Computer-Aided Civil and Infrastructure Engineering **26** (2011) 225–238

Influence of the Autoregressive Model Order on Damage Detection

Eloi Figueiredo & Joaquim Figueiras

Department of Civil Engineering, Faculty of Engineering of the University of Porto, Porto, Portugal

Gyuhae Park & Charles R. Farrar*

The Engineering Institute, Los Alamos National Laboratory, Los Alamos, NM, USA

&

Keith Worden

Department of Mechanical Engineering, University of Sheffield, Sheffield, United Kingdom

Abstract: *An important step for using time-series autoregressive (AR) models for structural health monitoring is the estimation of the appropriate model order. To obtain an optimal AR model order for such processes, this article presents and discusses four techniques based on Akaike information criterion, partial autocorrelation function, root mean squared error, and singular value decomposition. A unique contribution of this work is to provide a comparative study with three different AR models that is carried out to understand the influence of the model order on the damage detection process in the presence of simulated operational and environmental variability. A three-story base-excited frame structure was used as a test bed in a laboratory setting, and data sets were measured for several structural state conditions. Damage was introduced by a bumper mechanism that induces a repetitive impact-type nonlinearity. The operational and environmental effects were simulated by adding mass and by changing the stiffness properties of the columns. It was found that these four techniques do not converge to a unique solution, rather all require somewhat qualitative interpretation to define the optimal model order. The comparative study carried out on these data sets shows that the AR model order range defined by the four techniques provides robust damage detection in the presence of simulated operational and environmental variability.*

*To whom correspondence should be addressed. E-mail: *farrar@lanl.gov*

DOI: 10.1111/j.1467-8667.2010.00685.x

1 INTRODUCTION

The process of implementing a damage detection strategy for aerospace, civil, and mechanical engineering infrastructure is referred to as structural health monitoring (SHM). Here damage is defined as changes to the material and/or geometric properties of these systems, including changes to the boundary conditions and system connectivity, which adversely affect the system's current or future performance. The authors pose the SHM process in the context of a statistical pattern recognition paradigm. In this paradigm, the process can be broken down into four parts: (1) operational evaluation, (2) data acquisition, (3) feature extraction, and (4) statistical model development for feature classification. A more detailed discussion of this paradigm can be found in Farrar et al. (2001). The

current article focuses only on the third part of the process-feature extraction. Note that this paradigm is solely based on signal analysis of the measured vibration data.

Feature extraction is concerned with selection of damage-sensitive features, i.e., some quantities extracted from the measured system response data that can be correlated with the presence of damage in the structures. Ideally, a damage-sensitive feature will change in a consistent manner with increasing damage level. Identifying features that can accurately distinguish a damaged structure from an undamaged one is the focus of most the SHM technical literature (Doebling et al., 1996; Sohn et al., 2001; Sohn et al., 2004). Fundamentally, feature extraction is based on fitting some model, either physics-based (Moaveni et al., 2008) or data-driven (Sohn et al., 2008), to the measured system response data. The parameters of these models or the prediction errors associated with these models then become the damage-sensitive features. An alternate manner is to identify features that directly compare the sensor waveforms or spectra of these waveforms measured before and after damage. Many of the features identified for impedance-based and wave propagation-based SHM studies fall into this category (Park et al., 2003; Ihn and Chang, 2004).

In the SHM field, the ideal approach for feature selection (or more generally, extraction) is to choose features that are sensitive to damage, but are insensitive to operational and environmental variations. However, in real-world structures such an approach is rarely possible, and "intelligent" feature extraction procedures are usually required (Worden et al., 2007). This article addresses that issue by using time-series models to extract features to detect damage, even under the presence of simulated operational and environmental condition changes. Time-series analysis based on the use of autoregressive (AR) models have been extensively used in the SHM process as a feature extraction technique (Sohn et al., 2001; Farrar et al., 2007). This technique is typically applied to experimentally measured time-series data where future data values are predicted from past values. The residual errors have had considerable applications as damage-sensitive features. Furthermore, in addition to the widely used residual errors, the estimated AR parameters are also directly used as damage-sensitive features. This approach intends to extract features capable of being sensitive to the effects caused by damage and (quasi-) insensitive to the effects due to operational and environmental variations.

A key to employing the AR models efficiently as a feature extraction technique is to determine the appropriate model order. Traditionally, the model order selection is done in an *ad hoc* manner. To choose the optimal model order, this article presents four techniques based on the Akaike information criterion (AIC), partial autocorrelation function (PAF), root mean squared (RMS) error, and singular value decomposition (SVD). Furthermore, to emphasize the importance of the optimal model order in the damage detection process, a comparative study is carried out using three AR models of different orders. One is a lower order (less than ideal as determined by the four techniques) model and the other two correspond to the upper and lower bounds of the order range indicated by the four techniques applied to the baseline condition data.

The applicability of the proposed approach is investigated using time-series data sets from a three-story base-excited frame structure tested in a laboratory environment. The effect of damage is simulated by nonlinear phenomena introduced by a bumper mechanism that simulates a repetitive impact-type nonlinearity. The nonlinearities are intended to produce a small perturbation to an essentially stationary process, causing a nonlinear phenomenon called intermittency (Kantz and Schreiber, 2004), i.e., the system alternates between two conditions in an irregular way. This mechanism is intended to simulate a crack that opens and closes under dynamic loads (a so-called "breathing" crack), or loose connections that rattle. Several real-world examples have been published reporting bend and shear cracks that open and close under dynamic loads. For instance, in the prestressed bi-cellular box–girder bridge, such as the N. S. da Guia Bridge in Portugal, cracks were measured that open and close 0.12 mm under the loads produced by two vehicles passing side-by-side at 20 km/h (Pimentel et al., 2008).

Real-world structures have operational and environmental variability, which imposes difficulties in detecting and identifying structural damage (Xia et al., 2006; Peeters and De Roeck, 2001; Alampalli, 2000). For instance, and in the case of bridges, the influence of traffic loading on modal parameters was investigated by Kim et al. (2003) and reported that the measured natural frequencies of a 46 m long simply supported plate–girder bridge decreased by 5.4% as a result of heavy traffic. In a box–girder concrete bridge located in Irvine, CA, Soyoz and Feng (2009) found variations in the first natural frequency in the order of $\pm 10\%$. The authors attribute those variations to changes in the mass of the bridge caused by traffic and environmental effects. Further investigation concluded that the vehicles changed the total mass of the bridge in the order of 10% and the modal frequency of the bridge in the order of 5%. Several investigations have suggested that the temperature plays the major role on the modal parameters variability. In the case of highway bridges, those variations can reach 5–10% (Ko and Ni, 2005). In a study performed

in the Alamosa Canyon Bridge, Farrar et al. (2000) noted that natural frequencies were found to vary approximately 5% during a 24-h time period. Thus, in this article, to incorporate those variations into the frame structure, the operational and environmental effects are simulated through different mass and stiffness conditions, which will only cause structural change in a linear manner. Those changes were designed to introduce variability in the fundamental natural frequency up to approximately 7% from the baseline condition (Figueiredo et al., 2009).

The layout of this article is as follows. Section 2 provides a summary description of the test structure that relates the sources of simulated damage to real-world damage in structures, a brief description of the data acquisition system, and experimental procedure as well as the damage scenarios. Section 3 describes the AR model as a damage-sensitive feature extractor and describes four techniques to find the optimal order of the AR model. Section 4 shows the analysis carried out on the data sets from the test structure. The stationarity property of the data sets is also introduced in this section. Finally, Section 5 presents a summary and discussion of the analyses carried out in the article.

2 EXPERIMENTAL PROCEDURE: TEST STRUCTURE, DATA ACQUISITION SYSTEM, AND DATA SETS

The three-story frame structure shown in Figure 1 (indicating the basic dimensions) is used as a damage detection test bed structure. The structure consists of aluminum columns and plates assembled using bolted joints. The structure can slide on rails that allow movement in the x-direction only. At each floor, four columns ($17.7 \times 2.5 \times 0.6$ cm) are connected to the top and bottom plates ($30.5 \times 30.5 \times 2.5$ cm) forming a (essentially) four degree-of-freedom system. Additionally, a center column ($15.0 \times 2.5 \times 2.5$ cm) is suspended from the top floor. This column can be used to simulate damage by inducing nonlinear behavior when it contacts a bumper mounted on the next floor. The position of the bumper can be adjusted to vary the extent of impacting that occurs at a particular excitation level. It should be emphasized that this structure is not a scale model of any prototype system, but rather it was designed as a standard test bed for SHM validation studies.

An electrodynamic shaker provides a lateral excitation to the base floor along the centerline of the structure. The structure and shaker are mounted together on an aluminum baseplate ($76.2 \times 30.5 \times 2.5$ cm) and the entire system rests on rigid foam. The foam is intended to minimize extraneous sources of unmeasured excitation from being introduced through the base of the structure. A load cell (Channel 1) with a nominal sensitivity of 2.2 mV/N was attached at the end of a stinger to measure the input force from the shaker to the structure. Four accelerometers (Channels 2–5) with nominal sensitivities of 1,000 mV/g were attached at the centerline of each floor on the opposite side from the excitation source to measure the structure response. A Dactron Spectrabook data acquisition system was used to collect and process the data. The output channel of this system, which provides the excitation signal to the shaker, is connected to a Techron 5530 Power Supply Amplifier that drives the shaker. The analog sensor signals were discretized into 8,192 data points sampled at 3.125 ms intervals corresponding to a sampling frequency of 320 Hz. These sampling parameters yield time histories 25.6 sec in duration. A band-limited random excitation in the range of 20–150 Hz was used to excite the structure. This excitation signal was chosen to avoid the rigid body modes of the structure that are present below 20 Hz. The excitation level was set to 2.6 V RMS in the Dactron system, which corresponds to, approximately, 20 N RMS measured from the load cell. Force and acceleration time histories for a variety of different structural state conditions were collected as shown in Table 1 along with information that describes the different states. For example, the state condition labeled "State#4" is described as "87.5% stiffness reduction in column 1BD," which means that there was an 87.5% stiffness reduction in the column located between the base and first floor at the intersection of plane B and D. For each structural state condition, 10 tests were performed to take into account the variability in the data. Thus, in each state condition and for each of the five transducers, 10 time histories were measured.

The structural state conditions can be categorized into four main groups. The first group is the baseline condition. The baseline condition is the reference structural state and is labeled State#1 in Table 1. The bumper and the suspended column are included in the baseline condition, but the spacing between them was maintained in such a way that there were no impacts during the excitation. The second group includes the states with simulated operational and environmental variability. Operational and environmental variations often cause changes in the stiffness or mass distribution of the structure. To simulate such changes, tests were performed with different mass-loading and stiffness conditions (States#2–9). The mass changes consisted of adding a 1.2 kg (approximately 19% of the total mass of each floor) to the first floor and to the base,

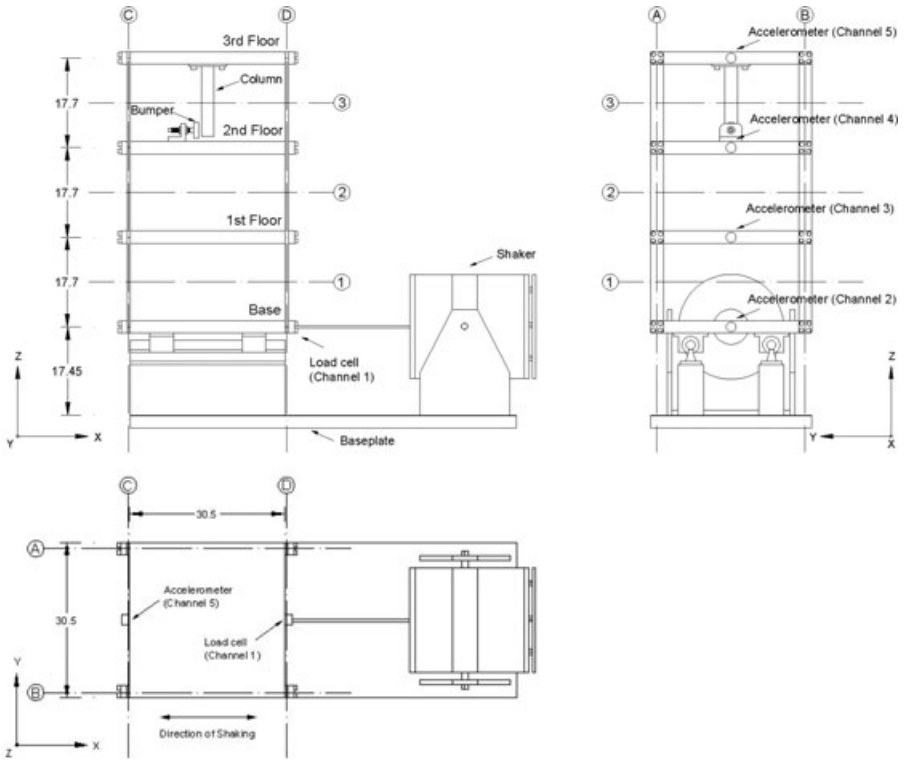

Fig. 1. Basic dimensions of the three-story test bed structure (all dimensions are in cm).

as shown in Figure 2a when the mass is at the base. The stiffness changes were introduced by reducing one or more of the column stiffnesses by 87.5%. This process was executed by replacing a column with one with half the cross-sectional thickness in the direction of shaking. Recall that those changes were designed to incorporate variations in the fundamental natural frequency up to 7% from the baseline condition. The third group includes the damaged state conditions (States#10–14); these were simulated by the introduction of nonlinearities (damage) into the structure using the bumper and the suspended column with different gaps between them, as shown in Figure 2b. The gap was varied (0.20, 0.15, 0.13, 0.10, and 0.05 mm) to introduce different levels of nonlinearity. Finally, to create more realistic conditions, the fourth group includes state conditions with nonlinearities in addition to the mass and stiffness changes used to simulate operational and environmental variations (States#15–17). Even though the simulated changes were designed to roughly match the variability expected in the natural frequencies for real-world structures, note that an 87.5% stiffness reduction in a column and/or the addition of 19% floor mass represents a very severe operational and environmental variation and are not considered damage in this study. More details about the test structure as well as data sets can be found in Figueiredo et al. (2009).

3 FEATURE EXTRACTION: THE AUTOREGRESSIVE MODEL

Time-series analysis takes into account the fact that data points sampled over time may have an internal structure, such as autocorrelations, trends or seasonal variations. The AR(p) model, where p defines the order or number of parameters in the model, is developed from response time-series data $x_1, x_2, \ldots, x_n$ and it can be written as

$$x_i = \sum_{j=1}^{p} \phi_j x_{i-j} + \varepsilon_i \tag{1}$$

where x_i is the measured signal at the time instant t_i. Regular sampling is assumed, i.e., $t_{i+1} = t_i + \Delta t$, where Δt is a constant sampling interval. The quantity ε_i is an unobservable random error (or residual error) at the signal value i. The unknown AR parameters, ϕ_j, can be estimated by using either the least-squares approach or the Yule-Walker equations (Box and Jenkins, 1976).

In SHM, the AR model can be used as a damage-sensitive feature extractor based on two approaches: (i) using the residual errors ε_i; and (ii) using the AR parameters ϕ_j. The first approach consists of using the AR model, with parameters estimated from the baseline condition, to predict the response of data obtained from

Table 1
Data labels of the structural state conditions

Label	*State condition*	*Description*
State#1	Undamaged	Baseline condition
State#2	Undamaged	Mass = 1.2 kg at the base
State#3	Undamaged	Mass = 1.2 kg on the 1st floor
State#4	Undamaged	87.5% stiffness reduction in column 1BD
State#5	Undamaged	87.5% stiffness reduction in column 1AD and 1BD
State#6	Undamaged	87.5% stiffness reduction in column 2BD
State#7	Undamaged	87.5% stiffness reduction in column 2AD and 2BD
State#8	Undamaged	87.5% stiffness reduction in column 3BD
State#9	Undamaged	87.5% stiffness reduction in column 3AD and 3BD
State#10	Damaged	Gap = 0.20 mm
State#11	Damaged	Gap = 0.15 mm
State#12	Damaged	Gap = 0.13 mm
State#13	Damaged	Gap = 0.10 mm
State#14	Damaged	Gap = 0.05 mm
State#15	Damaged	Gap = 0.20 mm and mass = 1.2 kg at the base
State#16	Damaged	Gap = 0.20 mm and mass = 1.2 kg on the 1st floor
State#17	Damaged	Gap = 0.10 mm and mass = 1.2 kg on the 1st floor

a potentially damaged structural condition. The residual error, which is the difference between the measured and the predicted signal, is calculated at time i as follows:

$$\varepsilon_i = x_i - \hat{x}_i \tag{2}$$

where $\hat{x}_i$ is the prediction at the ith sampling instant. For the baseline condition, the residual errors are generally assumed to be independent and normally distributed with zero mean and variance σ^2, $\varepsilon \sim N(0, \sigma^2)$. The first SHM approach (indicated above) is based on the assumption that damage will introduce either linear deviations from the baseline condition or nonlinear effects in the signal and, as a result, the linear model developed with the baseline data will no longer accurately predict the response of the damaged structure. As a consequence, the residual errors associated with the damaged structure will increase. The second SHM approach consists of fitting an AR model to signals from the undamaged and damaged structural conditions. In this approach, the AR parameters are used directly as damage-sensitive features, and some form of a multivariate classifier can be used to distinguish between the damage classes. Note that the parameters should be constant when obtained from time histories of a time-invariant structural system.

The appropriate AR model order is initially unknown. A higher order model may better fit the data, but may not generalize well to other data sets. On the other hand, a low-order model will not necessarily capture the underlying physical system dynamics. Model order estimation remains a very complex issue, and various techniques to address this problem have been proposed. Based on the mathematical formulation, the techniques can be classified into three major categories (Cassar et al., 2009). The first category requires an *a priori* estimate of the model parameters to find the optimal order. Among these techniques, the AIC and PAF have attracted much attention in the literature and the RMS error is a heuristic approach. Second, there are techniques that do not require an *a priori* estimate of the model parameters. This approach is generally based on the SVD of the data covariance matrix. Another category of techniques, but not reported in this study, includes Bayesian approaches that estimate the model order and model parameters, simultaneously, at the cost of more computational complexity. In this article, a

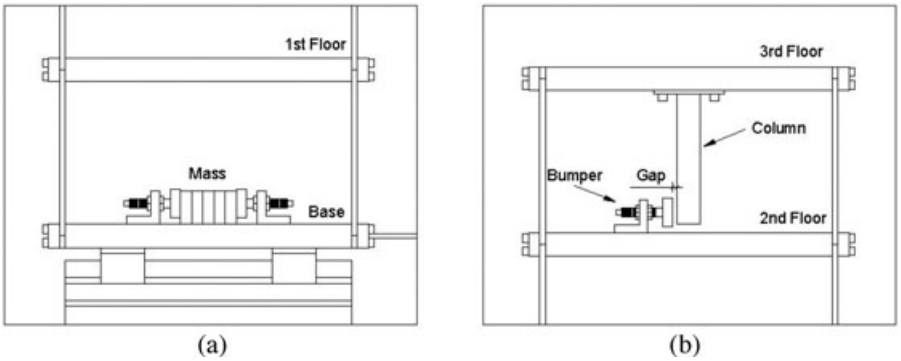

Fig. 2. Structural details of the sources of simulated operational and environmental changes. (a) Mass-loading added at the base. (b) Nonlinearity source.

brief explanation of the four mentioned techniques will be given for completeness.

The AIC has often been used in the past to assess the generalization performance of linear models (Akaike, 1974). This technique simply computes a statistic that is the sum of two terms as follows:

$$\mathrm{AIC}(p) = NLn(e) + 2p \tag{3}$$

where $e = SSE/N$ is an average sum of squared errors and $N = n - p$ is the number of observations used in fitting the model. It clearly represents a trade-off between the fit of the model and the model complexity. The first term is related to how well the model fits the data, i.e., if the model is too simple the residual errors increase. On the other hand, the second term is a penalty factor related to the complexity of the model, which increases as the number of parameters grows. The optimal order p is given by the AR model with the lowest AIC value.

The PAF has also been used effectively to estimate the AR model order by Box and Jenkins (1976). The PAF is defined by fitting AR models of successively increasing order p to the measured data and then plotting the last estimated parameter, ϕ_{pp}, as a function of the model order. Thus, the AR model of Equation (1) can be rewritten as follows:

$$x_i = \sum_{j=1}^{p} \phi_{pj} x_{i-j} + \varepsilon_i \tag{4}$$

For an AR model of a noise-free order-p process, the PAF coefficients, ϕ_{kk}, will be nonzero for $k \leq p$ and zero for $k > p$. For real-world structures with noise in the measurements, the partial autocorrelation coefficients of an actual AR(p) will not be exactly zero after lags greater than p, but will assume small random values. As a result, it is necessary to define upper and lower bounds such that the coefficients will be considered zero if they fall within these limits. The idea is to look for the points of the PAF that are essentially zero. An approach based on the standard deviation error, σ_ϕ, is defined as follows:

$$\sigma_\phi[\phi_{kk}] = \frac{1}{\sqrt{N}}, \quad k \geq p+1 \tag{5}$$

where the estimated partial autocorrelation coefficients of order $p + 1$ and higher are approximately independently distributed. Placing a confidence interval for statistical significance is helpful for this purpose. For example, assuming an approximate 95% confidence interval, the limits are placed at $\pm 2\sigma_\phi$.

A heuristic technique to establish the appropriate AR model order is by means of the RMS of the AR residuals for varying order p. For each model order, the RMS of the residuals becomes

$$\mathrm{RMS}(p) = \left(\frac{1}{N} \sum_{i=1}^{N} (\hat{x}_i - x_i)^2 \right)^{1/2} \tag{6}$$

In this case, the optimal model order will be at the convergence point of the RMS error values for varying order p.

Finally, an estimation of the optimal AR model order can also be given by using the SVD technique. In linear algebra, the SVD is a factorization of a rectangular matrix defined as follows:

$$\mathbf{M} = \mathbf{U}\Lambda\mathbf{V}^T \tag{7}$$

where the matrix Λ contains the singular values on the diagonal and zeros for the off-diagonal terms. The matrices $\mathbf{U}$ and $\mathbf{V}$ are both orthogonal. The SVD is used to determine the effective rank of a matrix $\mathbf{M}$ by counting the nonzero singular values. To conduct the analysis, Equation (1) can be written as

$$\begin{Bmatrix} x_{p+1} \\ x_{p+2} \\ \vdots \\ x_n \end{Bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_p \\ x_2 & x_3 & \cdots & x_{p+1} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n-p} & x_{n-p+1} & \cdots & x_{n-1} \end{bmatrix} \begin{Bmatrix} \phi_p \\ \phi_{p-1} \\ \vdots \\ \phi_1 \end{Bmatrix} \tag{8}$$

or

$$\mathbf{x} = \mathbf{M}\phi \tag{9}$$

where $\mathbf{x}$ and $\mathbf{M}$ contain the measured signal data and ϕ contains the AR parameters. One can also express Equation (8) in the form

$$\mathbf{z}_{p+1} = \phi_p \mathbf{z}_1 + \phi_{p-1}\mathbf{z}_2 + \cdots + \phi_1 \mathbf{z}_p \tag{10}$$

Hence, the SVD counts how many of the $\mathbf{z}$-vectors (the columns of $\mathbf{M}$) are linearly independent and therefore needed for an accurate model, which in turn defines the number of necessary AR parameters.

4 AUTOREGRESSIVE RESIDUAL ERRORS AND PARAMETERS AS DAMAGE-SENSITIVE FEATURES

4.1 Stationarity property of the data sets

Because no explicit mathematical equation can be written for the time histories produced by a random excitation, such as the measured data in this study, statistical procedures must be used to define their properties. In statistics, the data from a random process are said to be stationary and (essentially) ergodic when the

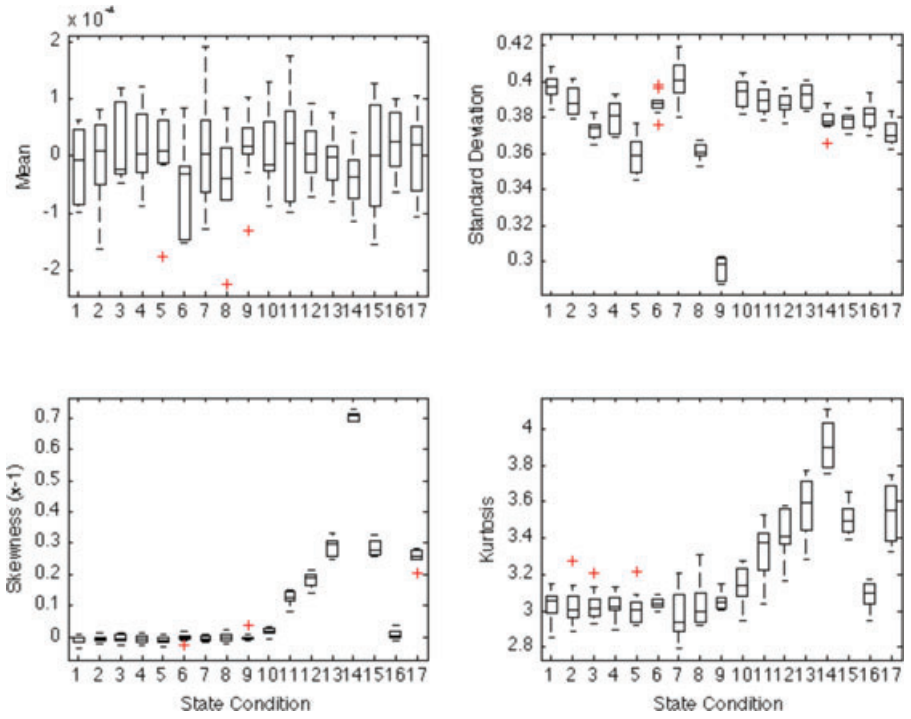

Fig. 3. Box plots of the first four statistical moments of all 10 time histories from Channel 5.

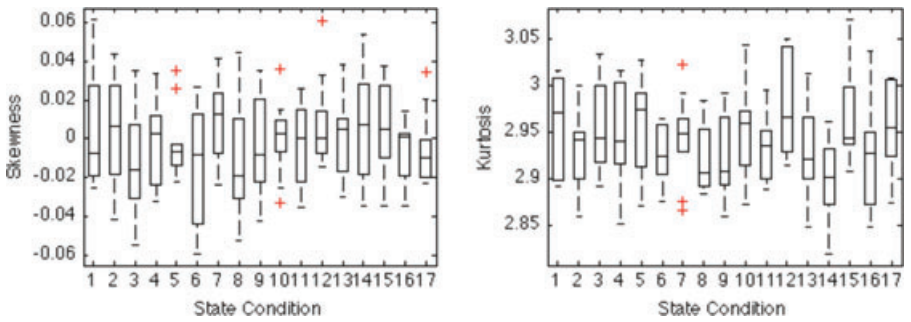

Fig. 4. Box plots of the skewness and kurtosis of all 10 time histories from Channel 1.

statistical and joint moments are time invariant and do not change when computed over different time histories (Bendat and Piersol, 2000). In this study, some of the basic statistical properties such as the first four statistical moments (mean, standard deviation, skewness, and kurtosis) as well as the autocorrelation functions (ACFs) are used to describe the data. The box plots in Figure 3 indicate the degree of dispersion and symmetry in the statistical moments computed over all 10 time histories from Channel 5 of each state condition. The bottom and top of each box are the lower and upper quartiles, respectively. (Note that the crossed signs correspond to outliers, i.e., values more than 1.5 times the interquartile range away from the bottom or top of each box.) The scatter highlighted in the mean is related to the arbitrary plot scale and is largely meaningless because the mean is approximately zero for all state conditions and cannot be used to scale the figure. Looking at the standard deviation, skewness, and kurtosis, one observes that, although there is variation in the statistics across states (between-class scatter), there is very little variation across the test sequence for a given state (within-class scatter). As, for each state, the statistics do not change significantly across the test sequence, the structural system in each state can be assumed stationary. As highlighted in the figure for the damaged states (States#10–17), the skewness and kurtosis actually prove to be features revealing the presence of nonlinearities (and hence the damage), as indicated by significant departures from the expected values of a Gaussian signal (zero for skewness and three for normalized kurtosis). Furthermore, the skewness and kurtosis of

the responses for Channel 1 in Figure 4 indicate that this channel is insensitive to the presence of damage. This result in itself is important because it shows that the departures from the Gaussian condition for Channel 5 cannot be blamed on the possibility of non-Gaussian input signals—the between-class scatter on the skewness and kurtosis for the different states of Channel 1 is not significant. As further evidence of stationarity for the various states, the ACFs of the responses for each state are relatively consistent across the sequence of 10 tests. For illustration purposes, the box plots in Figure 5 indicate the degree of dispersion and symmetry in the first 15 autocorrelation coefficients computed over all 10 time histories from Channel 5 of States#1 and 14.

In summary, based on the observations above, it can be assumed that the random processes associated with all channels are at least weakly stationary and may be stationary in the wide sense. Moreover, it was shown that the changes in the statistical moments of the response are not the result of changing inputs. This assumption implies that the statistics of a single time history are representative of the entire time-history ensemble corresponding to a specific damaged state condition. Based on this assumption, throughout this article and for each state condition, only one time history per state is used to discriminate features from the undamaged and damaged state conditions. Furthermore, for illustration purposes, only the responses from Channel 5 are used in the remainder of the article.

4.2 AR model estimation

The analysis performed to determine the optimal AR model order uses one measured time history of Channel 5 from the baseline condition (State#1). The plots in Figure 6 are produced based on the four techniques described in Section 3. The specific goal of using multiple techniques is to verify the consistency of the techniques to find the optimal number of parameters needed to fit the data and to detect damage in the form of nonlinearities. Figures 6a–c plot, respectively, results from the AIC, PAF, and RMS error techniques obtained by directly fitting AR models of increasing order $p = 1, 2, \ldots, 60$ to the measured response data. Additionally, Figure 6d plots the singular values of $\mathbf{M}$ in Equation (9). Assuming $p = 60$, the matrix $\mathbf{M}$ has a dimension of 8,132 $\times$ 60. Note that the abscissa of the plot is, in some sense, related to the AR model order as indicated by Equation (10). From these plots one can point out some facts that compromise the definition of the optimal model order: (i) the AIC function is not minimized in the plotted window; (ii) in the case of the PAF technique, after convergence the coefficients do not stay within the 95% confidence limits; and (iii) for the other two techniques, the point of convergence is not precisely defined. Because of these issues, optimal model order selection still requires interpretation as stated by (Al-Smadi and Wilkes, 2002). Even though it is not possible to establish a unique solution, the results suggest that the optimal order is in the range of 15–30. Additionally, it was confirmed that the AIC, PAF, and RMS error techniques suffer from the fact that all parameters of the models corresponding to different model orders first must be estimated to calculate these values. On the other hand, the SVD technique does not require prior estimation of the model parameters, which reduces the computational effort associated with this technique compared to the others. Even though the SVD technique gives an upper bound solution, the lower computational efforts can be an advantage for implementation on embedded hardware.

Based on this analysis, three AR models, namely AR(5), AR(15), and AR(30), are used throughout this article to indicate the influence of the AR model order on the damage detection process under operational and environmental variations. These values were chosen so the first is unequivocally too small, although the second and third bound the plausible range of model orders suggested by these four techniques.

4.3 AR residual errors as damage-sensitive features

The first approach to SHM based on the AR model uses parameters estimated from responses measured on the structure in its baseline condition to predict the responses from potentially damaged conditions. The AR parameters here are estimated by fitting the AR models to a time history from Channel 5 of the baseline condition using the least-squares technique. For comparison purposes, Figure 7 shows a 20-point window of the measured and predicted acceleration time histories using the three proposed models: AR(5), AR(15), and AR(30). From a qualitative point of view, all the models have captured the underlying physical structural response; however, as expected, the higher-order models predict the data better than the AR(5) model. (One should generally observe the usual caveat that the higher-order models are potentially overfitted.)

One indication that an AR model is fitting the data well is given by independent and normally distributed residual errors. For one time history of the baseline condition, Figure 8 shows histograms of the residual errors using 20 bins along with a superimposed Gaussian distribution based on the sample mean and standard deviation. Note that depending on the level of nonlinearity introduced by the bumper, the distribution of the residual errors may not be Gaussian when the AR model derived from the baseline condition is used to predict

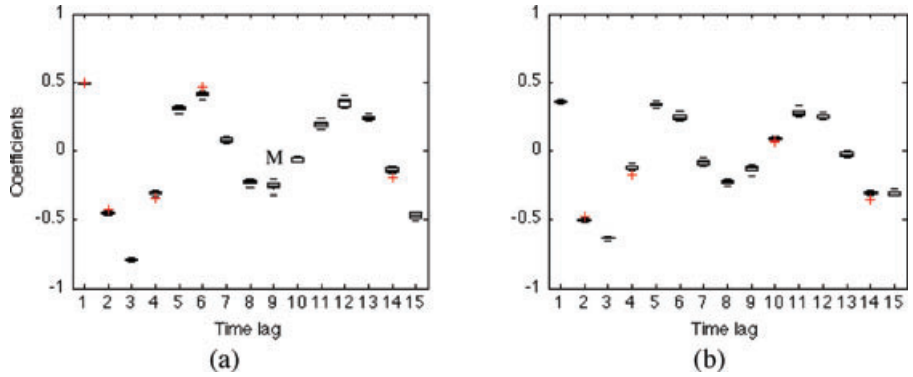

Fig. 5. Box plots of the ACFs of all time histories from Channel 5: (a) State#1 and (b) State#14.

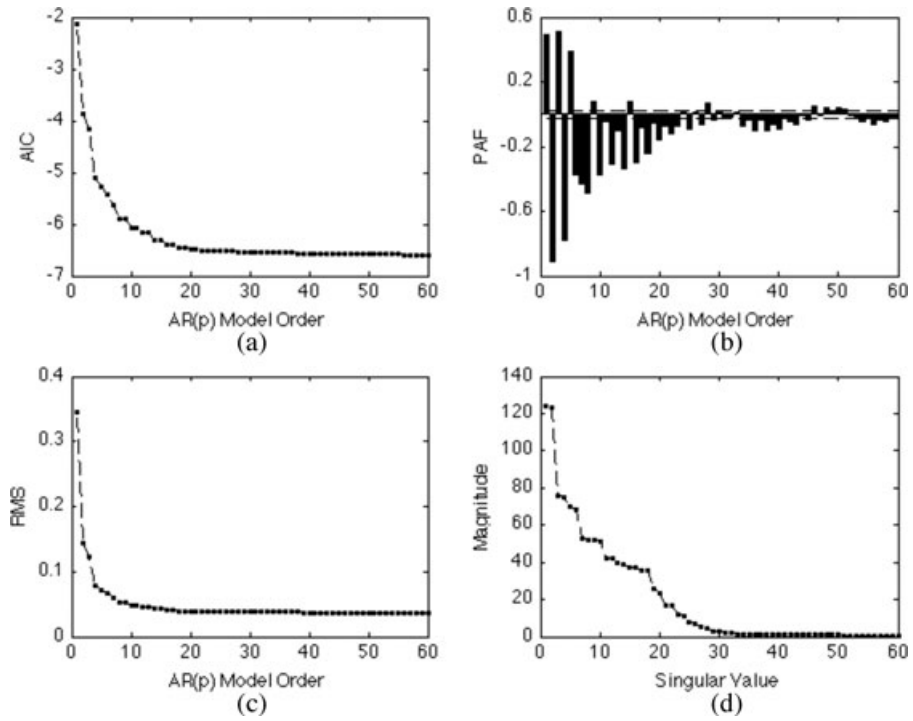

Fig. 6. AR order estimation based on one time history of Channel 5 from the baseline condition (State#1) using four different techniques: (a) AIC, (b) PAF, (c) RMS, and (d) SVD.

the responses of the damaged states. As expected, for the baseline condition, the histograms show that the higher the order of the AR model, the lower will be the variance of the residual errors.

Even though the residual error sequences of the three models have, apparently, an underlying normal distribution, other techniques might be used to verify the data independency. In this article, the power spectral density (PSD) functions of the residual errors are used to find correlations or the presence of periodic signals buried under noise. Figure 9 illustrates the PSDs of the residual errors from the three models. The plots show that the resulting residuals from the AR(5) model are still correlated, because it is possible to identify three natural frequencies of the structure at 30.7, 54.2, and 70.7 Hz (Figueiredo et al., 2009). On the other hand, the residual errors from the AR(15) and AR(30) models seem to be uncorrelated based on the relatively flat PSDs across the entire frequency range.

As indicated by the model order estimation techniques, the previous analysis suggests that an AR model order within the interval 15–30 is appropriate, because the residual errors given by Equation (2) are independent and normally distributed when the baseline response is calculated with parameters estimated from the same condition. The next step is to use the AR model

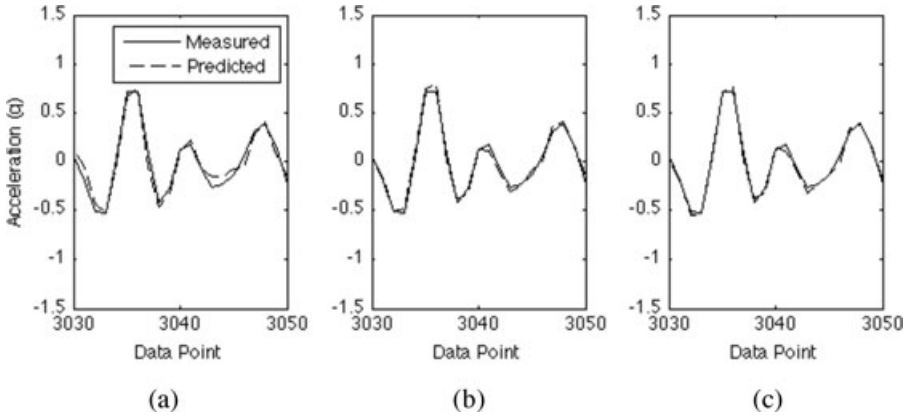

Fig. 7. Comparison of the measured and predicted time histories from the baseline condition: (a) AR(5), (b) AR(15), and (c) AR(30) models.

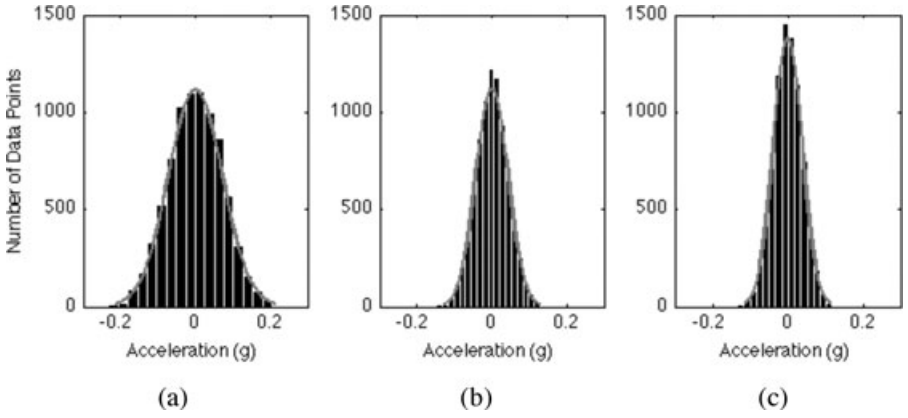

Fig. 8. Residual errors histograms along with the superimposed Gaussian distribution based on one time history from the baseline condition: (a) AR(5), (b) AR(15), and (c) AR(30) models.

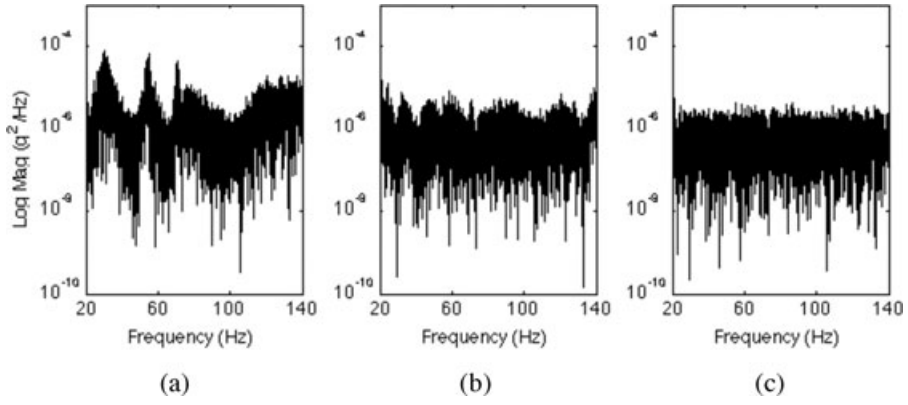

Fig. 9. Log PSDs of the residual errors based on one time history from the baseline condition: (a) AR(5), (b) AR(15), and (c) AR(30) models.

derived from the baseline condition to predict response data from the remainder conditions. If anomalies are present in the system, the residual error variance will generally increase. Note that this approach is based on the assumption that anomalies will introduce either linear deviation from the baseline condition or nonlinear effects in the signal; as a result, the linear baseline model will no longer accurately predict the response of the damaged structure. With such an indication of deviation from the baseline, the next action should be to launch

a full investigation to identify the cause of the anomalies. It is critical to note that, in this case, anomalies can result from either damage or from operational or environmental effects. To discriminate the latter anomalies from those caused by damage, this approach expects that the damage might introduce underlying nonrandom patterns in the measured data that will be evident from the residual errors.

The correlation analysis is one statistical technique used to find patterns within time series, such as the presence of periodic signals buried under noise due to damage (Figueiredo et al., 2009). Figure 10 shows the normalized ACFs (for the first 200 coefficients) of the acceleration time histories from Channel 5 of States#1, 7, 14, and 17. (Coefficients are scaled so that the ACF at zero time lag equals one.) The ACFs of the original time histories suggest the presence of strong correlation in the time histories. Moreover, it is difficult to make any inferences about the structural condition based on these plots.

Figure 11a shows the normalized ACFs of the residual errors from the AR(5) model for the four conditions mentioned above. Theoretically, if the AR model accurately represents the original time histories, the residual errors should be nearly uncorrelated and the ACF should approximate a delta function. Looking at the plots, one can conclude that the residual errors from the AR(5) model are still correlated. However, Figures 11b and c indicate that increasing the AR model order, into an interval spanning from 15 to 30, reduces the correlation among the residual errors for the undamaged conditions (States#1 and 7) and, at the same time, it shows correlations in the damaged conditions (States#14 and 17) generated by the nonlinearities introduced by the bumper (the simulated damage) and that is difficult to detect visually in the original time histories. In addition, these results indicate that residual errors associated with the AR(15) and AR(30) models will be able to distinguish the damaged cases from the undamaged ones even under operational and environmental variability. Such distinctions are not as clear with the AR(5) model. This fact points out the need to choose an appropriate AR model order to capture the damage-related information buried within the response signals subjected to operational and environmental variability.

4.4 AR model parameters as damage-sensitive features

The underlying idea here is to exploit the AR parameters themselves as damage-sensitive features. One might expect changing amplitudes of these parameters in the presence of nonlinearities caused by damage even under operational and environmental variations. Figure 12 shows the parameters from the AR(5), AR(15), and AR(30) models estimated on one time history of each structural state condition using the least-squares technique. For the sake of simplicity, the parameters are arranged into two groups; those states corresponding to the undamaged (States#1–9) and damaged (States#10–17) conditions. The figures suggest that increasing the number of impacts (by reducing initial gap at the impacting device) associated with the damaged conditions tends to decrease the AR model parameter amplitudes. Furthermore, Figure 12 clearly shows that the model parameters associated with the AR(15) and AR(30) models obtained from the undamaged structure data are distinct from those obtained from the damaged structure even in the presence of simulated operational and environmental variability. However, there is not such a clear distinction with the corresponding parameters of the AR(5) model. To better indicate these changes, Figure 13 shows the amplitude of the third AR parameter under each state condition for the three models. One can clearly identify (quasi-) linear changes in the amplitude of the parameter as a function of the initial gap for the damaged states without any simulated operational and environmental variations (States#10–14). However, the parameter amplitudes for the damaged states with simulated operational and environmental variability (States#15–17) do not clearly indicate such linear changes. Furthermore, the AR(5) model cannot accurately discriminate all the damaged and undamaged states when simulated operational and environmental variability is present. For example, the AR(5) parameter amplitude corresponding to the damaged State#16 is higher than the one for the undamaged State#7. This result is a clear indication that the operational and environmental variations can introduce changes in the structural response and mask changes in the responses related to damage when an inappropriate model order is used.

As discussed for the AR residual errors, these results highlight the importance of properly establishing the AR model order in an effort to avoid false-negative and false-positive indications of damage. In the case of the AR(15) and AR(30) models, the operational and environmental variations only affect the nearly linear relationship between the level of damage and model parameter amplitude. In the hierarchical structure of damage identification (Rytter, 1993), this fact highlights the usefulness of these damage-sensitive features to identify the existence of damage in the structure when the appropriate AR model order is used. To move to higher levels in the hierarchical structure, such as locating and quantifying damage, one needs to analyze relative data across the sensor network and to carry out statistical modeling for efficient feature classification.

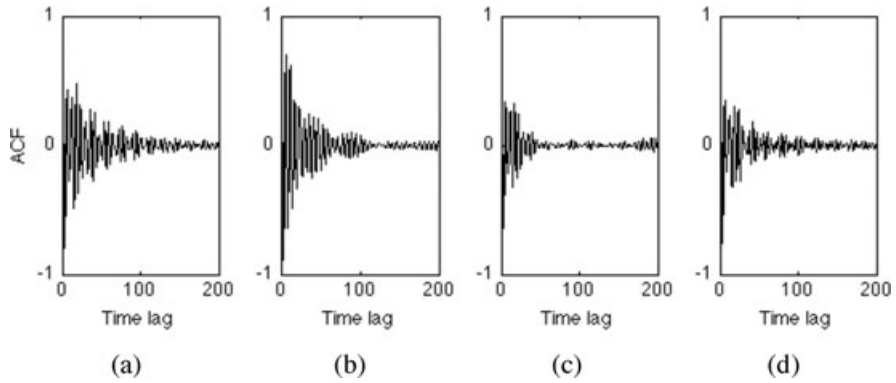

Fig. 10. ACFs of acceleration time histories for four structural conditions: (a) State#1, (b) State#7, (c) State#14, and (d) State#17.

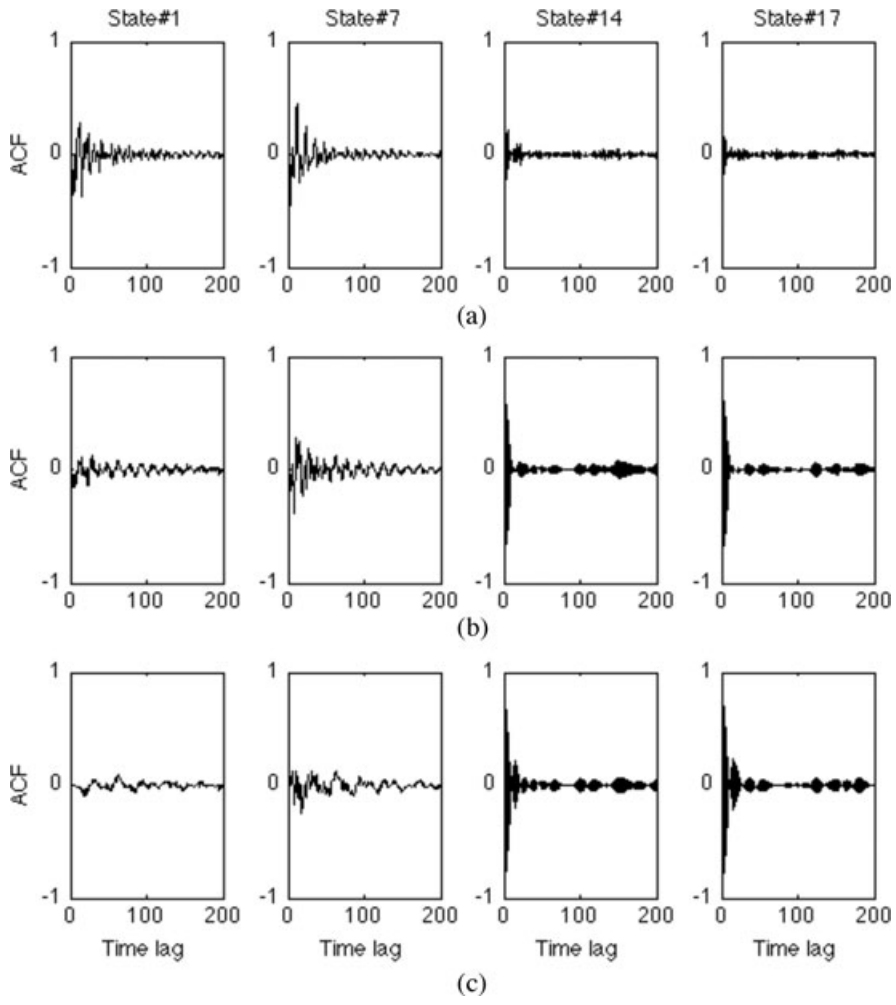

Fig. 11. ACFs of the residual errors from States#1, 7, 14, and 17: (a) AR(5), (b) AR(15), and (c) AR(30) models.

5 SUMMARY AND CONCLUSIONS

In general, real-world structures are subjected to operational and environmental condition changes, which complicate and obscure the detection of structural damage. The approach presented herein addresses the need for the development of feature extraction techniques capable of distinguishing the variations in the features related to damage from those related to operational and environmental effects. The damage is simulated by

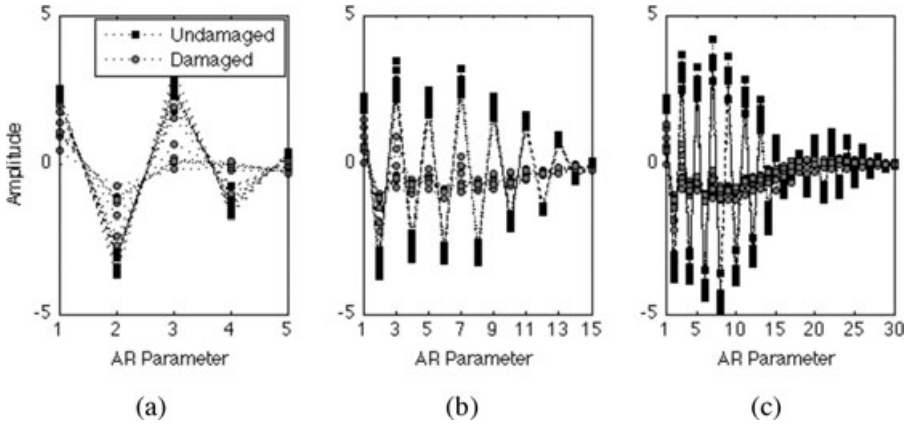

Fig. 12. AR model parameter amplitudes for all the state conditions: (a) AR(5), (b) AR(15), and (c) AR (30) models.

the nonlinear effect of introducing a bumper mechanism that simulates a repetitive impact-type nonlinearity. This mechanism is intended to simulate cracks that open and close or loose connections that rattle under dynamic loads. The operational and environmental effects were simulated by imposing different mass and stiffness distributions in such a way that the variability in the fundamental natural frequency is representative of the variability that has been observed in real-world structures.

The AR model was used to extract features from acceleration time-series data. However, the model order estimation process remains a very complex issue and various techniques to address this problem have been proposed. In this study, four techniques based on AIC, PAF, RMS error, and SVD were presented. It was observed that it is not possible to define a unique solution for the AR model order when the techniques are applied to baseline condition data. Nonetheless, the results suggest that the optimal order is in the range of 15–30. Additionally, it was confirmed that the AIC, PAF, and RMS error techniques suffer from the fact that all parameters from the models of each order must be estimated first to calculate those values. On the other hand, the SVD technique does not require prior estimation of the model parameters, which requires reduced computational effort compared to the other three techniques. Even though the SVD technique gives the upper bound solution, the lower computational effort was pointed out as an advantage for implementation on embedded hardware.

The unique contribution of this study was the comparison carried out between three different AR models that highlight the influence of the AR model order on the damage detection process. The analyses performed on data sets from a three-story base-excited frame structure showed that the lower bound of the range of AR model orders defined by the four techniques is capable of discriminating the undamaged and damaged state conditions when using either the AR residual errors or the model parameters as damage-sensitive features, even in the presence of simulated operational and environmental variability. The analysis also showed that using an AR model of order less than that suggested by

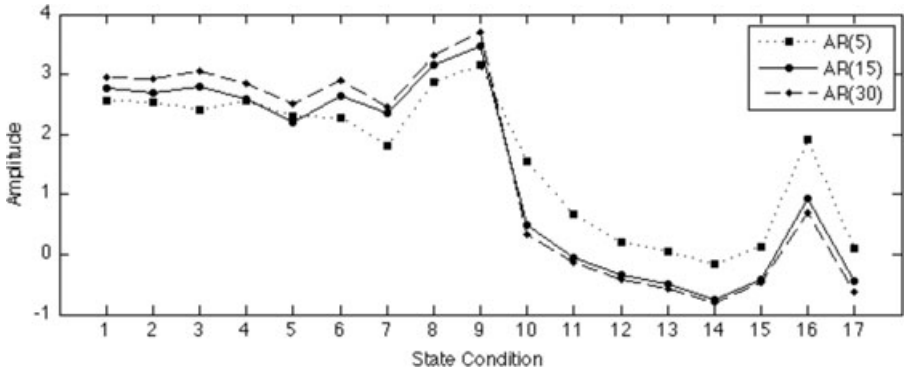

Fig. 13. Amplitude of the third parameter for all the state conditions from the three AR models.

the four techniques does not allow discrimination of all the undamaged and damaged state conditions when the operational and environmental variability was included in the experiments. Finally, the results reported herein indicate that AR modeling can be an effective feature extraction technique when the damage introduces nonlinearities and the operational and environmental variations manifest themselves as linear system changes.

REFERENCES

Akaike, H. (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **AC-19**(6), 716–23.

Alampalli, S. (2000), Effects of testing, analysis, damage, and environment on modal parameters, *Mechanical Systems and Signal Processing*, **14**(1), 63–74.

Al-Smadi, A. & Wilkes, D. M. (2002), Robust and accurate ARX and ARMA model order estimation of non-Gaussian processes, *IEEE Transactions on Signal Processing*, **50**(3), 759–63.

Bendat, J. S. & Piersol, A. G. (2000), *Random Data: Analysis and Measurement Procedures*, 3rd edn. John Wiley and Sons, Inc., New York, NY.

Box, G. & Jenkins, G. (1976), *Time Series Analysis: Forecasting and Control*, Prentice Hall, Englewood Cliffs, NJ.

Cassar, T. A., Camilleri, K. P. & Fabri, S. G. (2009), Order estimation of computational models for dynamic systems with application of biomedical signals, *Third International Conference on Advanced Engineering Computing and Applications in Sciences*, 163–8.

Doebling, S. W., Farrar, C. R., Prime, M. B. & Shevitz, D. (1996), Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review, Los Alamos National Laboratory: LA-13070-MS.

Farrar, C. R., Cornwell, P. J., Doebling, S. W. & Prime, M. B. (2000), Structural health monitoring studies of the Alamosa Canyon and I-40 bridges, Los Alamos National Laboratory Report: LA-13635-MS.

Farrar, C. R., Doebling, S. W. & Nix, D. A. (2001), Vibration-based structural damage identification, *Philosophical Transactions of the Royal Society: Mathematical, Physical & Engineering Sciences*, **356**(1778), 131–49.

Farrar, C. R., Worden, K., Todd, M. D., Park, G., Nichols, J., Adams, D. E., Bement, M. T. & Farinholt, K. (2007), Nonlinear system identification for damage detection, Los Alamos National Laboratory Report: LA-14353-MS.

Figueiredo, E., Park, G., Figueiras, J., Farrar, C. & Worden, K. (2009), Structural health monitoring algorithm comparisons using standard data sets, Los Alamos National Laboratory: LA-14393.

Ihn, J. B. & Chang, F. K. (2004), Detection and monitoring of hidden fatigue crack growth using a built-in piezoelectric sensor/actuator network: II. Validation using riveted joints and repair patches, *Smart Materials and Structures*, **13**(3), 621–30.

Kantz, H. & Schreiber, T. (2004), *Nonlinear Time Series Analysis*, 2nd edn. Cambridge University Press, Cambridge, UK.

Kim, C.-Y., Jung, D.-S., Kim, N.-S., Kwon, S.-D. & Feng, M. Q. (2003), Effect of vehicle weight on natural frequencies of bridges measured from traffic-induced vibration, *Earthquake Engineering and Engineering Vibration*, **2**(1), 109–15.

Ko, J. M. & Ni, Y. Q. (2005), Technology developments in structural health monitoring of large-scale bridges, *Engineering Structures*, **27**, 1715–25.

Moaveni, B. He, X. & Conte, J. P. (2008), Damage identification of a composite beam based on changes of modal parameters, *Computer-Aided Civil and Infrastructure Engineering*, **23**(5), 339–59.

Park, G., Soon, H., Farrar, C. R. & Inmar, D. J. (2003), Overview of piezoelectric impedance-based health monitoring and path forward, *Shock and Vibration Digest*, **35**(6), 451–63.

Peeters, B. & De Roeck, G. (2001), One-year monitoring of the Z24-bridge: environmental effects versus damage events, *Earthquake Engineering and Structural Dynamics*, **30**, 149–71.

Pimentel, M., Santos, J. & Figueiras, J. (2008), Safety appraisal of an existing bridge via detailed modelling, *International FIB Symposium 2008: Tailor Made Concrete Solutions, New Solutions for our Society*. Amsterdam, Netherlands.

Rytter, A. (1993), Vibration based on inspections of civil engineering structures, PhD dissertation, Department of Building Technology and Structural Engineering, Alborg University, Denmark.

Sohn, H., Farrar, C., Hunter, N. & Worden, K. (2001), Applying the LANL statistical pattern recognition paradigm for structural health monitoring to data from a surface-effect fast patrol boat, Los Alamos National Laboratory Report: LA-13761-MS.

Sohn, H., Farrar, C. R., Hemez, F. M., Shunk, D. D. & Stinemates, D. W. (2004), A review of structural health monitoring literature from 1996–2001, Los Alamos National Laboratory: LA-1397-MS.

Sohn, H., Kim, S. D. & Harries, K. (2008), Reference-free damage classification based on cluster analysis, *Computer-Aided Civil and Infrastructure Engineering*, 23(5), 324–38.

Soyoz, S. & Feng, M. Q. (2009), Long-term monitoring and identification of bridge structural parameters, *Computer-Aided Civil and infrastructure Engineering*, **24**, 82–92.

Worden, K., Farrar, C. R., Manson, G. & Park, G. (2007), The fundamental axioms of structural health monitoring, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **463**(2082), 1639–64.

Xia, Y., Hao, H., Zanardo, G. & Deeks, A. (2006), Long term vibration monitoring of an RC slab: temperature and humidity effect, *Engineering Structures*, **28**, 441–52.