

A Stochastic Bias Corrected Response Surface Method and its Application to Reliability-Based Design Optimization

Zhenfei Zhan
Chongqing Univ.

Yan Fu and Ren-Jye Yang
Ford Motor Co.

ABSTRACT

In vehicle design, response surface model (RSM) is commonly used as a surrogate of the high fidelity Finite Element (FE) model to reduce the computational time and improve the efficiency of design process. However, RSM introduces additional sources of uncertainty, such as model bias, which largely affect the reliability and robustness of the prediction results. The bias of RSM need to be addressed before the model is ready for extrapolation and design optimization. This paper further investigates the Bayesian inference based model extrapolation method which is previously proposed by the authors, and provides a systematic and integrated stochastic bias corrected model extrapolation and robustness design process under uncertainty. A real world vehicle design example is used to demonstrate the validity of the proposed method.

CITATION: Zhan, Z., Fu, Y., and Yang, R., "A Stochastic Bias Corrected Response Surface Method and its Application to Reliability-Based Design Optimization," *SAE Int. J. Mater. Manf.* 7(2):2014, doi:10.4271/2014-01-0731.

INTRODUCTION

In automotive industry, Finite Element (FE) models, as the most commonly used computer models, are increasingly employed in various stages of vehicle design and engineering. 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 Computational Aided Engineering (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. However, even with increasing speed of computers, the simulation time of full FE models is still too time-consuming to use it for direct design optimization or robust design. It sometimes takes hours or even days to complete a single FE simulation run. Hence, response surface model (RSM) is commonly used as a surrogate of the FE model to reduce the development time. Various RSM techniques have been developed in recent years including radial basis function (RBF) [7], artificial neural network (ANN), support vector regression (SVR), and Kriging (KRG) [8, 9]. The effectiveness of different RSM techniques varies based on the different modeling criteria, amount of available samples, and the nonlinear nature of the physical phenomena itself [10, 11]. No matter what RSM method is used, there is always some discrepancy between the results of the true FE model and the RSM. This is especially true when the RSMs are used to predict new designs in terms of interpolation and extrapolation. This calls for the development

of stochastic model interpolation and extrapolation method that can address the discrepancy between the RSM and FE models and provide bias-corrected prediction intervals of RSM model responses under uncertainty.

Recently, several statistical inference techniques have been developed [12]. However, these methods did not fully address the real-world needs for design interpolation and extrapolation under uncertainty. With these considerations, three interpolation and extrapolation methods are investigated by Zhan et al.[21]. They are Bayesian inference-based method [22], Gaussian process modeling (GPM)-based method [23], and Copula-based method [24]. Three stochastic validation metrics, such as area metric, reliability-based metric, and Bayesian confidence metric, are used to evaluate the predictive capability of these three methods. It is found that all three methods are capable of providing good predictions for the purpose of interpolation and extrapolation. Among the three methods, the Bayesian inference-based method is more straightforward, effective, has less tuning parameters, and is easy for engineering implementation. The Gaussian process method shows superiority in capturing the mean of the responses when less training data are available. However, it might be difficult to determine the hyper-parameters of the Gaussian process when solving high-dimensional problems. In addition, the drastically changing prediction intervals might have some potential negative effects on identification of a better design. Copula-based method is able to capture general

statistical relationships between model prediction and model bias, and between design variables and model bias. This method may be good to handle large bias noisy situation. However, the implementation process is complex.

Reliability-based design optimization (RBDO) is a frequently used approach in design under uncertainty aiming at obtaining a design solution that is most reliable under the variations of design variables. Beyond deterministic design optimization, which often drives the optimal design to the limit of design constraints and leaves little or no latitude for uncertainty, RBDO considers the uncertainty of controllable or uncontrollable variables associated with an existing CAE model. However, most of the traditional RBDO formulations do not account for the uncertainty from the CAE model; a CAE model is believed to be accurate enough for reliability assessment before it is implemented in RBDO process. RBDO fails if the model is not accurate and is affected with large model uncertainty. Therefore, model bias correction is critical before implementing RBDO.

The rest of the paper is organized as follows: The Bayesian inference based stochastic interpolation and extrapolation method is briefly introduced, followed by the comparison of three stochastic model validation metrics. Next, RBDO formulation is introduced and a vehicle frontal impact case study are used to investigate the capability of the proposed approach, and the FE simulations of the robust design are used to confirm the results. Finally, some discussions and summary are given in the end.

BAYESIAN INFERENCE-BASED BIAS CORRECTED RSM METHOD

Zhan et al [21] proposed a Bayesian inference-based model interpolation and extrapolation method. This method utilizes the advantages of Bayesian inference and response surface model (RSM). 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.

For the convenience of formulation, let $\mathbf{x}$ represents design variables, Y_t and Y_m stand for FE model result (test) and RSM model result (model) respectively. The relationship between $Y_t(\mathbf{x})$ and $Y_m(\mathbf{x})$ in $\mathbf{x}$ can often be generalized as follows:

$$Y_t(\mathbf{x}) = Y_m(\mathbf{x}) + \delta(\mathbf{x}) + \varepsilon(\mathbf{x}) \quad (1)$$

where ε is the random variable representing the test error that may depend on $\mathbf{x}$, and δ is the bias 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 the test error, and the CAE model prediction plus the model bias, as shown in Equation 2.

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

Following Equation 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 true 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:

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

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. For example, the Gaussian family is 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 $p(\mathbf{x})$ as Equation 5:

$$p(\theta | \mathbf{x}) = \frac{p(\mathbf{x} | \theta)p(\theta)}{\int p(\mathbf{x} | \theta)p(\theta)d\theta} \quad (5)$$

Given the assumption of normally distributed conjugate prior, let n denote the number of test observations, the posterior distribution of τ_μ can be expressed as

$$\tau_{\mu 1} \sim N(\mu_1, \sigma_1^2) \quad (6)$$

$$\text{Where } \mu_1 = \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n x_i}{\sigma^2} \right) / \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right),$$

$$\sigma_1^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)^{-1}$$

Table 1. Comparison of different stochastic bias correction methods

	Bayesian Inference-based Method	GPM-based Method	Copula-based Method
Implementation	Straightforward, effective, less tuning parameters, and is easy for engineering implementation	Difficult to determine the hyper-parameters when solving high-dimensional problems	The implementation process is complex
Uncertainty quantification	Can quantify repeated test data (need enhancement to consider CAE uncertainties)	Can quantify repeated test and CAE data	Can quantify repeated test and CAE data
Assumptions	Normal distribution is assumed for the bias data, prior distribution needs to be acquired or assumed	Gaussian process is assumed for the bias function, bias and CAE are assumed to be independent	No specific distribution needs to be assumed
Passing through the training points?	No	Yes	No
Advantages	Generic, effective, has good predictive capability in most situations.	Shows superiority in capturing the mean of the responses when less training data are available	Good to handle large bias noisy situation, can capture general statistical relationships between model prediction and model bias, and between DVs and model bias.
Disadvantages	Results are largely affected by the prior information/evidences, and available training data	The prediction interval increases dramatically when there is no known design points in the neighborhood, it might have potential negative effects on identification of a better design	the predicted model bias could have a wide confidence interval because it predicts the model bias in a more global sense

Based on the test and CAE data, the bias distribution at each design point $\mathbf{x}$ is obtained as: $\tau(\mathbf{x}) \sim N(\tau_{\mu 1}(\mathbf{x}), \tau_{\sigma^2}(\mathbf{x}))$. It is noted that this method can be also be used for conjugate prior of different distributions and non-conjugate prior as well.

Two RSMs are built for the mean and standard deviation of the prediction bias. After obtaining the CAE simulation result, 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 results of a new design are within the corresponding PIs, 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 optimization or robust design. On the other hand, if the prediction is rejected, additional DOE will be added and repeat the process until good quality prediction is achieved [21].

Based on the above case studies, several observations can be drawn. These discoveries can provide some insights into the specific characteristics of different methods, and help to identify their advantages and disadvantages in different application scenarios.

RELIABILITY-BASED DESIGN OPTIMIZATION UNDER MODEL AND DATA UNCERTAINTIES

Reliability-based design optimization (RBDO) is a popular approach in design optimization under uncertainty, RBDO shows its advantages in obtaining an optimal design that is most reliable under the uncertainties of design variables. The traditional design optimization is mostly deterministic, which often results in a optimal design that is pushed to the limit of design constraints and leaves little or no space for uncertainty, RBDO considers the uncertainty of controllable or uncontrollable variables associated with an existing CAE model. A typical RBDO formulation is:

$$\text{Minimize } E[f_0(X)]$$

$$\text{subject to } \Pr\{f_i(X) \leq 0\} \geq \alpha_i\%, i = 1, \dots, k$$

X is a random vector that follows a known distribution, such as normal distribution, $E[\cdot]$ is the expectation of a function, $f_i(X)$, $i = 1, \dots, k$ is a set of functional assessment of interest, and $\alpha_i\%$, $i=1, \dots, k$, is the reliability level.