

An Integrated Validation Method for Nonlinear Multiple Curve Comparisons

Zhenfei Zhan and Junqi Yang
Chongqing University

Xueqian Chen
Institute of Systems Engineering CAEP

Zhanpeng Shen
Engineering & Design

ABSTRACT

In automobile industry, computational models built to predict the performances of the prototype vehicles are on the rise. To assess the validity or predictive capability of the model for its intended usage, validation activities are conducted to compare computational model outputs with test measurements. Validation becomes difficult when dealing with dynamic systems which often involve multiple functional responses, and the complex characteristics need to be appropriately considered. Many promising data analysis tools and metrics were previously developed to handle data correlation and evaluate the errors in magnitude, phase shift, and shape. However, these methods show their limitations when dealing with nonlinear multivariate dynamic systems. In this paper, kernel function based projection is employed to transform the nonlinear data into linear space, followed by the regular principal component analysis (PCA) based data processing. The agreement between the reduced test and model outputs are then evaluated based on the validation metric of EEARTH. The proposed strategy helps to better represent the original data after dimension reduction, which improves the credibility of the afterwards multivariate validation. A vehicle design case is employed to demonstrate the effectiveness and advantages of the proposed method.

CITATION: Zhan, Z., Yang, J., Chen, X., and Shen, Z., "An Integrated Validation Method for Nonlinear Multiple Curve Comparisons," *SAE Int. J. Mater. Manf.* 9(2):2016, doi:10.4271/2016-01-0288.

INTRODUCTION

With the ever shortening time 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 these models can be used, their 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 in model validation is to select a validation metric that has desirable properties for assessing dynamic system with multiple functional responses [1, 2]. Developing comprehensive validation methods has attracted considerable researchers' interest in recent years [3]. Statistical theory is a workable approach to provide quantitative assessment in model validation [4, 5]. For instance, Jiang et al. [6, 7], Fu et al. [8] and Zhan et al. [9] exploited the Bayesian based validation methods for dynamic systems in various applications. However, these methods considered the whole distribution of interested time history, the agreements of the important features were not specifically addressed. In order to evaluate the agreement of crucial features, Schwer [3] and Sarin et al. [10] evaluated two categories of metrics for dynamic systems, namely, Sprague and

Geers metric and Knowles and Gear metric. These metrics extract the magnitude and phase features based on the time integration operation for the waveforms. More recently, Zhan et al. [11] proposed an enhanced method based on error assessment of response time history (EEARTH) metric. These methods, however, aim to quantitatively compare two curves (i.e., model prediction and test data) for a single response. It is difficult for these metrics to handle multivariate and correlated data directly. Hence, how to effectively perform validation for multivariate and complex dynamic systems, such as vehicle crash tests, is still a challenging task. A straight forward solution is to average all the assessments into one evaluation. However, it fails to provide useful information when the agreements of different quantities are largely different, and data correlation is not neglectable. To address such issue, Zhan et al. [12] proposed an integrated validation method, namely M-EARTH, for multivariate dynamic systems, in which Principal Component Analysis (PCA) is adopted to address the conflicting inferences caused by multivariate correlation and improve the effectiveness and efficiency of validation. However, conventional PCA is a linear projecting tools, and it is not appropriate to be utilized to tackle nonlinear data.

In this paper, an integrated method is proposed for the model validation of nonlinear multivariate dynamic systems. The remainder of the paper is organized as: Firstly, the PCA is introduced, followed by the proposed kernel function based PCA (K-PCA). Then the previously proposed EEARTH based dynamic curve comparison method is illustrated briefly. The proposed multivariate validation method is applied to a vehicle design case. Summary is provided in the end.

PROPOSED NONLINEAR MULTIVARIATE VALIDATION METHOD

Figure 1 presents the flowchart of the proposed multivariate validation method. It consists of several major steps: (1) kernel function based data projection, (2) PCA based dimension reduction, (3) error assessments for dynamic responses, (4) and Subject Matter Experts (SMEs) based response evaluation. Firstly, comparable multivariate data from test and CAE are normalized. The K-PCA is then applied to reduce data dimension and also to address multivariate data correlation. Then the error assessments are conducted for the reduced test and CAE data. Three independent errors including phase, magnitude, and shape are calculated. Based on the SMEs' knowledge, the three EARTH errors are transformed and combined as intuitive score ranging from 0 to 100% for each obtained principal component. Finally, the scores of all the reduced responses can be combined through the PCA coefficient into an overall score. Based on the results, the decision maker (DM) then decides to accept or reject the model.

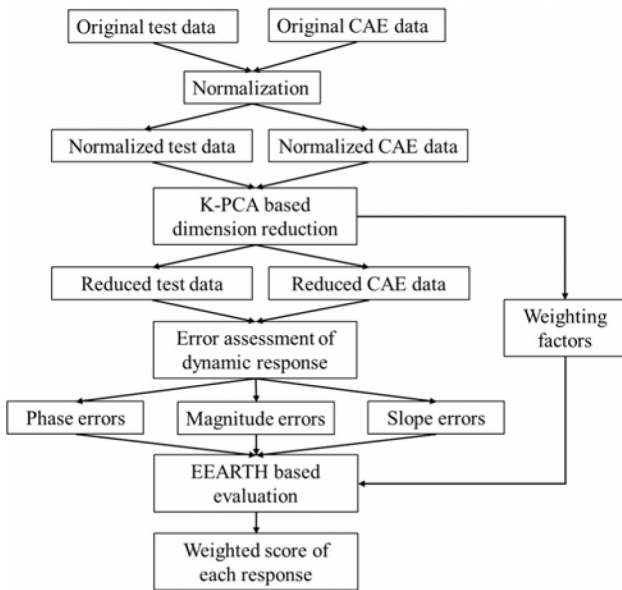


Figure 1. Framework of the proposed nonlinear multivariate validation method.

PCA Based Dimension Reduction

PCA is a well-established statistical method for dimensionality reduction and has been widely applied in various fields including data compression, image processing, exploratory data analysis, pattern recognition, and time series prediction [13, 14]. 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. The important property of PCA is that the principal component projection minimizes the squared reconstruction error in dimensionality reduction. PCA transforms highly correlated original responses into uncorrelated lower dimensional responses. 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 $T=[t_1, t_2, \dots, t_m]^T$, $C=[c_1, c_2, \dots, c_m]^T$ represent the $m \times n$ physical test data and CAE model predictions, t_i , c_i are the i th row of T and C , respectively, and $i \in [1, 2, \dots, m]$, $t_i, c_i \in R^n$.

$\Phi_T=[\phi_{T_1}, \phi_{T_2}, \dots, \phi_{T_p}]^T$ is a $p \times n$ data matrix with $\Phi_T \in R^p$ ($p \leq m$) 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, threshold of 95% information of the original data should be retained after PCA transformation, which could be decided by the eigenvalues of covariance matrix of T .

$$\sum_{i=1}^p \lambda_i / \sum_{i=1}^m \lambda_i \geq 95\% \quad (1)$$

W consists of the corresponding eigenvectors $\lambda_1, \lambda_2, \dots, \lambda_p$. The correlation between the original difference matrix and the reduced principal components can be expressed as:

$$T = W\Phi_T + \mu_T \quad (2)$$

The parameter vector μ_T consists of m mean values of data matrix T , each mean value contains the corresponding n sample points in the reduced space. Φ_T can be expressed as:

$$\Phi_T = W^T (T - \mu_T) \quad (3)$$

For comparison, the CAE data should be transformed into the same reduced space with test data. Hence, the PCA coefficient matrix W of test data is also applied to CAE data. So the reduced CAE data Φ_C can be calculated as:

$$\Phi_C = W^T (C - \mu_C)$$