The 6th International Workshop on Advanced Smart Materials and Smart Structures Technology
ANCRiSST2011
July 25-26, 2011, Dalian, China

Statistical Modeling Methods for Structural Damage Identification

Ruigen Yao[1], Shamim N. Pakzad[1]

[1]*Department of Civil and Environmental Engineering, Lehigh University, Bethlehem, PA, 18015.*
ruy209@lehigh.edu, pakzad@lehigh.edu

ABSTRACT

Early damage detection is very important for in-time repair and retrofit of civil structures. The past years have seen a growing interest in automatic/programmable damage detection techniques using measurements from structural monitoring because of its potential economic benefit for maintenance practices. Recently, statistical pattern recognition (SPC) has been brought forward as a promising way to realize this goal. The essence of SPC is to use well-defined concepts in statistics for data classification, which is an intuitive way to account for the inherent uncertainty in the data acquisition process. In this paper, two types of statistical damage detection algorithm are described and applied to the acceleration responses collected from a scaled two-bay steel frame subjected to random excitation. Damage is introduced into this specimen by replacing two members near the right joint with more flexible ones. The influence coefficient method for damage detection is based on correlation analysis between responses measured at two different locations, while the autoregressive (AR) modeling method evaluates the change in coefficients and residual characteristics of the AR model estimated from response from only one sensor node. Influence coefficients here are basically linear regression coefficients of one node's response on another, and AR model, as indicated by its name, regress the current response on those from several time steps back. Both techniques have successfully identified the damage existence in the steel frame. It is shown that the amount of change occurred in the influence coefficient values from different node combinations can be used to localize damage. For the AR modeling method, the autocorrelation function of residuals at each sensing unit is proved to be a more effective damage location indicator than Mahalanobis distance of model coefficients at each sensing unit in this case.

INTRODUCTION

In recent years, data-driven structural health monitoring (SHM) has been actively investigated by the civil engineering community as an important tool for future infrastructure maintenance.[2] The availability of affordable sensing and data transmission/storage systems has made continuous structural monitoring physically feasible [11]. In addition, the occasional failures of important transportation infrastructures (e.g. Minnesota I-35W) have attracted such ample media coverage and public attention that more effective structural assessment and repair/retrofit methods are much sought after to ensure public safety. The development of data-driven SHM will lead to automatic structural diagnosis, which will hopefully be less biased and less expensive than traditional human inspection methods.

Generally, data-driven structural health monitoring consists of two steps: periodically sampling data from the structure being monitored and compressing the large quantity of data for possible structural damage information. Most existing structural monitoring systems measures only vibration responses, as they are the easiest to obtain. A paradigm for vibration-based damage detection has been proposed by Farrar et al. [3], where this procedure in divided into four stages: 1) Operational evaluation; 2) Data acquisition and cleansing; 3) feature selection and data compression, and 4) statistical model development/decision making. The first two stages take care of data collection details, while the last two are about decision making based on the data collected. For a long time, step 4) of this paradigm has been overlooked; a lot of literatures can be found on feature extraction, but in comparison few researches address the problem of statistical damage threshold construction. Realizing this problem, a series of studies has been conducted by Farrar and his colleagues on hypothesis testing [6] for damage threshold determination, which is a part of their statistical pattern recognition (SPC)[5] scheme for structural damage identification.

SPC is a family of methods that performs data/pattern classification using statistical analysis concepts. This approach concerns both feature extraction and threshold evaluation, is well suited to process large amount of data, and automatically accommodates the inherent uncertainty of the signal acquisition process inside its framework. To obtain an objective evaluation of its potential for damage identification, however, results from its application to various types of structures are needed. Also, to understand the pros and cons of different statistical approaches they need to be compared and contrasted. In this paper, two statistical damage detection methods will be used on acceleration measurements collected from a two-bay steel frame; the influence coefficients method is based on regression between the responses from two nodes, and the autoregressive(AR) [1] modeling method evaluates the change in estimated model coefficients and residuals. Data-driven thresholds from change point analysis and cross validation [4] will be used for the features extracted from these two methods.

The organization of this paper is as follows; Section 2 presents the influence coefficient method for structural damage identification, which is basically regression of one node’s response on another; Section 3 is dedicated to description of univariate autoregressive modeling and damage indices from the model parameters and residuals. Section 4 focus on the threshold evaluation schemes for the damage features using resampling techniques. In section 5, the damage detection algorithms are applied to acceleration measurements collected from a two-bay steel frame specimen, and the results are presented and compared.

STATISTICAL METHOD 1: LINEAR REGRESSION BETWEEN RESPONSES FROM TWO SENSOR NODES

If a linear structure is under static/quasi-static loading, then the structural response u at any two locations should be linearly correlated:

$$u_j(t_k) = \alpha_{j,i} \cdot u_i(t_k) + \epsilon_{ij}(t_k) \tag{1}$$

Here i and j indicate the response location, and k is a time label. The correlation coefficient is determined by both the force distribution and structural stiffness properties.

In practice, most structures are subjected to dynamic loads. However, if only a small part of the structure (with large stiffness and insignificant mass) is monitored, then it can be assumed that the local behavior could be captured by a static model.

This pair-wise regression method has been applied to detect damage in a beam column specimen and a simulated model of the steel frame here [7, 8]. The feature extraction methodology can be summarized into the following steps; 1) for each pair of u_i and u_j, the correlation coefficients $\alpha_{i,j}$ and $\alpha_{j,i}$ and the corresponding residuals are evaluated using the least squares method (Fig.1); 2) to check the stability of estimation, two indices are calculated: accuracy factor $EA = \alpha_{i,j} \cdot \alpha_{j,i}$ and normalized estimation error $\gamma_{ij} = \sigma_{\alpha_{ij}}/\alpha_{ij}$; 3) correlation coefficients from those node pairs with EA close to 1 and γ_{ij} close to zero are selected as damage indicators.

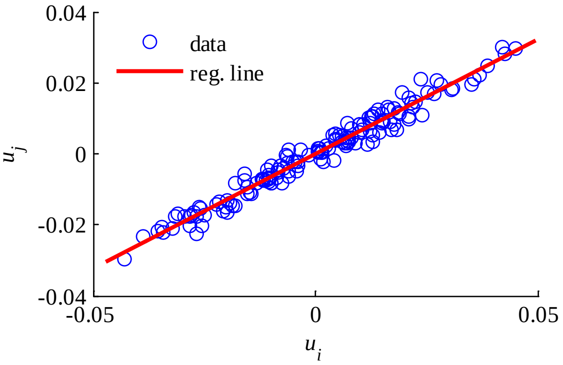

Figure 1 Regression for influence coefficients

This approach is straightforward in principle and efficient for computation. Also, the change in the values of regression coefficients can reflect to some extent the location of damage. However, when the dynamic response content becomes prominent (i.e. regression is performed for two far-apart nodes), the algorithm performance will degrade.

STATISTICAL METHOD 2: AUTOREGRESSIVE MODELING OF RESPONSE AT ONE LOCATION

Univariate autoregressive model[1] is among the most widely applied time series analysis tools. It basically expresses the value of the signal at a certain time point as a linear combination of its previous values and a random error term $e(t)$:

$$x(t) = \sum_{i=1}^{p} \phi_i x(t-i) + e(t). \tag{2}$$

Here $\phi_i s$ are the model coefficients, and p is the model order. AR coefficients can be estimated from collected signals using one of the standard algorithms such like Yule-Walker and Burg, and model residuals/errors $e(t)$ can then be obtained as the difference between model prediction and the real signal. Autoregressive damage detection algorithms can be based on either model coefficients or residual characteristics. In the remainder of this section, one example from either category will be presented.

The Mahalanobis distance of AR coefficients

It has been proved that if the signal is really an AR process, then any regular coefficients estimator$\{\phi_{xj}\}$ from the signal is asymptotically unbiased and normally distributed with covariance matrix $\sigma_e^2 \Gamma_p^{-1}$[1]. Therefore, a metric that represents the coeffcients' deviation in the probability space of normal distribution seems a good choice of damage feature.

Mahalanobis distance[10] is such a metric defined from the definition of multivariate normal distribution. The estimator of the Mahalanobis distance between a potential outlier vector x_ξ and baseline sample set can be obtained as

$$D_\zeta = (x_\xi - \bar{x})\hat{\Sigma}^{-1}(x_\xi - \bar{x}). \quad (3)$$

where $\bar{x}$ is the average of the baseline sample feature vectors, and $\hat{\Sigma}$ the estimated covariance matrix. When applying this method, the baseline signals are first segmented (often with large overlap) and for each segment an AR coefficient vector are estimated. Signals from current structural state are processed likewise and for each coefficient vector obtained its Mahalanobis distance to the baseline coefficients set will be computed. These Mahalanobis distance features are then compared with the Mahalanobis distances within baseline set. When the structural system is damaged, it is expected that the Mahalanobis distance feature for AR coefficients will increase significantly (Fig. 2).

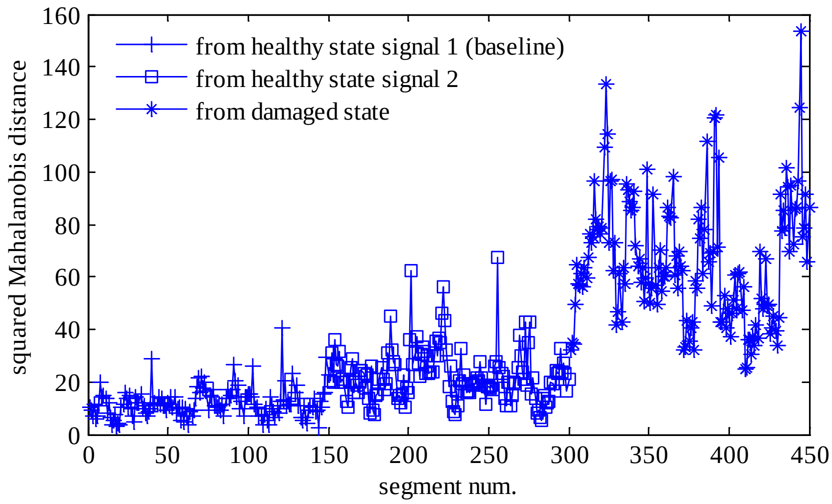

Figure 2 Mahalanobis distance for structural damage identification

Q-statistic trace of AR model residuals

If the structural condition has undergone changes, then the baseline model will no longer provide a good fit to the new signals collected. As the result, the residual series, instead of resembling a white noise, shall carry some identifiable patterns over time.

Ljung and Box[1] proposed a statistic to measure the difference between the residual series and white noise:

$$Q = n(n+2)\sum_{j=1}^{h}\frac{\rho_j^2}{n-j} \quad (4)$$

where n is the sample size, h is the number of lags, and ρ_j is the autocorrelation at the j^{th} lag. Since this Q −statistic follows a χ^2 distribution under the normality assumption of the input, a statistical test can be devised at significance level α, by setting the rejection threshold at $\chi^2_{1-\alpha,h}$, which is the α −quantile of the χ^2 distribution with h degrees-of-freedom.

The Q statistic trace is a plot of the Q statistic against the baseline AR model order (Fig. 2). For undamaged state signal, a downward trend is expected in the plot. Thus if the trace is oscillating, stopping decreasing at short or going upward as the model order increase, the system is recognized as damaged.

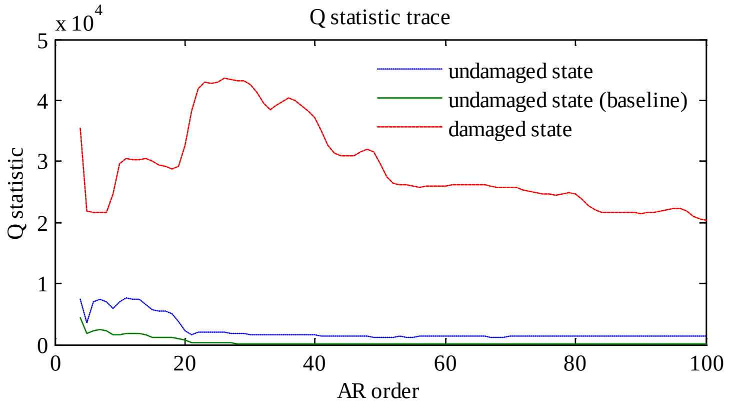

Figure 3 Ljung-Box statistic for structural damage identification

THRESHOLD CONSTRUCTION METHODS

After feature selection, appropriate damage threshold construction schemes are needed to reach a decision on the current state of the structure being monitored. If the feature distribution can be assumed, then statistical hypothesis testing could be used for damage identification. However, in most cases the feature distribution is unknown and cannot be approximated by an analytical function. As such, the damage threshold will be determined based on the data solely using one of the resampling techniques.

Change point analysis using cumulative sum

Change point analysis is used to find the point in a data sequence where the data characteristics change and the confidence level associated with this change point. It is an effective way of identifying thresholds and detecting subtle variations. A variety of approaches is available for performing change point analysis, such as cumulative sum, deviance reduction, and least squares. Here, the *cumulative sum* method [12] will be adopted for damage threshold construction for the influence coefficients features.

The cumulative sum $\{S_i\}$ for a data sequence $\{x_i\}$ is calculated as below:

1) Subtract every value in this sequence by its mean $\bar{x}$.
2) Compute the cumulative sum at step i by adding up the values occurred before and at i. (i.e. $S_i = \sum_{j<i}(x_j - \bar{x})$)

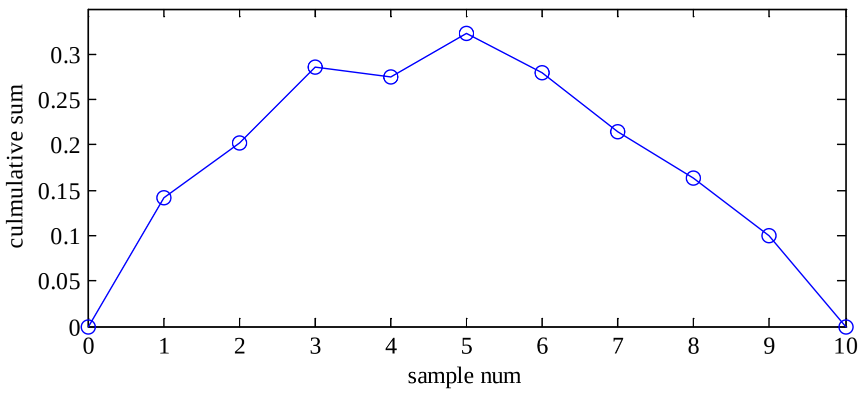

Figure 4 the cumulative sum plot from the implementation in section 5.

Experienced individuals can identify the change point directly as the place where the trend of the cumulative sum plot changes.(Fig.4) But as our aim is automatic damage threshold construction, a bootstrap analysis is performed; the original signal sequence is randomly permuted for N times and each time the maximum absolute value of the cumulative sum of the new sequence is recorded. The threshold for the maximum absolute accumulative sum of a data sequence of constant properties is then set at the point above which $N(1-\alpha)$ values occurred. The significance level α is set to 5% in all the applications in this paper.

Cross validation

As described in the previous section, the Mahalanobis distance feature is computed between a certain coefficients vector and a set of reference vectors. Its theoretical distribution is hard to derive and the feature itself often exhibit large fluctuations even within the healthy state. Thus, a cross-validation approach is devised for its damage threshold construction. It is essentially an empirical estimation of the feature distribution through recomputing the statistic for many a time by leaving out a certain portion of the observations, and can be viewed as a variation of standard bootstrapping technique [4].

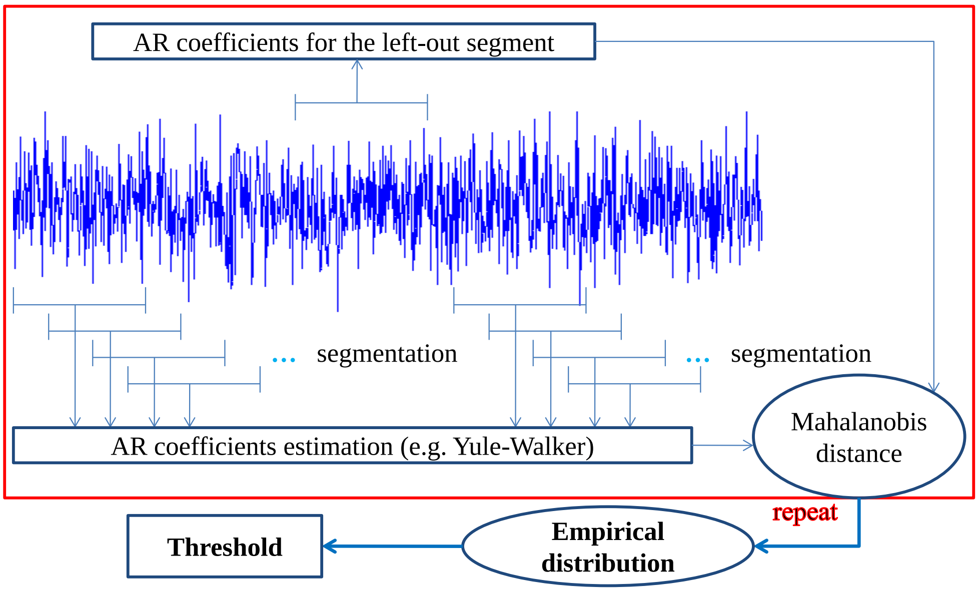

Figure 5 the 'cross-one-out' threshold evaluation method for Mahalanobis distance of model coefficients

Fig. 5 is a flow chart of this approach. It can also be described in words as follows: First a segment is cut from the baseline signal at a random time point and reserved for testing, and sample segments of the same size are cut with a preset overlap from the remaining signal. The Mahalanobis distance between the AR model coefficients of left-out segment and those of the sample set is then computed and stored. This process is repeated for a large number of times and the value beyond which 5% of the tests occur is used as threshold in subsequent analysis.

This method has already been applied in several previous researches [13, 14], and is shown to demonstrate superior performance to other threshold evaluation techniques such as hypothesis testing based on Gaussian assumption of model coefficients and Mont Carlo simulation [11, 13, 14].

EXPERIMENTAL RESULTS

The aforementioned influence coefficients and AR modeling method, together with their respective threshold evaluation approaches, are applied to detect damage in a scaled two-bay steel frame constructed from steel tubes. (Fig 6) The structure, instrumented with 21 accelerometers, is excited from the left beam-column joint by an electro-dynamic shaker (Fig 7). To simulate a structural damage, the potion between sensor R1 and R3 and that between R4 and R6 are switched out by 20% less stiff tubes. For each structural scenario, five random vibration tests are performed and the acceleration responses collected are used as input to the algorithms.

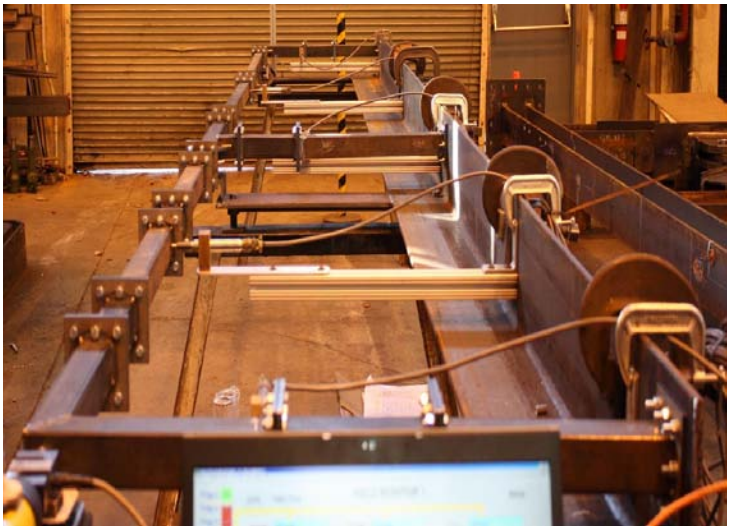

Figure 6 the two-bay steel frame

The first algorithm produces for each dataset $21 \times 20 = 420$ influence coefficients, each from a particular node combination. To facilitate the subsequent decision making process, only those coefficients with average evaluation accuracy factor EA>0.9 and normalized estimation error $\gamma < 0.003$ are examined. Out of the 24 coefficients selected (Table xx), the cumulative-sum-based change point analysis successfully identified damage for 21 of them over the 10 tests. It is observed, however, that the 3 coefficients that failed to report damage are all from regression between distant nodes, which are by nature not quite reliable.

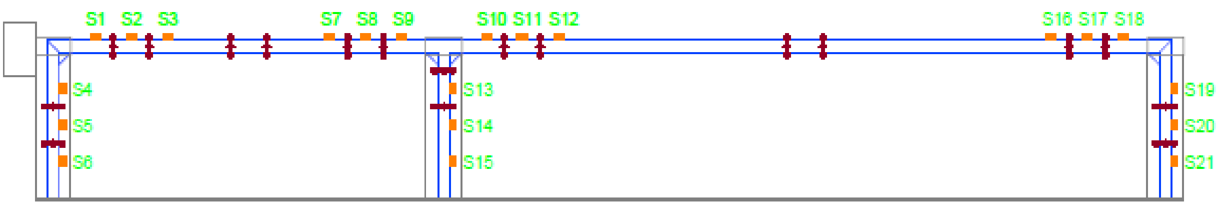

Figure 7 a schematic plot of the girder mounted with accelerometers

Fig 7 displays the estimation results from 4 influence coefficients, all indicating damage except for the one showed in the lower right corner. One interesting fact noticed is that though the nodes 5 and 6 are nowhere near damage, their corresponding influence coefficient value still increased abruptly as a result of structural change. This is probably because these correlation coefficients reflect the ratio of structural vibration ‘shape function’ values evaluated at two different locations, and those combinations with most affected coefficients are determined by not only the damage location but also the structural layout. Yet still, generally there will be a higher chance of observing changes near the damage for a properly restrained regular structure . Also if an analytical model of the structure is available, the influence coefficients can be used as input to a model updating scheme to find out the damage location.

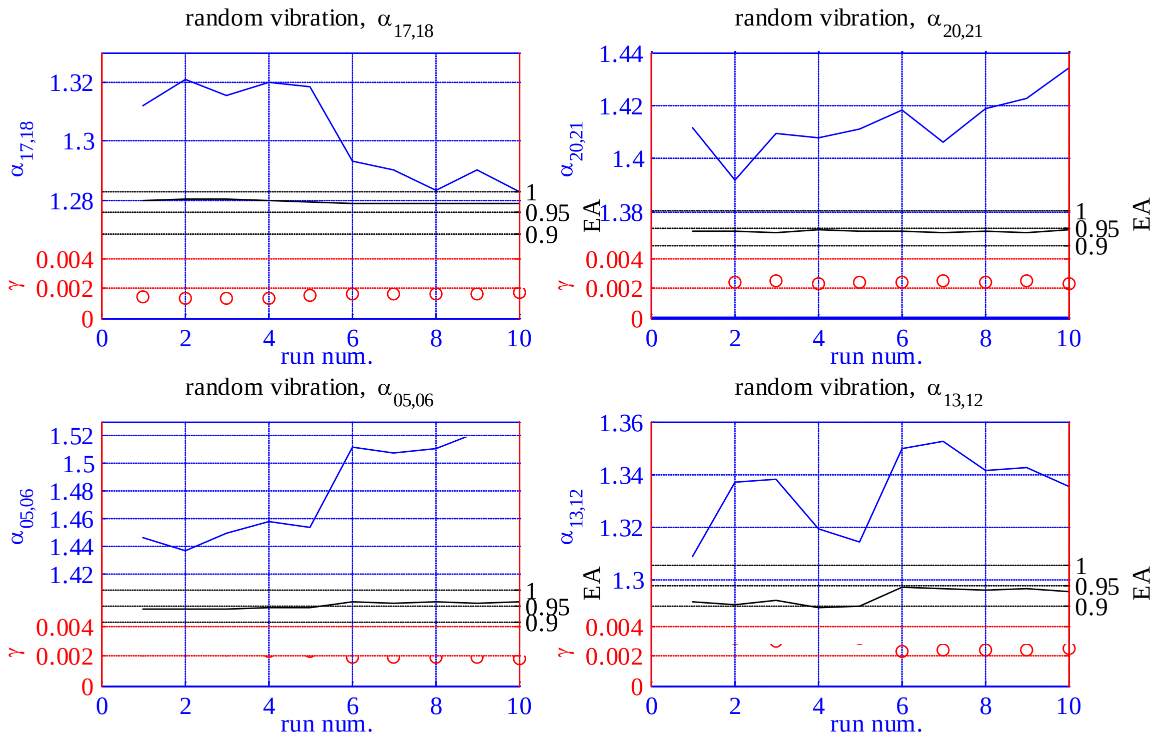

Figure 8 influence coefficients from different node pairs

The AR modeling based algorithms are also proved effective for damage detection in this case. While the influence coefficients capture the static behavior of the system, time series modeling deals with the structural dynamic effects. For has the system been all static, the response will only be a scaled white noise series with no distinguishable patterns for modeling. Figure 9 shows the results from AR coefficients based Mahalanobis distance method using data from sensor 2 and 17. To save space here, only two datasets per structural state are used. In both graphs dataset 3 is used as the baseline and 4 the false positive testing. Damage is clearly indicated in both plots, even though one node is much farther from the damage location. It may thus be concluded that here the local damage does affect the global dynamic properties in a certain way.

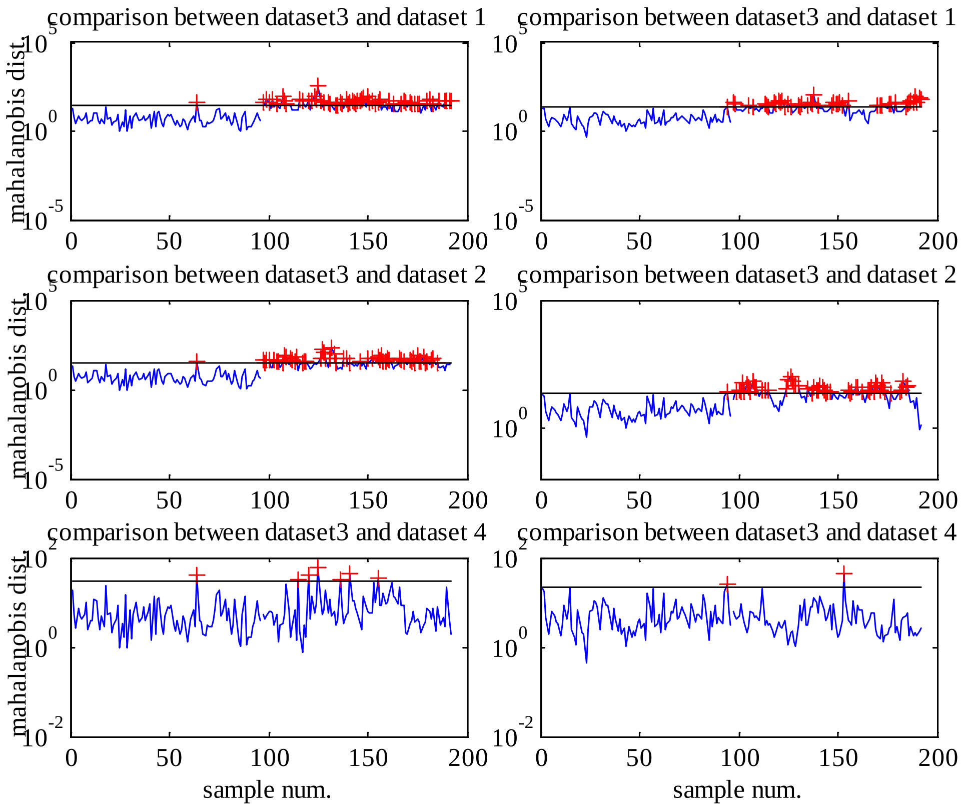

Figure 9 Mahalanobis distance from signals collected at node 2 (left) and node 17(right)

Fig. 10 is the Ljung-Box/Q statistic traces for node 2 and 17. The two damaged state traces float high above those undamaged ones as expected. The threshold from hypothesis testing is conservative for data from both sensors, suggesting that the theoretical assumptions do not quite represent reality here. Also, damaged state Q statistics from measurements at node 17 is much larger than that from node 2. After examining Q traces from all 21 nodes, it is found that the traces from nodes on the right beam show greatest change, indicating the damage location.

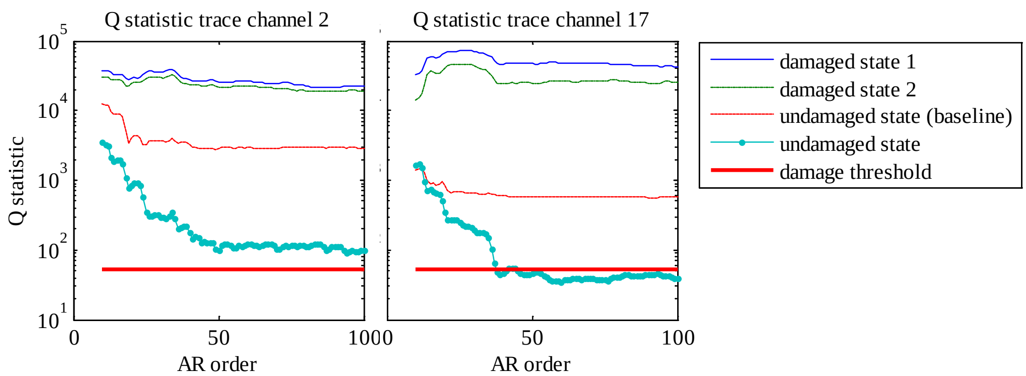

Figure 10 Ljung Box trace for damage detection

CONCLUSION

This paper focuses on validating the performance of statistical pattern recognition algorithms through their application to detect damage in an artificially excited two-bay steel frame using acceleration measurements collected. These algorithms are advantageous in that the decision on the structural state is made using well-established statistical concepts instead of human expertise, thereby eliminating possible individual biases. Besides, since they are mostly data driven, very few assumptions are needed regarding the physical structure.

The pair-wise regression and AR modeling method adopted here are complementary in the sense that while the former works on getting a static relation between signals from two different locations, the latter concerns itself with analyzing the dynamic pattern of the signal from a single sensor node. The damage features presented are influence coefficients from the 1st method, and Mahalanobis distance of model coefficients and Ljung-Box statistic from the 2nd method. Their corresponding damage threshold evaluation techniques are cumulative sum analysis, cross-one-out validation and hypothesis testing. It can be seen here that different types of features may need different ways to establish the damage threshold.

Judging from the results obtained in the previous section, both algorithms have successfully detected the existence of artificially introduced damage. It shall be noted, however, that the change in the influence coefficient values are not the ideal damage location indicator hinted in some earlier works[7,8]; they only suggest a possibility which need to be verified through other means of observation. The Mahalanobis distance is no fit for damage location detection in this case: at least as much change is observed from the features at node distant to damage as that from a nearby node. The Ljung-Box static from AR residuals seems to do better in this aspect, through no clear reasons can be provided for now. Also noticed is that the data-driven threshold evaluation methods based on resampling yield a superior performance to hypothesis testing, as they suffer less from statistical modeling errors.

To summarize, the statistical algorithms have been altogether effective for identifying damage existence in the lab specimen and two of them have been able to suggest the possible damage locations. Each feature extraction scheme and the associated threshold construction technique operate as a whole to bring forward the final result. To further investigate the capabilities of the algorithms, however, more experiments and numerical analysis on various types of structures are needed.

ACKNOWLEDGMENTS

The research described in this paper is partially supported by the National Science Foundation through Grant No. CMMI-0926898 by Sensors and Sensing Systems program, and by a grant from the Commonwealth of Pennsylvania, Department of Community and Economic Development, through the Pennsylvania Infrastructure Technology Alliance (PITA).

REFERENCES:

[1] P. J. Brockwell and R.A. Davis, Time Series: Theory and Methods, 2nd Edition, Springer, NY, 2009.

[2] C. R. Farrar, and K. Worden, an Introduction to Structural Health Monitoring, *Phil. Trans. R. Soc.,* Vol. 365, No. 1851, pp. 303-315, 2007.

[3] C. R. Farrar, T. A. Duffey, S. W. Doebling, D. A. Nix, A Statistical Pattern Recognition Paradigm for Vibration-based Structural Health Monitoring, *2nd International Workshop on Structural Health Monitoring* ,Standford , CA, 1999.

[4] P. I. Good, *Resampling Methods: A Practical Guide to Data Analysis*, Birkhauser, 1999.

[5] A.K. Jain, R.P.W. Duin, and J. Mao, Statistical Pattern Recognition: A Review, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 1, pp4-37, 2000.

[6] K. Koch, Parameter Estimation and Hypothesis Testing in Linear Models, Springer, 2nd Edition, 1999.

[7] E. L. Labuz, M. Chang, and S. Pakzad, Local Damage Detection in Beam-Column Connections Using a Dense Sensor Network, *Proc. of 19th Annual Structures Congress*, Orlando, FL, May 2010.

[8] E. L. Labuz, S. N. Pakzad, and L. Cheng, Damage Detection and Localization in Structures: A Statistics Based Algorithm Using a Densely Clustered Sensor Network, *Proc. of 20th Annual Structures Congress*, Las Vegas, NV, April 2011.

[9] Ljung, G. M. and Box, G.E.P., 1978, On A Measure of Lack of Fit in Time Series Models, *Biometrika*, Vol. 65, No. 2, pp. 297-303.

[10] P.C. Mahalanobis, on the Generalized Distance in Statistics. *Proceedings National Institute of Science*, India, Vol. 2, No. 1, pp. 49-55,1936.

[11] H. Sohn., C.R. Farrar, F. M. Hemez, D. D. Shunk, S. W. Stinemates, B. R. Nadler and J. J. Czarnecki, A Review of Structural Health Monitoring Literature form 1996-2001, *Los Alamos National Laboratory report LA-13976-MS*, 2004.

[12] WA Taylor., Change-point Analysis: A Powerful New Tool for Detecting Changes. *Deerfield, IL: Baxter Healthcare Corporation*, 2000, Online. May be accessed at http://www.variation.com/cpa/tech/changepoint.html

[13] R. Yao and S. N. Pakzad., Data-Driven Methods for Threshold Determination in Time-Series Based Damage Detection, *Proc. of 20th Annual Structures Congress*, Las Vegas, NV, April 2011.

[14] R. Yao and S. N. Pakzad., Autoregressive Statistical Pattern Recognition Algorithms for Damage Detection in Civil Structures, Manuscript submitted for publication, *Mechanical Systems and Signal Processing*, January 2011.