Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

DATA-DRIVEN SIMULATION OPTIMIZATION IN THE AGE OF DIGITAL TWINS: CHALLENGES AND DEVELOPMENTS

Enlu Zhou[1]

[1]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA

ABSTRACT

A digital twin is a virtual representation of a physical system, continuously updated with data from the physical system, enabling dynamic information exchange and decision-making. This bidirectional interaction, the key distinction from traditional simulations, introduces unique challenges to simulation optimization. These challenges encompass streaming data, system nonstationarity, and real-time decision-making. This tutorial reviews recent advancements and methodologies aimed at addressing these issues, with a focus on data-driven simulation optimization under streaming data.

1 INTRODUCTION

A digital twin is "a set of virtual information constructs that mimics the structure, context, and behavior of a natural, engineered, or social system (or system of systems), is dynamically updated with data from its physical twin, has a predictive capability, and informs decisions that realize value." (National Academies of Sciences, Engineering, and Medicine 2024). Digital twin emerges as an appealing and perhaps the sole feasible tool for modeling, comprehending, and optimizing complex real-life systems. It has significant potential in many engineering and science domains.

At the core of many digital twins is a stochastic simulation model or a system of such models. More detailed descriptions of digital twins and their connection with simulation are available in recent tutorial papers, such as Li et al. (2020), Biller et al. (2022), Matta and Lugaresi (2023). The defining feature of a digital twin, distinguishing it from conventional simulations, is its bidirectional interaction with the physical system. The physical system continually provides its digital twin with real-time data, enhancing the fidelity of the simulation model(s), while predictions and decisions based on these models are provided to the physical system. This process occurs repeatedly at a much faster pace than conventional simulation analysis and optimization. Therefore, digital twins present new challenges to simulation analysis and optimization, necessitating data-driven techniques that previously have not been much explored in the simulation literature.

This tutorial aims to provide an overview of the rapidly evolving research area of data-driven simulation optimization, outlining the challenges and opportunities presented by digital twins. It introduces some recent developments and primarily focuses on the types of data used for estimating probability distributions in stochastic simulation models. While other data types, such as contextual information or environmental variables, are also of interest, they are beyond the scope of this tutorial.

2 CHALLENGES IN DATA-DRIVEN SIMULATION OPTIMIZATION

Traditional simulation optimization, which assumes access to a fixed simulator, already faces numerous challenges such as expensive and noisy evaluation of objective functions and lack of knowledge about the objective function structures. The literature on simulation optimization has been advancing the frontier in this field (Fu 2015). However, the emergence of digital twins and data-driven applications brings additional challenges to simulation optimization. In the following sections, we will discuss these new challenges in greater detail and review existing work that addresses them.

2.1 Input Data vs. Observable Contexts

Let's begin by discussing the types of data commonly encountered in data-driven simulation optimization. The first type comprises data or observations stemming from randomness within the physical system. Examples of such data include customer demands and lead times in revenue management applications, and passenger arrival times and transit times in transportation networks. These data are frequently utilized to estimate the input distributions of the simulator, hence often referred to as input data. The second type consists of observable contexts, often termed contextual variables, covariates, or side information. For instance, personal characteristics such as tumor biomarkers and gene expressions serve as observable contexts for personalized treatment decisions in cancer treatment (Shen et al. 2021); in finance, underlying risk factors like stock prices and interest rates act as contexts for monitoring and minimizing portfolio risks (Jiang et al. 2019).

These two types of data play distinct roles in simulation optimization. Input data, observed from an unknown underlying distribution, introduce uncertainty for decision-makers, who must therefore consider hedging against this input uncertainty in the decisions deployed to the physical system. Conversely, covariates are observed beforehand, and decisions are made based on the realized covariate. Typically, decision-makers seek to identify an optimal decision for each possible realization of the covariate. Mathematically, the first problem entails finding a single decision that accounts for the uncertainty of the unknown underlying distribution, while the second involves identifying an optimal policy — a function that maps from the covariate space to the decision space.

In recent years, active research has focused on ranking and selection (R&S) under a fixed batch of input data. R&S is a type of simulation optimization problems with categorical decision variables. One line of research aims to screen out as many inferior designs as possible in the presence of input uncertainty, as demonstrated by works such as Corlu and Biller (2013), Corlu and Biller (2015), Song and Nelson (2019). Another stream of work takes a distributionally robust optimization (DRO) approach by assuming that the true input distribution is contained in a finite set of known distributions (i.e., ambiguity set), with a goal of selecting the design with the best worst-case performance over the ambiguity set, such as Gao et al. (2017), Xiao and Gao (2018), Xiao et al. (2020), Fan et al. (2020). Different from the mainstream DRO approach, Wu and Zhou (2017) aims to select the true best design instead of the worst-case optimal. Recognizing the overly conservative nature of DRO, which solely considers the worst case, Wu et al. (2018) proposes a Bayesian risk optimization (BRO) approach for general simulation optimization problems. Unlike constructing an ambiguity set, BRO utilizes the Bayesian posterior distribution on the input distribution parameter to quantify input uncertainty and subsequently applies a risk functional (with respect to the posterior) of the original objective function. There is also active research ongoing for contextual-dependent R&S and general simulation optimization, such as Shen et al. (2021), Gao et al. (2019), Jin et al. (2020), Li et al. (2020), Liu et al. (2021), Li et al. (2022), Cakmak et al. (2024).

2.2 Streaming Input Data

A critical aspect of digital twin is the synchronization of the physical system and its digital twin via the use of streaming data in real time (Biller et al. 2022). Streaming data arrive sequentially in batches, often of random and varying sizes. Figure 1 compares the process of traditional simulation optimization and digital-twin-based simulation optimization. Specifically, traditional simulation assumes a fixed simulation model, which is built with a batch of pre-collected data. Digital twin, in contrast, is a dynamic simulator that is frequently updated with the new incoming input data. This leads to the need of running more simulations under time-heterogeneous models and dynamically update the decisions. Moreover, the decisions may impact the data (i.e., data are endogenous to the decision), creating an even more complicated relationship between data, model, and decisions.

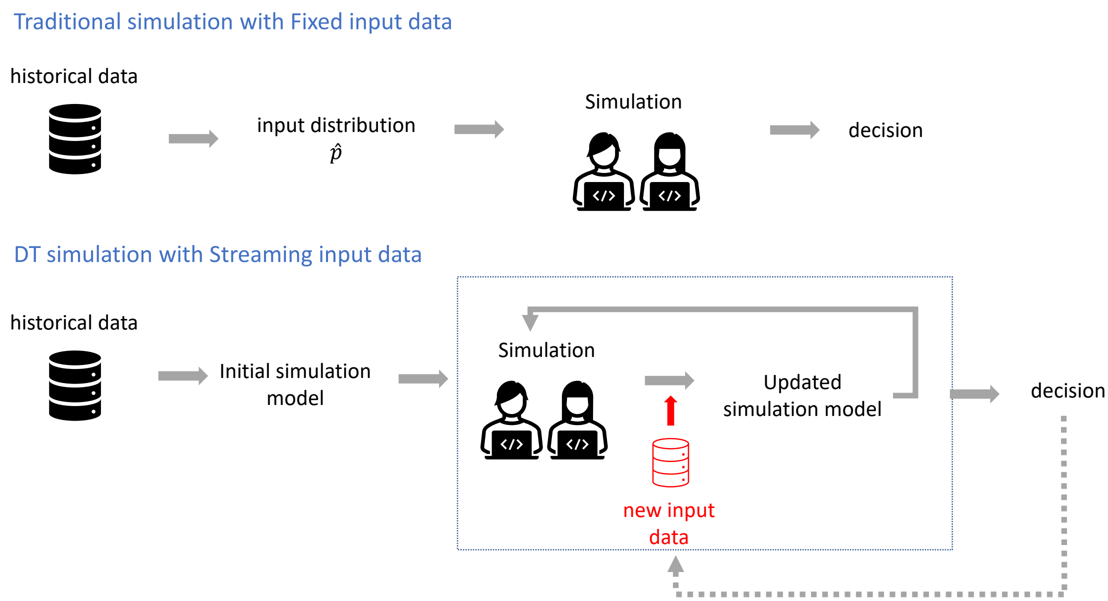

Figure 1: Traditional versus Digital-Twin-based Simulation Optimization.

Despite extensive research on simulation optimization with input data, most existing work focuses on fixed input data, as summarized in the previous subsection. The study of streaming data in this context is relatively limited and recent. Wu et al. (2022) consider streaming input data in R&S problems, and propose a data-driven approach that aims to achieve a specified probability of correct selection (PCS) of the best design. Wang and Zhou (2022) consider the R&S problem with streaming input data give a fixed simulation budget at each time stage, and develop an optimal budget allocation procedure to maximize the PCS. Additionally, He et al. (2024) and Liu et al. (2024) study continuous simulation optimization with streaming input data. The former paper estimates the unknown input parameter via a maximum likelihood estimator with i.i.d. (independent and identically distributed) streaming data, while the latter adopts a Bayesian approach and also accounts for decision-dependent streaming data.

Furthermore, some data can be actively acquired using financial or temporal resources, leading to questions about how to allocate budgets for data collection across different input distributions and how to allocate resources jointly to data collection and simulation runs. With this consideration, Wu and Zhou (2017) and Xu et al. (2020) explore budget allocation for input data collection and simulation under shared resources. Wu and Zhou (2017) adopt a two-stage approach that first collects sufficient input data and then allocates the remaining budget to running simulations. Xu et al. (2020) consider scenarios where the cost of data collection is random and compare situations where simulation costs are negligible or significant compared to data collection costs. Later, Kim and Song (2022) and Wang and Zhou (2023) consider two separate budgets for data collection and simulation. Kim and Song (2022) solve the two budget allocation problems separately and develop a sequential allocation procedure with the selection criterion of the most probable best design. In contrast, Wang and Zhou (2023) jointly formulate and optimize the two resource allocations.

2.3 Decision-dependent Uncertainty and System Nonstationarity

Most of the relevant work in the simulation optimization literature focus on exogenous distributions that are independent of the decision variables. However, as illustrated in Figure 1, decisions may affect some random factors in the system, meaning that these random factors are endogenous to decisions. This implies that data observed from endogenous uncertainty over time are not i.i.d. but rather correlated and differently distributed. For instance, in revenue management problems, the price (decision) of a product can influence the distribution of customer demands, and the customer demand data may, in turn, affect the price if dynamic pricing is adopted by the retailer. In a broader context, endogenous uncertainty or decision-dependent data can be seen as a special case of system nonstationary. Other cases of system nonstationarity include abrupt changes, structural changes, and small and possibly random perturbations in the system.

System nonstationarity has been seldom considered in simulation optimization. To the best of our knowledge, the only work in simulation optimization that considers decision-dependent input distribution/data are Liu et al. (2024) and its earlier conference paper Liu et al. (2021). They propose an algorithm that jointly estimates the input distribution parameter via Bayesian posterior distribution and updates the decision by applying stochastic gradient descent (SGD) iterations on the Bayesian average of the objective function. They show that the Bayesian posterior distribution is a consistent estimator of the input distribution parameters even with non-i.i.d. decision-dependent data, and consequently the algorithm converges to the optimal solution under the true input distribution.

2.4 Online Decision Making

Traditional simulation optimization is proficient in offline decision making, but is almost impossible to use for online decision making due to the long computation time to run simulations of a complex system. Some pioneering works, including (Nelson 2016; Lin and Nelson 2016; Hong and Jiang 2019; Jiang et al. 2019; Shen et al. 2021; Jin et al. 2020; Liu et al. 2021), address the problem of online decision-making with offline estimation. In this approach, the simulation model is fixed and run offline to simulate multiple scenarios. Then, in online applications, when a new scenario is observed, a predictive model is utilized to determine the optimal decision for the observed scenario based on the offline simulation results of known scenarios.

This approach, which combines online decision-making with offline estimation or simulation, is primarily suited for contextual data. However, it is challenging to apply to many digital twin applications where the simulator dynamically changes over time. A possible adaptation involves parameterizing the simulator using input parameters, typically continuous variables. This adaptation, though, could result in a high-dimensional parameter space, and the predicted simulation outputs might exhibit significant bias. In contrast, most studies dealing with streaming input data take the approach of offline decision-making with online estimation and simulation, as summarized in Section 2.2. However, the problem becomes much more challenging when online decision-making is coupled with online estimation, as online estimation (with streaming input data) modifies the underlying simulation model. Consequently, online decision-making within a dynamically evolving simulation model remains a largely open area. To address this, we anticipate a need to not only develop sophisticated simulation optimization algorithms but also to harness emerging techniques and infrastructures, such as machine learning tools and high performance computing.

2.5 Uncertainty Quantification

It is important to quantify the uncertainty of simulation outputs and the decisions made out of those outputs. There have been extensive study on input uncertainty quantification with a fixed batch of input data. These methods include Bayesian methods (e.g., Chick (2001), Zouaoui and Wilson (2003), Zouaoui and Wilson (2004), Xie et al. (2014)), frequentist methods (e.g., Barton and Schruben (1993), Barton and Schruben (2001), Cheng and Holloand (1997)), delta methods (e.g., Cheng and Holloand (1997)), meta-model assisted methods (e.g., Barton et al. (2013), Xie et al. (2014)), and some more recent ones (e.g., Lam (2016), Zhu et al. (2020), Lam and Qian (2016), Lam and Qian (2018), Lin et al. (2015), Feng and Song (2019)).

In spite of the flourishing research on input uncertainty quantification, there is almost no work considering streaming input data. Probably the only exception is by Liu and Zhou (2019), who propose a two-layer importance sampling framework that incorporates streaming data for online input uncertainty quantification. Moreover, little work has been done on quantifying the uncertainty of a solution or objective value of a simulation optimization problem with input data. In this regard, Lam and Zhou (2015) consider replacing the unknown input distribution by the empirical distribution of the input data, namely sample average approximation (SAA), and propose several approaches to quantify the uncertainty in the optimal objective value, including the empirical likelihood method, nonparametric Bayesian approach, and bootstrap approach. Later, Lam and Zhou (2017) delve more into the empirical likelihood method.

Lastly, we want to conclude this section by noting that there are more challenges than those mentioned here. For example, data could be high-dimensional, come from multiple sources, have different resolutions, and exist in different formats. Additionally, models of different scales and different fidelity levels may be available for complex systems. Integrating these diverse data and models is a challenging task for the successful application of digital twins to real-world problems.

3 CURRENT RESEARCH IN DATA-DRIVEN SIMULATION OPTIMIZATION

In this section, we provide a gentle and brief introduction to some recent developments in the aforementioned areas. We apologize that the focus will be on the author's own work. Notations and technicality are kept to a minimum, while emphasis is placed on intuitive understanding.

3.1 Simulation Optimization

We start with describing the problem setting. A general simulation optimization problem can be mathematically formulated as

$$\max_{x\in\mathscr{X}} \quad h(x) := \mathbb{E}[H(x,\xi)], \tag{1}$$

where x is the decision variable, $\mathscr{X}$ is the feasible region for the decision variable, H is a deterministic *performance function* specified by a simulation model, h is the *mean performance measure* of the system, and ξ is a random vector that models the randomness in the system. The feasible region $\mathscr{X}$ could be an explicitly specified set or implicitly determined by constraints on the decision variable. The performance function H usually lacks a closed-form expression and must be evaluated through simulation that is often computationally expensive — a key feature and challenge of simulation optimization.

Ranking and Selection (R&S) is a special type of simulation optimization problems defined on a finite solution space. R&S deserves its own dedicated study, distinct from continuous simulation optimization. This is due to the finite solution space inherent to R&S, which permits the simulation of every potential solution multiple times. As a result, R&S can employ different methodologies that yield stronger convergence results compared to those typically achieved in continuous simulation optimization. Moreover, the finite solutions are often categorical, distinguishing R&S from a discretized version of a continuous simulation optimization problem. Specifically, R&S wants to identify the solution with the smallest expected performance among K solutions (also often called *alternatives* or *designs*). Denote by $\mathscr{K} := \{1,2,\ldots,K\}$ the enumeration of all alternatives, and assume all the alternatives share the same randomness. Then, with other notations same as (1), the optimal solution or the best alternative is defined as

$$b := \arg\min_{i\in\mathscr{K}} \mathbb{E}[H(i,\xi)], \tag{2}$$

where b is often assumed to be unique to avoid technicality. Since it is impossible to identify the best alternative with 100% confidence unless all alternatives are simulated for an infinite number of times, the realistic goal of R&S often falls into two types: fixed confidence and fixed budget. The fixed confidence setting aims to minimize the simulation effort to achieve a pre-specified probability of correct selection (PCS) of the best alternative, while the fixed budget setting aims to maximize PCS when exhausting a given simulation budget.

Please note that the performance measure considered in this tutorial is the mean/expectation of the performance function. There could be other performance measures, such as quantiles (also often called Value-at-Risk (VaR) in finance and risk analysis), and Conditional Value-at-Risk (CVaR), or a probability of the performance function.

3.2 Input distribution, Input Data, Input Uncertainty

We now introduce some important concepts and notions regarding data-driven simulation optimization. The expectation in (1) or (2) is taken with respect to the distribution of the random vector ξ. In practice, the "correct" or true distribution F^c is almost never known and often needs to be estimated from *input data*. The estimated distribution $\hat{F}$, called the *input model*, is the actual distribution used to run the simulation.

As a result, a typical stochastic simulation experiment faces two types of uncertainties: 1) *simulation uncertainty* due to the simulation noise in performance evaluation (also referred to as the intrinsic uncertainty or aleatoric uncertainty, because it is caused by the stochastic nature of the system); and 2) *input uncertainty* in the input model $\hat{F}$ (also referred to as model/parameter uncertainty or epsidemic uncertainty). The performance estimation error due to simulation uncertainty can be "simulated away", i.e., we can decrease the error by increasing the number of simulation replications. In contrast, the input uncertainty is often uncontrollable, subject to the limited amount of input data available and/or the arrival rate of streaming data.

This first raises the question of how to address the input uncertainty when solving simulation optimization. Secondly, as *streaming input data* continues to arrive over time, it brings up additional questions about how to assimilate this data to improve the accuracy of the input model, and how to update the simulation-based performance estimates and decisions accordingly. In the next few subsections, we will introduce recent works that tackle these questions.

3.3 Simulation Optimization with Fixed Input Data

We first discuss the problem of how to find an "optimal" solution with a fixed batch of input data. As the correct input distribution is unknown, the optimal solution to the true problem (1) is also not possible to obtain; instead we need to re-define the optimality. One choice is the optimal solution to the empirical problem:

$$\min_{x\in\mathscr{X}} \mathbb{E}_{\hat{F}}[H(x,\xi)]. \tag{3}$$

However, such a solution to this empirical problem (3) may perform poorly for the true problem (1), especially when the size of input data is small, because the empirical problem completely ignores the input uncertainty and treats the empirical distribution as if it were the true distribution.

To account for the input uncertainty, a popular class of approaches under the name of "distributionally robust optimization (DRO)" aims to find a robust solution that optimizes the worst-case scenario in an ambiguity set, which is often constructed to contain the true distribution with a prescribed high probability. That is,

$$\min_{x\in\mathscr{X}} \max_{F\in\mathscr{F}} \mathbb{E}_F[H(x,\xi)], \tag{4}$$

where $\mathscr{F}$ denotes the ambiguity set. A significant amount of efforts have been spent on the construction of ambiguity sets. However, DRO's reliance on the ambiguity set can be a double-edged sword: it may lead to overly conservative solutions, as the worst case usually has a small probability to occur in reality.

In between the range of options from optimistically ignoring the distributional uncertainty as the empirical optimization (3) to pessimistically fixating on the worst case as DRO (4), Zhou and Xie (2015) and Wu et al.

(2018) propose a Bayesian risk optimization (BRO) approach that strikes the middle ground. Compared with DRO, BRO has the following two features: 1) it adopts the Bayesian posterior distribution, instead of an ambiguity set, to characterize the input uncertainty; and 2) it uses a general risk functional, rather than the worst-case, to provide more flexibility towards the risk attitude. For computational convenience, let's first assume the input distribution takes a parametric form $\{F_\theta \mid \theta \in \Theta\}$ that admits the probability density function (pdf) $f(\cdot|\theta)$, where Θ is the parameter space and $\theta^c \in \Theta$ is the unknown true parameter. Then they take a Bayesian perspective to estimate the unknown parameter. That is, the parameter vector θ is assumed to be random whose probability distribution is supported on the set Θ and follow a prior probability density $p(\theta)$. Given the data (sample) $\xi^{(N)} = (\xi_1, ..., \xi_N)$, the posterior distribution is determined by Bayes' rule

$$p(\theta|\xi^{(N)}) = \frac{f(\xi^{(N)}|\theta)p(\theta)}{\int_\Theta f(\xi^{(N)}|\theta)p(\theta)d\theta},$$

where $f(\xi^{(N)}|\theta) = \prod_{i=1}^N f(\xi_i|\theta)$ is the conditional density of the data by assuming ξ_i's are i.i.d.. Let θ_N denote a random vector that follows the posterior distribution $p(\theta|\xi^{(N)})$. This leads to the following Bayesian Risk Optimization (BRO) formulation

$$\min_{x\in\mathscr{X}} \rho_{\theta_N}\left(\mathbb{E}_{\xi|\theta_N}[H(x,\xi)]\right), \tag{5}$$

where ρ_{θ_N} is a risk functional (such as expectation, mean-variance, Value-at-Risk, Conditional Value-at-Risk) taken with respect to the posterior distribution $p(\theta|\xi^{(N)})$, and $\mathbb{E}_{\xi|\theta}$ is the expectation taken with respect to $f(\xi|\theta)$ conditional on θ. For example, if ρ is chosen to be the expectation, then the objective function in (5) equals $\int_\Theta \int_\Xi H(x,\xi) f(\xi|\theta) p(\theta|\xi^{(N)}) d\xi d\theta$.

Leveraging the well-known Bayesian consistency and Bayesian central limit theorem (a.k.a. the Bernstein-von Mises theorem), Wu et al. (2018) show that the BRO formulation converges almost surely to the true problem (1) as the data size N goes to infinity and that the BRO objective roughly equals a weighted sum of the posterior mean performance and the confidence interval width of the actual performance. Furthermore, numerical methods for solving (5) have been developed in Cakmak et al. (2021) and Cakmak et al. (2020). The former method is based on stochastic gradient descent, and the latter is based on Bayesian optimization.

The parametric assumption on the input distribution may bring in additional model uncertainty in some application problems where we do not have a strong reason for the chosen parametric family. To hedge against this model uncertainty (mis-specification of the parametric family) while still maintaining the advantage of Bayesian estimation when data are limited, Shapiro et al. (2023) propose a new formulation termed Bayesian Distributionally Robust Optimization (Bayesian-DRO), which poses robustness against the model uncertainty of the assumed parametric distributions. Bayesian-DRO constructs an ambiguity set by taking the parametric distribution as the reference distribution and optimizes the worst-case of the Bayesian average of the true problem. More specifically, for every $\theta \in \Theta$ let $\mathscr{M}^\theta$ be a set of probability distributions centered around $f(\xi|\theta)$. For example, $\mathscr{M}^\theta$ can be constructed as a ball with radius defined by the Kullback-Leiber (KL) divergence from the reference distribution $f(\xi|\theta)$. Then the Bayesian-DRO formulation is as follows:

$$\min_{x\in\mathscr{X}} \mathbb{E}_{\theta_N}\left[\sup_{Q\in\mathfrak{M}^{\theta_N}} \mathbb{E}_{Q|\theta_N}[G(x,\xi)]\right],$$

where $\mathbb{E}_{\theta_N}$ denotes the expectation with respect to the posterior distribution $p(\theta|\xi^{(N)})$, $\mathbb{E}_{Q|\theta}$ is the expectation with respect to distribution Q of ξ conditional on θ, and $\mathfrak{M}^\theta$ is referred to as the ambiguity set of input distribution familiy. Please note that the posterior distribution depends on choice of the prior density $p(\theta)$ and parametric model $f(\cdot|\theta)$. The choice of $p(\theta)$ and $f(\cdot|\theta)$ could both be subject to ambiguity, and in that paper they mainly deal with ambiguity with respect to the parametric model.

It should be mentioned that there have been other attempts to extend the BRO framework to non-parametric input distributions. For example, Wang et al. (2021) uses Dirichlet processes as the posterior process for estimating the input distribution in a non-parametric way.

3.4 Simulation Optimization with Streaming Input Data

The Bayesian updating offers a natural way to assimilate data that sequentially arrive over time, and thus, the BRO framework mentioned in the last section can be naturally extended to solve simulation optimization problems with streaming input data and possibly for online decision making. Liu et al. (2024) consider such a problem in two scenarios: decision-independent (exogenous) data, and decision-dependent (endogenous) data. For the scenario of decision-dependent uncertainty, the true problem is

$$\min_{x\in\mathscr{X}} \mathbb{E}_{f(\cdot;x,\theta^c)}[H(x,\xi)] \qquad \text{(decision-dependent uncertainty)} \tag{6}$$

As the distribution parameter θ^c is unknown, the decision maker observes data $\mathbf{y}_t = \{y_{t,1},\cdots,y_{t,D} \overset{\text{iid}}{\sim} f(\cdot|x_t,\theta^c)\}$ that come at each time stage t. To solve this problem, Liu et al. (2024) propose an iterative approach with the following main steps at each time stage t (only the algorithm for the decision-dependent case is shown below):

1. Data: at time t, receive a batch of data $\mathbf{y}_t = \{y_{t,1},\cdots,y_{t,D} \overset{\text{iid}}{\sim} f(\cdot|x_t,\theta^c)\}$.
2. Estimation: update the Bayesian posterior

$$p_t(\theta|\mathbf{y}_1,\ldots,\mathbf{y}_t) \propto p_{t-1}(\theta|\mathbf{y}_1,\ldots,\mathbf{y}_{t-1})\cdot\prod_{j=1}^{D} f(y_{t,j}|x_t,\theta). \tag{7}$$

3. Optimization: consider the Bayesian average problem

$$\mathbb{E}_{\theta_t}[\mathbb{E}_{\xi|\theta_t}[H(x,\xi)]], \tag{8}$$

where $\theta_t \sim p_t(\theta|\mathbf{y}_1,\ldots,\mathbf{y}_t)$. Take one or more iterates of SGD for solving (8), so the decision is updated via

$$x_{t+1} := \text{Proj}_{\mathscr{X}}\left\{x_t - \alpha_t \underbrace{\left(\nabla_x H(x_t,\xi_t) + H(x_t,\xi_t)\frac{\nabla_x \widehat{f}_t(\xi_t;x_t)}{\widehat{f}_t(\xi_t;x_t)}\right)}_{\text{unbiased gradient estimator}}\right\},$$

where $\widehat{f}_t(\cdot;x) = \mathbb{E}_{\theta_t}[f(\cdot|x,\theta)]$, $\xi_t \sim f(\cdot|x_t,\theta_t)$.

In the approach outlined above, Step 1 receives a batch of data $\mathbf{y}_t$ that follow a distribution $f(\cdot|x_t,\theta^c)$ dependent on both the current decision x_t and the unknown parameter θ^c. Step 2 estimates the unknown parameter via Bayesian posterior p_t, which is updated from the previous posterior p_{t-1} and the newly received batch of data. Since the true objective function (6) is not available, Step 3 considers an alternative objective — the Bayesian average problem (8), which is a special case of the BRO formulation (5) when the risk functional ρ is set as the expectation. It then employs stochastic gradient descent (SGD) to solve the Bayesian average problem, but given the limited time in each time stage, only a few iterates of SGD may be carried out to yield an online decision x_{t+1} for the next time stage. If decision-independent uncertainty is considered, the above algorithm simplifies by replacing $f(\cdot|x_t,\theta^c)$ with $f(\cdot|\theta^c)$ and setting the unbiased gradient estimator simply as $\nabla_x H(x_t,\xi_t)$.

A careful reader may notice that the Bayesian updating in (7) is questionable, since classical Bayes' rule assumes data are generated from the same distribution, whereas here the data $\{\mathbf{y}_1,\ldots,\mathbf{y}_t\}$ come from different and correlated distributions. Hence, a key question is whether the Bayes updating rule remains valid in the decision-dependent case. Liu et al. (2024) provide an affirmative answer to this question by showing the strong consistency of the Bayesian posterior in (7) under some mild conditions. The main condition can be intuitively interpreted as that for any decision x, the data generated from distributions of different θ's are distinguishable. Based on the consistency of the posterior distributions, the convergence of the decision sequence can be guaranteed.

3.5 Ranking and Selection with Streaming Input Data

Streaming data are also common in applications of R&S. Wu et al. (2022) consider a fixed confidence ranking and selection (R&S) problem where new input data arrives over time in batches of possibly varying and random sizes. In each time stage, the number of allowable simulation replications is limited, due to the computational time of each simulation replication and the time constraint of each stage. The new data batch can be used to update the input distribution of the simulation model so as to reduce input uncertainty. Because of the limited simulation replications at each time stage, it is necessary to aggregate the simulation outputs across different time stages to reduce estimation error in the performance estimate. More specifically, during each stage the input parameter estimate is updated to $\hat{\theta}_t$, and incremental simulations are run under the input distribution $F_{\hat{\theta}_t}$. Their goal is to design a procedure which terminates after a number of stages and outputs the true best design with probability at least $(1-\alpha)$, where $\alpha \in (0,1)$.

Figure 2 illustrates how simulation outputs are generated for a single design. Here are a few observations.

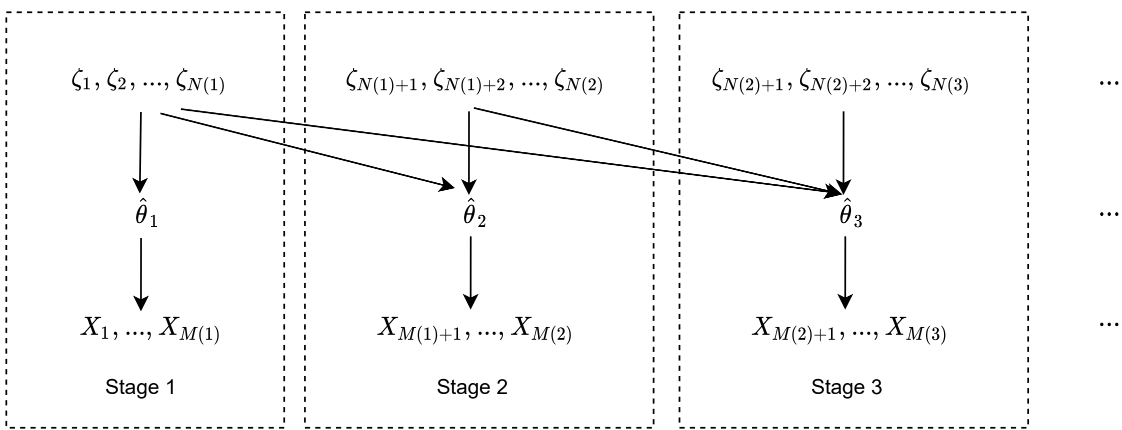

Figure 2: Illustration of simulating a single design with streaming input data, where ζ denotes input data, $\hat{\theta}$ is the input parameter estimate, and X denotes simulation output.

First, the parameter estimates $\{\hat{\theta}_t\}$ are correlated since they are computed using the same stream of input data. Second, since $\{\hat{\theta}_t\}$ are random variables that generally take different values, the simulation outputs X are not identically distributed across different stages. Third, conditioned on $\{\hat{\theta}_t\}$, the outputs X are i.i.d. within the same stage and independent across stages, but unconditional independence no longer holds since X are jointly affected by the correlation among $\{\hat{\theta}_t\}$. In summary, the simulation outputs are neither independent nor identically distributed.

To address the challenge associated with non-i.i.d. simulation outputs, they propose a moving-average performance estimator. Let $\{X_{i,r}(\hat{\theta}_t)\}_{r=1}^{m(t)}$ denote the batch of simulation outputs of the ith design, generated at stage t under the input distribution $P_{\hat{\theta}_t}$, where $m(t)$ denote the number of simulation replications at time t. Let $M(t) = \sum_{\ell=1}^{t} m(\ell)$ denote the total number of simulation replications up to time t. Then, the moving average estimator is as follows:

$$\hat{\mu}_{i,t} := [M(t) - M(t_\eta)]^{-1} \sum_{\ell=t_\eta+1}^{t} \sum_{r=1}^{m(\ell)} X_{i,r}(\hat{\theta}_\ell), \tag{9}$$

where $\eta \in [0,1)$ is called the *drop rate* which controls the amount of effective samples used for computation, and $t_\eta := \lfloor \eta t \rfloor$ is the number of stages "discarded". In words, the estimator in (9) only averages the latest $(1-\eta)$ portion of all the simulation outputs, hence the name "moving average". The idea of throwing away some earlier samples is motivated by the following two extreme cases.

(i) $\eta = 0$: Keeping all the outputs helps reduce simulation uncertainty but also retains all the biases $\{\mu_i(\hat{\theta}_t) - \mu_i(\theta^c)\}$ that accumulate over time.

(ii) $\eta = 1$: This roughly corresponds to keeping only the latest output $X_{i,m(t)}(\hat{\theta}_t)$, which has larger variance due to the limited number of outputs in each time stage.

Hence, the drop rate η characterizes the trade-off between input uncertainty and simulation uncertainty, and can be optimized to minimize the asymptotic variance of the estimator.

Using the moving average estimator, they then build on the sequential elimination (SE) framework to perform R&S. SE was first developed by Even-Dar et al. (2002) and Even-Dar et al. (2006) for multi-armed bandit problems. Ideally, design i will be eliminated if it is inferior to another alternative, i.e., $\eta_i < \eta_j$, for some $j \neq i$. Considering the estimation error in performance estimates, design i would be eliminated if

$$\hat{\delta}_{ij,t} + w_{ij,t} = \hat{\mu}_{i,t} - \hat{\mu}_{j,t} + w_{ij,t} < 0, \textit{ for some } j \neq i,$$

where $\hat{\delta}_{ij,t} = \hat{\mu}_{i,t} - \hat{\mu}_{j,t}$ is the estimated performance difference between design i and design j at time t, and $w_{ij,t}$ is a confidence band of the estimate $\hat{\delta}_{ij,t}$. Hence, design i will be eliminated if the upper confidence of performance difference δ_{ij} is below 0. Then, the key question is to derive the confidence bands $\{w_{ij,t}\}$ to guarantee that the final PCS is at least $1-\alpha$. To do this, notice that the false selection happens at any stage when the true optimal design is eliminated. Therefore, the probability of false selection (PFS) can be written as:

$$\begin{aligned}
\text{PFS} =& \mathbb{P}(\text{The optimal design } b \text{ is eliminated at some } t) \\
=& \mathbb{P}\left(\bigcup_{t=1}^{\infty}\{\text{The optimal design } b \text{ is eliminated at } t\}\right) \\
=& \sum_{t=1}^{\infty} \mathbb{P}\left(\text{The optimal design } b \text{ is eliminated at } t\right) \\
\leq& \sum_{t=1}^{\infty} \mathbb{P}\left(\bigcup_{i<j}\{\mid \hat{\delta}_{ij,t} - \delta_{ij}(\theta^c) \mid > w_{ij,t}\}\right),
\end{aligned}$$

where the second equality follows from the fact that the events {The optimal design b is eliminated at t}, $t = 1,2,\ldots$ are disjoint, and the last inequality follows from

$$\{\text{The optimal design } b \text{ is eliminated at } t\} \subseteq \bigcup_{i<j}\{\mid \hat{\delta}_{ij,t} - \delta_{ij}(\theta^c) \mid > w_{ij,t}\}.$$

Hence, if one can find $w_{ij,t}$ such that

$$\mathbb{P}\left(\bigcup_{i<j}\{\mid \hat{\delta}_{ij,t} - \delta_{ij}(\theta^c) \mid > w_{ij,t}\}\right) \leq \frac{6\alpha}{\pi^2 t^2}, \tag{10}$$

then one has

$$\begin{aligned}
&\sum_{t=1}^{\infty} \mathbb{P}\left(\bigcup_{i<j}\{\mid \hat{\delta}_{ij,t} - \delta_{ij}(\theta^c) \mid > w_{ij,t}\}\right) \\
\leq& \sum_{t=1}^{\infty} \frac{6\alpha}{\pi^2 t^2} = \alpha
\end{aligned}$$

By deriving the exact confidence bands $\{w_{ij,t}\}$ that satisfy (10), PCS is achieved to be at least $1-\alpha$.

To further improve the efficiency of the SE procedure, Wu et al. (2022) also derive asymptotically valid confidence bands, by utilizing the result of Multiple Comparisons with the Best (MCB) that was first proposed by Chang and Hsu (1992). MCB is extended by Song and Nelson (2019) to construct confidence bands accounting for input uncertainty, where they exploit the jointly asymptotic normality of pairwise differences of performance estimators (referring to $\{\hat{\delta}_{ij,t}\}_{j\neq i}$) that consist of samples under the current estimated input distribution. Wu et al. (2022) extend this jointly asymptotic normality to the moving-average estimator. Compared with the SEIU procedure (with exact confidence bands), the SEIU-MCB procedure (with the asymptotically valid confidence bands) requires much less restrictive assumptions and achieves higher efficiency. Furthermore, the MCB framework helps to avoid the usage of Boole's inequality (or union bound) across designs, and as a result, the asymptotic normality result yields much tighter confidence bands to control the cumulative error across stages. In addition, the asymptotic normality result explicitly characterizes how the drop rate η affects the trade-off between input uncertainty and simulation uncertainty.

3.6 Simultaneous Resource Allocation for Data Collection and Simulation

Collecting input data can be costly, and running simulations can be time-consuming, especially for complex models. One direct approach is to wait until sufficient input data has been collected to accurately estimate the input distribution before starting the simulation. However, since input data collection and simulation require different resources (monetary/human resource and computing resource, respectively), conducting both activities simultaneously may be more efficient.

Motivated by this observation, Wang and Zhou (2023) consider a simultaneous resource allocation problem, where input data collection and simulation are conducted in parallel. They propose a multi-stage procedure: at each stage, the input distributions are updated using the data collected from the previous stage, and the design performance is updated using the simulation outputs from the previous stage. Based on the current estimates of input distributions and design performance, resource allocation policies for input data collection and simulation are computed and implemented within their respective stage-wise budgets. These two allocation policies are interdependent, as they jointly influence the PCS.

More specifically, suppose that input data are actively collected with a cost of c_s per unit for $s=1,2,\ldots,S$, where S is the total number of input distributions. Without loss of generality, assume running simulations incurs computing resource at a cost of 1 per unit. At each stage t, one computes a stage-wise allocation policy $\{m_{i,t}\}$ to allocate a stage-wise budget T_S to run the simulation for designs under $\widehat{\theta}_t$, which is the estimation of the input distributions at the beginning of stage t, to obtain samples $X_i^1(\widehat{\theta}_t),\ldots,X_i^{m_{i,t}}(\widehat{\theta}_t)$. At the same time, one also computes a stage-wise allocation policy $\{n_{s,t}\}$ to allocate a stage-wise budget T_I to collect input data and update the input parameter $\widehat{\theta}_{t+1}$. The moving average estimator (9) is used for estimating performance of each design.

To derive the stage-wise budget allocation policies, let $\delta_{ij}(\theta)=\mu_i(\theta)-\mu_j(\theta)$ be the performance difference between design i and j under input parameter θ, $\hat{\delta}_{ij,t}=\hat{\mu}_{i,t}-\hat{\mu}_{j,t}$ be the sample approximation of $\delta_{ij}(\theta^{\mathbf{c}})$, and $\sigma_i^2(\theta)=\mathbf{Var}(X_i(\theta))$ be the performance variance of design i under θ. Assume that in the long run, the budget allocated to a certain design i (or input distribution s) is approximately $t\bar{m}_i$ (or $t\bar{n}_s$). They apply the large deviations theory to derive the asymptotic exponential convergence rate of PCS, and then formulate an optimization problem to maximize this rate subject to the two budget constraints ($\sum_{s=1}^{S} c_s\bar{n}_s = T_I$, $\sum_{i=1}^{K}\bar{m}_i = T_S$). Ignoring the minor issue of $t\bar{m}_i$ and $t\bar{n}_s$ not being integers and relaxing them to be continuous variables, thes rate optimization problem is a convex problem. Hence, the Karush-Kuhn-Tucker (KKT) condition charaterzies its optimality condition, which presents the following three sets of equations:

- (Local Balance) $\frac{\delta_{bi}^2(\theta^{\mathbf{c}})}{\tilde{\sigma}_{bi}^2} = \frac{\delta_{bj}^2(\theta^{\mathbf{c}})}{\tilde{\sigma}_{bj}^2} \quad \forall i\neq j\neq b$

- (Global Balance) $\bar{m}_b^2 = \sigma_b^2(\theta^{\mathbf{c}}) \sum_{i \neq b} \frac{\bar{m}_i^2}{\sigma_i^2(\theta^{\mathbf{c}})}$
- (IU Balance) $\frac{1}{c_s \bar{n}_s^2} \sum_{i \neq b} \frac{\bar{m}_i^2}{\sigma_i^2(\theta^{\mathbf{c}})} g(i,s) = \frac{1}{c_{s'} \bar{n}_{s'}^2} \sum_{i \neq b} \frac{\bar{m}_i^2}{\sigma_i^2(\theta^{\mathbf{c}})} g(i,s') \quad \forall s \neq s'$,

where $g(i,s) = \partial_{\theta_s} \delta_{bi}(\theta^{\mathbf{c}})^\intercal \Sigma_{D,s} \partial_{\theta_s} \delta_{bi}(\theta^{\mathbf{c}})$, $\tilde{\sigma}_{ij}^2 = \lambda_{I,\eta} \sum_{s=1}^{S} \bar{n}_s^{-1} g(i,s) + \lambda_{S,\eta} \bar{m}_i^{-1} \sigma_i^2(\theta^c) + \lambda_{S,\eta} \bar{m}_j^{-1} \sigma_j^2(\theta^c))$, ∂_{θ_s} is the partial derivative taken with respect to θ_s, $\lambda_{I,\eta} = \left(\frac{2}{1-\eta} + \frac{2\eta \ln \eta}{(1-\eta)^2} \right)$ and $\lambda_{S,\eta} = \frac{1}{1-\eta}$. Please note that $\tilde{\sigma}_{ij}^2$ is the asymptotic variance of $\sqrt{t}\widehat{\delta}_{ij,t}$, as shown in the asymptotic normality result in their paper.

The "Local Balance" equation and the "Global Balance" equation, two of the three optimality equations mentioned above, are similar in form to those in previous works that do not consider input uncertainty (Glynn and Juneja (2004)) or use passively obtained streaming input data (Wang and Zhou (2022)). However, the rate function in the “Local Balance" equation now includes both the allocation policy for the simulation budget and the allocation policy for input data collection, since they jointly determine the variance term $\tilde{\sigma}_{bi}^2$. The third optimality equation, the “IU Balance” equation, can be used to allocate budget for input data collection to reduce input uncertainty. The “IU Balance” equation suggests that the amount of input data allocated to a particular distribution $F_{\theta_s^c}$ depends on three factors in addition to the cost: the simulation effort $\bar{m}_i$, the simulation noise $\sigma_i^2(\theta^{\mathbf{c}})$ for design $i \neq b$, and its sensitivity $g(i,s)$ to the input parameter θ_s^c. When the function $g(i,s)$ is large, the performance of design i is more sensitive to the input parameter θ_s^c. If additionally $\frac{\bar{m}_i^2}{\sigma_i^2(\theta^{\mathbf{c}})}$ is also large, which means the simulation error of design i is small, more effort should be devoted to data collection for input distribution θ_s to reduce input uncertainty.

Leveraging these optimality conditions (balance equations), they develop a fully sequential algorithm using a ”Balancing” approach (Chen and Ryzhov 2022). This approach iteratively reduces the gap between two sides of the balance equations, aiming to achieve the equations in the limit.

4 CONCLUSIONS

This tutorial discusses the new challenges in simulation optimization arising from digital twins. Distinct from traditional simulators, digital twins synchronize with the physical system via streaming data and decisions in real time. Therefore, data-driven simulation optimization plays an extremely important role in the application of digital twins to real-world problems. Research on data-driven simulation optimization is still in its infancy, and it can be envisioned that there will be a great need for new efficient methodologies in this area.

REFERENCES

Barton, R. R., B. L. Nelson, and W. Xie. 2013. “Quantifying input uncertainty via simulation confidence intervals”. *INFORMS Journal on Computing* 26(1):74–87.

Barton, R. R. and L. W. Schruben. 1993. “Uniform and bootstrap resampling of empirical distributions”. In *Proceedings of the 1993 Winter Simulation Conference*, 503–508 https://doi.org/10.1109/WSC.1993.718091.

Barton, R. R. and L. W. Schruben. 2001. “Resampling methods for input modeling”. In *Proceedings of the 2001 Winter Simulation Conference*, 372–378 https://doi.org/10.1109/WSC.2001.977303.

Biller, B., X. Jiang, J. Yi, P. Venditti and S. Biller. 2022. “Simulation: the Critical Technology in Digital Twin Development”. In *Proceedings of 2022 Winter Simulation Conference*, 1340–1355 https://doi.org/10.1109/WSC57314.2022.10015246.

Cakmak, S., R. Astudillo, P. Frazier, and E. Zhou. 2020. “Bayesian Optimization of Risk Measures”. In *Advances in Neural Information Processing Systems 34 (NeurIPS 2020) Proceedings*, 20130 – 20141.

Cakmak, S., Y. Wang, S. Gao, and E. Zhou. 2024. “Contextual Ranking and Selection with Gaussian Processes and OCBA”. *ACM Transactions on Modeling and Computer Simulation (special issue on “From Data to Decision-making: Contending with Uncertainty and Non-Stationarity in Simulation Theory”)* 34(2):8:1–24.

Cakmak, S., D. Wu, and E. Zhou. 2021. “Solving Bayesian Risk Optimization via Nested Stochastic Gradient Estimation”. *IISE Transactions*:1081–1093.

Chang, J. Y. and J. C. Hsu. 1992. “Optimal designs for multiple comparisons with the best”. *Journal of Statistical Planning and Inference* 30(1):45–62.

Chen, Y. and I. O. Ryzhov. 2022. "Balancing Optimal Large Deviations in Sequential Selection". *Management Science* 69(6):3457–3473.

Cheng, R. C. and W. Holloand. 1997. "Sensitivity of computer simulation experiments to errors in input data". *Journal of Statistical Computation and Simulation* 57(1-4):219–241.

Chick, S. E. 2001. "Input distribution selection for simulation experiments: accounting for input uncertainty". *Operations Research* 49(5):744–758.

Corlu, C. G. and B. Biller. 2013. "A subset selection procedure under input parameter uncertainty". In *Proceedings of the 2013 Winter Simulation*, 463–473 https://doi.org/10.1109/WSC.2013.6721442.

Corlu, C. G. and B. Biller. 2015. "Subset selection for simulations accounting for input uncertainty". In *Proceedings of the 2015 Winter Simulation Conference*, 437–446 https://doi.org/10.1109/WSC.2015.7408185.

Even-Dar, E., S. Mannor, and Y. Mansour. 2002. "PAC bounds for multi-armed bandit and Markov decision processes". In *International Conference on Computational Learning Theory*, 255–270.

Even-Dar, E., S. Mannor, and Y. Mansour. 2006. "Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems". *Journal of machine learning research* 7(Jun):1079–1105.

Fan, W., L. J. Hong, and X. Zhang. 2020. "Distributionally robust selection of the best". *Management Science* 66(1):190–208.

Feng, M. and E. Song. 2019. "Efficient Input Uncertainty Quantification via Green Simulation Using Sample Path Likelihood Ratios". In *Proceedings of the 2019 Winter Simulation Conference* https://doi.org/10.1109/WSC40007.2019.9004835.

Fu, M. 2015. *Handbook of Simulation Optimization*, Volume Part of the book series: International Series in Operations Research Management Science, volume 216. Springer https://doi.org/10.1007/978-1-4939-1384-8.

Gao, S., J. Du, and C.-H. Chen. 2019. "Selecting the Optimal System Design under Covariates". In *15^{th} International Conference on Automation Science and Engineering*.

Gao, S., H. Xiao, E. Zhou, and W. Chen. 2017. "Robust ranking and selection with optimal computing budget allocation". *Automatica* 81:30–36.

Glynn, P. and S. Juneja. 2004. "A Large Deviations Perspective on Ordinal Optimization". In *Proceedings of the 2004 Winter Simulation Conference, 2004.*, 585–593 https://doi.org/10.1109/WSC.2004.1371364.

He, L., U. V. Shanbhag, and E. Song. 2024. "Stochastic Approximation for Multi-period Simulation Optimization with Streaming Input Data". *ACM Transactions on Modeling and Computer Simulation* 34(24):6:1–27.

Hong, J. and G. Jiang. 2019, 12. "Offline Simulation Online Application: A New Framework of Simulation-Based Decision Making". *Asia-Pacific Journal of Operational Research* 36:1940015.

Jiang, G., L. Hong, and B. Nelson. 2019, 10. "Online Risk Monitoring Using Offline Simulation". *INFORMS Journal on Computing* 32(2):356–375.

Jin, X., Y. Shen, H. Li, L. H. Lee, E. P. Chew and C. Shoemaker. 2020, 08. "Combining Adaptive Budget Allocation with Surrogate Methodology in Solving Continuous Scenario-based Simulation Optimization". In *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, 1509–1513.

Kim, T. and E. Song. 2022. "Optimizing Input Data Acquisition for Ranking and Selection: A View Through the Most Probable Best". In *Proceedings of the 2022 Winter Simulation Conference*, 2258–2269 https://doi.org/10.1109/WSC57314.2022.10015453.

Lam, H. 2016. "Robust sensitivity analysis for stochastic systems". *Mathematics of Operations Research* 41(4):1248–1275.

Lam, H. and H. Qian. 2016. "The empirical likelihood approach to simulation input uncertainty". In *Proceedings of the 2016 Winter Simulation Conference (WSC)*, 791–802 https://doi.org/10.1109/WSC.2016.7822142.

Lam, H. and H. Qian. 2018. "Subsampling variance for input uncertainty quantification". In *2018 Winter Simulation Conference (WSC)*, 1611–1622 https://doi.org/10.1109/WSC.2018.8632256.

Lam, H. and E. Zhou. 2015. "Quantifying Uncertainty in Sample Average Approximation". In *Proceedings of Winter Simulation Conference*, 3846–3857 https://doi.org/10.1109/WSC.2015.7408541.

Lam, H. and E. Zhou. 2017. "The Empirical Likelihood Approach to Quantifying Uncertainty in Sample Average Approximation". *Operations Research Letters* 45(4):301–307.

Li, H., X. Cao, P. Sharma, L. H. Lee and E. P. Chew. 2020. "Framework of O2DES.NET Digital Twins for Next Generation Ports and Warehouse Solutions". In *Proceedings of 2020 Winter Simulation Conference (WSC)*, 3188–3199 https://doi.org/10.1109/WSC48552.2020.9384111.

Li, H., H. Lam, Z. Liang, and Y. Peng. 2020. "Context-Dependent Ranking and Selection under a Bayesian Framework". In *Proceedings of the 2020 Winter Simulation Conference*, 2060–2070 https://doi.org/10.1109/WSC48552.2020.9383965.

Li, H., H. Lam, and Y. Peng. 2022. "Efficient Learning for Clustering and Optimizing Context-Dependent Designs". *Operations Research*.

Lin, Y. and B. L. Nelson. 2016. "Simulation analytics for virtual statistics via k nearest neighbors". In *Proceedings of the 2016 Winter Simulation Conference*, 448–459 https://doi.org/10.1109/WSC.2016.7822111.

Lin, Y., E. Song, and B. L. Nelson. 2015. "Single-experiment input uncertainty". *Journal of Simulation* 9(3):249–259.

Liu, H., X. Jin, H. Li, L. H. Lee and E. P. Chew. 2021. "A Unified Offline-Online Learning Paradigm via Simulation for Scenario-Dependent Selection". In *Proceedings of the Winter Simulation Conference* https://doi.org/10.1109/WSC52266.2021.9715378.

Liu, T., Y. Lin, and E. Zhou. 2021. "A Bayesian approach to online simulation optimization with streaming input data". In *Proceedings of the 2021 Winter Simulation Conference* https://doi.org/10.1109/WSC52266.2021.9715392.

Liu, T., Y. Lin, and E. Zhou. 2024. "Bayesian Stochastic Gradient Descent for Stochastic Optimization with Streaming Input Data". *SIAM Journal on Optimization* (1):389–418.

Liu, T. and E. Zhou. 2019. "Online Quantification of Input Model Uncertainty by Two-Layer Importance Sampling". *arXiv preprint arXiv: 1912.11172*.

Matta, A. and G. Lugaresi. 2023. "Digital Twins: Features, Models, and Services". In *Proceedings of the 2023 Winter Simulation Conference*, 46–60 https://doi.org/10.1109/WSC60868.2023.10407260.

National Academies of Sciences, Engineering, and Medicine 2024. "Foundational Research Gaps and Future Directions for Digital Twins" https://doi.org/https://doi.org/10.17226/26894.

Nelson, B. L. 2016. "Some tactical problems in digital simulation' for the next ten years". *Journal of Simulation* 10(1):2–11.

Shapiro, A., E. Zhou, and Y. Lin. 2023. "Bayesian Distributionally Robust Optimization". *SIAM Journal on Optimization* 33(2):1279–1304.

Shen, H., L. Hong, and X. Zhang. 2021. "Ranking and Selection with Covariates for Personalized Decision Making". *INFORMS Journal on Computing* 33(4):1500–1519.

Song, E. and B. L. Nelson. 2019. "Input–output uncertainty comparisons for discrete optimization via simulation". *Operations Research* 67(2):562–576.

Wang, H., X. Zhang, S. H. Ng, and S. Wang. 2021. "A Nonparametric Bayesian Approach for Simulation Optimization with Input Uncertainty". *arXiv preprint arXiv: 2008.02154*.

Wang, Y. and E. Zhou. 2022. "Fixed Budget Ranking and Selection with Streaming Input Data". In *Proceedings of the 2022 Winter Simulation Conference*, 3027–3038 https://doi.org/10.1109/WSC57314.2022.10015327.

Wang, Y. and E. Zhou. 2023. "Input Data Collection versus Simulation: Simultaneous Resource Allocation". In *Proceedings of the 2023 Winter Simulation Conference*, 3657–3668 https://doi.org/10.1109/WSC60868.2023.10408130.

Wu, D., Y. Wang, and E. Zhou. 2022. "Data-driven Ranking and Selection under Input Uncertainty". *Operations Research* 72(2):781–795.

Wu, D. and E. Zhou. 2017. "Ranking and selection under input uncertainty: a budget allocation formulation". In *Proceedings of the 2017 Winter Simulation Conference*, 2245–2256 https://doi.org/10.1109/WSC.2017.8247956.

Wu, D., H. Zhu, and E. Zhou. 2018. "A Bayesian Risk Approach to Data-driven Stochastic Optimization: Formulations and Asymptotics". *SIAM Journal on Optimization* 28(2):1588–1612.

Xiao, H., F. Gao, and L. H. Lee. 2020. "Optimal computing budget allocation for complete ranking with input uncertainty". *IISE Transactions* 52(5):489–499.

Xiao, H. and S. Gao. 2018. "Simulation budget allocation for selecting the top-m designs with input uncertainty". *IEEE Transactions on Automatic Control* 63(9):3127–3134.

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.

Xu, J., Z. Zheng, and P. W. Glynn. 2020. "Joint Resource Allocation for Input Data Collection and Simulation". In *Proceedings of the 2020 Winter Simulation Conference*, 2126–2137 https://doi.org/10.1109/WSC48552.2020.9384050.

Zhou, E. and W. Xie. 2015. "Simulation optimization when facing input uncertainty". In *Proceedings of the 2015 Winter Simulation Conference*, 3714–3724 https://doi.org/10.1109/WSC.2015.7408529.

Zhu, H., T. Liu, and E. Zhou. 2020. "Risk quantification in stochastic simulation under input uncertainty". *ACM Transactions on Modeling and Computer Simulation* 30(1):1:1–24.

Zouaoui, F. and J. R. Wilson. 2003. "Accounting for parameter uncertainty in simulation input modeling". *IIE Transactions* 35:781–792.

Zouaoui, F. and J. R. Wilson. 2004. "Accounting for input-model and input-parameter uncertainties in simulation". *IIE Transactions* 36(11):1135–1151.

ACKNOWLEDGMENT

The author gratefully acknowledges the support by the Air Force Office of Scientific Research under Grant FA9550-22-1-0244 and the National Science Foundation under Grant NSF-DMS2053489.

AUTHOR BIOGRAPHIES

ENLU ZHOU is a Fouts Family Professor in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology. She received the B.S. degree with highest honors in electrical engineering from Zhejiang University,

China, in 2004, and received the Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 2009. Her research interests include simulation optimization, stochastic optimization, and stochastic control. Her email address is enlu.zhou@isye.gatech.edu, and her web page is https://www.enluzhou.gatech.edu/.