

An Enhanced Input Uncertainty Representation Method for Response Surface Models in Automotive Weight Reduction Applications

Bo Liu

Changan Auto & Vehicle Mfg Tech

Junqi Yang, Zhenfei Zhan, and Ling Zheng
Chongqing University

Bo Lu, Ke Wang, Zhentao Zhu, and Zhongcai Qiu
Changan Auto & Vehicle Mfg Tech

ABSTRACT

Vehicle weight reduction has become one of the viable solutions to ever-growing energy and environmental crisis. 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 affects 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. For the purpose of constructing and correcting the bias in RSMs, scheduling Design of Experiments (DOEs) must be conducted properly. This paper develops a method to arrange DOEs in order to build RSMs with high quality, considering the influence of input uncertainty. The proposed method focuses on assigning samples corresponding to the obtained Probability Density Function (PDF) of represented input, and it shows its advantage in the process of model bias correction. A real-world vehicle weight reduction design example is used to illustrate the validity of the proposed method.

CITATION: Liu, B., Yang, J., Zhan, Z., Zheng, L. et al., "An Enhanced Input Uncertainty Representation Method for Response Surface Models in Automotive Weight Reduction Applications," *SAE Int. J. Mater. Manf.* 8(3):2015, doi:10.4271/2015-01-0423.

INTRODUCTION

Lightweight design for vehicle is a well-known strategy to address growing concerns about harmful emissions, improve fuel economy and save energy. Since the body structure possesses about 40% weight of full vehicle, the lightweight design of body structure plays a quite important role in weight reduction of full vehicle. It is a systematical project that involved with crashworthiness, stiffness, modal and NVH (noise, vibration and harshness) performance. In practice, integrating structural optimization directly with expensive FE simulations is generally infeasible since it typically requires thousands or even millions of iterations in automotive design. As a consequence, there is a growing interest in using RSM to approximate the complicated nonlinear behaviors to manage the complexity in analysis and optimization for body weight reduction.

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 design optimization. For the purpose of correcting the bias in RSMs, scheduling Design of Experiments (DOEs) must be conducted properly.

Ideal DOEs for correcting bias of RSM should take the effect of input uncertainty into consideration. For the purpose of studying the statistical characteristics of sparse points in DOEs, it is critical to represent the input under uncertainty in appropriate way. Plenty of researchers contributed a lot in the investigation of representing the input uncertainty. Aven [1] provided an idea to interpret several alternative uncertainty representation methods, aiming merely at the interval uncertainty. Though Sankararaman and Mahadevan [2, 3] have investigated a representation framework for epistemic uncertainty including both interval and sparse point data and applied such an approach to model validation, they conducted the representation process without the consideration of the influence of input uncertainty on simulation output. To tackle such problem, this paper proposed an enhanced input uncertainty representation method, aiming at obtaining response surface models with high fidelity for accurate prediction.

The rest of the paper is organized as follows: The next section introduces the framework of input uncertainty representation. In the section, the methods of modified Bootstrap based sampling, entropy based input uncertainty assessment and parametric uncertainty representation for resampled data are included. In the last section, an inference based method for revising bias of RSM is introduced. An

analytical case is utilized for illustration. And the proposed method is applied to a vehicle weight reduction design case. At last, some conclusions are given in the end.

PROPOSED FRAMEWORK FOR INPUT UNCERTAINTY REPRESENTATION

In this paper, we focus on only one type of epistemic uncertainty, namely sparse samples. Figure 1 illustrates the general steps of the proposed method for representing input uncertainty and its application on RSM correction. In the framework of proposed method, given the sparse points, a modified Bootstrap sampling method is proposed to get adequate samples to reflect the statistical characteristics. The uncertainty of these resampled data is assessed through the criterion of entropy. With the weighting factors deriving from the values of entropy, a group of distributions are integrated into a single one. By conducting resampling based on the integrated distribution, DOEs with the consideration of input uncertainty is generated to correct the bias of RSM. Details of these steps are introduced in the following subsections.

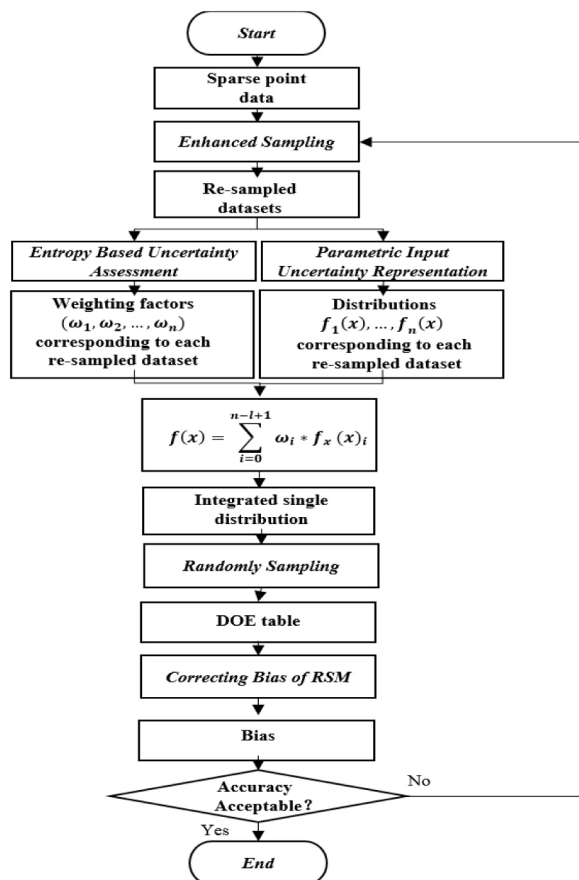


Figure 1. Framework of input uncertainty representation for correcting bias of RSM

Modified Bootstrap Based Resampling Method

In this section, a brief review of bootstrap is provided, based on which a modified version for resampling is introduced in detail. Given a computational model G , its input and corresponding output are indicated by X and Y , respectively. To input data, given the

original samples $X = (x_1, x_2, \dots, x_{n-1}, x_n)$ in size of n , Bootstrap randomly samples from X for b times with replacement. For the purpose of assessing the importance of every elements of the original dataset, standard Bootstrap is modified in some aspects. Here only the condition of single variable is considered for better illustration. In the original Bootstrap, every elements in X are replaced by the random quantity among $(x_1, x_2, \dots, x_{n-1}, x_n)$. After n elements are all replaced, a Bootstrap-based resampled data set, indicated by X_i^* ($i=1$ to b), is generated. This procedure is repeated for b times to get b Bootstrap sets $(X_1^*, X_2^*, \dots, X_b^*)$, which have the same size as X and allow one to compute b replications of $\hat{\theta}$, the parameter that need to be estimated to form a distribution.

$$\hat{\theta}_i = f(x_1, x_2, \dots, x_{n-1}, x_n) \quad (1)$$

where $i=1$ to b . The bootstrap estimate of standard error for $\hat{\theta}$ is given by:

$$\hat{\sigma}_{boot,b} = \sqrt{\sum_{i=1}^b (\hat{\theta}_i - \bar{\hat{\theta}})^2 / (b-1)} \quad (2)$$

However, what we need is to modify the data intentionally for investigating how the modification process effects the output. A novel sampling method is proposed. One main difference between our method and Bootstrap is that the proposed one adds elements intentionally according to a specified schedule while the original one conduct the process randomly. The enhanced resampling method can be illustrated as add- L method explicitly. In the input set $X = (x_1, x_2, \dots, x_{n-1}, x_n)$, a set $L = (x_i, \dots, x_{i+l})$ in the length of l is picked out, where $i=(1$ to $n-l+1)$. The following step is to append L to X for t times. With the same discipline, $n-l+1$ subsamples are generated. The new datasets are denoted by a matrix $X_{a,l}$ in the size of $(n-l+1) \times (n + (t-1) \cdot l)$, whose row vectors are expressed as:

$$\begin{aligned} X_{a,l}(1) &= [L_1 \dots L_1, x_{l+1}, x_{l+2}, \dots, x_{n-1}, x_n] \\ X_{a,l}(2) &= [x_1, x_2, L_2 \dots L_2, \dots, x_{n-1}, x_n] \\ &\vdots \\ X_{a,l}(n-l+1) &= [x_1, x_2, x_3, \dots, L_{n-l+1} \dots L_{n-l+1}] \end{aligned} \quad (3)$$

It can be inferred that if the length of L is 1 and we conduct the process randomly, it becomes the conventional methodology of Bootstrap.

Another major discrepancy is that, after obtaining the resampled datasets $X_{a,l}$, bootstrap-based method infers the statistical features (parameters and hyper-parameters) based on all the subsamples, as shown in Eq.2, while the proposed method assesses the additional uncertainty for each subsample, denoted by the row vectors of $X_{a,l}$. And the process of input uncertainty representation is conducted through integrating $n-l+1$ distributions for each subsample set rather than represent the distribution with the obtained $\hat{\theta}$, which will be illustrated in the following sections.

Entropy Based Input Uncertainty Assessment Method

The additional uncertainty caused by adding new samples to X needs to be assessed. Criteria of entropy is defined as the measurement of uncertainty of random variable. To a discrete random variable X with probability distribution:

$$P(X = x_i) = p_i, \quad i = 1, 2, 3, \dots, n$$

the entropy of X is expressed as:

$$H(X) = -\sum_{i=1}^n p_i \log p_i \quad (4)$$

To every row vectors in $X_{a,l}$, their values of entropy are calculated by:

$$H(X_{a,l}(i)) = -\sum_{m=1}^{n+(t-1) \cdot l} p(i)_m \log p(i)_m$$

where $n+(t-1) \cdot l$ is the length of row vector and $i=1$ to $n-l+1$

Large value of entropy refers for large uncertainty of variable. A ratio ω is introduced as weighting factor for distributions generated with resampled data sets.

$$\omega_i = \frac{H(X_i)}{H(X_1)+H(X_2)+\dots+H(X_{n-l+1})}, \quad i = 1 \text{ to } n-l+1 \quad (5)$$

The weight factors are beneficial to integrate several input distributions into a single one with fidelity, which will be introduced in the following.

Parametric Distribution Based Representation Method

After obtaining the datasets through the proposed sampling method, probability distribution is utilized to represent the uncertainty. Mahadevan and Rebba [4] introduced a methodology to deal with different sources of uncertainty, including interval inputs, distributions and sparse points. In parametric way, it is necessary to make an assumption for the type of distribution (normal, exponential, etc.) of $X_{a,l}(i)$, $i=1$ to $n-l+1$. P denotes the parameters (i.e. mean and standard deviation) of the assumed distribution. The prior probability density functions (PDFs) of the samples in the condition of parameter P is denoted by $f_x(x|P)$. It is critical to estimate P since this prior distribution depends upon the choice of P . The expression of likelihood for P can be derived from the first principles of mathematical statistics. Considering a single point x_i in data set X , as Sankararaman and Mahadevan [2, 3] illustrated, we can get the likelihood of this single observation x_i with the help of a constant ε . Pawitan [9] has defined the notion of likelihood as the probability of x_i given P . Under the assumption of a small enough and arbitrary constant ε , x_i can be expressed as an infinitesimally interval $[x - \varepsilon/2, x + \varepsilon/2]$. So the likelihood of P is

$$L(P) = p(x \in (x - \varepsilon/2, x + \varepsilon/2) | P) = \int_{x-\varepsilon/2}^{x+\varepsilon/2} f_x(x_i | P) dx \quad (6)$$

According to the mean value theorem of integration, it can be transformed to

$$L(P) = p(x \in (x - \varepsilon/2, x + \varepsilon/2) | P) = \varepsilon f_x(x_i | P) \quad (7)$$

For multiple points in the set, the overall likelihood function is expressed as:

$$L(P) \propto \prod_{i=1}^n f_x(x_i | P) \quad (8)$$

The observing points are all independent in this paper if without intentional note. In the theory of maximum likelihood estimation, after obtaining the likelihood function, parameter P can be estimated by maximizing the function in Equation 8.

The distribution of X with parameter P , namely $f_x(x)$ is calculated for every subsamples of $X_{a,l}$. Combining with the weight factors evaluated for each subsample, those distributions can be integrated into a single one as follow:

$$f(x) = \omega_1 * f_x(x)_1 + \dots + \omega_{n-l+1} * f_x(x)_{n-l+1} \quad (9)$$

Where $\omega_1 \dots \omega_{n-l+1}$ are the weighting factors assessed through criteria of entropy. The process of input uncertainty representation provided in the type of distribution is completed.

Bayesian Inference Based Method for Revising Bias

In the procedure of revising bias, Bayesian inference based method is utilized, difference between CAE and test is considered as evidence of prior distribution of bias. For the convenience of formulation, let x represents design variables, Y_t and Y_m stand for FE model result (test) and RSM model result (model) respectively. The relationship between $Y_t(x)$ and $Y_m(x)$ can often be generalized as follows[5, 6, 7]:

$$Y_t(x) = Y_m(x) + \delta(x) + \varepsilon(x) \quad (10)$$

where ε is the random variable representing the test error that may depend on 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 11.

$$Y_{true}(x) = Y_t(x) - \varepsilon(x) = Y_m(x) + \delta(x) \quad (11)$$

Following Equation 11, assume $\tau(x) = \delta(x) + \varepsilon(x)$. 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 $\tau(x) = \delta(x) + \varepsilon(x)$ follows a normal distribution with known variance as:

$$\tau(x) \sim N(\tau_\mu(x), \tau_{\sigma^2}(x)) \quad (12)$$