

AN ENTROPY BASED SEQUENTIAL CALIBRATION APPROACH FOR STOCHASTIC COMPUTER MODELS

Yuan Jun
Szu Hui Ng

Department of Industrial & Systems Engineering
National University of Singapore
SINGAPORE

ABSTRACT

Computer models are widely used to simulate complex and costly real processes and systems. In the calibration process of the computer model, the calibration parameters are adjusted to fit the model closely to the real observed data. As these calibration parameters are unknown and are estimated based on observed data, it is important to estimate it accurately and account for the estimation uncertainty in the subsequent use of the model. In this paper, we study in detail an empirical Bayes approach for stochastic computer model calibration that accounts for various uncertainties including the calibration parameter uncertainty, and propose an entropy based criterion to improve on the estimation of the calibration parameter. This criterion is also compared with the EIMSPE criterion.

1 INTRODUCTION

Computer models are commonly used as a tool to study and evaluate the performance of complex systems. When the computer model is applied to evaluate various policy scenarios or is used to predict the behavior of the real system, it is important to obtain accurate calibration parameters and also to improve the model's predictive accuracy and capability. In computer model analysis, both validation and calibration are closely related to the model's predictive performance. Validation is the process of ensuring that the computer model accurately represents the real process, while calibration is the process of adjusting the unknown input parameters of the computer model to fit the real observed data. This is done by comparing the computer model output with the observed data. These unknown input parameters are usually unobservable or unmeasurable in the real process but need to be specified in the computer model. Trucano et al. (2006) and Oberkampf and Roy (2010) provide comprehensive discussions about the differences between calibration and validation for computational science and engineering. In this paper, we focus on studying calibration procedures to update the model.

In many practical applications, calibration parameters can significantly influence the performance of the computer model. The importance of model calibration has been recognized in many practical problems, such as in macroscopic traffic simulation model calibration, Lee and Ozbay (2009), node calibration of wireless sensor networks, Cho et al. (2010), and healthcare microsimulation model calibration, Rutter et al. (2009).

Various approaches have been proposed for computer model calibration. One approach is to use an effective and efficient search algorithm that is directly applied to the computer model, such as stochastic approximation methods discussed by Yuan et al. (2013a). However, this approach may not be efficient

when the computer models are extremely time consuming and computationally expensive as it usually requires a relatively large number of simulation runs. Another popular and much more efficient approach is to use surrogates. These simpler and faster statistical approximations are used in the calibration process instead of the original complex computer models. More discussion on developing surrogates for computer models can be found in Kennedy and O'Hagan (2001), Santner et al. (2003).

In the development and use of the surrogate, various sources of uncertainties are introduced. These include parameter uncertainties (including calibration parameter uncertainty), model inadequacy, residual variability and observation error. When the model is subsequently used for evaluation and decision making, not only the point estimator but also the uncertainty information about the estimator is required to better reflect the overall confidence in the estimator and to provide for more informed decisions. Therefore, it is important to account for various uncertainties in predicting the behavior of the real process. Within these uncertainties, calibration parameter uncertainty sometimes can have significant effects on overall predictive uncertainty. Therefore, it is important to consider this uncertainty and its effects on the subsequent prediction in many applications. Yuan and Ng (2013b) proposed a Bayesian approach for stochastic computer model calibration using the Gaussian process (GP) as a surrogate model. Their approach accounts for all sources of uncertainty in the computer model analysis including the calibration parameter uncertainty. To provide for faster analytical calibration and prediction results, an approximation based on the estimation of several hyperparameters was also provided. The effects of these estimations under various data and noise scenarios however were not well studied.

The purpose of this article is to study in more detail the model calibration approach developed in Yuan and Ng (2013b), and study the effects of estimating certain parameters in the approximated approach with the fully Bayesian MCMC approach. In addition a new sequential criterion is proposed to improve on some of the shortcomings observed in the approximation method. This paper is organized as follows. In section 2, we review the model for calibration and prediction for stochastic computer models. In section 3, we study in detail the effects of noise and design size on the predictive performance. In section 4, we propose a new calibration design criterion to improve the performance of the approach. Finally in section 5, a numerical example is provided to illustrate the new criterion within the approach. A comparison is also made with the previously proposed EIMSPE criterion.

2 MODEL FORMULATION

2.1 Stochastic Model

Based on the model proposed by Kennedy and O'Hagan (2001), the relationship between the real observation and simulation output can be represented by

$$z_i = \zeta(x_i) + e_i = S(x_i, \theta) + \delta(x_i) + e_i,$$

where z_i is the i th real observation at input level x_i , $\zeta(x_i)$ is the true output from the real process, $S(x_i, \theta)$ represents the "true" simulation output at x_i and the optimum calibration parameter θ . Similar to previous works, the optimum θ is viewed as the value that best fits the computer model output to the real process. $\delta(x_i)$ is the model inadequacy or discrepancy term, which is considered to be independent of $S(x_i, \theta)$, and e_i is the observation error. More discussions about this relationship can be found in Kennedy and O'Hagan (2001); Yuan and Ng (2013b).

$S(x_i, \theta)$ is considered deterministic in most studies, where the computer model output is always identical with the same inputs x_i and θ . In many practical applications, the computer models built are stochastic. This means that simulations at the same input levels give different outputs. One typical way to represent the relationship between the observed and expectation of the stochastic simulation output (Yuan and Ng, 2013b) is

$$\bar{y}(x_i, \theta) = S(x_i, \theta) + \varepsilon_i,$$

where $S(x_i, \theta)$ denotes the expectation of the stochastic simulation output for a given x_i and θ , $\bar{y}(x_i, \theta)$ is the observed stochastic simulation output or the average of several stochastic simulation output replications under a specified x_i and θ , and ε_i is the sampling variability inherent in the stochastic simulation.

2.2 Gaussian Process Model

Gaussian process (GP) models have been widely used as surrogates of computer models because of their flexibility and convenience (Kennedy and O'Hagan, 2001), and comprehensive details of this form applied to computer models can be found in Santner et al. (2003). In this paper, the Gaussian process model is used as a surrogate model for the computer model and the following assumptions are made:

- 1) The computer model output $S(x, t)$ is a Gaussian process with mean $\mu_S(x, t) = h_S(x, t)^T \beta_S$ and covariance function $\sigma_S^2 R_S$, where t denotes the specified input calibration parameter value which can be controlled in the computer model. Here t is used to distinguish between the optimum unknown calibration parameter value θ that we want to calibrate for the computer model. $h_S(x, t) = (h_{S,1}(x, t), \dots, h_{S,k_S}(x, t))^T$ is a vector of k_S functions of x and t and $\beta_S = (\beta_{S,1}, \dots, \beta_{S,k_S})^T$ is a vector of k_S unknown coefficients. The correlation function R_S has an exponential form

$$R_S((x_i, t_i), (x_j, t_j)) = \prod_{u=1}^p \exp\{-\phi_{S_x, u}(x_{i,u} - x_{j,u})^2\} \cdot \prod_{v=1}^q \exp\{-\phi_{S_\theta, v}(t_{i,v} - t_{j,v})^2\},$$

where p is the dimension of x , q is the dimension of θ . The parameters $\phi_{S_x, u} > 0$ and $\phi_{S_\theta, v} > 0$. We write $\phi_S = (\phi_{S_x, 1}, \dots, \phi_{S_x, p}, \phi_{S_\theta, 1}, \dots, \phi_{S_\theta, q})^T$. ε_i is treated as a nugget effect with $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ for all i . It should be noted that this is just one possible form assumed for illustrative purpose and other forms can also be used. More details on various different correlation forms can be found in Cressie (1993).

- 2) The model inadequacy term $\delta(x)$ is a Gaussian process with mean $\mu_\delta(x) = h_\delta(x)^T \beta_\delta$ and covariance function $\sigma_\delta^2 R_\delta$, where $h_\delta(x) = (h_{\delta,1}(x), \dots, h_{\delta,k_\delta}(x))^T$ is a vector of k_δ functions of x , $\beta_\delta = (\beta_{\delta,1}, \dots, \beta_{\delta,k_\delta})^T$ is a vector of k_δ unknown coefficients, and R_δ has an exponential form

$$R_\delta(x_i, x_j) = \prod_{u=1}^p \exp\{-\phi_{\delta, u}(x_{i,u} - x_{j,u})^2\}, \text{ where } \phi_{\delta, u} > 0. \text{ We write } \phi_\delta = (\phi_{\delta,1}, \dots, \phi_{\delta,p})^T.$$

- 3) The observation error e has a normal distribution with mean 0 and variance σ_e^2 .
- 4) $S(x, \theta)$, $\delta(x)$, ε and e are mutually independent.

2.3 Bayesian Calibration

In the stochastic model defined in Section 2.1 and the GP model defined in Section 2.2, the parameters that need to be specified are: $\{\beta_S, \sigma_S^2, \phi_S, \beta_\delta, \sigma_\delta^2, \phi_\delta, \sigma_\varepsilon^2, \sigma_e^2, \theta\}$. One advantage of using the Bayesian approach is that we can use the prior knowledge about the computer model and the real process to quantify the uncertainties of these parameters by assigning prior distributions on these parameters. Therefore, we can quantify various uncertainties including model inadequacy, observation error, inherent stochastic error, surrogate model uncertainty and calibration uncertainty. Further assuming that θ is an independent parameter and with assumption 4 above, we have that $\{\theta\}$, $\{\beta_S, \sigma_S^2, \phi_S\}$, $\{\beta_\delta, \sigma_\delta^2, \phi_\delta\}$, $\{\sigma_\varepsilon^2\}$ and $\{\sigma_e^2\}$ are mutually independent. We further assume that $\{\beta_S, \sigma_S^2\}$ and $\{\phi_S\}$ are independent; $\{\beta_\delta, \sigma_\delta^2\}$ and $\{\phi_\delta\}$ are independent. Then we can write the prior distribution as

$$p(\theta, \beta_S, \sigma_S^2, \phi_S, \beta_\delta, \sigma_\delta^2, \phi_\delta, \sigma_\varepsilon^2, \sigma_e^2) = p(\theta) p(\beta_S, \sigma_S^2) p(\phi_S) p(\beta_\delta, \sigma_\delta^2) p(\phi_\delta) p(\sigma_\varepsilon^2) p(\sigma_e^2).$$

We further define $\tau_1^2 = \sigma_\varepsilon^2 / \sigma_S^2$, $\tau_2^2 = \sigma_\delta^2 / \sigma_S^2$, and $\tau_3^2 = \sigma_e^2 / \sigma_S^2$, which are reparameterized for the convenience of integrating out σ_S^2 in the posterior analysis. We let $\tau^2 = \{\tau_1^2, \tau_2^2, \tau_3^2\}$, $\beta = (\beta_S^T, \beta_\delta^T)^T$, $\sigma^2 = \{\sigma_S^2, \sigma_\varepsilon^2, \sigma_\delta^2, \sigma_e^2\}$ and $\phi = \{\phi_S, \phi_\delta\}$. There are two sets of observable data in the stochastic model calibration, which are the real observations $z = (z_1, z_2, \dots, z_N)^T$ at N input points from the real process and the ob-

served stochastic simulation outputs $\bar{y} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_n)^T$ at n input points from the stochastic computer model. Then the full set of observed data is $d^T = (\bar{y}^T, z^T)$, which is assumed to be normally distributed given $\{\theta, \beta, \sigma^2, \phi\}$. This will yield the likelihood function $p(d|\theta, \beta, \sigma^2, \phi)$. With the given prior distributions, the full posterior distribution of all the parameters can be obtained.

$$p(\theta, \beta, \sigma^2, \phi | d) \propto p(\theta, \beta_S, \sigma_S^2, \phi_S, \beta_\delta, \sigma_\delta^2, \phi_\delta, \sigma_e^2, \sigma_e^2) p(d | \theta, \beta, \sigma^2, \phi)$$

Furthermore, the marginal of θ , $p(\theta | d)$, describes the uncertainty surrounding the calibration parameter, and is important for determining the estimate of θ (e.g. in the setting of θ for future computer runs). Future predictions can be made for the computer model and the real process.

$$p(Z | d) \propto \int p(Z | \theta, \beta, \sigma^2, \phi, d) p(\theta, \beta, \sigma^2, \phi | d) d\theta d\beta d\sigma^2 d\phi \tag{1}$$

Different priors can be adopted in the Bayesian analysis, depending on the prior information available. The advantage of this Bayesian approach is that with the observed data, it provides a quantification of the uncertainty about θ , and accounts for the parameter uncertainties in the subsequent predictions with the model. In this paper, we adopt mathematical convenient priors, where β_S and β_δ are assumed to have multivariate normal distributions conditional on σ_S^2 and σ_δ^2 respectively, σ_S^2 , σ_δ^2 , σ_e^2 and σ_e^2 are assumed to have inverse gamma distributions, ϕ_S and ϕ_δ are assumed to have gamma distributions. A uniform prior is placed on the calibration parameter θ . More details about these prior distributions can be found in Yuan and Ng (2013b).

With the given prior distributions, a closed form joint posterior distribution of θ , τ^2 and ϕ $p(\theta, \phi, \tau^2 | d)$ can be obtained. However, it is not practical to further integrate out the parameters ϕ and τ^2 to obtain the posterior distribution of the calibration parameter $p(\theta|d)$ for a fully Bayesian analysis. Although the MCMC approach can be used to integrate numerically, it significantly increases the computational cost while the improvement on the model predictive performance may not be significant. Therefore, the parameters ϕ and τ^2 are first estimated and an empirical Bayes approach taken instead. These parameters can then be estimated by the modes of the joint density. Yuan and Ng (2013b) proposed an approach to estimate them in two steps.

- In the first step, data \bar{y} is used to estimate parameters ϕ_S and τ_1^2 by maximizing their posterior distribution.
- In the second step, data d is used to estimate the remaining parameters ϕ_δ , τ_2^2 and τ_3^2 by the EM algorithm given the estimated ϕ_S and τ_1^2 from the first step. In this procedure, θ is treated as a latent parameter. The complete procedure is detailed in Yuan and Ng (2013b).

With the estimated parameters $\hat{\phi}$ and $\hat{\tau}$, the conditional posterior distribution $p(\theta | \hat{\phi}, \hat{\tau}, d)$ can be used to make inference about θ , such as quantifying the calibration parameter uncertainty. This uncertainty can be accounted for in the subsequent prediction by integrating out θ with respect to $p(\theta | \hat{\phi}, \hat{\tau}, d)$ in (1). However, with θ present in the joint posterior density, and treated as a latent parameter in the EM approach, the accuracy of these estimated parameters is closely related to the accuracy of the estimated calibration parameter value. More specifically, it is closely related to the estimated posterior mode of the calibration parameter. Due to the dependency of the parameters in this estimation approach, the selection of data points and accuracy of the calibration parameter can affect estimated parameters and the predictive performance of the model.

In the next section, we study in detail the effects of estimating these parameters on the calibration parameter and on the subsequent use of the model for predictions under different initial data scenarios.

3 A NUMERICAL STUDY

In this section, we numerically examine the effects of different initial data and process on the calibration and prediction performance. Specifically, we vary the random noise level and design size.

3.1 A Kinetic Model Example

The model used in this example is based on Loeppky et al. (2006), which predicts the concentration $y(x)$ of the chemical as a function of time x , with mean response governed by the equation $\phi(x|T)=u+r_0\exp(-Tx)$, where u represents the residual concentration of the chemical at the end of the reaction process, r_0 is the initial concentration of the chemical and T is an unknown decay rate that is specific to the chemical reaction under consideration. The real process model is assumed to be $z=\phi(x|T=1.7)=1.5+3.5\exp(-1.7x)+e$, where $x\in[0,3]$ and $e\sim N(0,0.1)$. We further suppose that the stochastic computer model is coded as $y=\kappa(x,\theta)=5\exp(-\theta x)+\varepsilon$, where $\varepsilon\sim N(0,\sigma^2)$. In modeling, the Gaussian process surrogate model is assumed to have a constant mean, which is reasonable in many practical applications (Santner et al., 2003). Similar to Qian and Wu (2008), we select the “location-flat” priors $N(0,\sigma_S^2)$ and $N(0,\sigma_\delta^2)$ for β_S and β_δ . The “vague” prior $IG(2,1)$ is chosen for σ_S^2 , σ_δ^2 , σ_ε^2 and σ_e^2 , and $G(2,0.1)$ is chosen as prior for each element in ϕ .

We vary the value of the variance σ^2 and the number of initial data points to assess their effects on the calibration accuracy and the predictive performance. The experimental data for the real process is collected at different input points that are evenly distributed between 0 and 3. Only one observation is taken at each input point. The latin hypercube design method is used to collect the initial computer experimental design input sets with $x\in[0,3]$ and $t\in[0,2]$. Each input has 100 replications and the mean is taken as the simulation output.

First, we examine the effects of different values of the variance σ^2 on calibration and prediction performance. We choose initial 15 real process design points and 30 computer model design points. The value of variance σ^2 is set to be low (100) and high (10000). The posterior distribution of the calibration parameter for different variance values together with the target posterior distribution obtained with a large set of data (50 real process design points and 150 computer model design points) are shown in Figure 1a. As expected, the results show that the posterior distribution of the calibration parameter for low variance scenario is closer to the target posterior than the posterior for high variance scenario. To further assess the effects on the predictive performance, we first compute the average values and the variances of the estimated parameters of the surrogate model over 10 macroreplications for different variance levels. The results are given in Table 1 which shows that the accuracy of the parameters are significantly affected when the variance is high. The average root mean square prediction error (RMSPE) of the real process over 10 macroreplications of the design is also provided. The results show that the predictive performance is significantly worse in the high variance case. The average RMSPE is also computed when the MCMC approach is applied in both cases, with results 1.453 (low variance case) and 2.327 (high variance case). In summary, the results show that the inherent noise in the computer model will influence the accuracy of the calibration parameter and can affect the accuracy of the plug-in parameters and the model predictive performance.

Next, we assess the effects of the size of the initial data on the calibration and prediction performance. The value of the variance σ^2 is set at 100. We compare the calibration and prediction performance for small initial data design (with 10 real process design points and 20 computer model design points) and large initial data design (with 20 real process design points and 50 computer model design points). The posterior distributions of the calibration parameter for different set of initial data together with the target posterior are shown in Figure 1b. As expected, the posterior distribution of the calibration parameter with more initial data is closer to the target posterior than with less initial data. Table 2 gives the average value and variance of the estimated parameters. The average RMSPE over 10 macroreplications for different sets of initial data are also provided. Here we observe that with a larger initial design, more accurate model parameter estimates can be obtained, providing for a more accurate empirical Bayes approach in modeling and prediction. The average RMSPE is also computed for the MCMC approach with 0.213 (less initial data) and 0.105 (more initial data). It can be seen that more initial data results in better calibration performance and also results in more accurate estimated parameters. Hence better surrogate model

and better predictive performance can be obtained. When the initial data is small, the calibration accuracy will significantly influence the accuracy of the estimated parameters and also the predictive performance.

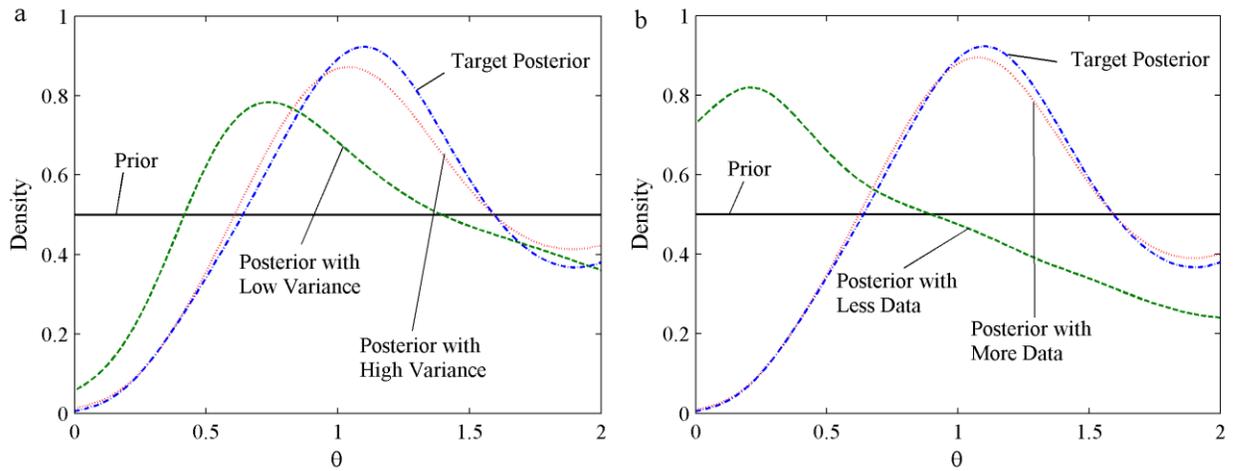


Figure 1. (a) Posterior distributions of the calibration parameter for different variances of the stochastic error together with target posterior. (b) Posterior distributions of the calibration parameter with different number of initial data together with target posterior

Table 1. Mean (variance) of the plug-in parameters estimated for different variance and the target values obtained with sufficient data.

Variance level	ϕ_{S_x}	ϕ_{S_θ}	ϕ_δ	τ_1	τ_2	τ_3	RMSPE
Low	0.625 [†] (0.0122)	1.587 [†] (0.0621)	4.595 [†] (0.6757)	0.152 [†] (7.7×10^{-4})	0.349 [†] (1.8×10^{-4})	0.293 [†] (6.3×10^{-4})	1.487 ^{**}
High	0.826 [*] (0.0164)	4.283 [*] (0.1137)	8.778 [*] (1.0123)	0.351 [*] (8.9×10^{-4})	0.098 [*] (1.9×10^{-4})	0.086 [*] (7.9×10^{-4})	2.364
Target	0.504	1.145	3.512	0.103	0.378	0.342	

^{*}Significantly different from Target at $\alpha=0.05$; [†]Significantly different from (High) at $\alpha=0.05$; ^{**}Significantly smaller than (High) at $\alpha=0.05$

Table 2. Mean (variance) of the plug-in parameters estimated for different number of initial data and the target values obtained with sufficient data.

No. of initial data	ϕ_{S_x}	ϕ_{S_θ}	ϕ_δ	τ_1	τ_2	τ_3	RMSPE
Large	0.525 [†] (0.0122)	1.139 [†] (0.0512)	3.682 [†] (0.8147)	0.109 [†] (6.9×10^{-4})	0.369 [†] (1.3×10^{-4})	0.348 [†] (6.2×10^{-4})	0.109 ^{**}
Small	0.893 [*] (0.0176)	5.843 [*] (0.1472)	7.176 [*] (1.1738)	0.289 [*] (8.3×10^{-4})	0.093 [*] (1.4×10^{-4})	0.087 [*] (7.2×10^{-4})	0.217
Target	0.504	1.145	3.512	0.103	0.378	0.342	

^{*}Significantly different from Target at $\alpha=0.05$; [†]Significantly different from (Small) at $\alpha=0.05$; ^{**}Significantly smaller than (Small) at $\alpha=0.05$

The above results indicate that when the initial set of data is sufficient and/or stochastic noise is small, the empirical Bayes approach (with the partially estimated model parameters) perform reasonably well, and the additional computational time required for a fully Bayesian MCMC approach is not necessary. However, when the initial data is insufficient (not enough simulation replications when the stochas-

tic error is large or the number of the initial data is limited), the accuracy of the calibration parameter and the partially estimated model parameters can be affected in the Bayesian approach. These will further affect the predictive performances of the model. In such cases, the MCMC approach may provide more accurate predictions (with slightly smaller RMSPE values), but the time required to obtain the calibration and prediction results is about 23 times longer.

From these observations, we see that if the set of initial data can be increased and/or noise reduced (by potentially increasing the replications), the model calibration and predictions can be improved without requiring time consuming MCMC analysis. In the next section, we propose a follow-up design criterion that can be used to improve the model and its predictive performance.

4 AN ENTROPY BASED CRITERION

As seen in the previous section, in situations where the results from initial experiments indicate that the data is insufficient (resulting in a poor fit) or the uncertainty in the calibration parameters are still high, further data can be collected to improve on the model before it is applied. Furthermore, due to the dependencies of the parameters in the EM estimation approach, the accuracy of the calibration parameters θ can affect the estimation and use of the model. In this section, we propose an entropy based criterion that balances the selection of computer design points and real process design points in a way that improves on the uncertainty in the calibration parameter. This not only provides for better model estimations (in the EM algorithm) but also provides for more accurate estimates of the calibration parameter in the eventual use of the computer model. Here for the purpose of calibration and subsequent prediction, we assume that both computer model and real process are available for additional sampling (although at different costs) in a follow-up experiment.

4.1 Formulation of Entropy Criterion

Entropy is intended to be a measure of the anticipated uncertainty of a random variable. Several authors (see e.g. Santner et al. (2003); Lindley (1956); Shewry and Wynn (1987)) proposed the use of the expected decrease in entropy given by an experiment as an optimal design criterion to select the values for the design parameters for the experiment. Here, we adopt this idea to select the experiments so as to maximize the gain in information of the calibration parameter. In this approach, the entropy criterion is used to quantify the knowledge about the calibration parameter and find the next design point that best improves the knowledge of the parameter.

The entropy of the calibration parameter θ before the experiment is defined as $H(\theta) = -\int p(\theta) \ln(p(\theta)) d\theta$, where $p(\theta)$ is the prior density function of θ . The conditional entropy of θ after the experiment using design D is defined as $H(\theta|D) = -\int p(\theta|D) \ln(p(\theta|D)) d\theta$. Then, the optimal entropy design can be obtained by maximizing the expected change in entropy $\max_D E[H(\theta) - H(\theta|D)]$. Since $H(\theta)$ is independent of design D , the optimal entropy design is obtained by minimizing the expected conditional entropy

$$\min_D E[H(\theta|D)].$$

As this design is applied as a follow-up after the initial experiment is conducted and d observed, this can be written as

$$\min_D E[H(\theta|D, d)].$$

4.2 Design Sampling

Here, the design decision involves deciding the allocation of the additional resources to the computer runs or the real process runs. In practical applications, the real process runs are generally preferred if they are

available. However, real process runs are usually expensive and can be quite time consuming to obtain. On the other hand, computer model data (although cheaper) does provide more information about the computer model itself. Therefore, a better surrogate model can be obtained. In addition, for the calibration purpose, more computer model data can also result in better calibration parameter values if it can provide for a better surrogate model. In this paper, we use the calibration parameter model, which is subsequently used for calibration and predictions of the system, as a means to determine allocation. The allocation procedure is as follows.

First, before the resource allocation is decided, the cost comparison between a computer model run and a real process run is required. Assume the cost of adding one real process follow-up design point is equal to the cost of adding ω computer model follow up design points.

Second, find the new real process design point and calculate the anticipated optimal entropy value if the additional resource is allocated to the real process. Let $p(\zeta_r|d)$ denote the predictive distribution of the real process output at new design point x_r conditional on the initial data d . Given this predictive distribution and the conditional posterior distribution of θ , the expected conditional entropy given the candidate design point x_r can be obtained by

$$E_{Z_r} [H(\theta|d, x_r, \zeta_r)] = \int p_{Z_r}(\zeta_r|d) \left(- \int_{\Theta} p(\theta|d, x_r, \zeta_r) \ln(p(\theta|d, x_r, \zeta_r)) d\theta \right) d\zeta_r, \quad (2)$$

where the expectation is taken with respect to the predictive random variable Z_r of the real process output ζ_r and Θ is the domain of θ . Then the optimal follow up design point for the real process can be found by

$$x_r^* = \underset{x_r \in X}{\operatorname{arg\,min}} E_{Z_r} [H(\theta|d, x_r, \zeta_r)], \quad (3)$$

where X is the design space of x .

Third, find the ω new computer model design points and calculate the anticipated optimal entropy value if the additional resource is allocated to the computer experiments. Let $D_c^\omega = \{(x_1, t_1), (x_2, t_2), \dots, (x_\omega, t_\omega)\}$ denote the ω new design points and $s_c^\omega = \{s_1, s_2, \dots, s_\omega\}$ denote the simulation outputs for these ω design points. Let $p(s_c^\omega|\bar{y})$ denote the predictive distribution of the simulation outputs at the new design points conditional on the initial observed simulation outputs \bar{y} . Then the expected conditional entropy given the candidate design points D_c^ω can be obtained by

$$E_{S_c^\omega} [H(\theta|d, D_c^\omega, s_c^\omega)] = \int p_{S_c^\omega}(s_c^\omega|\bar{y}) \left(- \int_{\Theta} p(\theta|d, D_c^\omega, s_c^\omega) \ln(p(\theta|d, D_c^\omega, s_c^\omega)) d\theta \right) ds_c^\omega \quad (4)$$

where the expectation is taken with respect to the predictive random vector S_c^ω of the simulation outputs. Then the optimal next evaluation points can be found by

$$D_c^{\omega*} = \underset{D_c^\omega \{x_i \in X, t_i \in \Theta, \text{ for all } i\}}{\operatorname{arg\,min}} E_{S_c^\omega} [H(\theta|d, D_c^\omega, s_c^\omega)]. \quad (5)$$

Finally, compare $E_{S_c^\omega} [H(\theta|d, D_c^{\omega*}, S_c^\omega)]$ and $E_{Z_r} [H(\theta|d, x_r^*, Z_r)]$ to decide the allocation of the resources. If $E_{S_c^\omega} [H(\theta|d, D_c^{\omega*}, S_c^\omega)] < E_{Z_r} [H(\theta|d, x_r^*, Z_r)]$, ω more computer experiments should be conducted. If $E_{S_c^\omega} [H(\theta|d, D_c^{\omega*}, S_c^\omega)] \geq E_{Z_r} [H(\theta|d, x_r^*, Z_r)]$, one more real process data point should be collected.

Here we consider only one additional real process run compared to ω computer model runs. It is easy to extend this to more than one real process run by replacing x_r with more points. It can be further extended to split the resources between the computer model and the real process either through allocating the resources point by point until the budget is exhausted or by comparing the performances of all possible resources allocation combinations. For situations where only computer experiments are available for sequential design, the optimal design points can be obtained using only Equation (5).

To search for the optimal follow-up design points, the expected conditional entropy (Equations 2 or 4) has to be computed, and Monte Carlo integration can be applied to evaluate this. Here, we adopt a discretization approach proposed by Vasicek (1976) for the evaluation of entropy to facilitate the computa-

tions. In addition, the minimization of Equations (3) and (5) can be done using standard optimization functions in MATLAB.

After the follow up design points with the minimized entropy are obtained, the computer model experiments or the real process experiments are run at these points to obtain the new experimental data. These new data are then added to the original data to implement the calibration procedure again. The predictive distribution is similarly updated with the new data.

5 NUMERICAL RESULTS

In this section, we illustrate the use of the proposed calibration criterion with the example from Section 3. In this example, we consider the case where $\sigma^2=100$. We adopt the same surrogate model form and prior settings as in Section 3. To assess its calibration and predictive performance, the proposed entropy criterion is also compared to the Expected Integrated Mean Squared Prediction Error (EIMSPE) criterion. This criterion was previously proposed by Yuan and Ng (2013b), and focused solely on reducing the overall prediction error based on the estimated model.

5.1 Kinetic Model Example

To consider the allocation of additional resources with the entropy based criterion, it is assumed that the cost of adding one real process design point is equal to the cost of adding two computer model design points. 4 initial real process points are evenly sampled between 0 and 3, and an initial 10 point LHD is used for the computer model with $x \in [0, 3]$ and $t \in [0, 2]$, and 100 replications taken at each design point.

Considering an additional 5 real process design points or 10 computer model design points can be selected, Equations (3) and (5) are used to determine the optimal real process design points and computer design points respectively, and compared to determine the optimal entropy design points. This resulted in additional data points collected at the optimal computer model design points. Figure 2 shows the posterior distributions of the calibration parameter for the two different allocation choices and the target posterior distribution. The figure indicates graphically that the posterior distribution with more computer model data is closer to the target posterior. For more accurate comparison, we calculated the average posterior mean, mode and variance of the calibration parameter for both allocation choices and the target posterior. The results are given in Table 3. It shows that the posterior mean and mode obtained with more computer model data are closer to the target posterior mean and mode than with more real process data. The posterior variance with more computer model data is also smaller than with more real process data. Hence more computer model design points are preferred in this example as they result in better calibration performance. It also can be seen that the posterior mean and mode obtained with more real process data are further from the target values than the values obtained with more computer model data. This indicates that for the calibration purpose, more computer model data are preferred when the initial data size is small. One possible reason is that the computer model points design on both the variable input x and the calibration parameter θ . This not only provides more information about the computer model itself with a more accurate surrogate model, but also provides more information about the calibration parameter.

Table 3. Average posterior mean, mode and variance for θ with different allocation choices

Posterior	Initial	With 10 more computer data points	With 5 more real data points	Target
Mean	0.635	1.141	0.596	1.147
Mode	0.182	0.919	0.283	1.096
Variance	0.272	0.204	0.211	

Next the calibration and prediction performance are compared for the entropy based criterion design (M1) and the EIMSPE based criterion design (M2). It is also compared with a single-stage approach (M4). 10 additional computer model design points are selected with the entropy criterion and the EIMSPE

criterion. For the single-stage approach, 4 real process design points are the same with the sequential follow-up approaches while 20 computer model design points are selected using the LHD method.

Figure 3 shows the posterior distribution of the calibration parameter after the experimental designs are applied. The figure shows that the posterior calibration parameter distribution from the entropy based design is closer to the target posterior distribution than the EIMSPE based design. The posterior distributions for both the sequential design approaches are closer to the target posterior distribution than the one-stage approach. For a better comparison, samples of posterior mean, mode and variance are obtained from 10 macroreplications of the design for each design approach. Then two-sample *t*-tests are used to compare the differences between the estimates for the different approaches. The results show that the entropy based criterion has closest posterior mean (1.141) and mode (0.919) to the target mean (1.147) and mode (1.096) and has the smallest posterior variance (0.204). This is followed by the EIMSPE based criterion (with mean 1.025, mode 0.723, variance 0.232). Both the follow-up design approaches perform better than the one-stage approach (with mean 0.937, mode 0.574, variance 0.248). The *t*-test results show that the differences between each pair of approaches are all significant at $\alpha=0.05$. These results indicate that the entropy based criterion provides the best calibration performance. For the predictive performance, the average root mean square prediction error (RMSPE) of the real process over 10 macroreplications of the design and two sample *t*-test are used to compare the differences among the approaches. The results show that the entropy based approach has an average RMSPE of (0.357), and this is significantly smaller than the RMSPE of the EIMSPE based approach (0.455). The values of both the follow-up design approaches are significantly smaller than the single-stage approach (0.503). This indicates that the entropy based sequential approach has the best predictive performance even though the EIMSPE criterion is directly used to reduce the overall prediction error.

To further assess the performance of the different design approaches, we calculate the average values and the variances of the estimated parameters over 10 macroreplications of each complete design. The results together with the target values (more accurate values estimated using a larger set of data) are given in Table 4. The results show that the estimated values obtained from the entropy based sequential design approach are closer to the target values. This indicates that when the initial data is not sufficient to obtain an accurate calibration parameter value, the entropy based design approach is able to provide for a more accurate calibration parameter estimate. This also results in more accurate values of the estimated model parameters, and hence, a more accurate surrogate model is obtained. This also explains why the entropy based sequential approach can provide better predictive performances than the EIMSPE design approach. This shows that the accuracy of the calibration parameter plays an important role in the predictive performance. As such, it is important to obtain accurate calibration parameter value first as it can provide for a more accurate surrogate model which results in better predictive performances.

The results in this example also show that the sequential follow-up design approaches perform better than the single-stage approach in both calibration and prediction. Overall, the entropy based sequential design has the best calibration and prediction performance.

Table 4. Mean (variance) of the plug-in parameters estimated for different sequential approaches and the target values obtained with sufficient data.

Sequential approach	ϕ_{S_x}	ϕ_{S_0}	ϕ_{δ}	τ_1	τ_2	τ_3
Entropy	0.574 (0.0124)	1.069 (0.0531)	4.732 (0.8744)	0.135 (7.3×10^{-4})	0.364 (1.9×10^{-4})	0.357 (6.7×10^{-4})
EIMSPE	0.941 (0.0185)	6.404 (0.1648)	9.903 (1.2063)	0.298 (8.6×10^{-4})	0.114 (1.7×10^{-4})	0.089 (7.5×10^{-4})
Target value	0.507	1.143	3.516	0.106	0.379	0.343

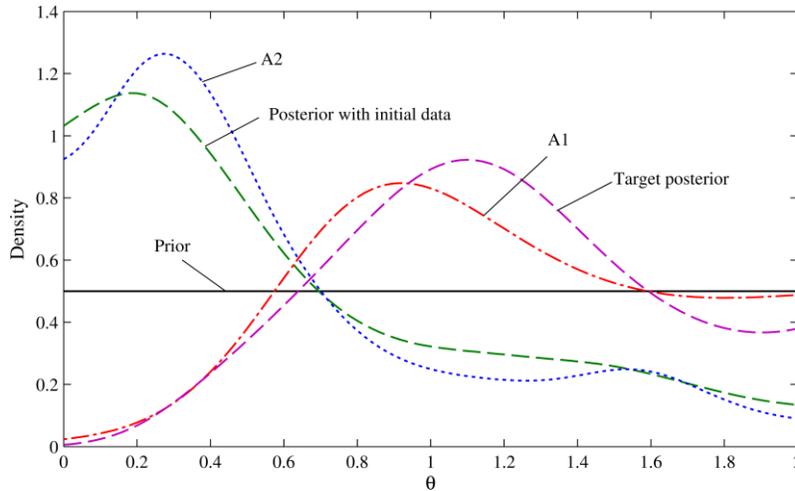


Figure 2. Posterior distributions of calibration parameter for the resources allocation choices with more computer model data (A1) and more real process data (A2) using the entropy.

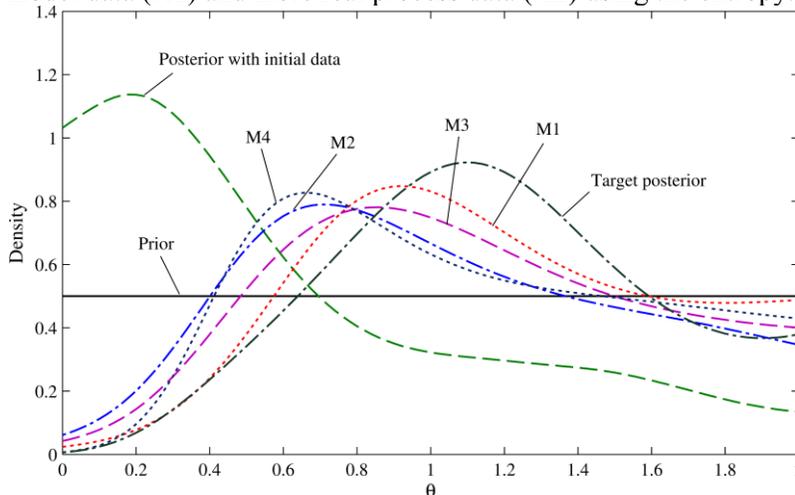


Figure 3. Posterior distribution of calibration parameter for different approaches.

6 CONCLUSION

In this paper, we study the effects of various noise and data scenarios on the model calibration approach developed in Yuan and Ng (2013b). We observe that the calibration and prediction performance of this empirical Bayes approach can be affected when the data size is small and/or noise is large. From these observations, we propose an entropy-based criterion to improve on some of the shortcomings observed in the approximation method. The numerical example illustrates the performance of the entropy based criterion over a EIMSPE based criterion. It shows that the entropy based criterion, applied in a sequential manner can have better calibration and predictive performances. One of the reasons is in the improvement of the calibration parameter, the estimation of some of the model parameters in this empirical Bayes approach also improves. In this criterion proposed, we focused mainly on improving the accuracy of the calibration parameter. A related extension is to improve the criterion to include the other estimated model parameters to further improve on the model accuracy and predictive performance. In addition, further work on developing an adaptive criterion to leverage on the objectives of the entropy criterion and the EIMSPE criterion together in a sequential approach is currently being explored.

ACKNOWLEDGMENTS

This research was partially supported by a research project grant (R-266-000-055-112) funded by the Ministry of Education Academic Research Fund in Singapore.

REFERENCES

- Cho J. J., Y. Ding, Y. Chen, and J. Tang. 2010. "Robust calibration for localization in clustered wireless sensor networks." *IEEE Transactions on Automation Science and Engineering*, 7:81-95.
- Kennedy M. C. and A. O'Hagan. 2001. "Bayesian calibration of computer models." *Journal of the Royal Statistical Society Series B* 63:425-450.
- Lee J. B. and K. Ozbay. 2009. "New calibration methodology for microscopic traffic simulation using enhanced simultaneous perturbation stochastic approximation approach." *Transportation Research Record*. 233-240.
- Lindley, D. V. 1956. "On a Measure of Information Provided by an Experiment." *Annals of Mathematical Statistics* 27:986-1006.
- Loeppky, J., D. Bingham, and W. Welch. 2006. "Computer Model Calibration or Tuning in Practice," Available at: <http://www.stat.ubc.ca/Research/TechReports/techreports/221.pdf>
- Oberkampf, W. L., and C. J. Roy. 2010. *Verification and Validation in Scientific Computing*. Cambridge University Press
- Qian, P. Z. G., and C. F. J. Wu. 2008. "Bayesian Hierarchical Modeling for Integrating Low-Accuracy and High-Accuracy Experiments." *Technometrics* 50:192-204.
- Rutter C. M., D. L. Miglioretti, and J. E. Savarino. 2009. "Bayesian calibration of microsimulation models." *Journal of the American Statistical Association* 104:1338-1350.
- Santner, T. J., B. J. Williams, and W. Notz. 2003. *The Design and Analysis of Computer Experiments*. New York: Springer.
- Shewry, M. C., and H. P. Wynn. 1987. "Maximum Entropy Sampling," *Journal of Applied Statistics* 14:165-170.
- Trucano, T. G., L. Swiler P., T. Igusa, W. L. Oberkampf, and M. Pilch. 2006. "Calibration, validation, and sensitivity analysis: What's what." *Reliability Engineering & System Safety* 91:1331-1357.
- Vasicek, A. 1976. "A Test for Normality Based on Sample Entropy." *Journal of the Royal Statistical Society Series B* 38:54-59.
- Yuan, J., S. H. Ng, and K. L. Tsui. 2013a. "Calibration of Stochastic Computer Models Using Stochastic Approximation Methods." *IEEE Transactions on Automation Science and Engineering* 10:171-186.
- Yuan, J., and S. H. Ng. 2013b. "A Sequential Approach for Stochastic Computer Model Calibration and Prediction." *Reliability Engineering & System Safety* 111:273-286.

AUTHOR BIOGRAPHIES

JUN YUAN received a B. E. degree in Industrial Engineering and Logistics Management from Shanghai Jiao Tong University, in 2008. He is currently a PhD candidate in the Department of Industrial and Systems Engineering at the National University of Singapore. His research interests are in the area of computer model calibration. His email address is jyuan@nus.edu.sg.

SZU HUI NG is an Associate Professor and Deputy Head (Undergraduate Studies) in the Department of Industrial and Systems Engineering at the National University of Singapore. She holds B.S., M.S. and Ph.D. degrees in Industrial and Operations Engineering from the University of Michigan. Her research interests include computer simulation modeling and analysis, design of experiments and quality and reliability engineering. She is a member of IEEE and INFORMS, and a senior member of IIE. Her email address is isensh@nus.edu.sg and her web page is <http://www.ise.nus.edu.sg/staff/ngsh/index.html>.