Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

BAYESIAN SEQUENTIAL CALIBRATION USING DETAILED SAMPLE PATHS

Bo Wang

Industrial and Systems Engineering
Rensselaer Polytechnic Institute
110 8th St.
Troy, NY 12180, USA

Qiong Zhang

Statistical Science and Operations Research
Virginia Commonwealth University
1015 Floyd Ave.
Richmond, VA 23284, USA

Wei Xie

Industrial and Systems Engineering
Rensselaer Polytechnic Institute
110 8th St.
Troy, NY 12180, USA

ABSTRACT

A simplified simulation model is often used to guide the decision-making for a real complex stochastic system. To faithfully assess the mean performance of the real system, it is necessary to calibrate the simulation model. Existing calibration approaches are typically built on the summary statistics of simulation outputs and ignore the dynamic information carried by the detailed sample paths. In this paper, we develop a calibration approach incorporating the detailed output sample paths in a sequential manner. Our theoretical development and empirical study demonstrate that we can efficiently use the simulation resources and achieve better calibration accuracy by exploring the system dynamic behaviors.

1 INTRODUCTION

There is relatively limited literature on calibration in the simulation literature because we tend to believe that the simulation model error has less impact on the optimization decision (Nelson 2016). However, when the simulation is required to make accurate predictions of system behaviors, calibration is necessary. For example, to support the production scheduling for semiconductor manufacturing, faced the orders of new products, the decision maker needs to accurately predict the production and delivery time to sign the contract. For the complex production system, simulation could be used to guide the decision making so that we can maximize the overall profit and still guarantee *on-time delivery*. Thus, it is necessary to calibrate the simulation system so that its output matches well with the real system. *In this paper, we focus on the stochastic simulation calibration for the system steady-state mean performance.*

Kennedy and O'Hagan (2001) introduce a Bayesian calibration framework for the deterministic computer model. They model the unknown response surface of the computer model and the model inadequacy with Gaussian processes. Given the data collected from the real and simulation systems, the posterior distribution of calibration parameters is developed to characterize the belief of the optimal calibration setting, and the posterior predictive distribution is used to quantify the overall prediction uncertainty of the real system response. Built on Kennedy and O'Hagan (2001), various approaches, e.g., Gramacy et al. (2015), Plumlee et al. (2016), Plumlee (2016), Tuo and Wu (2016), and Wong et al. (2016), have been developed for

calibrating deterministic simulation models, and the studies, e.g., Yuan et al. (2013), Yuan and Ng (2013b) and Yuan and Ng (2013a), focus on the stochastic simulation calibration for the mean response.

In the current interconnected world, the decision makers are often faced stochastic systems in large scale. Each simulation run could be computationally expensive. In addition, real-world systems have to evolve rapidly to be competitive. To make sure that the simulation system correctly represents the updated real system, we need to calibrate the simulation model frequently. Thus, we need to efficiently use the data from each simulation run to speed up the calibration process.

It is desired to sequentially allocate the simulation budget so that we can gradually find the promising calibration setting and allocate more simulation runs there. Various EI-type (Expected Improvement) sequential experimental design procedures are introduced in the simulation literature, e.g., Frazier et al. (2009) and Ryzhov (2017).

However, existing stochastic simulation calibration approaches are typically developed based on the summary statistics of the simulation outputs, e.g., the sample mean of customer waiting times in an $M/M/1$ queue, and ignore the dynamic information carried by the detailed sample paths. The recent advances in data storage and computing makes it easy to generate and save the comprehensive data of the output (Nelson 2016). The retained simulation output sample paths can be fully exploited to support the decision making and system diagnostics, such as the simulation analytics, e.g., Plumlee and Lam (2016), Jiang et al. (2016), and Lin and Nelson (2016).

Motivated by recent researches in simulation analytics, we introduce a new calibration approach that fully adopts the detailed sample paths and runs simulations in a sequential manner. Different from existing calibration approaches built on the summary statistics of simulation outputs, our approach can make efficient use of each simulation run. Both theoretical and empirical studies indicate given a finite simulation budget, our approach has better performance than the calibration approach based on the summary statistics.

In sum, the main contributions of our study are as follows.

- We propose a new calibration approach exploring the detailed output sample paths. Gaussian process (GP) is used to capture the dynamic behaviors carried in sample paths.
- Based on the proposed calibration model, we further develop a sequential design of experiments. It could efficiently allocate the simulation resources to important settings of calibration parameters delivering mean response close to that of the real system.
- We provide both theoretical and empirical study to demonstrate that our approach could achieve better calibration accuracy compared to the existing calibration approach built on the summary statistics. The empirical study indicates that our approach is also robust to the GP assumption on the output sample paths.

The remaining of this article is organized as follows. Section 2 provides the problem description and background. Section 3 introduces the calibration approach exploring detailed sample paths. Under the assumption that output sample paths follow GP, we compare our approach with the calibration approach built on the summary statistics. An AR process and an $M/M/1$ queue are used to study the finite sample performance and the robustness of our approach in Section 4. We conclude this paper in Section 5.

2 PROBLEM DESCRIPTION AND BACKGROUND

We consider a complex real system with unknown mean response $\mu^p(x)$, where x is the decision variable. Since each simulation run of the detailed simulation model is computationally prohibitive, a simplified simulation model is often used to guide the decision making. The mean response of the simulation model $\mu(x,\boldsymbol{\theta})$ often depends on the calibration parameters $\boldsymbol{\theta}$. Thus, it is critical to find the optimal calibration setting $\boldsymbol{\theta}^\star$ minimizing the distance between $\mu^p(x)$ and $\mu(x,\boldsymbol{\theta})$ over the entire decision space of x; see more details in Tuo and Wu (2015) and Wong, Storlie, and Lee (2016).

For example, for a complex semiconductor production process involving thousands of steps, given a scheduling policy x, we consider the expected cycle times of different types of products. To guide the decision making, we construct a simplified queueing network as the simulation model. Due to changing product mixes and unexpected break down of tools in the stations, near-bottlenecks may become a bottleneck. Since queue times at bottlenecks and near-bottlenecks dominate the cycle times, we model each bottleneck or near-bottleneck as one queue and aggregate the remaining stations to be a single queue. To correctly guide the decision making for the real system, we first calibrate $\boldsymbol{\theta}$, i.e., parameters specifying the service distributions of queues in the simulation model, so that the expected cycle time estimated from the simulation system matches well with that of the real system. Then, the *calibrated* simulation system could be used to search for the optimal scheduling decision if the product mixes change.

In this paper, we focus on calibrating the simulation model under a fixed decision policy x_0. Thus, the mean performances of the real system and the simulation model can be reduced to μ^p and $\mu(\boldsymbol{\theta})$, respectively. As noted earlier, the goal of calibration is to find the optimal calibration setting $\boldsymbol{\theta}^\star$, which minimizes the distance between μ^p and $\mu(\boldsymbol{\theta})$:

$$\boldsymbol{\theta}^\star \in \arg\min_{\boldsymbol{\theta}\in\Theta} [\mu^p - \mu(\boldsymbol{\theta})]^2, \tag{1}$$

where Θ denotes the space of the calibration parameters. In this paper, Θ contains a finite number of candidates, i.e., $\Theta = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M\}$. Since both μ^p and $\mu(\boldsymbol{\theta})$ are unknown, the optimization problem in (1) can not be solved directly. Alternatively, the optimal calibration setting can be attained by minimizing the expected value of the objective in (1) with regard to our beliefs about $\mu(\boldsymbol{\theta})$ and μ^p:

$$\hat{\boldsymbol{\theta}}^\star \in \arg\min_{\boldsymbol{\theta}\in\Theta} \mathrm{E}\,[\mu^p - \mu(\boldsymbol{\theta})]^2. \tag{2}$$

According to Proposition 2.1 in Ryzhov (2017), we can equivalently assume that μ^p is a fixed constant. Thus, the expectation in (2) is only taken with regard to the uncertainty of $\mu(\boldsymbol{\theta})$. Under this calibration framework, it is critical to accurately assess our belief about $\mu(\boldsymbol{\theta})$ using data collected from the simulation system. Given a tight simulation budget, it is more efficient to collect data in a sequential manner. In each step of the sequential procedure, we choose a calibration setting to obtain simulation outputs, and update our belief about $\mu(\boldsymbol{\theta})$, which guides us to further run simulations at the "promising" candidate calibration settings. There are two key components in this procedure, (1) a surrogate model for our belief about $\mu(\boldsymbol{\theta})$, and an updating scheme to update this surrogate model by incorporating new data, and (2) a criterion to determine the calibration setting for new simulation replications. Classical simulation approaches tend to focus on summary statistics of simulation outputs. Based on the development in Ryzhov (2017), we first introduce these two components using summary statistics of simulation outputs.

2.1 Summary Statistics Approach: Surrogate Model for Mean Performance

Under the calibration setting $\boldsymbol{\theta}$, a single run of the simulation model generates a sample path

$$\mathbf{Y}(\boldsymbol{\theta}) = (Y_1(\boldsymbol{\theta}), \ldots, Y_L(\boldsymbol{\theta}))^\top. \tag{3}$$

In the semiconductor production example mentioned above, $Y_\ell(\boldsymbol{\theta})$ could be the cycle time of the ℓth order. For simplification, we fix the run-length as L, and sequentially allocate simulation runs to the M different calibration settings in Θ.

Classical simulation approaches use the sample mean $\bar{Y}(\boldsymbol{\theta}) = \sum_{\ell=1}^{L} Y_\ell(\boldsymbol{\theta})/L$, and ignore the detailed sample paths. The simulation summary output $\bar{Y}(\boldsymbol{\theta})$ is often assumed to be a sample from a normal distribution with mean $\mu(\boldsymbol{\theta})$ and variance $\sigma^2(\boldsymbol{\theta})$. Under the finite calibration space $\Theta = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_M\}$, we denote $\mu_i \equiv \mu(\boldsymbol{\theta}_i)$ and $\lambda_i^2 \equiv \sigma^2(\boldsymbol{\theta}_i)$. We assume that μ_i and λ_i^2 are unknown, and independent across

different calibration settings. For each $i = 1, \ldots, M$, we use the normal-inverse-gamma distribution to surrogate our beliefs about μ_i and λ_i^2:

$$\mu_i|\lambda_i^2 \sim \mathcal{N}\left(\mu_i^{(0)}, \frac{\lambda_i^2}{\tau_i^{(0)}}\right), \quad \lambda_i^2 \sim \text{InvΓ}\left(a_i^{(0)}, b_i^{(0)}\right). \tag{4}$$

Let $\bar{Y}^{(n)} = \bar{Y}(\boldsymbol{\theta}_{i^{(n)}})$ be the summary statistic obtained at the n-th step, where $i^{(n)}$ is the calibration setting chosen to run simulation in the n-th step. Assume $i^{(n)} = i$. According to the conjugacy property of the normal-inverse-gamma model (Gelman et al. 2014), the parameters of the i-th calibration setting in (4) can be simply updated as follows

$$\begin{aligned} &\mu_i^{(n)} = \mu_i^{(n-1)} + \frac{\bar{Y}^{(n)} - \mu_i^{(n-1)}}{\tau_i^{(n-1)} + 1}, \quad b_i^{(n)} = b_i^{(n-1)} + \frac{\tau_i^{(n-1)}(\bar{Y}^{(n)} - \mu_i^{(n-1)})^2}{2(\tau_i^{(n-1)} + 1)}, \\ &\tau_i^{(n)} = \tau_i^{(n-1)} + 1, \quad a_i^{(n)} = a_i^{(n-1)} + \frac{1}{2}. \end{aligned} \tag{5}$$

Due to the independence assumption across different calibration settings, the parameters of the i-th calibration setting maintain the same as the previous step for $i \neq i^{(n)}$.

After the n-th step, the posterior mean of μ_i is $\mu_i^{(n)}$ as in (5), and the posterior variance of μ_i is

$$\sigma_i^{2,(n)} = \frac{b_i^{(n)}}{\tau_i^{(n)}\left(a_i^{(n)} - 1\right)}. \tag{6}$$

Therefore, based on our beliefs about the mean performance at the n-th step, the estimated optimal calibration setting in (2) can be expressed by

$$\hat{\boldsymbol{\theta}}_s^\star(n) = \underset{\boldsymbol{\theta}_i \in \Theta}{\arg\min} \left\{ \left(\mu_i^{(n)} - \mu^p\right)^2 + \sigma_i^{2,(n)} \right\}. \tag{7}$$

2.2 The Local Time Method for Sequential Design

Ryzhov (2017) introduced *the local time method* to select calibration setting based on our belief about the mean performances. Given $\mu_i^{(n-1)}$ and $\sigma_i^{2,(n-1)}$ as the posterior mean and variance of μ_i at the $(n-1)$-th step, the local time method selects the calibration setting to allocate simulation at the n-th step,

$$i^{(n)} = \underset{\boldsymbol{\theta}_i \in \Theta}{\arg\max} \ell(i) \text{ with } \ell(i) = \sigma_i^{(n-1)} f\left(-\frac{|\mu_i^{(n-1)} - \mu^p|}{\sigma_i^{(n-1)}}\right), \tag{8}$$

where the function $f(z) = z\Phi(z) + \phi(z)$ with ϕ and Φ being the density and the cumulative distribution of the standard normal distribution. The local time criterion measures how frequently μ_i (as a stochastic process based on our belief) visit the target value μ^p over an infinite time horizon. The calibration setting selected by the local time method is potentially closer to the target mean performance μ^p. Ryzhov (2017) demonstrates that the local time method is statistically consistent, and yields asymptotic rates equal to those obtained by the optimal computing budget allocation; see Chen and Lee (2011).

By combining the surrogate model for the mean performance and the local time criterion in (8), the calibration can be done in a sequential manner. However, the summary statistics approach ignores the inherent dependences between the elements in the sample path of the simulation output. To fully explore the sample path information, we propose a new sequential calibration approach in Section 3.

3 DETAILED SAMPLE PATH APPROACH FOR CALIBRATION

We propose a sequential calibration approach using the detailed sample paths of stochastic simulation outputs. Section 3.1 provides the surrogate model, and the model updating formulas for the mean performance, and Section 3.2 gives theoretical comparison of the detailed sample path approach and the summary statistic approach introduced in Section 2.1.

3.1 Surrogate Model for Mean Performance with Detailed Sample Paths

We consider the detailed sample paths generated from *a steady-state simulation system*. For any $\boldsymbol{\theta} \in \Theta$, we assume that the sample path output $\mathbf{Y}(\boldsymbol{\theta})$ in (3) follows a stationary GP with mean $\mathrm{E}[Y_\ell(\boldsymbol{\theta})] = \mu(\boldsymbol{\theta})$ for $\ell = 1, \cdots, L$, and covariance $\mathrm{Cov}[Y_\ell(\boldsymbol{\theta}), Y_{\ell+h}(\boldsymbol{\theta})] = \sigma^2(\boldsymbol{\theta}) R(h;\boldsymbol{\theta})$ for $\ell = 1, \cdots, L-h$, where σ^2 denotes the variance, and $R(\cdot;\boldsymbol{\theta})$ denotes the correlation function the under the calibration setting $\boldsymbol{\theta}$. Throughout this paper, we assume that correlation function $R(\cdot;\boldsymbol{\theta}_i)$ for $i = 1, \ldots, M$ is known. Developing an efficient estimation procedure for the correlation structure is included in the future research. We further denote the correlation matrix for sample path outputs under calibration setting $\boldsymbol{\theta}_i$ to be R_i. Let $\mu_i \equiv \mu(\boldsymbol{\theta}_i)$ and $\sigma_i^2 \equiv \sigma^2(\boldsymbol{\theta}_i)$. Then the detailed simulation output $\mathbf{Y}(\boldsymbol{\theta}_i)$ is a sample from a multivariate normal distribution with a mean vector $\mu_i \mathbf{1}_L$ and covariance matrix $\sigma_i^2 R_i$, where $\mathbf{1}_L$ is L-dimension vector with all entries to be 1. For $i = 1, \ldots, M$, we model our beliefs about the μ_i and σ_i^2 by a normal-inverse-gamma conjugate prior

$$\mu_i | \sigma_i^2 \sim \mathcal{N}\left(\mu_i^{(0)}, \frac{\sigma_i^2}{q_i^{(0)}}\right), \quad \sigma_i^2 \sim \mathrm{Inv}\Gamma\left(\alpha_i^{(0)}, \beta_i^{(0)}\right). \tag{9}$$

Let $\mathbf{Y}^{(n)} = \mathbf{Y}(\boldsymbol{\theta}_{i^{(n)}})$ be the sample path collected at the n-th step, where $i^{(n)}$ is the calibration setting chosen to run simulation in n-th step through the local time method. Based on the conjugacy property of this normal-inverse-gamma model, the updating formulas are provided by Proposition 1; see the derivation in Appendix A.

Proposition 1 Let $i = i^{(n)}$ be the calibration setting selected at the n-th step, the parameters for the i-th calibration setting in (9) can be updated by

$$\begin{aligned}
\mu_i^{(n)} =& \mu_i^{(n-1)} + \frac{\mathbf{1}_L^\top R_i^{-1}(\mathbf{Y}^{(n)} - \mu_i^{(n-1)}\mathbf{1}_L)}{q_i^{(n-1)} + \mathbf{1}_L^\top R_i^{-1}\mathbf{1}_L}, \quad q_i^{(n)} = q_i^{(n-1)} + \mathbf{1}_L^\top R_i^{-1}\mathbf{1}_L, \quad \alpha_i^{(n)} = \alpha_i^{(n-1)} + \frac{L}{2}, \\
\beta_i^{(n)} =& \beta_i^{(n-1)} + \frac{q_i^{(n-1)}(\mathbf{Y}^{(n)} - \mu_i^{(n-1)}\mathbf{1}_L)^\top R_i^{-1}(\mathbf{Y}^{(n)} - \mu_i^{(n-1)}\mathbf{1}_L) + \mathbf{1}_L^\top R_i^{-1}(\mathbf{1}_L(\mathbf{Y}^{(n)})^\top - \mathbf{Y}^{(n)}\mathbf{1}_L^\top)R_i^{-1}\mathbf{Y}^{(n)}}{2(q_i^{(n-1)} + \mathbf{1}_L^\top R_i^{-1}\mathbf{1}_L)}.
\end{aligned} \tag{10}$$

The parameters of the i-th calibration setting maintain the same as the previous step for $i \neq i^{(n)}$. After the n-th step, the posterior mean of μ_i is $\mu_i^{(n)}$ as in (10) and the posterior variance of μ_i is

$$\sigma_i^{2,(n)} = \frac{\beta_i^{(n)}}{q_i^{(n)}(\alpha_i^{(n)} - 1)}. \tag{11}$$

Therefore, based on our beliefs about the mean performance at the n-th step, the estimated optimal calibration setting in (2) can be expressed by

$$\hat{\boldsymbol{\theta}}_d^\star(n) = \underset{\boldsymbol{\theta}_i \in \Theta}{\arg\min} \left\{ \left(\mu_i^{(n)} - \mu^p\right)^2 + \sigma_i^{2,(n)} \right\}. \tag{12}$$

Also, according to the local time method in Ryzhov (2017), we allocate new experiments by choosing the calibration setting that maximizes the local time criterion in (8). Given the simulation budget to be N replications, our sequential mean performance calibration procedure with detailed sample path output is summarized in Algorithm 1.

Algorithm 1: Sequential calibration with detailed sample path

Input: Initial parameters $\mu_i^{(0)}$, $q_i^{(0)}$, $\alpha_i^{(0)}$, and $\beta_i^{(0)}$ for $i = 1, \ldots, M$ in (9).
Output: The estimated optimal calibration setting $\hat{\boldsymbol{\theta}}_d^{\star}(N)$.
for $n \leftarrow 1$ **to** N **do**
  Choose $i^{(n)}$ that maximizes $\ell(i)$ in (8);
  Generate the simulation sample path $\mathbf{Y}^{(n)}$ under the $i^{(n)}$-th calibration setting;
  for $i \leftarrow 1$ **to** M **do**
    if $i = i^{(n)}$ **then**
      Update $\mu_i^{(n)}$, $q_i^{(n)}$, $\alpha_i^{(n)}$, and $\beta_i^{(n)}$ as in Proposition 1.
    end
    else
      Set $\mu_i^{(n)} = \mu_i^{(n-1)}$, $q_i^{(n)} = q_i^{(n-1)}$, $\alpha_i^{(n)} = \alpha_i^{(n-1)}$, and $\beta_i^{(n)} = \beta_i^{(n-1)}$, respectively.
    end
  end
  Compute $\hat{\boldsymbol{\theta}}_d^{\star}(n)$ as in (12).
end

3.2 Theoretical Comparison

This section provides theoretical comparison between the summary statistics approach and the detailed sample path approach. The performances of these two methods are evaluated by the *Probability of Correct Selection* (PCS):

$$\text{PCS}(\hat{\boldsymbol{\theta}}^{\star}) = \Pr(\hat{\boldsymbol{\theta}}^{\star} = \boldsymbol{\theta}^{\star}), \tag{13}$$

where $\hat{\boldsymbol{\theta}}^{\star}$ is the estimated optimal calibration setting in (2), $\boldsymbol{\theta}^{\star}$ is the optimal calibration setting defined in (1), and the probability is taken with regard to the posteriors of mean responses μ_i for $i = 1, \ldots, M$. The larger PCS indicates that the calibration procedure is more likely to select the true best calibration setting $\boldsymbol{\theta}^{\star}$.

The comparison of PCS between the two approaches are provided in Theorem 2. This theorem is developed under the following conditions:

- (**C1**) The variances, λ_i^2 in summary approach, and σ_i^2 in detailed approach are known as a prior.
- (**C2**) The optimization problem in (1) has a unique true optimal point.
- (**C3**) The prior parameters in (4) and (9) are non-informative, with mean equal to 0, and variance equal to ∞.

Theorem 2 Under the conditions (**C1**)–(**C3**), for $\hat{\boldsymbol{\theta}}_d^{\star}(n)$ in (12), and $\hat{\boldsymbol{\theta}}_s^{\star}(n)$ in (7), we have that

$$\text{PCS}\left(\hat{\boldsymbol{\theta}}_d^{\star}(n)\right) \geq \text{PCS}\left(\hat{\boldsymbol{\theta}}_s^{\star}(n)\right),$$

if n is large enough.

This theorem indicates that the mean calibration procedure with detailed sample path outputs outperform the one with summary outputs in terms of asymptotic probability of correct selection. We provide the proof of Theorem 2 in Appendix B. Notice that the posterior development of the proposed method is based on the multivariate normal model assumption of $\mathbf{Y}(\boldsymbol{\theta}_i)$. Although this normality assumption may not hold in general, our empirical study in Section 4 indicates that the proposed approach is robust for non-normal cases.

4 EMPIRICAL STUDY

In this section, we compare the finite sample performances of the detailed sample path approach and the summary statistic approach. The opportunity cost (OC) of the optimization problem in (1) is used to evaluate the performances of these two approaches in additional to PCS. For an estimated optimal calibration setting $\hat{\boldsymbol{\theta}}^\star$, OC can be calculated by

$$\mathrm{OC}(\hat{\boldsymbol{\theta}}^\star) = \left(\mu(\hat{\boldsymbol{\theta}}^\star) - \mu^p\right)^2 - (\mu(\boldsymbol{\theta}^\star) - \mu^p)^2. \tag{14}$$

Since $\boldsymbol{\theta}^\star$ attains the optimal value of $(\mu(\boldsymbol{\theta}) - \mu^p)^2$, $\mathrm{OC}(\hat{\boldsymbol{\theta}}^\star)$ is always non-negative. A smaller $\mathrm{OC}(\hat{\boldsymbol{\theta}}^\star)$ indicates that $\hat{\boldsymbol{\theta}}^\star$ is a better estimator of $\boldsymbol{\theta}^\star$.

For each sequential calibration approach, the prior parameters in (4) and (9) are specified as $\mu_i^{(0)} = 0$, $\tau_i^{(0)} = q_i^{(0)} = 3$, $a_i^{(0)} = \alpha_i^{(0)} = 1.5$ and $b_i^{(0)} = \beta_i^{(0)} = 500$, which ensures that we start from the same non-informative beliefs on the mean performances. The conditional OC and PCS are estimated by using 200 micro-replications. The first order autoregressive model AR(1) and the $M/M/1$ system are used to generate simulation outputs, respectively. The normal assumption of simulation output holds for the AR(1) example, whereas it does not hold for the $M/M/1$ example.

4.1 Example 1: AR(1) Model

We first consider using the AR(1) model to generate the simulation outputs. Assume that the calibration settings in Θ have different mean performances μ_i for $i = 1, \ldots, M$ with the same variance. For the i-th calibration setting in Θ, a realization of the sample path $\mathbf{Y}(\boldsymbol{\theta}_i)$ in (3) is generated from

$$Y_\ell(\boldsymbol{\theta}_i) - \mu_i = \varphi\,(Y_{\ell-1}(\boldsymbol{\theta}_i) - \mu_i) + e_\ell(\boldsymbol{\theta}_i), \tag{15}$$

where φ is the correlation parameter, and $\{e_\ell(\boldsymbol{\theta}_i)\}_{\ell=1}^L$ are white noise from a normal distribution with mean zero and variance σ_e^2. Under this model, the marginal variance $\sigma^2 = \sigma_e^2/(1-\varphi^2)$, and φ is associated with the correlation function $\mathrm{Corr}[Y_\ell(\boldsymbol{\theta}), Y_{\ell+h}(\boldsymbol{\theta})] = \varphi^h$ for $\ell = 1, \cdots, L-h$. Larger $|\varphi|$ leads to stronger serial correlation within a sample path.

The true mean performance of the real system is specified as $\mu^p = 0$. We generate $M = 20$ calibration settings as the candidates in Θ, and the mean performances μ_i are taken as a equally spaced sequence from -10 to 10. Table 1 reports the average OC and empirical PCS of the estimated optimal calibration setting at different steps under different parameter settings of the AR(1) model in (15). To compare performances of the two approaches at each step of the sequential procedure, the average OC and empirical PCS at each step under the parameter setting $\sigma^2 = 25$ and $\varphi = 0.9$ are depicted in the top panel of Figure 1.

4.2 Example 2: $M/M/1$ Queue

We then consider the simulation outputs of the steady-state customer waiting times in an $M/M/1$ queue. In this example, we fix the arrival rate to be 1, and let the service rate as the calibration parameter. We generate $M = 20$ calibration settings with utilization equally spaced on $[0.2, 0.95]$. The sample path under each calibration setting is generated with run-length $L = 50$, after 200 warm-up samples.

The mean performances of three target real systems are generated with the $M/M/1$ queue with utilization equal to 0.5, 0.7, 0.9. Table 2 reports the average OC and empirical PCS of the estimated optimal calibration setting at different steps. To compare performances of the detailed and summary approaches at each step of the sequential procedure, the average OC and empirical PCS with target utilization 0.7 are depicted in the bottom panel of Figure 1.

We summarize the observations from the empirical study. From Figure 1 and Tables 1–2, as the simulation budget N increases, the average OC from both detailed and summary approaches converges to

Table 1: Average OC and empirical PCS of the estimated optimal calibration setting for AR(1).

			summary statistics approach			detailed sample path approach		
σ^2	φ		N=50	N=100	N=500	N=50	N=100	N=500
25	-0.5	OC	0.028	0.027	0.005	0.019	0.013	<0.001
		PCS	0.735	0.745	0.955	0.825	0.880	0.995
	0.2	OC	0.037	0.035	0.020	0.032	0.022	0.013
		PCS	0.650	0.665	0.815	0.695	0.790	0.875
	0.9	OC	0.665	0.283	0.039	0.328	0.106	0.037
		PCS	0.390	0.475	0.630	0.485	0.585	0.655
100	-0.5	OC	0.040	0.034	0.017	0.035	0.030	0.009
		PCS	0.620	0.675	0.840	0.670	0.720	0.920
	0.2	OC	0.230	0.122	0.035	0.087	0.049	0.026
		PCS	0.505	0.545	0.665	0.485	0.540	0.750
	0.9	OC	3.815	2.031	0.126	2.092	0.959	0.099
		PCS	0.160	0.295	0.590	0.250	0.360	0.585

Table 2: Average OC of the estimated optimal calibration setting for the $M/M/1$ queue.

		summary statistics approach			detailed sample path approach		
target utilization		N=50	N=100	N=500	N=50	N=100	N=500
0.9	OC	46.141	32.056	2.024	51.829	35.764	0.976
	PCS	0.460	0.600	0.960	0.390	0.600	0.990
0.7	OC	0.865	0.470	0.021	0.198	0.108	0.010
	PCS	0.275	0.370	0.755	0.400	0.510	0.805
0.5	OC	0.080	0.031	0.002	0.009	0.005	0.001
	PCS	0.270	0.405	0.805	0.495	0.555	0.730

zero, and the empirical PCS converges to one. The detailed sample path approach converges faster than the summary statistics approach. In the $M/M/1$ example, as the target utilization decreases, the average OC decreases. Notice that, for the case with target utilization equal to 0.9, the correlation of lag one is approximate 0.99, and marginal variance is large, which leads inferior performances for both approaches under limited simulation budget.

The benefit of our proposed approach is obvious from asymptotic and finite sample behavior: it converges faster than the corresponding approach using summary statistics. More specifically, the proposed procedure achieves small OC faster, indicating we need less runs to obtain a well-calibrated simulation model. In previous $M/M/1$ queue example with target utilization equal to 0.7, in order to reduce OC to 0.01, using proposed method we need 440 simulation runs on average, whereas in summary statistics approach the number of runs increases to 808. Due to the fact that each simulation run could be computationally expensive, our approach can accelerate the simulation based decision making.

5 CONCLUSION

We propose a new Bayesian sequential calibration approach using the detailed sample paths of stochastic simulation outputs. Compared to the classical calibration approach with summary statistics, our approach has been shown to be effective through theoretical comparison and numerical examples. Now we remark on the directions for future research. First, we will generalize the stationary setting in this paper, and develop a detailed sample path based calibration framework that can be used for non-stationary system

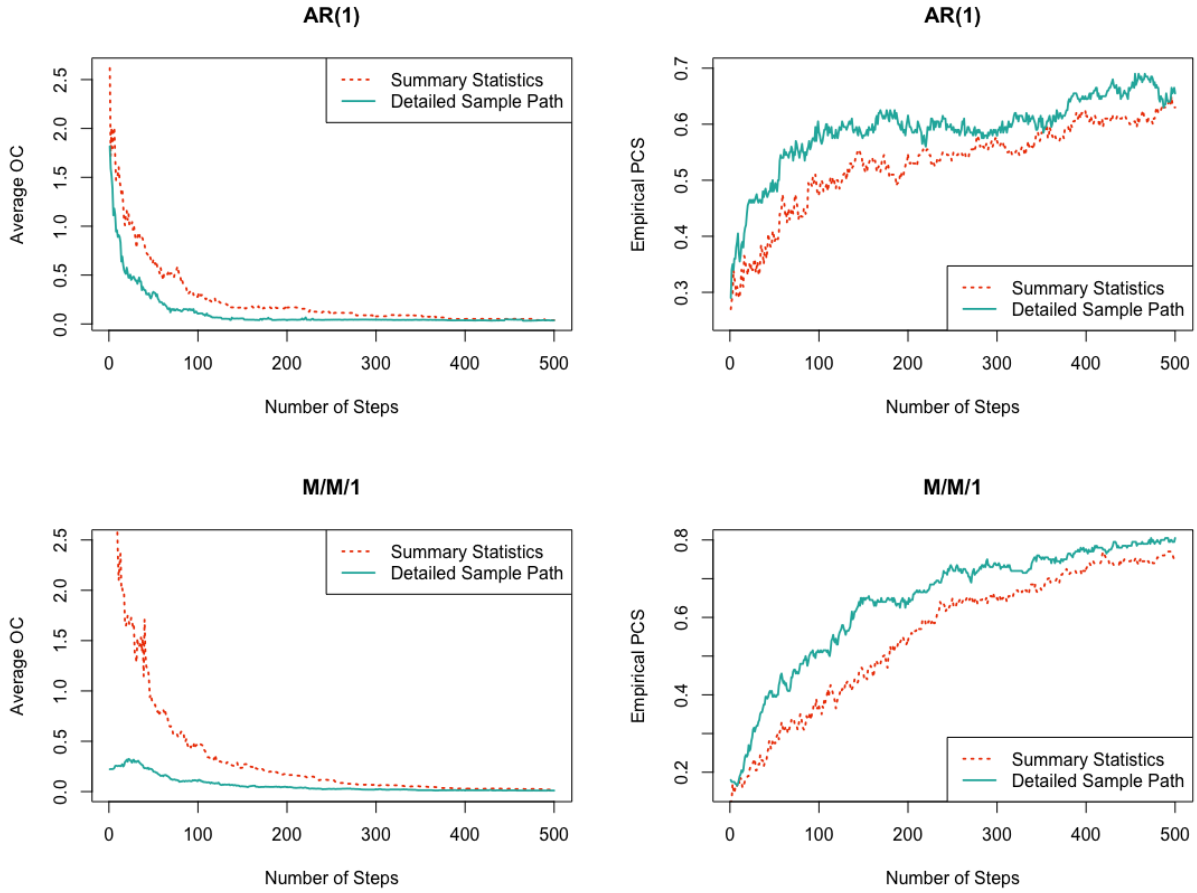

Figure 1: Average OC and empirical PCS at each step of two cases from each of the two examples. Top panel: the AR(1) example with $\sigma^2 = 25$ and $\varphi = 0.9$; Bottom panel: the $M/M/1$ example with target utilization 0.8.

performance. Second, we will develop efficient estimation or sequential learning procedure for correlation structure. Third, we will apply our approach to calibrate complex stochastic systems, such as supply chains and manufacturing systems, for efficient decision making.

A Proof of Proposition 1

Assume the current belief at $\boldsymbol{\theta}_i$ at the beginning of the n-th step:

$$\mu_i|\sigma_i^2 \sim \mathcal{N}\left(\mu_i^{(n-1)}, \frac{\sigma_i^2}{q_i^{(n-1)}}\right), \quad \sigma_i^2 \sim \text{InvGamma}\left(\alpha_i^{(n-1)}, \beta_i^{(n-1)}\right).$$

We select calibration setting i at the n-th step, and get sample path output $\mathbf{Y}^{(n)}$. Combining the prior distribution with multivariate normal distribution $\text{p}(\mathbf{Y}^{(n)}|\mu_i, \sigma_i^2)$, the posterior distributions of μ_i and σ_i^2

are

$$\mathrm{p}(\mu_i|\sigma_i^2,\mathbf{Y}^{(n)}) \propto (\sigma_i^2)^{-\frac{1}{2}} \exp\left[-\frac{q_i^{(n-1)}+\mathbf{1}_L^\top R_i^{-1}\mathbf{1}_L}{2\sigma_i^2}\Big(\mu_i - \frac{\mathbf{1}_L^\top R_i^{-1}(\mathbf{Y}^{(n)}-\mu_i^{(n-1)}\mathbf{1}_L)}{q_i^{(n-1)}+\mathbf{1}_L^\top R_i^{-1}\mathbf{1}_L}\Big)^2\right]$$

$$\mathrm{p}(\sigma_i^2|\mathbf{Y}^{(n)}) \propto (\sigma_i^2)^{-\frac{L}{2}-\alpha_i^{(n-1)}-1} \exp\Bigg[-\frac{1}{\sigma_i^2}\Big(\beta_i^{(n-1)} + \frac{q_i^{(n-1)}(\mathbf{Y}^{(n)}-\mu_i^{(n-1)}\mathbf{1}_L)^\top R_i^{-1}(\mathbf{Y}^{(n)}-\mu_i^{(n-1)}\mathbf{1}_L)}{2(q_i^{(n-1)}+\mathbf{1}_L^\top R_i^{-1}\mathbf{1}_L)}$$
$$+\frac{\mathbf{1}_L^\top R_i^{-1}(\mathbf{1}_L(\mathbf{Y}^{(n)})^\top - \mathbf{Y}^{(n)}\mathbf{1}_L^\top)R_i^{-1}\mathbf{Y}^{(n)}}{2(q_i^{(n-1)}+\mathbf{1}_L^\top R_i^{-1}\mathbf{1}_L)}\Big)\Bigg],$$

which indicates the results in Proposition 1.

B Proof of Theorem 2

Under the conditions (**C1**)–(**C3**), the parameter updating formulas of the summary statistics approach and the detailed summary statistics approach can be reduced to

$$\mu_i^{(N)} = \frac{1}{N_i}\sum_{n=1}^{N} \bar{Y}^{(n)}\mathbb{I}_{\{i^{(n)}=i\}},\ \sigma_i^{2,(N)} = \frac{\lambda_i^2}{N_i} \tag{16}$$

and

$$\mu_i^{(N)} = \frac{\mathbf{1}_L^\top \Lambda_i^{-1}(\sum_{n=1}^{N}\mathbf{Y}^{(n)}\mathbb{I}_{\{i^{(n)}=i\}})}{N_i\mathbf{1}_L^\top\Lambda_i^{-1}\mathbf{1}_L},\ \sigma_i^{2,(N)} = \frac{1}{N_i\mathbf{1}_L^\top\Lambda_i^{-1}\mathbf{1}_L}, \tag{17}$$

where $\Lambda_i = \sigma_i^2 R_i$ is the covariance matrix at $\boldsymbol{\theta}_i$, then the variance of summary output $\lambda_i^2 = (\mathbf{1}_L^\top\Lambda_i\mathbf{1}_L)/L^2$. Let $N_i = \sum_{n=1}^{N}\mathbb{I}_{\{i^{(n)}=i\}}$ be the number of simulations allocated to alternative $\boldsymbol{\theta}_i$ by time N.

Thus, we have $\mu_i^{(N)} \sim \mathcal{N}(\mu_i, \eta_i^2/N_i)$, where η_i^2 represents λ_i^2 under summary approach, $1/(\mathbf{1}_L^\top\Lambda_i^{-1}\mathbf{1}_L)$ under detailed approach, respectively. We assume that $\mu_i \neq \mu^p$ for all $i \in \{1,2,\cdots,M\}$, without loss of generality, we further assume that $\mu^p = 0 < \mu_1 < \mu_2 < \cdots < \mu_M$ (which implies $\boldsymbol{\theta}^\star = \boldsymbol{\theta}_1$). Asymptotic sampling ratio between optimal and other alternatives can be derived as in Theorem 4.3 of Ryzhov (2017),

$$\lim_{N\to\infty}\frac{N_i}{N_1} = \frac{\eta_i^2(\mu_1-\mu^p)^2}{\eta_1^2(\mu_i-\mu^p)^2} = \frac{\eta_i^2 r_i^2}{\eta_1^2 r_1^2},$$

where $0 < r_i = \mu_1/\mu_i < 1$ for $i \neq 1$, and $r_1 = 1$. After simple calculation, we have

$$N_i = \frac{\eta_i^2 r_i^2}{\sum_{k=1}^{M}\eta_k^2 r_k^2}N,\quad \sigma_i^{2,(N)} = \frac{\sum_{k=1}^{M}\eta_k^2 r_k^2}{N r_i^2}. \tag{18}$$

The PCS defined by (13) can be expressed by:

$$\mathrm{PCS} = \Pr\left\{(\mu_1+\sigma_1^{(N)}z_1)^2 + \sigma_1^{2,(N)} < \min_{i\neq 1}\left[(\mu_i+\sigma_i^{(N)}z_i)^2+\sigma_i^{2,(N)}\right]\right\},$$

where z_i's are independent standard normal random variable

$$\mathrm{PCS} = \Pr\left\{\mu_1^2 + 2\mu_1\sigma_1^{(N)}z_1 + O(\frac{1}{N}) < \min_{i\neq 1}\left[\mu_i^2+2\mu_i\sigma_i^{(N)}z_i + O(\frac{1}{N})\right]\right\}$$
$$= \Pr\left\{\mu_1^2+2\mu_1\sigma_1^{(N)}z_1 < \mu_2^2+2\mu_2\sigma_2^{(N)}z_2, \cdots, \mu_1^2+2\mu_1\sigma_1^{(N)}z_1 < \mu_M^2+2\mu_M\sigma_M^{(N)}z_M\right\}.\ (as\ N\to\infty)$$

We define $\mathbf{W} = (W_1, W_2, \cdots, W_{M-1})^\top$ as $W_i = \mu_{i+1}^2 - \mu_1^2 + 2\mu_{i+1}\sigma_{i+1}^{(N)} z_{i+1} - 2\mu_1\sigma_1^{(N)} z_1$ (for $i = 1, 2, \cdots, M-1$), then PCS can be written as

$$\text{PCS} = \Pr(\mathbf{W} > \mathbf{0}) = \Pr(\boldsymbol{\nu}^\top \mathbf{W} > 0 \ for\ all\ \boldsymbol{\nu} \in \mathbb{R}^{M-1},\ \boldsymbol{\nu} > \mathbf{0}).$$

We assume $\boldsymbol{\nu} = (\nu_1, \nu_2, \cdots, \nu_{M-1})^\top > \mathbf{0}$, and from the definition of $\mathbf{W}$. Since $\mathbf{W}$ follows multivariate normal distribution, $\boldsymbol{\nu}^\top \mathbf{W}$ would follow univariate normal

$$\boldsymbol{\nu}^\top \mathbf{W} \sim \mathscr{N}\left(\sum_{i=1}^{M-1} \nu_i(\mu_{i+1}^2 - \mu_1^2),\ 4\mu_1^2\sigma_1^{2,(N)}(\sum_{i=1}^{M-1}\nu_i)^2 + 4\sum_{i=1}^{M-1}\nu_i^2\mu_{i+1}^2\sigma_{i+1}^{2,(N)} \right).$$

And then we get

$$\Pr(\boldsymbol{\nu}^\top \mathbf{W} > 0) = \Phi\left(\frac{\sum_{i=1}^{M-1}\nu_i(\mu_{i+1}^2 - \mu_1^2)}{2\sqrt{\mu_1^2\sigma_1^{2,(N)}(\sum_{i=1}^{M-1}\nu_i)^2 + \sum_{i=1}^{M-1}\nu_i^2\mu_{i+1}^2\sigma_{i+1}^{2,(N)}}} \right).$$

So now we only need to prove the posterior variance given by detailed data approach is less than or equal to posterior variance by summary statistics. Further from (18), we only need to show that $1/(\mathbf{1}_L^\top \Lambda_i^{-1}\mathbf{1}_L) \le \lambda_i^2$ for all $i \in \{1, 2, \cdots, M\}$.

Since Λ_i is covariance matrix, let $\Lambda_i = \mathbf{U}\mathbf{D}\mathbf{U}^\top$, where $\mathbf{D} = diag(\xi_1, \xi_2, \cdots, \xi_L)$ ($\xi_\ell > 0\ for\ \ell = 1, 2, \cdots, L$) and $\mathbf{U}\mathbf{U}^\top = \mathbb{I}_L$. We have the inverse $\Lambda_i^{-1} = \mathbf{U}\mathbf{D}^{-1}\mathbf{U}^\top$ and $\mathbf{D}^{-1} = diag(1/\xi_1, 1/\xi_2, \cdots, 1/\xi_L)$. Define $\mathbf{U}^\top\mathbf{1}_L \triangleq \mathbf{u} = (u_1, u_2, \cdots, u_L)^\top$, then

$$\mathbf{1}_L^\top \Lambda_i^{-1}\mathbf{1}_L = \mathbf{1}_L^\top \mathbf{U}\mathbf{D}^{-1}\mathbf{U}^\top\mathbf{1}_L = \mathbf{u}^\top\mathbf{D}^{-1}\mathbf{u} = \sum_{\ell=1}^{L}\frac{u_\ell^2}{\xi_\ell},\quad \mathbf{1}_L^\top\Lambda_i\mathbf{1}_L = \mathbf{1}_L^\top\mathbf{U}\mathbf{D}\mathbf{U}^\top\mathbf{1}_L = \mathbf{u}^\top\mathbf{D}\mathbf{u} = \sum_{\ell=1}^{L}u_\ell^2\xi_\ell.$$

According to *Cauchy-Schwarz* inequality,

$$(\mathbf{1}_L^\top \Lambda_i^{-1}\mathbf{1}_L)(\mathbf{1}_L^\top\Lambda_i\mathbf{1}_L) = (\sum_{\ell=1}^{L}\frac{u_\ell^2}{\xi_\ell})(\sum_{\ell=1}^{L}u_\ell^2\xi_\ell) \ge (\sum_{\ell=1}^{L}u_\ell^2)^2 = (\mathbf{u}^\top\mathbf{u})^2 = (\mathbf{1}_L^\top\mathbf{U}\mathbf{U}^\top\mathbf{1}_L)^2 = L^2.$$

Recall that $\lambda_i^2 = (\mathbf{1}_L^\top\Lambda_i\mathbf{1}_L)/L^2$, consequently, $1/(\mathbf{1}_L^\top\Lambda_i^{-1}\mathbf{1}_L) \le \lambda_i^2$ for all $i \in \{1, 2, \cdots, M\}$, which indicates the conclusion.

REFERENCES

Chen, C.-h., and L. H. Lee. 2011. *Stochastic simulation optimization: an optimal computing budget allocation*, Volume 1. World scientific.

Frazier, P., W. B. Powell, and H. P. Simaão. 2009. "Simulation model calibration with correlated knowledge-gradients". In *Proceedings of the 2009 Winter Simulation Conference*, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 339–351. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2014. *Bayesian data analysis*, Volume 2. Chapman & Hall/CRC Boca Raton, FL, USA.

Gramacy, R. B., D. Bingham, J. P. Holloway, M. J. Grosskopf, C. C. Kuranz, E. Rutter, M. Trantham, R. P. Drake et al. 2015. "Calibrating a large computer experiment simulating radiative shock hydrodynamics". *The Annals of Applied Statistics* 9 (3): 1141–1168.

Jiang, G., L. J. Hong, and B. L. Nelson. 2016. "A simulation analytics approach to dynamic risk monitoring". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 437–447. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kennedy, M. C., and A. O'Hagan. 2001. "Bayesian calibration of computer models". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (3): 425–464.

Lin, Y., and B. L. Nelson. 2016. "Simulation analytics for virtual statistics via k nearest neighbors". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 448–459. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Nelson, B. L. 2016. "Some Tactical Problems in Digital Simulation for the Next 10 Years". *Journal of Simulation* 10:2–11.

Plumlee, M. 2016. "Bayesian calibration of inexact computer models". *Journal of the American Statistical Association* (just-accepted).

Plumlee, M., V. R. Joseph, and H. Yang. 2016. "Calibrating functional parameters in the ion channel models of cardiac cells". *Journal of the American Statistical Association* 111 (514): 500–509.

Plumlee, M., and H. Lam. 2016. "Fourier trajectory analysis for identifying system congestion". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 401–412. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Ilya O. Ryzhov 2017. "The local time method for targeting and selection".

Tuo, R., and C. J. Wu. 2015. "Efficient calibration for imperfect computer models". *The Annals of Statistics* 43 (6): 2331–2352.

Tuo, R., and C. J. Wu. 2016. "A theoretical framework for calibration in computer models: parametrization, estimation and convergence properties". *SIAM/ASA Journal on Uncertainty Quantification* 4 (1): 767–795.

Wong, R. K., C. B. Storlie, and T. Lee. 2016. "A frequentist approach to computer model calibration". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Yuan, J., and S. H. Ng. 2013a. "An entropy based sequential calibration approach for stochastic computer models". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 589–600. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Yuan, J., and S. H. Ng. 2013b. "A sequential approach for stochastic computer model calibration and prediction". *Reliability Engineering & System Safety* 111:273–286.

Yuan, J., S. H. Ng, and K. L. Tsui. 2013. "Calibration of stochastic computer models using stochastic approximation methods". *IEEE Transactions on Automation Science and Engineering* 10 (1): 171–186.

AUTHOR BIOGRAPHIES

BO WANG is a Ph.D. candidate in the Department of Industrial and Systems Engineering at Rensselaer Polytechnic Institute. His research interests are calibration of stochastic simulation and data analytics. His email address is wangb13@rpi.edu.

QIONG ZHANG is an Assistant Professor of statistics at Virginia Commonwealth University, Richmond, VA. She holds Ph.D. degree in statistics from University of Wisconsin-Madison. Her research interests include computer experiments, uncertainty quantification and spatial and spatial-temporal modeling. She is a member of ASA and INFORMS. Her email address is qzhang4@vcu.edu.

WEI XIE is an assistant professor in the Department of Industrial and Systems Engineering at Rensselaer Polytechnic Institute. She received her M.S. and Ph.D. in Industrial Engineering and Management Sciences at Northwestern University. Her research interests are in computer simulation, risk management and data analytics. Her email address is xiew3@rpi.edu and her web page is http://homepages.rpi.edu/~xiew3/.