Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

SIMULATION METAMODELING IN THE PRESENCE OF MODEL INADEQUACY

Xiaowei Zhang
Lu Zou

Department of Industrial Engineering and Logistics Management
The Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong, CHINA

ABSTRACT

A simulation model is often used as a proxy for the real system of interest in a decision-making process. However, no simulation model is totally representative of the reality. The impact of the model inadequacy on the prediction of system performance should be carefully assessed. We propose a new metamodeling approach to simultaneously characterize both the simulation model and its model inadequacy. Our approach utilizes both simulation outputs and real data to predict system performance, and accounts for four types of uncertainty that arise from the unknown performance measure of the simulation model, simulation errors, unknown model inadequacy, and observation errors of the real system, respectively. Numerical results show that the new approach provides more accurate predictions in general.

1 INTRODUCTION

Simulation models are often used to facilitate a decision-making process related to complex stochastic systems. However, no simulation model is totally representative of the real system due to various reasons. For instance, input distributions of the model are mis-specified because of lack of data; or certain structural details of the real system are missing from the model. We hereafter refer to the discrepancies between the model's behavior and the reality as *model inadequacy*. Clearly, failing to account for model inadequacy in simulation studies may lead to misinformed decisions and significant economic loss.

In the current simulation literature, the assessment and reduction of model inadequacy is typically conducted through the model calibration process. In general, one needs to perform several iterations of model validation and adjustment, until the discrepancies between the simulation outputs and the real data are reduced to an acceptable level. This process usually involves *ad hoc* techniques and relies heavily on the modeler's modeling experience and domain knowledge of the real system; see Sargent (1999).

Nevertheless, even a "calibrated" simulation model may still have non-negligible model inadequacy, if the time and cost budget does not allow a highly accurate model to be developed. Note that once a simulation model is adopted, its model inadequacy will generally be neglected in the subsequent studies, in which case it is difficult to characterize the impact of the model inadequacy on the predicted system performance. In this paper, we propose a new approach for explicitly quantifying the model inadequacy and integrating it into system performance predictions, thereby significantly improving prediction accuracy.

In recent years, metamodeling techniques have drawn increasing academic interest in the simulation community in light of the fact that large-scale simulation of complex systems is computationally expensive. The basic idea is that one runs simulations only at a few selected "design points" of the simulation model and then uses the simulation outputs at these points to fit a tractable metamodel. System performance at other points are then predicted by interpolating the previous simulation outputs under the metamodel, thereby achieving considerable reduction in computational expenses. There is a vast literature on metamodeling and we refer to Barton and Meckesheimer (2006) for a review on the subject.

We adopt metamodeling techniques since model inadequacy is potentially more substantial for complex systems. Specifically, we use two Gaussian random fields to respectively characterize the uncertainty about the mean performance measure of the chosen simulation model and the uncertainty about its model inadequacy. The metamodel that characterizes the unknown performance measure of the real system of interest takes the form of an additive combination of the two random fields. By using the simulation outputs and the real data, i.e. physical observations of the real system's performance, we can both quantify the model inadequacy and measure the fidelity of the simulation model relative to the reality; see equation (1) for our particular definition of inadequacy and fidelity of the simulation model.

The proposed approach extends existing metamodeling techniques, particularly the stochastic kriging metamodel (Ankenman, Nelson, and Staum 2010), in two aspects. First, prediction of system performance in our approach is driven by both simulation outputs and the real data, rather than relying on only the former as most current metamodels do. For the proposed approach to work in practice, one needs to be able to collect real data at different "design points" of the system, which may be challenging for some systems. For example, in an automobile manufacturing system, the design variables such as machine capacity and process cycle times are mostly fixed during its daily operations. Changing such input parameters would interrupt the operations, which is too costly for the purpose of data collection. However, there are many systems whose input parameters are varying by nature, such as telephone call centers, financial portfolios, etc., and our approach is applicable in their settings.

Second, the predicted system performance in our approach reflects four types of uncertainty:

(i) uncertainty about the mean performance measure of the simulation model,
(ii) uncertainty about simulation errors,
(iii) uncertainty about the model inadequacy, and
(iv) uncertainty about observation errors in the real system.

For illustration, suppose that the real system is a single-server queue and we decide to use an M/M/1 queue as our simulation model. Prior to simulation, we do not know the performance measure of the M/M/1 queue, thereby uncertainty (i). While performing simulation, each simulation sample is random, thereby uncertainty (ii). Because the real system is unknown, even after the simulation we have no knowledge about the difference between the M/M/1 queue and the real system, thereby uncertainty (iii). The real data are generally collected with random errors, thereby uncertainty (iv).

The last two types of uncertainty are missing from the stochastic kriging metamodel. Hence, our approach allows more comprehensive uncertainty analysis so that the decision-making process that follows will be well supported by a more accurate assessment of the real system.

The rest of the paper is organized as follows. Section 2 introduces the new metamodel for simultaneously characterize both the response surface of a simulation model and its model inadequacy relative to the real system. Section 3 derives the best linear unbiased predictor for predicting the performance of the real system at an arbitrary location. Section 4 develops the maximum likelihood estimation for estimating the unknown parameters of the new metamodel. Section 5 provides numerical results and Section 6 concludes. The proofs of the main results are collected in the Appendix.

2 THE METAMODEL

Let $\mathbf{x} = (x_1, \ldots, x_d)^\intercal$ denote the vector of decision variables of a real system. We are interested in charactering an unknown performance measure of the system $\zeta(\mathbf{x})$. For example, in a critical care facility simulation $\mathbf{x}$ may represent the collection of the arrival rate to the facility, the service rates at different care units and the routing probabilities among the units, while ζ could be the steady-state expected number of patients per day that are denied entry to the facility; see Xie, Nelson, and Barton (2014). In a typical simulation study, one builds a valid simulation model and expects that its performance measure $\eta(\mathbf{x})$, which is estimated via simulation, can approximate $\zeta(\mathbf{x})$ reasonably well.

One of our objectives in this paper is to quantify the discrepancy between η and ζ. To that end, we conceptually characterize the model inadequacy via

$$\zeta(\mathbf{x}) = \rho\,\eta(\mathbf{x}) + \delta(\mathbf{x}), \tag{1}$$

where ρ is an unknown regression parameter that measures the fidelity of the simulation model $\eta(\cdot)$ relative to the real system $\zeta(\cdot)$, and $\delta(\cdot)$ is a model inadequacy function which we will model explicitly later. Equation (1) extends the formulation in Kennedy and O'Hagan (2001) that is used to calibrate *deterministic* computer models. The critical difference is that in our context $\eta(\cdot)$ must be computed via stochastic simulation and thus has random simulation errors. This is also the key difference between kriging (Cressie 1993) and stochastic kriging.

Because of our uncertainty about η, we represent it as a random field as follows

$$\eta(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} + M(\mathbf{x}), \tag{2}$$

where $\mathbf{f}(\mathbf{x})$ is a vector of known functions of $\mathbf{x}$, $\boldsymbol{\beta}$ is a vector of unknown parameters having the same dimension as $\mathbf{f}(\mathbf{x})$, and $M(\cdot)$ is a zero-mean Gaussian random field with covariance $\Sigma_M(\mathbf{x},\mathbf{x}') \coloneqq \mathrm{Cov}[M(\mathbf{x}), M(\mathbf{x}')]$. Likewise, we represent the unknown δ as

$$\delta(\mathbf{x}) = \mathbf{g}(\mathbf{x})^\top \boldsymbol{\gamma} + W(\mathbf{x}), \tag{3}$$

where $\mathbf{g}(\mathbf{x})$ is a vector of known functions of $\mathbf{x}$, $\boldsymbol{\gamma}$ is a vector of unknown parameters having the same dimension as $\mathbf{g}(\mathbf{x})$, and $W(\cdot)$ is a zero-mean Gaussian random field with covariance $\Sigma_W(\mathbf{x},\mathbf{x}') \coloneqq \mathrm{Cov}(W(\mathbf{x}), W(\mathbf{x}'))$.

The functions $\mathbf{f}$ and $\mathbf{g}$ may be chosen from basis functions such as polynomial basis functions or radial basis functions; see, e.g., Rasmussen and Williams (2006). From a Bayesian perspective, $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta}$ reflects one's prior belief on $\eta(\mathbf{x})$. A common choice in practice is to take $\mathbf{f}(\mathbf{x}) \equiv 1$ and $\mathbf{g}(\mathbf{x}) \equiv 1$, meaning that one has no prior knowledge on the structure of $\eta(\mathbf{x})$ or $\delta(\mathbf{x})$, although we do not necessarily assume this.

In light of (2) and (3), the formulation (1) can be rewritten as

$$\zeta(\mathbf{x}) = \rho\,[\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} + M(\mathbf{x})] + \mathbf{g}(\mathbf{x})^\top \boldsymbol{\gamma} + W(\mathbf{x}). \tag{4}$$

Existence of the two random fields suggests that two types of data are relevant in order to apply the metamodel (4) in practice. Specifically, simulation outputs are used to quantify the metamodel (2), while real data (e.g. physical observations of the real system), in conjunction with simulation outputs, are needed to quantify (3).

We assume that the real system is observed at a set of locations $(\mathbf{x}_i : i = 1,\ldots,k)$ with corresponding observations $(z_i : i = 1,\ldots,k)$ subject to zero-mean observation errors $(e(\mathbf{x}_i) : i = 1,\ldots,k)$, namely,

$$z_i = \zeta(\mathbf{x}_i) + e(\mathbf{x}_i), \quad i = 1,\ldots,k.$$

Moreover, let $y_j(\mathbf{x})$ denote the simulation output on replication j at location $\mathbf{x}$. Then the relationship between $y_j(\cdot)$ and $\eta(\cdot)$ can be expressed as

$$y_j(\mathbf{x}) = \eta(\mathbf{x}) + \varepsilon_j(\mathbf{x}), \tag{5}$$

where $(\varepsilon_j(\mathbf{x}) : j = 1,2,\ldots)$ are the zero-mean simulation errors at location $\mathbf{x}$. Further, we assume that simulation is run at locations $(\mathbf{x}_i : i = 1,\ldots,l)$ with $l \geq k$. This implies that we run simulation at least at the locations where the real system is observed and possibly at more locations. This is a reasonable assumption since running simulation is often considered to be cheaper than collecting real data. For $i = 1,\ldots,l$, let

$$\overline{y}(\mathbf{x}_i) \coloneqq \frac{1}{n_i}\sum_{j=1}^{n_i} y_j(\mathbf{x}_i) \quad \text{and} \quad \overline{\varepsilon}(\mathbf{x}_i) \coloneqq \frac{1}{n_i}\sum_{j=1}^{n_i} \varepsilon_j(\mathbf{x}_i),$$

where n_i is the number of simulation replications at $\mathbf{x}_i$.

We are interested in predicting the response $\zeta(\mathbf{x}_0)$ at an arbitrary point $\mathbf{x}_0$ based on the two data sets, i.e. the simulation outputs $\overline{\mathbf{y}} := (\overline{y}(\mathbf{x}_1),\ldots,\overline{y}(\mathbf{x}_k))^\intercal$ and the real data $\mathbf{z} := (z_1,\ldots,z_l)^\intercal$.

Before concluding this section, we highlight the key differences between the metamodel (4) and the stochastic kriging metamodel. First of all, the stochastic kriging metamodel attempts to characterize the simulation model $\eta(\mathbf{x})$ in the form of (2). It is thus part of our metamodel. Second, the stochastic kriging metamodel predicts $\eta(\mathbf{x}_0)$ by interpolating the simulation outputs $\overline{\mathbf{y}}$. By contrast, the metamodel (4) predicts $\zeta(\mathbf{x}_0)$ by interpolating the augmented data set $(\overline{\mathbf{y}}^\intercal,\mathbf{z}^\intercal)$, which will be shown in the next section.

3 BEST LINEAR UNBIASED PREDICTOR

We now derive a predictor for the response $\zeta(\mathbf{x}_0)$ given the data $(\overline{\mathbf{y}},\mathbf{z})$. We consider a linear predictor of the form

$$a(\mathbf{x}_0)+\mathbf{b}(\mathbf{x}_0)^\intercal\overline{\mathbf{y}}+\mathbf{c}(\mathbf{x}_0)^\intercal\mathbf{z}, \tag{6}$$

where $a(\mathbf{x}_0)\in\mathbb{R}$, $\mathbf{b}(\mathbf{x}_0)\in\mathbb{R}_l$, and $\mathbf{c}(\mathbf{x}_0)\in\mathbb{R}_k$ are weights that depend on $\mathbf{x}_0$ and are chosen to minimize the mean squared error (MSE) for predicting $\zeta(\mathbf{x}_0)$.

To facilitate the presentation of our main result, we define the following notations. Let $\mathbf{M}(l) = (M(\mathbf{x}_1),\ldots,M(\mathbf{x}_l))^\intercal, \overline{\boldsymbol{\varepsilon}}(l) = (\overline{\varepsilon}(\mathbf{x}_1),\ldots,\overline{\varepsilon}(\mathbf{x}_l)^\intercal, \mathbf{W}(k) = (W(\mathbf{x}_1),\ldots,W(\mathbf{x}_k))^\intercal$, and $\mathbf{e}(k) = (e(\mathbf{x}_1),\ldots,e(\mathbf{x}_k))^\intercal$. Let $\Sigma_{\mathbf{M}(l)}$ denote the covariance matrix of $\mathbf{M}(l)$. Likewise, we define $\Sigma_{\overline{\boldsymbol{\varepsilon}}(l)}$, $\Sigma_{\mathbf{W}(k)}$, and $\Sigma_{\mathbf{e}(k)}$. Let $\Sigma_{\mathbf{M}(l),\mathbf{M}(k)} := (\mathrm{Cov}[M(\mathbf{x}_i),M(\mathbf{x}_j)])_{1\le i\le l,1\le j\le k}$ denote the covariance matrix between $\mathbf{M}(l)$ and $\mathbf{M}(k)$. Moreover, let $\Sigma_{\mathbf{M}(l)}(\mathbf{x}_0,\cdot)$ denote the vector $(\mathrm{Cov}[M(\mathbf{x}_0),M(\mathbf{x}_1)],\ldots,\mathrm{Cov}[M(\mathbf{x}_0),M(\mathbf{x}_l)])^\intercal$ and define $\Sigma_{\mathbf{W}(k)}(\mathbf{x}_0,\cdot)$ likewise. Finally, let $\mathbf{F}(l)$ denote the matrix with rows $\mathbf{f}(\mathbf{x}_1)^\intercal,\ldots,\mathbf{f}(\mathbf{x}_l)^\intercal$, and $\mathbf{G}(k)$ denote the matrix with rows $\mathbf{g}(\mathbf{x}_1)^\intercal,\ldots,\mathbf{g}(\mathbf{x}_k)^\intercal$.

We impose the following assumption in the sequel.

Assumption 1 $M(\cdot)$, $W(\cdot)$, $\varepsilon(\cdot)$, and $e(\cdot)$ are mutually independent.

Theorem 1 Under Assumption 1, the best linear unbiased predictor (BLUP) of $\zeta(\mathbf{x}_0)$ that minimizes the MSE of prediction is

$$\rho\,\mathbf{f}(\mathbf{x}_0)^\intercal\boldsymbol{\beta}+\mathbf{g}(\mathbf{x}_0)^\intercal\boldsymbol{\gamma}+\mathbf{C}^\intercal\mathbf{V}^{-1}\left[\begin{pmatrix}\overline{\mathbf{y}}\\ \mathbf{z}\end{pmatrix}-\begin{pmatrix}\mathbf{F}(l) & \mathbf{0}\\ \rho\,\mathbf{F}(k) & \mathbf{G}(k)\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix}\right], \tag{7}$$

where $\mathbf{C}$ and $\mathbf{V}$ are block matrices as follows

$$\mathbf{C} := \begin{pmatrix}\rho\Sigma_{\mathbf{M}(l)}(\mathbf{x}_0,\cdot)\\ \rho^2\Sigma_{\mathbf{M}(k)}(\mathbf{x}_0,\cdot)+\Sigma_{\mathbf{W}(k)}(\mathbf{x}_0,\cdot)\end{pmatrix} \tag{8}$$

and

$$\mathbf{V} := \begin{pmatrix}\Sigma_{\mathbf{M}(l)}+\Sigma_{\overline{\boldsymbol{\varepsilon}}(l)} & \rho\Sigma_{\mathbf{M}(l),\mathbf{M}(k)}\\ \rho\Sigma^\intercal_{\mathbf{M}(l),\mathbf{M}(k)} & \rho^2\Sigma_{\mathbf{M}(k)}+\Sigma_{\mathbf{W}(k)}+\Sigma_{\mathbf{e}(k)}\end{pmatrix}. \tag{9}$$

Moreover, the MSE of the BLUP (7) is

$$\rho^2\Sigma_M(\mathbf{x}_0,\mathbf{x}_0)+\Sigma_W(\mathbf{x}_0,\mathbf{x}_0)-\mathbf{C}^\intercal\mathbf{V}^{-1}\mathbf{C}.$$

We remark here that if there were no simulation errors, $\Sigma_{\overline{\boldsymbol{\varepsilon}}(l)}$ would disappear and the BLUP (7) would reduce to the predictor for deterministic computer models as derived in Kennedy and O'Hagan (2001). In addition, (7) has a similar structure as the stochastic kriging predictor for predicting $\eta(\mathbf{x}_0)$,

$$\hat{\eta}(\mathbf{x}_0) = \mathbf{f}(\mathbf{x}_0)^\intercal\boldsymbol{\beta}+\Sigma_{\mathbf{M}(l)}(\mathbf{x}_0,\cdot)[\Sigma_{\mathbf{M}(l)}+\Sigma_{\overline{\boldsymbol{\varepsilon}}(l)}]^{-1}[\overline{\mathbf{y}}-\mathbf{F}(l)\boldsymbol{\beta}];$$

see Ankenman, Nelson, and Staum (2010). Indeed, (7) is the *conditional* expectation of the response given the data, i.e. $\mathbb{E}[\zeta(\mathbf{x}_0)|\overline{\mathbf{y}},\mathbf{z}]$. In particular, $\rho\,\mathbf{f}(\mathbf{x}_0)^\intercal\boldsymbol{\beta}+\mathbf{g}(\mathbf{x}_0)^\intercal\boldsymbol{\gamma}$ represents the *unconditional* expectation of

the response, i.e. $\mathbb{E}[\zeta(\mathbf{x}_0)]$. The last term in (7) reflects the contribution to the predicted value from the correlation between $\zeta(\mathbf{x}_0)$ and the data $(\overline{\mathbf{y}}^\intercal, \mathbf{z}^\intercal)$. Specifically, $\mathbf{C}$ characterizes the covariance between $\zeta(\mathbf{x}_0)$ and $(\overline{\mathbf{y}}^\intercal, \mathbf{z}^\intercal)$, whereas $\mathbf{V}$ characterizes the covariances of within $(\overline{\mathbf{y}}^\intercal, \mathbf{z}^\intercal)$.

4 PARAMETER ESTIMATION

The BLUP (7) was derived under the premise that $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ and any other parameters involved in the covariance matrices are given. However, these parameters are generally unknown in practice. In this section, we therefore focus on parameter estimation, particularly maximum likelihood estimation (MLE). To that end, we adopt the following assumptions.

Assumption 2 The Gaussian random fields $M(\cdot)$ and $W(\cdot)$ are both second-order stationary, i.e. their covariance functions satisfy

$$\begin{aligned} \Sigma_M(\mathbf{x}, \mathbf{x}') &= \tau_M^2 R_M(\mathbf{x} - \mathbf{x}'; \boldsymbol{\theta}_M) \\ \Sigma_W(\mathbf{x}, \mathbf{x}') &= \tau_W^2 R_W(\mathbf{x} - \mathbf{x}'; \boldsymbol{\theta}_W), \end{aligned}$$

where τ_M and τ_W are constants; R_M and R_W are the correlations depending only on $\mathbf{x} - \mathbf{x}'$ and may be functions of some unknown parameters $\boldsymbol{\theta}_M$ and $\boldsymbol{\theta}_W$, respectively. Further, we assume that $R_M(\mathbf{0}; \boldsymbol{\theta}_M) = R_W(\mathbf{0}; \boldsymbol{\theta}_W) = 1$ and

$$\lim_{\|\mathbf{x} - \mathbf{x}'\| \to \infty} R_M(\mathbf{x} - \mathbf{x}'; \boldsymbol{\theta}_M) = \lim_{\|\mathbf{x} - \mathbf{x}'\| \to \infty} R_W(\mathbf{x} - \mathbf{x}'; \boldsymbol{\theta}_W) = 0,$$

where $\|\cdot\|$ denotes the Euclidean norm.

Assumption 3 For each $i = 1, \ldots, l$, the simulation errors $(\varepsilon_j(\mathbf{x}_i) : j = 1, 2, \ldots)$ are i.i.d. normal random variables with mean 0 and variance $V_\varepsilon(\mathbf{x}_i)$, independent of $\varepsilon_{j'}(\mathbf{x}_{i'})$ for all j' and $i' \neq i$; the observation errors $(e_j(\mathbf{x}_i) : j = 1, 2, \ldots)$ are i.i.d. normal random variables with mean 0 and variance $V_e(\mathbf{x}_i)$, independent of $e_{j'}(\mathbf{x}_{i'})$ for all j' and $i' \neq i$.

Then the multivariate normality of $(\overline{\mathbf{y}}^\intercal, \mathbf{z}^\intercal)$ is straightforward. Its mean and covariance matrix will be calculated in the Appendix.

Proposition 1 Let $\mathbf{R}_{\mathbf{M}(l)}(\boldsymbol{\theta}_M)$ denote the correlation matrix of $\mathbf{M}(l)$, $\mathbf{R}_{\mathbf{W}(k)}(\boldsymbol{\theta}_W)$ denote the correlation matrix of $\mathbf{W}(k)$, and $\mathbf{R}_{\mathbf{M}(l),\mathbf{M}(k)}$ denote the correlation matrix between $\mathbf{M}(l)$ and $\mathbf{M}(k)$. Then, under Assumptions 1 - 3, $\begin{pmatrix} \overline{\mathbf{y}} \\ \mathbf{z} \end{pmatrix}$ has multivariate normal distribution with mean $\begin{pmatrix} \mathbf{F}(l)\boldsymbol{\beta} \\ \rho \mathbf{F}(k)\boldsymbol{\beta} + \mathbf{G}(k)\boldsymbol{\gamma} \end{pmatrix}$ and covariance matrix

$$\begin{pmatrix} \tau_M^2 \mathbf{R}_{\mathbf{M}(l)}(\boldsymbol{\theta}_M) + \mathrm{Diag}(\frac{V_\varepsilon(\mathbf{x}_1)}{n_1}, \ldots, \frac{V_\varepsilon(\mathbf{x}_l)}{n_l}) & \rho \tau_M^2 \mathbf{R}_{\mathbf{M}(l),\mathbf{M}(k)}(\boldsymbol{\theta}_M) \\ \rho \tau_M^2 \mathbf{R}^\intercal_{\mathbf{M}(l),\mathbf{M}(k)}(\boldsymbol{\theta}_M) & \rho^2 \tau_M^2 \mathbf{R}_{\mathbf{M}(k)}(\boldsymbol{\theta}_M) + \tau_W^2 \mathbf{R}_{\mathbf{W}(k)}(\boldsymbol{\theta}_W) + \mathrm{Diag}(V_e(\mathbf{x}_1), \ldots, V_e(\mathbf{x}_k)) \end{pmatrix}. \tag{10}$$

Let $\Xi := (\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho, \tau_M^2, \tau_W^2, \boldsymbol{\theta}_M, \boldsymbol{\theta}_W)$ denote the collection of the unknown parameters. Before deriving the MLE for Ξ, we first address the estimation of the variances associated with the simulation errors and the observation errors, namely $\Sigma_{\overline{\boldsymbol{\varepsilon}}(l)} = \mathrm{Diag}(\frac{V_\varepsilon(\mathbf{x}_1)}{n_1}, \ldots, \frac{V_\varepsilon(\mathbf{x}_l)}{n_l})$ and $\Sigma_{\mathbf{e}(k)} = \mathrm{Diag}(V_e(\mathbf{x}_1), \ldots, V_e(\mathbf{x}_k))$, neither of which depends on Ξ.

The covariance matrix (10) is indeed the matrix V in (8) under Assumption 2. To stress its dependence on the unknown parameters, we denote $V = V(\Xi)$, where $\Xi := (\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho, \tau_M^2, \tau_W^2, \boldsymbol{\theta}_M, \boldsymbol{\theta}_W)$. Note that in (10), $\Sigma_{\overline{\boldsymbol{\varepsilon}}(l)} = \mathrm{Diag}(\frac{V_\varepsilon(\mathbf{x}_1)}{n_1}, \ldots, \frac{V_\varepsilon(\mathbf{x}_l)}{n_l})$ and $\Sigma_{\mathbf{e}(k)} = \mathrm{Diag}(V_e(\mathbf{x}_1), \ldots, V_e(\mathbf{x}_k))$, neither of which depends on Ξ.

Similar as the stochastic kriging metamodel, one can estimate $\Sigma_{\overline{\boldsymbol{\varepsilon}}(l)}$ via the sample variance of the simulation outputs, i.e.

$$\hat{V}_\varepsilon(\mathbf{x}_i) = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_j(\mathbf{x}_i) - \overline{y}(\mathbf{x}_i))^2, \quad i = 1, \ldots, l.$$

Moreover, we assume that one has access to an estimate of $\Sigma_{\mathbf{e}(k)}$. This assumption is reasonable in many practical settings, where the observations z_i's are the results of a certain statistical estimation procedure, and thus one should be able to estimate the observation errors. For instance, z_i could be the average waiting time of customers during a fixed time period, in which case one could estimate the observation error associated with z_i via statistical analysis of the waiting times involved.

We now derive the maximum likelihood estimator for Ξ given $\Sigma_{\overline{\boldsymbol{\varepsilon}}(l)}$ and $\Sigma_{\mathbf{e}(k)}$. By Proposition 1, the log likelihood function for Ξ is

$$\mathscr{L}(\Xi) := -\ln\left[(2\pi)^{(l+k)/2}\right] - \frac{1}{2}\ln[|V(\Xi)|] - \frac{1}{2}\begin{pmatrix}\overline{\mathbf{y}}-\mathbf{F}(l)\boldsymbol{\beta}\\ \mathbf{z}-\rho\mathbf{F}(k)\boldsymbol{\beta}+\mathbf{G}(k)\boldsymbol{\gamma}\end{pmatrix}^{\mathsf{T}} V(\Xi)^{-1}\begin{pmatrix}\overline{\mathbf{y}}-\mathbf{F}(l)\boldsymbol{\beta}\\ \mathbf{z}-\rho\mathbf{F}(k)\boldsymbol{\beta}+\mathbf{G}(k)\boldsymbol{\gamma}\end{pmatrix}.$$

By standard results of matrix calculus,

$$\frac{\partial\mathscr{L}(\Xi)}{\partial\boldsymbol{\beta}} = \begin{pmatrix}\mathbf{F}(l)\\ \rho\mathbf{F}(k)\end{pmatrix}^{\mathsf{T}} V(\Xi)^{-1}\begin{pmatrix}\overline{\mathbf{y}}-\mathbf{F}(l)\boldsymbol{\beta}\\ \mathbf{z}-\rho\mathbf{F}(k)\boldsymbol{\beta}+\mathbf{G}(k)\boldsymbol{\gamma}\end{pmatrix}, \tag{11}$$

$$\frac{\partial\mathscr{L}(\Xi)}{\partial\boldsymbol{\gamma}} = \begin{pmatrix}\mathbf{0}\\ -\mathbf{G}(k)\end{pmatrix}^{\mathsf{T}} V(\Xi)^{-1}\begin{pmatrix}\overline{\mathbf{y}}-\mathbf{F}(l)\boldsymbol{\beta}\\ \mathbf{z}-\rho\mathbf{F}(k)\boldsymbol{\beta}+\mathbf{G}(k)\boldsymbol{\gamma}\end{pmatrix}, \tag{12}$$

$$\begin{aligned}\frac{\partial\mathscr{L}(\Xi)}{\partial\rho} &= -\frac{1}{2}\text{trace}\left[V(\Xi)^{-1}\frac{\partial V(\Xi)}{\partial\rho}\right] + \begin{pmatrix}\mathbf{0}\\ \mathbf{F}(k)\boldsymbol{\beta}\end{pmatrix}^{\mathsf{T}} V(\Xi)^{-1}\begin{pmatrix}\overline{\mathbf{y}}-\mathbf{F}(l)\boldsymbol{\beta}\\ \mathbf{z}-\rho\mathbf{F}(k)\boldsymbol{\beta}+\mathbf{G}(k)\boldsymbol{\gamma}\end{pmatrix}\\ &\quad -\frac{1}{2}\begin{pmatrix}\overline{\mathbf{y}}-\mathbf{F}(l)\boldsymbol{\beta}\\ \mathbf{z}-\rho\mathbf{F}(k)\boldsymbol{\beta}+\mathbf{G}(k)\boldsymbol{\gamma}\end{pmatrix}^{\mathsf{T}}\left[-V(\Xi)^{-1}\frac{\partial V(\Xi)}{\partial\rho}V(\Xi)^{-1}\right]\begin{pmatrix}\overline{\mathbf{y}}-\mathbf{F}(l)\boldsymbol{\beta}\\ \mathbf{z}-\rho\mathbf{F}(k)\boldsymbol{\beta}+\mathbf{G}(k)\boldsymbol{\gamma}\end{pmatrix}, \end{aligned}\tag{13}$$

where

$$\frac{\partial V(\Xi)}{\partial\rho} = \begin{pmatrix}\mathbf{0} & \tau_M^2\mathbf{R}_{\mathbf{M}(l),\mathbf{M}(k)}(\boldsymbol{\theta}_M)\\ \tau_M^2\mathbf{R}_{\mathbf{M}(l),\mathbf{M}(k)}(\boldsymbol{\theta}_M)^{\mathsf{T}} & 2\rho\tau_M^2\mathbf{R}_{\mathbf{M}(k)}(\boldsymbol{\theta}_M)\end{pmatrix}.$$

For $\eta = \tau_M^2, \tau_W^2, \boldsymbol{\theta}_{M,p}, \boldsymbol{\theta}_{W,p}$, where $\boldsymbol{\theta}_{M,p}$ and $\boldsymbol{\theta}_{W,p}$ are the p^{th} element of $\boldsymbol{\theta}_M$ and $\boldsymbol{\theta}_W$, respectively,

$$\begin{aligned}\frac{\partial\mathscr{L}(\Xi)}{\partial\eta} &= -\frac{1}{2}\text{trace}\left[V(\Xi)^{-1}\frac{\partial V(\Xi)}{\partial\eta}\right]\\ &\quad -\frac{1}{2}\begin{pmatrix}\overline{\mathbf{y}}-\mathbf{F}(l)\boldsymbol{\beta}\\ \mathbf{z}-\rho\mathbf{F}(k)\boldsymbol{\beta}+\mathbf{G}(k)\boldsymbol{\gamma}\end{pmatrix}^{\mathsf{T}}\left[-V(\Xi)^{-1}\frac{\partial V(\Xi)}{\partial\eta}V(\Xi)^{-1}\right]\begin{pmatrix}\overline{\mathbf{y}}-\mathbf{F}(l)\boldsymbol{\beta}\\ \mathbf{z}-\rho\mathbf{F}(k)\boldsymbol{\beta}+\mathbf{G}(k)\boldsymbol{\gamma}\end{pmatrix}, \end{aligned}\tag{14}$$

in which

$$\begin{aligned}\frac{\partial V(\Xi)}{\partial\tau_M^2} &= \begin{pmatrix}\mathbf{R}_{\mathbf{M}(l)}(\boldsymbol{\theta}_M) & \rho\mathbf{R}_{\mathbf{M}(l),\mathbf{M}(k)}(\boldsymbol{\theta}_M)\\ \rho\mathbf{R}^{\mathsf{T}}_{\mathbf{M}(l),\mathbf{M}(k)}(\boldsymbol{\theta}_M) & \rho^2\mathbf{R}_{\mathbf{M}(k)}(\boldsymbol{\theta}_M)\end{pmatrix},\\ \frac{\partial V(\Xi)}{\partial\tau_W^2} &= \begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_{\mathbf{W}(k)}(\boldsymbol{\theta}_W)\end{pmatrix},\\ \frac{\partial V(\Xi)}{\partial\boldsymbol{\theta}_{M,p}} &= \begin{pmatrix}\tau_M^2\frac{\partial\mathbf{R}_{\mathbf{M}(l)}(\boldsymbol{\theta}_M)}{\partial\boldsymbol{\theta}_{M,p}} & \rho\tau_M^2\frac{\partial\mathbf{R}_{\mathbf{M}(l),\mathbf{M}(k)}(\boldsymbol{\theta}_M)}{\partial\boldsymbol{\theta}_{M,p}}\\ \rho\tau_M^2\frac{\partial\mathbf{R}^{\mathsf{T}}_{\mathbf{M}(l),\mathbf{M}(k)}(\boldsymbol{\theta}_M)}{\partial\boldsymbol{\theta}_{M,p}} & \rho^2\tau_M^2\frac{\partial\mathbf{R}_{\mathbf{M}(k)}(\boldsymbol{\theta}_M)}{\partial\boldsymbol{\theta}_{M,p}}\end{pmatrix},\\ \frac{\partial V(\Xi)}{\partial\boldsymbol{\theta}_{W,p}} &= \begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \tau_W^2\frac{\partial\mathbf{R}_{\mathbf{W}(k)}(\boldsymbol{\theta}_W)}{\partial\boldsymbol{\theta}_{W,p}}\end{pmatrix}.\end{aligned}$$

Then, setting (11) - (14) equal to $\mathbf{0}$ and solving the equations simultaneously produces the MLE for Ξ. We refer to Fang, Li, and Sudjianto (2006) for numerical methods for searching the MLE.

5 NUMERICAL EXPERIMENTS

Consider an M/G/1 queue with arrival rate 1. Assume that its service time has mean $x \in (0,1)$ and variance σ^2. Let $\zeta(x)$ be the steady-state mean waiting time (excluding the service time) of the system. The Pollaczeck-Khinchine formula stipulates that $\zeta(x) = \frac{x^2+\sigma^2}{2(1-x)}$. In our numerical experiments, we attempt to apply the metamodel (4) to fit the above response surface over a large domain for x (we use $x \in [x_L, x_U] = [0.5, 0.95]$).

Now suppose that the service time has the Gamma distribution but due to lack of information, we mistakenly decide to model it using the exponential distribution. Namely, we use an M/M/1 queue with arrival rate 1 and mean service time x as a simulation model to study the real system, i.e. the M/G/1 queue. It is well-known that the steady-state mean waiting time of the M/M/1 queue is $\eta(x) = \frac{x^2}{1-x}$. But in the numerical experiments that follow, we act as if $\eta(x)$ were unknown and estimate it via simulation.

Suppose that observation z is the average waiting time of m customers of the M/G/1 queue in steady state. Let W_h denote the waiting time of the h^{th} customer. Then,

$$\begin{aligned} \mathrm{Var}[z] &= \mathrm{Var}\left[\frac{W_1+\cdots+W_m}{m}\right] = \frac{1}{m}\mathrm{Var}[W_1] + \frac{2}{m}\sum_{h=1}^{m-1}\left(1-\frac{h}{m}\right)\mathrm{Cov}[W_1, W_{1+h}] \\ &\sim \frac{1}{m}\left[\mathrm{Var}[W_1] + \sum_{h=1}^{\infty}\mathrm{Cov}[W_1, W_{1+h}]\right], \end{aligned}$$

as $m \to \infty$, where the summation in the brackets is often called *time average variance constant* (TAVC); see Asmussen and Glynn (2007). To stress its dependence on the mean service time x, we let $H(x)$ to denote the TAVC, i.e. $\mathrm{Var}[z(x)] \sim m^{-1}H(x)$. We show in the Appendix that

$$H(x) = \frac{(x^2+\sigma^2)[x^3(x+2)+x(x^3-4x^2+5x+4)\sigma^2+(2x^2-8x+9)\sigma^4]}{6x^2(1-x)^4}. \tag{15}$$

Note that the following central limit theorem holds:

$$m^{-1/2}(z-\eta(x)) \Rightarrow \mathcal{N}(0, H(x))$$

as $m \to \infty$; see Glynn (1994). This suggests that we can generate observations $(z_i : i = 1,\ldots,k)$ by generating normal random variables with mean $\zeta(x_i)$ and variance $H(x_i)/m$, rather than conducting steady-state simulation of the M/G/1 queue, which would cost substantially more computation time and introduce non-negligible initialization bias. Moreover, by doing so we can easily control the observation errors by adjusting the value of m.

We choose $k = 6$ and make $(x_i : i = 1,\ldots,k)$ evenly spaced, i.e. $x_1 = 0.5$, $x_2 = 0.59$, $x_3 = 0.68$, $x_4 = 0.77$, $x_5 = 0.86$, and $x_6 = 0.95$. For the design points for running simulation, we choose $l = 11$ and make $(x_i : i = 1,\ldots,l)$ evenly spaced as well, i.e. $x_7 = 0.545$, $x_8 = 0.635$, $x_9 = 0.725$, $x_{10} = 0.815$, and $x_{11} = 0.905$. At each of the l design points, for each simulation replication, we simulate the M/M/1 queue initialized with the steady state and set the run length to be 10^4 customers. For the number of replications n_i, we first set the total number of replications $N = \sum_{i=1}^{l} n_i$ and then apply the approach in Ankenman, Nelson, and Staum (2010) to determine each n_i, which aims at minimizing the integrated MSE (IMSE) of the predictor $\hat{\eta}(x)$ relative to $\eta(x)$. (This may not be the optimal strategy for allocating simulation efforts in our setting, which should take into account the observations z_i's and minimize the IMSE of the predictor $\hat{\zeta}(x)$ relative to $\zeta(x)$. This is part of our on-going research.) Moreover, we choose m and N such that $m = 10^4 N$ because we want the simulation errors and the observation errors are about of the same magnitude.

We compare three methods for predicting $\zeta(x_0)$ for an arbitrary point $x_0 \in [x_L, x_U]$.

- Method 1: apply the Gaussian process regression to the real data $\mathbf{z}$, i.e. $z_i = \mathbf{g}(x_i)^\top \boldsymbol{\gamma} + W(x_i) + e(x_i)$

- Method 2: apply the stochastic kriging metamodel to the simulation outputs $\mathbf{y}$ and use $\hat{\eta}(x_0)$ as the predictor for $\zeta(x_0)$
- Method 3: apply the metamodel (4)

In all the three methods, the "trend" functions are taken as constants, i.e. $\mathbf{f}(\mathbf{x}) \equiv 1$ and $\mathbf{g}(x) \equiv 1$; in addition, the correlation functions in Assumption 2 are

$$R_M(x-x';\theta_M) = \exp(-\theta_M|x-x'|) \quad \text{and} \quad R_W(x-x';\theta_W) = \exp(-\theta_W|x-x'|).$$

Clearly, Method 3 uses both the real data and the simulation outputs for prediction, whereas Method 1 uses only the former and Method 2 uses only the latter.

We predict $\zeta(x_0)$ for $K = 200$ different values of x_0 that are equally spaced over $[x_L, x_U]$ using the three methods. For each method, we compute the achieved IMSE, i.e.

$$\int_{x_L}^{x_U} \left[\hat{\zeta}(x_0) - \zeta(x_0)\right]^2 \mathrm{d}x_0 \approx \frac{x_U - x_L}{K} \sum_{i=1}^{K} \left[\hat{\zeta}(x_0^{(i)}) - \zeta(x_0^{(i)})\right]^2,$$

where the superscript (i) denotes the i^{th} point at which a prediction is performed.

In light of the observation errors, which are large if x is close to 1 due to the exploding TAVC, the achieved IMSE may have substantial random variability. Hence, we conduct $T = 2000$ macroreplications of the entire experiment, and compute the average achieved IMSE over the macroreplications, i.e.

$$\overline{\text{IMSE}} = \frac{1}{T} \sum_{t=1}^{T} \int_{x_L}^{x_U} \left[\hat{\zeta}_t(x_0) - \zeta(x_0)\right]^2 \mathrm{d}x_0,$$

where the subscript t denotes the t^{th} macroreplication.

We compare the three methods for two different settings of the M/G/1 queue, one with $\sigma^2 = 2.5$ and the other with $\sigma^2 = 1.1$. Note that the M/G/1 is reduced to the M/M/1 queue if $\sigma^2 = x^2 < 1$. Consequently, the model inadequacy of the simulation model is large in the former setting, whereas small in the latter. Moreover, for a given σ^2, we consider two scenarios with different magnitude of observation errors, one with $m = 10^7$ and the other with $m = 10^6$. Table 1 gives the estimated IMSE in various settings based on 2000 macrreplications of the numerical experiments, while Figure 1 shows the predicted values from an arbitrary representative macroreplication.

Table 1: Estimated IMSE of predicting $\zeta(x)$ using the three methods.

	$\sigma^2 = 2.5$		$\sigma^2 = 1.1$	
	$m = 10^7$, $N = 1000$	$m = 10^6$, $N = 100$	$m = 10^7$, $N = 1000$	$m = 10^6$, $N = 100$
Method 1	1.7303	2.0053	0.6301	0.5760
Method 2	13.4073	13.5421	0.5457	0.5928
Method 3	0.4573	0.9518	0.1508	0.3488

There are several findings from the experiments. First, Method 3 clearly performs the best in all the four cases. This is not surprising since it utilizes more data than either of the other two methods. Second, regardless of the magnitude of its model inadequacy, incorporating a simulation model in a prediction procedure can significantly improve the prediction accuracy than simply relying on physical observations of the real system, since the simulation outputs essentially provide additional information on the structure of the unknown response surface. Third, the magnitude of observation errors has a considerable impact on the prediction accuracy of our method. For instance, in the case where $\sigma = 1.1$ and $m = 10^6$, namely when the magnitude of observation errors is large and the model inadequacy is small, the prediction accuracy of the three methods are comparable with one another.

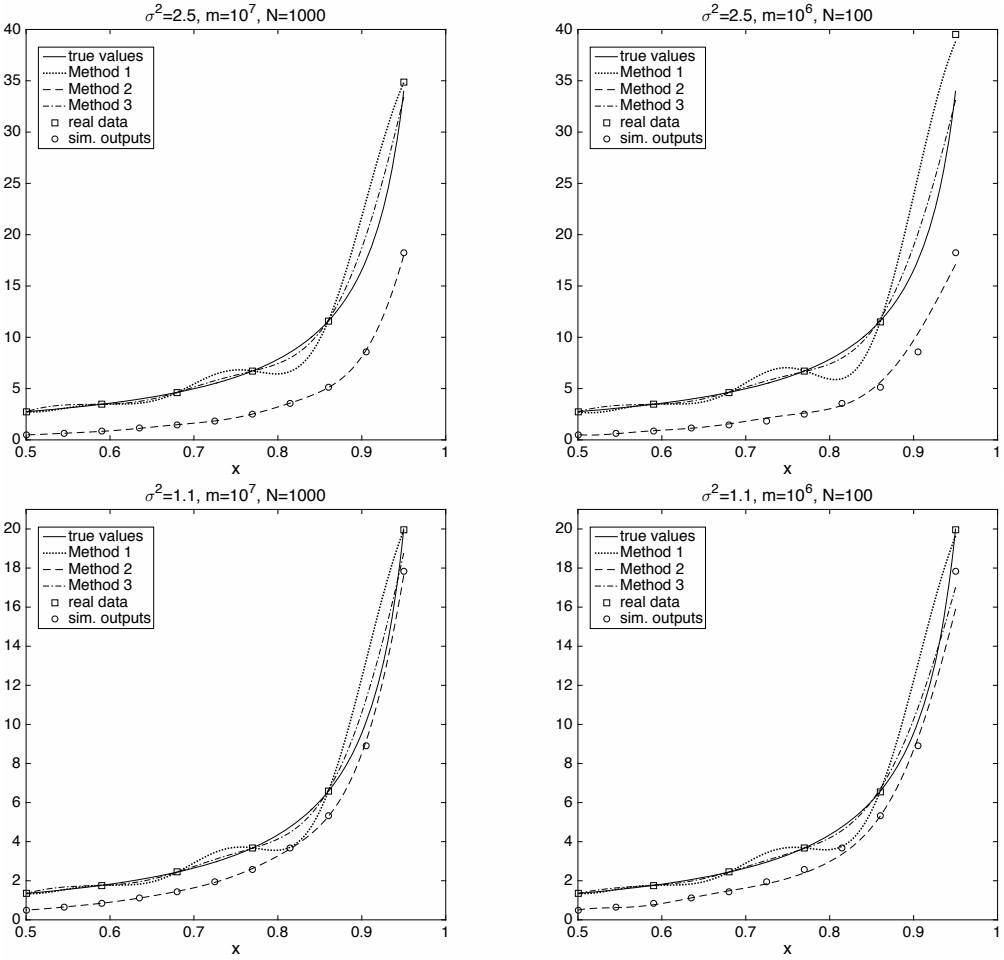

Figure 1: Predicted values of $\zeta(x)$ using the three methods.

6 CONCLUDING REMARKS

We have introduced a new simulation metamodel to simultaneously characterize both the response surface of a simulation model and its model inadequacy relative to the real system. This new metamodel extends the popular stochastic kriging metamodel by utilizing both simulation outputs and real data to predict the performance of the real system. The preliminary numerical experiments show that the new metamodel is promising and provides more accurate predictions. However, there are several important questions that are yet not addressed in this paper. For instance, given observations of the real system at locations $(\mathbf{x}_i : i = 1,\ldots,k)$, how should one select design points $(\mathbf{x}_i : i = 1,\ldots,l)$ at which simulation is run? Is it always a good choice to include $(\mathbf{x}_i : i = 1,\ldots,k)$ in the design points for simulation? Given a group of design points for simulation, how many replications should be run at each point? We leave the investigation of these issues for future research.

ACKNOWLEDGMENTS

The research is partially supported by the Hong Kong Research Grants Council under General Research Fund Project No. 624112.

A PROOF OF THEOREM 1

Consider the linear predictor (6) of $\zeta(\mathbf{x}_0)$

$$\hat{\zeta}(\mathbf{x}_0) = a + \mathbf{b}^\top \bar{\mathbf{y}} + \mathbf{c}^\top \mathbf{z},$$

where we have suppressed the dependence of a, $\mathbf{b}$, and $\mathbf{c}$ on $\mathbf{x}_0$ for notational simplicity. We first calculate its MSE. To that end, note that by (4),

$$\begin{aligned}
\zeta(\mathbf{x}_0) - (a + \mathbf{b}^\top \bar{\mathbf{y}} + \mathbf{c}^\top \mathbf{z}) &= \rho\left[\mathbf{f}(\mathbf{x}_0)^\top \boldsymbol{\beta} + M(\mathbf{x}_0)\right] + \mathbf{g}(\mathbf{x}_0)^\top \boldsymbol{\gamma} + W(\mathbf{x}_0) - [a + \mathbf{b}^\top \bar{\mathbf{y}} + \mathbf{c}^\top \mathbf{z}] \\
&= \underbrace{[\rho \mathbf{f}(\mathbf{x}_0)^\top \boldsymbol{\beta} + \mathbf{g}(\mathbf{x}_0)^\top \boldsymbol{\gamma} - a - \mathbf{b}^\top \mathbf{F}(l)\boldsymbol{\beta} - \mathbf{c}^\top \mathbf{G}(k)\boldsymbol{\gamma} - \rho \mathbf{c}^\top \mathbf{F}(k)\boldsymbol{\beta}]}_{:=I_1} \\
&\quad + \underbrace{[\rho M(\mathbf{x}_0) - \mathbf{b}^\top \bar{\mathbf{y}} + \mathbf{b}^\top \mathbf{F}(l)\boldsymbol{\beta}]}_{:=I_2} + \underbrace{[W(\mathbf{x}_0) - \mathbf{c}^\top \mathbf{z} + \mathbf{c}^\top \mathbf{G}(k)\boldsymbol{\gamma} + \rho \mathbf{c}^\top \mathbf{F}(k)\boldsymbol{\beta}]}_{:=I_3}.
\end{aligned}$$

Hence,

$$\mathbb{E}\left[\left[\zeta(\mathbf{x}_0) - (a + \mathbf{b}^\top \bar{\mathbf{y}} + \mathbf{c}^\top \mathbf{z})\right]^2\right] = \sum_{i=1}^{3} \mathbb{E}[I_i^2] + 2 \sum_{1 \le i < j \le 3} \mathbb{E}[I_i I_j] \tag{16}$$

By (2) and (5), we can write

$$\bar{\mathbf{y}} = \mathbf{F}(l)\boldsymbol{\beta} + \mathbf{M}(l) + \bar{\boldsymbol{\varepsilon}}.$$

Hence,

$$\begin{aligned}
\mathbb{E}[I_2^2] &= \mathbb{E}\left[\left[\rho M(\mathbf{x}_0) - \mathbf{b}^\top (\mathbf{M}(l) + \bar{\boldsymbol{\varepsilon}}(l))\right]^2\right] \\
&= \rho^2 \Sigma_M(\mathbf{x}_0, \mathbf{x}_0) + \mathbf{b}^\top [\Sigma_{\mathbf{M}(l)} + \Sigma_{\bar{\boldsymbol{\varepsilon}}(l)}]\mathbf{b} - 2\rho \mathbf{b}^\top \Sigma_{\mathbf{M}(l)}(\mathbf{x}_0, \cdot),
\end{aligned} \tag{17}$$

since $M(\cdot)$ and $\varepsilon(\cdot)$ are independent. In the same vein,

$$\begin{aligned}
\mathbb{E}[I_3^2] &= \mathbb{E}\left[\left[W(\mathbf{x}_0) - \mathbf{c}^\top (\rho \mathbf{M}(k) + \mathbf{W}(k) + \mathbf{e})\right]^2\right] \\
&= \Sigma_W(\mathbf{x}_0, \mathbf{x}_0) + \mathbf{c}^\top [\rho^2 \Sigma_{\mathbf{M}(k)} + \Sigma_{\mathbf{W}(k)} + \Sigma_{\mathbf{e}(k)}]\mathbf{c} - 2\mathbf{c}^\top \Sigma_{\mathbf{W}(k)}(\mathbf{x}_0, \cdot).
\end{aligned} \tag{18}$$

Moreover,

$$\begin{aligned}
\mathbb{E}[I_2 I_3] &= \mathbb{E}[(\rho M(\mathbf{x}_0) - \mathbf{b}^\top (\mathbf{M}(l) + \bar{\boldsymbol{\varepsilon}}(l)))(W(\mathbf{x}_0) - \mathbf{c}^\top (\rho \mathbf{M}(k) + \mathbf{W}(k) + \mathbf{e}(k)))] \\
&= \mathbb{E}[(\rho M(\mathbf{x}_0) - \mathbf{b}^\top \mathbf{M}(l))(-\rho \mathbf{c}^\top \mathbf{M}(k))] \\
&= -\rho^2 \mathbf{c}^\top \Sigma_{\mathbf{M}(k)}(\mathbf{x}_0, \cdot) + \rho \mathbf{b}^\top \Sigma_{\mathbf{M}(l), \mathbf{M}(k)} \mathbf{c}.
\end{aligned} \tag{19}$$

In addition, note that $\mathbb{E}[I_2] = \mathbb{E}[I_3] = 0$ and that I_1 is deterministic, so $\mathbb{E}[I_1 I_2] = \mathbb{E}[I_1 I_3] = 0$. It then follows from (16) - (19) that the MSE of the linear predictor (6) is a quadratic function in $(a, \mathbf{b}, \mathbf{c})$. Let $h(a, \mathbf{b}, \mathbf{c})$ denote such a function, i.e. $h(a, \mathbf{b}, \mathbf{c})$ is the left hand side of (16). To minimize $h(a, \mathbf{b}, \mathbf{c})$, note that

$$\begin{aligned}
\frac{\partial h}{\partial \mathbf{b}} &= 2[\Sigma_{\mathbf{M}(l)} + \Sigma_{\bar{\boldsymbol{\varepsilon}}(l)}]\mathbf{b} + 2\rho \Sigma_{\mathbf{M}(l), \mathbf{M}(k)} \mathbf{c} - 2\rho \Sigma_{\mathbf{M}(l)}(\mathbf{x}_0, \cdot), \\
\frac{\partial h}{\partial \mathbf{c}} &= 2\rho \Sigma_{\mathbf{M}(l), \mathbf{M}(k)}^\top \mathbf{b} + 2[\rho^2 \Sigma_{\mathbf{M}(k)} + \Sigma_{\mathbf{W}(k)} + \Sigma_{\mathbf{e}(k)}]\mathbf{c} - 2\rho^2 \Sigma_{\mathbf{M}(k)}(\mathbf{x}_0, \cdot) - 2\Sigma_{\mathbf{W}}(k)(\mathbf{x}_0, \cdot).
\end{aligned}$$

Setting $\frac{\partial h}{\partial \mathbf{b}} = \frac{\partial h}{\partial \mathbf{c}} = \mathbf{0}$ yields

$$\begin{pmatrix} \Sigma_{\mathbf{M}(l)} + \Sigma_{\bar{\boldsymbol{\varepsilon}}(l)} & \rho \Sigma_{\mathbf{M}(l), \mathbf{M}(k)} \\ \rho \Sigma_{\mathbf{M}(l), \mathbf{M}(k)}^\top & \rho^2 \Sigma_{\mathbf{M}(k)} + \Sigma_{\mathbf{W}(k)} + \Sigma_{\mathbf{e}(k)} \end{pmatrix} \begin{pmatrix} \mathbf{b}_* \\ \mathbf{c}_* \end{pmatrix} = \begin{pmatrix} \rho \Sigma_{\mathbf{M}(l)}(\mathbf{x}_0, \cdot) \\ \rho^2 \Sigma_{\mathbf{M}(k)}(\mathbf{x}_0, \cdot) + \Sigma_{\mathbf{W}(k)}(\mathbf{x}_0, \cdot) \end{pmatrix},$$

where $(\mathbf{b}_*, \mathbf{c}_*)$ denotes the optimal value. So

$$\begin{pmatrix}\mathbf{b}_*\\ \mathbf{c}_*\end{pmatrix} = \begin{pmatrix}\Sigma_{\mathbf{M}(l)} + \Sigma_{\overline{\boldsymbol{\varepsilon}}(l)} & \rho\Sigma_{\mathbf{M}(l),\mathbf{M}(k)}\\ \rho\Sigma^{\intercal}_{\mathbf{M}(l),\mathbf{M}(k)} & \rho^2\Sigma_{\mathbf{M}(k)} + \Sigma_{\mathbf{W}(k)} + \Sigma_{\mathbf{e}(k)}\end{pmatrix}^{-1} \begin{pmatrix}\rho\Sigma_{\mathbf{M}(l)}(\mathbf{x}_0,\cdot)\\ \rho^2\Sigma_{\mathbf{M}(k)}(\mathbf{x}_0,\cdot) + \Sigma_{\mathbf{W}(k)}(\mathbf{x}_0,\cdot)\end{pmatrix} = \mathbf{V}^{-1}\mathbf{C}. \quad (20)$$

Setting $\frac{\partial h}{\partial a} = 0$ yields

$$a_* = \rho\mathbf{f}(\mathbf{x}_0)^{\intercal}\boldsymbol{\beta} + \mathbf{g}(\mathbf{x}_0)^{\intercal}\boldsymbol{\gamma} - \mathbf{b}_*^{\intercal}\mathbf{F}(l)\boldsymbol{\beta} - \mathbf{c}_*^{\intercal}\mathbf{G}(k)\boldsymbol{\gamma} - \rho\mathbf{c}_*^{\intercal}\mathbf{F}(k)\boldsymbol{\beta}. \quad (21)$$

Therefore, the MSE-optimal predictor $\hat{\zeta}(\mathbf{x}_0)$ is

$$a_* + \mathbf{b}_*^{\intercal}\overline{\mathbf{y}} + \mathbf{c}^{\intercal}\mathbf{z} = \rho\mathbf{f}(\mathbf{x}_0)^{\intercal}\boldsymbol{\beta} + \mathbf{g}(\mathbf{x}_0)^{\intercal}\boldsymbol{\gamma} + \begin{pmatrix}\mathbf{b}_*\\ \mathbf{c}_*\end{pmatrix}^{\intercal}\left[\begin{pmatrix}\overline{\mathbf{y}}\\ \mathbf{z}\end{pmatrix} - \begin{pmatrix}\mathbf{F}(l) & \mathbf{0}\\ \rho\,\mathbf{F}(k) & \mathbf{G}(k)\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix}\right],$$

proving (7).

The unbiasedness of the linear predictor follows immediately from (16) and (21), because

$$\mathbb{E}[\zeta(\mathbf{x}_0) - (a_* + \mathbf{b}_*^{\intercal}\overline{\mathbf{y}} + \mathbf{c}_*^{\intercal}\mathbf{z})] = \rho\mathbf{f}(\mathbf{x}_0)^{\intercal}\boldsymbol{\beta} + \mathbf{g}(\mathbf{x}_0)^{\intercal}\boldsymbol{\gamma} - a_* - \mathbf{b}_*^{\intercal}\mathbf{F}(l)\boldsymbol{\beta} - \mathbf{c}_*^{\intercal}\mathbf{G}(k)\boldsymbol{\gamma} - \rho\mathbf{c}_*^{\intercal}\mathbf{F}(k)\boldsymbol{\beta} = 0.$$

Finally, by direct substitution, the optimal MSE is

$$\begin{aligned}
& \mathbb{E}[\zeta(\mathbf{x}_0) - (a_* + \mathbf{b}_*^{\intercal}\overline{\mathbf{y}} + \mathbf{c}_*^{\intercal}\mathbf{z})]\\
= \;& \mathbb{E}[I_2^2] + \mathbb{E}[I_3^2] + 2\mathbb{E}[I_2 I_3]\\
= \;& \rho^2\Sigma_M(\mathbf{x}_0,\mathbf{x}_0) + \Sigma_W(\mathbf{x}_0,\mathbf{x}_0) + \mathbf{b}_*^{\intercal}[\Sigma_{\mathbf{M}(l)} + \Sigma_{\overline{\boldsymbol{\varepsilon}}(l)}]^{-1}\mathbf{b}_* + \mathbf{c}_*^{\intercal}[\rho^2\Sigma_{\mathbf{M}(k)} + \Sigma_{\mathbf{W}(k)} + \Sigma_{\mathbf{e}(k)}]\mathbf{c}_*\\
& + 2\rho\mathbf{b}_*^{\intercal}\Sigma_{\mathbf{M}(l),\mathbf{M}(k)}\mathbf{c}_* - 2\rho\mathbf{b}_*\Sigma_{\mathbf{M}(l)}(\mathbf{x}_0,\cdot) - 2\rho^2\mathbf{c}_*^{\intercal}\Sigma_{\mathbf{M}(k)}(\mathbf{x}_0,\cdot) - 2\mathbf{c}_*^{\intercal}\Sigma_{\mathbf{W}(k)}(\mathbf{x}_0,\cdot)\\
= \;& \rho^2\Sigma_M(\mathbf{x}_0,\mathbf{x}_0) + \Sigma_W(\mathbf{x}_0,\mathbf{x}_0) - \mathbf{C}^{-1}\mathbf{V}\mathbf{C},
\end{aligned}$$

where the last equality follows from (20).

B PROOF OF PROPOSITION 1

The multivariate normality of $(\overline{\mathbf{y}}^{\intercal}, \mathbf{z}^{\intercal})$ is a direct result of the Gaussian assumptions on both the random fields (i.e. $M(\cdot)$ and $W(\cdot)$) and the noises (i.e. $\varepsilon(\cdot)$ and $e(\cdot)$). The mean of $(\overline{\mathbf{y}}^{\intercal}, \mathbf{z}^{\intercal})$ is also straightforward to calculate. We therefore focus on calculating its covariance matrix.

The calculation of the covariance matrix of $\overline{\mathbf{y}}$ follows the approach in the electronic companion of Ankenman, Nelson, and Staum (2010) but we include it in order that the proof be self-contained. Note that

$$\mathrm{Cov}[y_j(\mathbf{x}_i), y_{j'}(\mathbf{x}_{i'})] = \mathrm{Cov}[M(\mathbf{x}_i) + \varepsilon_j(\mathbf{x}_i), M(\mathbf{x}_{i'}) + \varepsilon_{j'}(\mathbf{x}_{i'})] = \begin{cases}\tau_M^2 + V_\varepsilon(\mathbf{x}_i), & i = i', j = j';\\ \tau_M^2, & i = i', j \neq j';\\ \tau_M^2 R_M(\mathbf{x}_i - \mathbf{x}_{i'}; \boldsymbol{\theta}_M), & i \neq i'.\end{cases}$$

Since $\varepsilon_j(\mathbf{x}_i)$'s are mutually independent for all i and j, it is easy to see that

$$\mathrm{Cov}[\overline{y}(\mathbf{x}_i), \overline{y}(\mathbf{x}_{i'})] = \begin{cases}\tau_M^2 + V_\varepsilon(\mathbf{x}_i)/n_i, & i = i';\\ \tau_M^2 R_M(\mathbf{x}_i - \mathbf{x}_{i'}; \boldsymbol{\theta}_M), & i \neq i'.\end{cases}$$

The mutual independence among $M(\cdot)$, $W(\cdot)$, $\varepsilon(\cdot)$, and $e(\cdot)$ implies that

$$\mathrm{Cov}[\overline{y}(\mathbf{x}_i), z_{i'}] = \mathrm{Cov}[M(\mathbf{x}_i) + \overline{\varepsilon}(\mathbf{x}_i), \rho M(\mathbf{x}_{i'}) + W(\mathbf{x}_{i'}) + e(\mathbf{x}_{i'})] = \rho\tau_M^2 R_M(\mathbf{x}_i - \mathbf{x}_{i'}; \boldsymbol{\theta}_M),$$

and

$$\begin{aligned}
\mathrm{Cov}[z_i, z_{i'}] &= \mathrm{Cov}[\rho M(\mathbf{x}_i) + W(\mathbf{x}_i) + e(\mathbf{x}_i), \rho M(\mathbf{x}_{i'}) + W(\mathbf{x}_{i'}) + e(\mathbf{x}_{i'})]\\
&= \begin{cases}\rho^2\tau_M^2 + \tau_W^2 + V_e(\mathbf{x}_i), & i = i';\\ \rho^2\tau_M^2 R_M(\mathbf{x}_i - \mathbf{x}_{i'}; \boldsymbol{\theta}_M) + \tau_W^2 R_W(\mathbf{x}_i - \mathbf{x}_{i'}; \boldsymbol{\theta}_W), & i \neq i'.\end{cases}
\end{aligned}$$

Hence, the covariance matrix of $(\overline{\mathbf{y}}^{\intercal}, \mathbf{z}^{\intercal})$ is indeed given by (10).

C VARIANCE OF AVERAGE WAITING TIME IN AN M/G/1 QUEUE

Assume that the service time distribution is Gamma with shape parameter α and rate parameter β. It is shown in Blomqvist (1967) that in steady state,

$$\begin{aligned}
\mathrm{Var}[W_1] &= \frac{(\alpha+1)\rho}{12\mu^2\alpha^2(1-\rho)^2}[4(1-\rho)(\alpha+2)+3\rho(\alpha+1)] \\
\sum_{h=0}^{\infty}\mathrm{Cov}[W_1,W_{1+h}] &= \frac{(\alpha+1)\rho}{24\mu^2\alpha^3(1-\rho)^4}[8\alpha(\alpha+2)-\rho(\alpha-1)(7\alpha+18) \\
&\qquad +2\rho^2(\alpha-1)(3\alpha+8)-\rho^3(\alpha-1)(\alpha+4)],
\end{aligned}$$

where $\mu = 1/x$ is the service rate and $\rho = \lambda/\mu$ is the utilization factor. Note that $\alpha = x^2/\sigma^2$ and $\beta = x/\sigma^2$, since the service time has mean x and variance σ^2. With $\lambda = 1$, direct calculation yields that the TAVC is given by (15).

REFERENCES

Ankenman, B., B. L. Nelson, and J. Staum. 2010. "Stochastic Kriging for Simulation Metamodeling". *Operations Research* 58 (2): 371–382.

Asmussen, S., and P. W. Glynn. 2007. *Stochastic Simulation: Algorithm and Analysis*. Springer-Verlag.

Barton, R. R., and M. Meckesheimer. 2006. "Metamodel-Based Simulation Optimization". In *Handbooks in Operations Research and Management Science*, edited by S. Henderson and B. Nelson, Volume 18. Elsevier.

Blomqvist, N. 1967. "The Covariance Function of the M/G/1 Queuing System". *Scandinavian Actuarial Journal* 1967 (3-4): 157–174.

Cressie, N. A. C. 1993. *Statistics for Spatial Data*. Wiley, New York.

Fang, K.-T., R. Li, and A. Sudjianto. 2006. *Design and Modeling for Computer Experiments*. Chapman & Hall/CRC.

Glynn, P. W. 1994. "Poisson's Equation for the Recurrent M/G/1 Queue". *Advances in Applied Probability* 26 (4): 1044–1062.

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

Rasmussen, C. E., and C. K. I. Williams. 2006. *Gaussian Processes for Machine Learning*. MIT Press.

Sargent, R. G. 1999. "Validation and Verification of Simulation Models". In *Proceedings of the 1999 Winter Simulation Conference*, edited by P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, 39–48. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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.

AUTHOR BIOGRAPHIES

XIAOWEI ZHANG is an Assistant Professor in the Department of Industrial Engineering and Logistics Management at the Hong Kong University of Science and Technology. He received his Ph.D. in Operations Research from Stanford University in 2011. He is a member of INFORMS and his research interests include input uncertainty, simulation optimization, rare-event simulation, and stochastic modeling in service engineering. His email address is xiaoweiz@ust.hk.

LU ZOU is a PhD candidate in the Department of Industrial Engineering and Logistics Management at The Hong Kong University of Science and Technology. She received her Bachelor in Mathematics from Nankai University in 2012. Her research interests include simulation metamodeling and input uncertainty. Her email address is lzou@connect.ust.hk.