Journal of Southeast University (English Edition) Vol. 23, No. 1, pp. 86 – 91 Mar. 2007 ISSN 1003—7985

Feature extraction and damage alarming using time series analysis

Liu Yi Li Aiqun Fei Qingguo Ding Youliang

(Key Laboratory of Concrete and Prestressed Concrete Structure of Ministry of Education, Southeast University, Nanjing 210096, China)

Abstract: Aiming at the problem of on-line damage diagnosis in structural health monitoring (SHM), an algorithm of feature extraction and damage alarming based on auto-regressive moving-average (ARMA) time series analysis is presented. The monitoring data were first modeled as ARMA models, while a principal-component matrix derived from the AR coefficients of these models was utilized to establish the Mahalanobis-distance criterion functions. Then, a new damage-sensitive feature index D_{DSF} is proposed. A hypothesis test involving the t-test method is further applied to obtain a decision of damage alarming as the mean value of D_{DSF} had significantly changed after damage. The numerical results of a three-span-girder model shows that the defined index is sensitive to subtle structural damage, and the proposed algorithm can be applied to the on-line damage alarming in SHM.

Key words: feature extraction; damage alarming; time series analysis; structural health monitoring

Structural health monitoring (SHM) has now received considerable attention in the civil engineering community in the past few years. The use of in-situ, nondestructive sensing and analysis of structural characteristics, including the structural response, for the purpose of detecting changes that may indicate damage or degradation is referred to as SHM[1], which involves the data acquisition, the extraction of damage-sensitive features from these data, and the statistical analysis of these features to determine the current state of the system' s health. The damage diagnosis based on on-line monitoring data is one of the key issues in the SHM process, and most currently proposed methods utilize the changes in modal parameters (frequencies, mode shapes, modal curvature, flexibility, etc.) or use the finite element (FE) model updating techniques to identify damage. Recent research indicates that the vibration-based damage diagnosis is essentially a problem of statistical pattern recognition[2-3]. From this point of view, methods have been developed that utilize signal processing techniques and statistical analysis in extracting damage-sensitive features directly from on-line monitoring data. It is noted that neither sophisticated FE models nor modal parameters are employed in such a statistical pattern recognition approach. Thus, these methods avoid the shortcomings of model-dependence in traditional damage diagnosis, because high-fidelity FE model and sophisticated modal analyses often require labor intensive tuning and always result in significant uncertainties caused by various errors.

In this paper, a novel algorithm of damage-sensitive feature extraction and damage alarming is presented based on auto-regressive moving-average (ARMA) time series analysis. The numerical results of a three-span-girder model show that the new damage-sensitive feature index defined in this paper is sensitive to subtle structural damage and the proposed methods can be used for structural on-line damage alarming. The primary objective of this study is to identify the existence of damage, and the localization and quantification of damage are not addressed in this paper.

1 Description of the Algorithm

1.1 Time series modeling of monitoring data

The time series modeling algorithm involved to analyze the statistical properties of dynamic data systems is one of the key branches in statistics and probability[4]. In the SHM, the monitoring data obtained by in-situ sensors is a typical time series. Thus, damage detection can be performed using the time series modeling algorithm of monitoring data measured from a structure before and after an event. Furthermore, this algorithm is solely based on signal analysis of output data making this approach very attractive for the development of on-line SHM[5-7]. In this study, simulated monitoring data are modeled as ARMA time series models, and a novel algorithm of feature extraction and damage alarming is presented, subsequently based on the coefficients of ARMA models.

First, all the monitoring data are standardized prior

Received 2006-11-13.

Foundation items: The National High Technology Research and Development Program of China (863 Program) (No. 2006AA04Z416), the National Natural Science Foundation of China (No. 50538020).

Biographies: Liu Yi (1980—), male, graduate; Li Aiqun (corresponding author), male, doctor, professor, aiqunli@seu.edu.cn.

to any analysis hereafter, such that

$$x_{ij}(t)=\frac{X_{ij}(t)-\mu_{ij}}{\sigma_{ij}} \tag{1}$$

where $X_{ij}(t)$ is the j-th stream of monitoring data obtained from sensor i, and μ_{ij} and σ_{ij} are the mean and standard deviation of $X_{ij}(t)$, respectively. This standardization procedure is applied to all the monitoring data in this paper.

Once the initial data pre-processing is complete, an ARMA (p, q) model with p auto-regressive (AR) terms and q moving-average (MA) terms is constructed as

$$x_{ij}(t) = \sum_{k=1}^{p}\varphi_k x_{ij}(t-k) + \sum_{k=1}^{q}\theta_k \varepsilon_{ij}(t-k) + \varepsilon_{ij}(t) \tag{2}$$

where $x_{ij}(t)$ is the standardized monitoring data; φ_k and θ_k are the k-th AR and MA coefficients, respectively; p and q are the model orders of the AR and MA processes, respectively; and $\varepsilon_{ij}(t)$ is the residual term. It is noted that the ARMA model consists of two parts, one is the auto-regressive part representing that $x_{ij}(t)$ is affected by its previous values, and the other is moving-average which represents that $x_{ij}(t)$ is also affected by previous noise interference. Backward shift operator B is applied to let $B^k x_{ij}(t) = x_{ij}(t-k)$, thus Eq. (2) can be written symbolically as

$$\varphi(B)x_{ij}(t) = \theta(B)\varepsilon_{ij}(t) \tag{3}$$

where φ and θ are the p-th and the q-th degree polynomials. From Eq. (3), it can be obtained that

$$x_{ij}(t) = \frac{\theta(B)}{\varphi(B)}\varepsilon_{ij}(t) \tag{4}$$

The ARMA modeling process can be considered as the output $x_{ij}(t)$ from a linear filter, whose transfer function is the ratio of two polynomials $\theta(B)$ and $\varphi(B)$, when the input is $\varepsilon_{ij}(t)$. Thus, the estimation for the terms and coefficients of the ARMA model is essentially a process of system identification.

The spectrum of ARMA(p, q) model is

$$S(f) = 2\sigma^2 \frac{|\theta(\mathrm{e}^{-\mathrm{i}2\pi f})|^2}{|\varphi(\mathrm{e}^{-\mathrm{i}2\pi f})|^2} = 2\sigma^2 \frac{|1-\theta_1\mathrm{e}^{-\mathrm{i}2\pi f}-\ldots-\theta_q\mathrm{e}^{-\mathrm{i}2\pi f}|^2}{|1-\varphi_1\mathrm{e}^{-\mathrm{i}2\pi f}-\ldots-\varphi_p\mathrm{e}^{-\mathrm{i}2\pi f}|^2} \tag{5}$$

where σ is the variance of system input $\varepsilon_{ij}(t)$. It can be seen in Eq. (5) that the roots existing both in the numerator and the denominator of the spectrum $S(f)$ indicate the zeros and poles of the system. The equation $\theta(B)=0$ and $\varphi(B)=0$ are thus called the characteristic equation for the process.

When the factor analysis is applied to the characteristic equation, it can be obtained that

$$\varphi(B) = (1-\lambda_1 B)(1-\lambda_2 B)\ldots(1-\lambda_p B) = \prod_{k=1}^{p}(1-\lambda_k B) = 0 \tag{6}$$

$$\theta(B) = (1-\eta_1 B)(1-\eta_2 B)\ldots(1-\eta_q B) = \prod_{k=1}^{q}(1-\eta_k B) = 0 \tag{7}$$

where λ_k and η_k are the characteristic roots of the AR and the MA processes, respectively. From the view of the system identification theory, λ_k represent the poles of transfer function representing the natural characteristics of system, and η_k are the zeros of transfer function representing the relationship between the system and the external environment.

The innovation algorithm is used for estimating the coefficients of the ARMA models. The optimal model order is obtained using Akaike information criteria. This criteria consists of two terms, one of which is a log-likelihood function and the other penalizes the number of terms in the ARMA model. Also, a cross validation analysis is carried out to check the accuracy of the results. The details of the estimation of the model are given in Ref. [4].

1.2 Feature extraction and damage alarming

Feature extraction is the process of identifying damage-sensitive properties derived from the measured monitoring data that allows one to distinguish between the undamaged and damaged structures. Generally speaking, systematic differences between time series from the undamaged and damaged structures are nearly impossible to detect by eye. Therefore, other features of the monitoring data must be extracted for damage alarming.

As discussed in section 1.1, the coefficients of the AR part of the ARMA model can represent the natural characteristics of the system. Thus, it indicates that the AR coefficients can be used in feature extraction. In this paper, the new damage-sensitive feature index and damage alarming methods are presented as follows:

① Populate a database with signals from the undamaged structure at each sensor location. Then, divide these data into two parts. One is used as "the training database" named S_1 and the other is used as "the reference database" named S_2. Meanwhile, obtain data from the unknown conditions of the structure and make these data as the third database named S_3. Let n_1, n_2 and n_3 be the number of samples in S_1, S_2 and S_3, respectively.

② Fit an ARMA(p, q) model to all data samples in database S_1, S_2 and S_3.

③ Let the j-th AR coefficients of all the ARMA (p, q) models in set $S_i (i=1,2,3)$ be the vector $\boldsymbol{X}_j (j=1, 2, \ldots, p)$. Then, the principal component analysis

(PCA) is applied to the matrix $\boldsymbol{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_p\}$. It obtains the k-th principal component of the matrix $\boldsymbol{X}$ as

$$\boldsymbol{Y}_k = \boldsymbol{e}_k^{\mathrm{T}} \boldsymbol{X} = e_{1k}\boldsymbol{X}_1 + e_{2k}\boldsymbol{X}_2 + \dots + e_{pk}\boldsymbol{X}_p \qquad k = 1, 2, \dots, p \tag{8}$$

and

$$\left.\begin{aligned} \operatorname{var}(\boldsymbol{Y}_k) &= \boldsymbol{e}_k^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{e}_k = \lambda_k & k &= 1, 2, \dots, p \\ \operatorname{cov}(\boldsymbol{Y}_k, \boldsymbol{Y}_l) &= \boldsymbol{e}_k^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{e}_l = 0 & k &\neq l \end{aligned}\right\} \tag{9}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{X}$, $\lambda_1 \geqslant \lambda_2 \geqslant \dots \geqslant \lambda_p$ and $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_p$ are the characteristics roots and corresponding eigenvectors of $\boldsymbol{\Sigma}$.

④ Select the prior m principal-components ($m < p$, m is usually set at 2 to 4, then these prior m principal-components can hold about 80% to 90% of the total information of matrices $\boldsymbol{X}$) from the results in step ③ to perform data compression. Thus, it gets $n_1 \times m$, $n_2 \times m$ and $n_3 \times m$ principal-component matrix from S_1, S_2 and S_3, respectively.

⑤ Let the $n_1 \times m$ principal-component matrix be the m-dimensional population of $\boldsymbol{G}$ with mathematics expectation of $\boldsymbol{\mu}$ and covariance matrices of $\boldsymbol{\Sigma}$. The Mahalanobis-distances[8] of m-dimensional vector $\boldsymbol{x}_i$ ($i = 1, 2, \dots, n_2$ and $i = 1, 2, \dots, n_3$) from the other two principal-component matrix to the m-dimensional population of G are defined as the damage-sensitive feature (DSF) index:

$$D_{\mathrm{DSF}} = [(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})]^{\frac{1}{2}} \tag{10}$$

⑥ Determine the statistical significance in the differences of mean values of the pre- and post-event data to report damage alarming. If $\mu_{\mathrm{D,\,damaged}}$ and $\mu_{\mathrm{D,\,undamaged}}$ are defined as the mean values of the D_{DSF} obtained from the damaged and undamaged cases, respectively, then a hypothesis t-test may be set up as follows to determine if their differences are significant:

$$\begin{aligned} \mathrm{H}_0 &: \mu_{\mathrm{D,\,damaged}} = \mu_{\mathrm{D,\,undamaged}} \\ \mathrm{H}_1 &: \mu_{\mathrm{D,\,damaged}} \neq \mu_{\mathrm{D,\,undamaged}} \end{aligned} \tag{11}$$

where H_0 and H_1 are the null and alternate hypotheses, respectively. H_0 represents the undamaged condition and H_1 represents the damaged condition. The significance level of this test is set to 0.05.

Fig. 1 shows the flowchart of the feature extraction and damage alarming processes presented in this paper using the ARMA time series modeling algorithm and statistical analysis.

2 Application Results

In order to test the validity of the presented algorithm, results from the numerical simulation of a uniform beam model containing some subtle damage are used. The model is a three-equal-span girder (see Fig. 2) with the span length of $L = 10$ m. And the model is meshed into 10 elements in each span. Some other important parameters of the model are: the Young's module $E = 2 \times 10^5$ MPa, the density $\rho = 7\,850$ kg/m^3, the section area $A = 1.671 \times 10^{-3}$ m^2, and the section inertia moment $I = 1.652 \times 10^{-6}$ m^4.

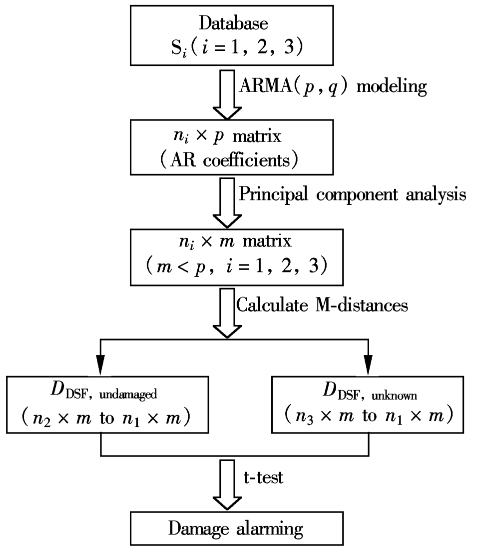

Fig. 1 Flowchart of feature extraction and damage alarming

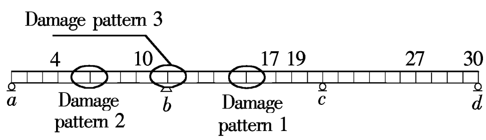

Fig. 2 Illustration of a three-span-girder model with damage patterns 1 to 3

A crack damage is simulated with an equivalent sub-beam having a reduced Young's module suggested in Ref. [9]. Thus, damage is simulated by reducing the Young's modules of elements, resulting in a loss of stiffness. Considering the symmetry of the three-span-girder model, damage patterns are designed in three different positions but with the same damage level. Damage patterns include:

① Damage pattern 1: reduce 10% of the Young's module of the two elements near the center of middle-span;

② Damage pattern 2: reduce 10% of the Young's module of the two elements near the center of left-side-span;

③ Damage pattern 3: reduce 10% of the Young's module of the two elements near restriction b.

Tab. 1 shows the prior five orders of natural frequencies of undamaged girder and damaged girders. It can be seen that the percentage deviations of natural frequencies for undamaged and damaged girders are all less than 1% due to the changes caused by subtle damage.

Tab. 1 Natural frequencies and percentage deviations for undamaged and damaged models

Mode number	Undamaged	Damage pattern 1		Damage pattern 2		Damage pattern 3	
	f_0/Hz	f_1/Hz	Deviation/%	f_2/Hz	Deviation/%	f_3/Hz	Deviation/%
1	2.493	2.475	0.72	2.475	0.72	2.493	0.02
2	3.195	3.194	0.03	3.170	0.79	3.180	0.48
3	4.664	4.625	0.83	4.653	0.23	4.621	0.92
4	9.970	9.960	0.10	9.960	0.10	9.960	0.10
5	11.364	11.351	0.11	11.338	0.23	11.325	0.34

Note: Deviation = $\left| f_i - f_0 \right| f_0 \times 100\%$, $i = 1, 2, 3$.

In order to obtain structural vibration response signal, the girder is loaded by a stochastic white-noise excitation. The output signal of vertical acceleration response is obtained as the simulated monitoring data from every node (27 nodes in total except for restrictions a, b, c and d). And the sampling frequency of the output signal is set to 100 Hz with a duration time of 450 s.

The simulated monitoring data obtained from undamaged condition is divided into two databases (S_1 and S_2) mentioned in section 1.2. One consists of the data obtained during the prior 300 s (S_1) and the other consists of the data obtained in the remaining 150 s (S_2). Meanwhile, the data obtained during the posterior 150 s from all the damaged patterns is used as database S_3. All the simulated monitoring data in S_1, S_2 and S_3 are further divided into 100, 50 and 50 segments to get sets of data samples, respectively. The amount of data in each segment is set to 2 048 with 87.5% of data overlapped among adjacent segments.

Fig. 3 shows the results from the application of the proposed feature extraction and the damage alarming method to the numerically simulated data samples. From Figs. 3 (a) to (f), it can be observed that there is a significant difference between the mean values of the D_{DSF} obtained from the damaged and undamaged cases. Then a hypothesis test involving the t-test method is further applied to obtain a decision of

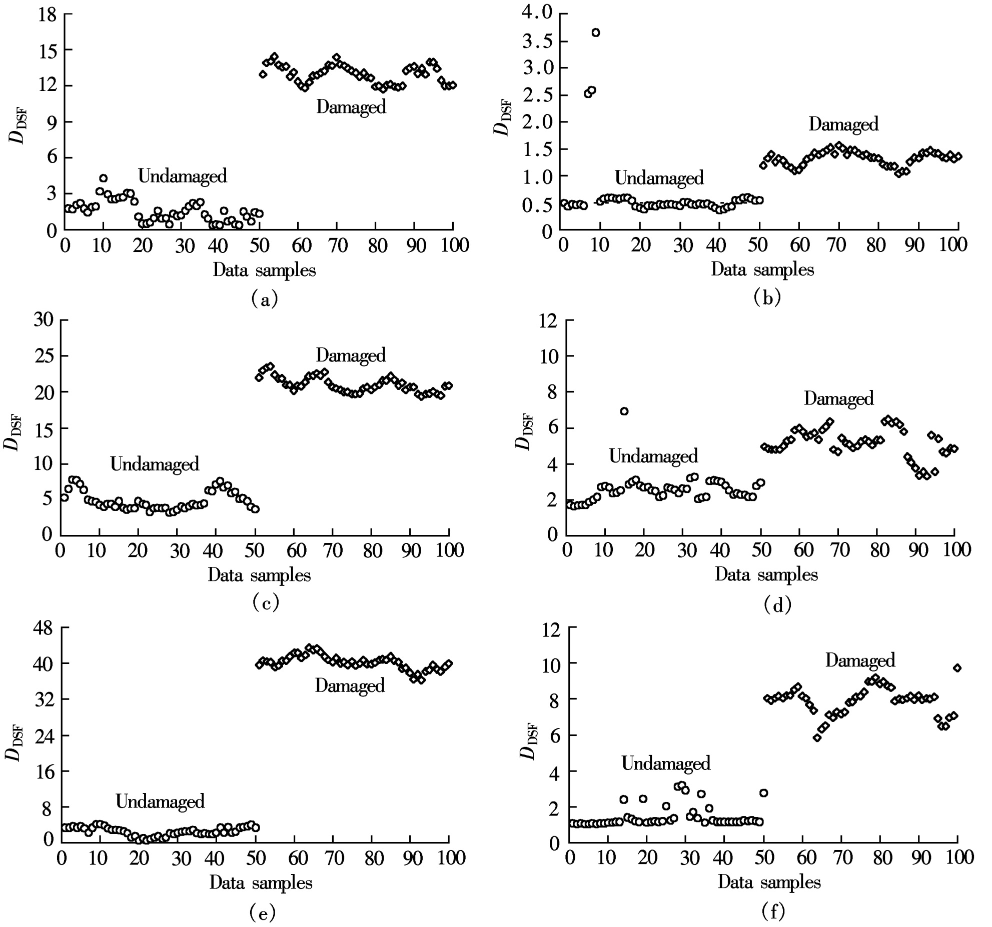

Fig. 3 Variation of D_{DSF} from part of the nodes for damage patterns 1 to 3. (a) Variation of D_{DSF} from node 4 for damage pattern 1; (b) Variation of D_{DSF} from node 10 for damage pattern 1; (c) Variation of D_{DSF} from node 17 for damage pattern 2; (d) Variation of D_{DSF} from node 27 for damage pattern 2; (e) Variation of D_{DSF} from node 19 for damage pattern 3; (f) Variation of D_{DSF} from node 30 for damage pattern 3

damage alarming if their differences are significant.

Tab. 2 shows the results of the damage decision for damage patterns 1 to 3 for the numerical simulation study. It is observed that for damage patterns 1 to 3, the damage alarming decision H_1 is given by 25, 25 and 22 nodes, respectively. These alarming nodes are in the majority of 92.6%, 92.6% and 81.5% of 27 nodes in total, respectively. The p-value is the probability that D_{DSF} does not predict damage, given the fact that there is damage in the structure. Since the p-values are all much less than the significance level of 0.05, the null hypothesis H_0 is rejected and the alternate hypothesis H_1 is accepted. In this numerical example, it can be concluded that the D_{DSF} index defined in this paper is sensitive to subtle structural damage and the proposed algorithm is able to detect the existence of damage.

Tab. 2 Results of damage alarming for damage patterns 1 to 3

Node number		Damage pattern 1		Damage pattern 2		Damage pattern 3	
		t-test	p-value	t-test	p-value	t-test	p-value
Left-side-span	2	H_1	0.000	H_1	0.030	H_0	0.329
	3	H_1	0.000	H_1	0.000	H_1	0.000
	4	H_1	0.000	H_1	0.000	H_1	0.000
	5	H_1	0.000	H_1	0.000	H_1	0.000
	6	H_1	0.000	H_1	0.002	H_1	0.003
	7	H_1	0.000	H_1	0.000	H_1	0.000
	8	H_1	0.000	H_1	0.000	H_0	0.157
	9	H_0	0.718	H_1	0.023	H_0	0.108
	10	H_1	0.000	H_1	0.000	H_1	0.043
Middle-span	12	H_1	0.000	H_1	0.000	H_1	
	13	H_1	0.000	H_1	0.000	H_1	0.000
	14	H_1	0.000	H_1	0.000	H_1	0.000
	15	H_1	0.000	H_1	0.000	H_1	0.000
	16	H_1	0.000	H_1	0.000	H_1	0.000
	17	H_1	0.000	H_1	0.000	H_1	0.000
	18	H_1	0.000	H_1	0.000	H_1	0.000
	19	H_1	0.000	H_1	0.000	H_1	0.000
	20	H_1	0.000	H_1	0.000	H_1	0.000
Right-side-span	22	H_1	0.000	H_0	0.240	H_0	0.996
	23	H_0	0.803	H_0	0.346	H_0	0.958
	24	H_1	0.000	H_1	0.000	H_1	0.000
	25	H_1	0.000	H_1	0.000	H_1	0.000
	26	H_1	0.000	H_1	0.000	H_1	0.000
	27	H_1	0.000	H_1	0.000	H_1	0.000
	28	H_1	0.000	H_1	0.000	H_1	0.000
	29	H_1	0.000	H_1	0.000	H_1	0.000
	30	H_1	0.000	H_1	0.000	H_1	0.000

Note: H_1 reports damaged and H_0 reports undamaged.

3 Conclusion

In this paper, a damage detection method based on the time series modeling algorithm is discussed. A damage-sensitive feature index, which is derived using the principal component analysis from the coefficients of the AR part of the model, is presented. A hypothesis test involving the t-test method is used to obtain a damage alarming decision. These methodologies were tested on a three-span-girder model. The results indicate that the algorithm of feature extraction and damage alarming presented in this paper can be used for on-line damage alarming in SHM because of its simplicity and output data only stream analysis.

References

[1] Housner G W, Bergman L A, Caughey T K, et al. Structural control: past, present and future [J]. *Journal of Engineering Mechanics*, 1997, **123**(9): 897 − 971.

[2] Bement M T, Farrar C R. Issues for the application of statistical models in damage detection [C]//*Proc of the 18th International Modal Analysis Conference*. San Antonio, USA, 2000: 1392 − 1398.

[3] Lei Y, Kiremidjian A S, Nair K K, et al. Statistical damage detection using time series analysis on a structural health monitoring benchmark problem [C]//*Proc of the 9th International Conference on Application of Statistics and Probability in Civil Engineering*. San Francisco, USA, 2003: 581 − 587.

[4] Brockwell P J, Davis R A. *Introduction to time series and forecasting* [M]. 2nd ed. New York: Springer-Verlag, 2002.

[5] Sohn H, Farrar C R, Hunter N F, et al. Structural health monitoring using statistical pattern recognition techniques [J]. *Journal of Dynamic Systems, Measurement and Control, ASME,* 2001, **123**(4): 706 - 711.

[6] Ou J P, He L, Xiao Y Q. Parameter identification of offshore platform structure using time series models and technology of extracting free vibration signal [J]. *Applied Mathematical and Mechanics*, 2003, **24**(4): 398 - 404.

[7] Nair K K, Kiremidjian A S, Law K H. Time series based damage detection algorithm with application to the ASCE benchmark structure [C]//Augusti G, Schueller G I, Ciampoli, eds. *The 9th International Conference on Structural Safety and Reliability* (*ICOSSAR* 2005). Rotterdam: Millpress, 2005: 761 - 768.

[8] Rice J A. *Mathematical statistics and data analysis* [M]. 2nd ed. Belmont: Duxbury Press, 1999.

[9] Gentile A, Messina A. On the continuous wavelet transforms applied to discrete vibrational data for detecting open cracks in damaged beams [J]. *International Journal of Solid and Structures*, 2003, **40**(2): 295 - 315.

一种利用时间序列分析的特征提取与损伤预警方法

刘 毅 李爱群 费庆国 丁幼亮

(东南大学混凝土及预应力混凝土结构教育部重点实验室,南京 210096)

摘要:针对结构健康监测中如何基于在线监测数据实现损伤诊断的问题,提出了一种利用时间序列分析 ARMA 模型的特征提取和损伤预警方法. 首先对所有监测数据样本建立 ARMA 模型,以模型中 AR 部分参数的主成分矩阵构建 Mahalanobis 距离判别函数,提出了一种新的结构损伤敏感指标 D_{DSF}. 然后,采用 t-检验考察该指标在损伤前后是否存在显著性变化,从而可以有效地实现结构损伤预警. 三跨连续梁数值算例表明,提出的结构损伤特征指标对结构的微小损伤具有敏感性,具备结构在线实时损伤预警的应用价值.

关键词:特征提取;损伤预警;时间序列分析;结构健康监测

中图分类号:TU311