Bayesian Calibration of Stochastic Computer Models

Jun Yuan, S. H. Ng
Department of Industrial and Systems Engineering, National University of Singapore, Singapore
(jyuan@nus.edu.sg, isensh@nus.edu.sg)

Computer models are widely used to simulate real processes. Within the computer model, there always exist some parameters which are unobservable in the real process but need to be specified in the computer model. The procedure to adjust these unknown parameters in order to fit the model to observed data and improve its predictive capability is known as calibration. In traditional calibration, once the optimal calibration parameter set is obtained, it is treated as known for future prediction. Calibration parameter uncertainty introduced from estimation is not accounted for. We will present a Bayesian calibration approach for stochastic computer models. We account for these additional uncertainties and derive the predictive distribution for the real process. Two numerical examples are used to illustrate the accuracy of the proposed method.

Keywords – **Bayesian calibration, stochastic computer models, Gaussian process, EM algorithm**

I. INTRODUCTION

Computer models are generally designed to simulate real processes. In the real processes, some parameters may be unknown or unobservable, but they need to be specified in the computer model in order to obtain the simulation outputs. Calibration is a procedure to adjust these unknown input parameters by comparing computer model output with real observed data. This is to ensure that the model is fit to the real process. Then the adjusted model is used for prediction.

However, there are various difficulties in real calibration problems. One significant difficulty is that running large and complicated computer model may be time consuming and computationally expensive [1]. One way to deal with this computational expense is to find an effective and efficient algorithm which can be directly applied with limited data resources, such as the stochastic approximation methods discussed by [2]. Another popular method to overcome this heavy computational expense is to use surrogates-based methods. These surrogate methods use simpler and cheaper models instead of the original large and complex models which make calibration more efficient. [1] discussed the using of Gaussian Process (GP) as a surrogate in the calibration problem. In our discussion, we also employ the GP as a surrogate model for the computer model calibration.

In practical applications, traditional calibration methods try to seek one optimal calibration parameter set and treat them as known and fix for the future prediction. Calibration parameter uncertainty introduced from estimation is not accounted for, and this problem is further aggravated when the computer model is stochastic. [1] proposed a Bayesian calibration approach which can account for all sources of uncertainty, including the calibration parameter estimation uncertainty in the prediction. However, they only consider the deterministic computer models. In most applications, the real systems of interest are often stochastic in nature, which require stochastic computer models to assess the output expectation and its distribution characteristics. Different from deterministic model, simulation outputs from a stochastic model may be different for the same input levels. This stochastic error should be considered in the calibration and prediction so as to improve the calibration accuracy and predictive capability. In our paper, we proposed a Bayesian calibration method for stochastic computer models. We account for the calibration parameter estimation uncertainty in the prediction and derive the predictive distribution for the real process which can be used for future prediction analysis. In the following sections, we first introduce the proposed Bayesian calibration method for stochastic computer model. Then we illustrate the accuracy of our proposed method using two numerical examples.

II. BAYESIAN CALIBRATION

A. Stochastic Model

Based on the model proposed by [1], the relationship between the real observation and simulation output can be represented by $z_i = S(x_i,\theta)+\delta(x_i)+e_i$, where z_i is the i-th real observation at input level x_i, $S(x_i,\theta)$ represents the "true" simulation output at x_i, $\delta(x_i)$ 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. $S(x_i,\theta)$ is treated as deterministic in most studies. However, in many practical applications, the computer models built are random / stochastic. This means that simulations of the input levels give different outputs. We employ the model used in [2]. The relationship between the observed and expectation of the stochastic simulation output is represented by $\bar{y}(x_i,\theta) = S(x_i,\theta)+\varepsilon_i$, where $S(x_i,\theta)$ denotes the expectation or deterministic term 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. ε_i is the stochastic simulation output error. Thus the stochastic model can be written as

$$z_i = \bar{y}(x_i,\theta)-\varepsilon_i+\delta(x_i)+e_i . \quad (1)$$

978-1-4577-0739-1/11/$26.00 ©2011 IEEE

B. Gaussian Process Model

Gaussian process models have been widely used as surrogates of computer models because of their flexibility and convenience [1]. A comprehensive study of Gaussian process is given by [3]. In our discussion, we use Gaussian process as surrogate. For the stochastic model (1), we assume the following:

1) The computer model output $S(x,t)$ is a Gaussian process with mean $\mu_S(x,t)=\mathrm{h}_S(x,t)^T\beta_S$ and covariance function $\sigma_S^2 R_S$, where t denotes the specified input calibration parameter value in the computer model. $\mathrm{h}_S(x,t)=(h_{S,1}(x,t),\ldots,h_{S,k_S}(x,t))^T$ is a vector of k_S functions of x and t, and 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 $\phi_S=(\phi_{S_x,1},\ldots,\phi_{S_x,p},\phi_{S_\theta,1},\ldots,\phi_{S_\theta,q})^T$.

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

$$R_\delta(x_i,x_j)=\prod_{u=1}^{p}\exp\{-\phi_{\delta,u}(x_{i,u}-x_{j,u})^2\},$$

where $\phi_{\delta,u}>0$. We write $\phi_\delta=(\phi_{\delta,1},\ldots,\phi_{\delta,p})^T$.

3) The stochastic error ε has a normal distribution with mean 0 and variance σ_ε^2. 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.

C. Prior Distributions for Parameters

One advantage of using Bayesian calibration is that we can use the prior knowledge about the real process. However, the priors should be selected appropriately as improper priors may lead to improper posteriors [4]. In our discussion, the parameters need to be specified are: $\{\beta_S,\sigma_S^2,\phi_S,\beta_\delta,\sigma_\delta^2,\phi_\delta,\sigma_\varepsilon^2,\sigma_e^2\}$. We assume the following proper priors according to [4]:

$$\beta_S\,|\,\sigma_S^2\sim N(b_S,\sigma_S^2V_S),\ \beta_\delta\,|\,\sigma_\delta^2\sim N(b_\delta,\sigma_\delta^2V_\delta);$$

$$\sigma_S^2\sim \mathrm{IG}(\alpha_S,\gamma_S),\ \sigma_\delta^2\sim \mathrm{IG}(\alpha_\delta,\gamma_\delta),$$
$$\sigma_\varepsilon^2\sim \mathrm{IG}(\alpha_\varepsilon,\gamma_\varepsilon),\ \sigma_e^2\sim \mathrm{IG}(\alpha_e,\gamma_e);$$
$$\phi_{S_x,i}\sim G(a_x,c_x),\ \text{for } i=1,\ldots,p,$$
$$\phi_{S_\theta,i}\sim G(a_\theta,c_\theta),\ \text{for } i=1,\ldots,q,$$
$$\phi_{S_\delta,i}\sim G(a_\delta,c_\delta),\ \text{for } i=1,\ldots,p.$$

Here $N(b,\sigma^2V)$ denotes a multivariate normal distribution with mean vector b and covariance matrix σ^2V. $\mathrm{IG}(\alpha,\gamma)$ denotes an inverse gamma distribution with density function

$$p(w;\alpha,\gamma)=\frac{\gamma^\alpha}{\Gamma(\alpha)}w^{-\alpha-1}e^{-\gamma/w},\ w>0,\alpha>0,\gamma>0.$$

$G(a,c)$ denotes the gamma distribution with density

$$G(w,a,c)=\frac{c^a}{\Gamma(a)}w^{a-1}e^{-cw},\ w>0,\alpha>0,\gamma>0.$$

we further assume that $\{\theta\}$, $\{\beta_S,\sigma_S^2,\phi_S\}$, $\{\beta_\delta,\sigma_\delta^2,\phi_\delta\}$, $\{\sigma_\varepsilon^2\}$ and $\{\sigma_\varepsilon^2\}$ are mutually independent; $\{\beta_S,\sigma_S^2\}$ and $\{\phi_S\}$ are independent; $\{\beta_\delta,\sigma_\delta^2\}$ and $\{\phi_\delta\}$ are independent. Then we can write the prior distribution as

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

D. Data

For the large and complex stochastic simulation models we consider here, the theoretical values of $S(x_i,\theta)$ are not observable or available. Due to the computational requirements, it is also practically impossible to run a large number of simulation replications for the specified inputs x_i and θ to accurately determine it. In reality, one would typically run several simulation replications at each input point and take the average of the replication outputs at each point as an approximate value. Here y_{ij} denotes the jth simulation replication output for the ith input set. It is assumed that there is a total of m replications for each i, then the average $\bar{y}_i(x_i,\theta)=(1/m)\sum_{j=1}^{m}y_{ij}(x_i,\theta)$ is taken as the observed stochastic simulation output for the ith input set. It is also assumed that the simulation observations for each replication are independent. Therefore, the two sets of observable data in the calibration of these stochastic models are:

1) Real observations: $z=(z_1,z_2,\ldots,z_N)'$,
2) Stochastic simulation outputs: $y=(\bar{y}_1,\bar{y}_2,\ldots,\bar{y}_n)'$.

Therefore, the full set of available data for analysis is $d^T=(y^T,z^T)$.

E. Posterior Distribution of Calibration Parameter

In the remaining subsections of this section, we explain the posterior analysis of the calibration problem and give the predictive distribution of the real process output. Before deriving the posterior distribution of the calibration parameter, we first specify some notations.

Let $D_S = \{(x_1,t_1),\ldots,(x_N,t_N)\}$ denotes the set of input points at which the computer model outputs are available. Similarly, let $D_z = \{x_1^*,\ldots x_n^*\}$ denotes the set of input points where the real observations are available. Let $D_z(\theta) = \{(x_1^*,\theta),\ldots(x_n^*,\theta)\}$ denotes the set of input points with the optimal calibration parameter θ. We further let $H_S(D_S) = (h_S(x_1^*,t_1),\ldots,h_S(x_N^*,t_N))^T$, Similarly we have $H_\delta(D_Z)$ and $H_S(D_Z(\theta))$. We define $R_S(D_S,D_S)$ to be the covariance matrix of $\bar{y}$ with (*i*, *j*) element $R_S((x_i,t_i),(x_j,t_j))$. Similarly we define $R_S(D_S,D_z(\theta))$, $R_S(D_z(\theta),D_z(\theta))$, and $R_\delta(D_z,D_z)$. We further define $\tau_1^2 = \sigma_\varepsilon^2/\sigma_S^2$, $\tau_2^2 = \sigma_\delta^2/\sigma_S^2$, and $\tau_3^2 = \sigma_e^2/\sigma_S^2$. For simplicity, we let $\beta = \{\beta_S,\beta_\delta\}$, $\sigma^2 = \{\sigma_S^2,\sigma_\delta^2,\sigma_\varepsilon^2,\sigma_e^2\}$, $\phi = \{\phi_S,\phi_\delta\}$ and $\tau^2 = \{\tau_1^2,\tau_2^2,\tau_3^2\}$.

The full data set *d* is normally distributed given $\{\theta,\beta,\sigma^2,\phi\}$. This will yield the likelihood function and further we can derive the posterior distribution of all the parameters. Following, we give the mean and the variance matrix of *d*.

$$E[d \mid \theta,\beta,\sigma^2,\phi] = \mu_d(\theta) = H(\theta)\beta,$$

where $H(\theta) = \begin{pmatrix} H_S(D_S) & 0 \\ H_S(D_z(\theta)) & H_\delta(D_z) \end{pmatrix}$.

$$Var(d \mid \theta,\beta,\sigma^2,\phi) = V_d(\theta) = \sigma_S^2 \bar{V}_d(\theta)$$
$$= \sigma_S^2 \cdot \begin{bmatrix} R_S(D_S,D_S)+\tau_1^2 I_N & R_S(D_S,D_z(\theta)) \\ R_S(D_z(\theta),D_S) & R_S(D_z(\theta),D_z(\theta))+\tau_2^2 R_\delta(D_z,D_z)+\tau_3^2 I_n \end{bmatrix}$$

With the prior distributions given in subsection *C*, we can derive the full posterior distribution of all the parameters. After integrate out β and σ_S^2, the posterior distribution of the remaining parameters are shown as:

$$p(\theta,\phi,\tau_1^2,\tau_2^2,\tau_3^2 \mid d) \propto p(\theta)p(\phi)(\tau_1^2)^{-\alpha_\varepsilon-1}(\tau_2^2)^{-\alpha_\delta-1}(\tau_3^2)^{-\alpha_e-1}$$
$$\cdot \left|\bar{V}_d(\theta)\right|^{-1/2} \left|V_{S+\delta}\right|^{-1/2} \left|\bar{A}\right|^{1/2} (\gamma_S + \gamma_\varepsilon/\tau_1^2 + \gamma_\delta/\tau_2^2 + \gamma_e/\tau_3^2 + (1/2)$$
$$\cdot (b_{S+\delta}{}^T (V_{S+\delta})^{-1} b_{S+\delta} + d^T \bar{V}_d(\theta)^{-1} d - \bar{\upsilon}^T \bar{A} \bar{\upsilon}))^{-\alpha_S-\alpha_\varepsilon-\alpha_\delta-\alpha_e-3-(N+n)/2}$$

where $V_{S+\delta} = \begin{pmatrix} V_S & 0 \\ 0 & \tau_2 V_\delta \end{pmatrix}$,

$$\bar{A} = \left[H(\theta)^T \bar{V}_d(\theta)^{-1} H(\theta) + (V_{S+\delta})^{-1} \right]^{-1},$$

$$\bar{\upsilon} = \left[H(\theta)^T \bar{V}_d(\theta)^{-1} d + (V_{S+\delta})^{-1} b_{S+\delta} \right].$$

To obtain the posterior distribution $p(\theta \mid d)$ of the calibration parameters, we can further integrate out ϕ, τ_1^2, τ_2^2 and τ_3^2. However, $p(\theta,\phi,\tau_1^2,\tau_2^2,\tau_3^2 \mid d)$ is a highly intractable function and it is computationally intensive to numerically integrate out these terms. Therefore, we estimate these parameters and treat them as fixed. Then the conditional posterior distribution given the estimated parameters is used for inference about θ. Following, we provide a method to estimate these parameters.

F. Parameters Estimation

The parameters are estimated in two steps. In the first step we use data $\bar{y}$ to estimate parameters ϕ_S and τ_1^2. Since data *z* depends on all the parameters and the number of observations in *z* is usually much smaller than the number of observations in $\bar{y}$, little information is lost when we use only $\bar{y}$ in the first step. In the second step we use data *d* to estimate remaining parameters ϕ_δ, τ_2^2 and τ_3^2 given estimated ϕ_S and τ_1^2 from the first step.

Setp1:

Since we have

$$\bar{y} \mid \beta_S,\sigma_S^2,\phi_S,\tau_1^2 \sim N(H_S(D_S)\beta_S, \sigma_S^2(R_S(D_S,D_S)+\tau_1^2)).$$

With the given prior distributions in subsection *C*, we can integrate β and σ_S^2 from the joint posterior distribution $p(\beta_S,\sigma_S^2,\phi_S,\tau_1^2 \mid \bar{y})$ to obtain

$$p(\phi_S,\tau_1^2 \mid \bar{y}) \propto p(\phi_S)(\tau_1^2)^{-\alpha_\varepsilon-1} \left|\bar{V}_S(D_S)\right|^{-1/2} \left|V_S\right|^{-1/2} \left|\bar{A}_S\right|^{1/2}$$
$$\cdot (\gamma_S + \frac{1}{\gamma_S} + \frac{\gamma_\varepsilon}{\tau_1^2} + \frac{1}{2}(b_S{}^T (V_S)^{-1} b_S + \bar{y}^T$$
$$\cdot \bar{V}_S(D_S)^{-1}\bar{y} - \bar{\upsilon}_S{}^T \bar{A}_S \bar{\upsilon}_S))^{-2\alpha_S-\alpha_\varepsilon-N/2}$$

where $\bar{V}_S(D_S) = (R_S(D_S,D_S)+\tau_1^2 I_N)$,

$$\bar{A}_S = \left[H_S(D_S)^T \bar{V}_S(D_S)^{-1} H_S(D_S) + (V_S)^{-1} \right]^{-1},$$

$$\bar{\upsilon}_S = \left[H_S(D_S)^T \bar{V}_S(D_S)^{-1} \bar{y} + (V_S)^{-1} b_S \right].$$

Then we can obtain the estimators of ϕ_S and τ_1^2 by maximizing $p(\phi_S,\tau_1^2 \mid \bar{y})$.

Step2:

In this step, we use the EM algorithm to estimate the remaining parameters ϕ_δ, τ_2^2 and τ_3^2. EM algorithm is a powerful computational method to find the maximizer of the posterior distribution. More details of the EM algorithm refer to [5]. Similarly to step1, we can derive the posterior distribution $p((\phi_\delta,\tau_2^2,\tau_3^2 \mid d,\theta)$ of the parameters given θ and the posterior distribution $p(\theta \mid d,\phi_\delta,\tau_2^2,\tau_3^2)$ of θ given the parameters. To obtain

the maximizer of the posterior $p((\phi_\delta, \tau_2^2, \tau_3^2 \mid d, \theta)$, we provide the following procedure.

1) Given the initial estimators $\phi_{\delta(0)}$, $\tau_{2(0)}^2$, $\tau_{3(0)}^2$,
2) E step: calculate posterior mode $\hat{\theta}$ with respect to $p(\theta \mid d, \phi_{\delta(0)}, \tau_{2(0)}^2, \tau_{3(0)}^2)$,
3) M step: maximize $p((\phi_\delta, \tau_2^2, \tau_3^2 \mid d, \hat{\theta})$ for given $\hat{\theta}$ to obtain the updated $\phi_{\delta(i)}$, $\tau_{2(i)}^2$, $\tau_{3(i)}^2$
4) The procedure terminate if the maximum iteration number reached or the difference between the two sequent estimators are small enough (see [5]). Otherwise go to E step.

The finally obtained estimators are taken as the best estimators of these parameters.

G. Calibration and Prediction

With the estimated parameters $\hat{\phi}$ and $\hat{\tau}$ from the previous subsection, we use the conditional posterior distribution $p(\theta \mid \hat{\phi}, \hat{\tau}, d)$ to make inference about θ. Either posterior mean or posterior mode can be used for the future simulation runs.

Practically, the purpose of calibration is to use the calibrated model for future prediction. Therefore, we derive the predictive distribution for the real process which can be used for prediction analysis.

Let $\zeta(x_0)$ be the real process output at x_0 to be predicted. Given the full set of available data *d*, we have that the real process prediction is a non-central *t* distribution shown by following:

$$[\zeta(x_0) \mid d] \sim T_1(\nu_{\zeta|d}, \mu_{\zeta|d}, \sigma_{\zeta|d}^2), \tag{2}$$

where $\nu_{\zeta|d} = n + N + 2\alpha_S$,

$$\mu_{\zeta|d} = h(x_0,\theta)^T \hat{\beta}_{\zeta|d} + V_\zeta(x_0,\theta)^T (\bar{V}_d(\theta))^{-1}(d - H(\theta)\hat{\beta}_{\zeta|d})$$

$$\sigma_{\zeta|d}^2 = (Q_{\zeta|d}^2 / \nu_{\zeta|d})\left\{1 - (h(x_0,\theta)^T, (1+\tau_2^2)^{-1} V_\zeta(x_0,\theta)^T) \cdot \begin{bmatrix} -V_{S+\delta}^{-1} & H(\theta)^T \\ H(\theta) & (1+\tau_2^2)^{-1}\bar{V}_d(\theta) \end{bmatrix}^{-1} \begin{pmatrix} h(x_0,\theta) \\ (1+\tau_2^2)^{-1} V_\zeta(x_0,\theta) \end{pmatrix}\right\}$$

$$\hat{\beta}_{\zeta|d} = (A_{\zeta|d} + V_{S+\delta}^{-1})^{-1}(B_{\zeta|d} d + V_{S+\delta}^{-1} b_{S+\delta})$$

$$A_{\zeta|d} = H(\theta)^T ((1+\tau_2^2)^{-1}\bar{V}_d(\theta))^{-1} H(\theta)$$

$$B_{\zeta|d} = H(\theta)^T ((1+\tau_2^2)^{-1}\bar{V}_d(\theta))^{-1}$$

$$Q_{\zeta|\bar{y}}^2 = 2(1+\tau_2^2)\gamma_S + d^T\left[((1+\tau_2^2)^{-1}\bar{V}_d(\theta))^{-1} - B_{\zeta|d}^T A_{\zeta|d}^{-1} B_{\zeta|d}\right] d + (b_S - A_{\zeta|d}^{-1} B_{\zeta|d} d)^T (V_{S+\delta} + A_{\zeta|d}^{-1})^{-1} (b_S - A_{\zeta|d}^{-1} B_{\zeta|d} d)$$

With the derived posterior distribution of the calibration parameter θ and (2), it is straightforward to assess the prediction uncertainty and quantify the effects of the calibration parameter uncertainty on the prediction uncertainty.

III. EXAMPLES

In this section, two examples are used to illustrate the calibration and prediction accuracy of our proposed method. The Gaussian process models of both examples are assumed to have a constant mean which is reasonable in many practical applications [3]. According to [4], 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.

A. The Simple Quadratic Example

We first use a simple quadratic function to illustrate the proposed method. The real process model is assumed to be $z = \theta^* x^2 + e$ with $e \sim N(0,4)$. The computer model is coded as $y = \theta x^2 + \varepsilon$. Here the discrepancy term is set to be 0 so that the influence of the stochastic error on the predictive performance can be clearly identified. We further assume that the true value of θ^* is 5, and θ has a uniform prior distribution between 3 and 7. For the real model, 11 experimental input points are evenly collected from $x \in [-3, 3]$ and one observation for each input point. For the computer model, Latin hypercube design method is used to collect 20 input sets with $x \in [-3, 3]$ and $\theta \in [3, 7]$. Each input set has 100 replications and mean is taken as the simulation output. We first assume a small variance (i.e. 10) for ε to assess the calibration accuracy. Then we compare with the large variance scenario (i.e. 100) to show the predictive capability of our method.

The parameter values estimated by the proposed approach are $\phi_S = 0.1197$, $\tau_1^2 = 0.4318$, $\phi_\delta = 6.8523$, $\tau_2^2 = 0.0762$ and $\tau_3^2 = 0.3124$. The posterior distribution of θ is shown by the left plot of Fig. 1. The posterior mean is 5.0123 and the posterior mode is 5.0282. Both are close to the true optimum $\theta = 5$. 95% confidence interval of θ is (4.6873, 5.3857). The results indicate that our approach can obtain an accurate parameter value.

Then we compare the predictive performance for both high and low variance scenarios. The results are shown in the right plot of Fig 1, which indicate that the proposed method can predict well and the predictive performance will not be significantly influenced by the value of the stochastic variance.

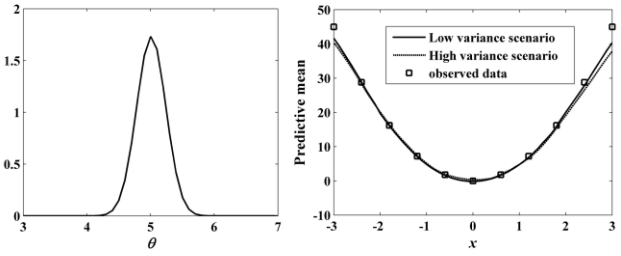

Fig. 1. Posterior distribution of θ (left plot) and predictive mean for both low and high variance scenarios(right plot) in quadratic example.

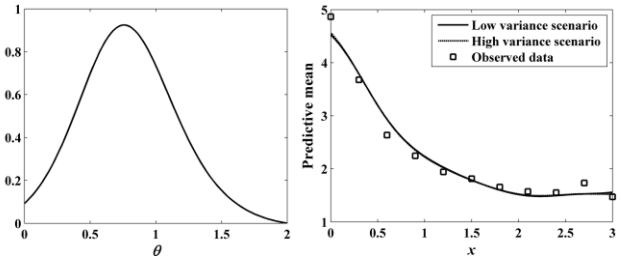

Fig. 2. Posterior distribution of θ (left plot) and predictive mean for both low and high variance scenarios(right plot) in chemical example.

B. Chemical Kinetic Model

The model used in this example is based on [6], which predicts the concentration $y(x)$ of the chemical as a function of time x, with mean response governed by the equation $\phi(x \mid T) = u + r_0 \exp(-Tx)$, 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 $e \sim N(0,0.4)$. The experimental data are collected from 11 input points which are evenly distributed between 0 and 3, and only one observation at each input point. We further suppose that the stochastic computer model is coded as $y = \kappa(x, \theta) = 5\exp(-\theta x)+\varepsilon$, where θ is assumed to have a uniform prior between 0 and 2. Latin hypercube design method is used to collect the initial 20 computer experimental design input set with $x \in [0,3]$ and $\theta \in [0,2]$. For the computer model, each input set has 100 replications and mean is taken as the simulation output. We first assume the variance of ε is 10. Later we compare the predictive performance with the higher variance (i.e. 100) to show the accuracy of the proposed method.

The estimated parameters are $\phi_S = 0.2696$, $\tau_1^2 = 0.0620$, $\phi_\delta = 3.6856$, $\tau_2^2 = 0.3348$ and $\tau_3^2 = 0.2549$. The posterior distribution of θ are given by the left plot of Fig 2. The posterior mean is 0.6387 and the posterior mode is 0.7578. 95% CI of the calibration parameter is (0.2634, 1.2173). The CI in this example is large because that the computer model can never fit well to the real model no matter which calibration parameter value is chosen. This is due to the discrepancy term between the real model and computer model. However, following predictive results indicate that our proposed method can predict well even the computer model itself cannot accurately represent the real model.

We compare the prediction performance of proposed method for both high and low variance scenarios. The results are shown in the right plot of Fig 2. The figure indicates that the proposed method can predict the real process well and the performance will not be significantly influenced by the value of the stochastic variance. The results also show that our method has a good predictive performance even the computer model cannot accurately represent the real model.

IV. DISCUSSION

In this paper, we proposed a Bayesian calibration method for stochastic computer models. We use GP as surrogate models for both computer model and the discrepancy term. We use two steps approach combining with the EM algorithm to estimate the parameters needed to be specified in the calibration procedure. We also derive the predictive distribution for the real process. The results of the two numerical examples indicate that our proposed method can calibrate accurately and the predictive performance will not be significantly influenced by the stochastic error. However, our discussion does not consider the sequential calibration which is important in many real applications. And we treat the stochastic error as nugget effect which has constant variance. This may not be true for many computer models. Therefore, our future work includes the sequential calibration and release the restriction of the stochastic error to make our method wider applicable.

ACKNOWLEDGMENT

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

[1] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statistical Society Series B-Statistical Methodology*, Vol. 63, pp. 425-450, 2001.

[2] J. Yuan, S.H. Ng, and K.L Tsui, "Application of stochastic approximation methods for stochastic computer models calibration," in *Industrial Engineering and Engineering Management (IEEM),* 2010 *IEEE International Conference on*, 2010, pp. 1606-1610.

[3] T. J. Santner, B. J. Williams, and W. I. Notz, *The design and analysis of computer experiments*, New York, Springer, 2003.

[4] P. Z. G. Qian, and C. F. J. Wu, "Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments," *Technometrics*, Vol. 50, pp. 192-204, 2008.

[5] G. C. G. Wei, and M. A. Tanner, "A monte-carlo implementation of the EM algorithm and the poor mans data augmentation algorithms," *Journal of the American Statistical Association*, Vol. 85, pp. 699-704, 1990.

[6] J. L. Loeppky, D. Bingham, and W. J. Welch, "Computer model calibration or tuning in practice," *Working Paper*, 2006.