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

A STOCHASTIC SIMULATION CALIBRATION FRAMEWORK FOR REAL-TIME SYSTEM CONTROL

Wei Xie, Pu Zhang

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

Qiong Zhang

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

ABSTRACT

A stochastic simulation model is often used to guide decision making for a complex real system, such as scheduling decisions for semiconductor production. To provide a reliable guidance, we propose a simulation calibration framework. We first develop a spatial-temporal metamodel to estimate the system dynamic behaviors at different settings of calibration parameters. Then, assisted by the metamodel, we introduce a calibration model so that the dynamic behaviors of the calibrated simulation model match with those of the real system. Thus, for any feasible decisions, the calibrated simulation model can predict the future outputs for the real system and deliver prediction intervals.

1 INTRODUCTION

Stochastic simulation is often used to guide decision making for complex real systems. Classical output analysis in the simulation literature tends to focus on predetermined summary performance measures, such as the expected production cycle time. They are not suitable for *the system control* which requires the prediction of future outputs for any feasible decisions. Future outputs depend on the current states and the system dynamic behaviors. For example, in the production control, given the current works-in-process (WIP) at the key working stations and the cycle times of past orders, the decision maker wants to find optimal scheduling decisions that can efficiently use the production resource and guarantee the on-time delivery for next orders. *Notice that the dynamic behaviors of the real system are critically important to predict the future outputs and further support the system control.*

However, the historical data collected from the real system are under certain restricted decision(s). To guide the real-time decision making, simulation can be used to answer the question, "how will the real system behave if?" For example, if we change to a new scheduling policy, could we finish the next orders on-time and what is the production cost? For a complex real system, a simplified computer model is often used. To deliver reliable decision guidance, we need to first calibrate the simulation model so that under the same decision policy, the *calibrated* simulation model has the dynamic behaviors matching

with the historical output data collected from the real system. Then, for any feasible decision, by running simulations starting from the current states, we can predict future outputs for the real system.

For deterministic simulation, Kennedy and O'Hagan (2001) proposed a Bayesian calibration framework. Gaussian processes (GP) are used to model the simulation outputs and the model discrepancy. Then, given the data collected from real and simulation systems, they use the posterior distribution and the posterior predictive distribution to quantify the estimation uncertainty of calibration parameters and the prediction uncertainty of the real system output, respectively. Most recent studies on calibration are built on this framework. Based on the methodologies quantifying the uncertainty, they can be divided into Bayesian approaches, e.g., Gramacy et al. (2015), Plumlee, Joseph, and Yang (2016), Plumlee (2016), and frequentist approaches, e.g., Tuo and Wu (2015b), Tuo and Wu (2015a), Wong, Storlie, and Lee (2016).

There are limited studies on calibration in the stochastic simulation literature because we tend to believe that the simulation model error has less impact on the optimization decision which is based on *relative* performance (Nelson 2016). In addition, these studies typically focus on the system mean performance; see for example Yuan, Ng, and Tsui (2013), Yuan and Ng (2013), Jun and Ng (2013). However, calibration becomes critically important for the system control when simulation is used to assess risk and make accurate predictions of system future behaviors (Nelson 2016).

A systematic approach to calibrate the stochastic simulation for the system control is open. *In this paper, we propose a calibration framework to support real-time system control.* Specifically, we first develop a spatial-temporal metamodel to model the dynamic behaviors of simulation output sample paths at different settings of calibration parameters. Then, under the assistance of the metamodel, we introduce a new calibration model so that the calibrated simulation system can capture the dynamic behaviors of the real system. Our framework can deliver credible intervals for calibration parameters and prediction intervals for future outputs of the real system. As a result, the calibrated simulation model can be used to provide a reliable guide for the real-time decision making.

In the next section, we use a semiconductor production control as an illustrative example to describe the problem of interest. We then briefly describe the proposed calibration framework. Since the queueing theory is often used to guide the decision making in production processes, an $M/M/1$ queue is studied to motivate the construction of a spatial-temporal metamodel capturing the system dynamic behaviors at different settings of calibration parameters in Section 3. We introduce a new calibration model in Section 4, study the finite sample performance of our approach in Section 5, and conclude this paper in Section 6.

2 PROBLEM DESCRIPTION AND PROPOSED RESEARCH

We use the production control in the semiconductor wafer manufacturing as an illustrative example. Since the production lines are capital-intensive, the decision maker wants to find real-time scheduling decisions that can efficiently utilize the facility and guarantee the on-time delivery. Considering that production processes involve thousands of steps and they are subject to unpredictable disruption, such as breakdowns of key equipments, a simplified simulation model is typically used to guide the production scheduling (Horiguchi et al. 2001). For example, since a few work stations with either expensive or unreliable equipments tend to dominate the flow of orders, the production process can be simplified by modeling each bottleneck or near-bottleneck work station as an $M/M/1$ queue and aggregating the remaining stations.

Denote the time index for the last finished order by T. We observe the historical data of cycle times collected from the real system, denoted by $\mathbf{y}^r_{[T]} = (y^r_1, \ldots, y^r_T)$, which could be collected under a set of scheduling decision settings. In this paper, for simplification, we assume that the historical data are collected under a fixed decision x_0. Suppose that we do not very frequently change the scheduling decisions (Horiguchi et al. 2001). Let S_T denote the current system state, e.g., WIP at important work stations and our belief of further machine breakdowns. To support the real-time scheduling decision making, we need to predict the cycle times of next orders $Y^r_{T+1}(x, S_T), Y^r_{T+2}(x, S_T), \ldots$ for any feasible decision x. For simplification, we only consider the next order cycle time $Y^r_{T+1}(x, S_T)$. *Thus, we are interested in the prediction distribution for* $Y^r_{T+1}(x, S_T)|\mathbf{y}^r_{[T]}$.

For any feasible x in the decision space, the simplified simulation model starting from the current state S_T could be used to estimate the predictive distribution for $Y^r_{T+1}(x, S_T)$. To correctly predict the future output of the real system, we need to first calibrate the simulation model so that the dynamic behaviors of the *calibrated* simulation model match with those of the real system in the entire decision space. We denote the calibration parameters by $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^\top$ and represent the simulation output sample path with runlength L by $\mathbf{Y}^s(x, \boldsymbol{\theta}) = (Y^s_1(x, \boldsymbol{\theta}), \ldots, Y^s_L(x, \boldsymbol{\theta}))$. In the simplified simulation model for the production control example mentioned above, the calibration parameters could be the service rates for $M/M/1$ queues representing bottleneck and near-bottleneck work stations and the outputs could be the order cycle times. Thus, we adjust $\boldsymbol{\theta}$ so that the dynamic behaviors of the output sample paths from the *calibrated* simulation model match with the historical data $\mathbf{y}^r_{[T]}$ under the same decision.

To characterize the dynamic behaviors of the simulation and real systems, we treat the real system outputs as a stochastic process with the decision variable x as the input, while the simulation outputs can be treated as a stochastic process with both decision variable x and calibration parameters $\boldsymbol{\theta}$ as the inputs. Thus, the prediction of $Y^r_{T+1}(x, S_T)$ using both simulation and real outputs is directly related to the output covariance structure of both stochastic processes. For computational and conceptual convenience,

(1) Suppose the correlation of simulation outputs is separable: $\mathrm{cor}(Y^s_i(x, \boldsymbol{\theta}), Y^s_j(x', \boldsymbol{\theta}')) = c_1(i, j) c_2(x, x') c_3(\boldsymbol{\theta}, \boldsymbol{\theta}')$, where c_1, c_2, and c_3 denote correlation functions. It implies that for fixed $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, $\mathrm{cov}(Y^s_i(x, \boldsymbol{\theta}), Y^s_j(x, \boldsymbol{\theta}')) \propto \mathrm{cor}(Y^s_i(x', \boldsymbol{\theta}), Y^s_j(x', \boldsymbol{\theta}'))$ holds for any feasible decisions x and x'.

(2) Suppose the correlation of real system outputs is separable: $\mathrm{cor}(Y^r_i(x), Y^r_j(x')) = c_1(i, j) c_2(x, x')$. It implies that for fixed $\boldsymbol{\theta}$, we have $\mathrm{cov}(Y^r_i(x), Y^r_j(x')) \propto \mathrm{cor}(Y^s_i(x, \boldsymbol{\theta}), Y^s_j(x', \boldsymbol{\theta}))$.

Under these assumptions, we ignore the interaction of decision variable and calibration parameters in the covariance structure of the simulation outputs. Thus, the simulation model calibrated at x_0 can be used to estimate the dynamic behaviors of the real system at any other feasible decision x.

In this paper, we propose a calibration framework for the system control. Based on the sample paths of simulation outputs, we first develop a spatial-temporal metamodel to capture the dynamic behaviors of the simulation model at different settings of calibration parameters $\boldsymbol{\theta}$ in Section 3. Then, under the assistance of the metamodel, we introduce a new calibration model to match the dynamic behaviors from real and simulation systems so that the calibrated simulation model can provide a reliable prediction of future outputs for the real system in Section 4. Since we focus on calibrating the simulation model at fixed decision x_0 here, for notational simplification, we drop the decision variable from the expressions of system outputs in the remaining paper.

3 A Spatial-Temporal Metamodel for Simulation System Dynamic Behaviors

When we calibrate the simulation model to capture the dynamic behaviors in the real data $\mathbf{y}^r_{[T]}$, we need to estimate the dynamic behaviors of simulation outputs at different settings of calibration parameters. Since each simulation run could be computationally expensive, we introduce a spatial-temporal metamodel to assist the calibration, which could efficiently use the simulation resources (Xie, Nelson, and Barton 2014). At any calibration parameters $\boldsymbol{\theta}$, let $(Y^s_{j,1}(\boldsymbol{\theta}), \ldots, Y^s_{j,L}(\boldsymbol{\theta}))$ represent the simulation outputs from the jth replication. There are two types of dependence in the simulation outputs: (1) the serial dependence of simulation outputs $Y^s_{j,i}(\boldsymbol{\theta})$ and $Y^s_{j,i'}(\boldsymbol{\theta})$ in the same sample path, and (2) the dependence of dynamic behaviors at different calibration settings $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$. Our metamodel can capture both types of dependence. Specifically, we first construct an autoregressive (AR) process to model the serial dependence of sample paths from simulation outputs $\{Y^s_{j,i}(\boldsymbol{\theta})\}$ in Section 3.1. Since the system dynamic behaviors are similar when the calibration parameters are close, we model the spatial dependence of the coefficients of the AR process with GP in Section 3.2.

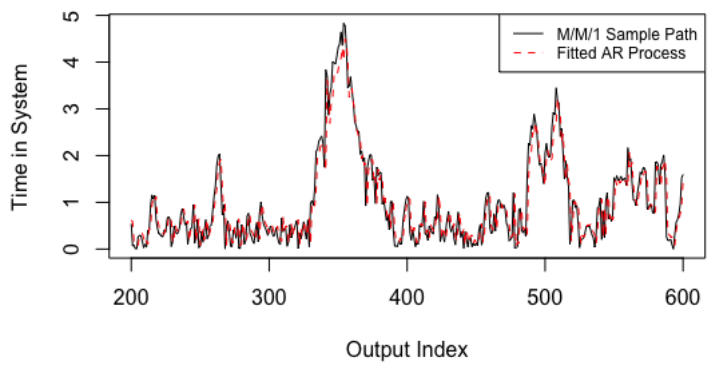


Figure 1: Real and fitted sample paths for outputs from an $M/M/1$ queue.

3.1 AR Model for Time Series Dependence in the Sample Paths

In this section, we propose to use an AR process to model sample paths of simulation outputs. Compared to classical output analysis that only investigates the summary statistics of simulation outputs, e.g., sample mean, our approach can capture the important dynamic properties and improve the prediction of future outputs. The queueing models are often used to study the performance of production systems, and they are also the state-of-the-art technique for many other applications, e.g., the inventory control. Here, we study the outputs of an $M/M/1$ queue with representative results shown in Figure 1. The solid line represents a single sample path of the time customer staying in the $M/M/1$ system. Then, given the simulation outputs, we fit an AR process with the historical data and then use it to do one-step prediction. The predicted outputs are shown with the dashed line. Figure 1 indicates that the AR process can capture the dynamic behaviors in the simulation outputs, and further provide a good forecast of future outputs.

At any calibration parameters $\boldsymbol{\theta}$, such as the service rates of the $M/M/1$ queues in the production example mentioned above, we model the simulation outputs as an AR process,

$$Y^s_{j,i}(\boldsymbol{\theta}) = c(\boldsymbol{\theta}) + \sum_{h=1}^{q} \alpha_h(\boldsymbol{\theta}) Y^s_{j,i-h}(\boldsymbol{\theta}) + e_{j,i}(\boldsymbol{\theta}), \tag{1}$$

where $e_{j,i}(\boldsymbol{\theta}) \sim \mathcal{N}(0, \sigma^2(\boldsymbol{\theta}))$ is Gaussian noise with $\mathcal{N}(a,b)$ standing for a Normal distribution with mean a and variance b. The serial dependence is characterized by $\boldsymbol{\alpha}(\boldsymbol{\theta}) = (\alpha_1(\boldsymbol{\theta}), \ldots, \alpha_q(\boldsymbol{\theta}))^\top$, and $c = (1 - \sum_{h=1}^{q} \alpha_h(\boldsymbol{\theta}))\mu(\boldsymbol{\theta})$ with $\mu(\boldsymbol{\theta}) = \mathrm{E}[Y^s_{j,i}(\boldsymbol{\theta})]$. Suppose that the order q ($< T$) is known here. The order selection for the AR process will be considered in the future research.

Given the simulation outputs from the jth replication, $(Y^s_{j1}, Y^s_{j2}, \ldots, Y^s_{jL})$, we can estimate AR process parameters $(\boldsymbol{\alpha}, c, \sigma^2)$ by using Yule-Walker estimation,

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_j &= (\hat{\alpha}_{1,j}, \ldots, \hat{\alpha}_{q,j})^\top = \hat{\Gamma}^{-1}_{q,j} \hat{\boldsymbol{\gamma}}_{q,j} && (2) \\ \hat{c}_j &= (1 - \mathbf{1}_{q\times 1} \hat{\Gamma}^{-1}_{q,j} \hat{\boldsymbol{\gamma}}_{q,j}) \bar{Y}^s_j && (3) \\ \hat{\sigma}^2_j &= \hat{\gamma}_j(0) - \hat{\boldsymbol{\alpha}}^\top_j \hat{\boldsymbol{\gamma}}_{q,j} \end{aligned}$$

where $\bar{Y}^s_j = \sum_{i=1}^{L} Y^s_{j,i}/L$, $\hat{\gamma}_j(h) = \frac{1}{L}\sum_{i=1}^{L-|h|}(Y^s_{j,i+|h|} - \bar{Y}^s_j)(Y^s_{j,i} - \bar{Y}^s_j)$, $\hat{\Gamma}_{q,j} = [\hat{\gamma}_j(i-i')]^q_{i,i'=1}$, $\hat{\boldsymbol{\gamma}}_{q,j} = (\hat{\gamma}_j(1), \ldots, \hat{\gamma}_j(q))^\top$. Then, the one-step predictor $\hat{Y}^s_{j,i} = \hat{c}_j + \sum_{h=1}^{q} \hat{\alpha}_{h,j} Y^s_{j,i-h}$ can be used to predict the next output.

3.2 Gaussian Process Metamodel for Spatial Dependence

As the calibration parameters $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ are close to each other, the dynamic behaviors of simulation output sample paths $\{Y_i^s(\boldsymbol{\theta})\}_{i=1}^{\infty}$ and $\{Y_i^s(\boldsymbol{\theta}')\}_{i=1}^{\infty}$ are similar. When we use the AR process in Equation (1) to model the simulation outputs, the dynamic behaviors are characterized by $(\boldsymbol{\alpha}, c, \sigma^2)$. According to Ankenman, Nelson, and Staum (2010), we construct a GP metamodel, denoted by $M(\boldsymbol{\theta})$, to model the spatial dependence of responses $(\boldsymbol{\alpha}, c)$ respectively, which are required in the AR process predictor. The variance σ^2 impacts on the estimation uncertainty for $(\boldsymbol{\alpha}, c)$. Thus, the simulation estimates of each response obtained from the jth replication can be modeled as

$$\hat{\xi}_j(\boldsymbol{\theta}) = \beta_0 + W(\boldsymbol{\theta}) + \varepsilon_j(\boldsymbol{\theta}) \tag{4}$$

with response $\hat{\xi}_j$ representing $\hat{c}_j$ or $\hat{\alpha}_{1,j}, \ldots, \hat{\alpha}_{q,j}$ obtained by using Equations (2)–(3). For each response, the simulation estimation noise caused by using a finite runlength is $\varepsilon(\boldsymbol{\theta}) \sim \mathcal{N}(0, \sigma_\varepsilon^2(\boldsymbol{\theta}))$. A mean-zero, second-order stationary GP, denoted by $W(\boldsymbol{\theta})$, accounts for the spatial dependence of $\xi(\boldsymbol{\theta})$. Thus, our belief on the response surface $\xi(\boldsymbol{\theta})$ is represented by the GP, $M(\boldsymbol{\theta}) = \beta_0 + W(\boldsymbol{\theta})$. The covariance between $W(\boldsymbol{\theta})$ and $W(\boldsymbol{\theta}')$ quantifies the spatial dependence of $\xi(\boldsymbol{\theta})$.

The estimation of the model in (4) for each response can be developed similarly. We describe the estimation procedure using the generic response $\hat{\xi}_j$. First, we choose an experiment design consisting of pairs $\mathscr{D} \equiv \{(\boldsymbol{\theta}_k, N_k, L_k), k = 1, \ldots, K\}$, where N_k and L_k denote the number of replications and runlength at the kth design point with calibration parameters $\boldsymbol{\theta}_k$. We run simulations at $\mathscr{D}$, and record the sample paths of simulation outputs, denoted by $\mathscr{Y}_{\mathscr{D}}^s$. Denote the sample means at the kth design point by $\bar{\xi}(\boldsymbol{\theta}_k) = \sum_{j=1}^{N_k} \hat{\xi}_j(\boldsymbol{\theta}_k)/N_k$. The sample means at all design points are denoted by $\bar{\xi}_{\mathscr{D}} = (\bar{\xi}(\boldsymbol{\theta}_1), \ldots, \bar{\xi}(\boldsymbol{\theta}_K))^\top$. The variance of $\bar{\xi}_{\mathscr{D}}$ is a $K \times K$ diagonal matrix $C = \mathrm{diag}\{\sigma_\varepsilon^2(\boldsymbol{\theta}_1)/N_1, \ldots, \sigma_\varepsilon^2(\boldsymbol{\theta}_K)/N_K)$.

Let Σ be the $K \times K$ spatial covariance matrix of the response at K design points with $\Sigma_{ih} = \mathrm{Cov}(M(\boldsymbol{\theta}_i), M(\boldsymbol{\theta}_h))$ for $i, h = 1, \ldots, K$. We use the Gaussian correlation function in the empirical study, $\mathrm{Cov}(M(\boldsymbol{\theta}_i), M(\boldsymbol{\theta}_h)) = \tau^2 \exp\left[-\sum_{l=1}^d \psi_l(\theta_{il} - \theta_{hl})^2\right]$, where τ^2 is the variance and $\boldsymbol{\psi} \equiv (\psi_1, \ldots, \psi_d)$ are covariance parameters. Let $\Sigma(\boldsymbol{\theta}, \cdot)$ be the $K \times 1$ spatial covariance vector between design points and a fixed prediction point $\boldsymbol{\theta}$. The metamodel uncertainty can be characterized by the posterior distribution of $M(\boldsymbol{\theta})$,

$$M_p(\boldsymbol{\theta}) \equiv M(\boldsymbol{\theta}) | \mathscr{Y}_{\mathscr{D}}^s \sim \mathrm{GP}(m_p(\boldsymbol{\theta}), \sigma_p^2(\boldsymbol{\theta})) \tag{5}$$

where $m_p(\cdot)$ is the minimum mean squared error (MSE) linear unbiased predictor

$$m_p(\boldsymbol{\theta}) = \hat{\beta}_0 + \Sigma(\boldsymbol{\theta}, \cdot)^\top (\Sigma + C)^{-1} (\bar{\xi}_{\mathscr{D}} - \hat{\beta}_0 \cdot 1_{K \times 1}) \tag{6}$$

and the corresponding variance is

$$\sigma_p^2(\boldsymbol{\theta}) = r^2 - \Sigma(\boldsymbol{\theta}, \cdot)^\top (\Sigma + C)^{-1} \Sigma(\boldsymbol{\theta}, \cdot) + \eta^\top [1_{K \times 1}^\top (\Sigma + C)^{-1} 1_{K \times 1}]^{-1} \eta \tag{7}$$

where $\hat{\beta}_0 = [1_{K \times 1}^\top (\Sigma + C)^{-1} 1_{K \times 1}]^{-1} 1_{K \times 1}^\top (\Sigma + C)^{-1} \bar{\xi}_{\mathscr{D}}$ and $\eta = 1 - 1_{K \times 1}^\top (\Sigma + C)^{-1} \Sigma(\boldsymbol{\theta}, \cdot)$ (Ankenman, Nelson, and Staum 2010). The intrinsic covariance matrix C is diagonal, and the kth diagonal element can be estimated by $\hat{C}_{k,k} = N_k^{-1}(N_k - 1)^{-1} \sum_{j=1}^{N_k} (\hat{\xi}_j(\boldsymbol{\theta}_k) - \bar{\xi}(\boldsymbol{\theta}_k))^2$ for $k = 1, \ldots, K$. Then, we substitute it into the likelihood function and obtain MLEs for τ^2 and $\boldsymbol{\psi}$.

In sum, the spatial-temporal metamodel for simulation outputs can be constructed in two main steps as described in Sections 3.1 and 3.2, respectively. To initialize the construction, we generate K design points $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K\}$ by using the Latin Hypercube design covering the space of the calibration parameters. Then, the metamodel can be constructed as follows. In the empirical study, we assign equal runlength and replication to each design point. Developing an experiment design $\mathscr{D} = \{(\boldsymbol{\theta}_k, N_k, L_k), k = 1, \ldots, K\}$ that can efficiently use the simulation resource to support the calibration will be considered in the future research.

1. For $k = 1, \ldots, K$
   1a. Run simulations at $\boldsymbol{\theta}_k$ with runlength L_k and replications N_k. Record the sample path $(Y^s_{j,1}, \ldots, Y^s_{j,L_k})$ for $j = 1, \ldots, N_k$.
   1b. Obtain $\hat{\boldsymbol{\alpha}}_{j,k}$ and $\hat{c}_{j,k}$ by applying Equations (2)–(3) for $j = 1, \ldots, N_k$.
2. Construct GP metamodels $M_p(\cdot)$ for responses $c, \alpha_1, \ldots, \alpha_q$ by applying Equations (5)–(7).

Remark: In this paper, we construct separate GP metamodels for $\alpha_1, \ldots, \alpha_q$, and c, respectively. Since these responses could be correlated, we can construct a multivariate GP model, which incorporates the correlation between different responses; see Qian, Wu, and Wu (2008).

4 A Calibration Model for the System Control

In this section, we calibrate the parameters $\boldsymbol{\theta}$ to match the dynamic behaviors of real and simulation systems. Given the historical data $\mathbf{y}^r_{[T]}$ and the simulation outputs $\mathscr{Y}^s_{\mathscr{D}}$, our belief of the calibration parameters is characterized by the posterior distribution $p(\boldsymbol{\theta}|M_p(\cdot), \mathbf{y}^r_{[T]})$. Then, the calibrated simulation model is used to predict the future output Y^r_{T+1}. Our approach provides the posterior predictive distribution $p(Y^r_{T+1}|M_p(\cdot), \mathbf{y}^r_{[T]})$ and also a prediction interval (PI) accounting for calibration parameters and simulation estimation uncertainty.

For the simulation-based system control, we want to calibrate the dynamic behaviors of output sample paths so that the calibrated simulation model can predict the future outputs of the real system. This goal can not be fulfilled directly by matching the data from real and simulation systems for deterministic simulation or matching the mean responses for stochastic system design; see for example Kennedy and O'Hagan (2001), Tuo and Wu (2015b), Yuan, Ng, and Tsui (2013). To support the system control, we propose a new calibration model,

$$y^r_i = \hat{y}^r_i(\boldsymbol{\xi}(\boldsymbol{\theta})) + \delta_i. \tag{8}$$

The predictor of y^r_i is expressed as

$$\hat{y}^r_i(\boldsymbol{\xi}(\boldsymbol{\theta})) = c(\boldsymbol{\theta}) + \sum_{h=1}^{q} \alpha_h(\boldsymbol{\theta}) y^r_{i-h}. \tag{9}$$

with $i = q+1, \ldots, T$ which depends on the historical data $\mathbf{y}^r_{[T]}$ and the serial dependence of the simulation outputs characterized by $\boldsymbol{\xi}(\boldsymbol{\theta}) = (c(\boldsymbol{\theta}), \boldsymbol{\alpha}(\boldsymbol{\theta}))$. Given $\mathscr{Y}^s_{\mathscr{D}}$, the estimation uncertainty of $\boldsymbol{\xi}(\cdot)$ is quantified by the GP metamodel $M_p(\cdot)$. The model inadequacy, denoted by $\{\delta_i\}$, in Equation (8) is a stochastic process measuring the difference between the historical data and the forecast based on the simulation model. In this paper, we assume that $\{\delta_i\}$ is independent of $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$, and modeled by an AR process,

$$\delta_i = \sum_{h=1}^{q'} \phi_h \delta_{i-h} + \varepsilon_i \tag{10}$$

with $\varepsilon_i \overset{i.i.d.}{\sim} \mathscr{N}(0, \sigma^2_\delta)$. Let $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_{q'})$. The first term on the right side of Equation (10) is used to model the remaining series dependence which is not captured by the predictor $\hat{y}^r_i$. The second term models the unpredictable noise, including the unpredicted error for the real system outputs. Next, we do inference on the calibration parameters in Section 4.1, and predict the future output Y^r_{T+1} in Section 4.2.

4.1 Inference for Calibration Parameters

In this section, we derive the posterior distribution $p(\boldsymbol{\theta}|M_p(\cdot), \mathbf{y}^r_{[T]})$ quantifying the estimation uncertainty of calibration parameters, and then provide a sampling procedure to generate the posterior samples of $\boldsymbol{\theta}$.

We first derive the conditional distribution for the predictors $\hat{\mathbf{y}}^r_{[q+1:T]}(\boldsymbol{\xi}(\boldsymbol{\theta})) = (\hat{y}^r_T(\boldsymbol{\xi}(\boldsymbol{\theta})), \ldots, \hat{y}^r_{q+1}(\boldsymbol{\xi}(\boldsymbol{\theta})))^\top$. According to Equation (9), we have

$$\begin{aligned} & p\left(\hat{\mathbf{y}}^r_{[q+1:T]}(\boldsymbol{\xi}(\boldsymbol{\theta})) \middle| \boldsymbol{\theta}, M_p(\cdot), \mathbf{y}^r_{[T-1]}\right) \\ \sim \quad & \mathscr{N}_{T-q}\left(\mathscr{Y}^r_{[T-1]}\mathrm{E}\left[\begin{array}{c} c(\boldsymbol{\theta}) \\ \alpha_1(\boldsymbol{\theta}) \\ \vdots \\ \alpha_q(\boldsymbol{\theta}) \end{array} \middle| \mathscr{Y}^s_{\mathscr{D}}\right], \mathscr{Y}^r_{[T-1]}\mathrm{Var}\left[\begin{array}{c} c(\boldsymbol{\theta}) \\ \alpha_1(\boldsymbol{\theta}) \\ \vdots \\ \alpha_q(\boldsymbol{\theta}) \end{array} \middle| \mathscr{Y}^s_{\mathscr{D}}\right] \mathscr{Y}^r_{[T-1]}{}^\top\right) \end{aligned} \tag{11}$$

where, $\mathscr{Y}^r_{[T-1]}$ is a $(T-q)\times(q+1)$ matrix

$$\mathscr{Y}^r_{[T-1]} = \begin{bmatrix} 1 & y^r_{T-1} & y^r_{T-2} & \cdots & y^r_{T-q} \\ \vdots & & & & \vdots \\ 1 & y^r_q & y^r_{q-1} & \cdots & y^r_1 \end{bmatrix}$$

and the conditional mean and variance of $(c(\boldsymbol{\theta}), \alpha_1(\boldsymbol{\theta}), \ldots, \alpha_q(\boldsymbol{\theta}))|\mathscr{Y}^s_{\mathscr{D}}$ can be obtained by using Equations (5)–(7) from the GP metamodels. At any $\boldsymbol{\theta}$, Equation (11) accounts for the impact of simulation estimation uncertainty on the predictors. Then, since $\delta_i = y^r_i - \hat{y}^r_i(\boldsymbol{\xi}(\boldsymbol{\theta}))$ follows the AR process as shown in Equation (10), we can derive the posterior distribution of calibration parameters,

$$\begin{aligned} p\left(\boldsymbol{\theta} \middle| M_p(\cdot), \mathbf{y}^r_{[T]}, \boldsymbol{\phi}, \sigma^2_\delta\right) & \propto p(\boldsymbol{\theta}) p\left((\delta_{q+q'+1}, \ldots, \delta_T) \middle| \boldsymbol{\theta}, M_p(\cdot), \mathbf{y}^r_{[T]}, \boldsymbol{\phi}, \sigma^2_\delta\right) \\ \propto \quad & p(\boldsymbol{\theta}) \prod_{i=q+q'+1}^{T} \exp\left[-\frac{(\delta_i - \sum_{h=1}^{q'} \phi_h \delta_{i-h})^2}{2\sigma^2_\delta}\right]. \end{aligned} \tag{12}$$

To obtain the posterior distribution of $\boldsymbol{\theta}$, we develop an empirical Bayesian procedure to efficiently draw posterior samples of $\boldsymbol{\theta}$, and then construct a $(1-\beta)100\%$ percentile credible interval (CrI) quantifying the calibration parameter estimation uncertainty as follows. To precisely estimate the percentile CrI for $\boldsymbol{\theta}$, it is recommended B to be a few thousands. In our empirical study, we use $B = 1000$.

1. Draw $\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$. Given the GP metamodel built in Section 3.2, draw the predictors $\hat{\mathbf{y}}^r_{[q+1:T]}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ by applying Equation (11). Obtain $(\delta_{q+1}, \ldots, \delta_T)$ with $\delta_i = y^r_i - \hat{y}^r_i(\boldsymbol{\xi}(\boldsymbol{\theta}))$ for $i = q+1, \ldots, T$.
2. Repeat Step 1 for B times. Then, use all B samples of $(\delta_{q+1}, \ldots, \delta_T)$ collected at different $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ to estimate the order q' by BIC and then obtain $\hat{\boldsymbol{\phi}}$ and $\hat{\sigma}^2_\delta$ by using Yule-Walker estimation.
3. Plug $\hat{\boldsymbol{\phi}}$ and $\hat{\sigma}^2_\delta$ into Equation (12). Calculate the likelihood $p((\delta_{q+q'+1}, \ldots, \delta_T)|\boldsymbol{\theta}, M_p(\cdot), \mathbf{y}^r_{[T]}, \hat{\boldsymbol{\phi}}, \hat{\sigma}^2_\delta)$ of all B samples of δ. Standardize the probability at all B samples of $\boldsymbol{\theta}$ to obtain the posterior distribution $p(\boldsymbol{\theta}|M_p(\cdot), \mathbf{y}^r_{[T]})$.
4. By applying the inverse CDF approach (Nelson 2013) to $p(\boldsymbol{\theta}|M(\cdot), \mathbf{y}^r_{[T]})$, generate B posterior samples of $\boldsymbol{\theta}$, and further construct a $(1-\beta)100\%$ percentile CrI quantifying the calibration parameter estimation uncertainty.

4.2 Prediction for Future Output Y^r_{T+1}

In this section, we derive the posterior predictive distribution of Y^r_{T+1}, and provide a $(1-\beta)100\%$ percentile PI accounting for the model inadequacy, calibration parameters and simulation estimation uncertainty. Built on the development in Section 4.1, the posterior predictive distribution for Y^r_{T+1} is

$$\begin{aligned} p\left(Y^r_{T+1} \middle| M_p(\cdot), \mathbf{y}^r_{[T]}, \boldsymbol{\phi}, \sigma^2_\delta\right) = & \int\int p\left(Y^r_{T+1} \middle| \hat{\mathbf{y}}^r_{[T]}(\boldsymbol{\xi}(\boldsymbol{\theta})), \mathbf{y}^r_{[T]}, \boldsymbol{\phi}, \sigma^2_\delta\right) \\ & \times p\left(\hat{\mathbf{y}}^r_{[T]}(\boldsymbol{\xi}(\boldsymbol{\theta})) \middle| \boldsymbol{\theta}, M_p(\cdot)\right) p\left(\boldsymbol{\theta} \middle| M_p(\cdot), \mathbf{y}^r_{[T]}, \boldsymbol{\phi}, \sigma^2_\delta\right) d\hat{\mathbf{y}}^r_{[T]}(\boldsymbol{\xi}(\boldsymbol{\theta})) d\boldsymbol{\theta}. \end{aligned} \tag{13}$$

It integrates out calibration parameters and simulation estimation uncertainty. By comparing to the best predictor with the underlying stochastic process of real system outputs, denoted by $\{Y_i^r\}$, known, we can assess the performance of the calibrated simulation model; see the empirical study in Section 5.2.

Then, the procedure to obtain the posterior predictive distribution and the $(1-\beta)100\%$ percentile PI for Y_{T+1}^r is described as follows.

1. By applying the inverse CDF approach, draw B posterior samples from $p(\boldsymbol{\theta}|M_p(\cdot),\mathbf{y}_{[T]}^r)$; see the procedure in Section 4.1.
2. At each posterior sample of $\boldsymbol{\theta}$,
   2a. Draw the sample of $(c,\alpha_1,\ldots,\alpha_q)$ by using Equations (5)–(7) from the GP metamodels. Then, obtain the sample of $(\hat{y}_{q+1}^r,\ldots,\hat{y}_{T+1}^r)$ by applying Equation (9).
   2b. Obtain $(\delta_{q+1},\ldots,\delta_T)$ with $\delta_i = y_i^r - \hat{y}_i^r$ for $i = q+1,\ldots,T$. Then, generate δ_{T+1} by applying Equation (10) with parameters $\hat{\boldsymbol{\phi}}$ and $\hat{\sigma}_\delta^2$, and obtain $Y_{T+1}^r = \hat{y}_{T+1}^r + \delta_{T+1}$.
3. Use B posterior samples obtained in Step (2) to construct the $(1-\beta)100\%$ percentile PI for Y_{T+1}^r.

5 EMPIRICAL STUDY

In this section, we use an $M/M/1$ queue to study the finite sample performance of our calibration framework. The arrival rate is set to be 1. Let the utilization be the calibration parameter θ. We are interested in the total time of the customer staying in the system. We first study the performance of the spatial-temporal metamodel in Section 5.1. Then, we study the performance of the proposed calibration model in Section 5.2.

5.1 Evaluate the Performance of Spatial-Temporal Metamodel

We first follow the procedure in Section 4.1 to construct a spatial-temporal metamodel capturing the dynamic behaviors of the $M/M/1$ queue at different calibration settings. We generate K design points of calibration parameter on $[0.2, 0.91]$ by using Latin-hypercube design. Here, we consider a simple AR(1) model to model the dynamic behaviors in the system output sample path. Thus, we construct the GP metamodels for parameters c and α. We evaluate these parameters estimation accuracy by using the integrated mean square error (IMSE). Specifically, we select 20 testing points with θ_ℓ for $\ell = 1,\ldots,20$ uniformly distributed on $[0.2, 0.91]$. At each θ_ℓ, a side experiment with run length 10^5 is used to estimate the exact AR(1) process parameters c and α. Then, 100 posterior samples of metamodel $M_p(\cdot)$ at θ_ℓ for $\ell = 1,\ldots,20$, denoted by $\tilde{c}^{(i)}$ and $\tilde{\alpha}^{(i)}$ for $i = 1,\ldots,100$, are used to estimate IMSE

$$\widehat{\text{IMSE}}(c(\boldsymbol{\theta})) = \frac{1}{20\times 100}\sum_{\ell=1}^{20}\sum_{i=1}^{100}(\tilde{c}^{(i)}(\theta_\ell) - c(\theta_\ell))^2 \text{ and } \widehat{\text{IMSE}}(\alpha(\boldsymbol{\theta})) = \frac{1}{20\times 100}\sum_{\ell=1}^{20}\sum_{i=1}^{100}(\tilde{\alpha}^{(i)}(\theta_\ell) - \alpha(\theta_\ell))^2.$$

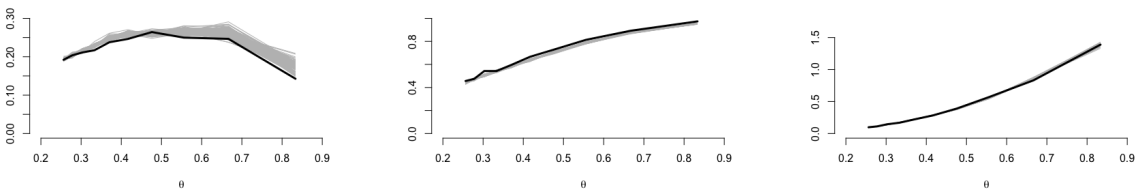
We evaluate the impact of the number of design points K and the runlength L on the IMSE and report the results for $K = 5, 15$ and $L = 50, 100$ in Table 1. When $K = 15$ and $L = 100$, the simulation estimation uncertainty for c and α is small. We also plot the response surfaces of parameters (c,α,σ^2) in Figure 2. The black solid line represents the exact parameter values estimated from the side experiment and the grey lines represent posterior samples from the GP metamodel. It can be seen that the samples from GP is close to the best AR(1) parameters estimated by side experiments.

Table 1: Estimated IMSE with AR(1) process

	c	α	σ^2		c	α	σ^2
$L=50, K=5$	0.0184	0.0152	0.1135	$L=100, K=5$	0.0157	0.0123	0.0768
$L=50, K=15$	0.0142	0.0131	0.0744	$L=100, K=15$	0.0109	0.0098	0.0460

Then, to study the prediction performance of the spatial-temporal metamodel, we collect 101 customers' times staying in system, $\mathbf{y}^r = (y_1^r,\ldots,y_{101}^r)$, at $\theta = 0.833, 0.667, 0.5$. We first derive the best predictor

Figure 2: Compare GP fitted AR(1) model parameters $(c, \boldsymbol{\alpha}, \sigma^2)$ with the best ones from side experiments



(a) The best c vs GP fitted c (b) The best α vs GP fitted α (c) The best σ^2 vs GP fitted σ^2

by using the underling output process. Let A_i and S_i denote the inter-arrival and service times of the ith customer in the $M/M/1$ queue. At any θ, the conditional expectation $\mathrm{E}[Y_i^r|\mathbf{y}_{[i-1]}^r, \theta]$ is the best mean square predictor (Brockwell and Davis 1991). Since $Y_i^r = \max\{Y_{i-1}^r - A_i, 0\} + S_i$, we have

$$\begin{aligned}\mathrm{E}[Y_i^r|\mathbf{y}_{[i-1]}^r, \theta] &= \mathrm{E}[S_i|\theta] + \mathrm{E}[\max\{y_{i-1}^r - A_i, 0\}|y_{i-1}^r] = \theta + \int_0^\infty \max\{y_{i-1}^r - A_i, 0\} e^{-A_i} dA_t \\ &= \theta + \int_0^{y_{i-1}^r} (y_{i-1}^r - A_i) e^{-A_i} dA_i = y_{i-1}^r + e^{-y_{i-1}^r} + \theta - 1.\end{aligned}$$

This predictor is used as the benchmark to study the performance of our metamodel-assisted prediction and a naive approach that uses the average of previous responses as a predictor, $\hat{y}_i = (i-1)^{-1} \sum_{h=1}^{i-1} y_{i-h}^r$. As for our metamodel-assisted prediction, we first draw a posterior sample of parameters $(\tilde{c}, \tilde{\alpha})$ from the GP metamodel $M_p(\theta)$. For the ith output, combining with the historical data, we can predict the future output by using $\hat{y}_i = \tilde{c} + \tilde{\alpha} y_{i-1}^r$ accounting for the simulation estimation uncertainty of parameters (c, α).

The relative mean absolute error (RMAE),

$$\mathrm{RMAE}(\hat{\mathbf{y}}) = \frac{1}{L-1} \sum_{i=2}^{L} \frac{|\hat{y}_i - y_i^r|}{\mathrm{E}(y_t)}$$

is used to assess the prediction performance. The results of RMAE obtained by using the metamodel are compared with the naive method and conditional expectation in Table 2. We record mean and standard deviation (in parentheses) of RMAE estimated from 100 macro-replications. The prediction accuracy of our metamodel-based approach becomes closer to the target, conditional expectation, when the runlength L and the number of design points K increase, and it performs obviously better than the naive method.

Table 2: RMAE obtained by spatial-temporal metamodel, naive approach and the conditional expectation

Utilization of test setting		$\theta = 0.833$	$\theta = 0.667$	$\theta = 0.5$
Metamodels	$L = 50, K = 5$	0.370 (0.034)	0.415 (0.040)	0.533 (0.048)
	$L = 50, K = 15$	0.316 (0.039)	0.388 (0.036)	0.504 (0.032)
	$L = 100, K = 5$	0.215 (0.051)	0.383 (0.053)	0.497 (0.041)
	$L = 100, K = 15$	0.197 (0.042)	0.342 (0.047)	0.490 (0.038)
Naive Method		0.475 (0.094)	0.524 (0.068)	0.631 (0.047)
Conditional Expectation		0.184 (0.055)	0.338 (0.056)	0.486 (0.044)

5.2 Evaluate the Performance of Calibration Framework

As a first step to study the performance of proposed calibration framework, we do not introduce any logic error in the simulation model here. We use the $M/M/1$ queue as the real and simulation models. The

unknown utilization for the real system is θ^r. The utilization is also the calibration parameter θ for the simulation system. Given the output data of the time staying in the real system $\mathbf{y}^r_{[T]}$, we calibrate θ so that the dynamic behaviors of the calibrated simulation model match with those of the real system.

To study the posterior $p(\theta|M_p(\cdot),\mathbf{y}^r_{[T]})$ for the calibrated simulation model, we first construct the GP metamodel $M_p(\cdot)$ with $K = 5,15,25$ and $L = 50,100$. Then, following the procedure in Section 4.1, we get the posterior $p(\theta|M(\cdot),\mathbf{y}^r_{[T]})$. We record the width for the 95% percentile CrI and the posterior mean $\mathrm{E}[\theta|M_p(\cdot),\mathbf{y}^r_{[T]}]$ in Table 3. The results of mean and SD (in parentheses) are estimated by using 100 macro-replications. As the simulation budget represented by L,K and the amount of real world data T increase, the posterior $p(\theta|M_p(\cdot),\mathbf{y}^r_{[T]})$ converges to the underlying utilization of the real system θ^r.

Table 3: CrI width and posterior mean of the calibration parameter θ

$\theta^r = 0.833$	$T = 50$		$T = 100$	
	$\lvert\mathrm{CrI}(\theta)\rvert$	$\mathrm{E}[\theta\lvert M_p(\cdot),\mathbf{y}^r_{[T]}]$	$\lvert\mathrm{CrI}(\theta)\rvert$	$\mathrm{E}[\theta\lvert M_p(\cdot),\mathbf{y}^r_{[T]}]$
$L = 50, K = 5$	0.106 (0.065)	0.892 (0.046)	0.096 (0.053)	0.875 (0.048)
$L = 50, K = 15$	0.099 (0.060)	0.874 (0.058)	0.089 (0.050)	0.866 (0.059)
$L = 50, K = 25$	0.094 (0.061)	0.865 (0.049)	0.085 (0.051)	0.858 (0.047)
$L = 100, K = 5$	0.101 (0.059)	0.883 (0.058)	0.092 (0.051)	0.869 (0.051)
$L = 100, K = 15$	0.095 (0.056)	0.868 (0.066)	0.089 (0.054)	0.853 (0.058)
$L = 100, K = 25$	0.090 (0.055)	0.855 (0.066)	0.084 (0.050)	0.848 (0.052)
$\theta^r = 0.667$	$T = 50$		$T = 100$	
	$\lvert\mathrm{CrI}(\theta)\rvert$	$\mathrm{E}[\theta\lvert M_p(\cdot),\mathbf{y}^r_{[T]}]$	$\lvert\mathrm{CrI}(\theta)\rvert$	$\mathrm{E}[\theta\lvert M_p(\cdot),\mathbf{y}^r_{[T]}]$
$L = 50, K = 5$	0.127 (0.068)	0.823 (0.085)	0.103 (0.058)	0.801 (0.083)
$L = 50, K = 15$	0.114 (0.057)	0.739 (0.080)	0.086 (0.044)	0.719 (0.075)
$L = 50, K = 25$	0.101 (0.052)	0.706 (0.074)	0.081 (0.043)	0.690 (0.072)
$L = 100, K = 5$	0.122 (0.064)	0.798 (0.082)	0.092 (0.048)	0.770 (0.082)
$L = 100, K = 15$	0.109 (0.053)	0.711 (0.075)	0.078 (0.041)	0.694 (0.069)
$L = 100, K = 25$	0.095 (0.048)	0.689 (0.072)	0.073 (0.052)	0.681 (0.071)
$\theta^r = 0.5$	$T = 50$		$T = 100$	
	$\lvert\mathrm{CrI}(\theta)\rvert$	$\mathrm{E}[\theta\lvert M_p(\cdot),\mathbf{y}^r_{[T]}]$	$\lvert\mathrm{CrI}(\theta)\rvert$	$\mathrm{E}[\theta\lvert M_p(\cdot),\mathbf{y}^r_{[T]}]$
$L = 50, K = 5$	0.134 (0.063)	0.727 (0.081)	0.106 (0.054)	0.693 (0.077)
$L = 50, K = 15$	0.120 (0.058)	0.634 (0.069)	0.094 (0.045)	0.568 (0.064)
$L = 50, K = 25$	0.113 (0.054)	0.572 (0.060)	0.091 (0.049)	0.543 (0.059)
$L = 100, K = 5$	0.122 (0.061)	0.685 (0.074)	0.102 (0.046)	0.586 (0.065)
$L = 100, K = 15$	0.107 (0.052)	0.568 (0.062)	0.087 (0.039)	0.517 (0.056)
$L = 100, K = 25$	0.098 (0.047)	0.526 (0.057)	0.084 (0.042)	0.506 (0.053)

Then, we study the prediction performance of the calibrated simulation model. Given the historical output data $\mathbf{y}^r_{[T]}$, we want to predict the next output Y^r_{T+1}. Following the procedure in Section 4.2, we can get the posterior predictive distribution $p(Y^r_{T+1}|M_p(\cdot),\mathbf{y}^r_{[T]})$. For simplification, we let the order $q' = 1$ here. The prediction error is defined as $\mathrm{Err}(\hat{y}^r_{T+1}) = |\hat{y}^r_{T+1} - y^r_{T+1}|$. For our approach, the predictor is $\hat{y}^r_{T+1} = \mathrm{E}[Y^r_{T+1}|M_p(\cdot),\mathbf{y}^r_{[T]}]$. For the real system, when θ^r and the underlying output process are known, the conditional expectation gives the best mean square predictor, $\mathrm{E}[Y^r_{T+1}|\mathbf{y}^r_{[T]},\theta^r] = y^r_T + e^{-y^r_T} + \theta^r - 1$. In Table 4, we report mean and SD for the prediction error $\mathrm{Err}(\hat{y}^r_{T+1})$ and the width of 95% percentile PI for Y^r_{T+1}. The results are based on 100 macro-replications. According to Table 4, as the size of historical

data T and the simulation budget represented by L and K increase, the prediction uncertainty reduces, and $\mathrm{E}[Y^r_{T+1}|M(\cdot),\mathbf{y}^r_{[T]}]$ obtained from the calibrated simulation model becomes closer to $\mathrm{E}[Y^r_{T+1}|\mathbf{y}^r_{[T]},\theta^r]$.

Table 4: Prediction error and PI width from our calibration framework and the conditional expectation.

$\theta^r = 0.833$	$T = 50$		$T = 100$	
	$\mathrm{Err}(\hat{y}^r_{T+1})$	$\lvert\mathrm{PI}(Y^r_{T+1})\rvert$	$\mathrm{Err}(\hat{y}^r_{T+1})$	$\lvert\mathrm{PI}(Y^r_{T+1})\rvert$
$L = 50, K = 5$	1.679 (0.853)	4.665 (1.138)	1.438 (0.713)	4.130 (1.058)
$L = 50, K = 15$	1.354 (0.748)	4.036 (1.092)	1.144 (0.562)	3.722 (0.943)
$L = 50, K = 25$	1.186 (0.657)	3.781 (0.959)	1.028 (0.534)	3.427 (0.908)
$L = 100, K = 5$	1.483 (0.735)	4.224 (1.117)	1.385 (0.545)	3.836 (0.952)
$L = 100, K = 15$	1.258 (0.633)	3.966 (0.983)	1.095 (0.507)	3.340 (0.824)
$L = 100, K = 25$	1.094 (0.515)	3.509 (0.885)	0.953 (0.422)	3.218 (0.836)
Conditional	0.880 (0.422)	3.639 (0.728)	0.871 (0.374)	3.602 (0.683)
$\theta^r = 0.667$	$T = 50$		$T = 100$	
	$\mathrm{Err}(\hat{y}^r_{T+1})$	$\lvert\mathrm{PI}(Y^r_{T+1})\rvert$	$\mathrm{Err}(\hat{y}^r_{T+1})$	$\lvert\mathrm{PI}(Y^r_{T+1})\rvert$
$L = 50, K = 5$	1.127 (0.441)	3.267 (0.708)	0.959 (0.415)	3.157 (0.604)
$L = 50, K = 15$	0.928 (0.409)	2.933 (0.674)	0.867 (0.360)	2.833 (0.534)
$L = 50, K = 25$	0.887 (0.386)	2.506 (0.621)	0.792 (0.344)	2.673 (0.517)
$L = 100, K = 5$	1.012 (0.435)	3.083 (0.686)	0.913 (0.385)	2.944 (0.562)
$L = 100, K = 15$	0.906 (0.418)	2.812 (0.647)	0.832 (0.364)	2.675 (0.498)
$L = 100, K = 25$	0.873 (0.377)	2.458 (0.602)	0.779 (0.338)	2.518 (0.470)
Conditional	0.814 (0.360)	2.617 (0.594)	0.723 (0.310)	2.595 (0.510)
$\theta^r = 0.5$	$T = 50$		$T = 100$	
	$\mathrm{Err}(\hat{y}^r_{T+1})$	$\lvert\mathrm{PI}(Y^r_{T+1})\rvert$	$\mathrm{Err}(\hat{y}^r_{T+1})$	$\lvert\mathrm{PI}(Y^r_{T+1})\rvert$
$L = 50, K = 5$	0.753 (0.324)	2.853 (0.517)	0.704 (0.283)	2.589 (0.463)
$L = 50, K = 15$	0.671 (0.290)	2.514 (0.442)	0.633 (0.258)	2.370 (0.428)
$L = 50, K = 25$	0.588 (0.276)	2.279 (0.403)	0.562 (0.209)	2.221 (0.413)
$L = 100, K = 5$	0.720 (0.306)	2.698 (0.493)	0.681 (0.266)	2.452 (0.437)
$L = 100, K = 15$	0.636 (0.283)	2.379 (0.438)	0.619 (0.235)	2.146 (0.409)
$L = 100, K = 25$	0.573 (0.250)	2.040 (0.376)	0.550 (0.207)	2.053 (0.392)
Conditional	0.494 (0.237)	1.975 (0.529)	0.507 (0.250)	1.989 (0.483)

6 CONCLUSIONS

In this paper, we propose a stochastic simulation calibration framework for the system control. We first introduce a spatial-temporal metamodel to estimate the system dynamic behaviors of the simulation model at different calibration settings. Then, assisted by the metamodel, we propose a Bayesian calibration framework so that the calibrated simulation model can capture the dynamic behaviors of the real system and improve the prediction of future outputs. The empirical study over an $M/M/1$ queue indicates the promising performances of our approach.

REFERENCES

Ankenman, B. E., B. L. Nelson, and J. Staum. 2010. "Stochastic Kriging for Simulation Metamodeling". *Operations Research* 58:371–382.

Brockwell, P. J., and R. A. Davis. 1991. *Time Series: Theory and Methods*. 2 ed. New York: Springer-Verlag.

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.

Horiguchi, K., N. Raghavan, R. Uzsoy, and S. Venkateswaran. 2001. "Finite-Capacity Production Planning Algorithms for a Semiconductor Wafer Fabrication Facility". *International Journal of Production Research* 39:825–842.

Jun, Y., and S. H. Ng. 2013. "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.

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.

Nelson, B. L. 2013. *Foundations and Methods of Stochastic Simulation: A First Course*. Springer-Verlag.

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* (To appear).

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* (To appear).

Qian, Z. G., H. Q. Wu, and C. F. J. Wu. 2008. "Gaussian Process models for Computer Experiments with Qualitative and Quantitative Factor". *Technometrics* 50:192–204.

Tuo, R., and C. F. J. Wu. 2015a. "Efficient Calibration for Imperfect Computer Models". *The Annals of Statistics* 43 (6): 2331–2352.

Tuo, R., and C. F. J. Wu. 2015b. "A Theoretical Framework for Calibration in Computer Models: Parametrization, Estimation and Convergence Properties". *arXiv preprint arXiv:1508.07155*.

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)*.

Xie, W., B. L. Nelson, and R. R. Barton. 2014. "A Bayesian Framework for Quantifying Uncertainty in Stochastic Simulation". *Operations Research* 62 (6): 1439–1452.

Yuan, J., and S. H. Ng. 2013. "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

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/.

PU ZHANG s a Ph.D. candidate of the Department of Industrial and Systems Engineering at Rensselaer Polytechnic Institute. His research interests are input modeling and uncertainty quantification in stochastic simulation. His email address is zhangp5@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.