ISAV2015-5245

A New Similarity Measure Method based on Statistical Pattern Recognition for Structural Health Monitoring

Alireza Entezami[1], Hashem Shariatmadar[2*]

[1] *Ph.D student, Department of Civil Engineering, Ferdowsi University of Mashhad, Mashhad, Iran, entezami@stu-mail.um.ac.ir*

[2] *Associate Professor, Department of Civil Engineering, Ferdowsi University of Mashhad, Mashhad, Iran, shariatmadar@um.ac.ir*

Abstract

This study presents a new similarity measure method in order to detect and locate structural damage based on statistical pattern recognition paradigm. The similarity method is Bhattacharyya distance (BD) method that enables to calculate the distance or similarity between two random distributions with aid of their statistical properties. Feature extraction is an important step in the statistical pattern recognition process, which is conducted through fitting autoregressive (AR) models to the time series data in order to extract residuals of the models as the damage- sensitive feature. Based on the proposed methods, the similarity between pair sensors in baseline and damaged conditions is calculated in such a way that a sensor location associated with the largest distance value is identified as a damaged area. The performance and effectiveness of the proposed approaches are then verified by acceleration time histories from a three-story laboratory frame. Results show that the BD method based on using the residuals of AR model can detect and locate the damage precisely. Furthermore, this method is approximately able to estimate the damage severity.

5th ISAV2015
International Conference on
Acoustics and Vibration
25-26 Nov 2015 Tehran - Iran
Iranian Society of Acoustics and Vibration
University of Tehran

A New Similarity Measure Method based on Statistical Pattern Recognition for Structural Health Monitoring

Alireza Entezami [a], Hashem Shariatmadar [b*]

[a] *Ph.D student in Structural Engineering, Department of Civil Engineering, Ferdowsi University of Mashhad, Azadi Square, 9177948974, Mashhad, Iran*

[b] *Associate Professor, Department of Civil Engineering, Ferdowsi University of Mashhad, Azadi Square, 9177948974, Mashhad, Iran*

* *Corresponding author e-mail: shariatmadar@um.ac.ir*

Abstract

This study presents a new similarity measure method in order to detect and locate structural damage based on statistical pattern recognition paradigm. The similarity method is Bhattacharyya distance (BD) method that enables to calculate the distance or similarity between two random distributions with aid of their statistical properties. Feature extraction is an important step in the statistical pattern recognition process, which is conducted through fitting autoregressive (AR) models to the time series data in order to extract residuals of the models as the damage- sensitive feature. Based on the proposed methods, the similarity between pair sensors in baseline and damaged conditions is calculated in such a way that a sensor location associated with the largest distance value is identified as a damaged area. The performance and effectiveness of the proposed approaches are then verified by acceleration time histories from a three-story laboratory frame. Results show that the BD method based on using the residuals of AR model can detect and locate the damage precisely. Furthermore, this method is approximately able to estimate the damage severity.

Keywords: Structural health monitoring; Time series analysis; Residual errors; Bhattacharyya distance method.

1. Introduction

Structural health monitoring (SHM) is a theoretical and operating procedure to detect damage in civil, mechanical, and aerospace structures by vibration data. The main aim of SHM is based on improving the safety and reliability of the structure by detecting the damage before it reaches critical states. In general, this process has got two phases: (1) hardware phase through sensing technology and, (2) software phase by theoretical-computational algorithms for damage detection. On the other hand, SHM can be carried out through constructing a finite element model and use the physical properties of the structure for the damage identification, which is known as model-based method. By contrast, it is possible to apply the raw measurement of vibration data with aid of statistical methods that fall into the data-based method. Although, the first method requires an appropriate FE model to reflect the actual behaviour of the real structure, there is a noteworthy difference between dynamic responses of analytical and real structure. Thus, some model updating techniques should be used to reduce their differences.

Nowadays, it is preferred to apply the data-based approach due to its applicability, simplicity, and capability in comparison with the model-based method. The data-based method, on the other hand, focuses on using statistical pattern recognition paradigm in the four-step process: (1) operational evaluation, (2) data acquisition, (3) feature extraction, and (4) statistical modelling for feature classification [1-3]. Even though all the steps are separately significant in the process of SHM, the feature extraction and the statistical modelling have got the most importance for the reason of the sensitivity of final decision of the damage detection procedure to the these steps. An inappropriate feature extraction algorithm, for instance, leads to undesirable results even if other steps to be performed properly. On the other hand, the statistical modelling is concerned with the implementation of algorithms that analyse statistical characteristics extracted from the feature extraction process in an effort to determine the damage state in a structure.

By considering the two last steps of statistical pattern recognition paradigm, numerous approaches have been published that make a new way or deal with existing limitations and drawbacks. In this regard, Sohn et al. [4] applied the coefficients of AR model as the damage feature to use in the statistical process control that only enables to detect the presence damage in the structure. Nair et al. [5] used an autocorrelation moving average (ARMA) model and applied the first three AR parameters as the damage-sensitive feature, which were able to identify and locate the damage. Omenzetter and Brownjohn [6] used autoregressive integrated moving average (ARIMA) models to analyse the static strain data from a bridge during construction and serviceability. Worden et al. [7] used transmissibility function for feature extraction along with Mahalanobis squared distance measure to detect the structural damage. Gul and Catbas [8] proposed random decrement technique for averaging the time series data to extract the damage-sensitive features and also used the Mahalanobis distance based outlier detection algorithms to identify difference types of structural changes on different test structures. Yao and Pakzad [9] proposed new feature extraction techniques using AR model spectra along with its residual autocorrelation function and verified their methods by several statistical pattern recognition algorithms including control charts, Ljung-Box test statistic, Mahalanobis distance and Cosh spectral distance measure.

The objective of this paper is to identify and locate the structural damage by a new similarity measure method with aid of the time series analysis of vibration measurements. To achieve these aims, Bhattacharyya distance method is introduced as a similarity measure method that is able to distinguish variations of statistical features between undamaged and damaged structures. The residual errors of AR model are chosen as the damage-sensitive feature because it is more efficient than the coefficients of the model in the distance techniques. The applicability and effectiveness of the proposed method are experimentally verified using acceleration measurements a three-story laboratory frame with different types of nonlinear damage cases. Results obtained from this structure show that the BD method is an appropriate tool for identification of damage location and also approximately enables to estimate the damage severity.

2. Time series analysis by autoregressive model

Time series is a sequence of data points that typically consists of successive measurements produced over a time interval. From a statistical point of view, these measurements are considered to be the realization of random variables that have certain probability distributions. Time series analysis, on the other hand, attempts to fit a model to the measurements for analysing time series data. Because of using the time series data in the most data-based methods, time series analysis should be applied to extract statistical characteristics of model such as coefficients and residual errors that are known as the damage-sensitive features [4].

Selection of a time series model, in general, depends on the type and property of time series data. The autoregressive model is usually used in the SHM process, because this model is consistence to vibration data in the time domain. The visual interpretation of time series observations such as using the autocorrelation function (ACF) and partial autocorrelation function (PACF) one can assist to choose adequate time series model. On the other hand, goodness of fit is an appropriate practice to check the adequacy of fitted model [10].

Autoregressive model (AR) is a linear time series model that has the simplest structure in comparison with other models. The AR model specifies that the response variable depends linearly on its own previous values, hence, this model one can estimate the value of a function based on linear combination of its prior value. The model order determines the number of past quantities applied to estimate the value at time *t*. For more details about the theory of time series analysis can be found in the technical literatures [10-12]. An AR model with *p* autoregressive terms, AR(*p*), can be written as:

$$x(t) = a + \sum_{j+1}^{p} \phi_j x(t-j) + e(t) \tag{1}$$

where *x(t)* is acceleration measurements observed at time *t* and *e(t)* denotes an unobservable random error (residual) with zero mean and constant variance. Furthermore, ϕ is coefficient of the AR model that should be estimated. The residual error is the difference between the measured and predicted signal which is calculated at time *t* as follows:

$$e(t) = x(t) - \hat{x}(t) \tag{2}$$

where $\hat{x}(t)$ is the predicted time series signal through AR model. The idea beyond the residual errors as the damage-sensitive feature is that an AR model is fitted to the healthy condition of structure (baseline condition). When this model is desirably modelled, the residuals between measured and predicted data are insubstantial quantities. Subsequently, the time series response of damaged structure is predicted by this model and can be observed that the residuals of this structure will be increased [13]. To put it another way, the damage leads to increasing in the residuals of damaged structure. Selection of the model order (*p*) and estimation of its coefficient are next steps after fitting a model to time series measurements in the time series analysis. Akaike information criterion [14] and Bayesian information criterion (BIC) are two common information criteria that can be applied to choose the model order [10].

3. Bhattacharyya distance method

In statistics, the Bhattacharyya distance [15] method measures the similarity of two discrete or continuous probability distributions. It is closely related to the Bhattacharyya coefficient which is a measure of the amount of overlap between two statistical samples or populations [16]. It is worthwhile remarking that even though the concepts of similarity and distance are not exactly the same, the principle of the similarity measure is related to the concept of statistical distance. Similarity is used to measure the common characteristics between two distribution data whereas distance is adopted to indicate the differences between them. Nonetheless, there is still a strong relation between similarity and distance, particularly in the context of SHM when the most data are based on time series. On the basis of BD method, zero distance measure relies on this fact that the structure is in the healthy condition whereas corresponding distance measure larger than zero indicates the damaged state. For discrete probability distributions *X* and *Y* over the same time domain *t*, the Bhattacharyya distance is defined as:

$$BD\left(X,Y\right)=-ln\left(\sum_{i\in t}\sqrt{X_iY_i}\right) \tag{3}$$

The time series data, particularly the residual errors, are continuous random distributions. For these distributions, therefore, the Bhattacharyya distance BD is expressed in the following form:

$$BD\left(X,Y\right)=-ln\left(\int\sqrt{X_iY_i}dt\right) \tag{4}$$

In the simplest form, the Bhattacharyya distance between two separate distributions under the normal distribution can be calculated by extracting the mean and variances of these distributions as follows:

$$BD\left(X,Y\right)=-\frac{1}{4}ln\left(\frac{1}{4}\left(\frac{\sigma_x^2}{\sigma_y^2}+\frac{\sigma_y^2}{\sigma_x^2}+2\right)\right)+\frac{1}{4}\left(\frac{\left(\mu_x-\mu_y\right)^2}{\sigma_x^2+\sigma_y^2}\right) \tag{5}$$

where σ_x and μ_x are the variance and mean of distribution X, respectively. Furthermore, σ_y and μ_y denote the corresponding variance and mean for random distribution Y, respectively. Based on Eq. (5), the first term of Bhattacharyya distance is concerned with the variances of two distributions. When these variances are same together, the first term of the distance is zero, that is, the distance depends on the mean of distributions. The Bhattacharyya distance is similar to the Mahalanobis distance, however, it is considered to be more reliable than the Mahalanobis distance. Indeed, the Mahalanobis distance is a particular case of the Bhattacharyya distance when the standard deviations of the two classes are the same. Therefore, when two classes have similar means but different standard deviations, the Mahalanobis distance would tend to zero, however, the Bhattacharyya distance would grow depending on the difference between the standard deviations.

4. Experimental verification

To evaluate the capability and precision of the proposed method, the acceleration measurements of a three-story laboratory frame is used. This frame and testing process belong to the Engineering Institute (EI) at Los Alamos National Laboratory [17] and can be downloaded from [17]. The schematic and sensor locations of laboratory frame are shown in Fig 1. A comprehensive details of descriptions of the structure are given in [13, 18]. The structure was instrumented with four accelerometers mounted at the centre line of each floor on the opposite side from the excitation source to measure the acceleration time history response. The sensor signals were sampled at 320 Hz for 25.6 seconds in duration, which are discretized into 8192 data sampled at 3.125 microsecond intervals. To induce the structural damage, a centre column was suspended from the third floor. This column was contacted a bumper mounted on the second floor, which the position of the bumper could be adjusted to define diverse structural damage. The source of the damage is a simulation of fatigue cracks to induce nonlinear behaviour that subsequently open and close under excitation forces. The structural state conditions can be categorized into four main groups: (1) based line condition (state #1), (2) simulation of operational and environmental variability by increasing the mass and decreasing the stiffness (states #2-#9), (3) the damaged conditions (states #10-#14) and, (4) the damaged conditions along with the operational and environmental variability by mass increasing (states #15-#17). Table 1 indicates some applicable and important states used in this study.

Table 1. Structural conditions for health monitoring in the laboratory frame

State No.	Condition	Description
1	Undamaged	Baseline condition
7	Undamaged	87.5% stiffness reduction in two columns of the 2nd inter-story
10	Damaged	Distance between bumper and column tip 0.20 mm
14	Damaged	Distance between bumper and column tip 0.05 mm
17	Damaged	Gap is 0.10 mm between column and bumper, 1.2 kg mass added at the base

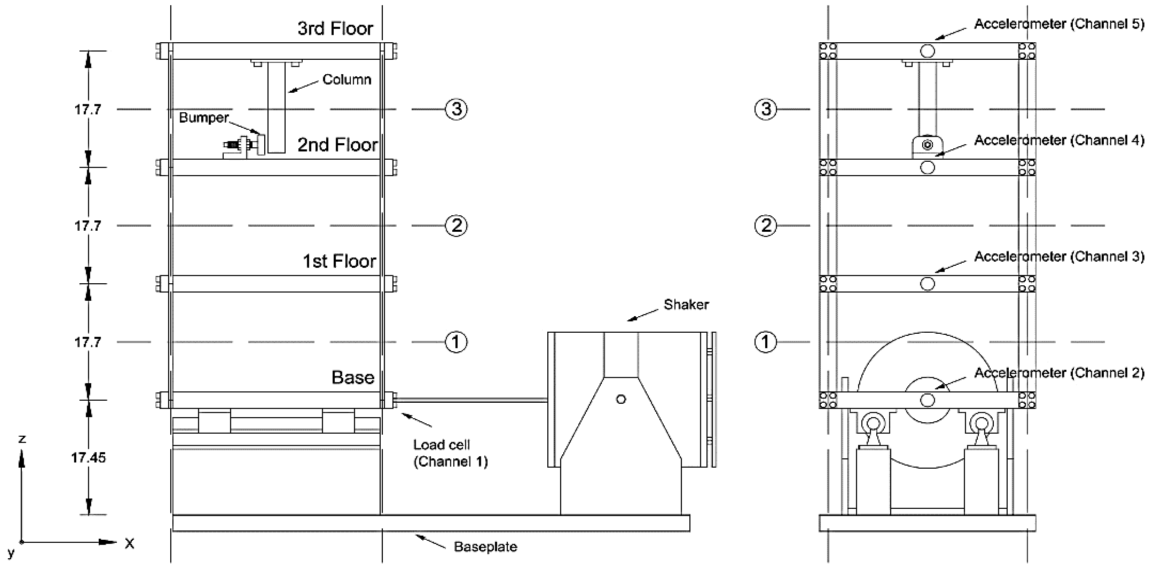


Figure 1. The three-story laboratory frame [13]

It is worth noting that the state #7 in [13] has been introduced as undamaged condition, which the stiffness reduction has been simulated as the influence of operational and environmental conditions on the vibration data. . Hence, in this paper, this state is considered to evaluate the capability of the distance method in the operational and environmental variability. In addition, the state #17 considers 1.2 kg mass added at the base as another type of operational and environmental variability. As stated in the Section 2, the autocorrelation and the partial autocorrelation functions are proper graphical tools in order to choose an appropriate time series model. In such cases, the constant variations in the ACF plot is an indicator of moving average (MA) model whereas the constant variation in the PACF implies the AR model for time series data [10]. Fig. 2 indicates the ACF and PACF plots at the channel #5 in the state#14, respectively.

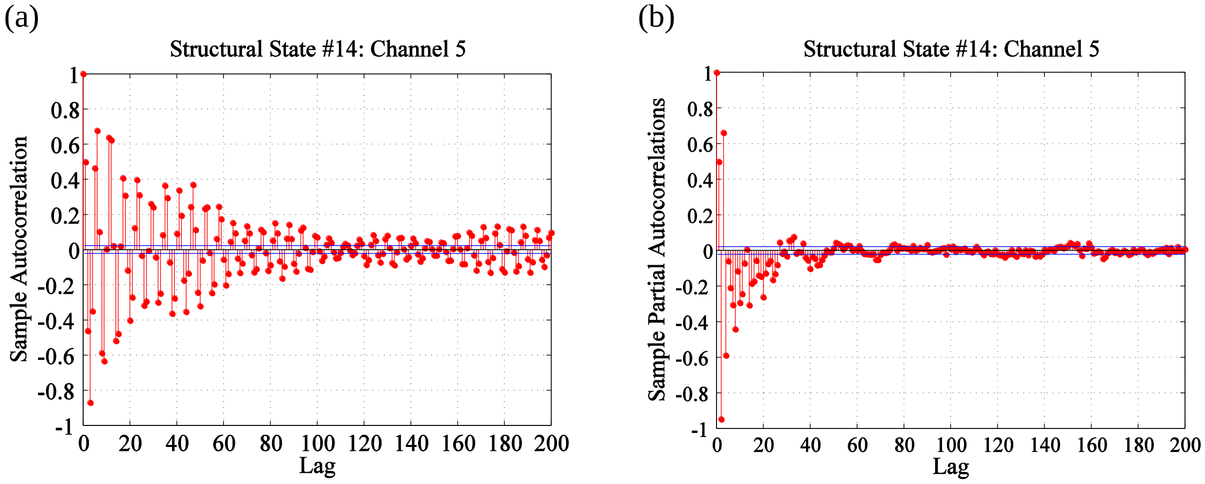


Figure 2. Graphical interpretation of choosing AR model in the structural state #14 at the channel 5: (a) autocorrelation function, (b) partial autocorrelation function

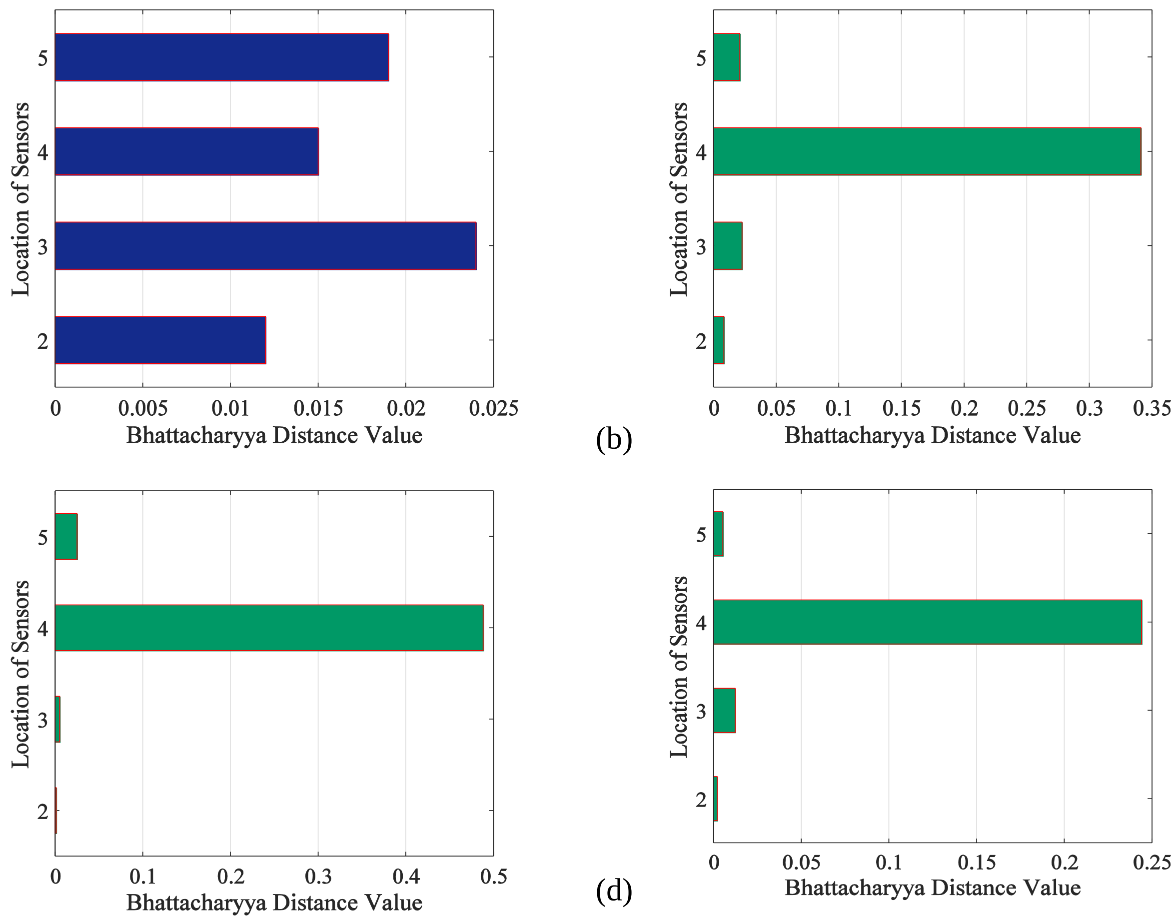
According to above figures, the autocorrelation function (ACF) has got exponential decreasing style and partial autocorrelation function (PACF) shows a constant variation after Lag 20. Such circumstances demonstrate that the AR model is an appropriate time series model for fitting the acceleration responses. It should be noted that other acceleration measurements have similar statistical property; thus, the AR model one can utilize for all the time series data.

After choosing the time series model, Bayesian information criterion (BIC) is applied to determine the model's orders. In the most techniques, the orders of time series model are estimated by choosing a reference channel (sensor) and then the response of damaged structure are predicted with this model [8, 18, 19]. It is a great ambiguity, because each sensor has unique information about itself and selection of a reference sensor cannot provide appropriate model for other sensors. In this study a new algorithm for this issue is investigated in such a way that all sensors of the baseline condition (healthy) are taken as reference channels and the residual errors of each sensor in the damaged conditions will be extracted with corresponding channel of the baseline condition. Table 2 represents the appropriate order for AR model obtained from Bayesian information criterion.

Table 2. The AR coefficients of the laboratory frame based on Bayesian information criterion

Structural state	Channel No.			
	Channel 2	Channel 3	Channel 4	Channel 5
State 1	36	28	12	16
State 7	33	28	12	14
State 10	37	28	11	17
State 14	37	31	11	16
State 17	35	27	12	18

After modelling the AR model for all channels and estimate their coefficients, the residuals of undamaged and damaged states are collected to use in the BD method. On the basis of Eq. (5), assuming that the vector X denotes the residuals of healthy structure and the vector Y represents the residual errors in the damaged states including states #10, #14, and #17. In order to evaluate the influence of environmental and operational conditions in the Bhattacharyya distance method, state #7 is considered as a damaged condition and its residuals are assigned to the vector Y.



(a) (b) (c) (d)

Figure 3. Identification of damage location by BD method in the laboratory frame: (a) state #7, (b) state #10, (c) state #14, (d) state #17

As can be seen, Figs. 3(b), 3(c), and 3(d) clearly show the locations of damage with the highest and striking columns in their bar figures. These columns obviously reveal that the damage has been occurred at the 2^{nd} story or at the location of the sensor number 4 where the bumper (location of induced damage) has been installed there. On the other hand, Fig. 3(a) belongs to the values of BD distance in the state #7. As expected, this figure does not yield the meaningful result about damage. Thus, it can be understood that the variations of its measurements along with the extracted features (residuals) do not pertain to the damage. In other words, the residual errors extracted from state #7 are not damage-sensitive feature.

One point that should be mentioned is that the BD method can approximately estimate the severity of induced damage. As stated in [13], the structural condition #14 is the highest level of nonlinear damage. Accordingly, the quantity in Fig. 3(c) at the sensor number 4 is the most damage severity in all damage cases. To put it another way, the Bhattacharyya distance method estimates the damage extent in addition to the detection and localization of structural damage.

5. Conclusions

The objective of this article is to identify and locate the structural damage based on a new similarity measure method with aid of the time series analysis of vibration data. In this regard, Bhattacharyya distance method as a similarity measure technique is applied to identify the damage location and probably estimate the extent of damage. Autoregressive model is utilized to fit a mathematical model to the vibration measurement in order to extract the residual errors as damage-sensitive feature. On the basis of this procedure, an improved feature extraction technique is proposed in such a way that all sensors of the healthy structure are chosen as reference points and then their AR models are adopted for the corresponding sensors in the damaged structure. This procedure is more reliable and accurate than choosing a sensor in the healthy state and generalizing its time series model to other sensors in the damaged state.

With respect to the residual errors of the undamaged condition (state #1) along with the residuals of the damaged conditions (states #7, #10, #14, and #17), the Bhattacharyya distance measures the similarity between residuals of the healthy and damaged states. On the basis of BD method, zero distance (similarity) value ideally indicates that no damage is available. By contrast, non-zero similarity quantity implied that the damage has been occurred in the structure. Comparing the distance values at all sensors with together provide more efficient results about the location of damage. Hence, it can be seen that BD values precisely identify the damage location in the states #10, #14, and #17 as striking columns in their bar figures. These columns belong to the Bhattacharyya distance value at the channel #4, which the nonlinear damage (the bumper gap) has been simulated there. Furthermore, the BD method is roughly able to estimate the extent of damage so that the most distance value among states is indicative of the most damage severity. However, in the state #7, BD cannot provide the meaningful information about damage due to the influence of operational and environmental variability in the structure, because the features extracted from this state are not sensitive to the damage.

Acknowledgement

The authors would like to acknowledge the Engineering Institute at Los Alamos National Laboratory for making available to the public domain the experimental data used in this study, and downloadable from http://institute.lanl.gov/ei/software-and-data/data.

REFERENCES

[1] M. L. Fugate, H. Sohn, and C. R. Farrar, "Vibration-based damage detection using statistical process control," *Mechanical Systems and Signal Processing,* vol. 15, pp. 707-721, 2001.

[2] C. R. Farrar and K. Worden, "An introduction to structural health monitoring," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* vol. 365, pp. 303-315, 2007.

[3] C. R. Farrar and K. Worden, *Structural Health Monitoring : A Machine Learning Perspective*: John Wiley & Sons Ltd, 2013.

[4] H. Sohn, J. A. Czarnecki, and C. R. Farrar, "Structural health monitoring using statistical process control," *Journal of Structural Engineering,* vol. 126, pp. 1356-1363, 2000.

[5] K. K. Nair, A. S. Kiremidjian, and K. H. Law, "Time series-based damage detection and localization algorithm with application to the ASCE benchmark structure," *Journal of Sound and Vibration,* vol. 291, pp. 349-368, 2006.

[6] P. Omenzetter and J. M. W. Brownjohn, "Application of time series analysis for bridge monitoring," *Smart Materials and Structures,* vol. 15, p. 129, 2006.

[7] K. Worden, G. Manson, and N. Fieller, "Damage detection using outlier analysis," *Journal of Sound and Vibration,* vol. 229, pp. 647-667, 2000.

[8] M. Gul and F. Necati Catbas, "Statistical pattern recognition for Structural Health Monitoring using time series modeling: Theory and experimental verifications," *Mechanical Systems and Signal Processing,* vol. 23, pp. 2192-2204, 2009.

[9] R. Yao and S. N. Pakzad, "Autoregressive statistical pattern recognition algorithms for damage detection in civil structures," *Mechanical Systems and Signal Processing,* vol. 31, pp. 355-368, 2012.

[10] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control (Fourth Edition)*. New Jersey: John Wiley & Sons, Inc, 2008.

[11] G. Kitagawa, *Introduction to Time Series Modeling*. Boca Raton: CRC Press, Taylor & Francis Group, 2010.

[12] D. S. G. Pollock, *A Handbook of Time-Series Analysis, Signal Processing and Dynamics*. London, UK and San Diego, California, USA: Academic Press, 1999.

[13] E. Figueiredo, G. Park, J. Figueiras, C. Farrar, and K. Worden, "Structural health monitoring algorithm comparisons using standard data sets," Los Alamos National Laboratory: LA-143932009.

[14] J. M. Caicedo, S. J. Dyke, and E. A. Johnson, "Natural excitation technique and eigensystem realization algorithm for phase I of the IASC-ASCE benchmark problem: Simulated data," *Journal of Engineering Mechanics*, 2003.

[15] A. M. Zoubir and J. F. Böhme, "Bootstrap multiple tests applied to sensor location," *Signal Processing, IEEE Transactions on,* vol. 43, pp. 1386-1396, 1995.

[16] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of the Calcutta Mathematical Society,* vol. 35, pp. 99–109, 1947.

[17] LANL. (2009). *Los Alamos National Laboratory: Experimental Data for Download*. Available: http://institute.lanl.gov/ei/software-and-data/data

[18] E. Figueiredo, J. Figueiras, G. Park, C. R. Farrar, and K. Worden, "Influence of the Autoregressive Model Order on Damage Detection," *Computer-Aided Civil and Infrastructure Engineering,* vol. 26, pp. 225-238, 2011.

[19] H. Zheng and A. Mita, "Localized damage detection of structures subject to multiple ambient excitations using two distance measures for autoregressive models," *Structural Health Monitoring,* vol. 8, pp. 207-222, 2009.