Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Multivariate error assessment of response time histories method for dynamic systems*

Zhen-fei ZHAN[1], Jie HU[†‡1], Yan FU[2], Ren-Jye YANG[2], Ying-hong PENG[1], Jin QI[1]

([1]*School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*)

([2]*Research and Innovation Center, Ford Motor Company, Dearborn, MI48124, USA*)

[†]E-mail: hujie@sjtu.edu.cn

Received Mar. 22, 2011; Revision accepted Sept. 6, 2011; Crosschecked Dec. 27, 2011

Abstract: In this paper, an integrated validation method and process are developed for multivariate dynamic systems. The principal component analysis approach is used to address multivariate correlation and dimensionality reduction, the dynamic time warping and correlation coefficient are used for error assessment, and the subject matter experts (SMEs)' opinions and principal component analysis coefficients are incorporated to provide the overall rating of the dynamic system. The proposed method and process are successfully demonstrated through a vehicle dynamic system problem.

Key words: Model validation, Multivariate dynamic responses, Principal component analysis, Dynamic time warping

doi:10.1631/jzus.A1100073 **Document code:** A **CLC number:** TP391.72

1 Introduction

With the ever shortening speed to market, computer aided engineering (CAE) has become a crucial tool for product development in automobile industry. Various computer programs and models are developed to simulate dynamic systems. Before using these models, their validity and predictive capabilities need to be assessed quantitatively. Model validation is the process of comparing CAE model outputs with test data in order to assess the validity or predictive capabilities of the CAE model for its intended usage. One of the critical tasks to achieve model validation is to select a validation metric that has the desirable metric properties for model assessment of a dynamic system with multiple functional responses (Oberkampf and Barone, 2006; Oberkampf and Trucano, 2006). Developing quantitative model validation methods has attracted considerable researchers' interest in recent years (Schwer, 2007; Ferson *et al*., 2008). Statistical hypothesis testing is one approach to provide quantitative model validation (Mahadevan and Rebba, 2005; Rebba and Mahadevan, 2006). Bayesian methods have been developed to determine the predictive capabilities of CAE models (Jiang and Mahadevan, 2008). They were applied to various model validation problems and showed significant potential. Jiang *et al*. (2009), Fu *et al*. (2010) and Zhan *et al*. (2011a) further exploited the Bayesian based validation methods for multivariate dynamic systems. However, these methods considered the whole distribution of interested time history, the agreements of the important features were not addressed. In order to evaluate the agreement of crucial features, Schwer (2007) and Sarin *et al.* (2010) evaluated two categories of metrics for model validation of dynamic systems, namely, Sprague and Geers metric and Knowles and Gear metric. In these metrics, the magnitude and phase features are calculated based on the time integration of the waveforms. More recently, Zhan *et al.* (2011b) proposed an enhanced method

‡ Corresponding author

* Project supported by the National Natural Science Fundation of China (No. 51075262), the New Century Excellent Talents in University Program (No. NCET-08-0361), and the Fund for the Doctoral Program of Higher Education (No. 200802480036), China

based on error assessment of response time history (EARTH) metric. These methods, however, aim to quantitatively compare two curves (i.e., model prediction and test data) for a single response. They cannot handle multivariate correlation directly, which may not be effective in cases of model validation for multivariate and complex dynamic systems, such as vehicle crash tests.

In this paper, a quantitative multivariate validation method and common process is proposed to quantitatively assess the agreements of important features of multiple dynamic responses simultaneously. It exploits several advanced techniques such as principal component analysis (PCA) (Joliffe, 2002) and dynamic time warping (DTW) (Lei and Govindaraju, 2003). The PCA approach is used to address multivariate correlation and dimensionality reduction, the DTW and correlation coefficient calculation are used for error assessment, the subject matter experts (SMEs)' opinions are incorporated to provide the overall rating of the dynamic system. In the following sections, the multivariate validation framework is first briefly introduced, and then the PCA based dimension reduction, error assessments of dynamic responses, SMEs based response scores calculation, and PCA based multivariate error assessment of response time histories (M-EARTH) score calculation are described in detail. Next, an application procedure is provided. Some conclusions are given in the end.

2 Multivariate error assessment of response time histories (M-EARTH) metric

Fig. 1 shows the flowchart of the proposed M-EARTH method. It contains multiple steps including: (1) PCA based dimension reduction, (2) error assessments of dynamic responses, (3) SMEs based response scores calculation, and (4) PCA based M-EARTH score calculation. The process starts with principal component analysis based dimension reduction. First, comparable multivariate data from test and CAE are normalized by test data peaks to dimensionless data. The PCA is then applied to the normalized test data to reduce data dimension and also to address multivariate data correlation. Using the PCA coefficients from test, the CAE data is transformed to the reduced PCA space for comparison with the test data. Then the error assessments are conducted to the PCA reduced test and CAE data. Three independent errors including phase, magnitude, and shape are calculated for each time history of reduced data. Using the SMEs knowledge, in each response, the three EARTH errors are transformed and combined as an intuitive score ranging from 0 to 100%. Finally, the scores for all responses of reduced data can be combined through the PCA coefficient into an overall score of the multivariate dynamic system. Based on the result, the decision maker (DM) then decides to accept or reject the model.

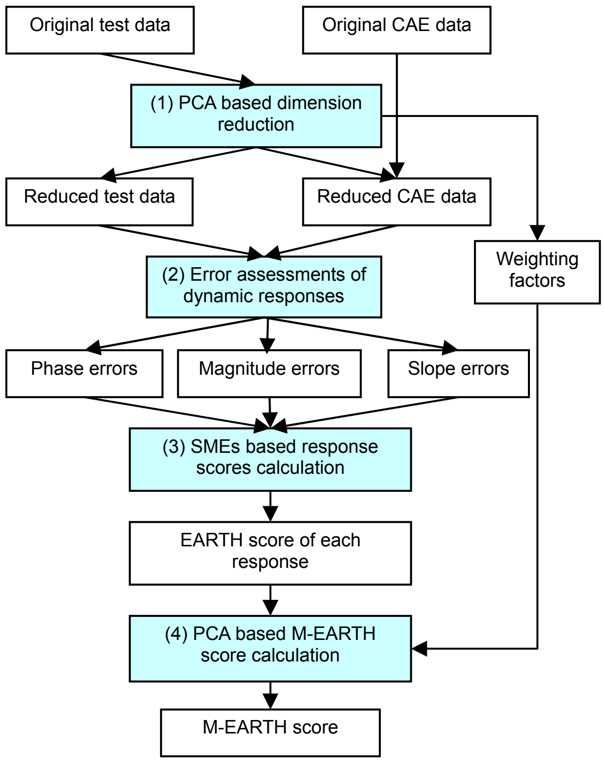

Fig. 1 M-EARTH validation process

2.1 PCA based dimension reduction

In multivariate data analysis, high numbers of variables and correlations between variables make it difficult to interpret and summarize the results as well as to apply multivariate statistics. One major target to deal with highly correlated multivariate data is to remove the dependence amongst variables and reduce the dimension of variables. Amongst all techniques, PCA is a well-established statistical method for dimensionality reduction and has been widely applied in data compression, image processing, exploratory data analysis, pattern recognition, and time series prediction (Hotelling, 1933; Joliffe, 2002). PCA is capable of reducing data dimensionality and addressing the multivariate correlation. It is selected in this study as the dimension reduction technique.

PCA involves a matrix analysis technique called eigenvalue decomposition. The eigenvalues and eigenvectors from decomposition represent the amounts of variation accounted for by each principal component and the weights for the original variables, respectively. Its main objective is to transform a set of correlated high dimensional variables to a set of uncorrelated lower dimensional variables (principal components). The important property of PCA is that the principal component projection minimizes the squared reconstruction error in dimensionality reduction. The PCA is not based on a probabilistic model, so no data distribution assumption is involved in a PCA transform process. In automotive safety applications, the output multiple responses are generated from the same crash event, hence, some of the responses are highly correlated. PCA could transform highly correlated original responses into uncorrelated lower dimensional responses. The judgment based on reduced principal components can be efficient and minimize the squared reconstruction error in dimension reduction.

To determine the proper number of principal components that should be retained is an important issue in PCA implementation. In our study, we use intrinsic dimensionality as the proper number. The intrinsic dimensionality is the minimum number of latent variables that is necessary to account for enough information in the original data. The eigenvalues corresponding to the principal components in the PCA represent the amount of variance explained by their corresponding eigenvectors. The first p eigenvalues are typically high, implying that most information is accounted for in the corresponding principal components. Thus, the intrinsic dimensionality p is obtained by calculating the cumulative percentage of the p eigenvalues (i.e., the total variability by the first p principal components) that is higher than a threshold value, say 95%. This implies that the retained p principal components account for 95% information of the original data.

Let $\boldsymbol{T}=[\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_m]^{\mathrm{T}}$, $\boldsymbol{C}=[\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_m]^{\mathrm{T}}$ represent the $m\times n$ physical test data and CAE model predictions, $\boldsymbol{t}_i$, $\boldsymbol{c}_i$ are the ith row of $\boldsymbol{T}$ and $\boldsymbol{C}$, respectively, and $i\in[1,2,\ldots,m]$, $\boldsymbol{t}_i$, $\boldsymbol{c}_i\in\mathbb{R}^m$, let $\boldsymbol{\Phi}_T=[\boldsymbol{\varphi}_{T_1}, \boldsymbol{\varphi}_{T_2}, \cdots, \boldsymbol{\varphi}_{T_p}]^{\mathrm{T}}$ be $p\times n$ data matrix with $\boldsymbol{\varphi}_{T_i}\in\mathbb{R}^p$ ($p\le m$) representing p principal components of test data, each containing the corresponding n positions in the reduced space. To determine the proper number of principal components, we defined that the threshold of 95% information of the original data should be retained after PCA transformation, which could be decided by eigenvalues of covariance matrix of $\boldsymbol{T}$ as

$$\sum_{i=1}^{p}\lambda_i \Big/ \sum_{i=1}^{m}\lambda_i \ge 95\%. \tag{1}$$

And the $m\times p$ weight matrix $\boldsymbol{W}$ consists of the corresponding eigenvectors of λ_1, λ_2, …, λ_p. The relation between original difference matrix and reduced principal components can be expressed as

$$\boldsymbol{T}=\boldsymbol{W}\boldsymbol{\Phi}_T+\boldsymbol{\mu}_T. \tag{2}$$

It describes the relationship between the two sets of variables $\boldsymbol{T}$ and $\boldsymbol{\Phi}_T$, the parameter vector $\boldsymbol{\mu}_T$ consists of m mean values of data matrix $\boldsymbol{T}$, each mean value contains the corresponding n positions in the reduced space. Hence, $\boldsymbol{\Phi}_T$ can be expressed as

$$\boldsymbol{\Phi}_T=\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{T}-\boldsymbol{\mu}_T). \tag{3}$$

For comparison, the CAE data should be transformed into the same reduced space with test data. Hence, the PCA coefficient matrix $\boldsymbol{W}$ from test data is also applied to CAE data. Thus, the difference of the resulting reduced time series data between test and CAE data stems only from the data of themselves, not from the PCA parameters. So the reduced CAE data $\boldsymbol{\Phi}_C$ can be calculated as

$$\boldsymbol{\Phi}_C=\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{C}-\boldsymbol{\mu}_C). \tag{4}$$

According to the mutual independence amongst principal components, the non-diagonal terms of the covariance matrix of reduced data should all be zero or approaching zero. The data matrices $\boldsymbol{\Phi}_T$ and $\boldsymbol{\Phi}_C$ will be applied in the quantitative assessment step as discussed in the following subsections, and the eigenvalues of the covariance matrix of $\boldsymbol{T}$ are used as a weighting factor when producing the overall rating of the dynamic system.

2.2 Error assessments of dynamic responses

To minimize the influence of the interactions among features such as phase, magnitude, and shape, an objective metric named EARTH is developed by Sarin *et al*. (2010). The EARTH validation metric is divided into global response error and target point response error. The global response error can be defined as the error associated with the complete functional response with equal weight on each point. The three main components of global response error are phase error n_ε, magnitude error $\varepsilon_{\text{magnitude}}$, and shape error $\varepsilon_{\text{shape}}$. Target point errors can be defined as the errors associated with certain localized phenomena of interest and are generally application dependent. Therefore, they are not the focus of this study.

The phase error deals with the overall error in timing between two functional responses when considering all the points of the response, and it is depicted in Fig. 2a. Magnitude error is defined as the difference in amplitude of the two functional responses when there is no time lag between the two and it is depicted in Fig. 2b. Shape error deals with error associated with the shape of the functional responses, such as the number of peaks, valleys, and slope, etc., and it is depicted in Fig. 2c. A unique feature of the EARTH metric is using DTW to separate the interaction of phase, magnitude, and shape errors. DTW is an algorithm for measuring discrepancy between time histories and was first used in context with speech recognition in the 1960s (Rabiner and Huang, 1993). The time warping technique aligns peaks and valleys as much as possible by expanding and compressing the time axis according to a given cost (distance) function (Lei and Govindaraju, 2003). When calculating the magnitude and shape errors, the DTW method is employed to minimize the effect of local or target point errors.

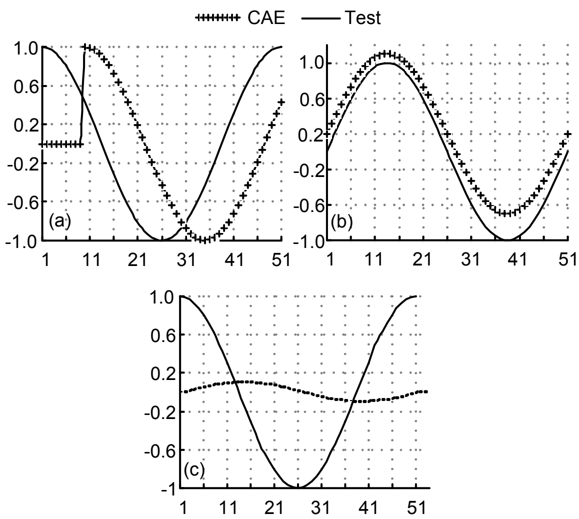

Fig. 2 Examples to illustrate the three types of global response errors
(a) Phase error; (b) Magnitude error; (c) Shape error

To calculate the phase error, the coefficient of correlation between test and CAE data is first calculated. The coefficient of correlation is a measure that indicates the extent of linear relationship between two time histories, i.e., to what extent can reduced CAE data $\boldsymbol{\Phi_C}$ be represented as $a_1\boldsymbol{\Phi_T}+b_1$ (where a_1 and b_1 are constants). The coefficient of correlation can range from −1 to +1. The value of +1 represents a perfect positive linear relationship between the time histories, which implies that they are both identical in shape but might be scaled. −1 would indicate a perfect negative linear relation which would indicate that the two time histories are scaled mirror images of each other. The coefficient of correlation is computed as

$$\rho_{\mathrm{E}}=\frac{\sum_{i=1}^{N}(\varphi_{C_i}-\bar{\boldsymbol{\Phi}}_{C})(\varphi_{T_i}-\bar{\boldsymbol{\Phi}}_{T})}{\left[\sum_{i=1}^{N}(\varphi_{C_i}-\bar{\boldsymbol{\Phi}}_{C})^2\sum_{i=1}^{N}(\varphi_{T_i}-\bar{\boldsymbol{\Phi}}_{T})^2\right]^{1/2}}. \tag{5}$$

The number of time steps shifted to maximize the coefficient of correlation, n_ε, is considered as the measure for error.

To calculate the magnitude error, the difference between the time histories caused by error in phase and shape need to be minimized. We can compensate for global time shift by shifting the time history by n_ε time steps. The resultant time histories after time shift are referred to as time shifted histories and are represented as $\boldsymbol{\Phi_C}^{\text{ts}}$ and $\boldsymbol{\Phi_T}^{\text{ts}}$. But even in these time shifted histories, there exist local timing errors between the histories. Also, error due to difference in slope cannot be treated as error in magnitude, and hence needs to be compensated for. In order to overcome these issues DTW (Lei and Govindaraju, 2003) is used.

The key idea of DTW is that any point of a time history can be (forward and/or backward) aligned with multiple points of the other time history that lie in different temporal positions, so as to compensate for temporal shifts (Capitani and Ciaccia, 2007). The cost function for warping is defined to penalize for distance and difference in slope between the two points. This ensures the mapping of a point to the closest point having similar slope on the other time history. To avoid scaling, the form of the cost function is (Sarin *et al.*, 2010)

$$d[i,j]=((\varphi_{C_i}^{\mathrm{ts}}-\varphi_{T_j}^{\mathrm{ts}})^2+(\varphi_{T_i}-\varphi_{T_j})^2)+\left|\left(\frac{\mathrm{d}\boldsymbol{\Phi}_C^{\mathrm{ts}}}{\mathrm{d}\boldsymbol{t}}\right)_{\boldsymbol{t}=\boldsymbol{t}_i}-\left(\frac{\mathrm{d}\boldsymbol{\Phi}_T^{\mathrm{ts}}}{\mathrm{d}\boldsymbol{t}}\right)_{\boldsymbol{t}=\boldsymbol{t}_j}\right|. \tag{6}$$

The warping path W defines an alignment between two time histories and, consequently, a cost to align them. The DTW distance is the minimum of such costs, i.e., the cost of the optimal warping path is give as

$$\mathrm{DTW}(\boldsymbol{\Phi}_C^{\mathrm{ts}},\boldsymbol{\Phi}_T^{\mathrm{ts}})^2=\min_W\left\{\sum_{[i,j]\in W}d[i,j]\right\}. \tag{7}$$

Fig. 3 depicts results before and after the DTW that are performed on the time histories. The warped time histories are now represented as $\boldsymbol{\Phi}_C^{\mathrm{ts+w}}$ and $\boldsymbol{\Phi}_T^{\mathrm{ts+w}}$. It can be observed that warping minimizes the local phase and shape effects. The L1 norm can now be used on the warped time shifted histories to isolate the contribution of magnitude error.

$$\varepsilon_{\mathrm{magnitude}}=\frac{\left\|\boldsymbol{\Phi}_C^{\mathrm{ts+w}}-\boldsymbol{\Phi}_T^{\mathrm{ts+w}}\right\|_1}{\left\|\boldsymbol{\Phi}_T^{\mathrm{ts+w}}\right\|_1}. \tag{8}$$

The shape error is a measure of discrepancy in shape of the two time histories. The shape of a time history is defined by the slope at each point. Therefore, the shape error is computed on the derivative of the time histories. In order to ensure that the effect of global time shift is minimized, the slope is calculated for the time shifted histories. Thus, by taking the derivative at each point, we obtain derivative time shifted histories represented by $\boldsymbol{\Phi}_C^{\mathrm{ts+d}}$ and $\boldsymbol{\Phi}_T^{\mathrm{ts+d}}$. Considering the derivative information ensures that the effect of magnitude is compensated for, as the derivative depends on the slope and not on the amplitude. The effect of localized time shifts though still exist (Fig. 4a). Thus, we use the same methodology to

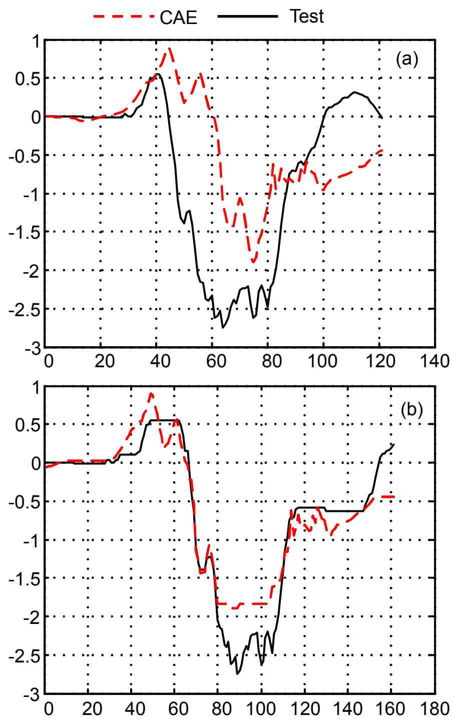

Fig. 3 Test vs. CAE time histories (a) before DTW and (b) after DTW

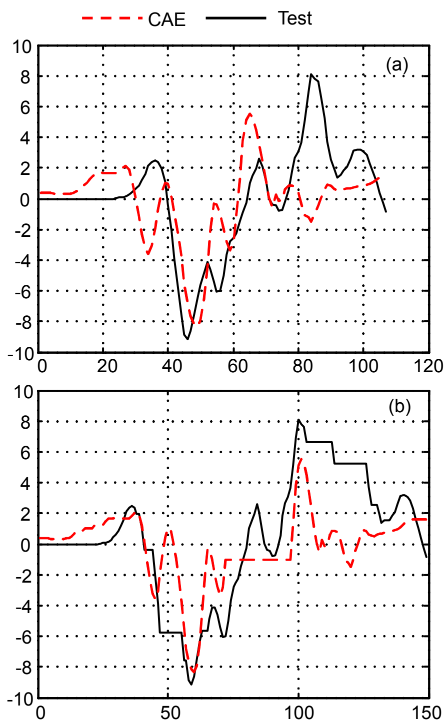

Fig. 4 Test vs. CAE derivative time histories (a) before DTW and (b) after DTW

evaluate the magnitude error on the derivative time shifted histories. The L1 norm of the warped derivative time shifted histories should quantify the isolated contribution of shape error. Fig. 4 depicts the derivative time histories before and after warping.

$$\varepsilon_{\text{shape}}=\frac{\left\|\boldsymbol{\Phi}_{C}^{\text{ts+d+w}}-\boldsymbol{\Phi}_{T}^{\text{ts+d+w}}\right\|_{1}}{\left\|\boldsymbol{\Phi}_{T}^{\text{ts+d+w}}\right\|_{1}}. \tag{9}$$

In addition to the inputs, there are two parameters that need to be set for executing this function and they are the time axis scale for warping (s), and the maximum warping window limit. It is a limit beyond which a point should not be warped in time. This also reduces the computation time for DTW.

2.3 SMEs' knowledge and PCA based model rating

The original EARTH metric (Sarin *et al.*, 2010) provides three independent values of phase, magnitude, and shape errors. Because the ranges of these errors are quite different, it is difficult for engineers to interpret how good or how bad a CAE model is based on these raw error data. To provide more intuitive rating based on the original EARTH metric, an enhanced EARTH rating score is proposed to combine the three errors of the original EARTH metric into one global score (Zhan *et al.*, 2011b). The enhanced EARTH score translates the original three errors into one score between 0 and 100%, so that it can provide an intuitive rating score and can be easily compared with other objective validation metrics.

Firstly, the original CAE curve is shifted one step at a time towards or away from the original test data, and the correlation coefficient between the truncated test data and CAE data is calculated until the maximum allowable time shift thresholds are reached. The number of the shifting steps resulted the maximum correlation coefficient is defined as the original EARTH phase error n_ε. In this study, Eq. (10) is used to calculate the enhanced EARTH phase score e_P, where n is the total number of the data points in the original functional responses, ε^*_P is the maximum allowable percentage of time shift, k_{E_p} defines the order of the regression. In this way, the best EARTH phase score is 100%, which means there is no need to shift CAE data to reach the maximum correlation coefficient between the original test and CAE data. If the shift is equal to or greater than the maximum allowable time shift threshold $\varepsilon^*_\text{P}\times n$, then the enhanced EARTH phase score is 0. In between, the EARTH phase score is calculated by regression method.

$$e_\text{P}=\begin{cases}100\%,\ n_\varepsilon=0,\\ 0,\ n_\varepsilon\ge\varepsilon_p^*\times n,\quad k_{E_\text{p}}\in\{1,2,3...\},\\ [(\varepsilon_\text{P}^*\times n-n_\varepsilon)/(\varepsilon_\text{P}^*\times n)]^{k_{E_\text{P}}},\ \text{otherwise.}\end{cases} \tag{10}$$

Secondly, the derivatives at each data point of the truncated test data, and the shifted and truncated CAE data are calculated, and they form the truncated test slope curve, and the shifted and truncated CAE slope curve. DTW is then performed on the truncated test data and the shifted and truncated CAE data. In this study, Eq. (11) is used to calculate the enhanced EARTH magnitude score e_M.

$$e_\text{M}=\begin{cases}100\%,\ \varepsilon_{\text{magitude}}=0,\\ 0,\ \varepsilon_{\text{magitude}}\ge\varepsilon_\text{m}^*,\ k_{E_\text{m}}\in\{1,2,3...\},\\ [(\varepsilon_\text{m}^*-\varepsilon_{\text{magitude}})/\varepsilon_\text{m}^*]^{k_{E_\text{m}}},\text{otherwise,}\end{cases} \tag{11}$$

where ε^*_m is the maximum allowable magnitude error, k_{E_m} defines the order of the regression.

Thirdly, DTW is performed on the truncated test slope curve and the shifted and truncated CAE slope curve. It results in the truncated and warped test slope curve, and the shifted, truncated, and warped CAE slope curve. Based on these two curves, the original EARTH shape error ε_shape is calculated (Sarin *et al.*, 2010). Eq. (12) is then used to calculate the enhanced EARTH shape score e_S.

$$e_\text{S}=\begin{cases}100\%,\ \varepsilon_{\text{shape}}=0,\\ 0,\ \varepsilon_{\text{shape}}\ge\varepsilon_\text{S}^*,\ k_{E_\text{s}}\in\{1,2,3...\},\\ [(\varepsilon_\text{S}^*-\varepsilon_{\text{shape}})/\varepsilon_\text{S}^*]^{k_{E_\text{s}}},\text{otherwise,}\end{cases} \tag{12}$$

where ε^*_S is the maximum allowable shape error, $k_{E\text{S}}$ defines the order of the regression.

The above three enhanced EARTH scores are then combined into one EARTH global score of a response by

$$E = w_P e_P + w_M e_M + w_S e_S, \quad (13)$$

where e_i and w_i represent the scores and weight factors of the phase, magnitude, and shape rating. An SMEs based calibration process was proposed in (Zhan *et al*., 2011b) to obtain the optimal setting of parameter values which have the least discrepancies with the experts' judgments. Hence, the knowledge from SMEs is incorporated into the evaluation results.

To combine the EARTH rating of all principal components, the variability of each principal component is used to calculate the weighting factor of each dimension. Let $E=[E_1, E_2, \ldots, E_p]$, so the multivariate overall rating E_M is calculated as

$$E_M = \sum_{i=1}^{p} (\lambda_i / \sum_{i=1}^{p} \lambda_i) \times E_i. \quad (14)$$

3 Case study

A driver side occupant restraint system MADYMO model (Fu *et al*., 2009; Zhan *et al*., 2011c) shown in Fig. 5 is used to demonstrate the proposed method. The model simulates a full frontal rigid barrier impact scenario at a speed of 35 m/h with a 50th percentile belted Hybrid III dummy (National Highway Traffic Safety Administration (NHTSA)) in a vehicle. This represents one of the USA New Car Assessment Program (NCAP) test modes.

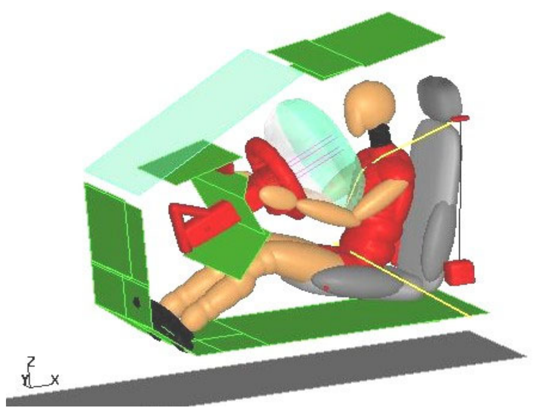

Fig. 5 A driver side occupant restraint system model

Fig. 6 and Fig. 7 show the time history plots of the test data and the corresponding results of model A and model B, with the 11 responses (Table 1). In model A, note that some CAE responses match well with the test (e.g., belt load at shoulder shown in Fig. 6e), but some do not (e.g., right femur load shown in Fig. 6g and upper neck moment shown in Fig. 6j). In model B, it can be observed that some responses, such as left femur load in *z*-direction (Fig. 7f), right femur load in *z*-direction (Fig. 7g), head acceleration in *x*-direction (Fig. 7h), and upper neck moment (Fig. 7j), are significantly better than model A. While some other responses are slightly worse comparing with the test data. Clearly, it is very difficult, if not impossible, to assess the quality, validity, and important feature agreements of the two CAE models with test data just using visual inspection.

Table 1 Eleven occupant responses

Response	Description
R_a	Chest deflection
R_b	Chest acceleration in *x*-direction
R_c	Belt load at anchor
R_d	Belt load at retractor
R_e	Belt load at shoulder
R_f	Femur load left in *z*-direction
R_g	Femur load right in *z*-direction
R_h	Head acceleration in *x*-direction
R_i	Upper neck load in *z*-direction
R_j	Upper neck moment
R_k	Pelvis acceleration in *x*-direction

The PCA method is then applied to the normalized test data, to reduce the dimensionality and take care of correlation of validation data. The proper number of the reduced dimensionality, which is the number of principal component, is first determined based on the 11×121 test data matrix $\boldsymbol{T}$ by predefining the amount of information in the original data to be considered. Assuming that the reduced data matrix in the PCA should account for at least 95% information in the original data matrix, the value of p=4 is obtained for the test data matrix, which accurately accounts for 95.6% information in the original data. Fig. 8 (p.129) shows the percentage of variability accounted for by each principal component. Clearly, the original eleven-dimensional data matrix is highly correlated and can be reduced to the data matrix with the dimension of four, which accounts for more than 95% information in the original data. In addition, it is observed that the first two principal components (p=2) contain more than 85% information in the original data matrix, while the first six principal

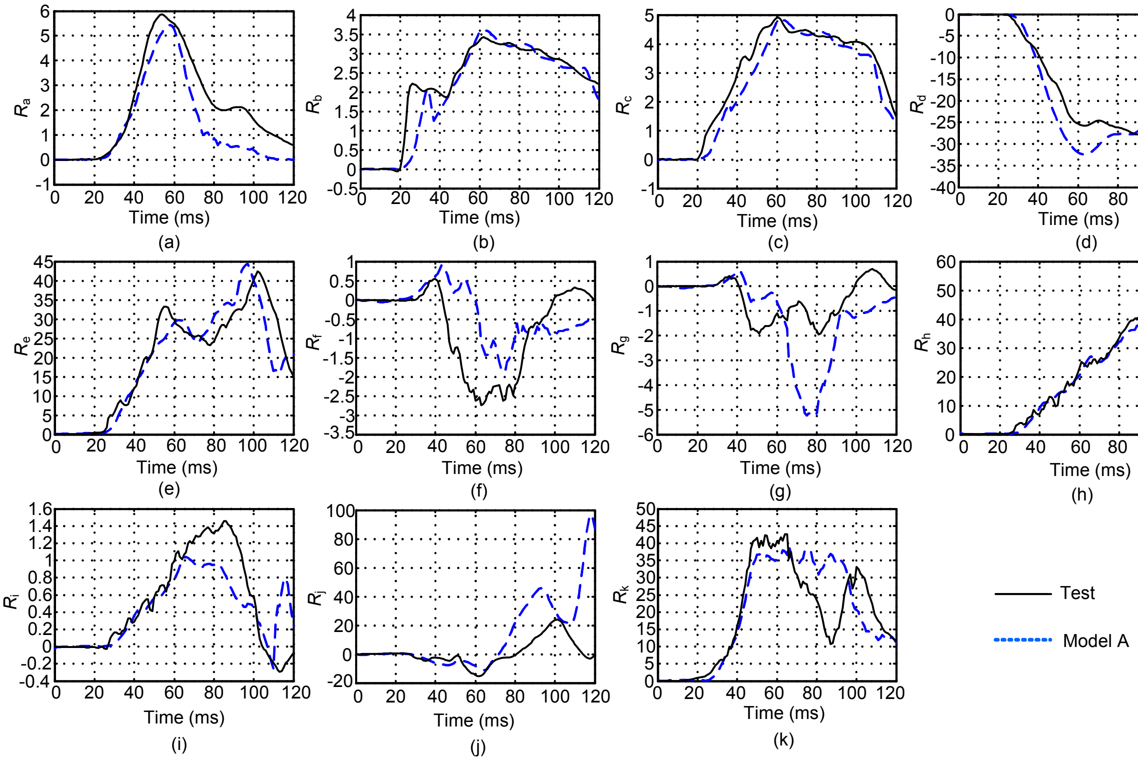

Fig. 6 Time history plots for the test and CAE model A with 11 responses (a)-(k)

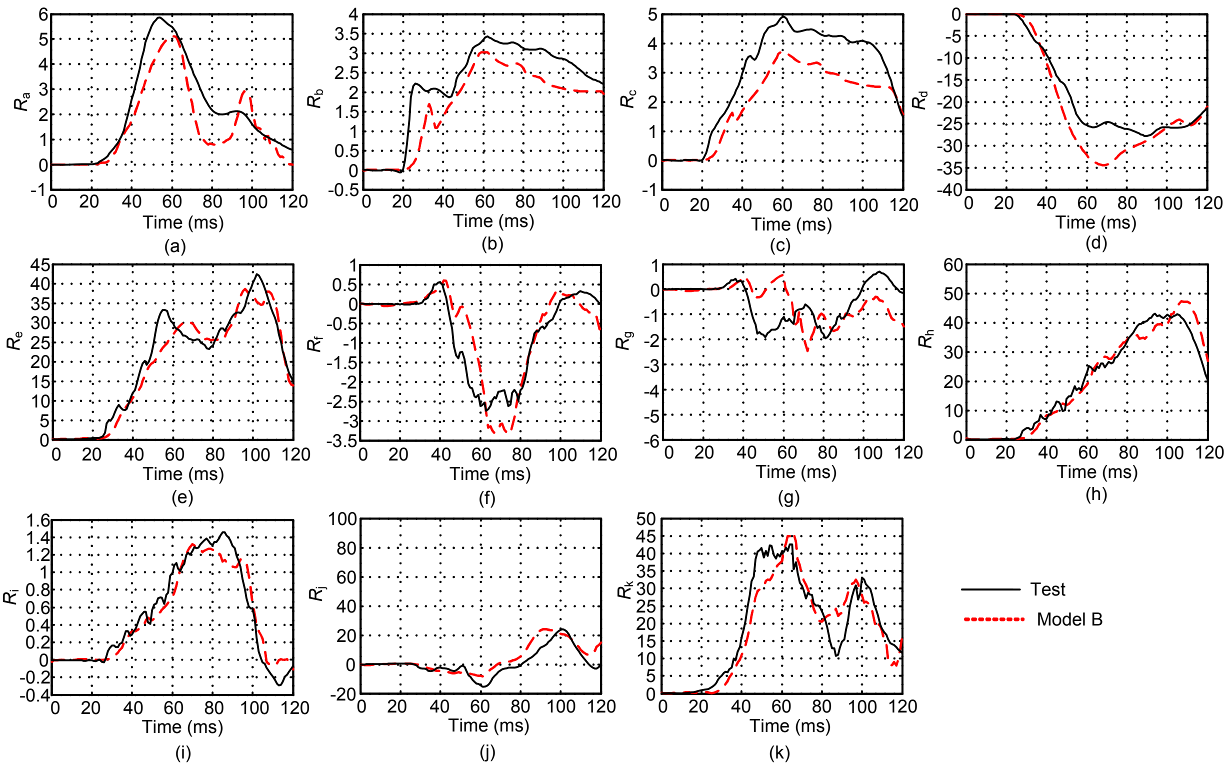

Fig. 7 Time history plots for the test and CAE model B with 11 responses (a)-(k)

components (p=6) contain or account for about 99% information. The information of difference principal components is used to rate the overall predictive capability of the CAE model.

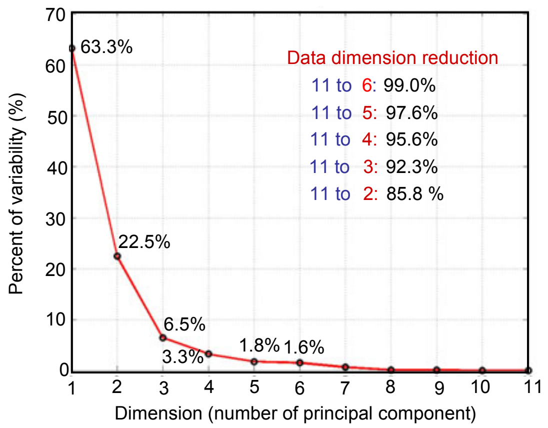

Fig. 8 PCA dimension reduction on test data

Fig. 9 and Fig. 10 show the first four principal components of the test data against the test PCA reduced data of two CAE models (p_1, p_2, p_3, and p_4), which account for more than 95% information of the original data. Note that the horizontal axis in each sub-figure represents the time interval, whilst the vertical axis represents the dimensionless magnitude. It is observed that both CAE models match well with the test data in terms of the first principal component which accounts for 63% of information in the original data, whilst for the next three principal components which account for 22.5%, 6.5%, and 3.3% of information respectively, the discrepancies between CAE model A and test data are significantly greater than those of the CAE model B by visual inspection.

Table 2 shows both the original EARTH metric errors and the enhanced EARTH ratings for two CAE models. Note that the phase error n_ε, magnitude error $\varepsilon_{\text{magnitude}}$, and shape $\varepsilon_{\text{shape}}$ given by the original EARTH metric are in different ranges, and they are not intuitive for users to interpret the goodness of the CAE model in a standard way. Extra knowledge or references are needed to understand the information provided by these errors and they are usually application dependent. The enhanced EARTH metric provides scores of phase, magnitude, shape errors, and the final global score in the range of 0 to 100%, which are consistent with the other validation metrics in the study for comparison (Zhan *et al*., 2011b). It is also noted that all the ratings from the enhanced EARTH metric are dependent on the values of the metrics parameters, and they can be customized by using different parameters for different applications (Eqs. (10)–(13)). The information accounted by different principal components is used to calculate the overall rating. As the bolded values shown in Table 2, model A is rated an overall MEARTH score of 56.6%, and model B is rated 64.1%. The results are consistent with the SMEs' opinions.

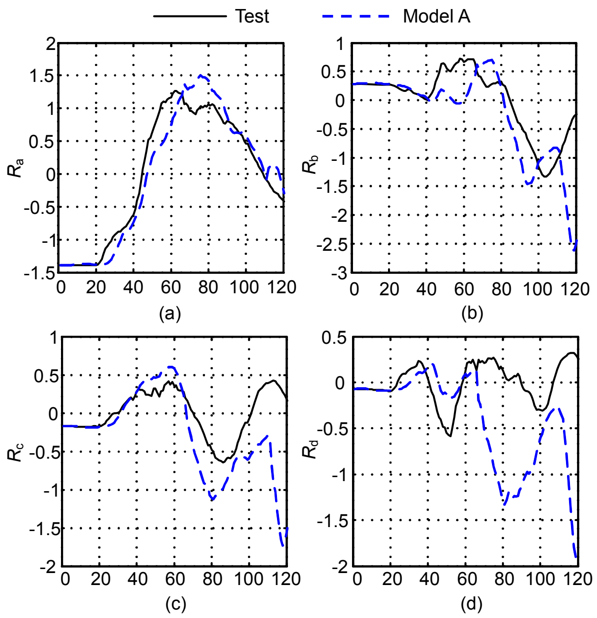

Fig. 9 Four principal components (a)-(d) of test and CAE model A

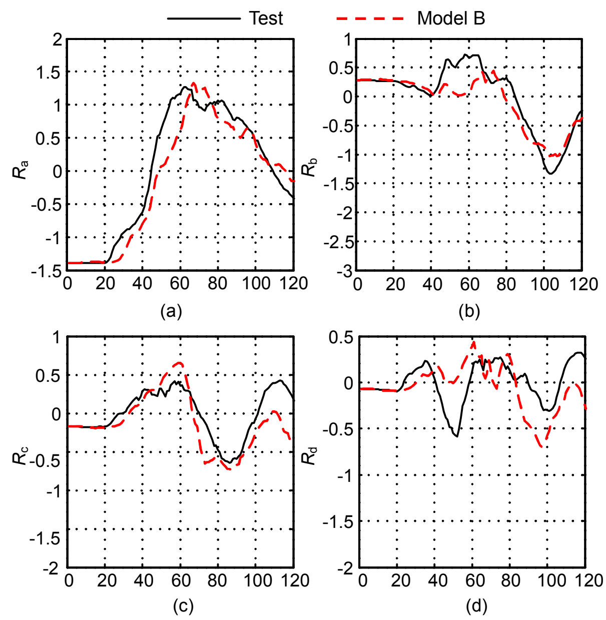

Fig. 10 Four principal components (a)-(d) of test and CAE model B

Table 2 Complete M-EARTH rating results for models A and B

Principal component		Phase (time step)	Magnitude error	Shape error	Phase score	Magnitude score	Shape score	M-EARTH score
Model A	p_1	5	0.1	0.6	78.30%	84.40%	48.60%	71.10%
	p_2	18	0.4	0.8	21.70%	29.30%	32.90%	27.60%
	p_3	5	0.6	0.8	78.30%	0.00%	31.80%	38.90%
	p_4	25	0.8	0.8	0.00%	0.00%	32.70%	10.10%
	Overall	8.7	0.2	0.7	62.20%	62.80%	43.20%	**56.60%**
Model B	p_1	6	0.1	0.6	73.90%	84.80%	50.00%	70.00%
	p_2	1	0.2	1.2	95.70%	58.60%	0.70%	54.60%
	p_3	2	0.4	0.8	91.30%	28.80%	37.20%	54.70%
	p_4	3	0.9	1.1	87.00%	0.00%	11.00%	35.80%
	Overall	4.4	0.2	0.8	80.70%	71.90%	36.20%	**64.10%**

5 Conclusions

Development and selection of an appropriate objective metric is one of the most important factors to achieve successful applications of model validation. There are several critical ideal characteristics to be considered when selecting a model validation metric, such as objectiveness, generalization, simplicity, with physical meaning and able to incorporate engineering knowledge. In addition, most dynamic systems require considering the differences of both the functional responses and the key features, such as phase, magnitude, and shape between test and CAE data. A metric named M-EARTH is developed in this paper to provide one intuitive score to assess the validity of the CAE model with multiple dynamic responses simultaneously for its intended usage. The PCA method is used to reduce data dimension and address multivariate correlation. The correlation coefficient and DTW are used to calculate the phase, magnitude, and shape error measures of each response. The physical-based thresholds, SMEs' knowledge, and PCA coefficients are incorporated to obtain the overall assessment of the model consisting of multiple dynamic responses. A real-world example with two CAE models is used to demonstrate the effectiveness and advantages of the M-EARTH metric. The results show that the M-EARTH metric not only maintains the existing advantages of the original EARTH method, but is also superior to the existing metric. The advantages of the M-EARTH are as follows: (1) it is capable of differentiating the goodness of CAE models with multiple dynamic responses; (2) it can provide an intuitive rating score; (3) it is scalable to other applications; and (4) it is consistent with SMEs' knowledge. Furthermore, since most of the M-EARTH metric parameters have clear physical meanings, they may be reused for similar applications. Note that the example in this paper is analytical in nature, and it has been published elsewhere (Fu *et al.*, 2009; Zhan *et al.*, 2011c).

References

Capitani, P., Ciaccia, P., 2007. Warping the time on data streams. *Data & Knowledge Engineering*, **62**(3):438-458. [doi:10.1016/j.datak.2006.08.012]

Ferson, S., Oberkampf, W.L., Ginzburg, L., 2008. Model validation and predictive capability for the thermal challenge problem. *Computer Methods in Applied Mechanics and Engineering*, **197**(29-32):2408-2430. [doi:10.1016/j.cma.2007.07.030]

Fu, Y., Jiang, X., Yang, R.J., 2009. Auto-Correlation of an Occupant Restraint System Model Using a Bayesian Validation Metric. SAE World Congress, Detroit, USA, SAE 2009-01-1402.

Fu, Y., Zhan, Z.F., Yang, R.J., 2010. A Study of Model Validation Method for Dynamic Systems. SAE World Congress, Detroit, USA, SAE 2010-01-0419.

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, **24**(6):417-441. [doi:10.1037/h0070888]

Jiang, X., Mahadevan, S., 2008. Bayesian wavelet method for multivariate model assessment of dynamical systems. *Journal of Sound and Vibration*, **312**(4-5):694-712. [doi:10.1016/j.jsv.2007.11.025]

Jiang, X., Yang, R.J., Barbat, S., Weerappuli, P., 2009. Bayesian probabilistic PCA approach for model validation of dynamic systems. *SAE International Journal of Materials & Manufacturing*, **2**(1):555-563. [doi:10.4271/2009-01-1404]

Joliffe, I.T., 2002. Principal Component Analysis. Springer, New York, USA.

Lei, H., Govindaraju, V., 2003. Synchronization of Batch Trajectory Based on Multi-Scale Dynamic Time Warping. Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xi'an, China.

Mahadevan, S., Rebba, R., 2005. Validation of reliability computational models using Bayes networks. *Reliability Engineering and System Safety*, **87**(2):223-232. [doi:10.1016/j.ress.2004.05.001]

Oberkampf, W.L., Barone, M.F., 2006. Measures of agreement between computation and experiment: validation metrics. *Journal of Computational Physics*, **217**(1):5-36. [doi:10.1016/j.jcp.2006.03.037]

Oberkampf, W.L., Trucano, T.G., 2006. Design of and Comparison with Verification and Validation Benchmarks. Technical Report Sand No. 2006-5376C, Sandia National Laboratories, Albuquerque, New Mexico, USA.

Rabiner, L.R., Huang, B.H., 1993. Fundamentals of Speech Recognition, Prentice Hall.

Rebba, R., Mahadevan, S., 2006. Model predictive capability assessment under uncertainty. *AIAA Journal*, **44**(10):2376-2384. [doi:10.2514/1.19103]

Sarin, H., Kokkolaras, M., Hulbert, G., Papalambros, P., Barbat, S., Yang, R.J., 2010. Comparing time histories for validation of simulation models: error measures and metrics. *Journal of Dynamic Systems Measurement and Control*, **132**(6):061401. [doi:doi:10.1115/1.4002478]

Schwer, L.E., 2007. Validation metrics for response histories: perspectives and case studies. *Engineering with Computers*, **23**(4):295-309. [doi:10.1007/s00366-007-0070-1]

Zhan, Z.F., Fu, Y., Yang, R.J., Peng, Y.H., 2011a. An enhanced Bayesian based model validation method for dynamic systems. *Journal of Mechanical Design*, **133**(4):041005. [doi:10.1115/1.4003820]

Zhan, Z.F., Fu, Y., Yang, R.J., 2011b. Enhanced Error Assessment of Response Time Histories (EEARTH) Metric and Calibration Process. SAE World Congress, Detroit, USA, SAE 2011-01-0245.

Zhan, Z.F., Fu, Y., Yang, R.J., Peng, Y.H., 2011c. An automatic model calibration method for occupant restraint systems. *Structural and Multidisciplinary Optimization*, **44**(6):815-822. [doi:10.1007/s00158-011-0671-6]