span Y-shape pylon cable stayed bridge with a main span of 1088m was chosen as a case study and the above proposed methods were applied. The result showed that the suggested method is very effective and can provide better and more accurate real life results in the continuous health monitoring of bridges.

Keywords: Cable stayed bridge, continuous modal parameter, covariance driven stochastic subspace identification, exploratory data analysis.

1. INTRODUCTION

The advancement in technology and growing needs for sophisticated structures, specially bridges, has evolved significantly providing people with an opportunity for safer trade and communication, Viecili *et al.* [1], Chen *et al.* [2]. In the past three decades, researchers have tried to establish effective techniques for monitoring civil structures. However, the unpredictable nature of civil structures due to the anomalous data obtained from these structures and the non-linear behavior of these structures makes the monitoring process more complex, Spark and Kasmarik [3], and so the data interpretation becomes more and more essential, Khan *et al.* [4]. The enormous amount of data obtained from these sophisticated measuring equipment installed at these civil structures, specially the long span cable stayed bridges, needs to be reliable, so that an accurate and detailed data analysis can be performed effectively, Cross *et al.* [5]. Since the data obtained from the health monitoring equipment is usually in a very raw form and because of the huge amount of time it takes to compute the data, it is not possible for the researcher to go through all the data which means that the researcher can very often ignore the initial check on the data, which in turn can lead to an erroneous analysis of the data. The inaccuracies in the data may be due to various reasons, like missing data, environmental effects on the data, and the malfunctioning of the sensors, etc. Wu *et al.* [6].

In order to detect the anomalies very quickly and to ensure an accurate long-term continuous operational data

analysis of long span cable stayed bridges; it is vital to investigate the initial raw data so that the inner secrets of the data can be depicted in a more professional way. Therefore, a tool for making sense of the data is introduced, which can help the user to effectively visualize the data before they start to carry out any further detailed data analysis. The tool, which aids the understanding of the data by providing visualization of the huge amount of data information, and enables the users to get a feel for the data in a short period of time, is a process called Exploratory Data Analysis (EDA).

Tukey [7] first presented the concept of EDA in 1971, and then several authors used it for data visualization, such as Chambers [8], Cleveland [9, 10], Tufté [11, 12], Buja *et al.* [13], Wainer *et al.* [14], etc. All of them, just focused on data visualization through making graphs, because the technology was not yet very advanced and fast modern processing computers and software were not very common, and so EDA was just a tool for some simple data representation. The focus of Brillinger *et al.* [15] study was on paths showing animals' movement in a forest, and he used Rowland *et al.* [16] data, which was just experimental data whose data size was very small and not continuous, in his work which was only studying the movement of elk and deer. Thus, his methods are not really suitable or applicable for anomaly detection and for quick continuous large scale data analysis. Therefore, this paper introduces a new concept for quick data visualization of large scale continuous real life monitoring data sets by detecting the defects and anomalies through the introduction of EDA, outlier analysis and auto and cross correlation.

The ambient excitation method for extracting the structural modal parameters in the frequency as well as in the

^{*}Address correspondence to this author at the School of Civil Engineering, Bridge Engineering Department, Southwest Jiaotong University, 610031, P. R. China; Tel: +86-18280158259; E-mail: inam_bunny@yahoo.com

time domain is of extreme significance for bridge engineering. In order to extract these operational modal parameters, a number of stochastic identification techniques have been established previously, but amongst them the stochastic subspace identification (SSI) is known to be a very effective and multivariate identification technique. The SSI algorithm is a sophisticated technique for identification of operational modal parameter, Liu *et al.* [17]. SSI uses the ambient dynamic data to extract the structure model through linear algebraic manipulations, Shan and Li [18]. The SSI consists of two methods: i) data driven stochastic subspace identification (DATA-SSI); and ii) covariance driven stochastic subspace identification (CO-SSI), Magalhães *et al.* [19]. The reference based SSI algorithm was developed by Peeters and De Roeck [20, 21] for the identification of steel transmitter masts and pre-stressed concrete bridges. It is not suitable for long span cable stayed bridges as the vibrations are more frequent and it is an offline technique unsuitable for continuous monitoring. This paper adopts the covariance driven stochastic sub-space (CO-SSI) because of its inherent advantages of numerical stability, robustness and because it avoids the computations of orthogonal projections which are replaced by converting raw time histories in an assemble of block covariances called Toeplitz matrix, from which the system dynamic characteristics can be extracted, Overschee and Moor. [22], Liu *et al.* [17], Bassville *et al.* [23], Weng *et al.* [24]. Bonyapinyo *et al.* [25] applied CO-SSI to extract derivatives, Goethals *et al.* [26] used CO-SSI to extract flutter, but the identified flutter derivatives were scattered and unstable. Another major drawback of CO-SSI is that, it is not an online technique and does not continuously process the data, thus making it unfeasible for continuous bridge health monitoring. Goethals *et al.* [26] come up with sliding window technique (SWT), which turned out to be not so effective, and which also had stabilization problems when extracting the continuous modal parameters from the actual long term continuous data. Thus, in this paper visualization of data is first carried out by employing EDA, to quickly explore the inner secrets of the enormous data sets, because one of the keys for continuous monitoring is quick data visualization. Then, based on the results of defective data, some treatments are done to the data using outlier analysis. This is followed by a further verification and visualization of the data through auto and cross correlation to cross check the data visualization. Finally, CO-SSI is used with some improvements, because CO-SSI is usually only ever used for offline analysis. To make CO-SSI more suitable for continuous modal parameter identification, SWT is introduced to identify the continuous modal parameter such as the frequency and the damping ratio.

2. DATA INTERPRETATION

2.1. Exploratory Data Analysis

Over the past few decades, statistical tools have become more and more important, but due to the standard paradigms of estimation and time consumption they cannot be used very effectively for the analysis of long-term continuous data. Therefore, EDA and data visualization have grown in importance as tools for real life continuous data monitoring. EDA is a philosophy on how to carry out data analysis in a more professional way. EDA is different from statistical

graphs, in that statistical graphs are all graphically based with an emphasis on data characterization, Tukey [7]. EDA encompasses a broad venue, and it helps in depicting the model selections and in choosing the calculation techniques according to data trends that are the most suitable by exploring within the data structures and helping in deciding which techniques are the most appropriate for the analysis, with a minimum of error and more accuracy. In case of EDA, high resolution graphs, and a more sophisticated interactive user interface with powerful software have given it more importance and space for graphical methods to help elaborate the data faster and more precisely.

In case of continuous data where little is known about the probability density functions (PDF's), it is usually recommended that the data should be first examined quickly to identify any underlying structures before engaging in more quantitative techniques. EDA can process physical reasoning, data filtering, hypotheses testing and anomalies detection, Huston [27]. The primary objective of EDA is to describe the overall appearance of data. EDA techniques are usually used for determining the location and extent of the spread of the data, and smoothing and transferring the data into a format which can be easily used for further viewing and analysis. There are various methods in graphical representation of data sets, and some of them are very diverse, like histograms, normal cumulative distribution function (CDF) plots, quartile-quartile (QQ) plots, box plots, scattered plots, etc. The main reason for its dependence on graphical representation is to explore the data more open-mindedly, and thus help the data to reveal its structural secrets, helping to gain some new and unsuspected insight into the data, Freehafer [28].

2.2. Outlier Analysis

An outlier is an observation in a data set that is astonishingly dissimilar from the remaining data and is supposed to influence the rest of data drastically, Sohn *et al.* [29], Farrar and Worden [30]. The treatment of outlier is a flourishing area in modern day probability and statistics and many techniques are available for detecting if the given observation is in fact an outlier, Sohn *et al.* [29], Farrar and Worden [30]. Outlier analysis is very important for anomaly detection, because anomalies in measured values may be due to reasons such as noise, operational and environmental variation, or sensor malfunctions, and will thus lead the matrices to exceed the thresholds and result in false data analysis. However, outliers often contain useful information, especially about the abnormal characteristic of system and so can impact the data generation process. After performing the EDA, outlier analysis is therefore carried out to further verify the data. This process will make the data flawless which will enable it to be used for the extraction of the continuous modal parameters. Additionally, the pauta criterion is adopted to further verify our collected data.

2.2.1. Pauta Criterion

In this paper, pauta criterion is adopted to determine abnormal values so that some possible treatment can be done on those values which are probably not outliers and which have a very little influence on our data analysis. These values are made useable after a little treatment and are used in con-

tinuous modal parameter identification. Finally, the Bessel's formula is used to point out any abnormal values in the sensor's data. Some improvements have been made to the Bessel's formula to help it suit the data. The details of the procedure are as follows:

Step (1): Let's suppose the original data obtained is X , then the average of it as A can be found, and then the standard deviation S is calculated using the Bessel formula as:

$$S = \sqrt{\frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - \frac{1}{n} \left(\sum_{i=1}^n X_i \right)^2 \right]} \quad (1)$$

where X_i is the original data ($i=1, 2, \dots, n$), and S is the standard deviation.

Step (2): If $|X_i - A| > 3S$, then X_i will delete one bit of data, $\frac{X(i-1) + X(i+1)}{2}$ is used in order to make up a new bit of data. The value obtained is compared with 3 times the standard deviation as this is considered to be a threshold for the discordancy value for an outlier detection.

2.3. Auto and Cross Correlation Functions

2.3.1. Auto Correlation

The auto correlation function of an ordered dataset is considered to be a time series correlation with its own past and future values, Karimi *et al.* [31], Fischer and Jensen [32]. It is also sometimes called a "lagged correlation" or a "serial correlation", which refers to the correlation between members of a series of numbers arranged in time, Farrar and Worden [31], and it can be written as:

$$\rho_{xx}(\tau) = E[x(t+\tau)x(t)] \quad (2)$$

where "E" is expected value and $\rho_{xx}(\tau)$ is the auto correlation function in the above equation. When a signal is shifted in the time domain by an amount τ , it can be used to find irregularities in the data. If the time lag $x(t)$ approaches zero, the variance equals the auto correlation, Farrar and Worden [30], and can be expressed as:

$$\rho_{xx}(0) = E[x(t)^2] = \sigma_x^2 \quad (3)$$

$$\rho_{xx}(\tau) = \rho_{xx}(-\tau) \quad (4)$$

Therefore, $\rho_x(\tau)$ is considered to be a function of " τ ".

2.3.2. Cross Correlation

The cross correlation function is considered to be relative information between data sets with a time lag, Sohn *et al.* [29]. Let say x and y are two data sets, then the cross correlation function $\rho_{yx}(\tau)$ can be written as:

$$\rho_{yx}(\tau) = E[y(t+\tau)x(t)] \quad (5)$$

$$\rho_{yx}(\tau) = \rho_{xy}(-\tau), \rho_{xy}(\tau) = \rho_{yx}(-\tau) \quad (6)$$

In this way, $\rho_{xx}(\tau)$ detects any unusual relationship between the sensor signals and can very accurately detect the anomalies in the sensors data.

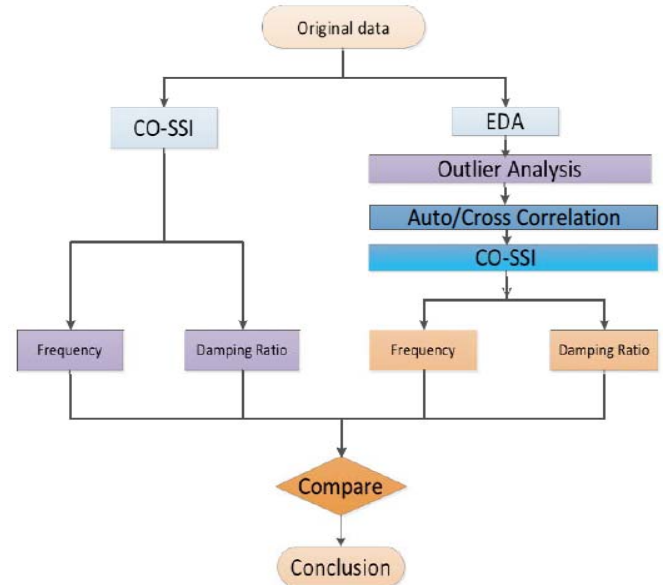


Fig. (1). Modal parameters identification process.

3. COVARIANCE DRIVEN STOCHASTIC SUBSPACE IDENTIFICATION

3.1. Mathematical Representation

The ambient vibrations of a linear time invariant dynamic system of a bridge structure can be written as, Khan *et al.* [33]:

$$M\ddot{x}(t) + c\dot{x}(t) + Kx(t) = Lu(t) \quad (7)$$

where "M", "c" and "K" are the mass, damping coefficient and the stiffness, respectively, and $\ddot{x}(t)$ is the acceleration vector, $\dot{x}(t)$ is the velocity vector and $x(t)$ is the displacement vector at time instant t. "L" is the input location matrix and $u(t)$ is vector that expresses "m" input as a function of time t.

By assuming that the white noise is the excitation and assuming that all data is sampled in a discrete time form, the CO-SSI utilizes the state space model to calculate the modal parameters. Therefore, the discrete time stochastic state space model, Magalhães *et al.* [19], can be written as:

$$\left. \begin{aligned} x_{k+1} &= Ax_k + w_k \\ y_k &= Cx_k + v_k \end{aligned} \right\} \quad (8)$$

where $w_k \in R^{2n \times 1}$ are anticipated as the zero mean, spatially white noise and x_{k+1} is a time state vector at the time instant k , y_k is a vector with the sampled outputs, A and C are the discrete state and output matrices, respectively and v_k is the modelling inaccuracies vectors, Khan *et al.* [33].

3.2. CO-SSI Algorithms

The CO-SSI techniques utilize the covariance matrix as a comparison to the data driven technique which starts from

the raw time series. Secondly, the results provided by the CO-SSI are very robust and exhibit quick computation efficiency. In the case of CO-SSI, the Hankel matrix plays a very important role, and thus the number of block rows and columns have a direct influence on the modal parameter identification process, Khan *et al.* [33].

$$H = \frac{1}{\sqrt{N}} \begin{bmatrix} y_1 & y_2 & \cdots & y_N \\ y_2 & y_3 & \cdots & y_{N+1} \\ \cdots & \cdots & \cdots & \cdots \\ y_i & y_{i+1} & \cdots & y_{i+N} \\ y_{i+1} & y_{i+2} & \cdots & y_{i+N+1} \\ y_{i+2} & y_{i+3} & \cdots & y_{i+N+2} \\ \cdots & \cdots & \cdots & \cdots \\ y_{2i} & y_{2i+1} & \cdots & y_{2i+N+1} \end{bmatrix} = \begin{bmatrix} Y_p \\ Y_f \end{bmatrix} \quad (9)$$

The block Toeplitz matrix can be obtained by multiplication between the future and the transpose of past measurement, and the Toeplitz matrix can be written as, Liu *et al.* [17]:

$$T = \begin{bmatrix} R_i & R_{i-1} & \cdots & R_1 \\ R_{i+1} & R_i & \cdots & R_2 \\ \cdots & \cdots & \cdots & \cdots \\ R_{2i-1} & R_{2i-2} & \cdots & R_i \end{bmatrix} = Y_f (Y_p)^T \quad (10)$$

$$T = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix} \begin{bmatrix} G & AG & A^2G & \cdots \end{bmatrix} = O_i \Gamma_i \quad (11)$$

Let $G = E[x_{k+1} y_k^T]$, $R_i = E[y_{k+i} y_k^T]$, and $R_i = CA^{i-1}G$ (12)

This is how CO-SSI can be derived, where the observability matrix $O_i \in R^{li \times 2n}$ and $\Gamma_i \in R^{2n \times li}$ is the reversed extended stochastic controllability matrix and E is expectation operator, (A and C) are the system matrices, $G \in R^{2n \times l}$ is the next step output covariance matrix and $R_i \in R^{l \times l}$ is the output covariance matrix of arbitrary time lag i , Liu *et al.* [17].

After performing the Singular Value Decomposition (SVD), for Eq. (11), it can be written as:

$$T = USV^T = (U_1 U_2) \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = U_1 S_1 V_1^T \quad (13)$$

After comparing Eq. (7) and Eq. (8), the SVD can be split into two parts and can be written as:

$$O_i = U_1 S_1^T T \quad \text{and} \quad \Gamma_i = T^{-1} S_1^{1/2} V_1^T \quad (14)$$

For original state space transformation, T can be simply a unit matrix, hence the above equation can be written as:

$$O_i = U_1 S_1^T \quad \text{and} \quad \Gamma_i = S_1^{1/2} V_1^T \quad (15)$$

where

$$C = O_i(1:l,:)\quad (16)$$

In MATLAB notation, the C matrix is just the first block of O_i and the structural discrete state matrix A can be written as follows:

$$A = \frac{1}{\sqrt{S_1}} U_1^T T_{\frac{2}{i+1}} V_1 \frac{1}{\sqrt{S_1}} \quad (17)$$

where $T_{\frac{2}{i+1}}$ is also called the Toeplitz matrix for covariance

from 2 to $i + 1$ Therefore, the eigenvalues and the eigenvectors can be obtained by the decomposition of the state matrix A.

$$A = \phi \ddot{E} \psi^{-1} \quad (18)$$

Hence, the values of A, G and C can be calculated by using MATLAB. After calculating these matrices, the next step is to identify the modal parameters. The diagonal matrix can be given as:

$$\Lambda = \text{diag}[\mu_i] \quad (19)$$

where the matrix Λ is composed of the complex eigenvalue μ_i with arbitrary time lag i , and the relation between the eigenvalues of matrix A and systematic one is given as:

$$\lambda_i = \frac{\ln \mu_i}{\Delta t} \quad (20)$$

where the eigenvalue of the system is λ_i , with a sampling time interval of Δt . The relation between the systematic eigenvalues and λ_i and the systematic intrinsic frequency ω_i and the systematic damping ratio ξ_i is given as:

$$\lambda_i, \bar{\lambda}_i = -\xi_i \omega_i \pm \sqrt{\xi_i^2 - 1} \quad (21)$$

Hence, the frequency and the damping ratio can be found from the equations given below:

$$f_i = \frac{|\lambda_i|}{2\pi} \quad (22)$$

$$\hat{1}_i = -\frac{Re \lambda_i}{2\pi f_i} \quad (23)$$

3.3. Sliding Window Technique (SWT)

The CO-SSI method is generally an offline identification technique and can only be used to process the data in one batch, and therefore it cannot be used as an online health monitoring technique. To overcome this drawback, a novel stabilization method to incorporate the continuous bridge monitoring data called SWT has been adopted to transform the CO-SSI into a continuous online modal parameter identification technique. In this paper, the length of sliding window is 6000 points (5 minutes), with an overlap of 3 minutes, and because it utilizes one hour data, so the total number of windows are 23.

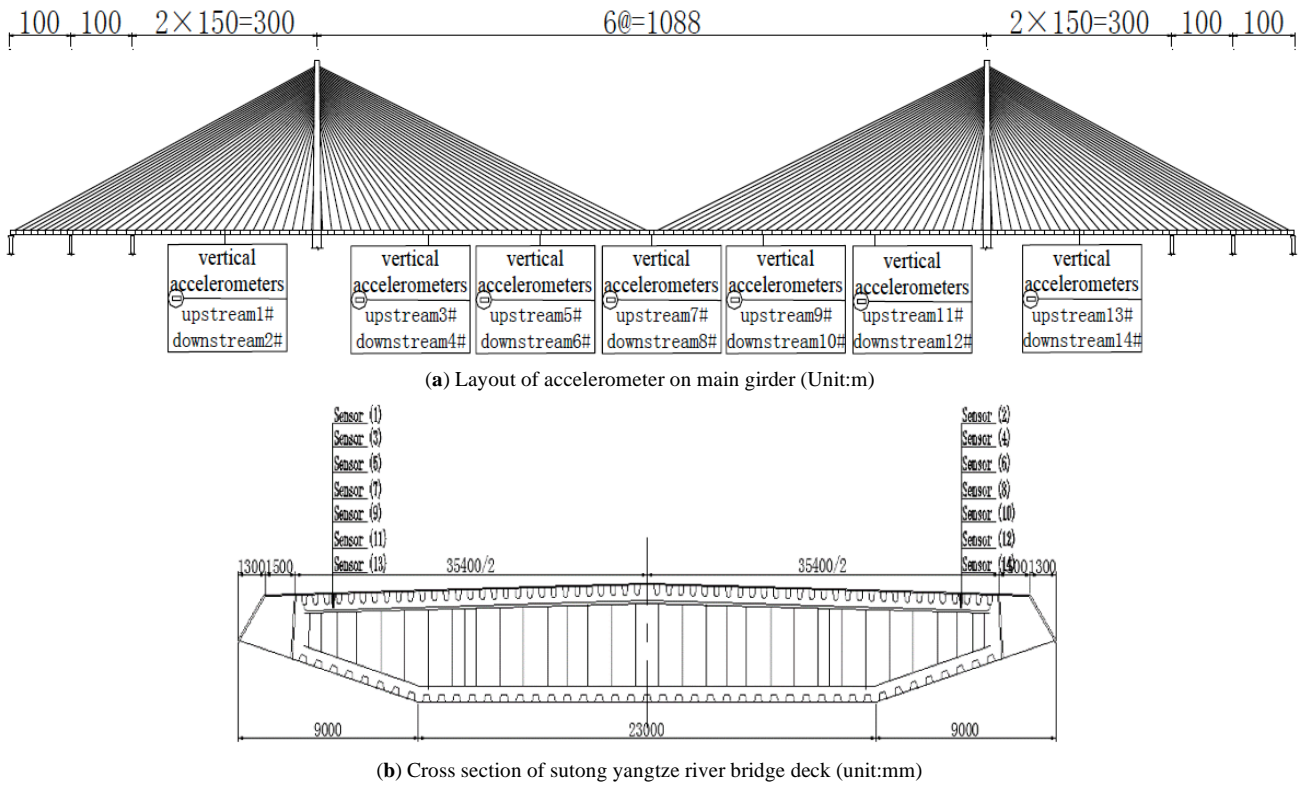


Fig. (2). Sutong yangtze river bridge.

Since the CO-SSI has been transformed so that it can handle the continuous monitoring data, it is therefore necessary that the old Hankel matrix also changed to incorporate the new continuous data. The new Hankel covariance matrix factorization form can thus be written as:

$$\hat{H} = \begin{bmatrix} R_1 & R_2 & \dots & R_i \\ R_2 & R_3 & \dots & R_{i+1} \\ \dots & \dots & \dots & \dots \\ R_i & R_{i+1} & \dots & R_{2i-1} \end{bmatrix} = O_i \Pi_i \quad (24)$$

where $O_i \in R^{li \times 2n}$ observability matrix and $\Pi_i \in R^{2n \times li}$ is the stochastic controllability matrix, which is similar to the Γ_i as discussed previously, but with its entries in reversed order.

The observability matrix can be obtained by performing SVD to the Hankel covariance matrix, which means that the system matrices and the modal parameters can be extracted in the same manner as presented in section 3.2, in the CO-SSI offline analysis. The Hankel covariance matrix can now be formulated by ordering the output measurement data vectors as follows:

$$y_k^+ \equiv \begin{bmatrix} y_{k-i+1} \\ y_{k-i+2} \\ \dots \\ y_k \end{bmatrix}, y_k^{-T} \equiv [y_{k-i}^T \quad y_{k-i-1}^T \quad \dots \quad y_{k-2i+1}^T] \quad (25)$$

where $y_k \in R^{l \times 1}$, $y_k^+ \in R^{il \times 1}$ and $y_k^{-T} \in R^{1 \times il}$ are the output measurements vectors, l is the number of sensors and “ i ” is the number of block rows which constitute the Hankel covariance matrix. Thus, the new Hankel matrix \hat{H} can be built and can be expressed as:

$$\hat{H} = E[y_k^+ y_k^{-T}] \frac{1}{p} \sum_{k=2i}^N y_k^+ y_k^{-T} \quad (26)$$

$$\hat{H} = \begin{bmatrix} y_{i+1} & y_{i+2} & \dots & y_{N-i+1} & y_i^T & y_{i-1}^T & \dots & y_1^T \\ y_{i+2} & y_{i+3} & \dots & y_{i+N+1} & y_{i+1}^T & y_i^T & \dots & y_2^T \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ y_{2i} & y_{2i+1} & \dots & y_N & y_{N-i}^T & y_{N-i-1}^T & \dots & y_{N-2i+1}^T \end{bmatrix}$$

$$\hat{H} = Y_k^+ Y_k^{-T} \quad (27)$$

where K is the limit of available data and p is an optional normalization parameter and is based on the order of the covariance matrix and the data length of the Hankel matrix is N .

In the end SWT is used on the data that is obtained by data interpretation. The new continuous data is assimilated into \hat{H} at the same rate as the old data is eradicated, so that the new formula can be written as:

$$\hat{H}_{N+1} = \hat{H}_N + y_{N+1-i}^+ (y_{N+1-i}^-)^T - y_i^+ y_i^{-T} \quad (28)$$

where all the manipulations of \hat{H} amount to update of the formula:

$$\hat{H}_{t+1} = \mu H_t + y_{t+1}^+ (y_{t+1}^-)^T \quad (29)$$

The above formula is used for the sliding window and is repeated twice to complete the subspace updating for each new bit of incoming data. The same formula as shown above *i.e.* Eq. (29), can be applied for downdating by setting the forgetting factor to $\mu=1$, thus the data vectors y_{t+1}^+ and y_{t+1}^- are replaced by the oldest data vectors in the moving window. For example, if the moving window length is L , these can be expressed as: y_{t-L+2i}^+ and y_{t-L+2i}^- . Thus, the updating task for the time varying subspace is accomplished and the system information can be extracted from the previously discussed section 3.2. The downdating can be expressed as:

$$\hat{H}_{t+1} = \mu H_{t+1} - y_{t-L+2i}^+ (y_{t-L+2i}^-)^T \quad (30)$$

4. RESULTS AND DISCUSSION

4.1. Sutong Yangtze River Bridge

The Sutong Yangtze river bridge is located in China on Yangtze River and is one of the longest spanning cable-stayed bridges and connects the cities of Suzhou and Nantong. The bridge is a seven span double Y-shape pylon and double cable plane steel box girder cabled-stayed bridge having a main span of 1088m, while the total length of the bridge is 2088m as shown in Fig. (2a), and the cross section is shown in Fig. (2b). In order to extract the continuous modal parameters of the Sutong Yangtze river bridge, data was obtained from 14 accelerometers sensors that were installed on the bridge for a period of 1 week from 01-07-2011 to 07-07-2011. This one week data is used for performing the EDA and one hour data is utilized for modal parameters identification to make the figures more clear. The layout of the vertical accelerometer sensors is shown in Fig. (2a). The sampling frequency for data collection was 20Hz.

The data was first visualized using EDA. When continuous monitoring is used, the collected data is always very large, generally in 100's of Giga Bytes, and so it is not possible for a researcher to go through the complete data set to find any anomalies in the data. A fast and accurate technique is therefore needed to help detect the anomalies in the data, which in turn helps avoiding any erroneous data analysis in the later stages. The technique proposed in this paper is to utilize EDA to deal with this kind of problem, which is the first time this concept has been introduced and used on long-term continuous monitoring data. After the EDA was done, based on the recognized faulty data, outlier analysis was then performed and any abnormal values were discarded and those values that had important information and were close to the limits were kept and used for the analysis. After that, auto and cross correlation was performed to further investigate the pattern of the data. Finally, after carrying out the data interpretation of the data, the continuous CO-SSI technique based on SWT was utilized in MATLAB to get the requisite continuous modal parameters, *i.e.* the frequency and damping ratios. The flow chart is shown in Fig. (1). The layout of the 14 accelerometer sensors installed on the Sutong Yangtze river cable stayed bridge is shown in Fig. (2a).

4.2. Exploratory Data Analysis

4.2.1. Histograms

Histograms are a graphical representation of the data to convey its distribution in vertical bars, and the height of bars correspond to the frequency. Histograms are very useful for the acquisition of a quick pictorial distribution of the data by superimposing the normal distribution over relative frequency histograms. The histograms shown in Fig. (3), are from the acceleration data of 4 sensors on Sutong Yangtze cable stayed bridge, although there are 14 accelerometer installed at the Sutong Yangtze cable stayed bridge, but due to space limitations only 4 sensors results are shown in Fig. (3). The results show that the histograms for sensor#3, sensor#4, sensor#5, sensor#6, sensor#7, sensor#8, sensor#10 and sensor#11 all follow a Gaussian distribution, whereas for sensor#1, sensor#2, sensor#9, sensor#12, sensor#13 and sensor#14 the bars do not follow a normal distribution, thus showing that the obtained data from these sensors is incor-

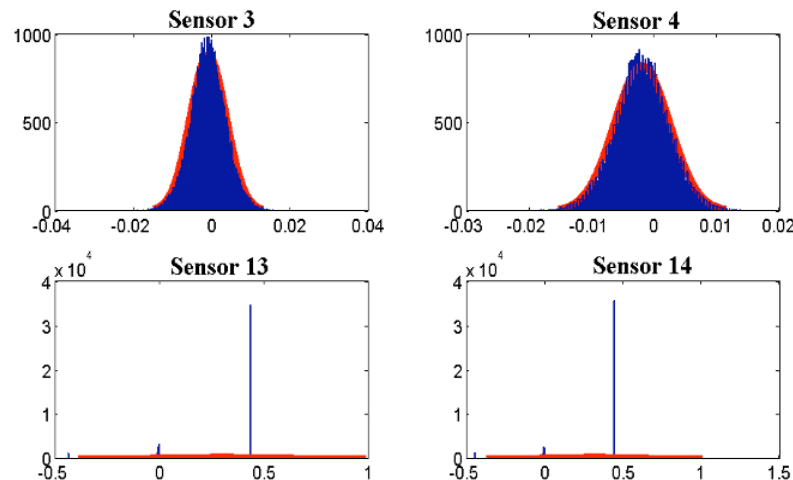


Fig. (3). Histograms plots for sensor #3, 4, 13 and sensor #14.

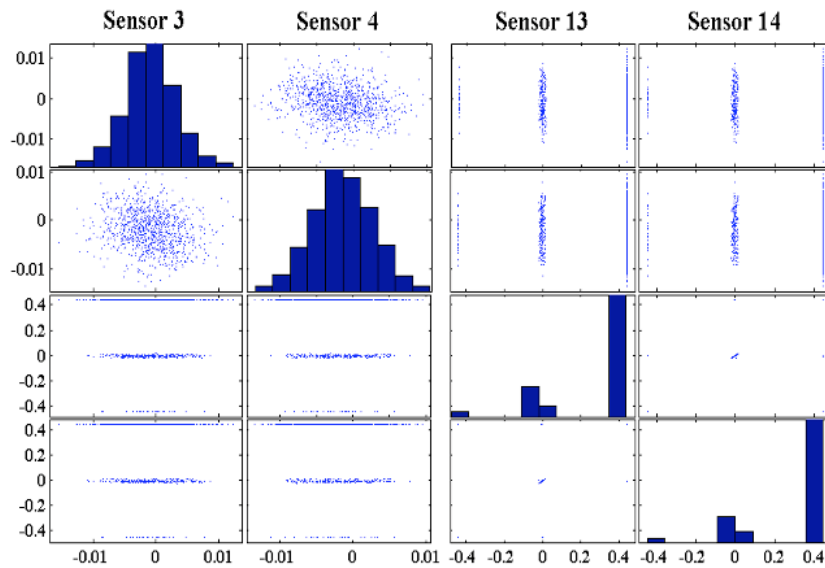


Fig. (4). Scattered plots for sensor #3, 4, 13 and sensor #14.

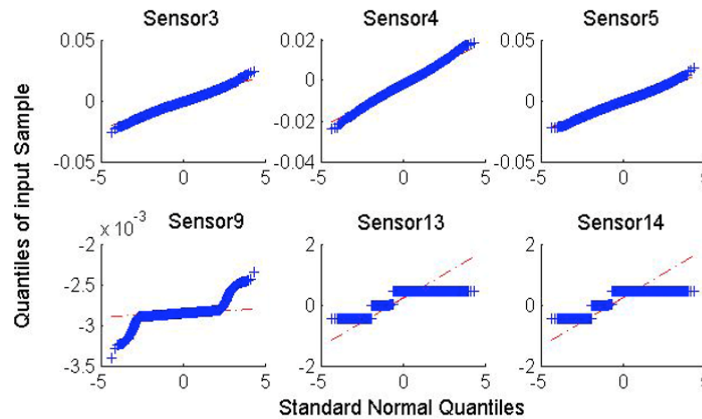


Fig. (5). QQ plots for sensor #1, 4, 5, 9, 13 and sensor #14.

rect due to malfunctioning or due to damaged sensors, which means this data cannot be used for the continuous modal parameter identification.

4.2.2. Scattered Plots

For multivariate data or higher level structured data where there are two or more variables, scattered plots are a very successful tool for the diagnosis of any anomalies in these multivariate continuous data. As shown in Fig. (4), QQ plots have been obtained for the 14 sensors, but due to space limitations results of only four sensors are shown here. There are some abnormal data patterns in the data obtained from sensor#1, sensor#2, sensor#9, sensor#12, sensor#13 and sensor#14, because the data of these sensors is scattered and the histograms in the diagonal of the figure prove that it does not follow the Gaussian distribution of the above mentioned sensors. Thus, from scattered charts a person can judge the pattern of the data, and can perform an analysis according to the quality of the data obtained from the sensors installed at these structures.

4.2.3. QQ Plots

QQ plots are used to visually evaluate two distributions by plotting the quantities of one against the quantities of the

other. QQ plots were drawn for the 14 accelerometers sensors installed at the Sutong Yangtze cable stayed bridge, but due to space limitation results of only six sensors are shown here in Fig. (5). The results show that sensor#1, sensor#2, sensor#9, sensor#12, sensor#13 and sensor#14 do not follow the normal distribution as compared to the other sensors such as sensor#3, sensor#4, sensor#5, sensor#6, sensor#7, sensor#8, sensor#10 and sensor#11, where the QQ plots distribution is normal. It can also be observed that, in case of sensor#13 and sensor#14 the data seems to be discontinuous that is why it does not follow the normal distribution trend; in sensor#9 and sensor#12 the data is not normally distributed, hence the data obtained from sensor#1, sensor#2, sensor#9, sensor#12, sensor#13 and sensor#14 cannot be used for continuous modal parameter identification.

4.2.4. Normal CDF Plots

Lognormal CDF plots are a very effective method for dealing simultaneously with more than one random variable such as with our continuous data. As shown in Fig. (6), the lognormal CDF plots show the explicit appearance of all the 14 accelerometer sensors installed at the Sutong Yangtze cable stayed bridge. It can be observed that there is no log-

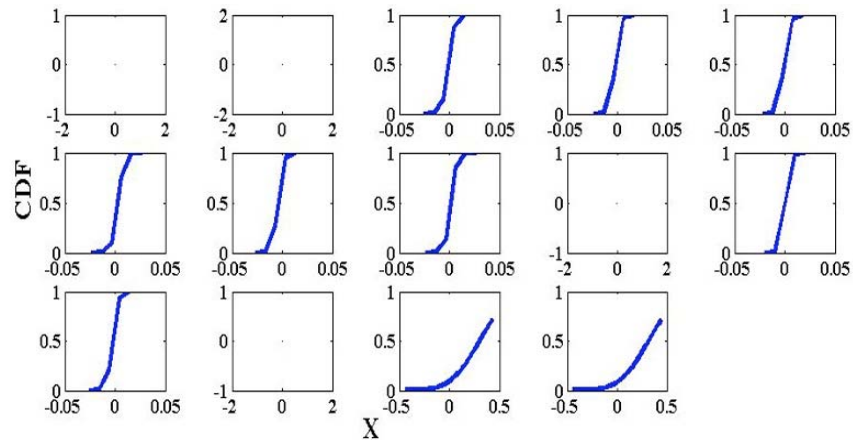


Fig. (6). Normal CDF plots for sensor #1 to sensor #14.

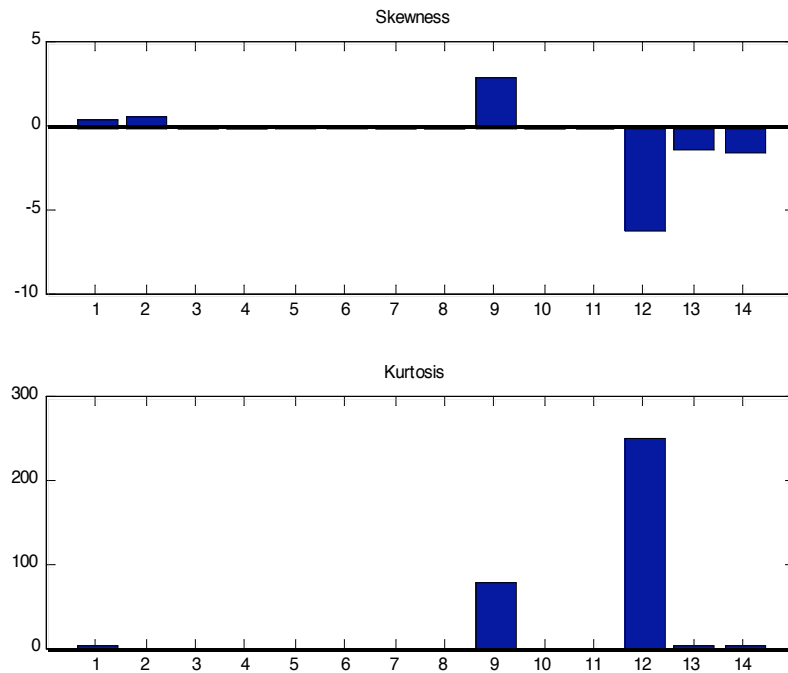


Fig. (7). Skewness and kurtosis for sensor #1 to sensor #14.

normal CDF for sensor#1, sensor#2, sensor#9 and sensor#12, because the data is so poor that the lognormal CDF cannot even be drawn. For sensor#13 and sensor#14, the lognormal CDF curve is incomplete and also the tail is very heavy at the beginning of the curve, and it also suddenly ruptures before completion. Therefore, the above mentioned six sensors cannot be used for further continuous data analysis.

4.2.5. Skewness and Kurtosis

Skewness is the measure or lack of symmetry, whereas kurtosis is the extent to which the data peaks or is in a flat form relative to the normal distribution. As shown in Fig. (7), the skewness of sensor#1, sensor#2, sensor#9, sensor#12, sensor#13 and sensor#14 is abnormal and does not follow the normal distribution, because their values are greater than zero. The value of sensor#3, sensor#4, sensor#5, sensor#6, sensor#7, sensor#8, sensor#10 and sensor#11 are ideal and are almost zero. As for as the kurtosis of the sen-

sors' data is concerned, it shows a value of 3 for all the sensors except sensor#9, where the kurtosis is 80 and for sensor#12 where the kurtosis is 240, which is extremely abnormal. Thus, these two sensors show some anomalies or defects, and so their data cannot be used for modal parameter identification.

4.2.6. Box and Whisker Plot

Box and whisker plots are also called boxplots. Boxplots are an excellent way to visualize the median, data distribution and to supplement multivariate displays with univariate information. The boxplots illustrated for the 14 accelerometer sensors of Sutong Yangtze river bridge are shown in Fig. (8). It can be perceived that sensor#1, sensor#2, sensor#9 and sensor#12, do not show any median and also their data distribution is poor and so the data spread is misleading, whereas for sensor#13 and sensor#14 the data distribution and spread is abnormal, and so these sensors cannot be used for further data analysis. However, the data obtained from

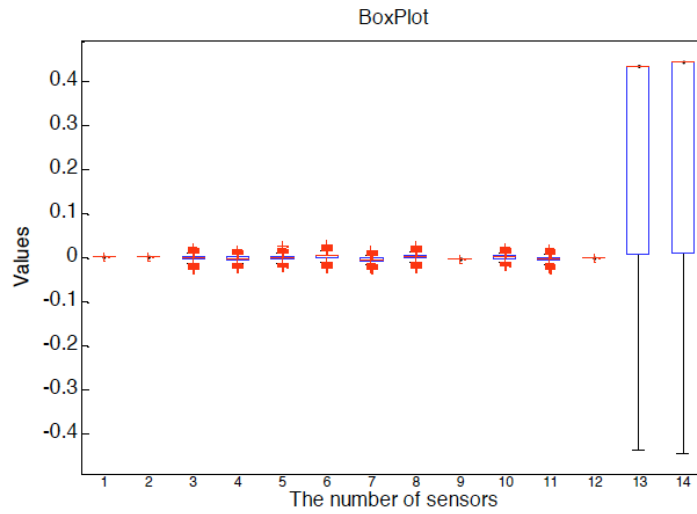
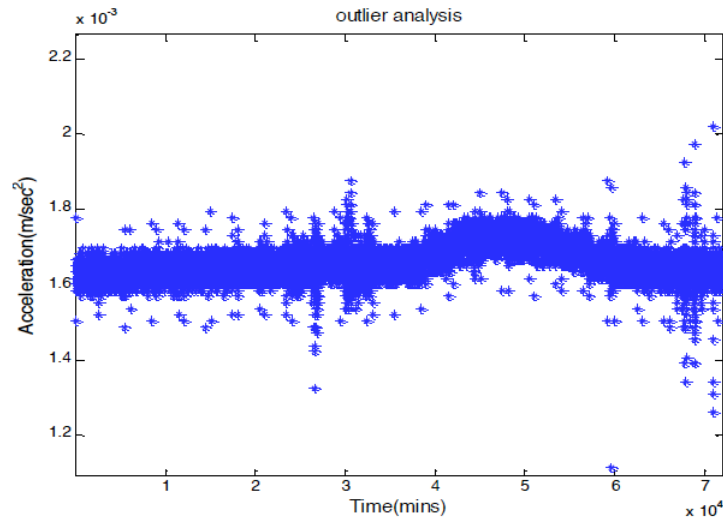
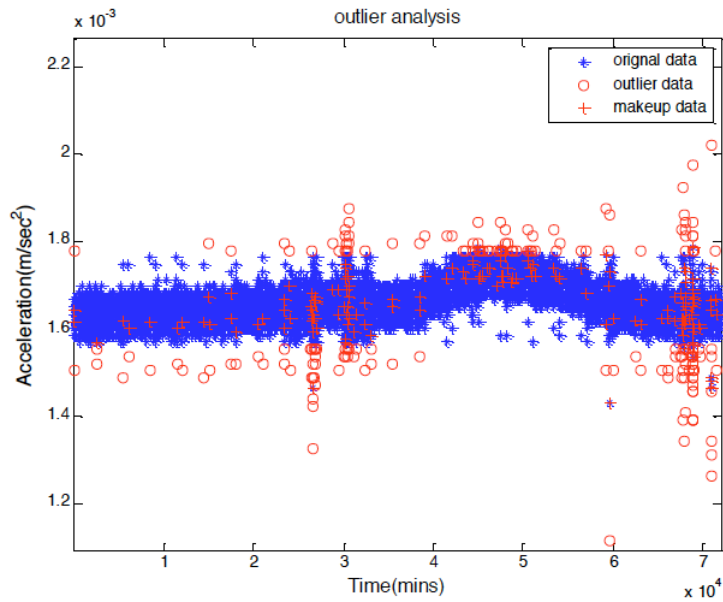


Fig. (8). Box plot for sensor #1 to sensor #14.



(a) Original data of sensor #8



(b) Outlier analysis of sensor #8

Fig. (9). Outlier analysis.

sensor#3, sensor#4, sensor#5, sensor#6, sensor#7, sensor#8, sensor#10 and sensor#11 show relatively good results, and so the data obtained from these sensors will be used for continuous modal parameter identification.

4.3. Outlier Analysis

In this paper, data was collected from the 14 accelerometers installed at the Sutong Yangtze river bridge, but for outlier analysis only the data treatment of sensor#8 will be exhibited in this paper due to space limitations. Outlier analysis is quantified by the extent to which the values do not fall outside the thresholds limits, otherwise they will generate huge problems in the data analysis. In the outlier analysis in this paper, the accelerometers data were investigated with the pauta criterion and those values which had some sort of information were treated and the useful information was utilized in the data analysis. However, those outliers which were of no use and which affected the data analysis were discarded as shown in Fig. (9b). The original data for sensor#8 is shown in Fig. (9a). Thus, an accurate and reliable data analysis can be performed by properly treating the outliers.

4.4. Auto and Cross Correlation Functions

Auto and cross correlation was performed on the data obtained from the 14 accelerometers installed on the girders of the Sutong Yangtze river bridge. Their purpose was to detect anomalies in a very speedy way and introduce the novel concept of auto and cross-correlation, as this technique has never been used to detect anomalies in huge amounts of continuous monitoring data. In Fig. (10), the autocorrelation for sensor #1 to sensor#14 are shown, and the auto correlation in the sensor data can depict the inner secrets of the data. It can easily be seen that, the data obtained from sensor#3, sensor#4, sensor#5, sensor#6, sensor#7, sensor#8, sensor#10 and sensor#11 are all good, whereas the data obtained from the sensor#1, sensor#2, sensor#9, sensor#12, sensor#13 and sensor#14 is in an abnormal condition. Therefore, anomalies can be identified at first glance by performing the auto correlation. Thus, based on this autocorrelation technique, anomalies can be detected and any abnormal data can be discarded, so that the analysis is free from false and defected data.

Similarly, a 3D cross correlation coefficient analysis was performed on the data collected from the 14 accelerometers sensors installed on the girders of Sutong Yangtze river bridge. As can be seen in Fig. (11), the cross correlation coefficients are very good except for sensor#1, sensor#2, sensor#13 and sensor#14, where the cross correlation coefficient are more than 0.5. The diagonal is considered to be perfect if the cross correlation coefficient is 1, and 0 for the other values not in the diagonal. In Fig. (11), there are also correlations and the values other than diagonal which are not zero, especially for sensor#1, sensor#2, sensor#13 and sensor#14, thus showing anomalies in sensor#1, sensor#2, sensor#13 and sensor#14. This proves that the cross correlation coefficient analysis is an easy non-physical model technique which can be used to ascertain the calculated retort from time, and which may be a first step in identifying any damage. At the same time, all the correlation values must be steady which is indicative of the normal condition, as is the case in the diagonal as shown in Fig. (11). When damage or the malfunctioning of the sensors occurs, these values will change, as can be seen in Fig. (11) for sensor#1, sensor#2, sensor#13 and sensor#14.

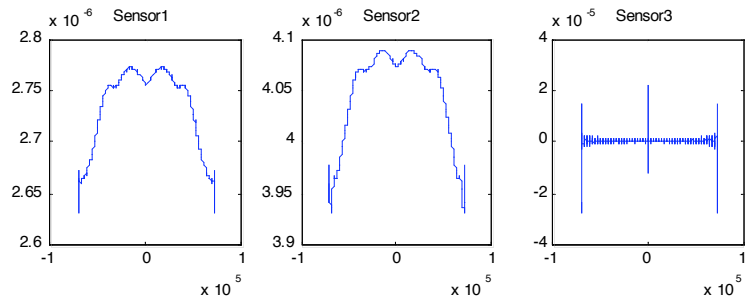
tioning of the sensors occurs, these values will change, as can be seen in Fig. (11) for sensor#1, sensor#2, sensor#13 and sensor#14.

4.5. Modal Parameter Identification

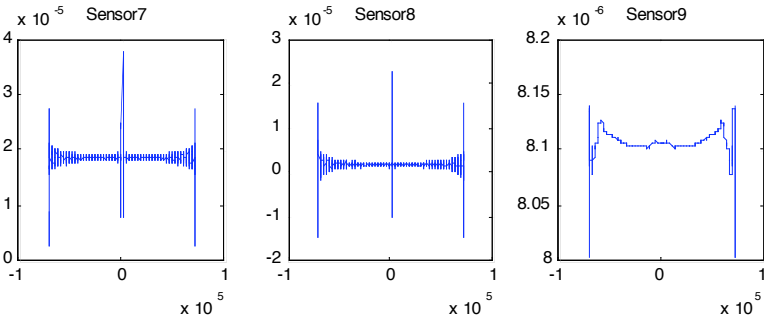
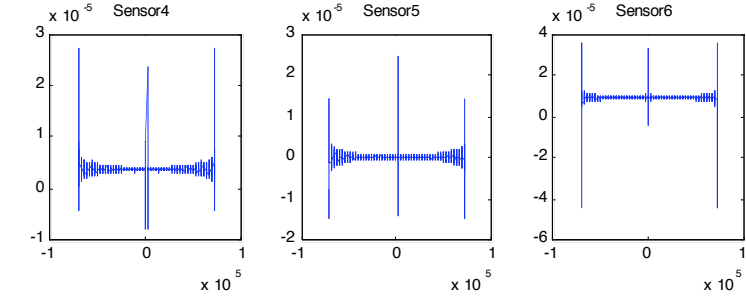
In order to identify the continuous modal parameters, the collected data from the 14 sensors were first processed by performing EDA. Then, outlier analysis was performed to detect the outliers in the data and possible treatments were carried out on the data which were not outliers in order to make the data usable. After this, auto and cross correlation was performed to further identify the anomalies and defects in the sensors' data. Finally, the continuous modal parameters, such as frequency and damping ratios were identified for a period of one hour using the CO-SSI technique based on SWT. For more stability and accuracy, the window length was kept at 6000 points and block row i at 200.

Fig. (12) shows the frequency values obtained before and after the newly suggested method through using MATLAB. Fig. (12a) shows that the frequency is scattered and unstable, and thus the frequency order cannot be determined very accurately. In addition, the lower modes cannot be identified due to anomalies in the data, and so false and unwanted frequency estimates are obtained, which will eventually lead to wrong structural information due to the absence of first few modes. The main reason behind its instability is because of the wrong data from these sensors, thus detection of such anomalies at early stage is very important, because it will make the whole identification process nastiest at later stage as can be seen in Fig. (12a). In case of Fig. (12b), the frequencies are stable and frequency modes are clear and visible and a regular trend can be seen from left to right and we can very accurately determine the frequency orders, for instance the first 12 orders are shown in Table 1. It thus shows us that the obtained results are very accurate and useful. As the identified accuracy of the frequency gradually decreases along with the increasing order in the engineering measurement, the first few orders therefore carry more information and are considered to be more reliable when compared to the higher order frequencies. In our proposed method, we can identify the first few orders very accurately, which illustrate the supremacy of our novel suggested method, whereas before our proposed method the first six order cannot be identified as shown in Table 1. It can also be deduced from the successful outcomes that sensor's data is more robust in terms of vibration, because there are generally more vibration in cable stayed bridges, and so the results are an accurate depiction of the real life conditions.

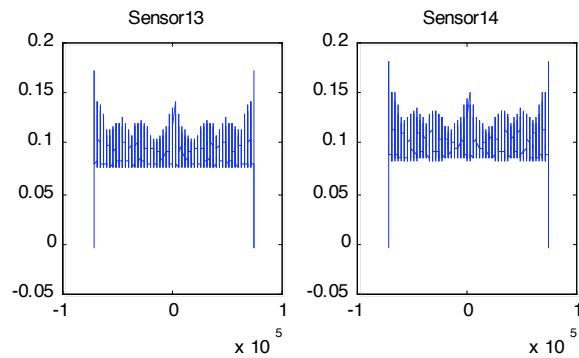
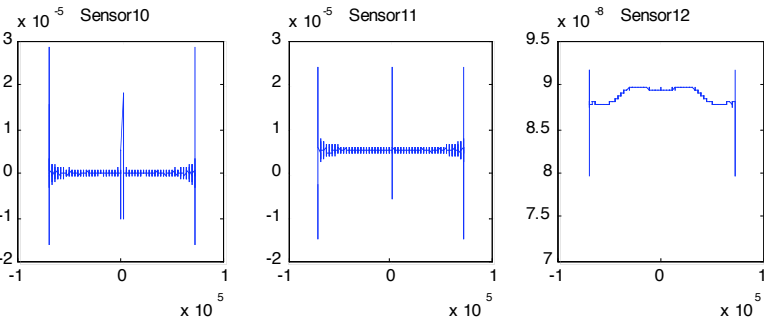
Fig. (13) shows the damping ratio estimates obtained before and after the proposed method by using MATLAB. The results obtained in Fig. (13b) show that the damping ratio is not more than 2% and is regular and stable when compared to the damping estimates obtained in Fig. (13a). In Fig. (13b), the damping ratio is stable and more pronounced, where as in Fig. (13a) the damping ratio is scattered and unstable. Visually it may not be seen clearly, but the first six damping ratios cannot be identified before the proposed method as shown in Table 1. Although damping ratios for real life continuous data are hard to be traced clearly, even if the window length and block rows are increased.



(a) Auto correlation for sensor #1 to #6



(b) Auto correlation for sensor #7 to #12



(c) Auto correlation for sensor #13 to #14

Fig. (10). Auto correlation for sensor #1 to sensor #14.

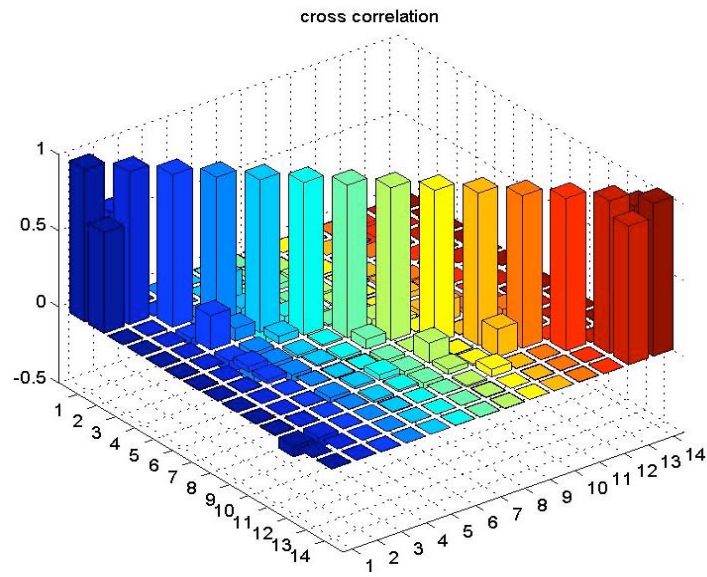


Fig. (11). 3D cross correlation coefficient form sensor #1 to sensor #14.

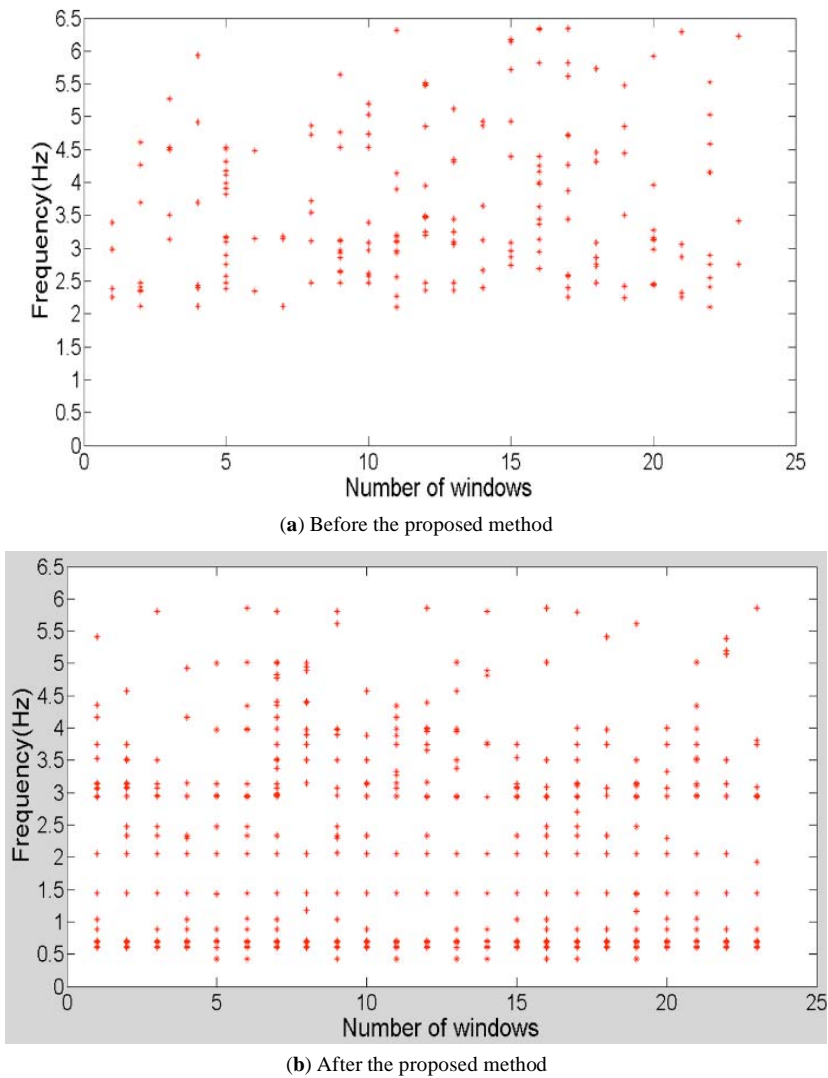
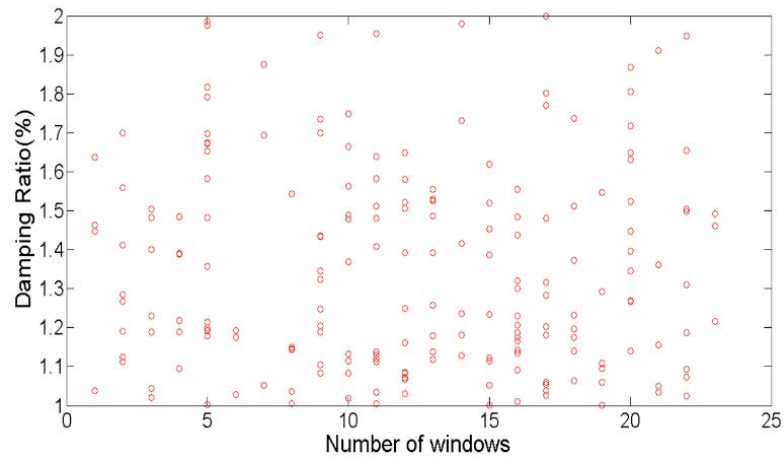


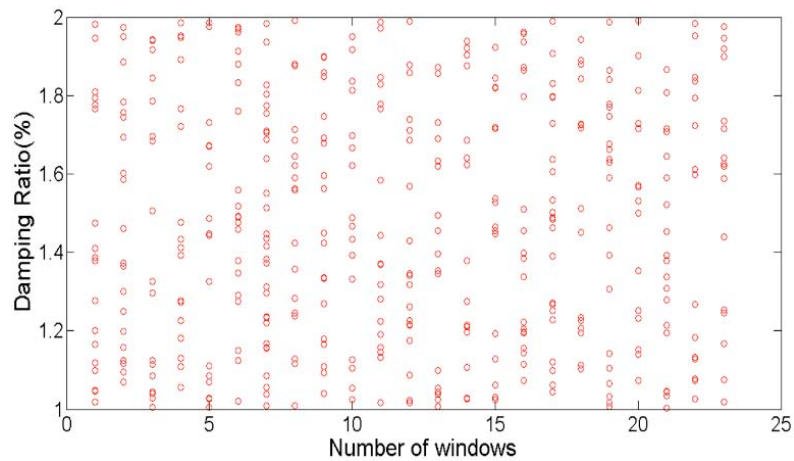
Fig. (12). Frequency estimates.

Table 1. Frequency and damping ratio before and after the proposed method.

Order	Frequency		Damping	
	Before	After	Before	After
1	Frequency Cannot be Identified	0.43	Damping Cannot be Identified	1.005
2		0.62		1.01
3		0.69		1.019
4		0.89		1.028
5		1.04		1.041
6		1.44		1.047
7	2.09	2.06	1.049	1.056
8	2.26	2.34	1.094	1.069
9	2.38	2.48	1.112	1.10
10	2.47	2.94	1.175	1.118
11	2.98	3.07	1.215	1.2
12	3.14	3.15	1.267	1.277



(a) Before the proposed method



(b) After the proposed method

Fig. (13). Damping ratio estimates.

Achieving this much minute accuracy in the Frequency estimates and the damping ratios obtained in Figs. (12) and (13) is very meaningful and coherent, as it shows the efficiency and robustness of our proposed method for identifying the continuous modal parameters. In the end, this paper suggests that more detailed research is required to better understand the influence of the environmental factors, especially the effect of temperature on the natural frequencies and the damping estimates in order to better understand the identification process.

CONCLUSION

In this study, continuous modal parameter identification for a cable stayed bridge was carried out by exploiting the approach to the interpretation of data. Firstly, the data was collected from the 14 accelerometers sensors installed at the Sutong Yangtze river cable stayed bridge. EDA was then performed on the data to quickly detect any anomalies in the long term continuous data, by plotting the histograms, scattered, QQ's, normal CDF's, skewness and kurtosis and box plots, to visualize any abnormal sensor's data. The results showed that sensor#3, sensor#4, sensor#5, sensor#6, sensor#7, sensor#8, sensor#10 and sensor#11 were reliable and were free from anomalies, whereas sensor#1, sensor#2, sensor#9, sensor#12, sensor#13 and sensor#14 were defective and probably unsuitable for continuous monitoring. Outlier analysis was then performed as a pre-treatment of the data and those data points which exhibited abnormal behavior were detected as outlier and discarded, then to further verify the data by detecting the anomalies and malfunctioning of the sensors' data, auto and cross correlation was performed. This paper thus introduces a new concept for the quick long-term data visualization of long term health monitoring data, because if such anomalies are not detected at the early stages, they will instigate a huge amount of complications at the later stages.

Finally, in order to identify the continuous modal parameters, CO-SSI was used because CO-SSI is a robust, clear and more accurate technique which suffer from only one drawback, that of processing the data in one batch, and therefore it cannot be used as a continuous identification technique. SWT was introduced in this paper to remedy this drawback and to accurately identify the continuous modal parameter. The above proposed method combines the advantages of two methods; firstly, any anomalies can be detected through quick visualization to ensure that only accurate data is used to ensure an accurate analysis; secondly, it can be used to perform the continuous modal parameter identification in a very robust manner due to its numerical stability, by employing the SWT based CO-SSI method. In the end, continuous modal parameters such as the frequency and the damping ratios were identified using the above technique which indicates the increase in tracking stability with the introduction of SWT. This shows that the obtained results of the new method are very effective as they can provide constructive information regarding the long-term identification of the state of civil structures; it also means that the new method can be substantiated as a very effective tool in the real life health monitoring of bridges.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

The research reported here has been conducted as a part of the result of a series of research projects granted by the National Key Basic Research and Development Plan (973 Program) with 2013CB036302, National Science Foundation with 51078316, Sichuan Science and Technology Program with 2011JY0032, Railway Ministry Science and Technology Research and Development Program with 2011G026-E & 2012G013-C.

REFERENCES

- [1] G. Vecili, A.H.O. Halim, A. Braimah, and O. E. Desouky, "Transportation optimization of ribbon floating bridges: analytical and experimental investigation", *The Open Civil Engineering Journal*, vol. 8, pp. 42-56, 2014.
- [2] L. Chen, L. Jiang, W. Long, and L. Wang, "Research on seismic response and damping effect for high-speed railway seismic isolated bridge", *The Open Civil Engineering Journal*, vol. 5, pp. 163-167, 2011.
- [3] R. Sparks, and A. Kasmarik, "Monitoring deterioration in a catchment's sewerage system", *The Open Civil Engineering Journal*, vol. 7, pp. 149-158, 2013.
- [4] I. U. Khan, D. S. Shan, and A. Q. Bhatti, *Temperature Effect Analysis of Viscoelastic Damper and Magnetorheological Damper for Vibration Control of Stayed-Cable*, In: Bridge Maintenance Safety Management & Life Extension (A. Chen, M.D. Frangopol, X. Ruan, (Eds), CRC Press: Shanghai, China, 2014, pp. 2167-2175.
- [5] E. J. Cross, K. Y. Koo, J. Brownjohn, and K. Worden, "Long-term monitoring and data analysis of the Tamar Bridge", *Mechanical Systems and Signal Processing*, vol. 35, pp. 16-34, 2013.
- [6] N. Wu, C. Liu, Y. Guo, and J. Zhang, "On-board computing for structural health monitoring with smart wireless sensors by modal identification using hilbert-huang transform", *Mathematical Problems in Engineering*, vol. 2012, (Article ID 509129) pp. 1-9, 2013.
- [7] J. W. Tukey, *Exploratory Data Analysis*, Addison-Wesley Publishing Company, 1977.
- [8] J. M. Chambers, *Programming With Data*. Springer-Verlag, New York, 1998.
- [9] W. S. Cleveland, *The Elements of Graphing Data*. Wadsworth Monterey: CA, 1985.
- [10] W. S. Cleveland, and R. McGill, "Graphical perception: theory, experimentation, and application to the development of graphical methods", *Journal of the American Statistical Association*, vol. 79, pp. 531-554, 1984.
- [11] E. R. Tufte, *The Visual Display of Quantitative Information*, Graphics Press Cheshire: CT, 1983.
- [12] E. R. Tufte, *Envisioning Information*, Graphics Press Cheshire: CT, 1990.
- [13] A. Buja, D. Cook, and D. Swayne, "Interactive high-dimensional data visualization", *Journal of Computational and Graphical Statistics*, vol. 5, pp. 78-99, 1996.
- [14] H. Wainer, *Visual Revelations*, Springer-Verlag: New York, 1997.
- [15] D. R. Brillinger, H. K. Preisler, A. A. Agerc, and J. G. Kiec, "An exploratory data analysis (EDA) of the paths of moving animals", *Journal of Statistical Planning and Inference*, vol. 122, pp. 43-63, 2004.
- [16] M. M. Rowland, L. D. Bryant, B. K. Johnson, J. H. Noyes, M. J. Wisdom, and J. W. Thomas, "The Starkey Project: History, Facilities, and Data Collection Methods for Ungulate Research", General Technical Report, PNW-GTR-396:1-62, US Department of Agriculture, Forest Service, Pacific Research Station, 1997.
- [17] Y. C. Liu, C. H. Loh, and Y. Q. Ni, "Stochastic subspace identification for output-only modal analysis: application to super high-rise tower under abnormal loading condition", *Earthquake Engineering & Structural Dynamics*, vol. 42, pp. 477-498, 2012.

- [18] D. S. Shan, and Q. Li, "Modal parameter identification of girder bridge based on covariance based stochastic subspace algorithm," In: *International Conference on Environmental Vibrations: Prediction, Monitoring and Evaluation*, Beijing, 2009.
- [19] F. Magalhães, E. Caetano, and Á. Cunha, "Challenges in the application of stochastic modal identification methods to a cable-stayed bridge", *Journal of Bridge Engineering*, vol. 12, no. 6, pp. 746-754, 2007.
- [20] B. Peeters, and G. D. Roeck, "Reference-based stochastic subspace identification for output-only modal analysis", *International Journal of Mechanical Systems and Signal Processing*, vol. 13, no. 6, pp. 855-878, 1999.
- [21] B. Peeters, and G. D. Roeck, "Stochastic system identification for operational modal analysis: a review", *Journal of Dynamic Systems, Measurement, and Control*, vol. 123, pp. 659-666, 2001.
- [22] P. V. Overschee and B. D. Moor, "Subspace identification for linear systems: Theory – Implementation Applications," Katholieke Universiteit Leuven, Belgium, Kluwer: Academic Publishers, Boston, London, Dordrecht, 1996.
- [23] M. Bassville, A. Benveniste, M. Goursat, and L. Mevel, "In-flight vibration monitoring of aeronautical structures," *IEEE Control System Magazine*, vol. 27, no. 5, pp. 27-41, 2007.
- [24] J. H. Weng, C. H. Loh, J. P. Lynch, P. Y. Linn, and Y. Wang, "Output-only modal identification of a cable-stayed bridge using wireless monitoring systems", *Journal of Engineering Structures*, vol. 30, no. 2, pp. 1802-1830, 2008.
- [25] V. Bonyapinyo, and T. Janesupasaeree, "Data-driven stochastic subspace identification of flutter derivatives of bridge decks", *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 98, pp. 784-799, 2010.
- [26] I. Goethals, L. Mevel, and B. D. Moor, "Recursive output only subspace identification for in flight flutter monitoring," In: *22nd International Conference*, Dearborn, Michigan, US, 2004.
- [27] D. Huston, *Structural Sensing, Health Monitoring, Performance Evaluation*. Taylor and Francis Group: LLC: London, 2011, pp. 1235-4.
- [28] J. E. Freehafer, *Geometrical Optics, In Propagation of Short Radio Waves*. Dover Publications, Mineola: NY, 1965.
- [29] H. Sohn, J. J. Czarnecki, and C. R. Farrar, "Structural health monitoring using statistical process control", *Journal of Structural Engineering*, ASCE, vol. 126, pp. 1356-1363, 2000.
- [30] C. R. Farrar, and K. Worden, *Introduction to Probability and Statistics, Structural Health Monitoring: A Machine Learning Perspective*, John Wiley & Sons, 2012.
- [31] S. Karimi, and A. H. Darooneh, "Measuring persistence in a stationary time series using the complex network theory", *Physica A Statistical Mechanics and its Applications*, vol. 392, no. 1, pp. 287-293, 2013.
- [32] M. Fischer, and B. A. Jensen, "Taxation, transfer income and stock market participation", *Review of Finance*, vol. 19, no. 2, pp. 823-863, 2014.
- [33] I. U. Khan, D. S. Shan, and K. Malik, "Covariance driven subspace identification technique for continuous modal parameters identification of sutong bridge," In: *IABSE Madrid Symposium: Engineering for Progress, Nature and People*, 2014, pp. 2666-2673(8).

Received: April 01, 2015

Revised: May 15, 2015

Accepted: June 08, 2015

© Khan et al.; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.