Distance Metrics with Varying Statistical Information in Model Calibration and Validation

Sifeng Bi* and Scott Cogan†
Femto-ST Institute, University of Bourgogne Franche-Comté, Besancon 25000, France

Saurabh Prabhu‡ and Sez Atamturktur§
Clemson University, Clemson SC 29632

Test-analysis comparison metrics are mathematical functions that provide a quantitative measure of the agreement (or lack thereof) between numerical predictions and experimental measurements. While calibrating and validating models, the choice of a metric can significantly influence the outcome, yet the published research discussing the role of metrics, in particular varying levels of statistical information the metrics can contain, has been limited. This manuscript calibrates and validates the model predictions using alternative metrics formulated based on three types of distance-based criteria: (i) the Euclidian distance, i.e. the absolute geometric distance between two points; (ii) the Mahalanobis distance, i.e. the weighted distance that considers the correlations of two point clouds; and (iii) the Bhattacharyya distance, i.e. the statistical distance between two point clouds considering their probabilistic distributions. A comparative study is presented in the first case study, where the influence of various metrics, and the varying levels of statistical information they contain, on the predictions of the calibrated models is evaluated. In the second case study, an integrated application of the distance metrics is demonstrated through a cross validation process with regard to the measurement variability.

---

* Post-doc research fellow, Department of Applied Mechanics, sifeng.bi@femto-st.fr, AIAA member
† Third author. Senior Research Fellow, Department of Applied Mechanics
‡ Second author. Post-doc research fellow, Glenn Department of Civil Engineering
§ Professor, Glenn Department of Civil Engineering

Pre-print submitted to AIAA J

Nomenclature

BC = *Bhattacharyya coefficient*

$\mathbf{C}$ = *covariance matrix*

d_E = *Euclidian distance*

d_{M1} = *pont-to-population Mahalanobis distance*

d_{M2} = *population -to- population Mahalanobis distance*

d_B = *Bhattacharyya distance*

$\mathbf{e}_m$ = *unit vector with m dimension*

$f(\cdot)$ = *simulator in an uncertain system*

l, m = *number of points in the feature sample*

P = *probability density function*

p = *number of parameters in an uncertain system*

q = *number of features an uncertain system*

s^2 = *variance value of one feature variable*

u = *number of bins in Bhattacharyya distance calculation*

$\mathbf{X}$ = *parameter matrix*

$\mathbf{x}$ = *observation vector of one parameter variable*

$\mathbf{Y}$ = *feature matrix*

$\mathbf{y}$ = *feature vector*

β = *coefficient of the agent model*

μ = *mean of normal distribution*

σ^2 = *variance of normal distribution*

Subscripts

e = *experimental data*

i, j, k = *index of parameters and features*

n = *numerical analyzed data*

I. Introduction

Numerical models, widely used for the design, optimization, and assessment of aerospace engineering systems, are approximate representations of reality in that their predictions exhibit a level of disagreement from experimental measurements. The main sources of this disagreement can be classified as: *parameter uncertainties* due to imprecisely known input parameters (e.g. Young's modulus, mass density, geometric size, and spring stiffness) introducing uncertainties to the model predictions; *model form bias* due to unavoidable simplifications and idealizations (e.g. linearized representations of nonlinear behavior, frictionless joint approximation) introducing systematic errors to the model predictions; and *test variability* due to hard-to-control random effects (e.g. environment noise, measurement uncertainties) making the experimental measurements only partially reproducible [1]. These first two sources are related to the model and can be mitigated through the well-known process of model calibration, which infers the *likely* parameter values and model bias from experimental measurements that improve the agreement between predictions and measurements [2, 3]. The calibrated models become conditioned upon the data that are calibrated against and therefore must be checked against an independent set of measurements for *model validation* [4-6].

As a concise summary of the vast research on model calibration, this paper classifies the published calibration techniques into three paradigms depending on how uncertainties are characterized: (i) deterministic, (ii) stochastic, and (iii) robust calibration. Paradigm (i) includes the traditional sensitivity-based procedures which involve adjusting a nominal computer model based on a single set of experimental measurement [7, 8]. Over the last two decades, the significant role of uncertainty has been widely acknowledged [9, 10] allowing paradigms (ii) and (iii) to gain a growing research interests. Stochastic approaches to deal with uncertainty include Monte Carlo simulation (MCS) [11], covariance adjustment [12], and Bayesian calibration [13, 14]. The recently developed robust calibration [15, 16] formulates model calibration as a multi-objective problem with two distinct objectives: fidelity and robustness. Under this approach, the model is calibrated with the maximum allowable uncertainty, while providing acceptable fidelity to the measurements.

Regardless of which paradigm is proposed, one must establish an appropriate, quantitative *test-analysis comparison metric* to assess the degree of similarity (or dissimilarity) between predictions and measurements [10, 17]. There are three basic aspects of this assessment [18]: (i) the quantities to be compared (i.e. system response features); (ii) the manner in which the comparison is made (i.e. mathematical function); and (iii) acceptable accuracy for the comparison (i.e. model adequacy). The focus of this paper is the second aspect, while the fact is all three aspects are interrelated. For instance, the response feature of interest can be a scalar value or vector representation (i.e. a series of natural frequencies or temporal/spatial variations), which in turn would dictate the mathematical functions to be used for comparison. Due to the widely varying types of response features, a unified metric is currently unavailable in published literatures [17].

Earlier studies [1, 4, 19] termed the general comparison between predictions and measurements as (test-analysis) *correlation*. In the field of statistics, the concept of correlation refers to the dependency relationship, more specifically the linear relationship, between two variables [20]. In this context, correlation metrics refer in particular to the degree of linear relationship between predictions and measurements. Classical correlation metrics in vibration testing are modal assurance criterion (MAC) [21] for modal vectors and frequency response assurance criterion (FRAC) [22] for frequency response functions (FRFs). This paper, on the other hand, proposes using "*distance* metrics" to compare predictions with measurements, without focusing solely on correlations. Note that the distance studied herein is not only the geometric distance, but also generalized distance quantifying the difference between statistical populations. When information such as correlation, covariance, and distribution of the population is considered, another key characteristic of metric raises: *deterministic* or *stochastic*.

The mathematical function used for a metric can be deterministic or stochastic in nature. Test variability, along with the uncertainties associated with input parameters and model form makes deterministic approaches prone to misleading comparisons [1]. Hence, it is necessary to consider uncertainty during the comparison of predictions and measurements. Furthermore, it is valuable to evaluate the influence of various uncertainty levels on the calibration results, which is also the objective of this paper. Various distance concepts can be found in the literature from the simplest geometric distance to concepts containing more statistical information such as Bhattacharyya and Kullback-Leibler distances. Among this wide range, three classical and distinct distance-based criteria, i.e. Euclidian, Mahalanobis, and Bhattacharyya distances are compared here, as they are well-known, commonly used, and easily applicable for test-analysis comparison. More importantly, these criteria contain increasing levels of statistical information by calculating the distance as point-to-point, point-to-population, and population-to-population.

The Euclidian distance, perhaps the most widely used distance function, focuses only on the absolute geometric distance between two points. If information regarding the variances of populations is available, one can use the Mahalanobis distance [23], a weighted distance that considers the variance of the point clouds. If the entire probability distributions of populations are available, then one can use the Bhattacharyya distance [24], which provides a distance between two point clouds considering their probabilistic characteristics. Based on the above three distance criteria, this paper proposes five distance metrics whose principle, application and impact on model calibration and validation are elaborated through two case studies focusing on a laboratory scale steel frame.

The main contributions of this paper includes: illustrating the impact of increased statistical information on the model calibration results; demonstrating the integrated application of both deterministic and stochastic metrics in a cross validation process; and proposing a MCS based approach for model calibration which is capable of using metrics with different levels of statistical information. Case studies with both simulated and real measurements are presented, after which key conclusions of the comparison study are discussed.

II. Mathematical formulations

A. Statistical description of the uncertain structural system

In this paper, the model of a structural system is calibrated using a stochastic approach, in that the uncertain input parameters of the model are represented as random variables obeying given probabilistic distributions [9, 20]. Such distributions of parameters are determined based on design manual, modelling characteristics, and engineering expertise. The uncertain system is characterized by three major elements: input parameters **x**, output features **y**, and simulator $f(\cdot)$ [25]:

$$\mathbf{y}_i = f_i(\mathbf{x}_1,\ \mathbf{x}_2,\ \cdots,\ \mathbf{x}_p),\quad \forall i = 1,\ 2,\ \cdots,\ q \tag{1}$$

where i is the index of the features; p and q are respectively the dimensionality of **x** and **y**. The simulator $f(\cdot)$ is usually given as a sophisticated computer code (e.g. finite element model) or simplified function relationship (agent model).

In deterministic calibration, measurements are often represented by only one observation of the features. The stochastic framework however allows the consideration of a population of feature observations in the measurements. Assuming *m* measurements are available, the feature sample is expressed as a matrix $\mathbf{Y} \in \Re^{m \times q}$:

$$\mathbf{Y} = [\mathbf{y}_1 \ \mathbf{y}_2 \ \cdots \ \mathbf{y}_q] \tag{2}$$

where $\mathbf{y}_i = (y_{1i}, y_{2i}, \cdots, y_{mi})^{\mathrm{T}}, \ \forall i = 1, 2, \cdots, q.$

In the following discussions, the subscripts *e* and *n* respectively denotes *experimental* measurements and *numerical* predictions. $\mathbf{Y}_e$ is the feature sample obtained through repeated measurements, while $\mathbf{Y}_n$ is predicted by numerous simulator analysis. Ideally, a sufficient number of identical prototypes of the structure are constructed using identical materials and procedures but with inevitable manufacturing tolerances and material heterogeneity. $\mathbf{Y}_e$ obtained from this multi-structure multi-measurement strategy can represent various sources of test variability. The random effect of environment noise and measurement errors is represented by the repeated experimental measurements. The remaining inherent variation of the physical system can be numerically matched using MCS where the parameter configurations are randomly sampled from their distribution [26]. The resulting features $\mathbf{Y}_n$ is a random sample obeying a feature distribution.

The exact probabilistic distribution of the feature population is difficult to determine. However, observation samples of the population can be utilized to estimate the probabilistic characteristics via either moment estimation or maximum likelihood estimation techniques. The mean $\bar{y}$ and variance s^2 of the *i*-th feature are obtained as

$$\begin{aligned} \bar{y}_i &= \frac{1}{m}\sum_{k=1}^{m} y_{ki} = \frac{1}{m}\mathbf{y}_j^{\mathrm{T}}\mathbf{e}_m; \\ s_i^2 &= \frac{1}{m-1}\sum_{k=1}^{m}(y_{ki} - \bar{y}_i)^2 \\ &= \frac{1}{m-1}(\mathbf{y}_i - \bar{y}_i\mathbf{e}_m)^{\mathrm{T}}(\mathbf{y}_i - \bar{y}_i\mathbf{e}_m) \end{aligned} \tag{3}$$

where $\mathbf{e}_m = (1, \cdots, 1)^{\mathrm{T}}; \mathbf{e}_m^{\mathrm{T}}\mathbf{e}_m = m.$ The mean vector $\bar{\mathbf{y}} \in \Re^{q \times 1}$ and covariance matrix $\mathbf{C} \in \Re^{q \times q}$ of the feature sample are determined as

$$\begin{aligned} \bar{\mathbf{y}} &= \frac{1}{m}\mathbf{Y}^{\mathrm{T}}\mathbf{e}_m; \\ \mathbf{C} &= \frac{1}{m-1}(\mathbf{Y} - \mathbf{e}_m\bar{\mathbf{y}}^{\mathrm{T}})^{\mathrm{T}}(\mathbf{Y} - \mathbf{e}_m\bar{\mathbf{y}}^{\mathrm{T}}). \end{aligned} \tag{4}$$

Once the probabilistic characteristics of the feature sample are estimated, various statistical distances between $\mathbf{Y}_e$ and $\mathbf{Y}_n$ can be computed as test-analysis comparison metrics.

B. Stochastic parameter calibration

The parameters are calibrated to obtain an associated numerical model so that the distance between the model prediction and the measurement is minimized. In this sense, parameter calibration is essentially an optimization problem where the uncertain parameters are the control variables and the distance metric is the objective function. Commonly used optimization methods, such as simplex algorithm and gradient-based algorithm, can be employed to optimize the deterministic metrics. However, these methods are unlikely to be effective for all metrics, especially for stochastic metrics based on the Mahalanobis and Bhattacharyya distances. The stochastic nature of the metrics leads to a lack of smoothness of the objective function, which renders the traditional derivative-based method unstable. Consequently, this paper proposes a MCS stochastic calibration technique, which is effective and robust when utilizing various statistical distances as metrics.

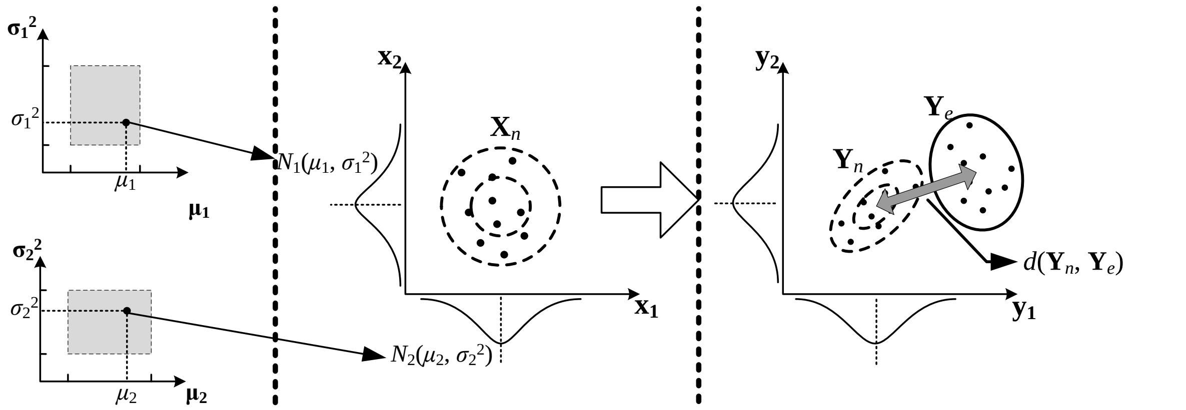
The MCS methodology is commonly used in the field of uncertainty analysis and some of the associated techniques (e.g. the return mapping approach or stepping technique [27]) have been successfully applied on practical engineering problems [28]. The return mapping approach is essentially an iterative modification method where the parameter dispersion is preset by a user-defined variation coefficient (ratio between the standard deviation and the mean) in each iteration. As a result, only the mean of the parameter is calibrated [29]. This paper expands the return mapping approach in two ways: first, the method proposed herein is no longer an iterative method but an approach containing two levels of MCS, which means the principle is more simple and direct; second, not only the mean but also the variance of the parameter can be calibrated with this proposed method.

This two level MCS stochastic calibration approach is illustrated in Fig. 1 where only two parameters and two features are presented for clarity without loss of generalizability. In this approach, the mean μ_i as well as the variance σ_i^2 of $\mathbf{x}_i$ is calibrated (i is either 1 or 2 in the following context). Intervals of μ_i and σ_i^2 are required as the starting point of the calibration. This approach is performed following the major steps:

Step I Randomly sample the configurations of (μ_i, σ_i^2) within their intervals.

Step II For each sampled data pair (μ_i, σ_i^2), a parameter sample $\mathbf{X}_n$ is randomly drawn from their distributions. Run the simulator (recall Eq. 1) for each parameter configuration inside $\mathbf{X}_n$ and the resulting features constitute the feature sample $\mathbf{Y}_n$.

Step III Calculate the distance metrics between $\mathbf{Y}_n$ and $\mathbf{Y}_e$. The specific instance (μ_i, σ_i^2) in **Step I** that gives the minimum metric value is identified and adopted as the calibrated value.



(a) **Step I**: The 1st level MCS (b) **Step II**: The 2nd level MCS (c) **Step III**: The calculation of distance metrics

Fig. 1 The MCS stochastic calibration logic

The shaded squares in Fig. 1(a) denote the intervals of μ_i and σ_i^2. The 1st level of MCS is performed within the shaded squares, and observations of the data pair (μ_i, σ_i^2) are obtained. No matter what probabilistic distribution $\mathbf{x}_i$ obeys, the uniform distribution is employed for the 1st level of MCS, so that every data pair (μ_i, σ_i^2) in the shaded squares can be searched with equal probability.

As shown in Fig. 1(b), the 2nd level of MCS is performed on each sampled (μ_i, σ_i^2) associated with the distribution of $\mathbf{x}_i$. The normal distribution $N_i(\mu_i, \sigma_i^2)$ is taken as an example. Once the parameter sample $\mathbf{X}_n$ is randomly drawn, the simulator is run on each parameter configuration inside $\mathbf{X}_n$ and the resulting features constitute the feature sample $\mathbf{Y}_n$. The uncertainties in $\mathbf{X}_n$ are thus propagated as the distribution of the feature sample $\mathbf{Y}_n$.

In Fig. 1(c), The distance between $\mathbf{Y}_n$ and $\mathbf{Y}_e$ is calculated as $d(\mathbf{Y}_n, \mathbf{Y}_e)$, where $d(\cdot)$ denotes various distance metrics (which will be discussed in following subsection). At the end of the process, the data pair (μ_i, σ_i^2) with the minimum distance is identified as the calibrated value for that particular metric. A large number of numerical analyses are required in the process. For example, when the sample size in the 1st and 2nd levels of MCS are respectively 10^4 and 10^3, then totally 10^7 analyses are required. If the simulator is a computationally expensive model, the calculation cost can become unacceptable. Techniques such as design of experiments and agent modeling (i.e. response surface modeling, metamodeling) [4, 30] are suggested to reduce the computational burden.

This two level MCS process is specifically designed to allow for the calibration of not only means but also variances. If only the mean of a parameter needs to be calibrated, the 2nd level of MCS is not required. In this case, the parameter sample $\mathbf{X}_n$ in Fig. 1(b) becomes unnecessary and only the single mean point $\bar{\mathbf{x}}_n$ is required. Thus, instead of computing the sample $\mathbf{Y}_n$, only its mean point $\bar{\mathbf{y}}_n$ is calculated in Fig. 1(c). The distance between $\bar{\mathbf{y}}_n$ and $\mathbf{Y}_e$ is calculated with either the Euclidian distance (point-to-point) or the Mahalanobis distance (point-to-population). In other words, only when the variance needs to be calibrated, the population-to-population distance is calculated in **Step III** where the random sample $\mathbf{Y}_n$ is required.

C. Distance-based test-analysis comparison metrics

Three distance criteria (i.e. the Euclidian, Mahalanobis, and Bhattacharyya distances) are presented to calculate the test-analysis metric $d(\mathbf{Y}_n, \mathbf{Y}_e)$ in Fig. 1(c). The proposed metrics are either point-to-point, point-to-population, or population-to-population types, implying various levels of statistical information are involved. Overall five distinct metrics based on these three distance criteria are evaluated in Section III.

1. Euclidian distance

The Euclidian distance is an absolute geometric distance between two points. It is a commonly used distance function with an intuitive geometrical meaning and simple to calculate. It is expressed as

$$d_{\mathrm{E}}(\mathbf{y}_n, \mathbf{Y}_e) = \left[(\mathbf{y}_n - \bar{\mathbf{y}}_e)^{\mathrm{T}} (\mathbf{y}_n - \bar{\mathbf{y}}_e) \right]^{1/2} \tag{5}$$

where $\bar{\mathbf{y}}_e$ is mean vector of the experimental sample $\mathbf{Y}_e$; $\mathbf{y}_n$ is a row vector in the numerical sample $\mathbf{Y}_n$ (i.e. a single observation point). The Euclidian distance is a point-to-point distance in that Eq. 5 represents the geometric distance from a single numerical point to the center point of the experimental sample.

2. Mahalanobis distance

The Mahalanobis distance is a classical statistical distance widely used in cluster analysis and image processing [31]. This distance involves the covariance matrix of the data, implying it considers the correlation of variables in the feature sample [32]. In practical applications, this correlation relationship among features is general, and cannot be ignored within the stochastic framework. Hence the Mahalanobis distance is gaining increasing interest in uncertainty analysis and model validation [1, 33].

The Mahalanobis distance proposed herein can be further divided into point-to-population distance and population-to-population distance. The classical point-to-population Mahalanobis distance between the single numerical point $\mathbf{y}_a$ and the experimental sample $\mathbf{Y}_e$ is expressed as

$$d_{\mathrm{M1}}(\mathbf{y}_n, \mathbf{Y}_e) = \left[(\mathbf{y}_n - \bar{\mathbf{y}}_e)^{\mathrm{T}} \mathbf{C}_e^{-1} (\mathbf{y}_n - \bar{\mathbf{y}}_e) \right]^{1/2} \tag{6}$$

where $\mathbf{C}_e^{-1}$ is inverse of the covariance matrix of the experimental sample. Eq. 6 differs from the Euclidian distance in that it considers the correlations of features in the experimental sample. If the covariance matrix in Eq. 6 is an identity matrix, the Mahalanobis distance reduces to the Euclidian distance. Nevertheless, as the point-to-population Mahalanobis distance handles only one numerical point, it contains the covariance information only from the experimental sample but not from the numerical sample. Hence, when it is employed as a calibration metric, only the means of the parameters can be calibrated, and the variances are ignored. In this sense, the point-to-population Mahalanobis distance has a similar effect with the deterministic Euclidian distance.

The alternative population-to-population Mahalanobis distance is expressed as

$$d_{\mathrm{M2}}(\mathbf{Y}_n, \mathbf{Y}_e) = \left[(\bar{\mathbf{y}}_n - \bar{\mathbf{y}}_e)^{\mathrm{T}} \mathbf{C}_{pool}^{-1} (\bar{\mathbf{y}}_n - \bar{\mathbf{y}}_e) \right]^{1/2} \tag{7}$$

where $\bar{\mathbf{y}}_n$ is the mean vector of $\mathbf{Y}_n$. Eq. 7 is also termed as the “pooled” Mahalanobis distance as it contains the pooled covariance matrix $\mathbf{C}_{pool}$:

$$\mathbf{C}_{pool} = \frac{(l-1)\mathbf{C}_n + (m-1)\mathbf{C}_e}{l+m-2} \tag{8}$$

where l and m are respectively the number of points in $\mathbf{Y}_n$ and $\mathbf{Y}_e$; $\mathbf{C}_n$ and $\mathbf{C}_e$ are respectively the covariance matrices of $\mathbf{Y}_n$ and $\mathbf{Y}_e$. Unlike the point-to-population Mahalanobis distance, the calculation process of $\mathbf{C}_{pool}$ involves the variance information from both numerical and experimental samples [34] allowing the pooled Mahalanobis distance to calibrate the parameter variance.

3. Bhattacharyya distance

The Bhattacharyya distance, widely used in signal processing and pattern recognition [35], is a statistical distance capable of quantifying the dissimilarity between two probabilistic distributions [36]. The Bhattacharyya distance is applicable to any data samples regardless of their distribution functions [37]. This characteristic makes it specifically appropriate for the stochastic model calibration where the distribution of the feature cannot be exactly determined. The Bhattacharyya distance is expressed as

$$d_{\mathrm{B}}(\mathbf{Y}_n,\ \mathbf{Y}_e) = -\ln[\mathrm{BC}(\mathrm{P}_n,\ \mathrm{P}_e)] \tag{9}$$

where BC is the Bhattacharyya coefficient, quantifying the probabilistic overlap between P_n and P_e, the probability density functions (PDFs) of the numerical and experimental populations. For continuous distributions, the BC is expressed as

$$\mathrm{BC}(\mathrm{P}_n, \mathrm{P}_e) = \int_{\mathbf{Y}} \sqrt{\mathrm{P}_n(y)\mathrm{P}_e(y)}\,dy \tag{10}$$

where $\int_{\mathbf{Y}}$ denotes the integration over the feature space. When the number of features $q > 1$, P_n and P_e become q-dimension joint PDFs.

In practical applications, the exact distributions needed in Eq. 10 are likely to be unavailable. In such situations, the histogram formulation [36] is proposed to estimate the probability mass function (PMF) of a discrete distribution of the feature. The interval of the feature is first determined and subsequently divided into a pre-defined number of bins. This number, termed as u, can be approximately estimated as [38]:

$$u \approx 1.87(m-1)^{2/5} \tag{11}$$

where m is the number of points in the sample. The next step is to group the points into corresponding bins. Probability of the points belonging to a particular bin is determined as dividing the number of points in this bin by the total number of points. The discrete probability distribution can be listed once the probabilities of all bins are calculated. The Bhattacharyya coefficient between two discrete distributions is then defined as:

$$\mathrm{BC}(\mathrm{P}_n, \mathrm{P}_e) = \sum_{i=1}^{u} \sqrt{\mathrm{P}_n(y_i)\mathrm{P}_e(y_i)} \tag{12}$$

where $\mathrm{P}_n(y_i)$ and $\mathrm{P}_e(y_i)$ are respectively the PMFs in the i^{th} bin of the numerical and experimental samples.

For a multivariate distribution including q features, the calculation process becomes increasingly more complex due to the need for a joint PMF and the total number of elements in Eq. 12 is no longer u but u^q. When multiple features are considered in a study, the Bhattacharyya distance against a particular feature, namely the marginal Bhattacharyya distance, can be calculated using this feature's marginal distribution. In this case, $\mathrm{P}_n(y_i)$ and $\mathrm{P}_e(y_i)$ refer to the marginal PMFs of the particular feature.

D. Random sampling method for experimental sample preparation

The metrics based on the Mahalanobis and Bhattacharyya distances require sufficient experimental observations, which not always available in practical applications. The authors propose a probabilistic random sampling method based on a finite number of experimental measurements to generate a "semi-experimental" sample. As the ASME's verification and validation standard [39] indicates, a normal distribution of the uncertainty is a convenient and feasible assumption in the absence of sufficient experimental data. In this context, each feature is assumed to obey normal distribution. Furthermore, the correlation relationship among features is another consideration when generating the semi-experimental sample. A joint multivariate normal distribution of the whole feature space is proposed to represent the correlation.

Assuming only one prototype of the structure is available, a small amount of measurements constitute the original measurement sample. The mean vector $\bar{\mathbf{y}}$ and covariance matrix $\mathbf{C}$ of the original measurement sample can be estimated from Eq. 4. Then the joint PDF of the q-dimensional normal distribution is given as follows

$$\mathrm{PDF}(\mathbf{y}) = \frac{1}{\sqrt{|\mathbf{C}|(2\pi)^q}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\bar{\mathbf{y}})\mathbf{C}^{-1}(\mathbf{y}-\bar{\mathbf{y}})^{\mathrm{T}}\right\}, \tag{13}$$

where $\mathbf{y} = [y_1, y_2, \cdots, y_q]^{\mathrm{T}}$ is an arbitrary feature point belonging to the experimental population. Using the PDF, a semi-experimental sample containing any number of points is randomly sampled. It is demonstrated in Section IV that not only the distribution characteristic of each single feature, but also the correlation between different features is represented in the semi-experimental sample.

III. Case study I: Comparative application in parameter calibration

A. Description of the structure and the numerical model

The proposed structure is a two-storey frame constructed by vertical and horizontal beams with different cross-sections. All beams are made of A36 mild steel and connected by bolts, which meet the SAE-J429-Grade-5 standard. The laboratorial setup and geometry details of this frame are shown in Fig. 2. A simplified finite element (FE) model is developed in MSC.Nastran using one-dimensional elements. Detail information of the FE model is given in Table 1.

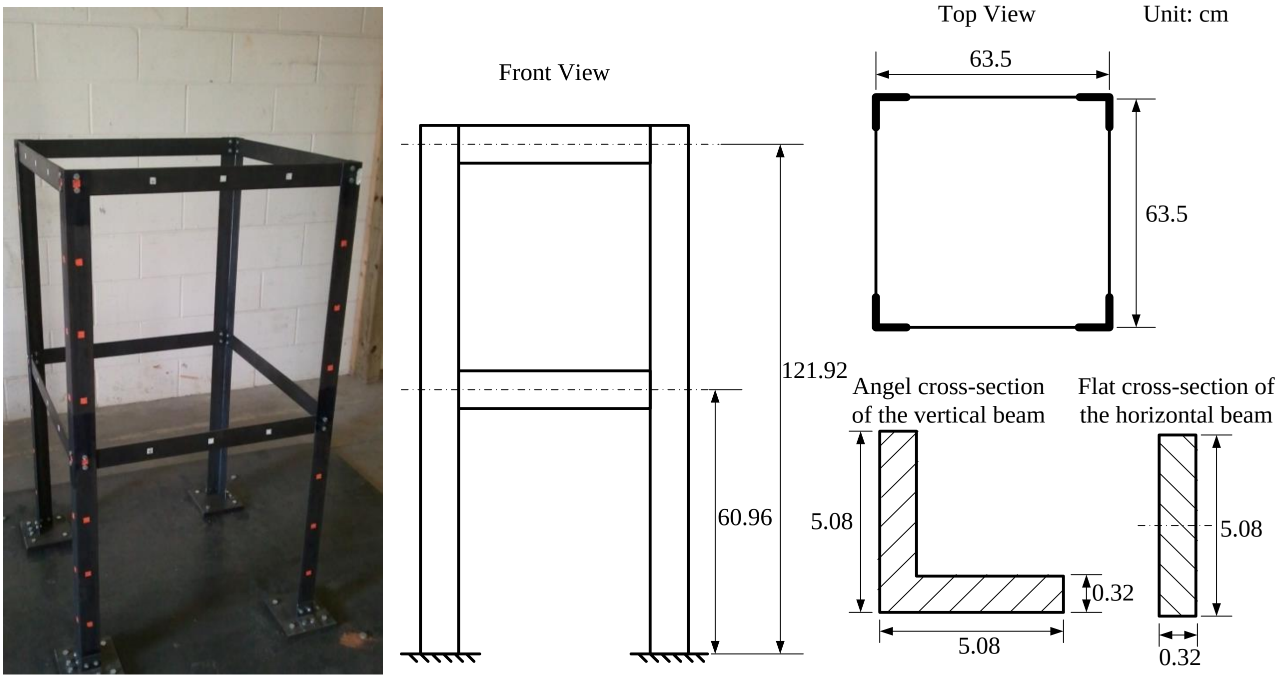


Fig. 2 The laboratoral setup of the frame structure and its geometry details

Table 1 Mesh information of the FE model in MSC.Nastran

Element type of the vertical beams (with angel cross-section)	CBEAM
Element type of the horizontal beams (with flat cross-section)	CBAR
Nodes number	316
Elements number	328

The difficulty of modeling the bolted connections results in significant model form uncertainties in their numerical representation. The vertical beams with angle cross-section are simplified by the one-dimensional elements, implying that the offset of the vertical centroid inside the cross-section is an important calibration parameter. The selected calibration parameters of this FE model are listed in Table 2. The parameters are assumed to follow normal distributions with both the means and the variances in need of calibration. The intervals on these values are determined based on characteristics of this structure, as detailed in Table 2. The first four natural frequencies of the structure are proposed as features.

Table 2 Calibration parameters of the FE model

Parameters	Mean Intervals	Variance Intervals	Descriptions
E_1 (10^{11}Pa)	[0.5, 2.5]	[0, 0.2]	Young's modulus of the elements of the upper connections
E_2 (10^{11}Pa)	[0.5, 2.5]	[0, 0.2]	Young's modulus of the elements of the middle connections
Offset (10^{-2}m)	[0.0, 3.0]	[0, 0.2]	Offset of the vertical elements
ρ (10^3Kg/m^3)	[7.0, 9.0]	[0, 0.2]	Overall mass density of the frame

As mentioned in Section II, in order to reduce the calculation burden of MCS approach, this example employs the quadratic polynomial agent model which is trained based on the FE model predictions. As the dimensionality of parameters and features are respectively four and three, the quadratic polynomial model between the k-th feature and the above four parameters is formulated as

$$y_k = \beta_0^k + \sum_{i=1}^{4} \beta_i^k x_i + \sum_{i<j}^{4} \beta_{ij}^k x_i x_j + \sum_{i=1}^{4} \beta_{ii}^k x_i^2 \qquad k = 1, 2, 3 \tag{14}$$

where k is the feature index; $x_{1\text{-}4}$ denote the four parameters; β is the undetermined coefficients. Application of this agent model follows the three typical steps:

1) Preparing the training sample from the FE predictions;

2) Estimating the undetermined coefficients using the training sample;

3) Assessing the precision of the agent model before it can be used.

Not actually being the focus of this paper, Steps 1) and 2) are omitted for clarity. A thorough description on the relative techniques can be found in Ref. [30]. For Step 3), the assessment result is given in Table 3 where five parameter configurations are randomly sampled within the intervals in Table 2. Percentage errors of the agent model predictions are evaluated with comparison to the FE predictions. As shown in Table 3, the mean absolute error (MAE) of each feature is extremely small, demonstrating that the quadratic polynomial agent model is feasible in the following calibration procedure.

Table 3 Precision assessment of the agent model (% error corresponding to the FE predictions in parentheses)

Assessment No.	Parameter configurations (x_1, x_2, x_3, x_4)	y_1 (Hz)		y_2 (Hz)		y_3 (Hz)		y_4 (Hz)	
		FE	Agent	FE	Agent	FE	Agent	FE	Agent
1	(1.50, 1.98, 1.62, 7.95)	20.61	20.62 (0.05)	27.46	27.81 (1.32)	30.62	30.52 (-0.33)	43.09	43.17 (0.20)
2	(1.07, 2.15, 2.86, 8.17)	21.78	21.76 (-0.11)	27.46	27.68 (0.78)	28.70	28.50 (-0.71)	38.93	38.92 (-0.03)
3	(2.16, 0.86, 0.63, 7.52)	20.15	20.19 (0.20)	26.55	26.47 (-0.29)	29.86	29.89 (0.11)	44.60	44.56 (-0.09)
4	(2.14, 0.63, 0.35, 7.17)	20.37	20.42 (0.27)	26.09	26.32 (0.91)	29.60	29.72 (0.42)	45.12	45.21 (0.19)
5	(2.38, 1.72, 1.94, 7.60)	21.53	21.53 (0.00)	29.60	29.82 (0.73)	31.11	31.22 (0.34)	44.00	44.09 (0.20)
Mean absolute errors			0.13%		0.81%		0.38%		0.14%

B. Parameter calibration

The simulated experimental feature sample is obtained by assigning nominal values to the parameter means (Table 4) and variances (Table 5), randomly generating a parameter sample containing 1000 points, and running a simulator (i.e. an agent model calculation using Eq. 14) at each sampled point. In the following description, the term "test" is utilized to denote the simulated experimental data, which serves as the target to be recovered through calibration. Another different set of parameter values are proposed as the initial values waiting to be calibrated. Due to the initial parameters being different from the test parameters, the initial features exhibit obvious errors compared with the test features. The mean absolute errors (MAE) of the initial feature sample in Table 4 (mean) and Table 5 (variance) are respectively 13.9% and 240%, implying a calibration process is required.

During the calibration, five distinct distance metrics are evaluated:

1. A point-to-point Euclidean distance between the numerical means and the test means, denoted Metric ED;
2. A point-to-population Mahalanobis distance between the numerical means and the test samples, denoted Metric MD-1;
3. A population-to-population Mahalanobis distance between the numerical and test samples, denoted Metric MD-2;
4. A population-to-population Bhattacharyya distance between the numerical and test samples, denoted Metric BD;
5. An integrated Euclidean and Bhattacharyya metric where the means are calibrated with the Euclidean distance and the variances are calibrated with the Bhattacharyya distance, denoted Metric ED-BD.

These metrics belong to either deterministic or stochastic categories depending on whether or not uncertainty in the numerical sample is considered. As point-to-point/population metrics, Metrics ED and MD-1 require only the numerical mean point to evaluate their values, implying only the parameter mean can be calibrated, and hence they become deterministic metrics. During the calculation of Metrics MD-2, BD, and ED-BD, numerical samples are generated based on the mean and variance. The randomly generated sample exhibits a slight difference in each calculation, yielding unequal calculated values even if the mean and variance are fixed. This is the reason for classifying Metrics MD-2, BD, and ED-BD as stochastic metrics.

The above deterministic metrics can be optimized using typical optimization methods such as simplex algorithm and gradient-based algorithm [40]. However, the calculation of Metrics MD-2, BD, and ED-BD involve a randomly generated sample. The stochastic nature of the metrics leads to a lack of smoothness of the objective function, which renders the standard local optimization methods inadequate. In the following evaluation, all of the five metrics are calibrated using the MCS stochastic calibration technique described in Section II.

Table 4 Means of the parameter and feature samples (% errors corresponding to the test values in parentheses).

		Test	Initial	Calibrated samples				
				ED	MD-1	MD-2	BD	ED-BD
Parameters	E_1 (10^{11}Pa)	2.0	1.0 (-50.0)	1.96 (-2.00)	1.84 (-8.00)	1.83 (-8.50)	2.39 (19.5)	1.96 (-2.00)
	E_2 (10^{11}Pa)	2.0	1.0 (-50.0)	1.93 (-3.50)	2.01 (0.50)	2.49 (24.5)	1.91 (-4.50)	1.93 (-3.50)
	Offset (10^{-2}m)	1.0	2.0 (100)	0.98 (-2.00)	1.03 (3.00)	1.09 (9.00)	0.94 (-6.00)	0.98 (-2.00)
	ρ (10^3Kg/m^3)	7.5	8.5 (13.3)	7.50 (0)	7.49 (-0.13)	7.56 (0.80)	7.46 (-0.53)	7.50 (0)
Mean absolute errors			53.3%	1.88%	2.91%	10.7%	7.63%	1.88%
Features	f_1 (Hz)	20.72	20.09 (-3.04)	20.70 (-0.10)	20.76 (0.19)	20.67 (-0.24)	20.75 (0.14)	20.68 (-0.19)
	f_2 (Hz)	28.44	24.69 (-13.2)	28.34 (-0.35)	28.32 (-0.42)	28.34 (-0.35)	28.62 (0.63)	28.28 (-0.56)
	f_3 (Hz)	32.55	26.24 (-19.4)	32.31 (-0.74)	32.31 (-0.74)	32.38 (-0.52)	32.80 (0.77)	32.43 (-0.37)
	f_4 (Hz)	46.49	37.29 (-19.8)	46.33 (-0.34)	46.25 (-0.52)	46.27 (-0.47)	46.90 (0.88)	46.38 (-0.24)
Mean absolute errors			13.9%	0.38%	0.47%	0.40%	0.61%	0.34%

Table 5 Variances of the parameter and feature samples (% errors corresponding to the test values in parentheses).

		Test	Initial	Calibrated samples				
				ED	MD-1	MD-2	BD	ED-BD
Parameters	E_1	0.12	0.05 (-58.3)	--	--	0.1871 (55.9)	0.1321 (10.1)	0.1061 (-11.6)
	E_2	0.02	0.15 (650)	--	--	0.1288 (544)	0.0350 (75.0)	0.0473 (137)
	Offset	0.12	0.05 (-58.3)	--	--	0.0476 (-60.3)	0.1927 (60.6)	0.0967 (-19.4)
	ρ	0.02	0.15 (650)	--	--	0.1510 (655)	0.0120 (-40.0)	0.0228 (14.0)
Mean absolute errors			354%	--	--	328%	46.4%	45.4%
Features	f_1	0.1133	0.2884 (155)	--	--	0.3216 (184)	0.1678 (48.1)	0.1124 (-0.79)
	f_2	0.5509	1.2248 (122)	--	--	1.3252 (141)	0.6847 (24.3)	0.5665 (2.83)
	f_3	0.3347	1.9514 (483)	--	--	1.1976 (258)	0.3903 (16.6)	0.3880 (15.9)
	f_4	0.9178	2.7696 (202)	--	--	2.2368 (144)	1.0890 (18.7)	0.8993 (-2.02)
Mean absolute errors			240%	--	--	181%	26.9%	5.39%

1. Results with Metric ED

Metric ED is calculated between the means of the numerical and test samples. The initial sample is calibrated towards the mean of the test sample without consideration of variance information. The calibration effect is illustrated by the positional relationship of the test, initial, and calibrated samples as shown in Fig. 3. The black, blue, and red point clouds are respectively the test, initial, and calibrated samples surrounded by ellipses which indicate the 95% confidence intervals of the samples. In the following context, the calibrated sample is termed as “Sample” followed by its corresponding metric. For example, the calibrated sample according to Metric ED is termed as Sample ED.

In Fig. 3, the positional relationship is separately illustrated in three sub-figures which represent three planes of f_1 *vs.* f_2, f_1 *vs.* f_3, and f_2 *vs.* f_3. For clarity, other planes (f_1 *vs.* f_4, f_2 *vs.* f_4, and f_3 *vs.* f_4) are omitted in the following context, but can be checked in Appendix A. The initial sample is shifted with respect to the test sample, while Sample ED is calibrated as its center point is mostly coincident with the test sample. Notice however, as the parameter variance remains unchanged, the 95% confidence interval ellipse of Sample ED remains different from the test sample.

The calibration result is also shown in Table 4 where MAEs of the calibrated parameters and features are respectively 1.88% and 0.38%, which are much smaller than the initial errors. Because Metric ED involves no information of variances, Sample ED has the same variance as the initial sample and consequently the variance of Sample ED is omitted in Table 5.

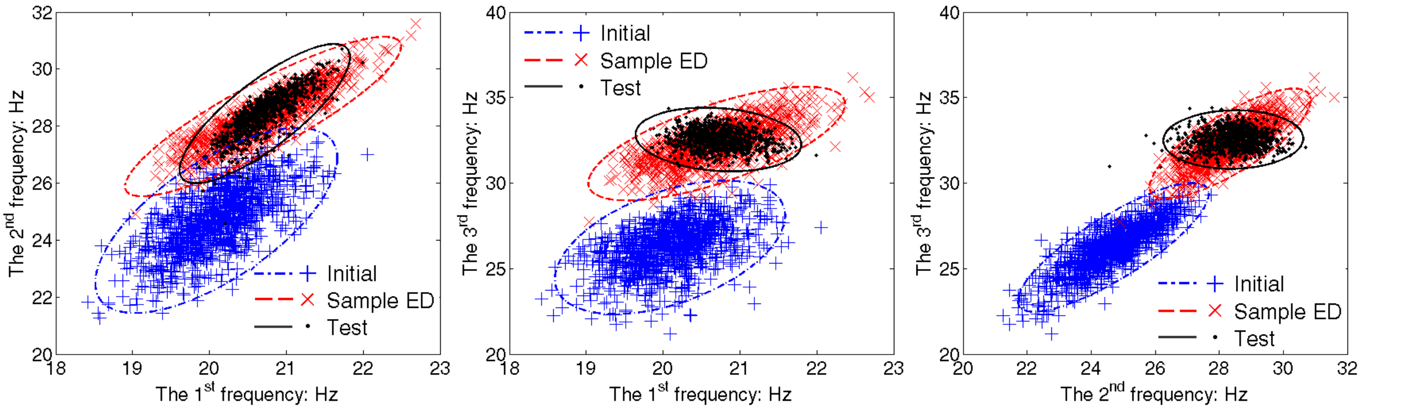


Fig. 3 The test, initial, and calibrated scatters in planes of different frequencies with Metric ED

2. Results with Metric MD-1

Metric MD-1 is the point-to-population Mahalanobis distance, where the point is the mean of the numerical sample and the population is the test sample. Metric MD-1 is actually a weighted Euclidean distance that considers the covariance only in the test sample and has no relationship with the numerical sample (recalling Eq. 6). Hence this metric is incapable of calibrating the parameter variance, which is similar to Metric ED.

As shown in Fig. 4, the center points of Sample MD-1 and the test sample are coincident, while their ellipse patterns are different. In Table 4, the MAEs of the calibrated parameters and features are respectively 2.91% and 0.47%, which are similar to the Metric ED results. This implies both the two metrics are capable of calibrating the means but not the variances.

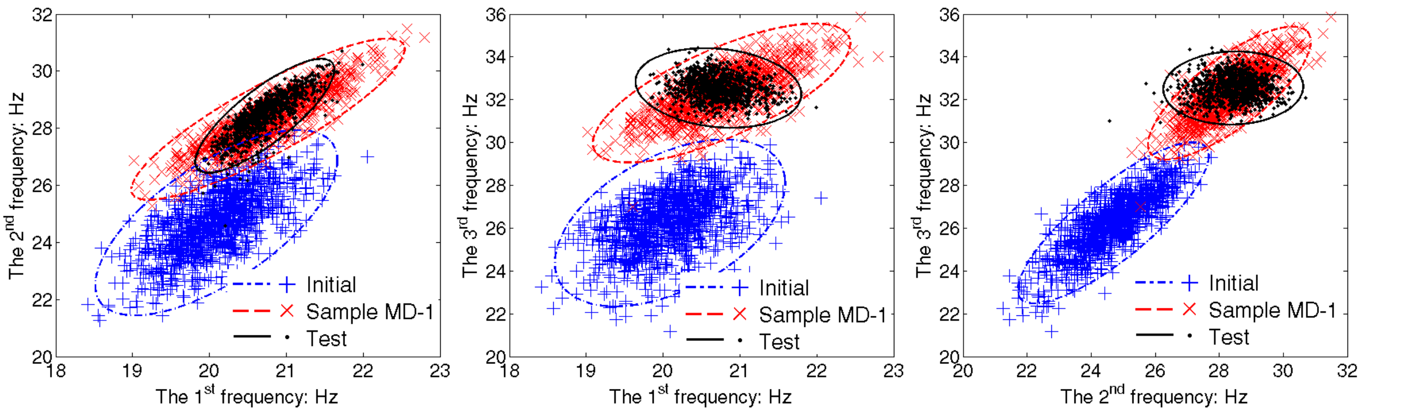


Fig. 4 The test, initial, and calibrated scatters in planes of different frequencies with Metric MD-1

3. Results with Metric MD-2

Unlike Metric MD-1, Metric MD-2 is the pooled Mahalanobis distance between two samples. The calculation process of Metric MD-2 involves the pooled covariance matrix which contains both test and numerical covariance information. It is thus possible to alter the variance of the numerical samples.

The MAEs of parameter and feature means are respectively 10.7% and 0.40%, as shown in Table 4. As for the variances in Table 5, MAEs of the parameter and feature are respectively 328% and 181%. The large errors for the variances show that Metric MD-2 can only "change", but cannot "calibrate" the variance. As clearly seen in Fig. 5, the center points of Sample MD-2 and test sample coincide, while the ellipse of Sample MD-2 is obviously larger than the ellipse of test sample.

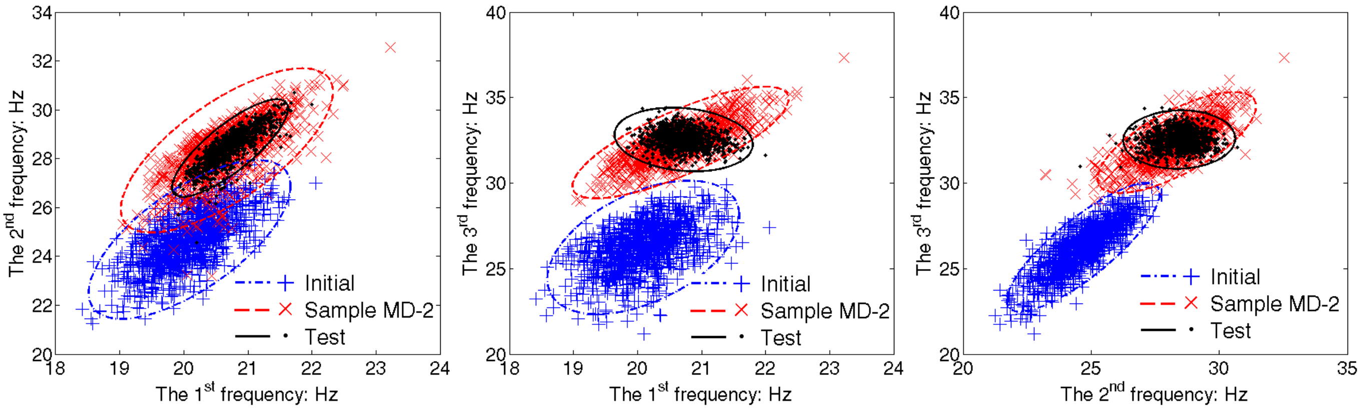


Fig. 5 The test, initial, and calibrated scatters in planes of different frequencies with Metric MD-2

4. Results with Metric BD

The results with Metric BD exhibit its capability for variance calibration, as shown in Table 5. The MAE of parameter variance is 46.4%, which is much less than the initial error 354%. The MAE of the feature variance is also reduced from 240% to 26.9%. As for the calibrated means, the MAE of parameters (7.63%) is higher than the previous metrics. The features means MAE (0.61%) is still low, although slightly higher than the previous metrics as shown in Table 4.

The result is also illustrated in Fig. 6, where the ellipse patterns between the test and calibrated samples show better agreement than previous metrics, although slight discrepancies still exist. Nevertheless, this discrepancy is reduced by Metric ED-BD in the following subsection.

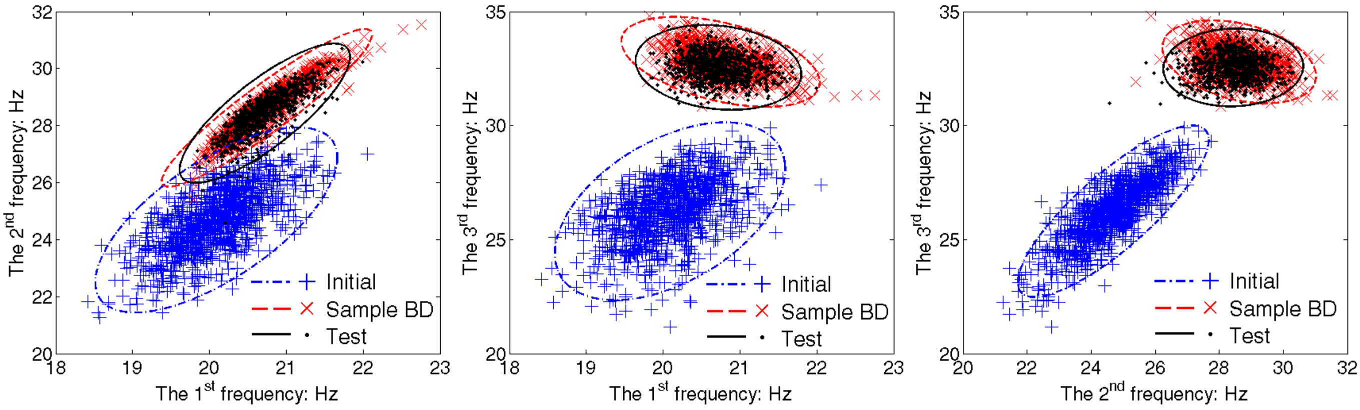


Fig. 6 **The test, initial, and calibrated scatters in planes of different frequencies with Metric BD**

5. Results with Metric ED-BD

Metric ED-BD is a combined metric that integrates the Euclidean and Bhattacharyya distances between two samples. As described in the previous subsection, Metric BD is capable of calibrating both mean and variance; however, the calibration precision of mean is lower than Metric ED. In this case, the means are first calibrated with the Euclidean distance and variances are subsequently calibrated with the Bhattacharyya distance.

An integrated advantage can be checked from Tables 4 and 5. The MAEs of parameter and feature means (1.88% and 0.34%) are both lower than the result with Metric BD (7.63% and 0.61%). Besides, the errors of feature variances are furthermore reduced compared with the result of Metrics BD. Metric ED-BD's improvement with respect to Metric BD could be demonstrated more obviously through Fig. 7 where Sample ED-BD is more coincident with the test sample than what in Fig. 6. Ellipses of Sample ED-BD and the test sample have almost the same positions, shapes, and orientations.

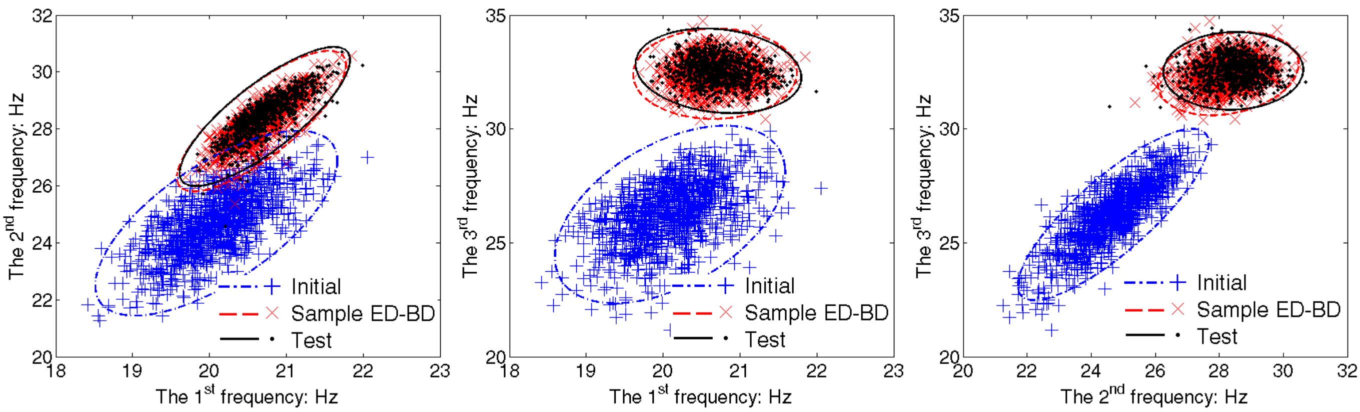


Fig. 7 **The test, initial, and calibrated scatters in planes of different frequencies with Metric ED-BD**

C. Cross comparison and result analysis

This subsection gives an overall comparison of the feature samples calibrated using the above five metrics. The distances between the calibrated samples and the test sample are detailed in Table 6. Note that all of the calibrated samples are randomly generated based on the given parameter means (Table 4) and variances (Table 5). The stochastic nature of the calibrated samples leads to a slight change in the distance values. Nevertheless, the constant tendencies are summarized as follows.

As can be seen from the row labeled as Euclidian distances in Table 6, Sample BD has the largest value, compared with the other calibrated samples whose values are at similar and lower levels. This is because the Euclidian and Mahalanobis distances are respectively the absolute and weighted distances between sample means, while the Bhattacharyya distance concentrates on the sample distributions. When only error of the mean is concerned, the performance of Bhattacharyya distance is not as good as the others by nature. However, as Sample ED-BD is first calibrated using Metric ED, it has the similar Euclidian distance as Sample ED.

The Mahalanobis distance is calculated based on Metric MD-2, which utilizes the pooled covariance matrix of two samples. As a result, Sample MD-2 has the smallest value in the corresponding row of Table 6. However, a calibrated sample with the smallest Mahalanobis distance does not mean it has a favorable representation of the test sample. Fig. 5 and Table 5 clearly show that the variance of Sample MD-2 is much larger than the reference. This increasing tendency can be explained by the formulation of Metric MD-2. The pooled covariance $\mathbf{C}_{pool}$ plays an important role in variance calibration, where a larger $\mathbf{C}_{pool}$ leads to a smaller distance (recalling Eq. 7). Because the experimental covariance is fixed as a target, $\mathbf{C}_{pool}$ increases only when the numerical covariance increases. As a result, the calibrated variance does not converge to the reference but obviously increases when Metric MD-2 is minimized.

The last row in Table 6 concerns the Bhattacharyya distance. The values of Samples BD and ED-BD are clearly reduced compared with other samples. It is necessary to state that Sample ED-BD possesses an even smaller value than Sample BD. This is because the calculation process of Bhattacharyya distance is also influenced by the coincidence of the means between two samples. A better performance in the Euclidian distance obviously brings a smaller Bhattacharyya distance between Sample ED-BD and the test sample. As shown in the last column of Table 6, Sample ED-BD has the advantages of the smallest Euclidian and Bhattacharyya distances, and its Mahalanobis distance is still relatively low. All of these advantages are visually illustrated in Fig. 7.

Table 6 Cross comparison of the distances among feature samples

	Calibrated samples *VS.* test sample				
	Sample ED	Sample MD-1	Sample MD-2	Sample BD	Sample ED-BD
Euclidian distance	0.3081	0.3646	0.3072	0.5102	0.2353
Mahalanobis distance	0.7692	0.6782	0.2374	0.6567	0.4773
Bhattacharyya distance	0.3618	0.3612	0.3252	0.2031	0.1644

Based on the above analysis, a general guideline on metric selection during model calibration is proposed as follows.

1) In case of single experimental measurement, only Metric ED is feasible among the above five metrics. A single set of parameter value is obtained through a deterministic calibration without any statistical treatment.

2) When multiple measurements are available, either deterministic or stochastic calibration is applicable depending on the practical requirement. a) If only parameter means are demanded, Metrics ED and MD-1 are proposed for deterministic calibration. Being simple and intuitive, the classical Metric ED is the first recommendation for this case. Metric MD-1 is the weighted geometric distance and it is sensitive to slight changes in the test variance (e.g. measurement errors). Consequently, it is proposed in certain applications [28, 30] to check the stability of the calibrated results. b) A stochastic calibration is necessary when both parameter means and variances are demanded. Metric MD-2 considered herein is not recommended as it can only change, but cannot calibrate the variance. Metric BD is a more comprehensive metric with a higher level of statistical information. It contains the probability distribution information of two samples, and hence the calibrated sample has a more similar distribution with the target sample. However, the calibration precision on the means is smaller than the Euclidian distance. The classical Euclidian distance together with measurement means should be the primary consideration even in stochastic calibration processes. Metric ED-BD is consequently proposed as the first choice because of its global advantages.

IV. Case study II: Validation with regard to measurement uncertainties

In practical experiments, measurements are acquired with hard-to-control random effects, such as environment noise and measurement errors. The distance metrics become useful for model selection when alternative calibrated models are found to demonstrate comparable fidelity with regard to uncertainties in multiple measurements. This section presents a cross validation process with an integrated application of both deterministic and stochastic metrics.

A. Experimental data preparation

This case study employs the same structure as Section III but with realistic experimental data. In order to represent test variability, the measurements are repeated four times by different people with different configurations of the sensors. The random sampling method in Section II.D is employed to generate the semi-experimental sample with 1000 points based on these four measurements, which are listed in Table 7. The semi-experimental sample and four original measurement points are illustrated in Fig. 8. A sketch of the probabilistic density function of the joint distribution between the 2nd and 4th frequencies is shown in Fig. 9. Further statistical information of the semi-experimental sample is presented in Table 8. The correlation coefficient between two of each frequency is defined as

$$\mathbf{R}(i,j)=\frac{\mathbf{C}(i,j)}{\sqrt{\mathbf{C}(i,i)\cdot\mathbf{C}(j,j)}}\qquad i,j=1,2,3,4 \tag{15}$$

where C is the covariance matrix. The correlation coefficient reflects the strength and direction of the linear relationship of the frequency pair, which is also illustrated by the orientation of the scatter ellipse in Fig. 8. Another statistical feature is the coefficient of variation, i.e. the ratio between the standard deviation and the mean. It is employed to quantify the degree of dispersion of the each frequency in the semi-experimental sample. As shown in Table 8, the coefficient of variation of the first frequency is relatively lower than the higher order frequencies. This tendency conforms to the practical experience that the higher order modes are always more sensitive to the test variability compared with the low order modes.

Table 7 The original 4 sets of experimental data

Participant No.	f_1 (Hz)	f_2 (Hz)	f_3 (Hz)	f_4 (Hz)
1	18.11	24.38	31.14	41.97
2	18.13	24.02	30.64	39.78
3	18.05	23.26	29.97	37.77
4	18.27	23.22	27.95	39.95

Table 8 The correlation matrix, mean, and coefficient of variation of the semi-experimental sample

		f_1	f_2	f_3	f_4
Correlation matrix	f_1	1	-0.33	-0.80	0.29
	f_2	-0.33	1	0.82	0.75
	f_3	-0.80	0.82	1	0.27
	f_4	0.29	0.75	0.27	1
Mean		18.14	23.70	29.88	39.79
Coefficient of variation		0.0052	0.0238	0.0476	0.0418

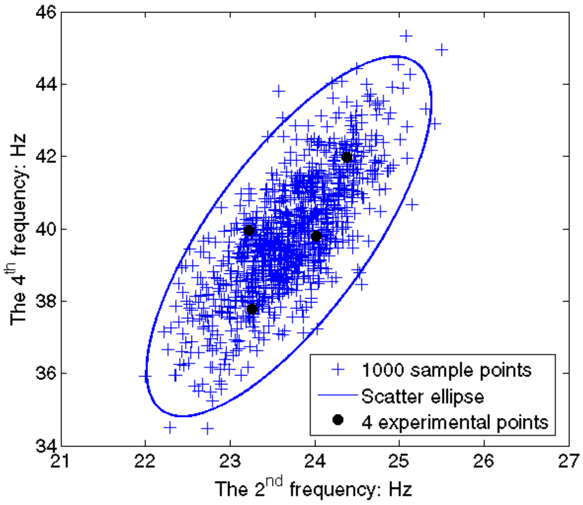


Fig. 8 The random sampling of the experimental data on the plane of the 2nd and 4th frequencies

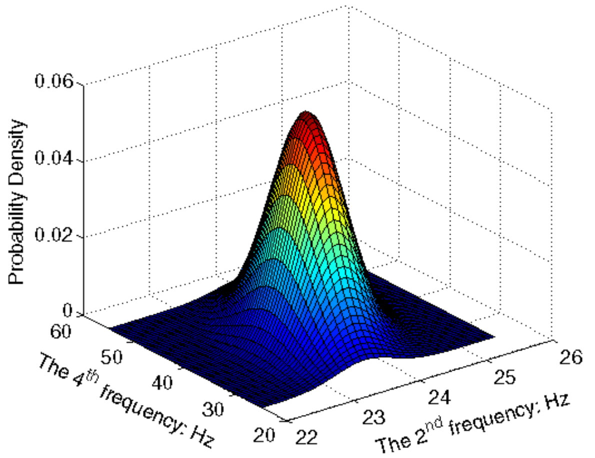


Fig. 9 The joint probabilistic density function of the 2nd and 4th frequencies

B. Cross validation with deterministic and stochastic metrics

A deterministic calibration using Metric ED is individually performed on the FE model developed in Section III, according to each of the four measurements. The corresponding four calibration results of this model (termed as Model-1) are listed in Table 9.

In additional to Model-1, another FE model (termed as Model-2) with a different parameter configuration is proposed for comparison. The calibration parameters of Model-2 and their intervals are detailed in Table 10. Similarly, Model-2 is separately calibrated using Metric ED according to each of the four measurements. The four calibration points of each model are extended to a random sample using the same method illustrated in Fig. 8. Means of the samples, as well as their relative errors compared with the semi-experimental sample, are presented in Table 9. The similar mean errors imply that the alternative Model-1 and Model-2 possess comparable fidelity with regard to the measurements. In additional to the means, the correlation matrices and coefficients of variation of the calibrated samples of Model-1 and Model-2 are presented in Table 11. A more comprehensive validation process considering the statistical information is performed to assist in model selection.

Table 9 Calibrated means according to the four measuring participants in Table 7

Participant No.	Model-1				Model-2			
	f_1 (Hz)	f_2 (Hz)	f_3 (Hz)	f_4(Hz)	f_1 (Hz)	f_2 (Hz)	f_3 (Hz)	f_4(Hz)
1	18.37	24.24	30.20	43.16	19.22	24.15	30.21	42.66
2	18.33	23.29	28.90	41.48	18.97	23.63	29.01	41.29
3	18.22	22.06	27.99	39.96	18.74	21.92	28.15	39.79
4	18.28	22.94	27.75	40.30	18.80	22.96	27.79	40.17
Sample mean (% error)	18.30 (0.87%)	23.09 (-2.55%)	28.68 (-3.99)	41.18 (3.49%)	18.93 (4.35%)	23.15 (-2.29)	28.77 (-3.68)	40.95 (2.92%)
Mean absolute error	2.73%				3.31%			

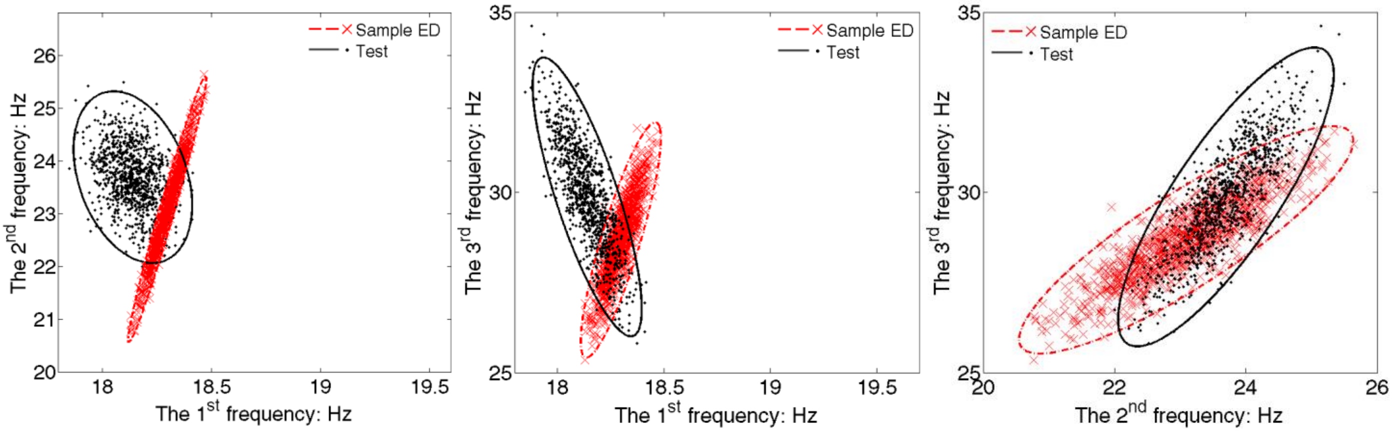
Table 10 Calibration parameters of Model-2

Parameters	Intervals	Descriptions
E_1 (10^{11}Pa)	[0.5, 2.5]	Young's modulus of the upper connections
E_2 (10^{11}Pa)	[0.5, 2.5]	Young's modulus of the middle connections
T_1 (10^{-3}m)	[2.0, 5.0]	Thickness of the upper connections
T_2 (10^{-3}m)	[2.0, 5.0]	Thickness of the middle connections

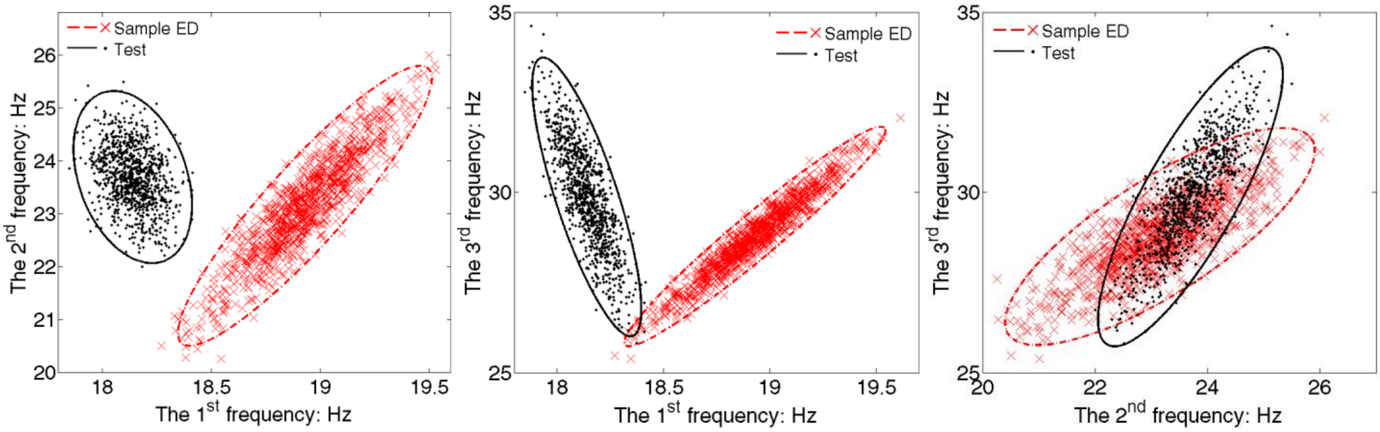
Table 11 Correlation matrices and coefficients of variation of the calibrated samples of Model-1 and Model-2

		Model-1				Model-2			
		f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4
Correlation matrix	f_1	1	0.98	0.87	0.94	1	0.91	0.97	0.9992
	f_2	0.98	1	0.87	0.95	0.91	1	0.79	0.92
	f_3	0.87	0.87	1	0.98	0.97	0.79	1	0.96
	f_4	0.94	0.95	0.98	1	0.9992	0.92	0.96	1
Coefficient of variation		0.0034	0.0378	0.0377	0.0431	0.0115	0.0426	0.0375	0.0317

The cross validation considered herein is essentially a study on the robustness of the calibration results with regard to the test variability. The calibration sample with a lower degree of dispersion indicates a low sensitivity, i.e. high robustness, in the presence of the measurement uncertainty. The degree of dispersion is quantified by the coefficient of variation in Table 11, and visually illustrated by the position relations of the scatters in Fig. 10. Similarly, planes of only the first three frequencies are presented in the figure and the planes relative to the fourth frequency can be checked in Appendix B.



(a) Model-1



(b) Model-2

Fig. 10 Comparison of the test and calibrated scatters of Model-1 and Model-2

Note that the calibration samples are no longer expected to coincide with the measurement samples. In this case study, the multiple measurements are repeated on the same structure implying the structural characteristics are not significantly changed. The test variability considered herein is primarily caused by the uncertainties in experimental procedure, e.g. environment noise and measurement errors. The attempt to compensate the test variability by modifying the parameters of the FE model is not the objective of this validation process. However, different FE models, with different levels of model form error, have different sensitivity with regard to test variability. This phenomenon is clearly illustrated by the comparison between Fig. 10 (a) and (b), where Sample Model-1 and the test sample have more overlap, compared with Sample Model-2. This means Model-1 is less sensitive (i.e. more robust) than Model-2 in presence of test variability. Ideally, a calibrated model exhibits maximum robustness when its calibration sample completely falls into the region of the test sample. However, because of the hard-to-control randomness in the test procedure, this ideal case is normally not seen in practical application. A quantitative assessment of the sensitivity is provided by Metric BD as listed in Table 12, where a smaller Bhattacharyya distance of Model-1 reveals its lower sensitivity and higher robustness compared with Model-2. Metric BD is utilized here as an appropriate tool for cross validation in the presence of test variability.

Table 12 Bhattacharyya distances between the test sample and calibration samples of Model-1 and Model-2

	Model-1 *VS.* test sample	Model-2 *VS.* test sample
Bhattacharyya distance	1.2599	3.0594

V. Conclusions

Based on the Euclidian, Mahalanobis, and Bhattacharyya distance-based criteria, five test-analysis comparison metrics (Metrics ED, MD-1, MD-2, BD, and ED-BD) are evaluated and compared in model calibration. Emphasis is given to the impact of increased statistical information on the calibration results. An integrated application of both deterministic and stochastic metrics is proposed in a cross validation process. Following a deterministic calibration using Metric ED, the sensitivity and robustness of the calibration results are assessed by Metric BD with regard to measurement uncertainties.

This work provides the framework to utilize and evaluate the distance metrics with varying levels of statistical information during model calibration and validation. The stochastic metric is more comprehensive for uncertainty treatment as the contained statistical information is increasing, while the calculation burden is naturally rising. With the rapid development of computer technology, it is expected that these distance metrics ensure the quantitative uncertainty analysis and robust prediction of variability, which are highlights in computational mechanics engineering.

Appendix

A. Supplements of Section III: The scatters relative to the fourth frequency

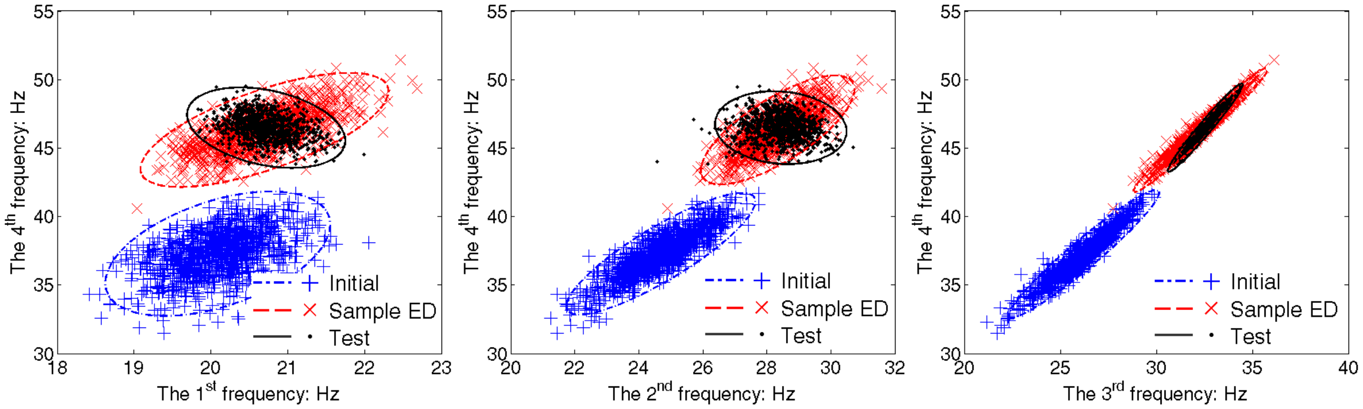


Fig. A.1 Scatters of Sample ED in the planes relative to the 4th frequency

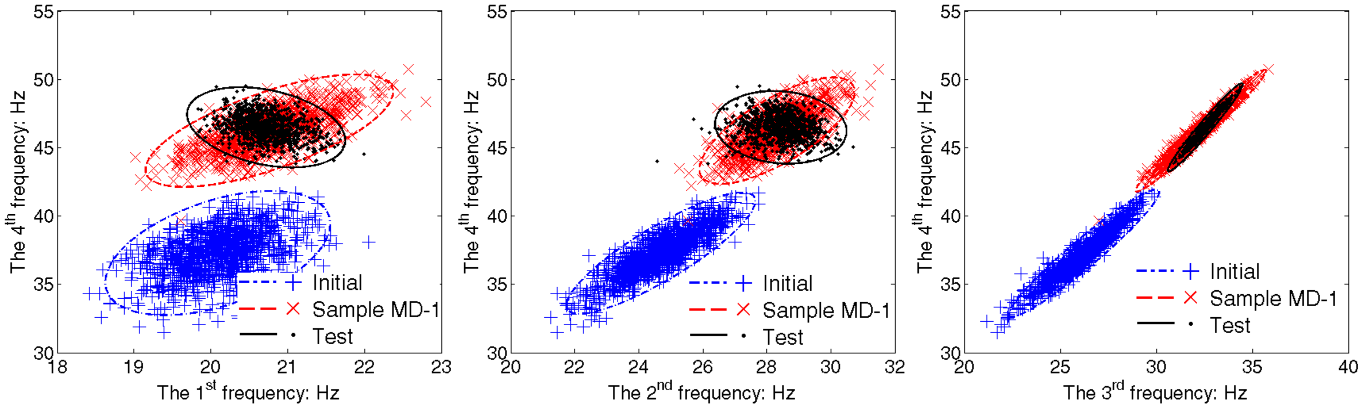


Fig. A.2 Scatters of Sample MD-1 in the planes relative to the 4th frequency

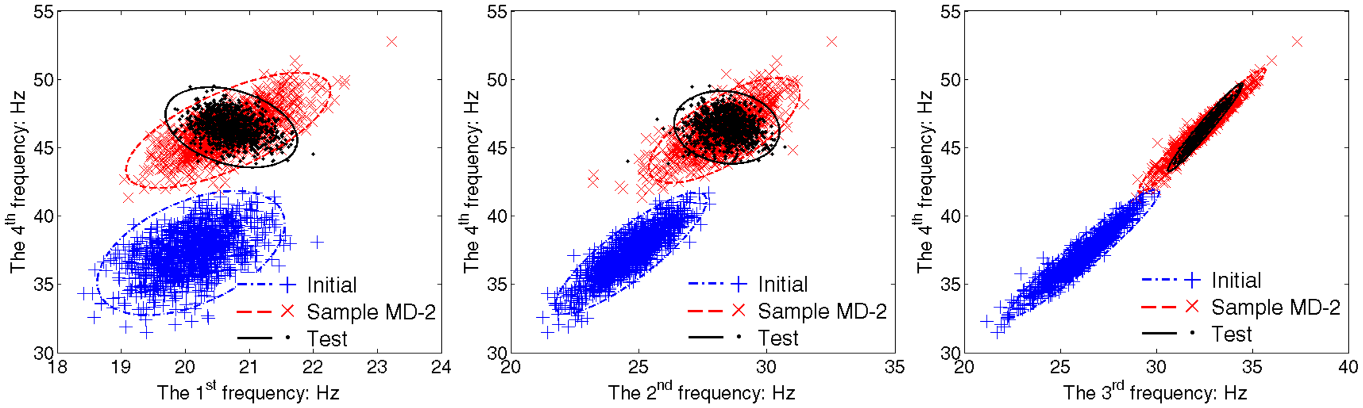


Fig. A.3 Scatters of Sample MD-2 in the planes relative to the 4th frequency

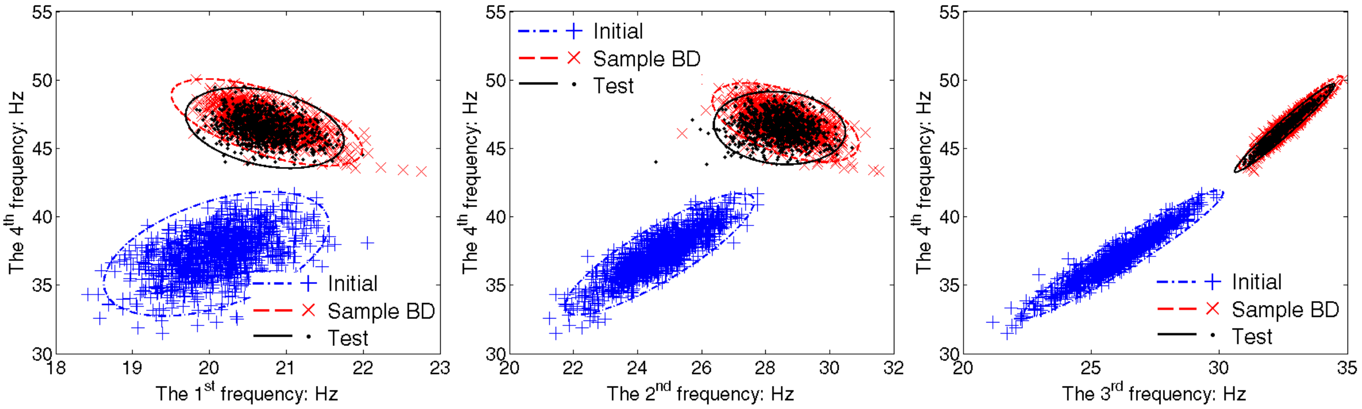


Fig. A.4 Scatters of Sample BD in the planes relative to the 4th frequency

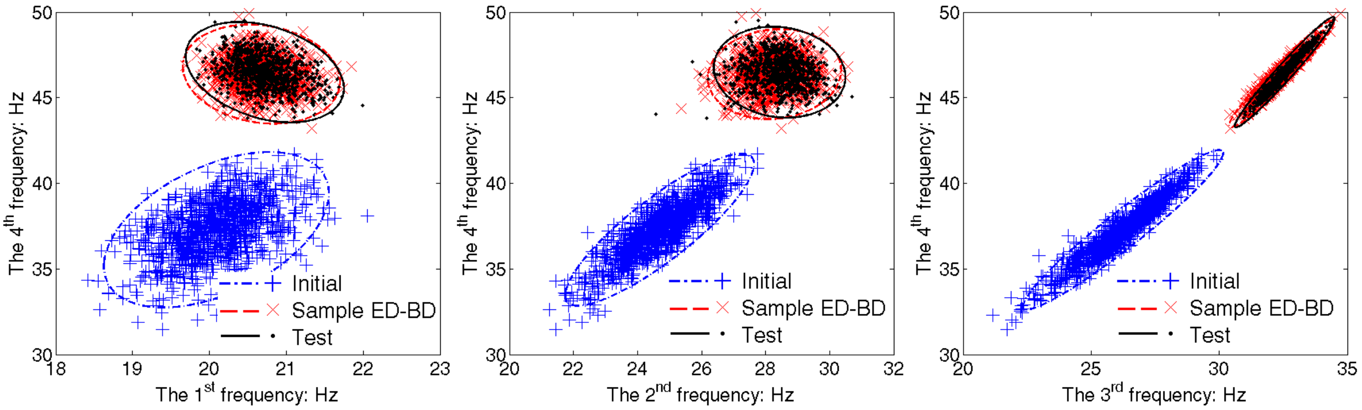
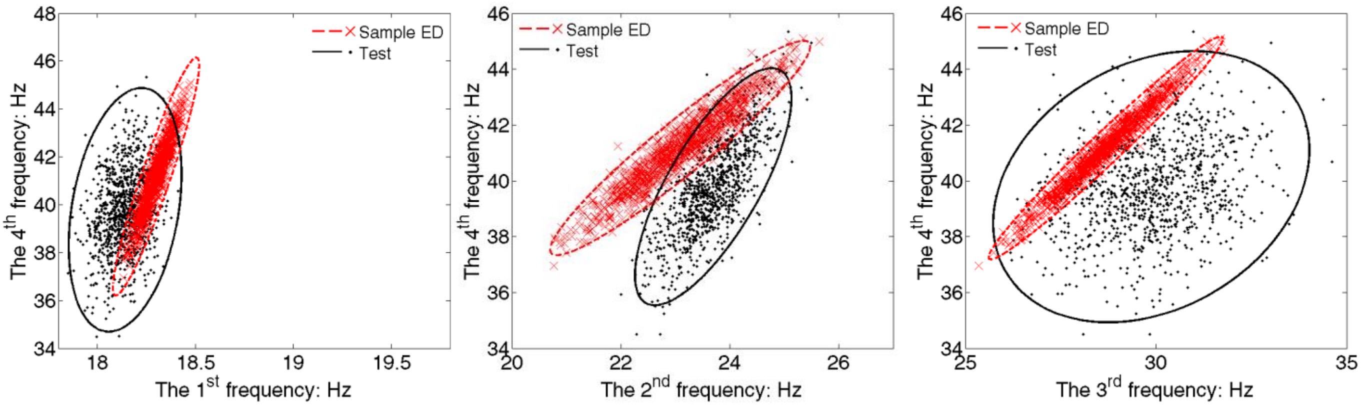
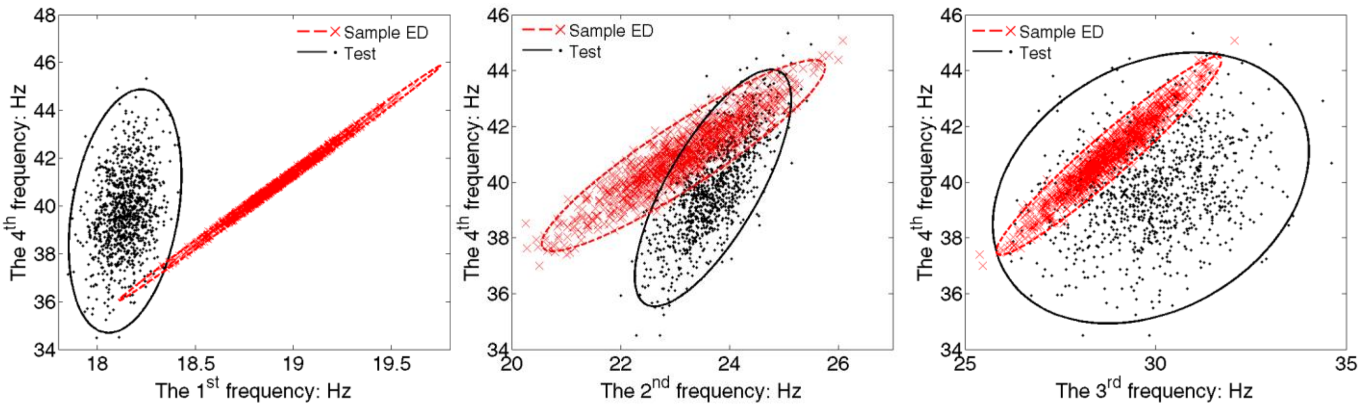


Fig. A.5 Scatters of Sample ED-BD in the planes relative to the 4th frequency

B. Supplements of Section IV: The scatters relative to the fourth frequency



(a) Model-1



(b) Model-2

Fig. B The test and calibrated scatters of Model-1 and Model-2 in the planes relative to the 4th frequency

References

[1] Calvi, A. "Uncertainty-based loads analysis for spacecraft: Finite element model validation and dynamic responses," *Computers & Structures* Vol. 83, No. 14, 2005, pp. 1103-1112.

doi: 10.1016/j.compstruc.2004.11.019

[2] Farajpour, I., and Atamturktur, S. "Error and uncertainty analysis of inexact and imprecise computer models," *Journal of Computing in Civil Engineering* Vol. 27, No. 4, 2012, pp. 407-418.

doi: 10.1061/(ASCE)CP.1943-5487.0000233

[3] Doebling, S. W. "Minimum-rank optimal update of elemental stiffness parameters for structural damage identification," *AIAA Journal* Vol. 34, No. 12, 1996, pp. 2615-2621.

doi: 10.2514/3.13447

[4] Doebling, S. W., Hemez, F. M., Schultze, J. F., and Girrens, S. P. "Overview of structural dynamics model validation activities at Los Alamos National Laboratory," *43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. Denver, Colorado, 2002.

[5] "Guide for the Verification and Validation of Computational Fluid Dynamics Simulations (AIAA G-077-1998(2002))," *AIAA Standards*, 1998.

doi: 10.2514/4.472855.001

[6] Chen, W., Baghdasaryan, L., Buranathiti, T., and Cao, J. "Model validation via uncertainty propagation and data transformations," *AIAA Journal* Vol. 42, No. 7, 2004, pp. 1406-1415.

doi: 10.2514/1.491

[7] Friswell, M. I., and Mottershead, J. E. *Finite Element Model Updating in Structural Dynamics*. Dordrecht: Kluwer Academic Press, 1995.

[8] Mottershead, J. E., Link, M., and Friswell, M. I. "The sensitivity method in finite element model updating: A tutorial," *Mechanical Systems and Signal Processing* Vol. 25, 2011, pp. 2275-2296.

doi: 10.1016/j.ymssp.2010.10.012

[9] Schuëller, G. I. "On the treatment of uncertainties in structural mechanics and analysis," *Computers & Structures* Vol. 85, No. 5–6, 2007, pp. 235-243.

doi: 10.1016/j.compstruc.2006.10.009

[10] Rebba, R., and Mahadevan, S. "Model predictive capability assessment under uncertainty," *AIAA Journal* Vol. 44, 2006, pp. 2376–2384.

doi: 10.2514/1.19103

[11] Sairajan, K. K., and Aglietti, G. S. "Robustness of System Equivalent Reduction Expansion Process on Spacecraft Structure Model Validation," *AIAA Journal* Vol. 50, No. 11, 2012, pp. 2376-2388.

doi: 10.2514/1.j051476

[12] Govers, Y., Khodaparast, H. H., and Link, M. "A comparison of two stochastic model updating methods using the DLR AIRMOD test structure," *Mechanical Systems and Signal Processing* Vol. 52, 2015, pp. 105-114.

doi: 10.1016/j.ymssp.2014.06.003

[13] Atamturktur, S., Hemez, F. M., Unal, C., Tome, C., and Williams, B. "A forecasting metric for predictive modeling," *Computers & Structures* Vol. 89, No. 23-24, 2011, pp. 2377-2387.

doi: 10.1016/j.compstruc.2011.06.010

[14] McFarland, J., Mahadevan, S., Romero, V., and Swiler, L. "Calibration and Uncertainty Analysis for Computer Simulations with Multivariate Output," *AIAA Journal* Vol. 46, No. 5, 2008, pp. 1253-1265.
doi: 10.2514/1.35288

[15] Atamturktur, S., Liu, Z., Cogan, S., and Juang, H. "Calibration of imprecise and inaccurate numerical models considering fidelity and robustness: a multi-objective optimization-based approach," *Structural and Multidisciplinary Optimization* Vol. 51, No. 3, 2014, pp. 659-671.
doi: 10.1007/s00158-014-1159-y

[16] Stevens, G., Buren, K. V., Wheeler, E., and Atamturktur, S. "Evaluating the fidelity and robustness of calibrated numerical model predictions," *Engineering Computations* Vol. 32, No. 3, 2015, pp. 621-642.
doi: 10.1108/EC-09-2013-0217

[17] Liu, Y., Chen, W., Arendt, P., and Huang, H. Z. "Toward a Better Understanding of Model Validation Metrics," *Journal of Mechanical Design* Vol. 133, No. 7, 2011.
doi: 10.1115/1.4004223

[18] Schwer, L. E. "Validation metrics for response histories: perspectives and case studies," *Engineering with Computers* Vol. 23, 2007, pp. 295-309.
doi: 10.1007/s00366-007-0070-1

[19] Hemez, F. M., and Ben-Haim, Y. "Info-gap robustness for the correlation of tests and simulations of a non-linear transient," *Mechanical Systems and Signal Processing* Vol. 18, No. 6, 2004, pp. 1443-1467.
doi: http://dx.doi.org/10.1016/j.ymssp.2004.03.001

[20] Ang, A. H.-S., and Tang, W. H. *Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering, 2nd Edition*: John Wiley & Sons, Inc., 2007.

[21] Allemang, R. J. "The modal assurance criterion: twenty years of use and abuse," *Sound and vibration* Vol. 8, 2003, pp. 14-21.

[22] Heylen, W., and Lammens, S. "FRAC: A Consistent Way of Comparing Frequency Response Functions," *International Conference on Identification in Engineering*. Swansea, UK, 1996, pp. 48-57.

[23] Mahalanobis, P. C. "On the generalised distance in statistics," *Proceedings of the National Institute of Sciences of India*. 1936, pp. 49-55.

[24] Bhattacharyya, A. "On a Measure of Divergence between Two Multinomial Populations," *The Indian Journal of Statistics* Vol. 7, No. 4, 1946, pp. 401-406.

[25] Oberkampf, W. L., Helton, J. C., Joslyn, C. A., Wojtkiewicz, S. F., and Ferson, S. "Challenge problems: uncertainty in system response given uncertain parameters," *Reliability Engineering & System Safety* Vol. 85, No. 1, 2004, pp. 11-19.

doi: 10.1016/j.ress.2004.03.002

[26] Wei, P., Lu, Z., and Song, J. "Extended Monte Carlo Simulation for Parametric Global Sensitivity Analysis and Optimization," *AIAA Journal* Vol. 52, No. 4, 2014, pp. 867-878.

doi: 10.2514/1.j052726

[27] ST-ORM. *User's manual, version 2.0*. Alzenau, Germany, 2002.

[28] Doltsinis, I. *Stochastic Analysis of Multivariate Systems in Computational Mechanics and Engineering*. Barcelona, Spain: International Centre for Numerical Methods in Engineering (CIMNE), 1999.

[29] Marczyk, J. *Principles of simulation-based computer-aided engineering*: Fim Publications Barcelona, 1999.

[30] Deng, Z., Bi, S., and Atamturktur, S. "Stochastic model updating using distance discrimination analysis," *Chinese Journal of Aeronautics* Vol. 27, 2014, pp. 1188-1198.

doi: 10.1016/j.cja.2014.08.008

[31] Xiang, S., Nie, F., and Zhang, C. "Learning a Mahalanobis distance metric for data clustering and classification," *Pattern Recognition* Vol. 41, No. 12, 2008, pp. 3600-3612.

doi: 10.1016/j.patcog.2008.05.018

[32] De Maesschalck, R., Jouan-Rimbaud, D., and Massart, D. L. "The Mahalanobis distance," *Chemometrics and Intelligent Laboratory Systems* Vol. 50, No. 1, 2000, pp. 1-18.

doi: 10.1016/S0169-7439(99)00047-7

[33] Guratzsch, R. F., and Mahadevan, S. "Structural Health Monitoring Sensor Placement Optimization Under Uncertainty," *AIAA Journal* Vol. 48, No. 7, 2010, pp. 1281-1289.

doi: 10.2514/1.28435

[34] Brereton, R. G., and Lloyd, G. R. "Re-evaluating the role of the Mahalanobis distance measure," *Journal of Chemometrics*, Vol. 34, No. 4, 2016, pp. 134-143.

doi: 10.1002/cem.2779

[35] Choi, E., and Lee, C. "Feature extraction based on the Bhattacharyya distance," *Pattern Recognition* Vol. 36, No. 8, 2003, pp. 1703-1709.

doi: 10.1016/S0031-3203(03)00035-9

[36] Patra, B. K., Launonen, R., Ollikainen, V., and Nandi, S. "A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data," *Knowledge-Based Systems* Vol. 82, 2015, pp. 163-177.

doi: 10.1016/j.knosys.2015.03.001

[37] Aherne, F. J., Thacker, N. A., and Rockett, P. I. "The Bhattacharyya metric as an absolute similarity measure for frequency coded data," *Kybernetika* Vol. 34, No. 4, 1998, pp. 363-368.

[38] Wang, Y., and Sui, L. *Design of experiment and matlab data analysis*. Beijing: Tsinghua University Press, 2012.

[39] ASME. "An illustration of the concepts of verification and validation in computational solid mechanics." *the American Society of Mechanical Engineers*, Three Park Avenue, New York, 10016 USA, 2012.

[40] Zhang, Y. "Solving large-scale linear programs by interior-point methods under the Matlab Environment," *Optimization Methods and Software* Vol. 10, No. 1, 1998, pp. 1-31.

doi: 10.1080/10556789808805699