 ICCES, vol.6, no.3, pp.183-188

Influence of the Regression Error of the Response Surface to the Diagnostic Accuracy of the Unsupervised Statistical Damage Diagnostic Method

A.Iwasaki[1], K.Yuguchi[2], A.Todoroki[3], Y.Shimamura[4]

Summary

The present study is about study on the diagnostic accuracy of the unsupervised damage diagnosis method named SI-F method. For the health monitoring of existing structures, modeling of entire structure or obtaining data sets after creating damage for training is almost impossible. This raises significant demand for development of a low-cost diagnostic method that does not require modeling of entire structure or data on damaged structure. Therefore, the present study proposes a low-cost unsupervised statistical diagnostic method for structural damage detection. The proposed method statistically diagnoses structural condition by means of investigating the change of a response surface which conducts the system identification between sensor outputs. The response surface is calculated as a regression model of relationship between multiple sensors. The shape of the response surface is changed reflecting the change of the structural condition. In this method, the change of the response surface is statistically investigated with the F-test. In the F-test, the threshold of normal or damaged condition is decided with only theoretical F-probability distribution. This theoretical F-distribution is easily calculated using the response surface parameters. Therefore, diagnosis is conducted by means of only intact data used for the reference data. This means the proposed method doesn't require information about the damaged condition.

Since the SI-F method is able to detect the damage in the structure by judging the deviation from the normal state, it is important to reduce the false positive detection for raising the reliability of the structure. In the present study, to clarify the relationship between the condition of the false positive detection, diagnostic accuracy and the regression error of the response surface, several numerical simulations were carried out.

Introduction

A large number of the structures including the bridge, the power plant, etc. are reaching their lifetime considered when they designed. Since a lot of the structures

[1]Department of mechanical engineering, Gunma university, 1-5-1 Tenjincho, Kiryu, Gunma, Japan

[2]Grad. school of engineering, Gunma university, Japan (Thermos K.K. at present)

[3]Dept. of Mech. Sciences and engineering., Tokyo Tech, 2-12-1 Ookayama, Meguro-ku, Tokyo, Japan

[4]Dept. of Mechanical Engineering, Shizuoka university, 3-5-1 Johoku, Hamamatsu, Shizuoka, Japan

deteriorate due to aging, cost for maintenance is increasing year by year. Previously, the integrity of the structure was evaluated by the in service inspection (ISI) using the visual inspection, or the kinds of non-destructive test such as CT, PT etc. However, the cost of the ISI is remarkably high and those kind of structures must be suspended during the inspection. Moreover, since they require the highly qualified technicians, these types of inspecting might not be possible after major event. Thus the Structural health monitoring attracts global attention as the technology raising the reliability of structure and reducing the maintenance cost. Structural health monitoring will be achieved by judging the integrity of the structure from analysis of the physical parameter obtained from the sensor installed to the structure.

Although many researchers have developed the signal processing technology to estimate the integrity of the structures ever, most of the previous methods were not practical from several reasons. For instance, the complex physical model is necessary to utilize the FEM analysis, and the large number of data acquired from the fracture test is required to apply the neural networks. The SI-F method was suggested by Iwasaki et al[1-2], as the reasonable damage detection algorithm which doesn't use the data of damaged state. By the SI-F method, damages are detected by judging the statistical difference of data of intact state and present state. Therefore, the SI-F method is novel damage detection algorithm which doesn't use the complex physical model for FEM and the training data for the neural networks. The difference between the normal data and present data will be judged by the difference of the correlation among the sensors attached. The correlation among the sensors will be identified by the response surface methodology. The normal response surface will be obtained from data acquired in the intact state, and the detection response surface will be obtained from data acquired in present state. By testing the similarity between the normal response surface and the detection response surface using the F-test, it is possible to estimate the damage occurrence in the structure. If the hypothesis of the similarity were rejected, damage would be detected. Otherwise, integrity of the structure would be proved.

Since the SI-F method is able to detect the damage in the structure by judging the deviation from the normal state, it is important to reduce the false positive detection for raising the reliability of the structure. In the present study, to clarify the relationship between the condition of the false positive detection, diagnostic accuracy and the regression error of the response surface, several numerical simulations were carried out.

Procedures for the damage diagnosis

Damage diagnosis method using SI-F method

By the SI-F method[1-2], damages are detected by judging the statistical difference of data of intact state and present state. The difference between the normal data and present data will be judged by the difference of the correlation among the output of the sensors equipped. The correlation among the sensor output will be identified by the response surface methodology. The normal response surface (NRS) will be obtained from data acquired in intact state, and the detection response surface (DRS) will be obtained from data acquired in present state. By testing the similarity between the NRS and the DRS using the F-test, it is possible to estimate the damage occurrence in the structure. If the hypothesis of the similarity were rejected, damage would be detected. Otherwise, integrity of the structure would be proved.

System identification using response surface methodology

Response surface methodology is use for the system identification in this method. Response surface methodology is employed for the process optimization in a quality engineering field. Response surface methodology consists of a design of experiments to select the most suitable points for fitting the surfaces effectively and the least-square-method to regress response surfaces. Response surface is the approximation function that expresses the relationship between a response and predictors. Generally, a response surface is represented with the following formula.

$$y = f\left(x_1, x_2, \cdots, x_l\right) + \varepsilon \tag{1}$$

Where x are predictors, y is a response, ε is a regression error and l is a number of predictors. In general, 1st or 2nd degree polynomial is used for response surface. In the SI-F method, Response surface is applied to identify the correlation among sensor output. By using the response surface, it is possible to detect the damage of the structure of interest accurately regardless of the change of the boundary condition.

When the polynomial is used for the response surface, in terms of n observations, the equation (1) can be written in matrix form as follows.

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \tag{2}$$

Unbiased estimator of $\beta(\mathbf{b})$ is obtained using the least-square-method as follows.

$$\mathbf{b} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y} \tag{3}$$

Since the response surface is regressed from least-square method, sum of the squares error (*SSE*) is defined as follow.

$$SSE = \mathbf{Y}^T\mathbf{Y} - \mathbf{b}^T\mathbf{X}^T\mathbf{Y} \tag{4}$$

When the regression model for the response surface is perfect, the residual error follows the normal distribution.

Similarity test of the response surface using statistical F-test

The similarity of the NRS and the DRS is tested by the F test that is generally used for test of the similarity of the two distributions. By assuming the regression error of each response surface is independent, and follows normal distributions of the same variance, the test statistic F_0 is defined as follows:

$$F_0 = \frac{SSE_0 - (SSE_1 + SSE_2)}{(SSE_1 + SSE_2)} \frac{n-2p}{p} \tag{5}$$

Where subscript 0 means the response surface with all data. Subscript 1 and 2 mean the NRS and DRS respectively. If two response surfaces are similar, the test statistic F_0 follows the theoretical F-distribution of degree of freedom $(p,n\text{-}2p)$. Thus critical limit for the hypothesis of similarity is defined as follows:

$$F_0 > F^{\alpha}_{p,n-2p} \tag{6}$$

Influence of the regression error of the response surface to the diagnostic accuracy of the statistical damage diagnostic method

Simulation model

As mentioned before, the influence of the regression accuracy of the response surface to the accuracy of the damage diagnosis is considered here. For simulation, 3 sensor response surface model is considered. Sine wave with gaussian white noise is used as the sensor measurement. The sensor measurement at arbitrary time is shown by the following formula.

$$N_i = A_i \sin(2\pi f_i t + \varphi_i) + e(t) \quad (i = 1, 2, \cdot\cdot, k) \tag{7}$$

Where N is sensor measurement, t is time. A is the amplitude, f is frequency (=10), φ (= 0, π/16, π/8, π/4, π/2) is the phase of the sine wave. $e(t)$ shows a gaussian white noise with 1 for variance, k is the number of the sensor. Since the polynomial is used for the response surface, this "shift of the phase" means the irrelevancy of the model of the response surface.

When the residual error only caused by the effect of the noise, the residual error follows the normal distribution. As shown in the previous section, only the case of the residual error follows the normal distribution, F_0 distribution follows theoretical distribution. Therefore, F_0 distribution of the undamaged condition follows the theoretical distribution when the model of the response surface is perfect. However, when the regression model of the response surface is not suitable for the regression of the relation between the sensors, the residual error may not follow the normal distribution. This section address the effect of residual error caused by the irrelevancy of the model of the response surface. The irrelevancy of the model is simulated by the gap of the phase of the sine wave. Frequency Quadratic polynomial is used for the response surface.

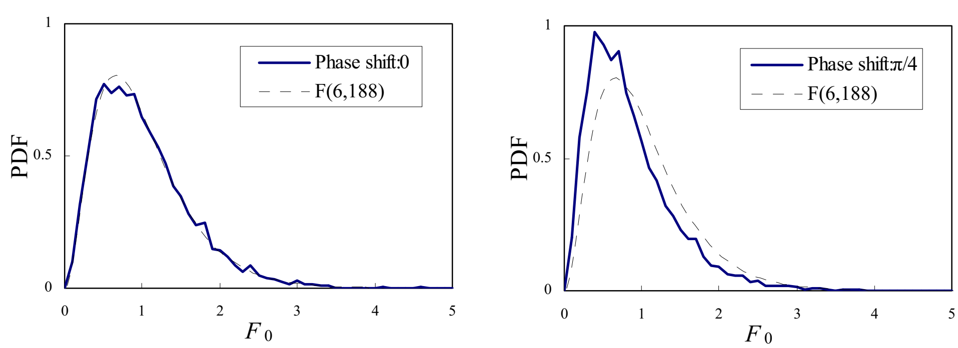
$$N_1 = \beta_0 + \beta_1 N_2 + \beta_2 N_3 + \beta_3 N_2^2 + \beta_4 N_3^2 + \beta_5 N_2 N_3 \tag{8}$$

Number of the data for regression is 100. In this case, theoretical distribution follows $F(6{,}188)$. Damage is simulated as setting the amplitude A to the twice value.

Effect of residual error caused by the irrelevancy of the model of the response surface

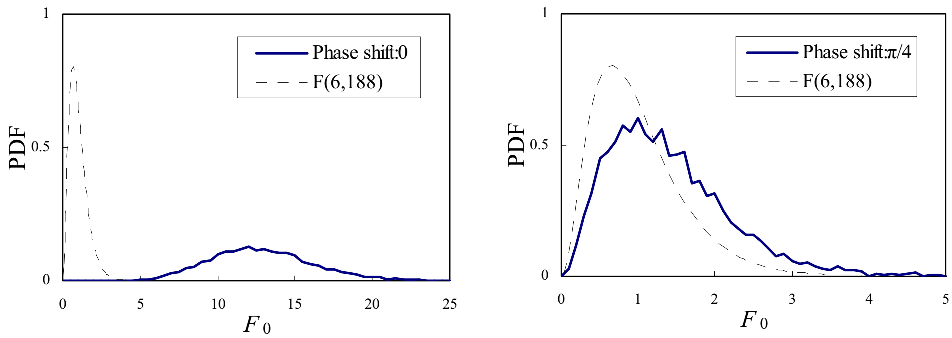
Figure 1 and 2 shows the probability distribution of F_0 of the damage diagnosis of the undamaged state and the damaged state when the phase shift is 0 and $\pi/4$ for example. As shown in the figures, when the response surface model is not suitable, the regression error does not follow the normal distribution and probability distribution function of the F_0 doesn't follows the theoretical distribution. Table 2 shows the result of test of goodness of fit between F_0 distribution and theoretical distribution using statistical χ^2 test. 99% lower confidential interval of the test is 63.7. As shown in the table, with the decrease of the regression accuracy, F_0 distribution deviate from the theoretical distribution and the probability of the false positive detection increases. Table 2 shows together the average of F_0 distribution of each phase shift conditions. As shown in the table, since the average of theoretical distribution is 1.01, at the damaged condition, deviation from the theoretical distribution decreases when the phase shift of N_1 increases. It means that the decrease of the relevancy of the response surface causes the difficulty of the damage diagnosis.

To reduce the false positive detection and increase the diagnostic accuracy, not only the model which maximizes the regression accuracy but also the model which the regression error follows a normal distribution is necessary to select as the response surface.



(a) $\varphi = 0$ (b) $\varphi = \pi/4$

Figure 1: Influence of phase shift for the PDF of F_0 (undamaged state)



(a) $\varphi = 0$ (b) $\varphi = \pi/4$

Figure 2: Influence of phase shift for the PDF of F_0 (damaged state)

Table 1: Summary of the result of the simulation

		0	π/16	π/8	π/4	π/2
Undamaged	R^2_{adj}	0.970	0.933	0.829	0.489	-0.004
	χ^2	36.4	200.0	350.0	547.0	654.0
Damaged	$E(F_0)$	12.9	5.64	2.54	1.41	1.02

Conclusions

By conducting similarity tests of the two identified system by the F-test, the present paper describes the new unsupervised damage diagnosis method. And relationship between the condition of the damage detection and the regression error of the response surface is clarified. As a result, to reduce the false positive detection and increase the diagnostic accuracy, not only the model which maximizes the regression accuracy but also the model which the regression error follows a normal distribution is necessary to select as the response surface.

Acknowledgement

Funding for this study was primarily provided by New Energy and Industrial Technology Development Organization. The authors express their appreciation.

References

1. A. Iwasaki, etc (2006): “Damage and Fault Diagnosis of In-service Structure via Statistical Comparison of Relation between Sensor measurements”, *Transactions of the JSME*, A72(724), pp.1955-1969 (in Japanese).
2. A. Iwasaki, etc (2005): “Unsupervised statistical damage diagnosis for SHM of existing civil structure”, *Smart Materials and Structures*, 14, pp.154-161.