

A GAUSSIAN PROCESS BASED ALGORITHM FOR STOCHASTIC SIMULATION OPTIMIZATION WITH INPUT DISTRIBUTION UNCERTAINTY

Haowei Wang
Szu Hui Ng
Xun Zhang

Department of Industrial Systems Engineering and Management
National University of Singapore
10 Kent Ridge Crescent
Singapore 119260, SINGAPORE

ABSTRACT

Stochastic simulation models are increasingly popular for analyzing complex stochastic systems. However, the input distributions driving the simulation models are typically unknown in practice and are usually estimated from real world data. Since the size of real world data tends to be limited, the resulting estimation of input distribution will contain errors. This estimation error is commonly known as input uncertainty. In this paper, we consider the stochastic simulation optimization problem when the input uncertainty is present and assume that both the family and parameters of the input distribution are unknown. Traditional efficient metamodel-based optimization approaches like Efficient Global Optimization (EGO) do not take the input uncertainty into account. This can lead to sub-optimal decisions when the input uncertainty level is high. Here, we adopt a nonparametric Bayesian approach to model the input uncertainty and propose an EGO-based simulation optimization algorithm that explicitly accounts for the input uncertainty.

1 INTRODUCTION

In many operational systems such as manufacturing, service and financial systems, the decision makers are typically required to make decisions to optimize some system performances. As the dimension and complexity of the systems increase, direct experiments conducted on real systems can be quite expensive. Simulation models are increasingly being used to study these complex systems because they are cheaper and less risky to implement. Simulation experiments enable users to choose and try different experiment settings to analyze the system performance, and is also widely used for optimizing these systems. This optimization via the simulation approach, often known as simulation optimization, aims to find the best settings that optimize some simulation output values. This in turn can guide the users on the optimal decisions in real systems. The classical stochastic simulation optimization (minimization) problem is formulated as:

$$\min_{x \in \mathcal{X}} \quad \mathbb{E}_{\xi \sim P^c} [h(x, \xi)], \quad (1)$$

where $x \in \mathbb{R}^d$ is the decision variable that we aim to minimize with a continuous box-constraint design space \mathcal{X} , and ξ is a random variable that accounts for the random effects of the system. h is the simulation output that depends on both x and ξ , and h is stochastic due to the randomness of ξ . For example, in an inventory problem, the $h(x, \xi)$ can be the simulated steady-state cost, which is stochastic as it depends on the random demand ξ . The expectation in (1) is taken with respect to the distribution of ξ , i.e. $\xi \sim P^c$. This P^c is usually referred to as the true input distribution (Zhou and Xie 2015; Wu et al. 2018).

In practice, however, the true input distribution P^c that drives the simulation run is unknown and can only be estimated from finite real world data. For example, in the inventory problem, the true distribution (input distribution) P^c for the customers' random demand is unknown and can only be estimated from

customers' historical demand. The estimated input distribution \hat{P} is usually used as if it were the true input distribution P^c in the simulation experiments, ignoring the estimation error due to the finite input data. The typical problem solved while ignoring the estimation error is

$$\min_{x \in \mathcal{X}} \quad E_{\xi \sim \hat{P}}[h(x, \xi)]. \quad (2)$$

If the size n of the real world dataset is small, a poorly estimated \hat{P} may result. Then the best decision found by solving (2) can be quite different from that of (1). This uncertainty arising from a finite data estimation is called input uncertainty. Ignoring the input uncertainty can have a non-negligible impact on the system performance evaluation as well as simulation optimization (Lam 2016; Wu et al. 2018).

There has been quite a sizeable of literature on solving the stochastic simulation optimization problem given by (2) without considering input uncertainty. There are three main streams of methods, including metaheuristics (Ólafsson 2006), gradient-based methods (Kushner and Yin 2003) and metamodel based methods (Barton and Meckesheimer 2006). Xu et al. (2015) provides a comprehensive review of these different methods. Metamodels are statistical approximation models of the simulation output response $h(\cdot)$ that require no strict assumptions on h , and they are usually easy and cheap to develop. Besides, metamodel-based methods have been shown to be efficient at locating optimal solutions when the simulation runs are limited. The Gaussian Process (GP) model is a type of metamodel that has received much attention due to its mathematical convenience and modeling flexibility (Kleijnen 2009; Ankenman et al. 2010).

Although most of the simulation optimization methods can only be applied directly to the case when input uncertainty is ignored, some recent research have started to take input uncertainty into account in the simulation optimization problem. Realizing the risk of solving problem (2), Zhou and Xie (2015) considered a simplified case that P^c can be fully characterized by some unknown input parameter λ^c , i.e. $P^c = P_{\lambda^c}$. They then considered the input uncertainty arising from not exactly knowing λ^c in simulation optimization and formulated the Bayesian Risk Optimization (BRO) problem as

$$\min_{x \in \mathcal{X}} \quad E_{\lambda \sim \pi} [E_{\xi \sim P_{\lambda}}[h(x, \xi)]] . \quad (3)$$

From a Bayesian perspective, Zhou and Xie (2015) viewed λ^c as a random variable λ . A prior is posed on λ and a posterior distribution π for λ is then computed conditional on the real world data. This π is used to quantify the uncertainty about the true input parameter λ^c . Zhou and Xie (2015) pointed out that BRO outperforms (2) when the real data size n is small and h is sensitive to λ . Wu et al. (2018) further derived the consistency and asymptotic normality of the BRO solution. Pearce and Branke (2017) and Wang et al. (2019) proposed Gaussian process based algorithms to solve this BRO problem. However, all of these proposed algorithms only consider the case when ξ follows a known distribution family and consider the input uncertainty in the input parameter. For example, in an inventory problem, the random demand ξ is assumed to follow some parametric distribution, e.g. $P^c = \exp(\lambda_c)$, where λ_c is the unknown true input parameter. The assumption of a known distribution family, however, can be too restrictive in practice especially for situations where no prior knowledge about the input distribution model is available.

Without assuming the input distribution family known, modeling input uncertainty in a nonparametric perspective has been proposed and studied over the years within the simulation community (Lam 2016). Existing methods, including direct resampling (Barton and Schruben 1993), bootstrap resampling and uniformly randomized empirical distribution function (Barton and Schruben 2001), and Dirichlet Process Mixtures (Xie et al. 2019), aim to quantify the input uncertainty and construct confidence intervals of the simulation output (Song and Nelson 2015). However, none of them considers the impact of this uncertainty on the results of simulation optimization. To the best of our knowledge, the inclusion of input uncertainty when neither the input distribution family nor the input parameter is known in the stochastic simulation optimization problem has not yet been comprehensively explored. In this paper, we focus on modifying a metamodel-based method, the popular Efficient Global Optimization (EGO) algorithm to efficiently solve the stochastic simulation optimization problem, when the distribution family of P^c is unknown. Other

GP-based algorithms, such as Knowledge Gradient (Frazier et al. 2009) and Informational Approach to Global Optimization (Villemonteix et al. 2009) can also be adjusted to account for input uncertainty. Here, we use EGO as it is straightforward to calculate and is the most commonly used GP-based simulation optimization algorithm. Our proposed algorithm can be viewed as an extension of Wang et al. (2019) whose algorithms can only be applied when the input distribution family is known.

The rest of this paper is organized as follows. Section 2 formulates the Nonparametric Bayesian Risk Optimization (NBRO) problem, and introduces the nonparametric Bayesian approach to modeling the unknown input distribution. Section 3 provides the modified EGO algorithm to solve NBRO. In Section 4, we conduct experiments on an inventory problem. Section 5 concludes and future directions are discussed.

2 NONPARAMETRIC BAYESIAN RISK OPTIMIZATION (NBRO)

In this section, we will extend the BRO problem to the case when the input distribution family is unknown and term the problem as Nonparametric Bayesian Risk Optimization (NBRO) problem.

2.1 Nonparametric Bayesian Risk Optimization (NBRO)

In order to hedge the risk of ignoring input uncertainty in solving (2), instead of using the estimated \hat{P} , we model P^c with a posterior $P \sim \pi$ and aim to solve the following problem:

$$\min_{x \in \mathcal{X}} g(x) = E_{P \sim \pi} [E_{\xi \sim P} [h(x, \xi)]] . \quad (4)$$

The outer expectation in (4) is taken with respect to the posterior of P . The inner expectation is taken with respect to the distribution of ξ . The problem given by (4) is similar to the BRO problem (3) in Wu et al. (2018) where the distribution family for P is assumed known. Different from Wu et al. (2018), we consider a more general case that the distribution family is unknown. As we do not have any parametric assumptions on P , we call problem (4) as Nonparametric Bayesian Risk Optimization (NBRO) problem.

2.2 Nonparametric Bayesian Modeling of the Input Distribution

In the following, we introduce the nonparametric Bayesian approach to modeling the unknown input distribution. Recall P^c is the true input distributions from which the real world data is collected. Denote the real world data set as $D_n = \{\xi_1, \dots, \xi_n\}$.

From the frequentist perspective, the empirical distribution function \hat{P} is used to estimate P^c : $\hat{P}(t) = \frac{1}{n} \sum_{i=1}^n I(\xi_i \leq t)$. However, the empirical distribution function \hat{P} can be rather inaccurate when the real world data size n is limited. An alternative way to estimate P^c is from a Bayesian perspective, where we model P^c as a random variable P . Specifically, a prior $\pi_0(P)$ is first assumed on P which reflects our initial belief on P^c . Then given the real world data D_n , the posterior distribution of P is updated as $\pi(P|D_n)$ which is our current knowledge on P^c after observing D_n . As n gets larger and larger, the posterior distribution $\pi(P|D_n)$ will be more and more concentrated around the true input distribution P^c . When n tends to be infinity, the posterior $\pi(P|D_n)$ will recover P^c (Gelman et al. 2013).

There exist several nonparametric Bayesian models for estimating a distribution, such as the Dirichlet distribution, the Dirichlet Process and the Dirichlet Process Mixtures (Xie et al. 2019). Here, we use the conjugate Dirichlet process prior as it is more flexible than the Dirichlet distribution and more computational efficient than the Dirichlet Process Mixtures (Hjort et al. 2010; Ferguson 1973). The Dirichlet process prior can be parameterized by the concentration parameter α and the base distribution P_0 , denoted as $DP(\alpha, P_0)$. The base distribution P_0 is a distribution that can be thought of as a prior guess for P^c . α controls how tightly concentrated the prior is around P_0 . The larger the concentration parameter, the more concentrated the prior is around P_0 . Therefore, high values of α imply high confidence in P_0 and low values of α represent vague beliefs. The concentration parameter α can be thought of as the number of “initial” data to obtain P_0 . Given the real world data $\xi_1, \dots, \xi_n \sim P^c$ and the chosen Dirichlet process $DP(\alpha, P_0)$ as the prior, we can derive the posterior distribution that is also a Dirichlet process: $\pi(P|\xi_1, \dots, \xi_n) \sim DP(\alpha + n, \frac{\alpha P_0 + \sum_{j=1}^n \delta_{\xi_j}}{\alpha + n})$.

With $P \sim \pi(P|\xi_1, \dots, \xi_n)$, our knowledge about P^c as well as the corresponding input uncertainty can be quantified. Popular sampling method such as the algorithm of the stick-breaking process (Hjort et al. 2010) can be used to efficiently generate the samples from the posterior.

3 AN EFFICIENT APPROACH TO NBRO: A GAUSSIAN PROCESS BASED ALGORITHM

If we write the inner expectation in (4) as $f(x, P) = E_{\xi \sim P}[h(x, \xi)]$, then problem (4) is equivalent to

$$\min_{x \in \mathcal{X}} g(x) = E_{P \sim \pi}[f(x, P)]. \quad (5)$$

Note that $f(x, P)$ can only be estimated through noisy simulation outputs $h(x, \xi)$. In addition, the design space \mathcal{X} is continuous. Hence it is impossible to evaluate at all possible design-distribution points (x, P) to get full information about the surface of $f(x, P)$. Here we focus on the goal of making the best use of the finite simulation runs to solve (5). With limited budget, Gaussian process based algorithms have been shown to be efficient and effective in solving simulation optimization problems (Frazier 2018).

In order to use the GP-based optimization algorithms to solve (5), a GP model for $g(x)$ is required. As we focus on the limited budget case, it is not possible to obtain observations of $g(x)$ from direct simulation methods such as those suggested in Chick (2001), Zouaoui and Wilson (2003) which require a large number of simulations at each design point. Therefore, the simulation output on $g(x)$ is not readily available and only the noisy simulation output on $f(x, P)$ is observable. In this case, we would like to first obtain the GP model $F(x, P)$ for $f(x, P)$, and develop an approximation for $g(x)$. Then we apply this approximation of $g(x)$ in the EGO algorithm to propose an efficient approach to solve problem (5).

3.1 Stochastic GP Model for $f(x, P)$

Let $y(x, P) \triangleq h(x, \xi)$ where $\xi \sim P$ be the stochastic simulation output at (x, P) , and we assume that $y(x, P)$ is a realization of a random process that can be described by the following model:

$$y(x, P) = f(x, P) + \varepsilon,$$

where ε is the stochastic noise that is assumed to follow a normal distribution with mean 0 and finite constant variance, i.e. $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. The expected simulation output $f(x, P) = E[y(x, P)]$ (for the simplicity of notation, the expectation without subscript is with respect to ε hereafter) is further assumed to be modeled as a Gaussian process (GP) F with mean β_0 and stationary covariance function Σ_F , denoted as $F \sim GP(\beta_0, \Sigma_F)$. The covariance function Σ_F is used to measure the spatial correlation of F between any two points (x, P) and (x', P') : $\Sigma_F((x, P), (x', P')) = \text{Cov}(F(x, P), F(x', P')) = \tau^2 R_F((x, P), (x', P'); \theta)$, where R_F is the correlation function, τ and θ are the model parameters.

Suppose the initial sample size is $s \cdot r$, i.e. we can evaluate $f(x, P)$ through simulation at s distinct design-distribution pairs, denote as $\{(x_1, P_1), \dots, (x_s, P_s)\}$. At each (x_i, P_i) , r replicates of the simulation are conducted. Denote the simulation output at (x_i, P_i) at the j th simulation replicate as $y_j(x_i, P_i)$, and the simulation output sample mean is computed as $\bar{y}(x_i, P_i) = \frac{1}{r} \sum_{j=1}^r y_j(x_i, P_i)$. Then denote the available observed simulation output sample mean vector as $\bar{Y}_s = \{\bar{y}(x_1, P_1), \dots, \bar{y}(x_s, P_s)\}$. Conditional on \bar{Y}_s , F can be updated and the posterior distribution $F_s(x, P) \triangleq F|\bar{Y}_s \sim GP(m_s, k_s)$ can be obtained. m_s and k_s are the conditional mean and conditional covariance of $F_s(x, P)$ respectively:

$$m_s(x, P) = \tau^2 R_F((x, P), \cdot; \theta)^T [\tau^2 R_F(\theta) + \Sigma_\varepsilon]^{-1} \bar{Y}_s,$$

$$k_s((x, P), (x', P')) = \tau^2 R_F((x, P), (x', P'); \theta) - \tau^4 R_F((x, P), \cdot; \theta)^T [\tau^2 R_F(\theta) + \Sigma_\varepsilon]^{-1} R_F((x', P'), \cdot; \theta).$$

The conditional mean m_s can be used to approximate $f(x, P)$. $k_s((x, P), (x', P'))$ is the conditional covariance between any two design-distribution pairs, and the conditional variance of F is $k_s((x, P), (x, P))$. Besides, $\tau^2 R_F((x, P), \cdot; \theta)^T = [\text{Cov}[F((x, P)), F((x_1, P_1))], \dots, \text{Cov}[F((x, P)), F((x_s, P_s))]]^T$ is a $s \times 1$ vector,

and $\tau^2 R_F(\theta)$ is the $s \times s$ covariance matrix across the s observed points. $\Sigma_\varepsilon = \text{Diag}\{\frac{1}{r}\sigma_\varepsilon^2 \dots, \frac{1}{r}\sigma_\varepsilon^2\} = \frac{1}{r}\sigma_\varepsilon^2 I$ is an $s \times s$ matrix. The σ_ε^2 can be estimated with pooled variance of the simulation outputs and the parameters β_0 , τ and θ can be estimated with the maximum log-likelihood method (Williams and Rasmussen 2006).

For the stochastic GP model $F_s(x, P)$, we also need to properly choose the correlation function between (x, P) and (x', P') . We assume x and P are independent of each other. Denote the correlation function as $R_F((x, P), (x', P')) = r_{\mathcal{X}}(x, x')r_{\mathcal{M}}(P, P')$ where $r_{\mathcal{X}}(x, x')$ and $r_{\mathcal{M}}(P, P')$ are correlation kernels. The correlation kernels need to be symmetric and positive definite. For $r_{\mathcal{X}}(x, x')$, there are many choices including the squared exponential, Matern correlation kernels etc. In this article, we adopt the most widely-used squared exponential correlation kernels, i.e. $r_{\mathcal{X}}(x, x') = \exp\{-(x - x')^2/2v_1^2\}$ (Williams and Rasmussen 2006). For $r_{\mathcal{M}}(P, P')$, we use the following form: $r_{\mathcal{M}}(P, P') = \exp\{-D^2(P, P')/2v_2^2\}$, where $D^2(P, P')$ is some closeness measure between P and P' that can lead to positive definite correlation kernels.

There are several candidates for the closeness measure, such as the total variation, the squared Hellinger distance, the Jensen-Shannon divergence (these three are f-divergence measures proposed by Hein and Bousquet (2005)) and the quadratic Wasserstein distance (Bachoc et al. 2017). In this paper, we choose the Wasserstein distance over the three f-divergence measures because it can more accurately capture the differences in the characteristics of different distributions. The f-divergence measures only consider the point-wise difference between input distributions, while the Wasserstein distance considers both the mass difference point-wisely and the difference between points. Here we use a simple example to illustrate. Let us consider the three distributions, denoted as A, B, C respectively in Figure 1. The distance between B and A and the distance between C and A are the same in terms of the three f-divergence measures, while the Wasserstein distance between C and A is larger than the Wasserstein distance between B and A. The Wasserstein distance intuitively is more reasonable as the difference of the mean values between C and A ($4 - 2 = 2$) is larger than that between B and A ($2.5 - 2 = 0.5$) and the difference of the median values between C and A ($5 - 1 = 4$) is also larger than that between B and A ($2.5 - 2 = 0.5$). Formally, the quadratic Wasserstein distance is defined as: $D^2(P, P') = \inf_{\tau \in \gamma(P, P')} E_{(z, z') \sim \tau} [|z - z'|^2]$, where $\gamma(P, P')$ is the set of the distribution of (z, z') whose marginal distributions are P and P' . The Wasserstein distance can be calculated efficiently with the algorithm provided by Peyré and Cuturi (2019).

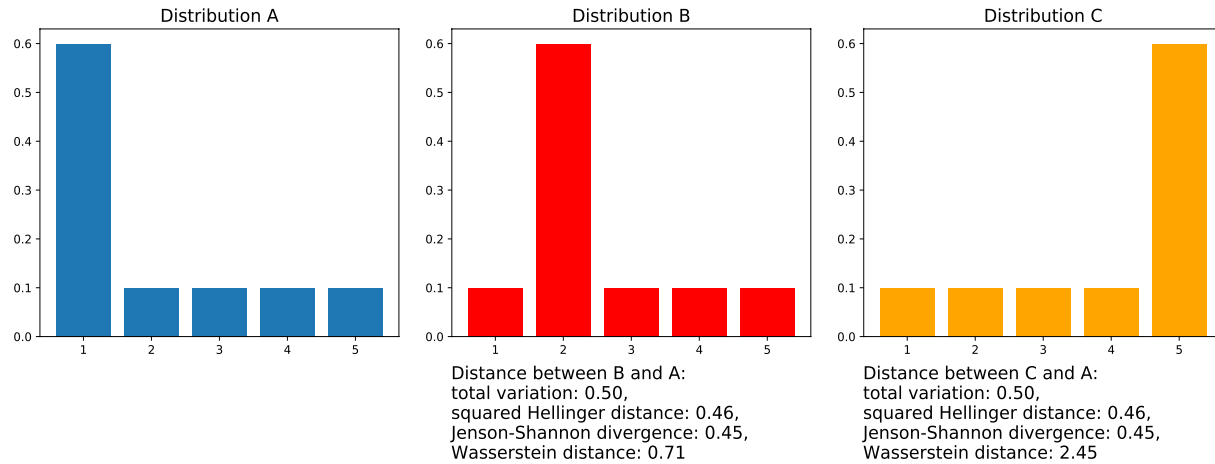


Figure 1: Comparison of Wasserstein distance and the three f-divergence measures.

3.2 Approximation for $g(x)$

With $F_s(x, P)$ obtained as an estimated model for $f(x, P)$, here, we further derive an approximation model for $g(x) = E_{P \sim \pi}[f(x, P)]$. Specifically, we consider the following process:

$$G_s(x) = E_{P \sim \pi}[F_s(x, P)]. \tag{6}$$

By rewriting (6) as the limit of Riemann sums (De Oliveira and Kone 2015), it can be shown that $G_s(x)$ is still a GP. The mean and covariance of $G_s(x)$ can be derived as follows.

$$\mathbb{E}[G_s(x)] = \int_P \mathbb{E}[F_s(x, P)] \cdot \pi(P|D_n) dP = \int_P m_s(x, P) \cdot \pi(P|D_n) dP, \quad (7)$$

$$\text{Cov}[G_s(x), G_s(x')] = \int_P \int_{P'} \pi(P|D_n) \pi(P'|D_n) k_s((x, P), (x', P')) dP' dP. \quad (8)$$

The integration in (7) and (8) can be estimated numerically. Specifically, samples of $\{P_1, \dots, P_{N_{MC}}\}$ can be generated from $\pi(P|D_n)$ to compute:

$$\mathbb{E}[G_s(x)] \approx \mu_s(x) = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} m_s(x, P_i),$$

$$\text{Cov}[G_s(x), G_s(x')] \approx c_s(x, x') = \frac{1}{N_{MC}^2} \sum_{i=1}^{N_{MC}} \sum_{j=1}^{N_{MC}} k_s((x, P_i), (x', P_j)).$$

In fact, we have $\mu_s(x) = \mathbb{E} \left[\frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} F_s(x, P_i) \right]$ and $c_s(x, x') = \text{Cov} \left(\frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} F_s(x, P_i), \frac{1}{N_{MC}} \sum_{j=1}^{N_{MC}} F_s(x', P_j) \right)$.

Furthermore, as

$$\hat{G}_s(x) = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} F_s(x, P_i) \quad (9)$$

is a finite sum of Gaussian random variable, $\hat{G}_s(x)$ is a Gaussian process with mean $\mu_s(x)$ and covariance $c_s(x, x')$. The above $\hat{G}_s(x)$ can then be used to approximate $G_s(x)$ and used as the approximation model for $g(x)$. $\hat{G}_s(x)$ is our current knowledge about the surface of $g(x)$ after the initial evaluations. The MC samples P_i in (9) can be generated from the posterior π with the algorithm of the stick-breaking process.

3.3 EGO with Input Uncertainty to Select (x_{s+1}, P_{s+1})

In this section, we aim to modify a popular global optimization algorithm, the Efficient Global Optimization (Jones et al. 1998), to guide the sequential selection of (x_{s+1}, P_{s+1}) for the next evaluation.

3.3.1 Overview of EGO

EGO was first introduced to solve the deterministic simulation optimization problem when the simulation outputs are deterministic. EGO starts by evaluating at an initial sets of points to estimate a GP model and then takes sequential steps to search the space. The sequential evaluation point is determined by maximizing the Expectation Improvement (EI) criterion based on the GP model. The EI criterion measures the expected improvement of the new function evaluation over the current best evaluations. It not only aims to choose points that can reduce the objective function value most, but also considers the uncertainty at unobserved points, enabling it to balance the search within local areas of current optimizer and unexplored areas away from previously observed points. The optimal is finally determined by the minimum of the simulation output at the already sampled points.

3.3.2 Modified EGO with Input Uncertainty

In order to adapt the EGO for stochastic optimization problem and account for the input uncertainty to solve (5), we need to answer two questions: 1) what is the current best values of $g(x)$ and 2) what is the predictive distribution of $g(x)$ after a new hypothetical evaluation is generated.

Recall that in section 3.2, we have derived a GP model $\hat{G}_s(x)$ for $g(x)$ with the predictive mean $\mu_s(x)$. Here, we estimate the current best values upon which we aim to improve, denoted as T , as the best predictive mean $\mu(x)$ over already sampled design points, i.e. $T = \min\{\mu_s(x_1), \dots, \mu_s(x_s)\}$.

To answer the second question, we need to derive the predictive distribution $\hat{G}_{s+1}(x)$ conditional on evaluation at any arbitrary point (x_{s+1}, P_{s+1}) , which is denoted as $\hat{G}_{s+1}(x|x_{s+1}, P_{s+1})$. Before obtaining the predictive distribution $\hat{G}_{s+1}(x|x_{s+1}, P_{s+1})$, we need to first derive the predictive distribution $F_{s+1}(x, P)$ conditional on evaluation at (x_{s+1}, P_{s+1}) , denoted here as $F_{s+1}((x, P_i)|x_{s+1}, P_{s+1})$. Then we have

$$\hat{G}_{s+1}(x|x_{s+1}, P_{s+1}) = \frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} F_{s+1}((x, P_i)|x_{s+1}, P_{s+1}). \quad (10)$$

Suppose we denote the evaluation value at (x_{s+1}, P_{s+1}) to be y_{s+1} , where $y_{s+1} = y(x_{s+1}, P_{s+1})$ is a sample value of the stochastic simulation output. Conditioning on this output, we can derive the conditional distribution $F_{s+1}((x, P)|x_{s+1}, P_{s+1}, y_{s+1}) \sim GP(m'_{s+1}, k'_{s+1})$ with the following mean and covariance:

$$\begin{aligned} m'_{s+1}(x, P) &= m_s(x, P) + \frac{k_s((x, P), (x_{s+1}, P_{s+1}))}{k_s((x_{s+1}, P_{s+1}), (x_{s+1}, P_{s+1})) + \frac{1}{r}\sigma_\varepsilon^2} [y_{s+1} - m_s(x, P)] \\ &= m_s(x, P) + \frac{k_s((x, P), (x_{s+1}, P_{s+1}))}{\sqrt{k_s((x_{s+1}, P_{s+1}), (x_{s+1}, P_{s+1})) + \frac{1}{r}\sigma_\varepsilon^2}} Z, \end{aligned}$$

$$k'_{s+1}((x, P), (x', P')) = k_s((x, P), (x', P')) - \frac{k_s((x, P), (x_{s+1}, P_{s+1}))k_s((x', P'), (x_{s+1}, P_{s+1}))}{k_s((x_{s+1}, P_{s+1}), (x_{s+1}, P_{s+1})) + \frac{1}{r}\sigma_\varepsilon^2},$$

where $Z \sim \mathcal{N}(0, 1)$ is the standard normal random variable. Then from equation (10), we can obtain the distribution of $\hat{G}_{s+1}(x|x_{s+1}, P_{s+1}, y_{s+1})$ whose conditional mean and covariance are

$$\begin{aligned} \mathbb{E}[\hat{G}_{s+1}(x|x_{s+1}, P_{s+1}, y_{s+1})] &= \frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} m'_{s+1}(x, P_i) \\ &= \frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} m_s(x, P_i) + Z \frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} \frac{k_s((x, P_i), (x_{s+1}, P_{s+1}))}{\sqrt{k_s((x_{s+1}, P_{s+1}), (x_{s+1}, P_{s+1})) + \frac{1}{r}\sigma_\varepsilon^2}} \\ &\sim \mathcal{N}\left(\frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} m_s(x, P_i), \left(\frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} \frac{k_s((x, P_i), (x_{s+1}, P_{s+1}))}{\sqrt{k_s((x_{s+1}, P_{s+1}), (x_{s+1}, P_{s+1})) + \frac{1}{r}\sigma_\varepsilon^2}}\right)^2\right), \end{aligned}$$

$$\text{Cov}[\hat{G}_{s+1}(x|x_{s+1}, P_{s+1}, y_{s+1}), \hat{G}_{s+1}(x'|x_{s+1}, P_{s+1}, y_{s+1})] = \frac{1}{N_{mc}^2} \sum_{i=1}^{N_{mc}} \sum_{j=1}^{N_{mc}} k'_{s+1}((x, P_i), (x', P_j)).$$

It can then be derived that $\hat{G}_{s+1}(x|x_{s+1}, P_{s+1}) \sim \mathcal{N}(\mu'_s(x), \sigma_s'^2(x))$, where

$$\mu'_s(x) = \frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} m_s(x, P_i),$$

$$\sigma_s'^2(x) = \left(\frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} \frac{k_s((x, P_i), (x_{s+1}, P_{s+1}))}{\sqrt{k_s((x_{s+1}, P_{s+1}), (x_{s+1}, P_{s+1})) + \frac{1}{r}\sigma_\varepsilon^2}}\right)^2 + \frac{1}{N_{mc}^2} \sum_{i=1}^{N_{mc}} \sum_{j=1}^{N_{mc}} k'_{s+1}((x, P_i), (x, P_j)).$$

If we only consider the updated value at x_{s+1} caused by a new evaluation at (x_{s+1}, P_{s+1}) , then the expected improvement over the current best values, the proposed EI based infill criterion, is given by

$$\begin{aligned} \text{EI}_T(x_{s+1}, P_{s+1}) &= \mathbb{E}_{\hat{G}_{s+1}(x_{s+1}|x_{s+1}, P_{s+1})}[(T - \hat{G}_{s+1}(x_{s+1}|x_{s+1}, P_{s+1}))^+ | \bar{Y}_s] \\ &= \Delta \Phi\left(\frac{\Delta}{\sigma'_s(x_{s+1})}\right) + \sigma'_s(x_{s+1}) \phi\left(\frac{\Delta}{\sigma'_s(x_{s+1})}\right), \end{aligned}$$

where $\Delta = T - \mu'_s(x_{s+1})$, and $T = \min\{\mu_s(x_1), \dots, \mu_s(x_s)\}$ is the estimated current best value for $g(x)$ as mentioned above. Although the function value of $g(x)$ at x_{s+1} is unknown, we have the distribution $\hat{G}_{s+1}(x_{s+1}|x_{s+1}, P_{s+1})$ to account for the uncertainty of $g(x_{s+1})$. Hence, the improvement given by $T - \hat{G}_{s+1}(x_{s+1}|x_{s+1}, P_{s+1})$ is averaged with respect to this marginal distribution $\hat{G}_{s+1}(x_{s+1}|x_{s+1}, P_{s+1})$. Since $\hat{G}_{s+1}(x_{s+1}|x_{s+1}, P_{s+1})$ is a normal random variable, the EI function can be computed analytically.

The next design-distribution pair is then selected as $\arg \max_{(x_{s+1}, P_{s+1})} \text{EI}_T(x_{s+1}, P_{s+1})$, i.e. the point that can on average improve the objective function value most. r simulation replications is used to evaluate this point (x_{s+1}, P_{s+1}) to obtain simulation output $\bar{y}(x_{s+1}, P_{s+1})$. The GP model is updated and next sequential evaluation point is selected again based on the EI criterion. This sequential evaluation is iterated until the total simulation budget $N \times r$ is exhausted. Denote $\mathcal{Y}_N = \{x_1, \dots, x_N\}$ as the set of all design points visited when the budget is exhausted, the minimizer of the predictive mean $\mu_N(x)$ of $\hat{G}_N(\cdot)$ over the \mathcal{Y}_N , is returned as the final approximated minimizer, i.e $\hat{x}^* = \arg \min_{x \in \mathcal{Y}_N} \mu_N(x)$. As more and more points are evaluated, the estimated model $\hat{G}(\cdot)$ can better approximate the objective function $g(x)$, and consequently the accuracy of \hat{x}^* is also improved. Algorithm 1 below outlines the general steps for our proposed algorithm.

Algorithm 1 Modified EGO Algorithm with Input Uncertainty

-
- 1: Given i.i.d. real world data $D_n = \{\xi_1, \xi_2, \dots, \xi_n\}$ of size n and the Dirichlet process prior $DP(\alpha, P_0)$ on the input distribution P , derive the posterior $\pi(P|D_n)$;
 - 2: **Initialization:** Generate $\{x_1, x_2, \dots, x_s\}$ using Latin Hypercube Sampling from Uniform distribution with support \mathcal{X} and generate $\{P_1, P_2, \dots, P_s\}$ randomly from $\pi(P|D_n)$ to obtain an initial set of design-distribution sample pairs $\{(x_1, P_1), (x_2, P_2), \dots, (x_s, P_s)\}$. Denote $\mathcal{V}_s = \{x_1, \dots, x_s\}$;
 - 3: Run the simulation experiment at these initial sample points with r replications at each point and obtain the observed output sample mean vector $\bar{Y}_s = [\bar{y}(x_1, P_1), \bar{y}(x_2, P_2), \dots, \bar{y}(x_s, P_s)]^T$;
 - 4: **Validation:** Based on \bar{Y}_s , construct a stochastic GP model $F_s(x, P)$; Perform cross-validation (e.g. Leave-one-out cross validation) to ensure that the metamodel is valid;
 - 5: **while** $s \leq N - 1$ **do** $\triangleright N \times r$ is the total budget
 - 6: Generate P_1, \dots, P_{NMC} from $\pi(P|D_n)$; Derive the model $\hat{G}_s(x, P) \sim GP(\mu_s, c_s)$ for $g(x)$;
 - 7: **Select sequential design-distribution point** (x_{s+1}, P_{s+1}) : use the modified EGO algorithm;
 - 8: Run simulation experiments at (x_{s+1}, P_{s+1}) with r replications and obtain the observed output mean $\bar{y}(x_{s+1}, P_{s+1})$, set $\bar{Y}_{s+1} = [\bar{Y}_s, \bar{y}(x_{s+1}, P_{s+1})]^T$;
 - 9: **Update:** update the stochastic GP model $F_s(x, P)$ based on \bar{Y}_{s+1} ;
 - 10: Set $\mathcal{V}_{s+1} = \mathcal{V}_s \cup x_{s+1}$ and set $s = s + 1$;
 - 11: **return** $\hat{x}^* = \arg \min_{x \in \mathcal{V}_N} \mu_N(x)$.
-

4 EMPIRICAL PERFORMANCE: AN INVENTORY PROBLEM

In this section, we focus on the inventory problem of Fu and Healy (1997) that has been widely used to test the empirical performance of simulation optimization algorithms (Jalali et al. 2017; Wang et al. 2019). We will first examine the empirical convergence of our proposed modified EGO algorithm in obtaining the optimal value of the NBRO objective function. We will then compare the performance of our approach which explicitly accounts for input uncertainty with the approach which overlooks input uncertainty. This is to further highlight the importance of considering the NBRO problem when input uncertainty is high.

4.1 The Inventory Simulator and Experiment Setting

In the inventory problem of Fu and Healy (1997), a company manages the inventory of a single product with a periodic review policy. At the end of each period (e.g. each week), the company will check the inventory position with the following rule: if the inventory position is above the basic ordering level (denoted as s), the company will not order; If the inventory position is below s , the company will order the difference between the order-up-to level (denoted as S) and the inventory position. The decision variable $x = [s, S]$ is two dimensional. The customer demand, denoted as ξ is stochastic and is independently identically distributed according to a distribution P across different periods. The cost include the fixed ordering cost = 100, unit cost = 1, holding cost = 1, and back-order cost = 100. The candidate decision space is $\{x = [s, S] | s \in [10000, 22500], S \in [22600, 35000]\}$. All these parameters used in this experiment are set the same as in Jalali et al. (2017). The simulator is coded in Python, and has been validated by comparing with that of Jalali et al. (2017). The length of simulation run is set to be 1000 periods with a warm-up length of 100 periods per simulation. The simulator output is the steady-state cost which is the cost averaged over the 900 (= 1000 – 100) periods. For a given decision x and a given demand distribution P , each simulation run produces an output $y(x, P) = h(x, \xi)$, $\xi \sim P$. The expected steady-state cost (expected simulation output) for a given decision x and a given demand distribution P is denoted as $f(x, P)$.

We assume the true distribution of the demand P^c as $\exp(\lambda^c)$, where $\lambda^c = 0.0002$ is the true input parameter value. The true expected steady-state cost function is denoted as $f(x, P^c)$. Ideally, we aim to solve the problem of $\min_x f(x, P^c)$. When the true demand distribution P^c is exponential, the expected cost function has a closed-form and the optimal decision can be calculated analytically. The expected cost

function is given by

$$f(x, P^c = \exp(\lambda^c)) = \frac{1}{100} \left[\frac{1}{\lambda^c} + \left[100 + s - \frac{1}{\lambda^c} + 0.5\lambda^c(S^2 - s^2) + \frac{101}{\lambda^c} e^{-\lambda^c s} \right] / [1 + \lambda^c(S - s)] \right]. \quad (11)$$

In practice however, the P^c is unknown (i.e. both the exponential distribution and the value of λ^c is unknown). Instead, only a finite set of real data $D_n = \{\xi_1, \dots, \xi_n\} \sim P^c$ can be observed. Therefore, we cannot use the analytical form in (11) to calculate the optimal decision $x^* = \operatorname{argmin}_x f(x, P^c)$ directly. What is available is the simulator, from which the simulation output $y(x, P)$ can be obtained for a given decision x and a given demand distribution P .

Under the NBRO approach, we model P^c as a random variable P . We assume a Dirichlet prior for P with $\alpha = 1$ and $P_0 = \text{uniform}(0, \max(\xi_1, \dots, \xi_n))$ (as recommended in Gelman et al. (2013)). Then we solve the problem of $\min_x g(x) = E_{P \sim \pi}[f(x, P)]$. As the exact form of $f(x, P)$ and $g(x)$ is unknown, our proposed GP algorithm first builds a GP model $F_s(x, P)$ for $f(x, P)$ and then constructs a GP model $\hat{G}_s(x)$ for $g(x)$. Both $F_s(x, P)$ and $\hat{G}_s(x)$ are updated sequentially as additional points are selected to be evaluated by our proposed EGO algorithm. When the total budget is exhausted, the optimal solution is returned based on $\mu_N(x)$, the predictive mean for $g(x)$, as follows: $\hat{x}^* = \operatorname{argmin}_{x \in \mathcal{Y}_N} \mu_N(x)$.

4.2 Empirical Convergence of the Proposed Algorithm

We first test the efficiency of our proposed algorithm in solving the NBRO problem given by (5). We define $\text{gGAP} = |\min_x g(x) - g(\hat{x}^*)|$. As $g(x)$ is not directly observable, we approximate it with $\frac{1}{10000} \sum_{i=1}^{10000} \bar{y}(x, P_i)$, where $\bar{y}(x, P_i)$ is the average of 1000 simulation outputs at (x, P_i) , and we use a grid search over a large set of the design points to obtain $\min_x g(x)$. The gGAP is used to test whether the solution of our proposed GP algorithm is able to converge to the optimal solution of $g(x)$.

We follow the suggestion of Jones et al. (1998) to use the number of $10 \times d = 30$, where $d = 3$ is the dimension of the function, as the initial design sample size and sequentially select additional 100 points. For evaluation on each point (x, P) , we use $r = 10$ replications to determine the value of $\bar{y}(x, P)$. Therefore, the total budget is $(30 + 100) \times 10 = 1300$. We conduct 100 macro-replications and calculate the mean of the gGAP values and its 95% confidence interval. The number of real world data is $n = 10$. The results are shown in Figure 2. The figure shows that as the iteration (budget) increases, the optimal values returned by our algorithm is able to get close to the optimal values of the $g(x)$. At the end of the experiment, the mean of the gGAP is less than 1, which has reduced more than 95% compared with that at beginning of the experiment.

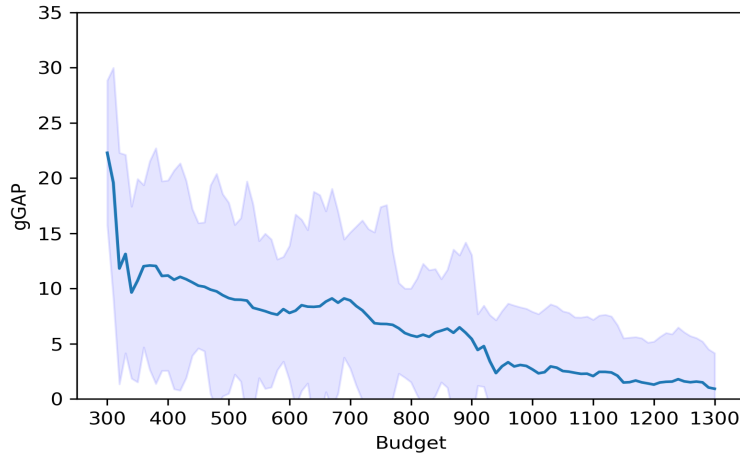


Figure 2: Empirical performance of our algorithm with respect to the number of budget in terms of gGAP.

4.3 Comparison of the NBRO with the Approach Ignoring the Input Uncertainty

As mentioned in section 1, when P^c is unknown, it is typical in practice to ignore the input uncertainty and directly use an estimated \hat{P} as if it were P^c . This approach then solves (2), ignoring the input uncertainty. In the following, we term this approach as the Plug-In (PI) approach. The empirical cumulative distribution computed on the real data: $\hat{P}(t) = \frac{1}{n} \sum_{i=1}^n 1_{\xi_i \leq t}$ is commonly chosen. The PI approach can be solved using the sequential kriging optimization (SKO, an variant of EGO proposed by Huang et al. (2006)), in which a GP model is built for $f(x, P)$ with predictive mean denoted as $\hat{f}(x, P)$. The obtained optimizer is given by $\hat{x}^* = \arg \min_{x \in \mathcal{V}'_N} \hat{f}(x, \hat{P})$, where \mathcal{V}'_N is the set of all evaluated design points.

In the following, we will compare the NBRO approach which explicitly considers the input uncertainty with the PI approach to see the potential benefits of the NBRO. We define $GAP = |\min_x f(x, P^c) - f(\hat{x}^*, P^c)|$. GAP directly measures the loss incurred from the estimated minimizer \hat{x}^* . We use also 100 macro-replications with different initial Latin Hypercube design points for each macro-replication. For a given macro-replication, the two approaches start with the same initial design points. 20 points are selected sequentially after the initial design. Therefore, the total budget is $(30 + 20) \times 10 = 500$. Here we consider the performance of the two approaches under different input uncertainty levels, which can be roughly controlled by the number of the real world data. We consider four cases of this number n , with n set at 10, 50, 100 and 1000. The comparison results are summarized and shown in the box plots in Figure 3.

We use Mood's median test (Mood 1954) with a significant level of 0.05 to test whether there are significant differences between the GAP medians of the two approaches. The mood's median test shows that the difference is not significant when $n = 1000$, but is significant for $n = 10, 50$ and 100. When $n = 1000$, the GAP values for both approaches tend to be zero, and the variance of this measure is small. This can be expected because as n increases, the input uncertainty level will decrease. When n tends to infinity (i.e. the input uncertainty vanishes), both the $\pi(P|D_n)$ and \hat{P} will be similar to P^c . As a result, both the NBRO and the PI approaches will reduce to the true problem (1), and hence it is expected that the minimizers from both approaches converge. However, when $n = 10$ and 50, the performance of the NBRO approach is significantly better than the PI approach. When $n = 100$, the proposed NBRO approach achieves both a smaller median and variance than its counterpart. Although for this data size, the median of PI approach is not very large, we observe that its variance is large, indicating that this approach is highly dependent on \hat{P} . This result is consistent with our motivation of proposing NBRO in section 1.

In summary, the proposed NBRO approach with the modified EGO algorithm which explicitly considers input uncertainty is more robust to the input uncertainty level compared with the PI approach which ignores it. This proposed approach is able to provide a smaller overall median GAP and variance performances. This

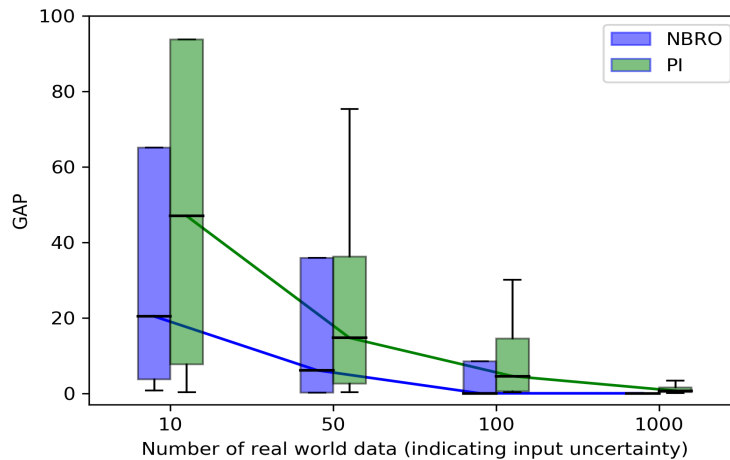


Figure 3: Boxplot of the GAP values for the two approaches (NBRO and PI) under different levels of input uncertainty. The black solid lines within the boxes represent the medians.

further illustrates that our approach can be effective in hedging the risk that arises from input uncertainty in simulation optimization, especially when this uncertainty is high.

5 CONCLUSION

In this paper, we consider the stochastic simulation optimization problem when the input distribution is unknown. We first formulate the nonparametric Bayesian risk optimization (NBRO) problem and propose to use a nonparametric Bayesian approach to model the input uncertainty. Then the Efficient Global Optimization algorithm is extended to solve this NBRO problem. From the results of the empirical experiment, we find that: 1) Our proposed algorithm can empirically converge to the optimal value of the NBRO objective function $g(x)$. This essentially shows that our proposed algorithm which uses a fast approximation, can locate the global minimizer of the NBRO problem accurately, and can be applied to efficiently solve the NBRO problem; 2) Compared with an approach that ignores the input uncertainty, our algorithm is able to obtain solutions closer to the optimal value of $f(x, P^c)$. Besides, NBRO appears to be more robust to the input uncertainty level, which illustrates the benefits of considering the NBRO as the objective in stochastic simulation optimization when input uncertainty is present.

There are several future work directions. Firstly, in this work, we use the quadratic Wasserstein distance as the distance measure for two distributions in the GP model. The sensitivity of the performance of our GP algorithm with respect to different distance measures will need further investigation. Secondly, it will be interesting to compare the performance of our approach with other parametric approaches to modeling the input uncertainty.

REFERENCES

- Ankenman, B., B. L. Nelson, and J. Staum. 2010. "Stochastic Kriging for Simulation Metamodeling". *Operations Research* 58(2):371–382.
- Bachoc, F., F. Gamboa, J.-M. Loubes, and N. Venet. 2017. "A Gaussian Process Regression Model for Distribution Inputs". *Institute of Electrical and Electronics Engineers Transactions on Information Theory* 64(10):6620–6637.
- Barton, R. R., and M. Meckesheimer. 2006. "Metamodel-based Simulation Optimization". *Handbooks in Operations Research and Management Science* 13:535–574.
- Barton, R. R., and L. W. Schruben. 1993. "Uniform and Bootstrap Resampling of Empirical Distributions". In *Proceedings of the 1993 Winter Simulation Conference*, edited by G. Evans, M. Mollaghasemi, E. Russell, and W. Biles, 503–508. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Barton, R. R., and L. W. Schruben. 2001. "Resampling Methods for Input Modeling". In *Proceedings of the 2001 Winter Simulation Conference*, edited by B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, 372–378. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Chick, S. E. 2001. "Input Distribution Selection for Simulation Experiments: Accounting for Input Uncertainty". *Operations Research* 49(5):744–758.
- De Oliveira, V., and B. Kone. 2015. "Prediction Intervals for Integrals of Gaussian Random Fields". *Computational Statistics and Data Analysis* 83:37–51.
- Ferguson, T. S. 1973. "A Bayesian Analysis of Some Nonparametric Problems". *The Annals of Statistics* 1(2):209–230.
- Frazier, P., W. Powell, and S. Dayanik. 2009. "The Knowledge-gradient Policy for Correlated Normal Beliefs". *Institute for Operations Research and the Management Sciences Journal on Computing* 21(4):599–613.
- Frazier, P. I. 2018. "Bayesian Optimization". In *Tutorials in Operations Research: Recent Advances in Optimization and Modeling of Contemporary Problems*, edited by E. Gel and L. Ntamo, 255–278. Catonsville, Maryland: Institute for Operations Research and the Management Sciences.
- Fu, M. C., and K. J. Healy. 1997. "Techniques for Optimization via Simulation: An Experimental Study on an (s, S) Inventory System". *Institute of Industrial Engineers Transactions* 29(3):191–199.
- Gelman, A., H. S. Stern, J. B. Carlin, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. *Bayesian Data Analysis*. 3rd ed. Boca Raton: Chemical Rubber Company Press.
- Hein, M., and O. Bousquet. 2005. "Hilbertian Metrics and Positive Definite Kernels on Probability Measures". In *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, edited by R. G. Cowell and Z. Ghahramani, 136–143. New Jersey: Society for Artificial Intelligence and Statistics.
- Hjort, N. L., C. Holmes, P. Müller, and S. G. Walker. 2010. *Bayesian Nonparametrics*. Cambridge, England: Cambridge University Press.

- Huang, D., T. T. Allen, W. I. Notz, and N. Zeng. 2006. "Global Optimization of Stochastic Black-box Systems via Sequential Kriging Meta-models". *Journal of Global Optimization* 34(3):441–466.
- Jalali, H., I. Van Nieuwenhuysse, and V. Picheny. 2017. "Comparison of Kriging-based Algorithms for Simulation Optimization with Heterogeneous Noise". *European Journal of Operational Research* 261(1):279–301.
- Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-box Functions". *Journal of Global Optimization* 13(4):455–492.
- Kleijnen, J. P. 2009. "Kriging Metamodeling in Simulation: A Review". *European Journal of Operational Research* 192(3):707–716.
- Kushner, H., and G. Yin. 2003. *Stochastic Approximation and Recursive Algorithms and Applications*. New York: Springer-Verlag.
- Lam, H. 2016. "Advanced Tutorial: Input Uncertainty and Robust Analysis in Stochastic Simulation". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 178–192. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Mood, A. M. 1954. "On the Asymptotic Efficiency of Certain Nonparametric Two-sample Tests". *The Annals of Mathematical Statistics* 25(3):514–522.
- Ólafsson, S. 2006. "Metaheuristics". *Handbooks in Operations Research and Management Science* 13:633–654.
- Pearce, M., and J. Branke. 2017. "Bayesian Simulation Optimization with Input Uncertainty". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 2268–2278. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Peyré, G., and M. Cuturi. 2019. "Computational Optimal Transport: With Applications to Data Science". *Foundations and Trends® in Machine Learning* 11(5-6):355–607.
- Song, E., and B. L. Nelson. 2015. "Quickly Assessing Contributions to Input Uncertainty". *Institute of Industrial Engineers Transactions* 47(9):893–909.
- Villemonteix, J., E. Vazquez, and E. Walter. 2009. "An Informational Approach to the Global Optimization of Expensive-to-evaluate Functions". *Journal of Global Optimization* 44(4):509.
- Wang, H., J. Yuan, and S. H. Ng. 2019. "Gaussian Process Based Optimization Algorithms with Input Uncertainty". *Institute of Industrial and Systems Engineers Transactions* 52(4):377–393.
- Williams, C. K., and C. E. Rasmussen. 2006. *Gaussian Processes for Machine Learning*. Cambridge, Massachusetts: Massachusetts Institute of Technology press.
- Wu, D., H. Zhu, and E. Zhou. 2018. "A Bayesian Risk Approach to Data-driven Stochastic Optimization: Formulations and Asymptotics". *Society for Industrial and Applied Mathematics Journal on Optimization* 28(2):1588–1612.
- Xie, W., C. Li, Y. Wu, and P. Zhang. 2019. "A Bayesian Nonparametric Framework for Uncertainty Quantification in Simulation". *arXiv preprint arXiv:1910.03766*.
- Xu, J., E. Huang, C.-H. Chen, and L. H. Lee. 2015. "Simulation Optimization: A Review and Exploration in the New Era of Cloud Computing and Big Data". *Asia-Pacific Journal of Operational Research* 32(03):1–34.
- Zhou, E., and W. Xie. 2015. "Simulation Optimization When Facing Input Uncertainty". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 3714–3724. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Zouaoui, F., and J. R. Wilson. 2003. "Accounting for Parameter Uncertainty in Simulation Input Modeling". *Institute of Industrial Engineers Transactions* 35(9):781–792.

AUTHOR BIOGRAPHIES

Haowei Wang is currently a Ph.D. candidate in the Department of Industrial Systems Engineering and Management at National University of Singapore. He received his B.Eng. degree in Industrial Engineering from Nanjing University in 2016. His research interests include uncertainty quantification, simulation metamodeling and optimization. His email address is haowei_wang@u.nus.edu.

Szu Hui Ng is an Associate Professor in the Department of Industrial Systems Engineering and Management at the National University of Singapore. She holds B.S., M.S. and Ph.D. degrees in Industrial and Operations Engineering from the University of Michigan. Her research interests include computer simulation analysis and optimization, applications of simulation to maritime transportation and quality engineering. She is a member of IEEE and INFORMS, and a senior member of IISE. Her e-mail address is isensh@nus.edu.sg.

Xun Zhang is a Ph.D. candidate in the Department of Industrial Systems Engineering and Management at National University of Singapore. He received his B.S. in Department of Mathematics from Nanjing University in 2017. His research interests include optimization algorithms and data-driven operations management. His email address is e0225088@u.nus.edu.