Chen Liujie, Yu Ling, Fu Jiyang, Ng Ching-Tai

Nonlinear damage detection using linear ARMA models with classification algorithms

Smart Structures and Systems, 2020; 26(1):23-33

Journal article:

Liujie Chen, Ling Yu, Jiyang Fu, Ching Tai Ng. (2020). Nonlinear damage detection using linear ARMA models with classification algorithms. Smart Structures and System, 26(1):23-33.

Nonlinear damage detection using linear ARMA models with classification algorithms

Liujie Chen[1, 2a], Ling Yu[2b], Jiyang Fu[*3c], Ching Tai Ng[4d]

[1]School of Civil Engineering, Guangzhou University, No.230, Wai Huan Xi Road, Guangzhou Higher Education Mega Center, Guangzhou, China
[2]MOE Key Lab of Disaster Forecast and Control in Engineering, Jinan University, No.601, West Huangpu Avenue, Guangzhou, China
[3]Guangzhou University-Tamkang University Joint Research Center for Engineering Structure Disasters Prevention and Control, No.230, Wai Huan Xi Road, Guangzhou Higher Education Mega Center, Guangzhou, China
[4]School of Civil, Environmental & Mining Engineering, The University of Adelaide, North Terrace, Adelaide, SA 5005, Australia

(Received , Revised , Accepted)

Abstract. Majority of the damage in engineering structures is nonlinear. Damage sensitive features (DSFs) extracted by traditional methods from linear time series models cannot effectively handle nonlinearity induced by structural damage. A new DSF is proposed based on vector space cosine similarity (VSCS), which combines K-means cluster analysis and Bayesian discrimination to detect nonlinear structural damage. A reference autoregressive moving average (ARMA) model is built based on measured acceleration data. This study first considers an existing DSF, residual standard deviation (RSD). The DSF is further advanced using the VSCS, and then the advanced VSCS is classified using K-means cluster analysis and Bayes discriminant analysis, respectively. The performance of the proposed approach is then verified using experimental data from a three-story shear building structure, and compared with the results of existing RSD. It is demonstrated that combining the linear ARMA model and the advanced VSCS, with cluster analysis and Bayes discriminant analysis, respectively, is an effective approach for detection of nonlinear damage. This approach improves the reliability and accuracy of the nonlinear damage detection using the linear model and significantly reduces the computational cost. The results indicate that the proposed approach is potential to be a promising damage detection technique.

Keywords: nonlinear damage detection; time series analysis; linear autoregressive moving average model; vector space cosine similarity; classification algorithms

---

1. Introduction

In the last two decades, there have been considerable developments in the area of Structural Health Monitoring (SHM). Worden *et al.* (2007) stated that the fundamental problem of SHM is damage identification. Different damage detection and identification techniques have been developed and investigated in the literature, for example using modal parameters (Yin *et al.* 2017; Ng and Au 2018) and time domain vibration data (Lam *et al*. 2017). One of the important and rapidly evolving approaches in the area of damage detection using vibration data is time series analysis. Originally, it was developed to analyze regularly sampled long sequences data and is inherently suitable to SHM. Time series analysis has been employed to extract damage sensitive features from measured vibration data. The time series analysis algorithms aim at fitting a time series model to vibration data, and the damage features can be extracted from the constructed time series models for damage detection purpose. These algorithms make use of linear Auto-Regressive (AR) (Gul and Catbas 2009; Jayawardhana *et al.* 2015), Auto-Regressive models with eXogenous outputs (ARX) (Zhang 2007; Fasel *et al.* 2010), and/or Auto-Regressive moving average (ARMA) models (Carden and Brownjohn 2008; Bao *et al.* 2013; Fan *et al.* 2016) to provide a statistical damage detection. In general, the time series analysis algorithms have been used to diagnose either linear or nonlinear damage in structures.

---

* Corresponding author, Professor, E-mail: jiyangfu@gzhu.edu.cn

[a] Ph.D., E-mail: cechenliujie@gzhu.edu.cn

[b] Professor, E-mail: celyu@jnu.edu.cn

[d] Associate Professor, E-mail: alex.ng@adelaide.edu.au

A number of researchers have employed traditional linear time series analysis algorithms, e.g. AR, ARX and ARMA models, and developed damage sensitive features (DSFs) using the coefficients of these models, as well as the mean square deviation of the residual errors, to detect the linear damage in structures. Noh *et al.*(2009) developed time series based damage detection algorithms using both acceleration and strain data to model AR processes, and the DSF was defined using the first three AR coefficients. Carden and Brownjohn (2008) proposed a statistical classification algorithm. In their algorithm, the structure's time-series responses are fitted with ARMA models, while classifier was fed with the ARMA coefficients. Lautour and Omenzetter (2010) proposed an algorithm consisting of AR models and artificial neural network (ANN) for damage classification and estimation. The coefficients of the AR models were treated as DSF and treated as inputs of the ANN in damage detection. The performance of the algorithm was verified and evaluated using experimental data of a 3-story shear building structure from Los Alamos National Laboratory (Figueiredo *et al.* 2009). Other researchers used the residual error generated by the time series model for damage detection. Lu *et al.* (2008) used AR and ARX to determine damage in two near full-scale single-story reinforced concrete frames. The measure of damage was residual error calculated by the ARX model. Rao and Ratnam (2012) presented an AR model for health monitoring of welded structures by determining residual errors through Shewhart and exponentially weighted moving average control charts. Roy *et al.* (2015) proposed different DSFs based on ARX models, such as ARX model coefficients, Kolmogorov–Smirnov (KS) test statistical distance, and model residual error.

In addition to detecting the existence of damage, studies have focused on determining the damage location. Gul and Catbas (2011) presented two approaches to extract DSFs from ARX models. The first approach is to quantify simple ARX models and noise free data, by which the coefficients of the ARX models are directly employed as the DSF. The second approach is to use the ARX model fitting ratios as the DSF, which was successfully applied to different cases for locating and identifying the damage based numerical and experimental data under noisy condition. Zheng and Mita (2007, 2008, 2009) presented a two-stage damage detection method for detecting and locating damage using ARMA models. Two distance measures were introduced using the cepstral metric and subspace angles of ARMA models, respectively.

All the aforementioned studies for damage identification are inherently limited to linear models and ignored the nonlinearities of the structural. The structure may originally act linear but subsequently behave nonlinear because of the inchoation of damage (Prawin and Rao 2018). One of the challenges in SHM is to distinguish and categorize linear and nonlinear damage, and nonlinear damage and nonlinearities in the healthy structures (Adams and Farrar 2002). Linear damage is defined as the situation when the initially linear-elastic structure remains linear-elastic after damage. Nonlinear damage is defined as the situation when the initially linear-elastic structure behaves in a nonlinear manner after the damage has been appeared. An example of nonlinear damage is the formation of a fatigue crack that subsequently opens and closes under the normal operation environment (Doebling et al. 1996). Therefore, the nonlinear characteristics of structural response need to be taken into damage detection. Sohn et al. (2003) demonstrated that although nonlinear responses of structures have often been overlooked in the developments of SHM, they can provide useful information for damage detection.

Traditional linear time series analysis methods using AR, ARX, or ARMA models are unable to reliably det ect nonlinear damage. Because the linear time series methods assume that the residual error gained from the m odels follows normal distribution, and adopts standard statistical analysis. However, when nonlinear damage o ccurs, the distribution of the responses no longer follows the normal distribution. As a result, the damage detec tion accuracy is affected by the nonlinear effect of the responses and this may lead to false alarm in damage de tection if the nonlinearities are neglected (Fan and Yao 2006; Farrar and Lieven 2007).

In this study, two algorithms are proposed to detect nonlinear damage using the linear ARMA models. The existing residual standard deviation (RSD) employed in this study is defined as the ratio of RSD in the unknown state to that in the benchmark state. First, the cosine similarity of the DSF is proposed to improve the performance of the existing DSF in damage detection. To enhance the capability of the nonlinear damage detection, this advanced DSF is then combined with either cluster analysis based on K-means or Bayes discriminant approach to further improve the reliability of the nonlinear damage detection based on the linear model and a new damage index. To verify and compare the performance of the proposed algorithms, experimental data of a three-story shear structure is used to construct the ARMA models, and then the ARMA models with existing DSF is compared with the performance of the advanced DSF. The results demonstrate that the proposed advanced DSF proposed can effectively diagnose the nonlinear damage.

2. Two proposed algorithms based on linear ARMA model

This section describes the background of the ARMA model and a damage index. The proposed DSF and associated classification algorithms are then presented.

2.1 Description of ARMA model

The equation below gives the conditional mean in a general linear ARMA model.

$$y_t = c + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=1}^{q} \beta_j y_{t-j} + \varepsilon_t \tag{1}$$

The time series to be modeled is represented with y_t. c in the right portion is a constant. The order of the AR and moving average (MA) processes are expressed by p and q, respectively. AR and MA coefficients are expressed by α_i and β_j, respectively. And ε_t is the error term. At the right-hand side of equation (1), the first three terms are deterministic, and can be taken as the prediction of the current state based on previous observations and errors. The error term ε is a random variable that stands for the random component in the mean of y_t. Commonly, ε is taken as a variable whose mean is zero and variance is constant, i.e., (i) $E(\varepsilon_t) = 0$, (ii) $E(\varepsilon_t \varepsilon_T) = 0$ for $t \neq T$, and (iii) $E(\varepsilon_t^2) = \sigma^2$.

2.2 Definition of RSD

When any damage occurs in the structure, the previously constructed model based on the signals of the benchmark state would be unable to reproduce the time series acquired from damage conditions. The residual errors in damage or nonlinear states are assumed to be large and exhibit greater variance when it is compared to the benchmark model. Therefore, the standard deviation of the residual errors from the ARMA model can be defined as a damage factor. Chen *et al*. (2013, 2015) proposed the residual standard deviation (RSD), which is defined as the ratio of the standard deviation of residual error in the unknown condition to that in the benchmark condition and is given by

$$\mathrm{RSD} = \frac{\mathrm{std}\left(\varepsilon_t^{\mathrm{Test}}\right)}{\mathrm{std}\left(\varepsilon_t^{\mathrm{Ref}}\right)} \tag{2}$$

when RSD=1, the structure is deemed as healthy. RSD>1 means damage exists in the structure.

2.3 The proposed VSCS based on vector space cosine similarity

It is ineffective if the nonlinear damage detection only relies on the existing damage index extracted using linear models due to the loss of nonlinear damage information. In order to enhance nonlinear damage detection, the advanced VSCS is proposed in this study, which is derived from the existing RSD using cosine similarity. Cosine similarity in the vector space is defined as a weighted sum of the similarity between two high order vectors. The fundamental nature of the cosine similarity is the cosine of the angle between two vectors. Rather than weighted distance or length, cosine similarity stands for the difference in direction between two vectors Thus, cosine similarity of two vectors X and Y is given by Zhu *et al.* (2011) as follow.

$$\mathrm{sim}(X,Y) = \cos\theta = \frac{\langle X \cdot Y \rangle}{\|X\| \cdot \|Y\|} \tag{3}$$

In Eq. (3), sim is used as an abbreviation for similarity, thus sim (X, Y) represents the cosine similarity between the vectors X and Y. The numerator is the inner product of two vectors and denominator is the product of the L_2 norm of the two vectors. The cosine similarity always has a range of values from -1 to 1, where 1 means the two vectors are very similar. The value in [-1,1] means there is a certain degree of variation between the two vectors.

In this study, a new DSF is proposed and it is defined as VSCS between the RSD of benchmark condition and test condition, which is given by

$$\text{VSCS}=\frac{\sum_{k=1}^{m}\text{RSD}_{ik}^{\text{Ref}}\text{RSD}_{jk}^{\text{Test}}}{\sqrt{\left(\sum_{k=1}^{m}\left(\text{RSD}_{ik}^{\text{Ref}}\right)^2\right)\left(\sum_{k=1}^{m}\left(\text{RSD}_{jk}^{\text{Test}}\right)^2\right)}} \tag{4}$$

where $\text{RSD}_{ik}^{\text{Ref}}$ is the RSD of benchmark model from the *k*-th component of the characteristic vector in *i*-th category, and $\text{RSD}_{jk}^{\text{Test}}$ is the RSD of test model from the *k*-th component of the characteristic vector in *j*-th category.

2.4 Classification algorithms based on K-means cluster analysis and Bayes discriminant analysis

Generally, the DSF need to be used in a pattern classification framework to detect damage, which assumes the distribution of response data from the structures are normal. However, structural damage affects dynamic properties of structures, resulting a change of the extreme values in the data (Fan and Yao 2006). Thus, the assumption of normality imposed may lead to improper damage detection, especially when the damage is nonlinear. Two approaches are proposed in this study. The first approach is to use K-means cluster analysis. The second approach is to use Bayes discriminant analysis, which is based on the value of the posterior probability to distinguish between health and damage state. The following two sub-sections describe the details of K-means cluster analysis and Bayes discriminant analysis.

2.4.1 K-means cluster analysis

K-means cluster analysis algorithm was originally introduced by McQueen (1967), which has been one of the most popular and widely used cluster analysis methods. The basic idea of K-means cluster analysis algorithm is to group similar data points together and determine the underlying patterns. The word “means” in the name of the algorithm refers to averaging of the data, i.e. finding the centroid. In the K-means cluster analysis algorithm, K number of centroids is first identified. Each individual data point is then allocated to the nearest cluster according to certain similarity measure standard, while keeping the centroids as small as possible. The centers are re-identified as centers of mass of their assigned points. This process is repeated until it is stabilized or maximum number of iterations is reached. The aim of the K-means cluster analysis is to partition the *m* data in multivariate data set into K clusters, where each data in the dataset is assigned to a specific cluster. K-means cluster analysis is a hard-partitioning algorithm and an iterative process.

Firstly, data are assigned to groups. After calculating the mean of each group, the data is assigned by allocating each datum to its nearest means cluster position (Weatherill and Burton 2009; Novianti *et al.* 2017). The process is summarized as below.

1) The K-means method aims to determine the cluster centers $(\lambda_1,\lambda_2,\cdots,\lambda_k)$ to minimize the sum of the squared distances, i.e. distortion, of each data point (x_m) to its nearest cluster center (λ_k) as

$$\min\sum_{i=1}^{K}\sum_{x_m\in C_i}\left\|x_m-\lambda_i\right\|^2 \tag{5}$$

where λ_i is the center of the cluster C_i, M is the number of objects in the cluster C_i, x_m is the *m*-th object of the *i*-th cluster.

2) To reduce the squared error, the database is put into the cluster whom represented by the nearest centroid. An instance $x_m\in C_t$ in the relocation step can change its cluster membership $x_m\in C_k$, if $\left\|x_m-\lambda_k\right\|\le\left\|x_m-\lambda_j\right\|$ for all $j=1,2,\cdots m, j\neq t$.

3) The centroids of the cluster C_t and C_k, and the squared error is recomputed. The entire process is continuously repeated until no further reduction can be achieved for the squared error, when its cluster membership cannot be further changed by any instance.

In the subsequent clustering analysis, K-means cluster will be chosen. The similarity of the characteristic vectors of benchmark and test states will be calculated compared with the center of each category (one cluster representing damage and the other cluster representing health), then each test state will be assigned to the category with the highest similarity.

2.4.2 Bayesian discrimination

Bayesian methods take into account the structural prior information including historical data or experience of expert and the measured data of the structure comprehensively, and determine the posterior probability distribution of the structure parameter based on the optimal probability model (Yuen 2010; Xin et al. 2019). This means it allows to determine the probability by combining expectation based on previous experience (prior probability) with information from measured data. The advantages of Bayesian methods are that they fully utilize the prior information and update the probability distribution (posterior probability) of structural parameters based to the measured data. Finally, the condition of the structure can be judged based on the posterior probability distribution of the structural parameters. The executive process of Bayesian methods is consistent with the ideology of on-line structural health monitoring.

As the core of Bayesian theorem, the Bayes formula can be expressed as:

$$P\left(w_i|X\right)=\frac{p(X,w_i)}{p(X)}=\frac{p(X|w_i)P(w_i)}{p(X)}, \quad i=1,2\cdots c \tag{6}$$

where $P\left(w_i|X\right)$ is the posterior probability of w_i under the condition of X. The parameter w_i is defined as a random variable represented for the condition of the structure. X is the measured data for observations. $P(w_i)$ is the prior probability known or artificial hypothesis. $p(X|w_i)$ is the conditional probability of X under the condition of w_i, e.g., the probability of the observations of X fall into the *i*-th cluster. The overall density is represented by $p(X)$. The joint probability density of X and w_i is expressed as $p(X,w_i)$.

Following the Eq. (6), the essence of Bayesian inference is that under the conditions of parameter w_i, $P(w_i)$ is revised continuously according to the measured data X and their conditional probability $p(X|w_i)$, and finally to get the estimated value of $P\left(w_i|X\right)$. Thus, Bayesian decision can assign the observations to their clusters of the highest posterior probability based on Bayesian formula, by which we can keep the overall error rate minimum when the conditional probability and priori probability are obtained (Zhang 2009). For instance, decision criterion of the cluster problem for structural damage detection is described in Eq. (7) as follow:

$$\text{If } p(X|w_1)P(w_1)>p(X|w_2)P(w_2)\text{, then } X\in w_1\text{, otherwise, } X\in w_2 \tag{7}$$

where X is the test state, w_1 is the cluster “Healthy”, w_2 is the cluster “Damaged”.

2.5 Discussion and comparison of the two proposed algorithms

When a structure is healthy, the time-domain responses under normal operating conditions are generally expected to follow stationary random process and the performance of structural characteristic is almost linear. Thus, the responses can be constructed as stable ARMA models, while the prediction error from ARMA model follows the white noise distribution. When the damage exists in the structure, the structure may still behave linearly with only change of geometric dimensions, but the structure may also demonstrate nonlinear characteristics with structural response being non-stationary and nonlinear, e.g. the cracks open and close under the loading condition. If the benchmark state ARMA model is used to predict the response signals that contain nonlinearities, the prediction errors will increase. The prediction errors may follow colored noise with nonlinear characteristics rather than white noise distribution (Zhu and Yu 2012; Wang 2013).

Colored noise can be approximated by linear regression using white noise, which means that the prediction error from ARMA model can be expressed as a linear combination of white noises, i.e., the MA model. The process can only ensure the prediction error of the benchmark state follows the white noise distribution, but it cannot guarantee the prediction error of damage state follows the white noise distribution. Carden and Brownjohn (2008) used ARMA model coefficients as damage features. Although AR coefficient can reflect the linear features of structures when the AR model residuals follows colored noise distribution, the MA coefficient may reflect some other characteristics. Unfortunately, the ARMA linear models are not applicable to nonlinear phenomena (Fan and Yao 2006).

Mean and variance values are common statistical description of the white noise. Equation (4) expresses the DSF based on the statistical properties of white noise in residual signal. On the contrary, it could not reflect the structural characteristics when the residual signal does not obey the white noise distribution. Therefore, the traditional method can effectively detect the structural linear damage. But if there is nonlinear damage, the traditional method may misdiagnose due to ignoring the nonlinear characteristics of the responses and leakage of nonlinear damage information.

To improve the performance of nonlinear damage detection using linear model, the proposed method is based on linear ARMA model and proposed damage index VSCS. Then it is classified by the K-means cluster analysis or Bayes discriminant analysis, respectively. Vector space cosine similarity measures the difference between the two individuals by comparing the cosine value of the angles between two vectors, and paying more attention to the difference of the two vectors in the direction rather than the distance or length.

3. Identification of structural nonlinear damage using ARMA model

This section presents a process for constructing the ARMA model based on the time series data of the acceleration response and applies the proposed DSF to identify the structure damage. The procedure using the ARMA model to diagnose the structural nonlinear damage is summarized below, and the procedure is shown in Fig 1.

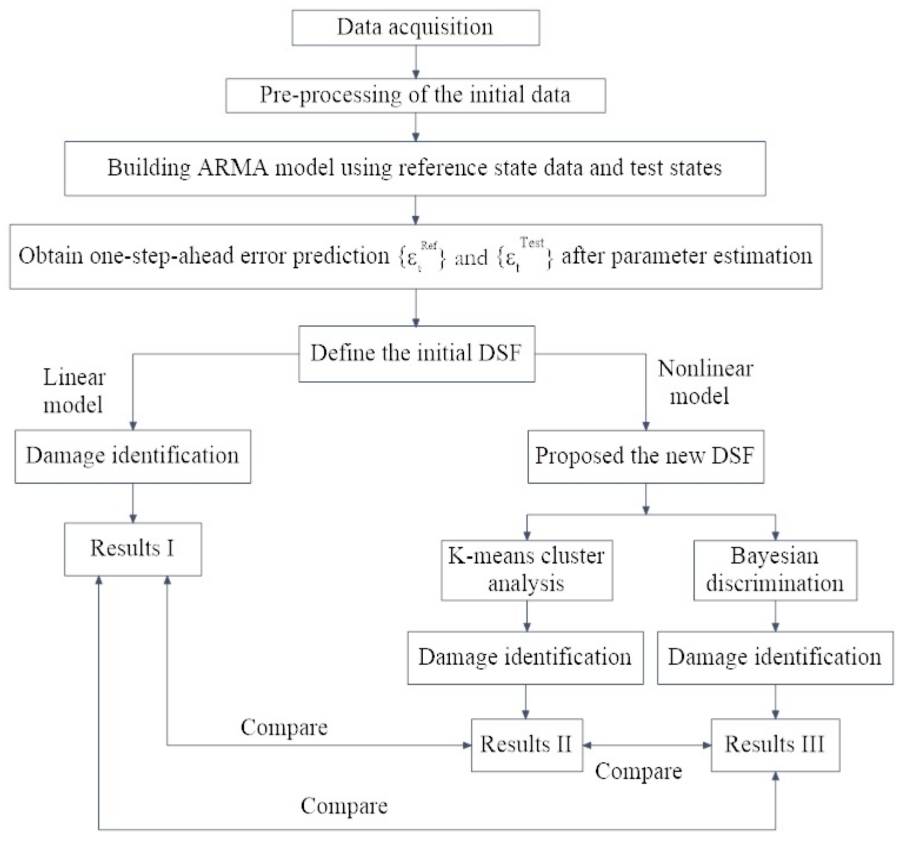


Fig. 1 Procedure of damage identificaiton using proposed method

First, the ARMA model is built using the acceleration responses of the structure benchmark state. To eliminate the influence of environmental factors on the amplitude of the response data, all data are standardized as below:

$$x_{ij}(t) = \frac{X_{ij}(t) - \mu_{ij}}{\sigma_{ij}} \tag{8}$$

where μ_{ij} and σ_{ij} is the mean and standard deviation of the *j*-th stream of sensor *i*, respectively. $X(t)$ is the original time series data and $x(t)$ is time series data after normalization. This standardization procedure is applied to all the response data in this paper.

Once the pre-processing of the initial data is completed, the optimal ARMA order can be determined using the Akaike Information Criteria (AIC) (Chen and Yu 2013), and its parameters are estimated based on the prediction error method in The Mathworks (2014). ARMA(14,15) in Chen and Yu (2013), which was chosen based on the AIC plots for the same damage detection problem, is adopted in this study. After that, substituting this into the benchmark ARMA model using standardized y_t obtained from unknown state according to Eq. (1), we can get

$$\varepsilon_t = y_t - \left(c + \sum_{i=1}^{R} AR_i y_{t-i} + \sum_{j=1}^{M} MA_j y_{t-j} \right) \tag{9}$$

If the test state is a healthy state, time series data y_t from the unknown state will satisfy the benchmark model, and hence, there is no significant difference between $\varepsilon_t^{\text{Ref}}$ and $\varepsilon_t^{\text{Test}}$. However, y_t cannot satisfy the benchmark model when the unknown state has damage in the structure, which leads to significant difference between $\varepsilon_t^{\text{Ref}}$ and $\varepsilon_t^{\text{Test}}$.

4. Application to three-story shear building structure

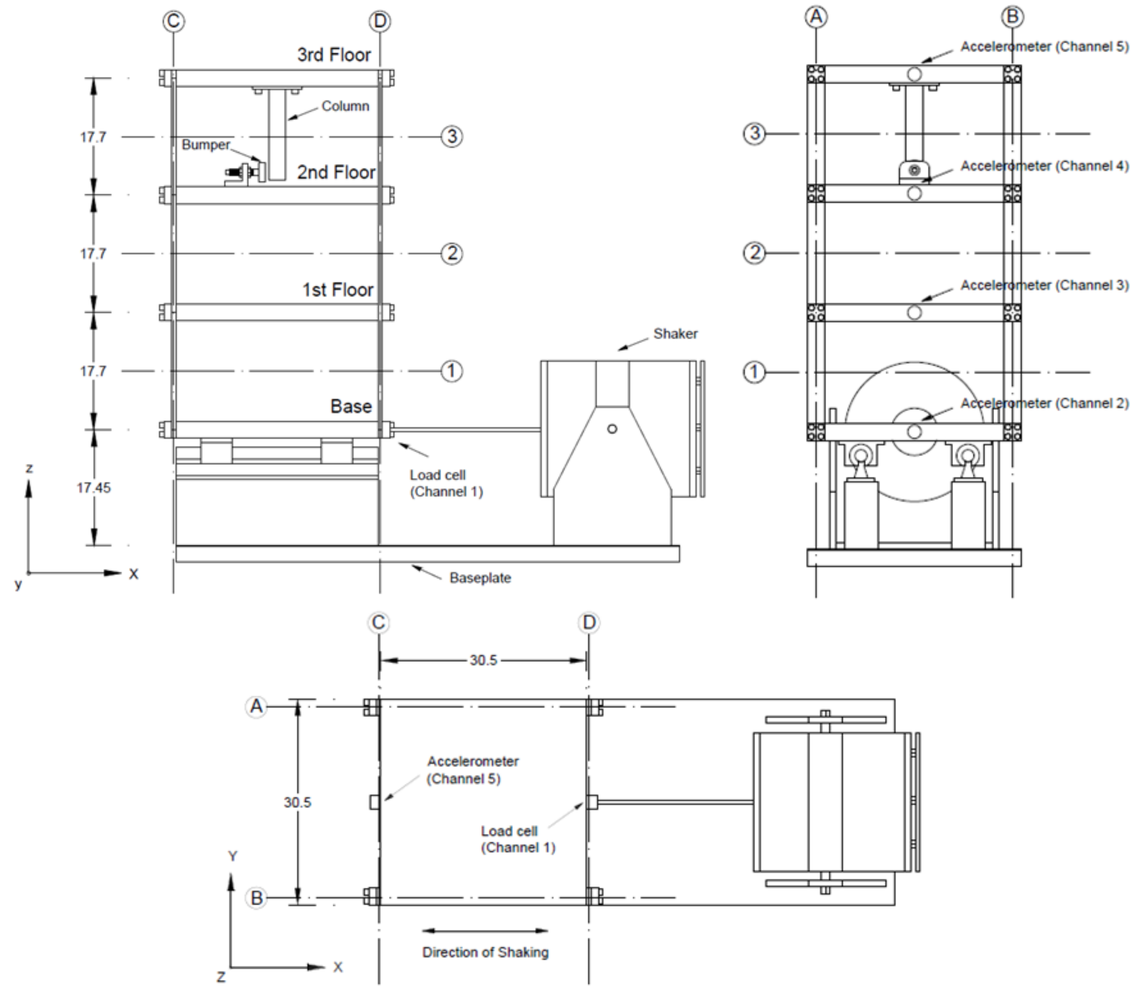
In this section, the experimental data of a three-story shear building from Los Alamos National Laboratory (Figueiredo *et al.* 2009) is employed to validate the proposed methods.

4.1 Description of the shear building structure

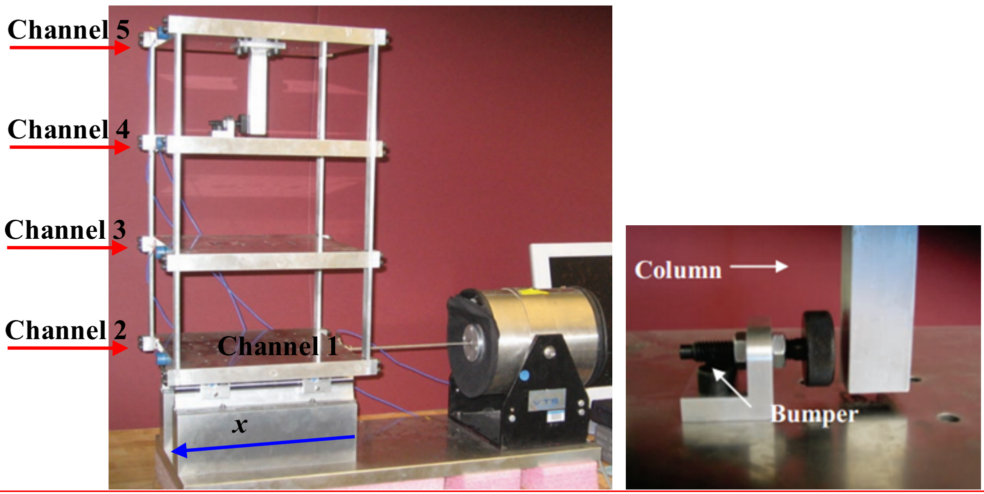
The three-story shear building structure shown in Fig. 2 is composed of aluminum columns and plates assembled by bolted joints, which only allow sliding on rails in the *x*-direction. A center column (15.0 × 2.5 × 2.5 cm^3) is suspended from the top floor when it contacts a bumper mounted on the next floor, and this is the source of nonlinear damage. The bumper's position can be adjusted to vary the extent of impact for a particular excitation level. The purpose of this is to simulate the effect of fatigue crack that opens and closes under operational and environmental conditions. (Figueiredo et al. 2009). The environmental and operational uncertainties were simulated by reducing stiffness and adding mass at several locations of the shear building structure. The time domain data of force and acceleration are recorded.

In the experiment, ten tests were carried out for each case in order to take into account the variability of the measured data. A Hanning window was employed in the time-domain data for the purpose of leakage reduction and five averages were used to decrease the influence of random effect. The measured real-world data always contains measurement noise that can obscure the actual state condition of the structure. The reliability of identification results may be influenced if the real data measured from the structures contains measurement noise. Thus, the robustness of the proposed method against the measurement noise effect is very important (Ding et al. 2019).

As listed in Table.1, various structural conditions are considered in this study and they are classified into four groups. The first group (Stated #1) is taken as the baseline condition. States #2-#9 can be taken as the second group, in which different mass and stiffness conditions were tested to simulate the variation of environmental and operational condition. An example can be found in Fig.1a, "Sated #4" in Table. 1 stands for the case that the stiffness of the columns located between the base and 1st floor was reduced to 87.5% of its original value. Sated #4 is abbreviated as 1BD since these columns are at the intersection of planes B and D, and they can be abbreviated in similar way. To perform the stiffness reduction, cross-section thickness of the column in the direction of excitation is reduced to half. Nonlinear damage States #10-#14 formed the third group. Nonlinearities was introduced into this group by using a bumper and a suspended column. The nonlinearity is variable since gaps between the bumper and suspended column is changeable. Other than the nonlinear damage as in the third group, the fourth group (States #15-#17) additionally involves variation of the mass and stiffness, as in the second group.



(a)



(b)

Fig. 2 (a) Details of the shear building structure and (b) experiment setup (Figueiredo *et al.* 2009)

Table 1 Cases and their corresponding structural state conditions (Figueiredo *et al.* 2009)

	State#	**Cases**	**Condition**	**Configuration**				
				Perturbation			**Damage**	
				Content	**Magnitude**	**Location**	**Content**	**Magnitude**
Benchmark	1	1-9	Healthy	/	/	/	/	/
Operating conditions	2	10-19	Healthy	Mass	1.2 kg	Base	/	/
	3	20-29	Healthy	Mass	1.2 kg	1st floor	/	/
	4	30-39	Healthy	Stiffness	-87.50%	Column 1BD	/	/
	5	40-49	Healthy	Stiffness	-87.50%	Column 1AD and 1BD	/	/
	6	50-59	Healthy	Stiffness	-87.50%	Column 2BD	/	/
	7	60-69	Healthy	Stiffness	-87.50%	Column 2AD and 2BD	/	/
	8	70-79	Healthy	Stiffness	-87.50%	Column 3BD	/	/
	9	80-89	Healthy	Stiffness	-87.50%	Column 3AD and 3BD	/	/
Damage conditions	10	90-99	Damaged	/	/	/	Gap	0.20 mm
	11	100-109	Damaged	/	/	/	Gap	0.15 mm
	12	110-119	Damaged	/	/	/	Gap	0.13 mm
	13	120-129	Damaged	/	/	/	Gap	0.10 mm
	14	130-139	Damaged	/	/	/	Gap	0.05 mm
Operating+ Damaged	15	140-149	Damaged	mass	1.2 kg	Base	Gap	0.20 mm
	16	150-159	Damaged	mass	1.2 kg	1st floor	Gap	0.20 mm
	17	160-169	Damaged	mass	1.2 kg	1st floor	Gap	0.10 mm

4.2 Damage detection results of the proposed DSF based on ARMA model

In this section, the results using the existing DSF and the proposed DSF based on ARMA model are compared and discussed.

4.2.1 Results of nonlinear damage detection based on linear ARMA model and the existing RSD

Fig. 2 shows the existing RSD based on linear ARMA model for nonlinear damage detection in the three-story shear building structure. When RSD=1, the test state is determined as a healthy condition. When RSD>1, indicating the structure is in damage condition. The following conclusions are drawn from the results presented in Fig. 3.

1) The location of the damaged potion is determined between Channel 4 and Channel 5. State #1-#9 are identified as health states, while States #10-#17 are assumed to be damage states.

2) States #5 and States #9 are obviously misjudged in Channel 2 for health states; States #6 and States #7 are obviously misjudged in Channel 3 for health states, and four significant misjudgments (States #13-#14, States #15 and States #17) for damage states. It indicates that the RSD based on this algorithm are sensitive to environmental and operational condition changes.

3) State #10 and State #16 are easily misjudged in Channels 4 and 5 for damage states. The results show that this method and RSD are not sensitive to small damage.

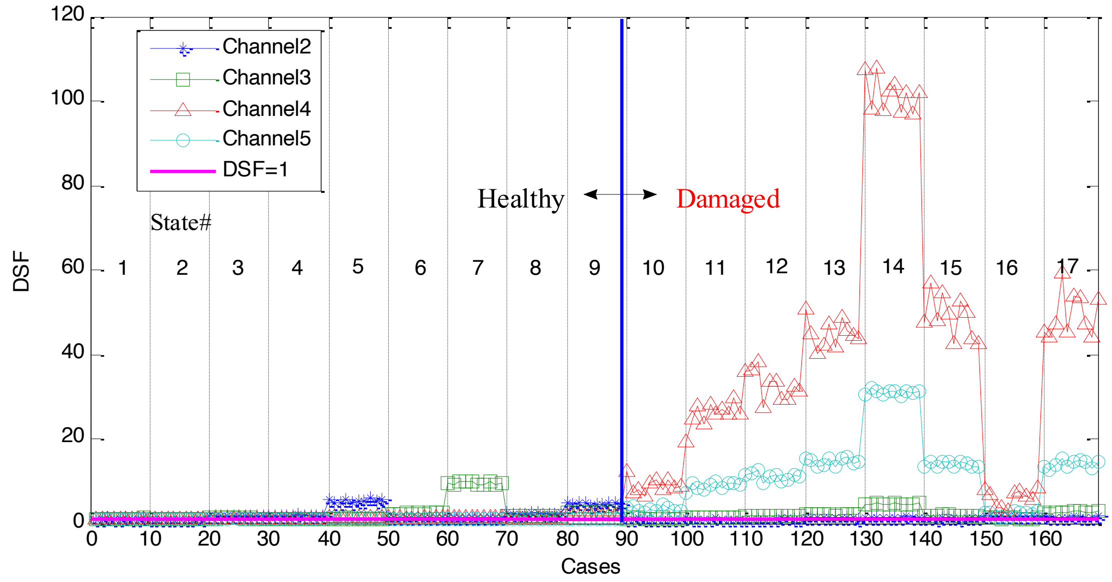


Fig. 3 The existing RSD based on linear ARMA model for States #1-#17 from Channels 2 – 5.

4.2.2. Approach I: Improvement of nonlinear damage detection by combining the proposed VSCS and K-means cluster analysis

Taking the advantage of K-means cluster analysis, the RSD's cosine similarity in vector space is computed and compared, subsequently. Fig. 4 illustrates the cosine similarity of test case for each cluster using K-means cluster analysis. As can be seen, in Fig. 4, values of similarity all fall between -1 and 1. It's worthy mention that the algorithm used in K-means cluster analysis sorts the structural states with data set. And, in the K-means cluster analysis, all data are configured in groups. This mechanism benefits the classification by eliminate the necessity of providing the threshold value to distinguish between healthy and damaged states.

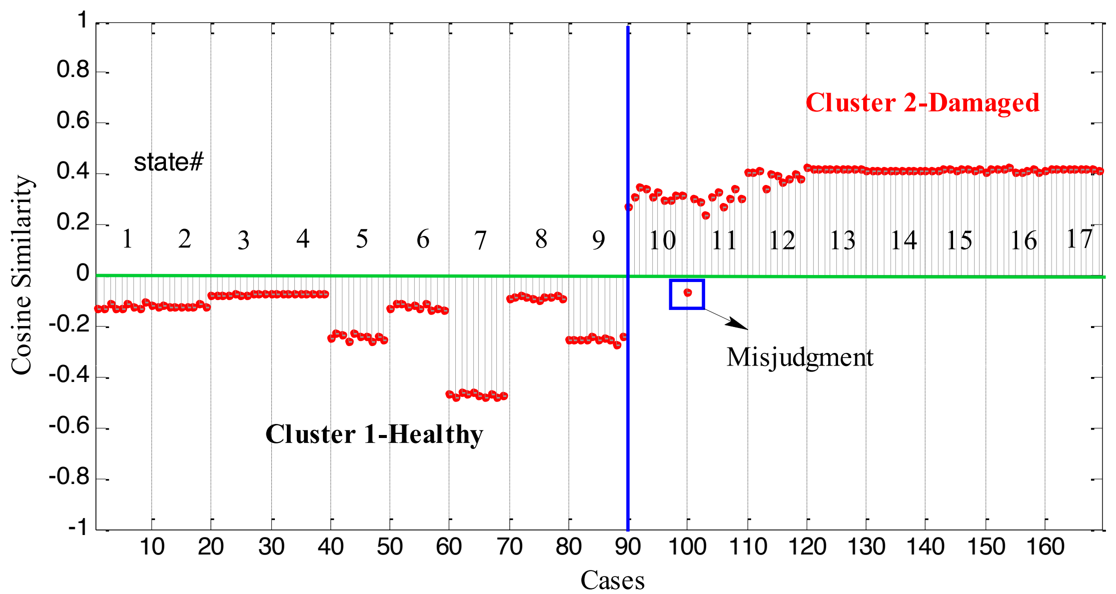


Fig. 4 The cosine similarity of test case for each cluster using K-means cluster analysis

Table 2 Damage identification results with the cosine similarity

State / Damage Index	#1	#2	#3	#4	#5	#6	#7	#8	#9	False Positive
Cluster 1 (*/#)	0/9	0/10	0/10	0/10	0/10	0/10	0/10	0/10	0/10	0/89
State / Damage Index	**#10**	**#11**	**#12**	**#13**	**#14**	**#15**	**#16**	**#17**	**/**	**False Negative**
Cluster 2 (*/#)	0/10	1/10	0/10	0/10	0/10	0/10	0/10	0/10	/	1/80

Note:
False Positive manifests misjudged health state from damage.
False Negative manifests misjudged damage state from health.
*/# manifests that there are * cases misjudgment among # cases.

The procedure, which is based on the ratio of the standard deviation of residual errors for ARMA model due to the loss of nonlinear damage information, has lower reliability in nonlinear damage detection. The proposed VSCS on account of linear time series analysis combined cosine similarity with K-means cluster analysis improves the reliability of the nonlinear damage detection and this is evidenced by only one case of State #11 is misjudged.

4.2.3 Approach II: Improvement of nonlinear damage detection by combining the advanced DSF and Bayesian discrimination

The analysis results in Section 4.2.2 uses 90 as condensation point. The data of the advanced VSCS fall into two categories, such as "Healthy" and "Damaged" using the Bayes discriminant analysis. The obtained sort results are listed in Table 3.

Table 3 Square matrix (CLMat) of two categories

Observations / Predicted results	State #1-State #9	State #10- State #17
State #1-State #9	89	0
State #10- State #17	0	80
Condition	**Healthy**	**Damaged**

As can be noticed from above, CLMat is a square matrix whose size equals the number of categories. This is a count of observations known to be in category *i* but it is predicted to be in category *j*. Diagonal elements of CLMat is the correct classified categories number. Diagonal elements in CLMat stands for the correct classified categories number. As can be seen in Table.1, health states are numbered into Cases 1–89 (States #1-#9), while damage states are numbered into 90–169 (States #10-#17).The classification results show that all discriminations for the structure states are correctly determined using the Bayesian discrimination. Table 4 shows the results that classify every state into a category.

Table 4 Square matrix (CLMat) with size equal to the total number of categories

Observations / Predicted results	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17
#1	***8***	***1***	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#2	***2***	***8***	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#3	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#4	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0	0	0
#5	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0	0
#6	0	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0
#7	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0

#8	0	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0
#9	0	0	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0	0
#10	0	0	0	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0
#11	0	0	0	0	0	0	0	0	0	0	**10**	0	0	0	0	0	0
#12	0	0	0	0	0	0	0	0	0	0	***1***	***9***	0	0	0	0	0
#13	0	0	0	0	0	0	0	0	0	0	0	0	***8***	0	***2***	0	0
#14	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**	0	0	0
#15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**	0	0
#16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**	0
#17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**
Condition	**Healthy**									**Damaged**							

Note:
#1-#17 represent State #1-State #17.

States#1-#9 are "health" and States #10-#17 are "damage". It shows that one case of State #1 is incorrectly sort into State #2; two cases in State #2 are incorrectly sort into State #1; one case in State #12 is incorrectly sort into State #11; and two cases in State#13 are incorrectly sort into State #15. But the cases from State #1-#9 are all classified as "health" and the cases from State #10-#17 are all classified as "damage" correctly.

According to Bayesian decision theory, when the class conditional density and priori probability are obtained (or estimated), it can assign the sample to its cluster with the highest posterior probability. Thus, we can take the posterior probability distribution into consideration.

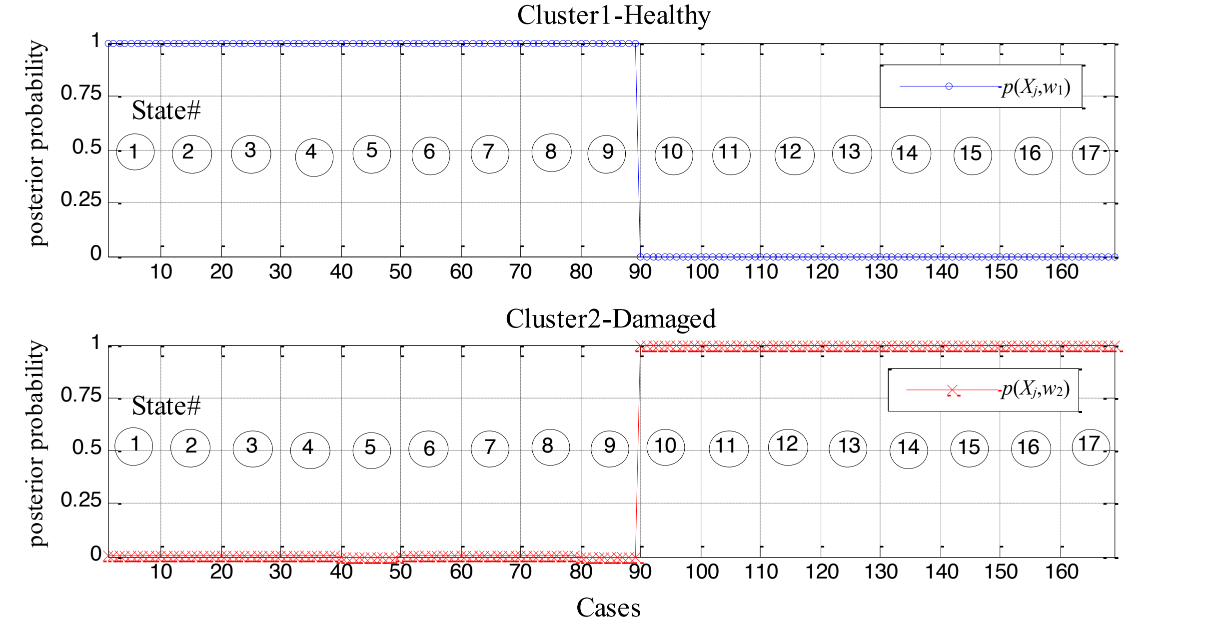


Fig. 5 Posterior probability distribution of each category based on Bayesian discrimination

It should be noted that $p(X_j, w_1)$ indicates the posterior probability of *j*-th test sample belonging to Category 1 (health) and $p(X_j, w_2)$ indicates the posterior probability of *j*-th test sample belonging to Category 2 (damage). The *j*-th sample will be grouped into the category, in which the *j*-th sample carries the highest posterior probability. It is shown in Fig. 5 that $p(X_j, w_1) >> p(X_j, w_2),\ (j = 1, 2, \cdots, 89)$, $p(X_j, w_2) >> p(X_j, w_1),\ (j = 90, 91, \cdots, 169)$ and $p = [0, 1]$.

As mentioned above, combining the VSCS with K-means cluster analysis and Bayes discriminant analysis provides effective approach for nonlinear damage estimation and classification, meanwhile reduces the computational cost. Only one case of State #11 is misjudged using Approach I, and all cases are correctly classified by using Approach II. This study has analyzed the existing damage index RSD based on linear

ARMA model, and has demonstrated that the reliability of the nonlinear identification results is influenced by the leakage of the damage information. Thus, this study has developed the advanced VSCS and classified algorithms, K-means cluster analysis or Bayesian discrimination, based on linear ARMA model. The results have shown that the proposed approaches improve the efficiency and reliability in identifying nonlinear damage over the existing methods in the literature.

5 Conclusions

This study has presented two algorithms for improving nonlinear damage detection using linear ARMA model. As an improved detection methodology of nonlinear damage, DSF is extracted using linear time series analysis. An approach that combines cosine similarity with K-means cluster analysis and Bayes discriminant analysis has been proposed in this study. The performance of the algorithms has been verified and evaluated using the experimental data of a three-story shear building structure. The current study demonstrated that by combine linear ARMA model and the advanced DSF with cluster analysis or Bayes discriminant analysis, effective approaches can thus be formed for damage detection in nonlinear situation. Furthermore, the accuracy is improved and the computational cost is reduced in the proposed two approaches.

Main advantages of the proposed two approaches are that no sophisticated finite element is required, and the complicated nonlinear damage detection can be complicated with simple linear time series model. Knowledge gained from the two approaches and with VSCS is that, rather than the distance or length, majority of the nonlinear damage information lies in direction of the feature vectors. When Bayesian discrimination is utilized to calculate the posterior probability of the structural parameters from different state conditions, the results of this algorithm have indicated that the information of the structural states can be distinguished through the posterior probability. The proposed algorithms have to implement the structural damage identification by evading extreme values that are related to imposed noise or singular values, which indicates the methodologies put forward in this paper has better robustness to noise.

The study has mainly focused on determining damage existence. Further work can extend the proposed methods to determine the location and severity of the damage.

Acknowledgement

The research described in this paper was financially supported by the open project foundation (Grant No. 20160626005) from MOE Key Lab of Disaster Forecast and Control in Engineering at Jinan University in China. In addition, the first author expresses her personal gratitude to A.P. Cui Z. of Institute of Rock and Soil Mechanics, Chinese Academy of Sciences, for his continuous support and proofreading work for the current study.

References

Adams, D.E. and Farrar, C.R. (2002), "Classifying linear and nonlinear structural damage using frequency domain ARX models", *Struct. Health Monit*, **1**(2), 185-201. https://dx.doi.org/10.1106/147592102028970.

Bao, C.X., Hao, H. and Li, Z.X. (2013), "Integrated ARMA model method for damage detection of subsea pipeline system", *Eng Struct*, **48**, 176-192. https://doi.org/10.1016/j.engstruct.2012.09.033.

Carden, E.P and Brownjohn, J.M.W. (2008), "ARMA modelled time-series classification for structural health monitoring of civil infrastructure", *Mech Syst Signal Pr*, 22(2), 295-314. https://doi.org/10.1016/j.ymssp.2007.07.003.

Chen, L.J. and Yu, L. (2013), "Structural nonlinear damage identification algorithm based on time series ARMA/GARCH model", *Adv in Struct Eng*, **16**(9), 1597-1609. https://doi.org/10.1260%2F1369-4332.16.9.1597.

Chen, L.J., Yu, L. and Xiao, T. (2015), "The structural nonlinear damage detection based on linear time series algorithm", *Appl Mech Mater*, **744-746**, 345-350. https://doi.org/10.4028/www.scientific.net/AMM.744-746.345.

Ding, Z.H., Li, J., Hao, H. and Lu, Z.R. (2019), "Nonlinear hysteretic parameter identification using an improved tree-seed algorithm", *Swarm Evol Comput*, **46**, 69-83. https://doi.org/10.1016/j.swevo.2019.02.005.

Doebling, S.W., Farrar, C.R., Prime, M.B. and Shevitz, D.W. (1996), *Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in Their Vibration Characteristics: A Literature Review*, Los Alamos National Lab., NM, USA, May.

Fan, J.Q. and Yao, Q.W. (2006), *Nonlinear Time Series:Nonparametric and Parametric Methods*, Springer-Verlag, Germany.

Fan, X., Li, J. and Hao, H. (2016), "Piezoelectric impedance based damage detection in truss bridges based on time frequency ARMA model", *Smart Struct Syst*, **18**(3), 501-523. https://doi.org/10.12989/sss.2016.18.3.501.

Farrar, C.R. and Lieven, N.A.J. (2007), "Damage prognosis: the future of structural health monitoring", *Philosophical Transactions of the Royal Society A: Mathematical, Phys Eng Sci*, **365**(1851), 623-632. https://doi.org/10.1098/rsta.2006.1927.

Fasel, T.R., Hoon, S., Gyuhae, P. and Farrar, C.R. (2010), "Active sensing using impedance-based ARX models and extreme value statistics for damage detection", *Earthq Eng Struct D*, **34**(7), 763-785. https://xs.scihub.ltd/https://doi.org/10.1002/eqe.454.

Figueiredo, E., Park, G., Figueiras, J., Farrar, C. and Worden, K. (2009), *Structural health monitoring algorithm comparisons using standard data sets*, Los Alamos National Laboratory (LANL), Los Alamos, NM, United States. https://doi.org/10.2172/961604.

Gul, M. and Catbas, F.N. (2009), "Statistical pattern recognition for Structural Health Monitoring using time series modeling: Theory and experimental verifications", *Mech Syst Signal Pr*, **23**(7), 2192-2204. https://doi.org/10.1016/j.ymssp.2009.02.013.

Gul, M. and Catbas, F.N. (2011), "Structural health monitoring and damage assessment using a novel time seri es analysis methodology with sensor clustering", *J Sound Vib*, **330**(6), 1196-1210. https://doi.org/10.1016/j. jsv.2010.09.024.

Jayawardhana, M., Zhu, X., Liyanapathirana, R. and Gunawardana, U. (2015), "Statistical damage sensitive feature for structural damage detection using AR model coefficients", *Adv in Struct Eng*, **18**(10), 1551-1562. https://doi.org/10.1260%2F1369-4332.18.10.1551.

Lam, H.F., Alabi, S.A. and Yang, J.H. (2017), "Identification of rail-sleeper-ballast system through time-doma in Markov chain Monte Carlo-based Bayesian approach", *Eng Struct*, **140**, 421-436. https://doi.org/10.1016/ j.engstruct.2017.03.001.

Lautour, O.D. and Omenzetter, P. (2010), "Damage classification and estimation in experimental structures using time series analysis and pattern recognition", *Mech Syst Signal Pr*, **24**(5), 1556-1569. https://doi.org/10.1016/j.ymssp.2009.12.008.

Lu, K.C., Loh, C.H., Yang, Y.S., Lynch, J.P. and Law, K.H. (2008), "Real-time structural damage detection using wireless sensing and monitoring system", *Smart Struct Syst*, 4(6), 759-778. https://doi.org/ 10.12989 /sss.2008.4.6.759.

McQueen, J.B. (1967), "Some methods of classification and analysis of multivariate observations", *Proceedings. of the 5th Berkeley Symposium on Mathematical Statistics & Probability*,. Berkeley, USA.

Ng, CT and Au, S.K. (2018), "Mode shape scaling and implications in modal identification with known input", *Eng Struct*, **156**, 411-416. https://xs.scihub.ltd/https://doi.org/10.1016/j.engstruct.2017.11.017.

Noh, H.Y., Nair, K.K., Kiremidjian, A.S. and Loh, C.H. (2009), "Application of time series based damage detection algorithms to the benchmark experiment at the National Center for Research on Earthquake Engineering (NCREE) in Taipei, Taiwan", *Smart Struct Syst*, 5(1), 95-117. https://doi.org/10.12989/ sss. 2009.5.1.095.

Novianti, P., Setyorini, D. and Rafflesia, U. (2017), "K-Means cluster analysis in earthquake epicenter clustering", *Int J Adv Intell Informatics*, 3(2), 81-89. https://doi.org/10.26555/ijain.v3i2.100.

Prawin, J. and Rao, A. R. M. (2018), "Nonlinear Structural Damage Detection Based on Adaptive Volterra Filter Model", *Int J Struct Stab Dyn*, **18**(2), 1-12. http://dx.doi.org/10.1142/S0219455418710037.

Rao, P.S. and Ratnam, C. (2012), "Health monitoring of welded structures using statistical process control", *Mech Syst Signal Pr*, **27**(1), 683-695. https://xs.scihub.ltd/https://doi.org/10.1016/j.ymssp.2011.09.023.

Roy, K., Bhattacharya, B. and Ray-Chaudhuri, S. (2015), "ARX model-based damage sensitive features for structural damage localization using output-only measurements", *J Sound Vib*, **349**, 99-122. https://xs.scihub.ltd/https://doi.org/10.1016/j.jsv.2015.03.038

Sohn, H., Farrar, C.R., Hemez, F.M., Shunk, D.D., Stinemates, D.W. and Nadler, B.R. (2003), *A review of structural health monitoring literature: 1996-2001*, Los Alamos National Laboratory Report, New Mexico.

The Mathworks, I. (2014), "MATLAB (R2015a), Econometrics Toolbox".

Wang X.M. (2013), *The Study on Movement Characteristics and Non-linear Model of CGCS2000 Framework*, Chinese Academy of Surveying and Mapping, Beijing, China.

Weatherill, G. and Burton, P.W. (2009), "Delineation of shallow seismic source zones using K-means cluster analysis, with application to the Aegean region", *Geophys J Int*, **176**(2), 565-588. https://doi.org/10.1111/j.1365-246X.2008.03997.x.

Worden, K., Farrar, C.R., Manson, G. and Park, G. (2007), "The fundamental axioms of structural health monitoring", *Pro Math Phy Eng Sci*, **463**(2082), 1639-1664. https://doi.org/10.1098/rspa.2007.1834.

Xin, Y., Hao, H., Li, J., Wang, Z.C., Wan, H.P. and Ren, W.X. (2019), "Bayesian based nonlinear model updating using instantaneous characteristics of structural dynamic responses", *Eng Struct*, **183**, 459-474. http://doi.org/10.1016/j.engstruct.2019.01.043.

Yin, T., Jiang, Q.H. and Yuen K.V. (2017), "Vibration-based damage detection for structural connections using incomplete modal data by Bayesian approach and model reduction technique", *Eng Struct*, **132**, 260-277. https://xs.scihub.ltd/https://doi.org/10.1016/j.engstruct.2016.11.035.

Yuen, K.V. (2010), *Bayesian Methods for Structural Dynamics and Civil Engineering*, John Wiley & Sons Pte Ltd., Singapore.

Zhang, Q.W. (2007), "Statistical damage identification for bridges using ambient vibration data", *Comput Struct*, **85**(7-8), 476-485. https://xs.scihub.ltd/https://doi.org/10.1016/j.compstruc.2006.08.071.

Zhang, X. (2009), *Pattern Recognition*, Tsinghua University Press, Beijing, China.

Zheng, H. and Mita, A. (2007), "Two-stage damage diagnosis based on the distance between ARMA models and pre-whitening filters", *Smart Mater Struct*, **16**(5), 1829-1836. http://dx.doi.org/10.1088/0964-1726/16/5/038

Zheng, H.T. and Mita, A. (2008), "Damage indicator defined as the distance between ARMA models for structural health monitoring", *Struct Control Hlth*, **15**(7), 992-1005. https://doi.org/10.1002/stc.235.

Zheng, H.T. and Mita, A. (2009), "Localized damage detection of structures subject to multiple ambient excitations using two distance measures for autoregressive models", *Struct. Health Monit*, **8**(3), 207-222. https://doi.org/10.1177%2F1475921708102145.

Zhu, J.H. and Yu L. (2012), "Damage detection based on time series analysis and higher statistical moments", *J Southeast U: Nat Sci Ed*, **42**(1), 137-143.

Zhu, S., Wu, J., Xiong, H. and Xia, G. (2011), "Scaling up top-K cosine similarity search", *Data Knowl Eng*, **70**(1), 60-83. https://xs.scihub.ltd/https://doi.org/10.1016/j.datak.2010.08.004