

A Bayesian Inference based Model Interpolation and Extrapolation

Zhenfei Zhan, Yan Fu and Ren-Jye Yang
Ford Motor Co.

Zhimin Xi
Univ. of Michigan

Lei Shi
Ford Motor Co.

ABSTRACT

Model validation is a process to assess the validity and predictive capabilities of a computer model by comparing simulation results with test data for its intended use of the model. One of the key difficulties for model validation is to evaluate the quality of a computer model at different test configurations in design space, and interpolate or extrapolate the evaluation results to untested new design configurations. In this paper, an integrated model interpolation and extrapolation framework based on Bayesian inference and Response Surface Models (RSM) is proposed to validate the designs both within and outside of the original design space. Bayesian inference is first applied to quantify the distributions' hyper-parameters of the bias between test and CAE data in the validation domain. Then, the hyper-parameters are extrapolated from the design configurations to untested new design. They are then followed by the prediction interval of responses at the new design points. A vehicle design of front impact example is used to demonstrate the proposed methodology.

CITATION: Zhan, Z., Fu, Y., Yang, R., Xi, Z. et al., "A Bayesian Inference based Model Interpolation and Extrapolation," *SAE Int. J. Mater. Manf.* 5(2):2012, doi:10.4271/2012-01-0223.

INTRODUCTION

Computer modeling and simulation have become one of the fundamental research areas of product design and development in the automotive industry. Computational Aided Engineering (CAE) is widely used to simulate vehicle structure interior, and restraint system in various impact modes. To maximize the benefits of these models, the validity and predictive capabilities of these models need to be assessed quantitatively. Model validation [1, 2, 3, 4, 5, 6] is the process of comparing CAE model outputs with test measurements in order to assess the validity and predictive capabilities of the CAE model for its intended usage. Successful implementation of model validation will increase the chance of virtual testing and may lead to significant reduction in prototype building and testing of vehicle designs.

In numerical analysis, interpolation is a method of constructing new data points within the range of a discrete set of known data points. Extrapolation is a process in which information is gained by obtaining function from the known data, and then estimating the value of a new data point which

is beyond the range of given data by using the extended (extrapolated) function as the source. Typically, the quality of a particular extrapolation method is limited by the assumptions about the function made by the method. For example, if a method assumes the data are smooth, then a non-smooth function will be poorly extrapolated. Even for proper assumptions about the function, the extrapolation can diverge strongly from the function. In general, CAE models are not developed solely for single design point analysis. It is required that the CAE model possesses dependable predictive capability under different design scenarios within a predefined design space. Additionally, CAE models can have a large amount of inputs. It is desirable that a CAE model is valid in a wide range of design variables (for design space exploration). A CAE model may have to be validated for the entire range of its inputs before it can be claimed as a "globally validated model", which will always yield designs that will perform as predicted. This requires a vast amount of resources to conduct multiple tests at each point of a sufficiently large data set in an extremely high-dimensional

space defined by the model inputs. However, due to limited resources and time constraints, the model validation activities are usually conducted in a well-controlled environment where only a limited number of physical tests are available [7]. All of these call for the development of an effective and systematic model interpolation and extrapolation method to rigorously and effectively interpolate and extrapolate the inference of CAE models to the new design points both within and beyond the design space.

To achieve model interpolation and extrapolation, various statistical inference techniques have been developed, such as χ^2 (Chi-square) test on residuals between model and probability density design objective experimental results [9, 10] require multiple evaluations of the model and experiments, and many statistical assumptions that are difficult to satisfy. Bayesian methods have been investigated extensively for validation and calibration of complex CAE models [5, 8, 9, 10, 11, 12, 13, 14, 15]. The Bayes network links the computational model from component level to system level in the context of both physics-based extrapolation and statistical extrapolation. Rebba and Mahadevan [11, 12] developed and implemented a Bayesian framework for drawing inferences for predictions in the untested domain using Bayes networks. A Bayes network approach is proposed for validating the reliability assessment made by computational models. However, it was numerically expensive to obtain sufficient model evaluations for the Bayes network, and the emphasis of the method was on validating the modeling accuracy at tested design points, but not in the context of a new design. Jiang and Mahadevan [13] proposed Bayesian-hypothesis-testing-based methodology for model validation and confidence extrapolation under uncertainty, using limited test data. The Bayesian network with Markov Chain Monte Carlo simulation and Gibbs sampling was used for extrapolating the inference from the validated domain at the component level to the untested domain at the system level. Kim and Kwak [16] introduced the concept of the design space optimization. By starting with a small design space and advancing to a larger space, the method eventually achieves an optimal design space. Chen et al [17], Li et al [18] and Kakkolaras et al [19] brought up the need for validating models from the perspective of engineering design, and provide stochastic assessment of the validity of a model. However, these methods didn't fully address the needs for design interpolation and extrapolation.

This paper proposes a Bayesian based model interpolation and extrapolation method which utilizes the advantages of Bayesian inference and response surface model (RSM). The predictive model of the bias between test and CAE are built in the design space, the prediction interval of the output value at the new design points are then calculated given the available test and CAE data. In the following sections, the proposed Bayesian based model interpolation and extrapolation method is first presented, followed by the introduction of three key techniques of the proposed method including the Bayesian updating of the prediction bias, the

RSM construction of the prediction bias, and the prediction interval calculation for the new design, and then a vehicle system example is used to demonstrate the effectiveness of the proposed method. Finally, some discussions and summary are given in the end.

BAYESIAN UPDATING BASED MODEL INTERPOLATION AND EXTRAPOLATION

Essential to the proposed method is the prediction bias which measures the difference between the CAE model output and the test value. Figure 1 shows the flowchart of the Bayesian based model interpolation and extrapolation method. The process starts with design of experiment (DOE) matrix for validation in the design space, followed by performing repeated physical tests and CAE simulations. The difference between test and CAE data are then calculated as evidence for the Bayesian updating of hyper-parameters of the bias distribution. After obtaining the prior information, the posterior distributions of prediction bias hyper-parameters are calculated. Then, two RSMs are built for the mean and standard deviation of the prediction bias. Next, the CAE simulations are performed. Then the Prediction Intervals (PIs) of the output at new designs are calculated. Confirmation tests at new designs are used to validate the prediction results. If the test result of a new design is within the corresponding PI, the prediction by interpolation or extrapolation is considered successful. Based on the result, the decision maker then decides to accept or reject the prediction result. If accepted, it is then passed to the downstream engineers for design iterations or design optimization. On the other hand, if the prediction is rejected, additional DOE will be added and repeat the process. Some of the major steps are described in details in the following subsections.

BAYESIAN UPDATING OF THE PREDICTION BIAS

This paper proposes a Bayesian based model interpolation and extrapolation method which can utilize the advantages of Bayesian inference and response surface model (RSM). For the convenience of formulation, let $\mathbf{x}$ represents design variables, Y_{test} and Y_{CAE} stand for test data and CAE model results, respectively. The relationship between the test data $Y_{\text{test}}(\mathbf{x})$ and the result generated by a CAE model $Y_{\text{CAE}}(\mathbf{x})$ can often be generalized as follows:

$$Y_{\text{test}}(\mathbf{x}) = Y_{\text{CAE}}(\mathbf{x}) + \delta(\mathbf{x}) + \varepsilon(\mathbf{x}) \quad (1)$$

where ε is the random variable representing the test error (relating to both experimental setup and measurement) that may depend on $\mathbf{x}$, and δ is the error of the CAE model. Hence, it can be drawn that the “reality” or “true value” of output, namely Y_{true} , is equal to the test observation minus

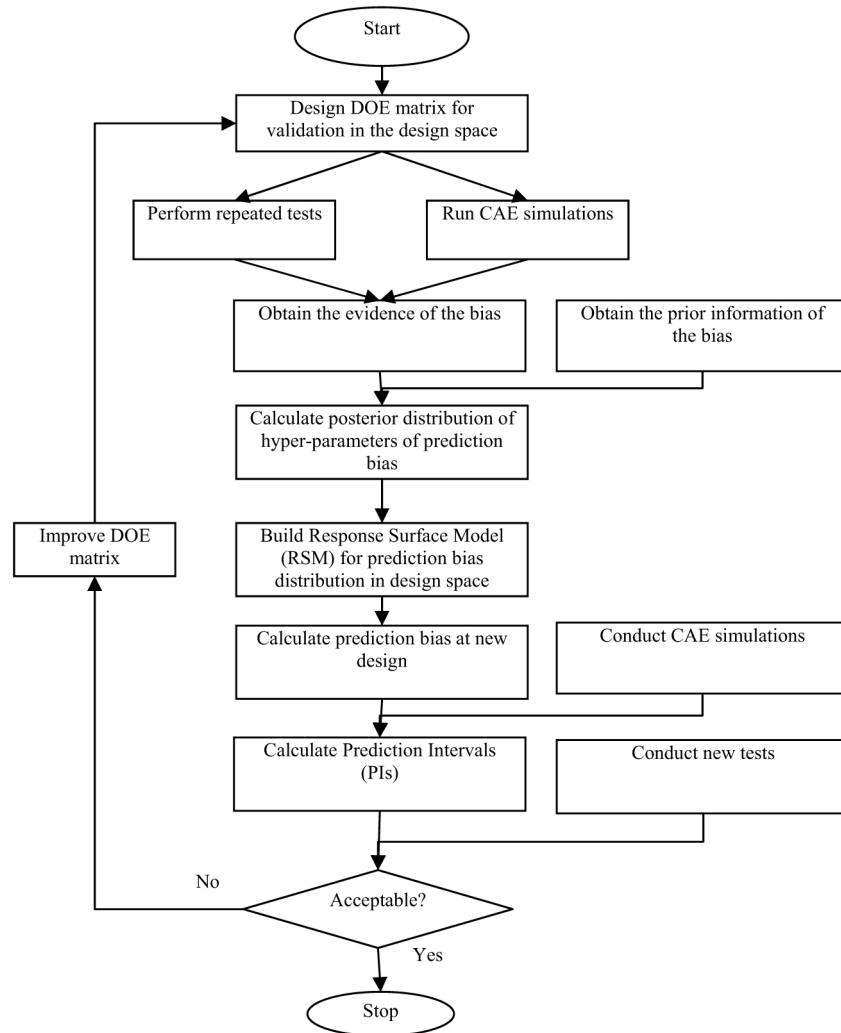


Figure 1. Flowchart of Bayesian based Model Interpolation and Extrapolation

the test error, and the CAE model prediction plus the model bias, as shown in Equation 2.

$$Y_{\text{true}}(\mathbf{x}) = Y_{\text{test}}(\mathbf{x}) - \varepsilon(\mathbf{x}) = Y_{\text{CAE}}(\mathbf{x}) + \delta(\mathbf{x}) \quad (2)$$

Assume $\tau(\mathbf{x}) = \delta(\mathbf{x}) + \varepsilon(\mathbf{x})$, it is noted that the prediction bias is an essential step for estimating the value of Y_{true} . For the convenience of formulation and method demonstration, it is reasonable to assume the bias $\tau(\mathbf{x}) = \delta(\mathbf{x}) + \varepsilon(\mathbf{x})$ follows a normal distribution with known variance as:

$$\tau(\mathbf{x}) \sim N(\tau_{\mu}(\mathbf{x}), \tau_{\sigma^2}(\mathbf{x})) \quad (3)$$

As shown previously, the bias $\tau(\mathbf{x}) = \delta(\mathbf{x}) + \varepsilon(\mathbf{x})$ follows a normal distribution with known variance as $\tau(\mathbf{x}) \sim N(\tau_{\mu}(\mathbf{x}), \tau_{\sigma^2}(\mathbf{x}))$. Given the known variance τ_{σ^2} , the mean value τ_{μ} needs to be inferred with available data. In Bayesian probability theory, if the posterior distributions $p(\theta | \mathbf{x})$ are in the same family as the prior probability distribution $p(\theta)$, the

prior and posterior are then called conjugate distributions, and the prior is called a conjugate prior for the likelihood. For example, the Gaussian family is conjugate to itself (or self-conjugate) with respect to a Gaussian likelihood function. If the likelihood function is Gaussian, choosing a Gaussian prior over the mean will ensure that the posterior distribution is also Gaussian. In this study, the prior distribution of τ_{μ} can be assumed as:

$$\tau_{\mu 0} \sim N(\mu_0, \sigma_0^2) \quad (4)$$

Consider the general problem of inferring a distribution for a parameter θ given some datum or data $\mathbf{x}$. From Bayes' theorem, the posterior distribution is equal to the product of the likelihood function and prior $p(\theta)$, normalized (divided) by the probability of the data $p(\mathbf{x})$: