

A DESIGN-DRIVEN VALIDATION APPROACH USING BAYESIAN PREDICTION MODELS

Wei Chen¹, Ying Xiong
Department of Mechanical Engineering
Northwestern University

Kwok-Leung Tsui, Shuchun Wang
School of Industrial & Systems Engineering
Georgia Institute of Technology

ABSTRACT

In most of the existing work, model validation is viewed as verifying the model accuracy, measured by the agreement between computational and experimental results. Due to the lack of resource, accuracy can only be assessed at very limited test points. However, from the design perspective, a good model should be considered the one that can provide the discrimination (with good resolution) between competing design candidates under uncertainty. In this work, a design-driven validation approach is presented. By combining data from both physical experiments and the computer model, a Bayesian approach is employed to develop a prediction model as the replacement of the original computer model for the purpose of design. Based on the uncertainty quantification with the Bayesian prediction and subsequently that of a design objective, some decision validation metrics are further developed to assess the confidence of using the Bayesian prediction model in making a specific design choice. We demonstrate that the Bayesian approach provides a flexible framework for drawing inferences for predictions in the intended, but maybe untested, design domain. The applicability of the proposed decision validation metrics is examined for designs with either a discrete or continuous set of design alternatives. The approach is demonstrated through an illustrative example of robust engine piston design.

KEYWORDS

Model validation, Bayesian prediction, Uncertainty quantification, Validation metrics, Design confidence, Engineering Design

¹ Corresponding author. Department of Mechanical Engineering, Northwestern University, 2145 Sheridan Road, Tech B224, Evanston, IL 60208-3111, Email: weichen@northwestern.edu, Phone: (847) 491-7019, Fax: (847) 491-3915.

NOMENCLATURE

$Y^e(\mathbf{x})$	physical experimental observation
$\varepsilon(\mathbf{x})$	experimental error
$Y^r(\mathbf{x})$	true response outcome
$Y^m(\mathbf{x})$	outcome of computer model
$\delta(\mathbf{x})$	the bias (or error) of computer model
$f(\mathbf{x})$	design objective function
$Z(\mathbf{x})$	difference between $f(\mathbf{x})$ and $f(\mathbf{x}^*)$
\mathbf{x}	$\mathbf{x} = (x_1, \dots, x_p)^T$, design in a p -dimensional space
D_e	$D_e = \{\mathbf{x}_1, \dots, \mathbf{x}_{n_e}\}$, for physical experiments
D_m	$D_m = \{\mathbf{x}'_1, \dots, \mathbf{x}'_{n_m}\}$, for computer experiments
n_m, n_e	size of D_m / D_e , the number of computer / physical experiments
σ_ε^2	variance parameter of $\varepsilon(\mathbf{x})$
σ_δ^2	variance parameter of the prior Gaussian process $\delta(\mathbf{x})$
ϕ_δ	correlation parameter of the prior Gaussian process $\delta(\mathbf{x})$
τ	ratio of σ_ε^2 to σ_δ^2
$n_{\delta e,m}$	degree of freedom of t distribution
$\mu_{\delta e,m}(\mathbf{x})$	noncentrality parameter of t distribution
$\sigma_{\delta e,m}^2(\mathbf{x})$	scale parameter of t distribution
$M_D(\mathbf{x}^*)$	decision validation metric
H	design tolerance
X^0	indifferentiable region
Ω_d	set of feasible design alternatives
K	number of design candidates or competing designs

1. INTRODUCTION

With rapid increase of computational power, modeling or simulation based design has been increasingly used for designing new engineering systems. However, it remains a challenge on assessing the risks and uncertainties associated with the use of computer models in engineering design. Even though there is growing interest from both government and industries in developing fundamental concepts and terminology for model validation (DoD; Ang et al. 1996, Doebbling, et al. 2002; Oberkampf et al, 2003; Cafeo and Thacker 2004; Gu and Yang, 2003), model validity and model validation are poorly understood in engineering design. Most of the existing model validation work (e.g., Marczyk et al. 1997; Freese, 1960; Reynolds, 1984; Gregoire and Reynolds, 1988;

Hills and Trucano 1999), is rooted in computational science where validation is viewed as verifying the model accuracy, i.e., a measure of the agreement between computational and experimental results. An extensive discussion of validation literature in computational mechanics can be found in Oberkampf and Trucano (2000). In most of the existing work, model validation has been primarily carried out from the perspective of model builders (or analysts) but not from that of designers (model users). Validation metrics are assessed based on very limited test points without considering the predictive capability at untested but potential design space and the various sources of uncertainties. In summary, the existing approaches for validating analysis models are not directly applicable for assessing the confidence of using predictive models in engineering design.

In the engineering design research community, special attentions have been given to how models and information are used in design decision making (McAdams and Dym, 2004). Preliminary efforts have been made on characterizing and assessing the validity of behavior models and their predictions in design (Malak and Paredis, 2004). Hazelrigg (2003) brought up the notion that validation of a predictive model can be accomplished only in the context of a specific decision, and only in the context of subjective input from a decision maker, including preferences. As noted by Hazelrigg (2003), what really matters to designers is whether a model generates design choices whose real outcomes are better than other design choices. To make such an assessment, as illustrated in Fig. 1, in the presence of model uncertainty, it is important to assess the probability P_{ij} of design alternative \mathbf{x}_i to produce an outcome that is preferred to or indifferent to another alternative \mathbf{x}_j , i.e., $P_{ij} = P[f(\mathbf{x}_i) < f(\mathbf{x}_j)]$ assuming the smaller-the-better scenario. It should be pointed out that design alternatives are compared against each other with regard to a specific design objective $f(\mathbf{x})$, which is a function of single or multiple responses $y(\mathbf{x})$ from computer model(s). To quantify the uncertainty of $f(\mathbf{x})$, statistical inference techniques must be developed to first quantify the uncertainty associated with the prediction of the response $y(\mathbf{x})$ based on the results from both models and physical experiments.

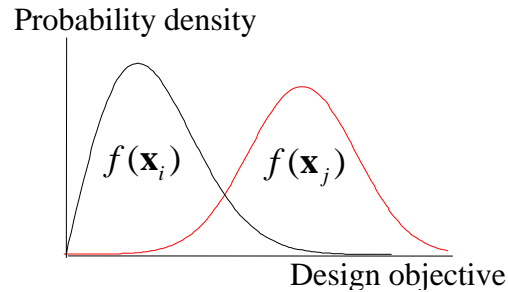


Figure 1. Design resolution

Recent approaches for quantitatively comparing computations and experiments to derive the probability distributions of model predictions can be divided into two categories, namely classical frequentist approach (Oberkampf and Barone, 2004) and Bayesian approach (Kennedy and O’Hagan, 2001; Bayarri et al., 2002; Buslik, 1994; Hanson, 1999). Easterling and Berger (2002) provided an extensive review on classical statistical approaches for model validation and a simple case study. A review of Bayesian approaches can be found in Bayarri et al. (2002). The fundamental difference between the frequentist and the Bayesian approach is that the former draws confidence intervals of prediction based on statistical data analysis, while the latter assumes that the model parameters themselves are random and follow a prior distribution, specified based on model builder/designers’ prior knowledge. The prior distribution will be updated once experimental data is available and becomes posterior distribution. The Bayesian approach is preferred to the classical statistical approach when it is too expensive to obtain a statistically sufficient amount of data, which is often the case in engineering design.

Existing model validation metrics are mostly associated with the measures of model accuracy based on limited tested points. Many of existing approaches cannot provide stochastic measurements with regard to the confidence in using a model. For instance, graphical comparisons through visual inspection of x-y plots, scatter plots and contour plots are often subjective and not sufficient (Oberkampf and Trucano, 2000). Quantitative comparisons (Marczyk et al. 1997) that rely on the measures of correlation coefficient and other weighted and non-weighted norms to quantify the distance between the two “clouds” cannot provide statistical judgment of model validity. Various statistical inference techniques, such as χ^2 (Chi-square) test on residuals between model and

experimental results (Freese, 1960; Reynolds, 1984; Gregoire and Reynolds, 1988) require multiple evaluations of the model and experiments, and many statistical assumptions that are difficult to satisfy.

Although the need for validating models from the perspective of engineering design has been brought up (Malak and Paredis, 2004; Hazelrigg 2003), few have developed quantitative means to define and to assess model validity in such context. In the author's earlier work, an approach was developed to provide stochastic assessment of the validity of a model (Chen et al. 2004; Buranathiti et al. 2004). However, the approach is more useful for rejecting (invalidating) a model rather than accepting (validating) a model. In the recent work of Mahadevan and Rebba (2005), a Bayes network approach is proposed for validating the reliability assessment made by computational models. Validation was treated as a hypothesis testing problem. However, the emphasis was on validating the modeling accuracy at tested design points, but not in the context of a new design.

In this paper, we present a design-driven validation approach (Section 2) to guide the use of predictive models in engineering design. A Bayesian procedure (Section 3) is employed to combine the data from physical experiments and computer models for generating prediction models and quantifying the prediction uncertainty. The Bayesian approach provides a framework for drawing inferences for predictions in the intended but untested design domain. Our approach is generic enough to handle cases where design settings of physical experiments and the computer model may or may not overlap. When limited amount of physical data is available, the approach is capable of taking into account scientific knowledge as well as designer's belief based on past experience in the form of prior distributions of model parameters. With the quantified uncertainty of Bayesian prediction models, we further develop some decision validation metrics (Section 4) to provide confidence measures of using Bayesian prediction in making a specific design choice for a given design objective. The implications of using such metrics are examined and the computational requirements are discussed for cases with either discrete or continuous design alternatives. Our approach is then demonstrated through an illustrative example of robust engine piston design in Section 5. Section 6 is the closure of this paper.

2. PROPOSED DESIGN-DRIVEN VALIDATION FRAMEWORK

In contrast from the traditional viewpoint where *model validation* is viewed as the means to assess the accuracy of a computer model, it is viewed in this work as *a process to improve designer's confidence in making a design choice based on the improved predictive model resulted from the validation process*. As shown in the comparison in Fig. 2, the traditional model validation framework (Fig. 2a) is restricted to validating (accepting) or invalidating (rejecting) a computer model by directly comparing the results from both computer and physical experiments. Once a computer model is “validated”, it will then be passed to designers for design purposes. Due to limited resources for physical experiments, it is unlikely that a model can be validated across the whole design space, and the validity of a “validated” model at limited settings of model inputs can hardly be extended to untested regions.

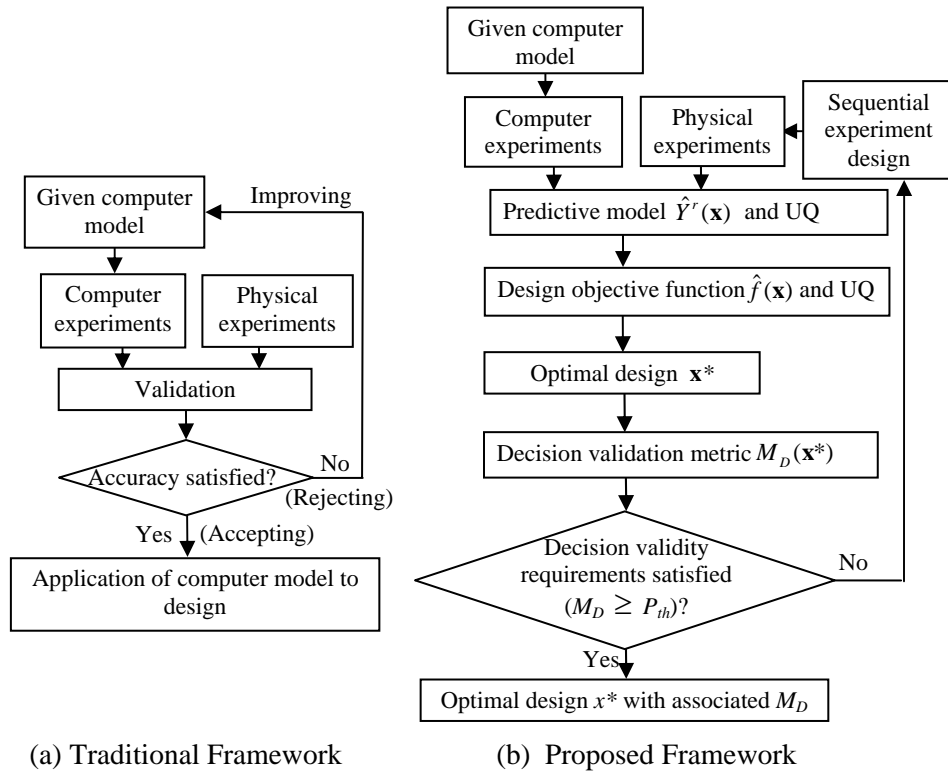


Figure 2. Comparison of Traditional and Proposed Validation Approaches (UQ - Uncertainty Quantification)

A design-driven model validation approach is proposed in this work with the emphasis on enhancing the predictive capability of a computer model for the purpose of

design decision making (Fig. 2b). A ***predictive model***, which is the prediction of a real model response, denoted as predictor $\hat{Y}^r(\mathbf{x})$, and the associated uncertainty quantification (UQ), are first obtained by combining the results from both computer and physical experiments. The predictive model can be viewed as a *corrected* computer model by characterizing the bias between the computer model and the reality. A Bayesian procedure for obtaining the predictive model, called Bayesian prediction model, with uncertainty quantification (UQ) is described in Section 3. To validate the use of the Bayesian prediction model for design, the next step is to construct a ***design objective function*** $f(\mathbf{x})$ based on specific design requirements and preference posed by designers. For example, when uncertainty of design variables/parameters is considered, $f(\mathbf{x})$ could be a typical robust design objective (Chen et al. 1996) as a function of both optimizing the mean μ_y and minimizing the standard deviation σ_y of performance y . Prediction of the design objective value $\hat{f}(\mathbf{x})$ will then be derived and its uncertainty will be quantified by considering model uncertainty in $y(\mathbf{x})$. Given a feasible set of design alternatives, an optimal design (denoted as \mathbf{x}^*) could be identified by optimizing $\hat{f}(\mathbf{x})$.

For a candidate optimal design \mathbf{x}^* , the ***decision validation metric*** (denoted as $M_D(\mathbf{x}^*)$) is used to assess the validity of using the Bayesian prediction model for choosing \mathbf{x}^* as the optimum. Some decision validation metrics proposed in this work are presented in Section 4. For a pre-specified confidence level (or threshold P_{th} , such as 90%), if $M_D \geq P_{th}$, the optimal design \mathbf{x}^* is concluded with the confidence level as high as M_D . If $M_D < P_{th}$, it means that the design objective function $\hat{f}(\mathbf{x})$ can not provide sufficient *resolution* to support the design decision \mathbf{x}^* , and more information needs to be gathered through additional physical experiments to reduce the uncertainty of $\hat{f}(\mathbf{x})$. When the physical experiments are added sequentially, the validation process is repeated until the validity requirement is satisfied. In the following sections, we will provide some details of the key elements of the proposed validation framework, i.e., the Bayesian procedure for model prediction and the evaluation of decision validation metrics. Sequential experiment design is not covered in this paper due to the space limit.

3. THE BAYESIAN PROCEDURE FOR MODEL PREDICTION

Central to the proposed validation framework is the prediction of the amount by which a model output may differ from the true value, which is often complicated by the presence of uncertainties and errors from various sources, such as model (lack of knowledge), parametric, algorithmic, computational, and system variability, as well as testing data that is used to compare with the model prediction. Different ways of classifying uncertainties in model prediction are seen in the literature (Apostolakis 1994; Trucano, 1998; Hazelrigg, 1999; Oberkampf et al., 1999). In this work we assume that the computational error has been satisfactorily resolved or eliminated through “verification” (Oberkampf and Trucano, 2000). Using \mathbf{x} to represent design variables and y stand for model response, the relationship between the experimental observation $Y^e(\mathbf{x})$ and the result generated by a computer model $Y^m(x)$ can often be generalized as follows:

$$Y^e(\mathbf{x}) = Y^m(\mathbf{x}) + \delta(\mathbf{x}) + \varepsilon(\mathbf{x}), \quad (1)$$

where $\varepsilon(\mathbf{x})$ is the random variable representing the experimental error (relating to both experimental setup and measurement) that may depend on \mathbf{x} , and $\delta(\mathbf{x})$ is the error of the model, or called the *prediction bias*, i.e.,

$$\delta(\mathbf{x}) = Y^r(\mathbf{x}) - Y^m(\mathbf{x}), \quad (2)$$

which captures the model inadequacy, where $Y^r(\mathbf{x})$ is the true output.

In the above equations, the prediction bias $\delta(\mathbf{x})$ is more closely related to the assessment of model accuracy, while the prediction of the true model output $Y^r(\mathbf{x})$ is essential to assess the probability that a design alternative will produce an outcome that is preferred to or indifferent to other alternatives. Most research in validating computer models has focused on estimating *prediction bias*, but much less work had been done on characterizing *prediction uncertainty* and *prediction bias* under general situations. It can be noted from Eqn. (2) that estimating the prediction bias $\delta(\mathbf{x})$ is an intermediate step for estimating the true model output $Y^r(\mathbf{x})$. Based on the experimental data, outputs of the computer model, and the experimental error $\varepsilon(\mathbf{x})$, the estimated prediction error, $\hat{\delta}(\mathbf{x})$, and its probability distribution is first obtained. The estimated prediction, $\hat{Y}^r(\mathbf{x})$, can

then be obtained by $\hat{Y}^r(\mathbf{x}) = Y^m(\mathbf{x}) + \hat{\delta}(\mathbf{x})$. $\hat{Y}^r(\mathbf{x})$ is the so-called *Bayesian prediction model* to obtain the candidate optimal design in the subsequent procedure, while uncertainty quantification (UQ) of $\hat{Y}^r(\mathbf{x})$ plays a critical role in validation.

In this work, a Bayesian approach is used to provide uncertainty quantification of both $\hat{\delta}(\mathbf{x})$ and $\hat{Y}^r(\mathbf{x})$. Theoretical details of the Bayesian approach can be found in Wang et al. (2006), while related references for Bayesian analysis could be found in Qian and Wu (2005) and Reese et al. (2004). Bayesian inferences are preferred as they require fewer assumptions and are more flexible for engineering applications where it may be too expensive to obtain experimental data. In addition, Bayesian methods may be preferable as additional information or designer's belief can be incorporated through prior distributions. Below, some details of each step in the Bayesian procedure are provided. It should be pointed out that the mathematical framework considered in this work is similar to the one in Kennedy and O'Hagan (2001), however, our work focuses on characterizing the behavior of the prediction bias $\delta(\mathbf{x})$ while the emphasis of Kennedy and O'Hagan's work is on the *calibration* of computer models based on physical observations, with the term $Y^m(\mathbf{x})$ in Eqn. (1) replaced by $\rho Y^m(\mathbf{x}, \Theta)$, where ρ is an unknown regression parameter, and Θ is the vector of calibration parameters.

(1) Collect both physical and computer model data

Both physical observations and computer model outputs are essential to model validation. Let $\mathbf{x} = (x_1, \dots, x_p)^T$ be a point in a p-dimensional design variable space. Let $D_e = \{\mathbf{x}_1, \dots, \mathbf{x}_{n_e}\}$ and $D_m = \{\mathbf{x}'_1, \dots, \mathbf{x}'_{n_m}\}$ be the design settings for physical experiments and computer experiments, respectively; $\mathbf{y}^e = (y^e(\mathbf{x}_1), \dots, y^e(\mathbf{x}_{n_e}))^T$ and $\mathbf{y}^m = (y^m(\mathbf{x}'_1), \dots, y^m(\mathbf{x}'_{n_m}))^T$ be the corresponding physical experimental observations and deterministic computer model outputs, respectively. Note that D_e and D_m may or may not overlap. Physical observations are desired to be as many as possible and close to the intended design region. Compared to physical observations, computer model outputs are less costly and should be simulated at design settings where the physical observations are available and close, if not within, the intended design regions.

When D_e and D_m do not overlap and computer simulations are expensive and time-consuming, a metamodel ($\hat{Y}^m(\mathbf{x})$) that interpolates the computer model data may be used to replace $Y^m(\mathbf{x})$. One approach to approximating $Y^m(\cdot)$ is to fit a Gaussian process model based on the available computer experiments (Santner et al., 2003).

(2) Determine priors of Gaussian process parameters for prediction bias

One advantage of the Bayesian approach is its ability to take into account scientific knowledge and past information in the form of prior distributions for model parameters. From Eqn (1), the prediction bias $\delta(\mathbf{x})$ could be formulated as $\delta(\mathbf{x}) = Y^e(\mathbf{x}) - Y^m(\mathbf{x}) - \varepsilon(\mathbf{x})$. In this paper $\delta(\mathbf{x})$ is treated as a Gaussian process, with the process parameters denoted as $\theta = (\sigma_\delta^2, \beta_\delta, \phi_\delta, \sigma_\varepsilon^2)$, which respectively represent the variance parameter (σ_δ^2), the location parameter (β_δ), the correlation parameter (ϕ_δ), and the variance parameter related to the experiment error $\varepsilon(\mathbf{x})$ (σ_ε^2). The following forms of priors for the variance parameter σ_δ^2 and location parameter β_δ are adopted (similar treatments could be found in Reese et al., 2004, Qian and Wu, 2005):

$$\sigma_\delta^2 \sim IG(\alpha_\delta, \gamma_\delta), \quad \beta_\delta | \sigma_\delta^2 \sim N(\mathbf{b}_\delta, \sigma_\delta^2 \mathbf{V}_\delta),$$

where $IG(\alpha, \gamma)$ denotes the inverse gamma distribution. As will be detailed in the following description of step (3), to simplify the Bayesian analysis, no priors are specified for ϕ_δ and σ_ε^2 which are instead treated as fixed and estimated directly from data.

(3) Compute the posterior of prediction bias

Based on the Bayes Theory, the posterior of the prediction bias $\delta(\mathbf{x})$ given physical observations \mathbf{y}^e and computer outputs \mathbf{y}^m can be obtained by integrating out θ through the following equation

$$p(\delta(\mathbf{x}) | \mathbf{y}^e, \mathbf{y}^m) = \int_{\theta} p(\delta(\mathbf{x}) | \mathbf{y}^e, \mathbf{y}^m, \theta) p(\theta | \mathbf{y}^e, \mathbf{y}^m) d\theta. \quad (3)$$

The density function $p(\delta(\mathbf{x}) | \mathbf{y}^e, \mathbf{y}^m, \theta)$ can be easily computed based on the data from both computer model and physical experiments. With some tedious mathematical derivations (refer to Appendix and Wang et al., 2006 for proof), it could be shown that the posterior of $\delta(\cdot)$ follows a t-distribution:

$$\delta(\mathbf{x}) | \mathbf{y}^e, \mathbf{y}^m, \phi_\delta, \tau \sim T(n_{\delta|e,m}, \mu_{\delta|e,m}(\mathbf{x}), \sigma_{\delta|e,m}^2(\mathbf{x})), \quad (4)$$

with the following degree of freedom, noncentrality, and scale parameters:

$$n_{\delta|e,m} = n_e + 2\alpha_\delta, \quad (5)$$

$$\mu_{\delta|e,m}(\mathbf{x}) = \mathbf{f}_\delta^T(\mathbf{x}) \mathbf{A}_\delta \mathbf{v}_\delta + \mathbf{r}_\delta^T(\mathbf{x}) (\mathbf{R}_\delta + \tau \mathbf{I}_{n_e})^{-1} (\mathbf{y}^e - \mathbf{y}_{n_e}^m - \mathbf{F}_\delta \mathbf{A}_\delta \mathbf{v}_\delta), \quad (6)$$

$$\sigma_{\delta|e,m}^2(\mathbf{x}) = \frac{Q_\delta^2}{n_{\delta|e,m}} \cdot \left(1 - \begin{bmatrix} \mathbf{f}_\delta(\mathbf{x}) \\ \mathbf{r}_\delta(\mathbf{x}) \end{bmatrix}^T \begin{bmatrix} -\mathbf{V}_\delta^{-1} & \mathbf{F}_\delta^T \\ \mathbf{F}_\delta & \mathbf{R}_\delta + \tau \mathbf{I}_{n_e} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{f}_\delta(\mathbf{x}) \\ \mathbf{r}_\delta(\mathbf{x}) \end{bmatrix} \right), \quad (7)$$

where

$$Q_\delta^2 = 2\gamma_\delta + (\mathbf{y}^e - \mathbf{y}_{n_e}^m)^T (\mathbf{R}_\delta + \tau \mathbf{I}_{n_e})^{-1} (\mathbf{y}^e - \mathbf{y}_{n_e}^m) + \mathbf{b}_\delta \mathbf{V}_\delta^{-1} \mathbf{b}_\delta - \mathbf{v}_\delta^T \mathbf{A}_\delta \mathbf{v}_\delta, \quad (8)$$

$$\mathbf{A}_\delta^{-1} = \mathbf{F}_\delta^T (\mathbf{R}_\delta + \tau \mathbf{I}_{n_e})^{-1} \mathbf{F}_\delta + \mathbf{V}_\delta^{-1}, \quad (9)$$

$$\mathbf{v}_\delta = \mathbf{F}_\delta^T (\mathbf{R}_\delta + \tau \mathbf{I}_{n_e})^{-1} (\mathbf{y}^e - \mathbf{y}_{n_e}^m) + \mathbf{V}_\delta^{-1} \mathbf{b}_\delta. \quad (10)$$

In the above equations, $\mathbf{F}_\delta = (\mathbf{f}_\delta(\mathbf{x}_1), \dots, \mathbf{f}_\delta(\mathbf{x}_{n_e}))^T$ is the $n_e \times q_\delta$ design matrix, \mathbf{R}_δ is the $n_e \times n_e$ correlation (parameterized by ϕ_δ) matrix of δ_{n_e} , and $\mathbf{r}_m(\mathbf{x})$ is the correlation (parameterized by ϕ_δ) between $\delta(\mathbf{x})$ and δ_{n_e} . Here, $\delta_{n_e} = \mathbf{y}^e - \mathbf{y}_{n_e}^m$ could be viewed as the ‘observations’ at setting D_e , for the Gaussian process $\delta(\mathbf{x})$, and $\mathbf{y}_{n_e}^m$ is the computer model output $Y^m(\cdot)$ (or the metamodel $\hat{Y}^m(\cdot)$ in the case that $Y^m(\cdot)$ is expensive to compute) at D_e . It naturally follows that δ_{n_e} is essentially the observed bias between physical experiments and the computer model (or metamodel) outputs.

We denote ϕ_δ as the correlation parameter underlying \mathbf{R}_δ and \mathbf{r}_δ^T ; τ as the ratio of σ_ε^2 to σ_δ^2 , i.e., $\tau = \sigma_\varepsilon^2 / \sigma_\delta^2$, where σ_δ^2 denotes the process variance of $\delta(\mathbf{x})$ while σ_ε^2 denotes the variance of $\varepsilon(\mathbf{x})$. Unlike $\delta(\mathbf{x})$ which is assumed to be the Gaussian process with spatial correlation structure, $\varepsilon(\mathbf{x})$ follows identical independent normal distribution at different design sites \mathbf{x} . To get the marginal posterior of $\delta(\mathbf{x})$, ϕ_δ and τ also need to be integrated out, which is computationally prohibitive. Alternatively, ϕ_δ and τ can be treated as their true values and estimated with methods such as the Cross Validation (CV) (Hastie et al., 2000), Maximum Likelihood Estimates (MLE) (Hastie et al., 2000),

Markov Chain Monte Carlo (MCMC) (Geyer, 1992), and Minimum Mean Squared Error Estimates (MMSE) (Hastie et al., 2000).

(4) Compute the prediction of the true behavior to obtain Bayesian prediction model

The true behavior $Y^r(\mathbf{x})$ is predicted using the following equations on the estimations of the mean and variance,

$$\hat{Y}^r(\mathbf{x}) = Y^m(\mathbf{x}) + \hat{\delta}(\mathbf{x}), \quad (11)$$

$$\text{Var}[Y^r(\mathbf{x})] = \text{Var}[\delta(\mathbf{x})] = \sigma_{\delta|e,m}^2(\mathbf{x}). \quad (12)$$

The covariance between $Y^r(\mathbf{x}_i)$ and $Y^r(\mathbf{x}_j)$ is given by:

$$\begin{aligned} \text{Cov}[Y^r(\mathbf{x}_i), Y^r(\mathbf{x}_j)] &= \text{Cov}[Y^m(\mathbf{x}_i) + \delta(\mathbf{x}_i), Y^m(\mathbf{x}_j) + \delta(\mathbf{x}_j)] \\ &= \text{Cov}[Y^m(\mathbf{x}_i), Y^m(\mathbf{x}_j)] + \text{Cov}[\delta(\mathbf{x}_i), \delta(\mathbf{x}_j)] = 0 + \sigma_{\delta|e,m}^2(\mathbf{x}_i, \mathbf{x}_j) = \sigma_{\delta|e,m}^2(\mathbf{x}_i, \mathbf{x}_j), \end{aligned} \quad (13)$$

where

$$\sigma_{\delta|e,m}^2(\mathbf{x}_i, \mathbf{x}_j) = \frac{Q_\delta^2}{n_{\delta|e,m}} \cdot (R_\delta(\mathbf{x}_i, \mathbf{x}_j) - \begin{bmatrix} \mathbf{f}_\delta(\mathbf{x}_i) \\ \mathbf{r}_\delta(\mathbf{x}_i) \end{bmatrix}^T \begin{bmatrix} -\mathbf{V}_\delta^{-1} & \mathbf{F}_\delta^T \\ \mathbf{F}_\delta & \mathbf{R}_\delta + \tau \mathbf{I}_{n_e} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{f}_\delta(\mathbf{x}_j) \\ \mathbf{r}_\delta(\mathbf{x}_j) \end{bmatrix}). \quad (14)$$

When $\mathbf{x}_i = \mathbf{x}_j = \mathbf{x}$, Eqn. (14) reduces to Eqn. (7) and Eqn. (13) reduces to Eqn. (12). The predictor of the true behavior $\hat{Y}^r(\mathbf{x})$, along with its uncertainty quantification, are referred to as the *Bayesian prediction model* in this work.

In the following section, we will present some design decision validation metrics that utilize the information of the predicted objective function $\hat{f}(\mathbf{x})$ at multiple design sites to select the best design candidate under model uncertainty and determine the confidence associated with the design decision.

4. SOME DECISION VALIDATION METRICS

Unlike the existing validation metrics that assess the predictive capability (accuracy) of a model, the *decision validation metrics* M_D are proposed and examined in this work to provide a probabilistic measure of whether a candidate optimal design is better than other design choices with respect to a particular design objective. Such metrics are desired to provide the confidence associated with a design decision with consideration of model uncertainty and to guide validation activities. If large uncertainty exists in model

response $Y^r(\mathbf{x})$, subsequently in the design objective $f(\mathbf{x})$, the achieved M_D may be too low to meet the design validity requirements, forcing designers to add new experiments to reduce model uncertainty or to lower the validity requirement.

Assuming a smaller design objective value is preferred, three types of decision validation metrics M_D , namely, M_D^M , M_D^A , and M_D^W , are presented as follows,

(1) The Multiplicative Metric:

$$M_D^M(\mathbf{x}^*) = \left\{ \prod_{\mathbf{x}_i \in \Omega_d, \mathbf{x}_i \notin X^0} P\{f(\mathbf{x}^*) < f(\mathbf{x}_i)\} \right\}^{1/K} = \left\{ \prod_{\mathbf{x}_i \in \Omega_d, \mathbf{x}_i \notin X^0} P\{Z(\mathbf{x}_i) > 0\} \right\}^{1/K} \quad (15)$$

(2) The Average (Additive) Metric:

$$M_D^A(\mathbf{x}^*) = \frac{1}{K} \sum_{\mathbf{x}_i \in \Omega_d, \mathbf{x}_i \notin X^0} P\{f(\mathbf{x}^*) < f(\mathbf{x}_i)\} = \frac{1}{K} \sum_{\mathbf{x}_i \in \Omega_d, \mathbf{x}_i \notin X^0} P\{Z(\mathbf{x}_i) > 0\} \quad (16)$$

(3) The Worst-Case Metric:

$$M_D^W(\mathbf{x}^*) = \min_{\mathbf{x}_i \in \Omega_d, \mathbf{x}_i \notin X^0} P\{f(\mathbf{x}^*) < f(\mathbf{x}_i)\} = \min_{\mathbf{x}_i \in \Omega_d, \mathbf{x}_i \notin X^0} P\{Z(\mathbf{x}_i) > 0\} \quad (17)$$

In Eqns. (15-17), the decision metrics are evaluated for a candidate optimal design \mathbf{x}^* , which may be obtained by optimizing the prediction (predictor or mean value) of objective function $\hat{f}(\mathbf{x})$. The probability $P\{f(\mathbf{x}^*) < f(\mathbf{x}_i)\}$ is evaluated for comparing \mathbf{x}^* against *individually* each other design alternatives, where \mathbf{x}_i ($i=1,2,\dots,K$) belongs to the set of feasible design alternatives Ω_d , excluding those in the indiffereniable region X^0 . A multiplicative, average, or worst-case evaluation will then be used for determining the overall confidence of choosing \mathbf{x}^* as the optimal design. For simplification, a random variable $Z(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{x}^*)$ (see Eqns. (15-17)) is used to represent the difference between $f(\mathbf{x})$ and $f(\mathbf{x}^*)$. The set of feasible design alternatives Ω_d could be either discrete or continuous. When Ω_d is continuous, competing designs \mathbf{x}_i will be selected across over Ω_d .

The concept of indiffereniable region X^0 is introduced to consider the fact that, with the consideration of model uncertainty, distinguishing designs with identical mean values might not be possible. This is especially true in a continuous design space, where designs in a small neighborhood can be considered as equally good. Specifically, we define the

indifferentiable region(s) X^0 as the region(s) (could be either single or disjoint) within which the design points are claimed indifferent to \mathbf{x}^* with a confidence level c (e.g. $c=95\%$), for a given tolerance H :

$$X^0 \triangleq \left\{ \mathbf{x} \mid P[|Z(\mathbf{x})| < H] > c \right\}. \quad (18)$$

Eqn. (18) means that when there is a high confidence that the difference (either positive or negative) between two designs becomes too small (smaller than the tolerance H), the two designs become indifferentiable.

The proposed decision validation metrics in Eqns. (15) and (16) provide averaged measures of the probability that the real outcome of \mathbf{x}^* is better than other design choices, representing the confidence of using a Bayesian prediction model to select \mathbf{x}^* as the optimal design. If $M_D=1$, it indicates that a designer should have full confidence of taking \mathbf{x}^* as the optimal design. The metric $M_D^W(\mathbf{x}^*)$ in Eqn. (17) stands for the worst-case of P to be used instead of the average. Therefore, it only concerns the most competitive design (2^{nd} best design \mathbf{x}_i) to \mathbf{x}^* , outside the indifferentiable region X^0 . With discrete design alternatives, the computations of all three metrics require the equal amount of effort. However, in the scenario with *continuous* design alternatives, the worst-case metric is the most straightforward and easiest to compute. It can be evaluated through optimization by treating the search of the minimum probability $P\{f(\mathbf{x}^*) < f(\mathbf{x}_i)\}$ (finding the 2^{nd} best, or most competing, design \mathbf{x}_i) as an objective and considering the boundary of the indifferentiable region as constraint. Note that the indifferentiable region can be multiple or disjoint in some cases.

By considering the joint distribution of $f(\mathbf{x})$ and $f(\mathbf{x}^*)$, the distribution of $Z(\mathbf{x})$ is obtained with the mean

$$\mu_z(\mathbf{x}) = E[f(\mathbf{x})] - E[f(\mathbf{x}^*)] \quad (19)$$

and the variance

$$\sigma_z^2(\mathbf{x}) = \text{Var}[f(\mathbf{x})] + \text{Var}[f(\mathbf{x}^*)] - 2\text{Cov}[f(\mathbf{x}), f(\mathbf{x}^*)]. \quad (20)$$

The covariance component in Eqn. (20) should not be ignored, especially when \mathbf{x} and \mathbf{x}^* are close to each other. When $\mathbf{x}=\mathbf{x}^*$, it follows that $\sigma_z^2(\mathbf{x})$ reduces to zero.

Compared to the Probability of Correct Selection (PCS) measure defined by Chen et al. 2000 in the work of ordinal optimization, the metric M_D used in this work shares the similar idea of assessing probabilistically whether one design option is better than another. However, the contexts of these two research works are different. Chen et al.'s work (2000) focused on the ordinal optimization for discrete-event systems; only the discrete design options are considered. In our work M_D is proposed to address the design confidence of choosing one design option over the others in cases with both continuous and discrete design spaces. In addition, the indiffereniable region X^0 used in the definition of M_D is a new concept introduced in our work.

5. EXAMPLE: ENGINE PISTON DESIGN

We use the vehicle engine piston design case study previously analyzed in Jin et al. (2005) as an illustrative example to demonstrate our approach. The Noise, Vibration and Harshness (NVH) characteristic of the vehicle engine is one of the critical elements of customer dissatisfaction. The design goal is to optimize the geometry of the engine piston to obtain the minimal piston slap noise (measured in dB). A robust design scenario is considered by treating the design variable also as a random variable considering the manufacturing variation. Technical background of the piston design problem can be found in Hoffman et al. (2003). To graphically illustrate the results and better explain the concepts of the proposed method, only one design variable, the skirt profile (SP), is considered. Nevertheless, the same approach can be applied to high-dimensional problems.

Skirt profile is represented by the unitless characteristic ratio of the shape of an engine piston, ranging continuously from 1 to 3. Previous results show that the skirt profile (SP) strongly affects the design performance, slap noise. Piston slap noise is the engine noise resulting from piston secondary motion, which can be simulated using ADAMS/Flex, a finite element based multi-body dynamics code. Nine (9) computer experiments are conducted using the finite element model, while nine (9) hypothetical physical experiments are considered. It should be pointed out that nine computer experiments are sufficient to capture the behavior of the computer model for this one-dimensional case, although normally the amount of computer experiments are expected to

be more than that of the physical tests. All data are plotted in Fig. 3, with details included in Tables A.1 and A.2. Note the design variable $\mathbf{x} = \text{SP}$ has been normalized to the unit interval $[0,1]$.

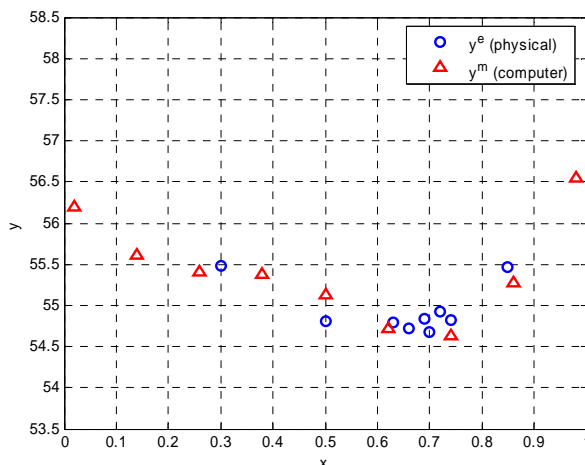


Figure 3. Physical and computer experiment data (circles: physical experiments; triangles: computer experiments)

5.1 Bayesian Prediction Model and Uncertainty Quantification

Based on the available data, the Bayesian approach described in Section 3 is implemented. For the purpose of comparison, the Bayesian prediction models are produced in two stages. *In the first stage, we only use the first six(6) points out of the nine(9) physical experiment points in Table A.1.* The remaining 3 points are added in the second stage to demonstrate the impact of adding more experiment data in a sequential process.

Prediction of computer model $\hat{Y}^m(\mathbf{x})$

From the data shown in Tables A.1 and A.2, it is found that there is no overlap between D_e and D_m , indicating that the settings of the design variable (x) for computer outputs are different from those for physical experiments. We first calculate the prediction of $\hat{Y}^m(\mathbf{x})$ by a normal Gaussian process model (e.g., Kriging). From Fig. 4, it is noted that $\hat{Y}^m(\mathbf{x})$ passes through all six computer experiment points. Because for this single-dimension case, six points are sufficient to capture the behavior of the computer model, the interpolation uncertainty (Apley et al., 2006) of using the normal Gaussian process model to replace the computer model can be neglected.

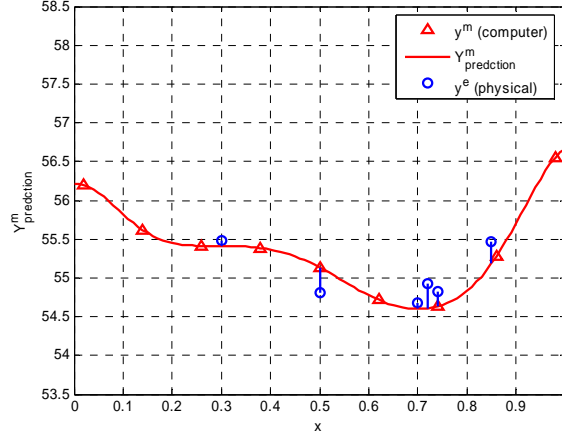


Figure 4. Prediction of $\hat{Y}^m(\mathbf{x})$, interpolating all nine computer experiments

Estimation of prediction bias $\hat{\delta}(\mathbf{x})$ and uncertainty quantification

From Eqn. (4), the prediction bias $\hat{\delta}(\mathbf{x})$ and the associated uncertainty is characterized by the posterior of $\delta(\mathbf{x})$, given ϕ_δ and τ . Ten-fold cross validation is used to determine the optimal values of ϕ_δ and τ in the similar way as in estimating ϕ_m . The results show the optimal setting at $\tau=2$, $\phi_\delta=22$. Since there is little a priori knowledge about σ_δ^2 and β_δ , vague priors are used for this example:

$$\sigma_\delta^2 \sim IG(2,1), \quad \beta_\delta | \sigma_\delta^2 \sim N(0_{q_\delta}, \sigma_\delta^2 \mathbf{I}_{q_\delta}),$$

where 0_k is a $k \times 1$ vector of zeros, and \mathbf{I}_k is a $k \times k$ identity matrix. Fig. 5 displays the predictor of $\hat{\delta}(\mathbf{x})$ and the 95% confidence interval. Note the experimental points illustrated in Fig. 5 represent the deviation between the physical experiments $y^e(\mathbf{x})$ and the model predictions $\hat{Y}^m(\mathbf{x})$ (the magnitude of the vertical line segments shown in Fig. 4). It is noted that $\delta(\mathbf{x})$ has a relatively smaller variance (uncertainty) in the range of $\mathbf{x} \in [0.7, 0.75]$, due to the fact that more physical experiments are available in that region.

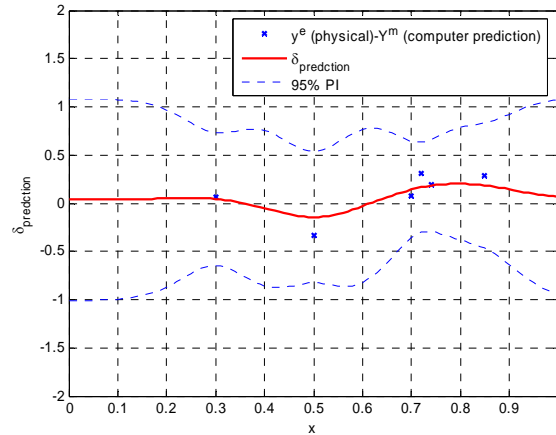


Figure 5. Prediction of $\hat{\delta}(\mathbf{x})$ (dB) and 95% confidence interval

Estimation of $\hat{Y}^r(\mathbf{x})$ and uncertainty quantification

Having obtained the predictions of $Y^m(\mathbf{x})$ and posteriors of $\delta(\mathbf{x})$, the Bayesian prediction model, which is the prediction of the true model response, denoted as $\hat{Y}^r(\mathbf{x})$, is simply the addition of $\hat{Y}^m(\mathbf{x})$ and $\hat{\delta}(\mathbf{x})$. In this case, since the interpolation uncertainty of using $\hat{Y}^m(\mathbf{x})$ to replace the computer model is ignored, the uncertainty of $\hat{Y}^r(\mathbf{x})$ is contributed by the posterior of $\delta(\mathbf{x})$. The predictor $\hat{Y}^r(\mathbf{x})$ and 95% confidence interval are illustrated in Fig. 6.

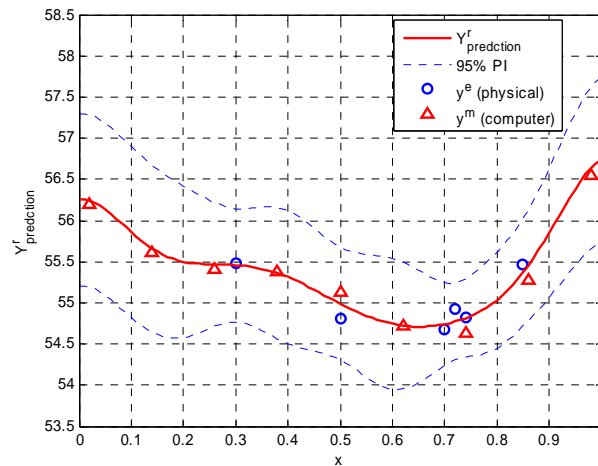


Figure 6. Prediction of $\hat{Y}^r(\mathbf{x})$ (dB) and 95% confidence interval

Prediction of $\hat{f}(\mathbf{x})$ and uncertainty quantification

In this work we consider a typical robust design objective, $f(\mathbf{x}) = \mu_y + k \cdot \sigma_y$, where μ_y and σ_y are the mean and standard deviation of y (piston slap noise), and the weighting factor k is chosen as 3. Note the unit of $f(\mathbf{x})$ is dB, the same with y . The robust design objective is utilized to reduce the impact of the uncertainty associated with the randomness of \mathbf{x} , which follows a normal distribution $\mathbf{x} \sim N(\mu_x, 0.05)$. Since the uncertainty of $Y^r(\mathbf{x})$ is reducible when more experiment data are added, it is essentially the uncertainty in design objective function $f(\mathbf{x})$, due to the model uncertainty, that influences the confidence in making a design decision.

Prediction of $\hat{f}(\mathbf{x})$ and the associated uncertainty quantification could be computationally challenging, especially for high-dimensional problems. Approximation of the mean and variance of $f(\mathbf{x})$ using analytical derivations was discussed by Apley et al. (2006). But, the covariance of any two processes $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$ was not addressed by their method. Due to the low dimension of this case study, Monte Carlo simulation approach is used. Based on the mean, variance and covariance of $Y^r(\mathbf{x})$ given in Eqns. (11)~(13), one can simulate a large amount (e.g. 100) of realizations of the random process $Y^r(\mathbf{x})$. For simplicity, only three of such realizations are selected and shown in Fig. 7. Each single realization of $Y^r(\mathbf{x})$ determines the corresponding realization of $f(\mathbf{x})$ subject to the randomness of \mathbf{x} . As a result, the prediction of $\hat{f}(\mathbf{x})$ and its uncertainty could be quantified as shown in the bold lines in Fig. 7.

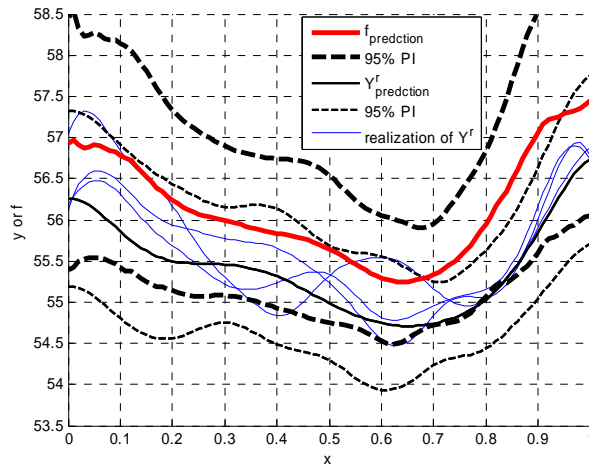


Figure 7. Prediction of $\hat{f}(\mathbf{x})$ (dB) and 95% confidence interval
(6 physical experiments)

5.2 Application of Decision Validation Metrics

In this section, we apply the decision validation metrics M_D proposed in Section 4 to the robust engine piston design. Two design scenarios, i.e., discrete design alternatives and continuous design space, are considered separately.

Scenario 1 - Discrete Design Alternatives

Suppose five design candidates have been identified as $\mathbf{x}_i = \{0.2, 0.4, 0.5, 0.65, 0.7\}$. By minimizing the mean of $f(\mathbf{x}_i)$, \mathbf{x}_4 is chosen as the candidate optimal design, i.e., $\mathbf{x}^* = 0.65$. Figs. 8 and 9 show the mean and 95% prediction interval of $f(\mathbf{x}_i)$ and $Z(\mathbf{x}_i) = f(\mathbf{x}_i) - f(\mathbf{x}^*)$, respectively, at five candidate points. Note both the mean and variance of $Z(\mathbf{x}_i)$ reduces to zero at $\mathbf{x}_i = \mathbf{x}^* = \mathbf{x}_4$, because there is no uncertainty in comparing the candidate optimal design \mathbf{x}^* with itself. Note that $\hat{Z}(\mathbf{x}_i)(i=1,2,3,5) > \hat{Z}(\mathbf{x}_4) = 0$, i.e., the mean of $Z(\mathbf{x}_i)$ at other designs other than \mathbf{x}_4 are all positive. However, due to the uncertainty of $Z(\mathbf{x}_i)(i=1,2,3,5)$, its value may become negative, which indicates that it is likely that \mathbf{x}_i may be better than \mathbf{x}_4 .

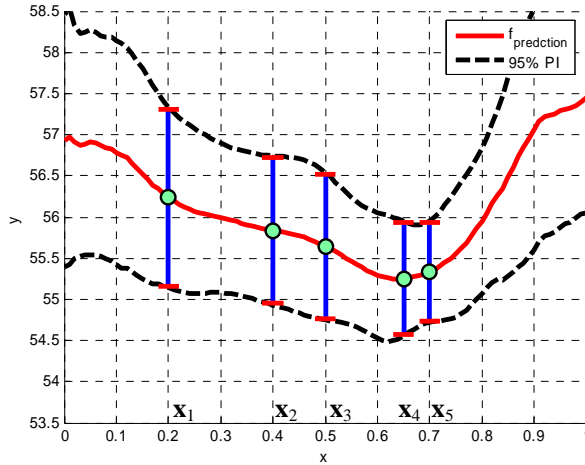


Figure 8. Mean and 95% confidence interval of $f(\mathbf{x}_i)$ (dB) at five design candidates
(6 physical experiments)

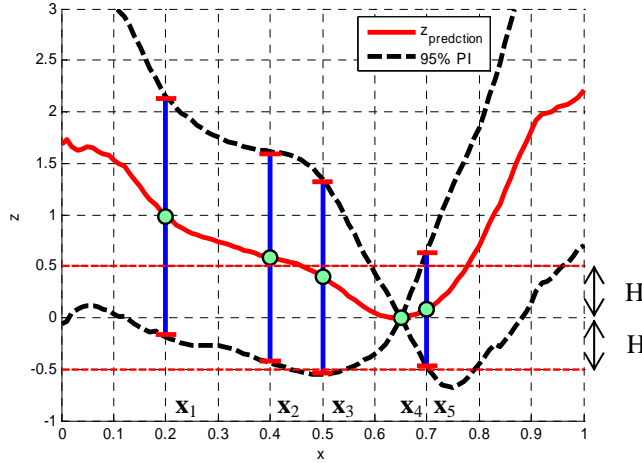


Figure 9. Mean and 95% confidence interval of $Z(\mathbf{x}_i)$ (dB) at five design candidates (6 physical experiments)

Table 1 provides some details in calculating the three different forms of decision validation metrics M_D using the Bayesian prediction model obtained from 6 physical experiments. The tolerance and confidence specified for the indiffereniable region X^0 in this example are taken as $H=0.5$ (dB) and $c=95\%$. The tolerance zone formed by the positive and negative H lines are marked in Fig. 9. By checking the probability value $P[|Z(\mathbf{x}_i)| < H]$, it is found that no point among the four design candidates (i.e., \mathbf{x}_1 , \mathbf{x}_2 , \mathbf{x}_3 , and \mathbf{x}_5) should be claimed indiffereniable to \mathbf{x}^* . Therefore \mathbf{x}_i is set at all these four design candidates in calculating the decision validation metrics M_D . The three types of metrics M_D proposed in Eqns. (15) to (17) are calculated as 0.8017, 0.8118, and 0.6204, respectively. When using the ‘worst-case’ metric, \mathbf{x}_5 is identified as the worst-case point or the most competing (2nd best) design to \mathbf{x}^* . The results of metrics M_D indicate the confidence of using the Bayesian prediction model based on 6 physical experiments for choosing \mathbf{x}_4 as the optimal design among five given design alternatives.

Table 1. Calculation of the decision validation metrics (H=0.5, c=95%, 6 physical experiments)

\mathbf{x}_i	\mathbf{x}_1	\mathbf{x}_2	\mathbf{x}_3	\mathbf{x}_4 (\mathbf{x}^*)	\mathbf{x}_5
$P[Z(\mathbf{x}_i) < H]$	0.1968	0.4152	0.5568		0.9114
$\mathbf{x}_i \in X^0?$	No	No	No		No
$P[Z(\mathbf{x}_i) > 0]$	0.9544	0.8742	0.7981		0.6204
M_D^M (multip)	0.8017				

M_D^A (additive)	0.8118
M_D^W (worst-case)	0.6204 (worst-case point: \mathbf{x}_5)

Table 2 shows the results of decision validation metrics M_D with 6+3 physical experiments. With the reduced uncertainty in $Z(\mathbf{x}_i)$ by the inclusion of 3 additional experiments, it is found $P[|Z(\mathbf{x}_5)| < H] = 0.9799$, indicating that \mathbf{x}_5 should be claimed indifferentiable to $\mathbf{x}^*=\mathbf{x}_4$, and excluded for calculating M_D . The three M_D quantities are now changed to 0.9080, 0.6619, and 0.7826, respectively, all higher than the results based on 6 experiments (see Table 1). The improved confidence of claiming \mathbf{x}_4 as the optimal design resulted from the improved resolution of $f(\mathbf{x})$. Because \mathbf{x}_5 is considered as indifferentiable to $\mathbf{x}^*=\mathbf{x}_4$, \mathbf{x}_3 becomes the most competing design to \mathbf{x}^* under the worst-case metric.

Table 2. Calculation of the decision validation metrics
($H=0.5$, $c=95\%$, 6+3 physical experiments)

\mathbf{x}_i	\mathbf{x}_1	\mathbf{x}_2	\mathbf{x}_3	\mathbf{x}_4 (\mathbf{x}^*)	\mathbf{x}_5
$P[Z(\mathbf{x}_i) < H]$	0.1910	0.3790	0.6111		0.9799
$\mathbf{x}_i \in X^0?$	No	No	No		Yes
$P[Z(\mathbf{x}_i) > 0]$	0.9616	0.9034	0.7826		
M_D^M (multip)	0.9080				
M_D^A (additive)	0.6619				
M_D^W (worst-case)	0.7826 (worst-case point: \mathbf{x}_3)				

Scenario 2 - Continuous Design Alternatives

Calculating M_D in a continuous design space is more challenging than in a discrete design space. By definition, the first two types of decision validation metrics essentially take average of \mathbf{x}^* versus \mathbf{x}_i chosen over the whole feasible design space Ω_d excluding the indifferentiable region X^0 . In high dimensional situations, this could be implemented by sampling a large amount of points of competing designs \mathbf{x}_i . However the sampling of \mathbf{x}_i is non-trivial especially because the boundary of X^0 could be irregular. For hypercubic regions, space-filling DOE approaches, e.g., the Optimal Latin Hypercube (Jin et al., 2005) could be utilized to generate the sampling points. Due to the difficulties brought up, we find that the ‘worst-case’ metric is the easiest to implement for problems with a

continuous design space. Optimization could be used to locate the worst-case point \mathbf{x}_i , by taking $P[Z(\mathbf{x}) > 0]$ as the objective to minimize and treating $P[|Z(\mathbf{x})| < H] < c$ (i.e., $\mathbf{x} \notin X^0$) and $\mathbf{x} \in \Omega_d$ as the constraints. Because only one design variable is considered in our example problem, we evenly sampled 101 grid points in the span of $[0, 1]$ of \mathbf{x} for illustration.

Figs. 10a and 10b show respectively the mean and 95% confidence interval of $Z(\mathbf{x})$ for two experimental sizes, 6 and 6+3, where the solid bold portions indicate the indiffereniable region X^0 in the small neighborhood of \mathbf{x}^* (identified by minimizing $\hat{f}(\mathbf{x}_i)$). It is noted that the indiffereniable region X^0 expands when the uncertainty in $Z(\mathbf{x})$ is reduced. For the ‘worst-case’ metric, the most competing design to \mathbf{x}^* is located at the boundary of X^0 . Table 3 provides the calculated M_D values based on all three proposed forms. The increased M_D values with more physical experiments reflect the improved confidence of claiming \mathbf{x}_4 as the optimal design after the resolution of $f(\mathbf{x})$ is improved.

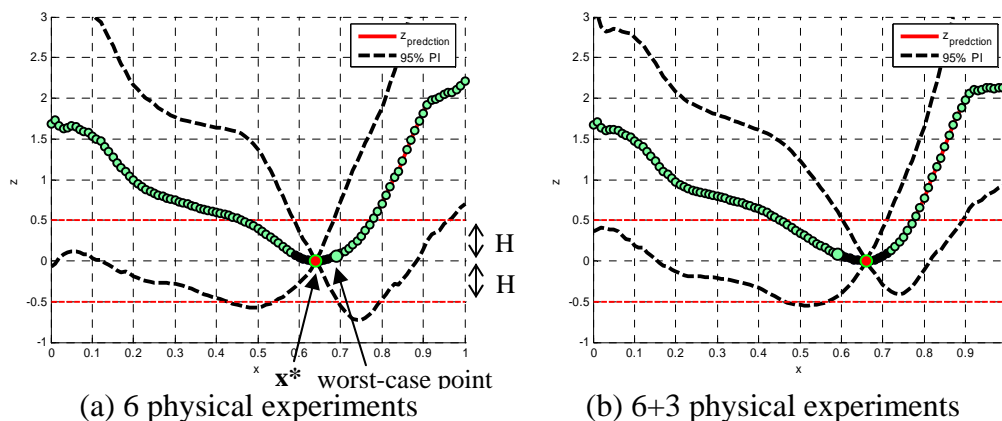
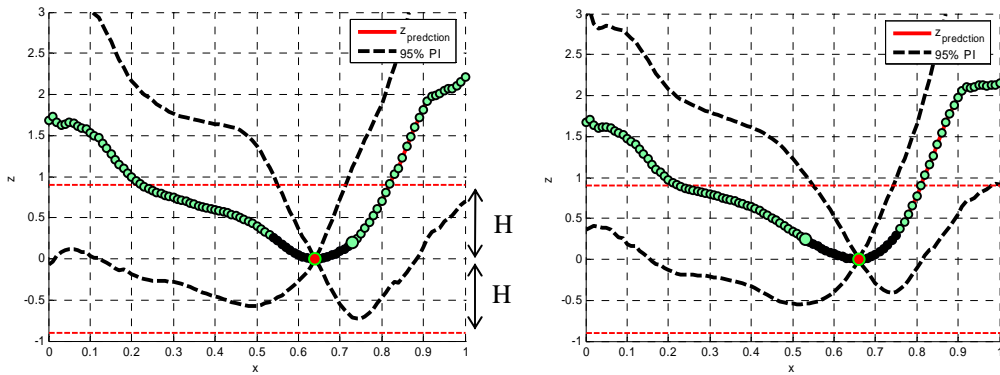


Figure 10. Mean and 95% confidence interval of $Z(\mathbf{x})$ (dB) ($H=0.5$, $c=95\%$)

Table 3. Calculation of desion validation metrics at different tolerance H and experiment size ($c=95\%$)

Phy. exp. #	H=0.5(dB)		H=0.9(dB)	
	6	6+3	6	6+3
M_D^M (multip)	0.8953	0.9203	0.9281	0.9535
M_D^A (additive)	0.8122	0.8161	0.7516	0.7457
M_D^W (worst-case)	0.5947	0.6289	0.6706	0.7341
Worst-case point	$\mathbf{x}=0.69$	$\mathbf{x}=0.59$	$\mathbf{x}=0.73$	$\mathbf{x}=0.53$

Fig. 11 also shows the results when the tolerance H is set at a higher value, 0.9 (dB). It is noted that the indiffereniable region X^0 expands with more designs previously considered as competing now being included as indiffereniable to \mathbf{x}^* . Multiplicative and Worst-case metrics increase as a result of the less strict tolerance, which is an expected outcome because when a larger tolerance H is specified by a designer, it implies that lower resolution of the model is demanded.



(a) 6 physical experiments (b) 6+3 physical experiments
 Figure 11. Mean and 95% confidence interval of $Z(\mathbf{x})$ ($H=0.9$, $c=95\%$)

6. CLOSURE

In this work, a design-driven validation approach is proposed along with a *Bayesian prediction procedure* that provides quantitative assessments of model uncertainty as well as some decision validation metrics that provide probabilistic assessments of designer's confidence in making a specific design choice. Unlike most of the existing model validation works that focus on the assessment of model accuracy, model validation is viewed in this work as a process to improve designer's confidence in making a design choice using the improved predictive model, which is the augmented model that includes both the original computer model as well as the estimate of the bias function.

Our work results in a full Bayesian analysis procedure for predicting computer model bias and the true model output. Since the analytical derivation is obtained for Bayesian model parameters, our approach is expected to be more accurate and economically sound compared to the conventional numerical approach to Bayesian analysis. In engineering applications where it is too expensive to obtain experimental data, the Bayesian inference approach offers much flexibility as additional design knowledge and information can be

easily incorporated through prior distributions. It also offers rigorous methods for quantifying the model uncertainty in an intended design domain that may interpolate as well as extrapolate from a tested domain. With the Bayesian approach, uncertainty in prediction related to the lack of experiment data can be captured by the magnitude of uncertainty of the bias function.

Even though our approach is used for validating deterministic models, various sources of uncertainties are considered. Uncertainties associated with physical experiment setup and measurements are modeled as experiment error. We have treated the model uncertainty derived from the bias function and the parameter uncertainty related to a design condition differently. The consideration of parameter uncertainty is embedded into the design objective formulation, such as the robust design objective, while the prediction uncertainty of a design objective is quantified with respect to only the model uncertainty. Treating model uncertainty separately from design variable/parameter uncertainty is especially useful for guiding the sequential experimentation process, as its goal is to reduce model uncertainty.

Our work offers a new and improved way of viewing model validation by relating its definition to a specific design choice related to a specific design objective. The approach can be used to overcome the limitations of many existing model validation approaches by providing direct estimate of the global impact of uncertainty sources on the confidence in a design decision. Our proposed decision validity metrics are generally applicable for both cases with either a discrete or continuous set of design candidates, with the worst-case metric demonstrated to be the most appropriate. As it has been illustrated, besides the model itself, the validation result highly depends on subjective inputs from designers, such as the construction of the design objective function, and the specification of tolerance and confidence level in identifying the indiffereniable region.

Even though our approach is demonstrated for a simplified one dimensional engineering design problem for ease of visualization, the same approach can be applied to problems with multidimensional design inputs and the interest is always to provide the probabilistic assessment on whether the design objective value of one particular design is better than the others. Since the analytical derivation has been obtained for implementing

the Bayesian approach, the proposed method can be easily extended to multidimensional problems.

Future research is planned for particularizing the proposed Bayesian procedure and statistical inferences for specific engineering applications where the natures of available experimental and computational data vary. Methods for incorporating designers' belief into Bayesian modeling based on prior knowledge and experience will be further examined. The obtained bias function can be further used to assess the global predictive capability of a model over both the tested and untested regions. Strategies of using decision validation metrics for guiding the sequential experimentation is also being developed. The role of decision validation metrics in engineering design will be further extended by introducing not only product design decisions but also decisions in allocating the resources for physical and computer experiments. This will require the incorporation of decision analysis techniques to study the tradeoffs involved in model refinement and uncertainty reduction.

ACKNOWLEDGMENTS

The grant support from National Science Foundation (NSF) to this collaborative research between Northwestern University (DMI – 0522662) and Georgia Tech (DMI-0522366) is greatly appreciated. The views expressed are those of the authors and do not necessarily reflect the views of the sponsors. We are grateful for the insights provided by anonymous reviewers, whose comments have helped us tremendously for improving the clarity of the paper.

APPENDIX

Table A.1 Nine (9) physical experiments

i	1	2	3	4	5	6	7
$\mathbf{x}_i \in D_e$	0.3000	0.5000	0.7000	0.7200	0.7400	0.8500	0.6300
$y^e(\mathbf{x}_i)$	55.4815	54.8042	54.6803	54.9198	54.8282	55.4664	54.7931
i	8	9					
$\mathbf{x}_i \in D_e$	0.6600	0.6900					
$y^e(\mathbf{x}_i)$	54.7150	54.8351					

Table A.2 Nine (9) computer experiments

i	1	2	3	4	5	6	7
$\mathbf{x}_i \in D_m$	0.0200	0.1400	0.2600	0.3800	0.5000	0.6200	0.7400
$y^m(\mathbf{x}_i)$	56.2011	55.6167	55.4132	55.3859	55.1350	54.7155	54.6321

i	8	9					
$\mathbf{x}_i \in D_m$	0.8600	0.9800					
$y^m(\mathbf{x}_i)$	55.2739	56.5551					

Derivations of Eqns. (4-10)

(1) The Posteriors of β_δ and σ_δ^2

The posteriors of β_δ and σ_δ^2 are

$$\beta_\delta | \mathbf{y}^e, \mathbf{y}^m, \theta_{-(\beta_\delta)} = \beta_\delta | \mathbf{y}^e, \mathbf{y}_{n_e}^m, \sigma_\delta^2, \phi_\delta, \sigma_\varepsilon^2 \sim N(\mathbf{A}_\delta \mathbf{v}_\delta, \sigma_\delta^2 \mathbf{A}_\delta)$$

and

$$\begin{aligned} \sigma_\delta^2 | \mathbf{y}^e, \mathbf{y}^m, \theta_{-(\beta_\delta, \sigma_\delta^2)} &= \sigma_\delta^2 | \mathbf{y}^e, \mathbf{y}_{n_e}^m, \phi_\delta, \tau \\ &\sim IG\left(\alpha_\delta + \frac{n_e}{2}, \gamma_\delta + \frac{1}{2} \left[(\mathbf{y}^e - \mathbf{y}_{n_e}^m)^T (\mathbf{R}_\delta + \tau \mathbf{I}_{n_e})^{-1} (\mathbf{y}^e - \mathbf{y}_{n_e}^m) + \mathbf{b}_\delta \mathbf{V}_\delta^{-1} \mathbf{b}_\delta - \mathbf{v}_\delta^T \mathbf{A}_\delta \mathbf{v}_\delta \right] \right), \end{aligned}$$

where

$$\begin{aligned} \mathbf{A}_\delta^{-1} &= \mathbf{F}_\delta^T (\mathbf{R}_\delta + \tau \mathbf{I}_{n_e})^{-1} \mathbf{F}_\delta + \mathbf{V}_\delta^{-1}, \\ \mathbf{v}_\delta &= \mathbf{F}_\delta^T (\mathbf{R}_\delta + \tau \mathbf{I}_{n_e})^{-1} (\mathbf{y}^e - \mathbf{y}_{n_e}^m) + \mathbf{V}_\delta^{-1} \mathbf{b}_\delta, \\ \tau &= \sigma_\varepsilon^2 / \sigma_\delta^2. \end{aligned}$$

and \mathbf{I}_{n_e} is an $n_e \times n_e$ identity matrix.

(2) The posterior of $\delta(\mathbf{x})$

It can be shown that

$$\begin{aligned} \delta(\mathbf{x}) | \mathbf{y}^e, \mathbf{y}^m, \theta \\ \sim N(\mathbf{f}_\delta^T(\mathbf{x}) \beta_\delta + \mathbf{r}_\delta^T(\mathbf{x}) (\mathbf{R}_\delta + \tau \mathbf{I}_{n_e})^{-1} (\mathbf{y}^e - \mathbf{y}_{n_e}^m - \mathbf{F}_\delta \beta_\delta), \sigma_\delta^2 \left[1 - \mathbf{r}_\delta^T(\mathbf{x}) (\mathbf{R}_\delta + \tau \mathbf{I}_{n_e})^{-1} \mathbf{r}_\delta(\mathbf{x}) \right]), \end{aligned}$$

and the posterior of $\delta(\mathbf{x})$ given ϕ_δ and τ is

$$\begin{aligned} p(\delta(\mathbf{x}) | \mathbf{y}^e, \mathbf{y}^m, \phi_\delta, \tau) \\ = \iint_{\beta_\delta, \sigma_\delta^2} p(\delta(\mathbf{x}) | \mathbf{y}^e, \mathbf{y}^m, \theta) \cdot p(\beta_\delta | \mathbf{y}^e, \mathbf{y}_{n_e}^m, \sigma_\delta^2, \phi_\delta, \tau) \cdot p(\sigma_\delta^2 | \mathbf{y}^e, \mathbf{y}_{n_e}^m, \phi_\delta, \tau) d\beta_\delta d\sigma_\delta^2 \end{aligned}$$

where

$$\begin{aligned} p(\delta(\mathbf{x}) | \cdot) &\propto (\sigma_\delta^2)^{-\frac{1}{2}} \cdot \exp \left\{ -\frac{\left[\delta(\mathbf{x}) - \mathbf{r}_\delta^T(\mathbf{x}) \mathbf{B} (\mathbf{y}^e - \mathbf{y}_{n_e}^m) - \mathbf{h}^T \beta_\delta \right]^2}{2\sigma_\delta^2 (1 - \mathbf{r}_\delta^T(\mathbf{x}) \mathbf{B} \mathbf{r}_\delta(\mathbf{x}))} \right\}, \\ p(\beta_\delta | \cdot) &\propto (\sigma_\delta^2)^{-\frac{p+1}{2}} \cdot \exp \left\{ -\frac{(\beta_\delta - \mathbf{A}_\delta \mathbf{v}_\delta)^T \mathbf{A}_\delta^{-1} (\beta_\delta - \mathbf{A}_\delta \mathbf{v}_\delta)}{2\sigma_\delta^2} \right\}, \\ p(\sigma_\delta^2 | \cdot) &\propto (\sigma_\delta^2)^{-\alpha_\delta - \frac{n_e}{2} - 1} \cdot \exp \left\{ -\frac{\gamma_\delta + \frac{1}{2} \left[(\mathbf{y}^e - \mathbf{y}_{n_e}^m)^T \mathbf{B} (\mathbf{y}^e - \mathbf{y}_{n_e}^m) + \mathbf{b}_\delta \mathbf{V}_\delta^{-1} \mathbf{b}_\delta - \mathbf{v}_\delta^T \mathbf{A}_\delta \mathbf{v}_\delta \right]}{\sigma_\delta^2} \right\}, \end{aligned}$$

where $\mathbf{B} = (\mathbf{R}_\delta + \tau \mathbf{I}_{n_e})^{-1}$, and $\mathbf{h} = \mathbf{f}_\delta(\mathbf{x}) - \mathbf{F}_\delta^T \mathbf{B} \mathbf{r}_\delta(\mathbf{x})$. Collecting all terms involving β_δ together gives

$$\exp\left\{-\frac{1}{2\sigma_\delta^2}[\beta_\delta^T \mathbf{A}^{-1} \beta_\delta + \mathbf{v}^T \beta_\delta]\right\},$$

where

$$\mathbf{A}^{-1} = \mathbf{A}_\delta^{-1} + \frac{\mathbf{h}\mathbf{h}^T}{1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})},$$

$$\mathbf{v} = \mathbf{v}_\delta + \frac{\delta(\mathbf{x}) - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}(\mathbf{y}^e - \mathbf{y}_{n_e}^m)}{1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})} \cdot \mathbf{h}.$$

Hence,

$$\int_{\beta_\delta} \exp\left\{-\frac{1}{2\sigma_\delta^2}[\beta_\delta^T \mathbf{A}^{-1} \beta_\delta + \mathbf{v}^T \beta_\delta]\right\} d\beta_\delta \propto (\sigma_\delta^2)^{\frac{p+1}{2}} \cdot \exp\left\{\frac{\mathbf{v}^T \mathbf{A} \mathbf{v}}{2\sigma_\delta^2}\right\}.$$

Collecting all terms involving σ_δ^2 gives

$$(\sigma_\delta^2)^{-\frac{1}{2} - \alpha_\delta - \frac{n_e}{2} - 1} \cdot \exp\left\{-\frac{\gamma}{\sigma_\delta^2}\right\},$$

where

$$\gamma = \gamma_\delta + \frac{(\mathbf{y}^e - \mathbf{y}_{n_e}^m)^T \mathbf{B}(\mathbf{y}^e - \mathbf{y}_{n_e}^m) + \mathbf{b}_\delta \mathbf{V}_\delta^{-1} \mathbf{b}_\delta}{2} + \frac{[\delta(\mathbf{x}) - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}(\mathbf{y}^e - \mathbf{y}_{n_e}^m)]^2}{2(1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x}))} - \frac{\mathbf{v}^T \mathbf{A} \mathbf{v}}{2}.$$

Performing the integration over σ_δ^2 yields

$$\int_{\sigma_\delta^2} (\sigma_\delta^2)^{-\frac{1}{2} - \alpha_\delta - \frac{n_e}{2} - 1} \exp\left(-\frac{\gamma}{\sigma_\delta^2}\right) d\sigma_\delta^2 \propto \gamma^{-\frac{1+2\alpha_\delta+n_e}{2}}.$$

Therefore,

$$p(\delta(\mathbf{x}) | \mathbf{y}^e, \mathbf{y}^m) \propto \gamma^{-\frac{1+2\alpha_\delta+n_e}{2}}.$$

If we can write γ in the form of

$$\gamma = C \cdot \left(1 + \frac{1}{2\alpha_\delta + n_e} \cdot \frac{(\delta(\mathbf{x}) - \mu_{\delta|e,m}(\mathbf{x}))^2}{\sigma_{\delta|e,m}^2(\mathbf{x})}\right),$$

where C is any constant, then

$$p(\delta(\mathbf{x}) | \mathbf{y}^e, \mathbf{y}^m) \propto \left(1 + \frac{1}{2\alpha_\delta + n_e} \cdot \frac{(\delta(\mathbf{x}) - \mu_{\delta|e,m}(\mathbf{x}))^2}{\sigma_{\delta|e,m}^2(\mathbf{x})}\right)^{-\frac{1+2\alpha_\delta+n_e}{2}},$$

which implies that $\delta(\mathbf{x}) | \mathbf{y}^e, \mathbf{y}^m$ has a noncentral t distribution with degree of freedom $2\alpha_\delta + n_e$, noncentrality parameter $\mu_{\delta|e,m}(\mathbf{x})$, and scale parameter $\sigma_{\delta|e,m}(\mathbf{x})$.

Expanding γ gives

$$\begin{aligned}
\gamma &= \gamma_\delta + \frac{1}{2} \left[(\mathbf{y}^e - \mathbf{y}_{n_e}^m)^T \mathbf{B} (\mathbf{y}^e - \mathbf{y}_{n_e}^m) + \mathbf{b}_\delta \mathbf{V}_\delta^{-1} \mathbf{b}_\delta \right] \\
&+ \frac{\delta^2(\mathbf{x}) - 2\mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}(\mathbf{y}^e - \mathbf{y}_{n_e}^m) \cdot \delta(\mathbf{x}) + [\mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}(\mathbf{y}^e - \mathbf{y}_{n_e}^m)]^2}{2[1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})]} \\
&- \frac{\mathbf{h}^T \mathbf{A} \mathbf{h} \cdot \delta^2(\mathbf{x})}{2[1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})]^2} - \left[\frac{\mathbf{h}^T \mathbf{A} \mathbf{v}_\delta}{1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})} - \frac{\mathbf{h}^T \mathbf{A} \mathbf{h} \cdot \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}(\mathbf{y}^e - \mathbf{y}_{n_e}^m)}{[1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})]^2} \right] \delta(\mathbf{x}) \\
&- \frac{\mathbf{v}_\delta^T \mathbf{A} \mathbf{v}_\delta}{2} - \frac{\mathbf{h}^T \mathbf{A} \mathbf{h} \cdot [\mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}(\mathbf{y}^e - \mathbf{y}_{n_e}^m)]^2}{2[1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})]^2} + \frac{\mathbf{h}^T \mathbf{A} \mathbf{v}_\delta \cdot \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}(\mathbf{y}^e - \mathbf{y}_{n_e}^m)}{1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})}.
\end{aligned}$$

As

$$\begin{aligned}
\mathbf{A} &= \mathbf{A}_\delta - \frac{\mathbf{A}_\delta \mathbf{h} \mathbf{h}^T \mathbf{A}_\delta}{\mathbf{h}^T \mathbf{A}_\delta \mathbf{h} + 1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})}, \\
\mathbf{h}^T \mathbf{A} \mathbf{h} &= \frac{1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})}{\mathbf{h}^T \mathbf{A}_\delta \mathbf{h} + 1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})} \cdot \mathbf{h}^T \mathbf{A}_\delta \mathbf{h}, \\
\mathbf{h}^T \mathbf{A} \mathbf{v}_\delta &= \frac{1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})}{\mathbf{h}^T \mathbf{A}_\delta \mathbf{h} + 1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})} \cdot \mathbf{h}^T \mathbf{A}_\delta \mathbf{v}_\delta, \\
\mathbf{v}_\delta^T \mathbf{A} \mathbf{v}_\delta &= \mathbf{v}_\delta^T \mathbf{A}_\delta \mathbf{v}_\delta - \frac{(\mathbf{h}^T \mathbf{A}_\delta \mathbf{v}_\delta)^2}{\mathbf{h}^T \mathbf{A}_\delta \mathbf{h} + 1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})},
\end{aligned}$$

we have

$$\begin{aligned}
\gamma &= \gamma_\delta + \frac{1}{2} \left[(\mathbf{y}^e - \mathbf{y}_{n_e}^m)^T \mathbf{B} (\mathbf{y}^e - \mathbf{y}_{n_e}^m) + \mathbf{b}_\delta \mathbf{V}_\delta^{-1} \mathbf{b}_\delta - \mathbf{v}_\delta^T \mathbf{A}_\delta \mathbf{v}_\delta \right] \\
&+ \frac{1}{2} \frac{[\mathbf{h}^T \mathbf{A}_\delta \mathbf{v}_\delta + \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}(\mathbf{y}^e - \mathbf{y}_{n_e}^m) - \delta(\mathbf{x})]^2}{\mathbf{h}^T \mathbf{A}_\delta \mathbf{h} + 1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mu_{\delta|e,m}(\mathbf{x}) &= \mathbf{h}^T \mathbf{A}_\delta \mathbf{v}_\delta + \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}(\mathbf{y}^e - \mathbf{y}_{n_e}^m) \\
\sigma_{\delta|e,m}^2(\mathbf{x}) &= \frac{Q_\delta^2}{2\alpha_\delta + n_e} \cdot [\mathbf{h}^T \mathbf{A}_\delta \mathbf{h} + 1 - \mathbf{r}_\delta^T(\mathbf{x})\mathbf{B}\mathbf{r}_\delta(\mathbf{x})]
\end{aligned}$$

where

$$Q_\delta^2 = 2\gamma_\delta + (\mathbf{y}^e - \mathbf{y}_{n_e}^m)^T \mathbf{B} (\mathbf{y}^e - \mathbf{y}_{n_e}^m) + \mathbf{b}_\delta \mathbf{V}_\delta^{-1} \mathbf{b}_\delta - \mathbf{v}_\delta^T \mathbf{A}_\delta \mathbf{v}_\delta$$

Substituting \mathbf{B} and \mathbf{h} into above equations for $\mu_{\delta|e,m}$ and $\sigma_{\delta|e,m}^2$, we have

$$\begin{aligned}
\mu_{\delta|e,m}(\mathbf{x}) &= \mathbf{f}_\delta^T(\mathbf{x})\mathbf{A}_\delta \mathbf{v}_\delta + \mathbf{r}_\delta^T(\mathbf{x})(\mathbf{R}_\delta + \tau \mathbf{I}_{n_e})^{-1}(\mathbf{y}^e - \mathbf{y}_{n_e}^m - \mathbf{F}_\delta \mathbf{A}_\delta \mathbf{v}_\delta) \\
\sigma_{\delta|e,m}^2(\mathbf{x}) &= \frac{Q_\delta^2}{2\alpha_\delta + n_e} \cdot \left(1 - \begin{bmatrix} \mathbf{f}_\delta(\mathbf{x}) \\ \mathbf{r}_\delta(\mathbf{x}) \end{bmatrix}^T \begin{bmatrix} -\mathbf{V}_\delta^{-1} & \mathbf{F}_\delta^T \\ \mathbf{F}_\delta & \mathbf{R}_\delta + \tau \mathbf{I}_{n_e} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{f}_\delta(\mathbf{x}) \\ \mathbf{r}_\delta(\mathbf{x}) \end{bmatrix} \right)
\end{aligned}$$

where

$$Q_\delta^2 = 2\gamma_\delta + (\mathbf{y}^e - \mathbf{y}_{n_e}^m)^T (\mathbf{R}_\delta + \tau \mathbf{I}_{n_e})^{-1} (\mathbf{y}^e - \mathbf{y}_{n_e}^m) + \mathbf{b}_\delta \mathbf{V}_\delta^{-1} \mathbf{b}_\delta - \mathbf{v}_\delta^T \mathbf{A}_\delta \mathbf{v}_\delta.$$

REFERENCES

- Ang, J.A., Trucano, T.G., and Luginbuhl, D.R., "Confidence in ASCI Scientific Simulations", Ninth Nuclear Explosives Code Developers' Conference, San Diego, CA, Oct. 22-25, 1996.
- Apley, D., Liu, J., and Chen, W., "Understanding the Effects of Model Uncertainty in Robust Design with Computer Experiments", *ASME Journal of Mechanical Design*, 128, pp. 745-958, July 2006.
- Apostolakis, G., "A Commentary on Model Uncertainty", Model Uncertainty: Its Characterization and Quantification, editors: Mosleh, A, Siu, N, Smidts, C. and Lui, C, NUREG/CP-0138, U.S. Nuclear Regulatory Commission, 1994.
- Bayarri, M.J., Berger, J.O., Higdon, D., Kennedy, M.C., Kottas, A., Paulo, R., Sacks, J., Cafeo, J.A., Cavendish, J., Lin, C.H., and Tu, J., "A Framework for Validation of Computer Models", Foundations for Verification and Validation in 21st Century Workshop, Johns Hopkins University, October 22-23, 2002.
- Bayarri, M.J., Berger, J.O., Higdon, D., Kennedy, M.C., Kottas, A., Paulo, R., Sacks, J., Cafeo, J.A., Cavendish, J., Lin, C.H., and Tu, J., "A Framework for Validation of Computer Models", Foundations for Verification and Validation in 21st Century Workshop, Johns Hopkins University, October 22-23, 2002.
- Buranathiti, T., Cao, J., Chen, W., Baghdasaryan, L., and Xia, Z.C., "Approaches for Model Validation: Methodology and Illustration on a Sheet Metal Flanging Process", *ASME Journal of Manufacturing Science and Engineering*, in press, 126, November, 2004.
- Buslik, A., "A Bayesian Approach to Model Uncertainty", Model Uncertainty: Its Characterization and Quantification, editors: Mosleh, A, Siu, N, Smidts, C. and Lui, C, U.S. Nuclear Regulatory Commission, NUREG/CP-0138, 1994.
- Cafeo, J.A., and Thacker, B.H., "Concepts and Terminology of Validation for Computational Solid Mechanics Models", SAE 2004 World Congress & Exhibition, Detroit, MI, March, 2004.
- Chen, C. H., Lin, J., Yucesan, E., and Chick, S. E., "Simulation Budget Allocation for Further Enhancing the Efficiency of Ordinal Optimization", *Discrete Event Dynamic Systems*, 10(3), 251-270, 2000.
- Chen, W., Baghdasaryan, L., Buranathiti, T., and Cao, J., "Model Validation via Uncertainty Propagation and Data Transformations", *AIAA Journal*, 42(7), 1406-1415, 2004.
- DoD, DoD Directive No. 5000.61, "Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A)", Defense Modeling and Simulation Office, www.dmsomil/docslib.
- Doebling, S.W., Hemex, F.M., Schultz, J.F., Girrens, S.P., "Overview of Structural Dynamics Model Validation Activities at Los Alamos National Laboratory", Proc AIAA/ASME/ASCE/AHS/ASC 43rd Structures, Structural Dynamics, and Materials Conf., AIAA 2002-1643, Denver, CO, April 22-25, 2002.
- Easterling, R. G. & Berger, J. O., Statistical Foundations for The Validation of Computer Models, presented at Computer Model Verification and Validation in the 21st Century Workshop, Johns Hopkins University, 2002.
- Freese, F., Testing Accuracy, *Forest Science*, 6(2), 139-145, 1960.

- Geyer, C. J., "Practical Markov Chain Monte Carlo," *Statistical Science*, 7(4):473-511, 1992.
- Gregoire, T., G., and Reynolds, M., R., *Accuracy Testing and Estimation Alternatives*, *Forest Science*, 34(2), 302-320, 1988.
- Gu, L., Yang, R.J., "Recent Applications on Reliability-based Optimization of Automotive Structures," SAE Technical Paper Series: 2003-01-0152, 2003.
- Hanson, K.M., "A Framework for Assessing Uncertainties in Simulation Predictions", *Physica D*. 133, 179-188, 1999.
- Hastie, T., Tibshirani, R., Friedman, J., *The Elements of Statistical Learning*, Springer-Verlag, 2000.
- Hazelrigg, G.A., "Thoughts on Model Validation for Engineering Design", ASME Design Technical Conference, DETC2003/DTM-48632, Chicago, IL, Sept. 2-6, 2003.
- Hazelrigg, G.A., "On the Role and Use of Mathematical Models in Engineering Design", *Transactions of ASME, Journal of Mechanical Design*, 121(3), 336-341, 1999.
- Hills, G. R., and Trucano, T. G., "Statistical Validation of Engineering and Scientific Models: Background", SAND99-1256, 1999.
- Hoffman, R. M., Sudjianto, A., Du, X., and Stout, J., "Robust Piston Design and Optimization Using Piston Secondary Motion Analysis," SAE World Congress, 2003-01-0148, Detroit, Michigan, Mar. 3-6, 2003.
- Jin, R., Chen, W., and Sudjianto, A., "An Efficient Algorithm for Constructing Optimal Design of Computer Experiments", *Journal of Statistical Planning and Inference*, 134(1), 268-287, 2005.
- Kennedy, M. C. and O'Hagan, A., "Bayesian Calibration of Computer Experiments", *Journal of the Royal Statistical Society*, B. 63, 425-464, 2001.
- Mahadevan, S., and Rebba, R., "Validation of Reliability Computational Models using Bayes Networks," *Journal of Reliability Engineering and System Safety*, 87, pp. 223-232, 2005.
- Malak, R.J. and Paredis, "On Characterizing and Assessing the Validity of Behavioral Models and Their Predictions", 2004 ASME Design Technical Conferences, DETC2004-57452, Salt Lake City, Utah, Sept. 28-Oct. 2, 2004.
- Marczyk, J., Holzner, M., et.al., "Stochastic Automotive Crash Simulation; A new Frontier in Virtual prototyping," Proceedings of the Pam 97 user Conference, Pragu, Czech Republic, October 16-17, 1997.
- McAdams, D.A. and Dym, C.L., "Modeling and Information in the Design Process", 2004 ASM Design Technical Conferences, DETC2004-57101, Salt Lake City, Utah, Sept. 28-Oct. 2, 2004.
- Oberkampf, W. and Barone, M., "Measures of Agreement Between Computation and Experiment: Validation Metrics", 34th AIAA Fluid Dynamics Conference and Exhibit, AIAA-2004-2626, Portland, Oregon, June 28-1, 2004.
- Oberkampf, W.L., Deland S.M., Rutherford, B.M, Diegert, K.V. and Alvin, K.F., "A New Methodology for the Estimation of Total Uncertainty in Computational Simulation", AIAA System Dynamics, Material, Conference, AIAA-99-1612, 3061 – 3083, 1999.
- Oberkampf, W.L. and Trucano, T.G., "Validation Methodology in Computational Fluid Dynamics", AIAA 2000-2549, Fluids 2000, Denver, CO, 2000.

- Oberkampf, W.L., Trucano, T.G., and Hirsch, C., "Verification, Validation, and Predictive Capability in Computational Engineering and Physics", SAND2003-3769, February, 2003.
- Qian, Z. and Wu, C. F. J., "Bayesian hierarchical modeling for integrating low-accuracy and highaccuracyexperiments," Georgia Tech technical Report, 2005.
- Reese C., Wilson A., Hamada M., Martz H., Ryan K., "Integrated Analysis of Computer and Physical xperiments", Technometrics, 46(2), pp. 153-164, 2004.
- Reynolds, M.R., Jr., "Estimating the Error in Model Predictions", Forest Science, 30(2), 454-469, 1984.
- Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P., "Design and analysis of computer experiments," Statistical Science, 4(4), pp. 409-435, 1989.
- Santner, T.J., B.J. Williams, and W.I. Notz (2003), "The Design and Analysis of Computer Experiments," New York: Springer-Verlag.
- Trucano, T. G., "Prediction and Uncertainty in Computational Modeling of Complex Phenomena, A Whitepaper", Sandia report, SAND98-2776, 1998.
- Wang, N., Ge, P., "Study of Metamodeling Techniques and Their Applications in Engineering Design," ASME-MED Manufacturing Science and Engineering, Vol. 10, pp. 89-95, 1999.
- Wang, S., Chen, W., and Tsui, K, "Bayesian Validation of Computer Models," Georgia Tech technical report, 2006.