Stochastic model updating using distance discrimination analysis

Deng Zhongmin a, Bi Sifeng a,*, Sez Atamturktur b

a School of Aeronautics, Beihang University, Beijing 100191, China
b Glenn Department of Civil Engineering, Clemson University, Clemson, SC 29632, USA

Received 9 September 2013; revised 3 March 2014; accepted 15 May 2014
Available online 20 August 2014

Abstract This manuscript presents a stochastic model updating method, taking both uncertainties in models and variability in testing into account. The updated finite element (FE) models obtained through the proposed technique can aid in the analysis and design of structural systems. The authors developed a stochastic model updating method integrating distance discrimination analysis (DDA) and advanced Monte Carlo (MC) technique to (1) enable more efficient MC by using a response surface model, (2) calibrate parameters with an iterative test-analysis correlation based upon DDA, and (3) utilize and compare different distance functions as correlation metrics. Using DDA, the influence of distance functions on model updating results is analyzed. The proposed stochastic method makes it possible to obtain a precise model updating outcome with acceptable calculation cost. The stochastic method is demonstrated on a helicopter case study updated using both Euclidian and Mahalanobis distance metrics. It is observed that the selected distance function influences the iterative calibration process and thus, the calibration outcome, indicating that an integration of different metrics might yield improved results.

© 2014 Production and hosting by Elsevier Ltd. on behalf of CSAA & BUAA.

1. Introduction

Finite element (FE) models, developed to analyze engineering systems in various fields, are often used in a predictive capacity at an untested setting. The common practice of comparing model predictions to measurements at tested settings invariably yields discrepancies due to uncertainties in models and variability in testing. As a result, model updating techniques have been used to obtain models that more closely match experimental measurements. There are three main sources of disagreement between FE analytical data and test measurements, all of which should be considered in model updating.

(1) Parameter uncertainty. An FE model generally involves a set of imprecise parameters of the physical structures (e.g. elastic modulus, mass density, geometric size, and spring stiffness).

(2) Modeling uncertainty. Unavoidable simplifications and idealizations (e.g. assuming a linear response, frictionless joints and the erroneous modeling of boundary
conditions) prevent FE models from accurately representing physical characteristics.

(3) Testing variability. Due to the uncontrollable random effects of the test system, measurements are only partially reproducible.

Uncertainty is either epistemic (reducible) or aleatory (irreducible). Epistemic uncertainty is due to a lack of knowledge and consists of imprecise parameters and inexact model performance (Categories 1 and 2). For example, typical aleatory uncertainties are manufacturing tolerance and test variability (Category 3). Regarding the treatment of epistemic uncertainty, inverse analysis techniques are developed to update models. For the treatment of aleatory uncertainty, statistical characteristics (e.g., mean and variance) as well as intervals analysis are typically employed.

Stochastic model updating is a procedure, in which various statistical algorithms are used to update the model to better predict measurements by considering parameter uncertainty, model form uncertainty, and test variability. Stochastic model updating includes problems such as error localization, parameter selection, feature extraction, and correlation analysis, which have been and continue to be extensively studied. Applications in the example demonstrate the performance of the proposed approach in obtaining an improved model at an acceptable computational expense.

2. Methods of analysis

2.1. Uncertain structural system

An uncertain structural system generally includes a set of random input variables \( \mathbf{x} \) with a nominal value \( \mathbf{x}_0 \) and variations \( \Delta \mathbf{x} \) around this nominal value:

\[
\mathbf{x} = \mathbf{x}_0 + \Delta \mathbf{x}. \tag{1}
\]

The random input variables of interest (termed as "parameters") are uncertain and have significant influence on the structural response of interest (termed as "features"). These parameters, selected based on structural characteristics and engineering judgments, are calibrated during the updating procedure. As the structures become more complex, it is not rare to encounter larger FE models involving a large amount of parameters. Parameter selection is consequently developed as a research focus to identify which parameters exhibit the most effect on the features of interest. Currently, techniques such as analysis of variance and MC-based sensitivity analysis have been employed in this field.

In this paper, the authors describe a novel stochastic model updating method integrating distance discrimination analysis (DDA) and advanced Monte Carlo (MC) sampling. DDA is used here to perform test-analysis correlation (TAC), which refers to the process of determining the degree of similarity between the experimental measurements and corresponding (or complementary) model predictions. The DDA-based test-analysis correlation procedure can provide a stochastic quantification of the disagreement based on statistical data samples.

MC, while regarded as a suitable stochastic analysis framework for the forward propagation of uncertainty due to its high precision, is demanding in regards to the computational resources necessary for its use. Consequently, techniques such as subset simulation, line sampling and parallel algorithm are employed to mitigate computational challenges. Although stochastic updating methods have been proposed to successfully cope with high computational resource demand, these demands have prevented applications to real life, non-trivial problems.

In this manuscript, the authors describe the use of a novel stochastic TAC procedure that integrates DDA and MC. Moreover, the authors analyze and compare the suitability of various distance functions used as correlation metrics during DDA. Furthermore, the authors enable more efficient MC by using response surface models to dramatically reduce the calculation cost. Applications in the example demonstrate the performance of the proposed approach in obtaining an updated model at an acceptable computational expense.

2.1. Uncertain structural system

An uncertain structural system generally includes a set of random input variables \( \mathbf{x} \) with a nominal value \( \mathbf{x}_0 \) and variations \( \Delta \mathbf{x} \) around this nominal value:

\[
\mathbf{x} = \mathbf{x}_0 + \Delta \mathbf{x}. \tag{1}
\]

The random input variables of interest (termed as “parameters”) are uncertain and have significant influence on the structural response of interest (termed as “features”). These parameters, selected based on structural characteristics and engineering judgments, are calibrated during the updating procedure. As the structures become more complex, it is not rare to encounter larger FE models involving a large amount of parameters. Parameter selection is consequently developed as a research focus to identify which parameters exhibit the most effect on the features of interest. Currently, techniques such as analysis of variance and MC-based sensitivity analysis have been employed in this field.

Assume a set of identical test structures are constructed using identical materials and procedures but with manufacturing tolerance and material heterogeneity. A measurement sample of features can be obtained through a multi-structure multi-measurement strategy with associated test variability. The input/output random variables account for the uncertainty in both parameters and features. An uncertain structural system can be characterized as a group of complex functional relationships between \( p \) parameters and \( q \) features:

\[
\begin{align*}
\mathbf{y} &= \mathbf{f} (\mathbf{x}) + \mathbf{e} \\
\mathbf{y} &= [y_1 \ y_2 \ \ldots \ y_q]^T \\
\mathbf{x} &= [x_1 \ x_2 \ \ldots \ x_p]^T
\end{align*} \tag{2}
\]

where \( \mathbf{e} \) is a zero mean random error.

Suppose the number of identical structures is \( u \), and for each of the structures, \( x \) repeated measurements are executed. Consequently, the size of the measurement data sample is \( m = u \times x \), with the matrix of this sample \( \mathbf{Y}_{test} \) assembled as

\[
\begin{align*}
\mathbf{Y}_{test} &= [\mathbf{y}_1 \ \mathbf{y}_2 \ \mathbf{y}_3 \ \ldots \ \mathbf{y}_u] \\
\mathbf{y}_j &= [y_{j1} \ y_{j2} \ y_{j3} \ \ldots \ y_{jq}]^T \ (j = 1, 2, \ldots , q)
\end{align*} \tag{3}
\]

where \( \mathbf{Y}_{test} \in \mathbb{R}^{m \times q} \), \( \mathbf{y}_j \in \mathbb{R}^{q \times 1} \).

For the feature sample, techniques such as moment estimation and maximum likelihood estimation can be implemented to estimate mean and variance of the data population. The mean \( \bar{y} \) and variance \( s^2 \) of each of the \( q \) features are obtained from

\[
\begin{align*}
\bar{y}_j &= \frac{1}{m} \sum_{k=1}^{m} y_{jk} = \frac{1}{m} \mathbf{y}_j^T \mathbf{e}_m; \\
\bar{s}^2 &= \frac{1}{m-1} \sum_{k=1}^{m} (y_{jk} - \bar{y}_j)^2 = \frac{1}{m-1} \left( \bar{y}_j - \bar{y}_j \mathbf{e}_m \right)^T (\bar{y}_j - \bar{y}_j \mathbf{e}_m)
\end{align*} \tag{4}
\]

where \( j = 1, 2, \ldots , q \); \( \mathbf{e}_m = [1 \ \ldots \ 1]^T \); \( \mathbf{e}_m^T \mathbf{e}_m = m \). Mean vector \( \bar{y} \) and covariance matrix \( \mathbf{C} \) of the feature sample are determined according to

\[
\begin{align*}
\bar{y} &= \frac{1}{m} [y_1 \ y_2 \ \ldots \ y_q]^T \mathbf{e}_m = \frac{1}{m} \mathbf{Y}_{test}^T \mathbf{e}_m \\
\mathbf{C} &= \frac{1}{m-1} (\mathbf{Y}_{test} - \bar{y} \mathbf{1})^T (\mathbf{Y}_{test} - \bar{y} \mathbf{1}^T)
\end{align*} \tag{5}
\]

The sample mean vector \( \bar{y} \in \mathbb{R}^{q \times 1} \) and covariance matrix \( \mathbf{C} \in \mathbb{R}^{q \times q} \) are found to be unbiased estimators of the population mean vector \( \bar{\mathbf{y}} \) and the population covariance matrix \( \mathbf{\Sigma} \).

2.2. MC simulation and response surface model

FE models of structural systems are available only with modeling simplifications and approximations. The selected

\[
\begin{align*}
\mathbf{x} &= \mathbf{x}_0 + \Delta \mathbf{x}. \tag{1}
\end{align*}
\]
parameters are assumed to obey the given probability distributions, which are subsequently propagated through FE models during MC to describe uncertainties of features. These prior distributions are typically determined with expert opinion, previously published literature or repeated tests on the components and/or materials.

Pairs of input and output data points generated through FE-runs during MC are illustrated in Fig. 1. After a sample of analytical data \( Y_{\text{analysis}} \) is obtained, test-analysis correlation is undertaken to quantify the degree of similarity (and dissimilarity) between \( Y_{\text{analysis}} \) and the test measurement \( Y_{\text{test}} \).

In MC, an increased demand for precision requires an increased number of statistically independent samples. On the other hand, for FE models that are computationally demanding, increasing the required sample size makes the direct MC prohibitively expensive. Thus, it is often necessary to implement techniques with the ability to reduce the calculation burden of direct MC procedure.

A response surface model, also known as the agent model or meta-model, is a purely mathematical model representing the complex relationship between parameters and features. A response surface model is thus particularly well-suited for sampling-based uncertainty analysis techniques (e.g. direct MC) because of its fast-running nature.

The first step to construct the agent model is to prepare a certain number of input/output FE-data points, i.e. training sample. Obviously, with greater size of the training sample, the generated agent model will be more precise, however, with more FE-runs and greater calculation burden. The technique known as design of experiment (DoE) can systematically navigate the parameter space in an optimal fashion. As a typical method of DoE, the orthogonal Latin square design is proposed here to efficiently arrange multiple FE-runs.

It is assumed that each parameter has \( r \) levels, and the total number of parameter is \( p \). If we want to get complete information of the parameter space, \( r^p \) times of FE-runs are needed, which require unrealistic calculation cost. However, with the orthogonal Latin square method, \( r^2 \) times of FE-runs are enough to get uniform and comprehensive information. A specific example with 4 parameters and 4 levels is proposed here to demonstrate the usage of this method.

In Fig. 2, the four parameters are respectively labeled as \( A, B, C \) and \( D \), with subscripts as their levels, e.g., \( C_4 \) means the 3rd parameter with its 4th level. The procedure of the orthogonal Latin square design follows these steps: (1) construct parameters \( A \) and \( B \) using the full factor design, as shown in the top left of Fig. 2, \( 4^2 = 16 \) times arrangements are generated; (2) as shown in the bottom left of Fig. 2, choose two orthogonal Latin squares respectively for parameters \( C \) and \( D \), (the orthogonal Latin squares can be consulted from most statistical manuals), (3) superpose the three blocks \( AB, C, \) and \( D \) to obtain the sampling strategy in the parameter space as shown in the right part of Fig. 2.

The second step, as soon as the training sample is generated, is to decide the format of the agent model and subsequently to train this model. The polynomial-based function is a typical model with advantages such as simple principle, high efficiency, etc. A quadratic polynomial-based agent model between the \( j \)th feature and all of the \( p \) parameters can be defined as

\[
y_j = \beta_0^j + \sum_{i=1}^{p} \beta_i^j x_i + \sum_{i<j}^{p} \beta_{ij}^j x_i x_j + \sum_{i=1}^{p} \beta_i^j x_i^2 = z^T \beta_j
\]

(6)

where

\[
\begin{aligned}
\mathbf{z} &= [1, x_1, x_2, \ldots, x_p, x_1 x_2, x_1 x_3, \ldots, x_1 x_p, x_1^2 x_2, \ldots, x_1^2 x_p]^T \\
\mathbf{\beta}_j &= [\beta_0^j, \beta_1^j, \beta_2^j, \ldots, \beta_p^j, \beta_{12}^j, \beta_{13}^j, \ldots, \beta_{1p}^j, \beta_{22}^j, \ldots, \beta_{pp}^j]^T
\end{aligned}
\]

(7)

The output index \( j \) is omitted in the following context for clarity. The unknown coefficient vector \( \mathbf{\beta} \) has the following dimension:

\[
w = \frac{(p + 1)(p + 2)}{2}
\]

(8)

Multiple regression analysis is employed here to determine the coefficient vector \( \mathbf{\beta} \). In the following regression analysis, suppose the size of the training sample is \( s \). After the \( s \) data points are substituted in Eq. (6), we obtain an \( s \)-dimensional simultaneous equation given as:
Stochastic model updating using distance discrimination analysis

The features predicted by the agent model and FE model, consequently transmitted to the next iteration step. The root mean square error (RMSE) is proposed as an index for various case studies. The evaluation of $\beta_{\text{test}}$, a real value less than 1, is not a fixed value but mainly depends on the context, the ratio $h$ gets lower, the iterative procedure achieves a higher convergence speed; however, the fidelity of the calibrated model predictions to the experiments degrades. A high $h$ value during each iteration, the sample of analytical features $Y_{\text{analysis}}$ is generated after advanced MC. Matrix $Y_{\text{analysis}}$ has a similar composition as $Y_{\text{test}}$, as assembled in Eq. (3). To quantify the degree of similarity (and dissimilarity) between $Y_{\text{test}}$ and $Y_{\text{analysis}}$, the distance between $Y_{\text{test}}$ and each row vector of $Y_{\text{analysis}}$ is calculated.

As shown in Fig. 4, all distances are calculated and sorted in ascending order during each iterative step. A truncation ratio $\theta$ is defined prior to the procedure to reserve a certain amount of analytical points which are “closer” to $Y_{\text{test}}$. In this context, the ratio $\theta$, a real value less than 1, is not a fixed value for various case studies. The evaluation of $\theta$ mainly depends on the initial FE-model’s closeness to the physical structure.

As $\theta$ gets lower, the iterative procedure achieves a higher convergence speed; however, the fidelity of the calibrated model predictions to the experiments degrades. A high $\theta$ value

Eq. (9) can be written in a vector form as

$$y = Z\beta$$

where $y \in \mathbb{R}^{n \times 1}$, $Z \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^{p \times 1}$.

The least squares or maximum likelihood estimate method can be used in the regression analysis to obtain an unbiased estimate of $\beta$

$$\beta = (Z^T Z)^{-1} Z^T y$$

The last step of agent modeling is precision evaluation. In order to evaluate the agent model’s precision compared with the FE analysis, some FE data points independent of the training sample are required, termed as the “evaluation sample”. The root mean square error (RMSE) is proposed as an index with an expression as

$$\text{RMSE} = \sqrt{\frac{1}{s'} \sum_{i=1}^{s'} (y_i - y'_i)^2}$$

where $s'$ is the size of the evaluation sample, and $y_i$ and $y'_i$ are the features predicted by the agent model and FE model, respectively.

2.3. TAC and iterative procedure

The proposed iterative procedure for uncertainty propagation, TAC and parameter calibration is shown in Fig. 3. A set of precise parameters and a satisfactory forecast of features compared with the test measurement is then obtained with acceptable calculation and time cost.

As shown in Fig. 3, a response surface model is used to separate the MC procedure into two parts. First, a certain amount of FE-runs is executed to generate a number of pairs of FE data points, which are used to generate the low-order model. An increased number of FE-runs can cover the input space more completely and get a better coverage in the response surface model, of course, with more calculation time and cost.

The response surface model is retrained in every iterative step during which new FE data is obtained based on the previous iteration. As the additional data approaches the practical value, the response surface model becomes more precise on the actual data interval of the structure system. With the use of the fast-running response surface model, large-scale MC becomes possible.

As the core of the iterative procedure, DDA based correlation, illustrated in Fig. 4, is performed to calibrate parameters of the initial FE model. By exploiting the results of DDA, the mean value of the parameter sample is calibrated and subsequently transmitted to the next iteration step.
will cause the procedure to converge slowly and consequently need more calculation time. Typically, if the initial model dissimilarity is obvious, a smaller $\theta$ value is required; on the contrary, a higher value is appropriate. Based on a huge amount of engineering experience on various structures and models, the truncation ratio $\theta$ typically falls in the interval of [0.01, 0.10].

After the value of $\theta$ is determined, the size of the analytical data sample is truncated to

$$n' = n \theta,$$  \hspace{1cm} (13)

where $n$ and $n'$ are the size of the data sample before and after truncation, respectively. As soon as the truncated feature sample is obtained, the corresponding input points sample is subsequently identified, an illustration of which is provided in Fig. 4. The mean of the reserved parameters sample, obtained through moment estimation or maximum likelihood estimation, is assumed as

$$\bar{x}^{(j+1)} = \frac{1}{n'} \sum_{k=1}^{n'} x_k^{(j)},$$  \hspace{1cm} (14)

where $\bar{x}^{(j+1)}$ is the mean of the reserved sample waiting to be utilized in the next iterative step; $x_k^{(j)}$ is the $k$th point of the parameter sample.

In the next iterative step, a new round of MC is performed with the calibrated parameter mean. It is assumed that the distribution of parameter is identical during each of the iterations, while variance indicates that the parameter dispersion is changing. The variance depends on the degree of dissimilarity between the initial FE prediction and the test measurement. If the initial dissimilarity is obvious, a greater variance of input sample is required. In order to quantitatively control this value, the authors propose coefficient of variation (CV), ratio of the standard deviation to the mean, during the iterative procedure. The CV herein is typically within the interval of [0.01, 0.10]. As CV gets higher, the iterative procedure achieves a higher convergence speed; however, the fidelity of the calibrated model predictions to the experiments degrades. A small CV value will cause the procedure to converge slowly and consequently need more calculation time.

A convergence of the iterative process occurs when the qualifications listed below are satisfied to a desired degree.

$$\begin{align*}
\bar{x}^{(j+1)} &\cong \bar{x}^{(j)} \\
\delta y_{j}^{(j)} &\cong 0 \\
\Delta y / y_{\text{test}} &\leq 5\%
\end{align*}$$  \hspace{1cm} (15)

where $\Delta y$ is the feature deviation between the FE prediction and test measurement, and $\delta y^{(j)}$ the distance from $Y_{\text{analysis}}$ to $Y_{\text{test}}$ in the $j$th iteration. These qualifications respectively indicate the mean value of parameter is no longer changing among different iterations; the output error (i.e. the distance) is approximating to a minimum and no longer changing; the analytical features have a sufficient agreement with the test measurements. A uniform standard for “sufficient” has not been presently developed, but there are examples of particular application fields that maintain standards for model predictive accuracy.$^{10,23}$

Herein the authors propose an error tolerance below 5%.

### 2.4. Various distance functions in DDA

Calculate the distance between $Y_{\text{test}}$ and each data point of $Y_{\text{analysis}}$ is a critical aspect of the iterative procedure. There are various distance functions to calculate the statistical distance, which are considered as the correlation metric. TAC involves the definition of an appropriate metric to facilitate the most efficient quantification of the correlation. Let $y^t$ denote each row vector of $Y_{\text{analysis}}$ and $\bar{y}$ denote the mean vector of $Y_{\text{test}}$. Some common distances are typically utilized in DDA:

**Absolute distance**

$$d(y, Y_{\text{test}}) = \sum_{i=1}^{q} \vert y_i - \bar{y} \vert$$  \hspace{1cm} (16)

**Euclidian distance (ED)**

$$d(y, Y_{\text{test}}) = \left[ \sum_{i=1}^{q} (y_i - \bar{y}_i)^2 \right]^{\frac{1}{2}}$$  \hspace{1cm} (17)

**Chebyshev distance**

$$d(y, Y_{\text{test}}) = \max_{1\leq i \leq q} \{ y_i - \bar{y}_i \}$$  \hspace{1cm} (18)

When considering the correlations among different features, the following distance functions may also be applicable:

**Variance weighted distance**

$$d(y, Y_{\text{test}}) = \left[ \sum_{i=1}^{q} (y_i - \bar{y}_i)^2 s_j^2 \right]^{\frac{1}{2}}$$  \hspace{1cm} (19)

where $s_j^2$ is variance of the $j$th output variable which can be obtained from Eq. (4).

**Mahalanobis distance (MD)**

$$d(y, Y_{\text{test}}) = (y - \bar{y})^T C^{-1} (y - \bar{y})$$  \hspace{1cm} (20)

where $C^{-1}$ is inverse of covariance matrix $C$ in Eq. (5).

### 3. Simulation example and discussion

#### 3.1. Description of FE model and synthetic test data

An FE model of a helicopter airframe$^{24}$ is employed in this simulation as shown in Fig. 5. The refined FE model is composed of 3905 nodes and 8167 elements. Decomposition views of this model are also shown in Fig. 5 where these two parts of frames are designed to have different material properties. Four parameters are selected to be calibrated and the first six natural frequencies are selected as features. A description of the parameters is presented in Table 1.
In this example, the “test” data is also simulated by providing nominal values for the parameters. These parameters are assumed to obey Gaussian distributions based on engineering experience and material properties. To incorporate variability in the test, noises are added to the parameter sample by assigning standard deviations of these distributions. The given distributions and numerical characteristics of the test parameters are detailed in Table 1.

A random sample with 50 parameter points obeying the specific distributions is proposed in this example. The responding feature sample is generated by executing the FE analysis 50 times using the 50 sets of parameters. This synthetic experimental feature sample is then utilized to estimate the probabilistic characteristics of the feature population in the following TAC and parameter calibration process.

3.2. Response surface modeling

Prior to the MC procedure, it is necessary to construct a response surface model to replace the time-consuming FE model. As shown in Table 1, intervals of these parameters are determined based on structural characteristics and engineering judgments. Herein, a 4-parameters-8-levels orthogonal Latin square design is used leading to a total 64 samples within this parameter space. After the training sample is generated, quadratic multinomial-based agent models between the 6 features and the 4 parameters are trained.

The response surface model between the first two parameters and the first feature is shown in Fig. 6. As it can be seen from the figure, the relationship between the feature and parameters displays an obvious nonlinearity. Before the response surface model can be utilized, precision evaluation is required. Herein, 20 FE data points independent from the training sample are employed. The maximum, minimum, and mean absolute errors of each frequency are detailed in Table 2. It can be seen that the mean error of most frequencies is less than 1%, except the 5th and 6th frequencies. The maximum errors of all frequencies are less than 5%. In practice, the required level of precision for a response surface model is a question on which there has not been an agreed standard. However, there are particular examples\(^{21,25}\) of response surface models’ applications which suggest that the agent model proposed here is appropriate.

3.3. Parameter calibration

The initial parameter values, shown in Table 3, differ from the synthetic test values in that the responding initial features have a significant error. Hence, parameter calibration is necessary to fine-tune the nominal parameter values. In this case study, the value of the truncation ratio \(\theta\) and CV are respectively assigned as 0.01 and 0.05. The iterative procedure converges after 25 iterations. Both the nominal and calibrated values of the parameters are provided in Table 3.

Herein, two distance functions, ED and MD are implemented as correlation metrics. The mean values of the updated features with these two metrics are provided in Table 3. The data in the penultimate column in Table 3 is the final result using ED as a metric, and the last column presents the updated data with MD. For each of the iterative steps, Figs. 7 and 8 respectively present the variation tendency of parameters and features in comparison to the test data.

As shown in Fig. 7, both iterative procedures yield acceptable precision. The results obtained with ED as a correlation metric, however, are closer to the test data than the results

---

**Table 1** Description of input variables in helicopter airframe case.

<table>
<thead>
<tr>
<th>Parameters Description</th>
<th>Nominal mean</th>
<th>Nominal standard deviation</th>
<th>Distribution</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elastic modulus of annular frame (10^11 Pa)</td>
<td>2.1</td>
<td>0.021</td>
<td>Gaussian</td>
<td>[1.0, 2.5]</td>
</tr>
<tr>
<td>Elastic modulus of longitudinal frame (10^11 Pa)</td>
<td>0.7</td>
<td>0.007</td>
<td>Gaussian</td>
<td>[0.5, 1.5]</td>
</tr>
<tr>
<td>Mass density of helicopter skin (10^3 kg/m^3)</td>
<td>2.7</td>
<td>0.027</td>
<td>Gaussian</td>
<td>[2.0, 4.0]</td>
</tr>
<tr>
<td>Sectional area of inner support rod (10^-2 m^2)</td>
<td>5.0</td>
<td>0.050</td>
<td>Gaussian</td>
<td>[3.0, 9.0]</td>
</tr>
</tbody>
</table>

---

**Table 2** Precision evaluation of response surface model.

<table>
<thead>
<tr>
<th>Mode no.</th>
<th>Absolute error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maximum</td>
</tr>
<tr>
<td>1</td>
<td>1.63</td>
</tr>
<tr>
<td>2</td>
<td>0.43</td>
</tr>
<tr>
<td>3</td>
<td>1.84</td>
</tr>
<tr>
<td>4</td>
<td>2.63</td>
</tr>
<tr>
<td>5</td>
<td>3.47</td>
</tr>
<tr>
<td>6</td>
<td>4.65</td>
</tr>
</tbody>
</table>
Table 3  Data of initial and updated models.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Synthetic test</th>
<th>Initial</th>
<th>Updated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Results with ED</td>
</tr>
<tr>
<td>$E_1$ ($10^{11}$Pa)</td>
<td>2.1</td>
<td>1.2 ($-42.86$)</td>
<td>2.08 ($-0.95$)</td>
</tr>
<tr>
<td>$E_2$ ($10^{11}$Pa)</td>
<td>0.7</td>
<td>1.2 (71.43)</td>
<td>0.70 (0)</td>
</tr>
<tr>
<td>$\rho$ ($10^3$kg/m$^3$)</td>
<td>2.7</td>
<td>3.5 (29.63)</td>
<td>2.64 ($-0.22$)</td>
</tr>
<tr>
<td>$S$ ($10^{-3}$m$^3$)</td>
<td>5.0</td>
<td>7.0 (40.00)</td>
<td>4.97 ($-0.60$)</td>
</tr>
<tr>
<td>Absolute mean error (%)</td>
<td>45.98</td>
<td>0.44</td>
<td>3.90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Synthetic test</th>
<th>Initial</th>
<th>Updated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Results with ED</td>
</tr>
<tr>
<td>$f_1$</td>
<td>10.41</td>
<td>8.86 ($-14.89$)</td>
<td>10.44 (0.29)</td>
</tr>
<tr>
<td>$f_2$</td>
<td>16.20</td>
<td>15.12 ($-6.67$)</td>
<td>16.19 ($-0.06$)</td>
</tr>
<tr>
<td>$f_3$</td>
<td>16.69</td>
<td>15.82 ($-5.21$)</td>
<td>16.69 (0)</td>
</tr>
<tr>
<td>$f_4$</td>
<td>23.01</td>
<td>16.72 ($-27.34$)</td>
<td>23.00 ($-0.04$)</td>
</tr>
<tr>
<td>$f_5$</td>
<td>26.34</td>
<td>19.66 ($-25.36$)</td>
<td>26.33 ($-0.04$)</td>
</tr>
<tr>
<td>$f_6$</td>
<td>34.19</td>
<td>28.96 ($-15.30$)</td>
<td>34.15 ($-0.12$)</td>
</tr>
<tr>
<td>Absolute mean error (%)</td>
<td>15.80</td>
<td>0.09</td>
<td>2.13</td>
</tr>
</tbody>
</table>

Note: Value in parenthesis are errors (%).

Fig. 7  Parameter mean value variation tendency with different metrics.
Fig. 8  Mean value variation tendency of $f_1$–$f_6$ with different metrics.
obtained with MD. As seen from Figs. 7 and 8, ED can accelerate the iterative procedure and converge faster than MD. Furthermore, the analytical data updated using these two metrics are observed to have different variation tendencies prior to convergence with the test data.

### 3.4. Scatter points

Here, we present the plots of scatter points in various feature planes to demonstrate the proposed iterative distance discrimination procedure. The scatter points of the first two natural frequencies \( f_1 \) and \( f_2 \) are presented in Figs. 9 and 10. As shown in Fig. 9, the cloud of light grey points represents 10000 MC samples generated with the initial FE model, with the light grey ellipse illustrating the 95% confidence interval of the cloud. The test points along with the 95% confidence interval are shown in black. Fig. 10 shows the data points along with the 95% confidence interval of the updated FE models using both ED and MD as correlation metrics. It can be seen in Fig. 10 that the central point of the ED ellipse is closer to the test than the ellipse with MD, indicating a greater precision of the updated outcome with ED than the results with MD, as compared to the mean value of the synthetic test.

Fig. 11 show the convergence of the analytical and test scatter points in the plane of different output variables. The plots provide a visual explanation for the basic principle of the proposed iterative method. For different correlation metrics, the analytical scatter points follow different tracks prior to arrival at their final position. As seen in Fig. 11(e)-(f), the 2nd and the 3rd natural frequencies exhibit a strong linear relationship. This phenomenon is also evident in Fig. 8 where the two variables exhibit a consistent tendency of change during the iterative procedure.

This analysis clearly indicates the characteristic of ED in obtaining an updated FE model that is closer to mean value of the test data. Also the convergence speed of the procedure is more rapid with ED than a procedure using MD. This finding can potentially be explained by the fact that ED is a directly geometric distance between the analytical points to the central point of the test sample, while MD is a weighted distance that considers the correlations among the features.

### 4. Conclusions

Both uncertainties in FE models and variability in testing can be taken into account by integrating DDCs and MC sampling into the stochastic model updating method. Because of the techniques of response surface model and design of experiment, outcomes with satisfied fidelity are available with an acceptable calculation cost. This proposed method makes it possible to analyze and compare the influence of different metrics on the iterative model calibration process and the calibrated parameters. The results of the helicopter case study clearly indicate that the ED can generate a more precise outcome with respect to mean values. It does not, however, necessarily indicate that the ED is a superior metric than the MD when aspects such as variable correlations and the degree of variable dispersion are considered. The integrated use of various correlation metrics has the potential to yield improved model calibration outcomes and thus must be studied further.
Figure 11: Variation tendency of scatter ellipses.
Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 10972019), the Innovation Foundation of BUAA for Ph.D. Graduates of China, and the China Scholarship Council.

References