Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

ROBUSTNESS REVISITED: SIMULATION OPTIMIZATION VIEWED THROUGH A DIFFERENT LENS

Susan M. Sanchez
Paul J. Sanchez

Operations Research Department
Naval Postgraduate School
1 University Circle
Monterey, CA 93943, USA

ABSTRACT

We start by introducing key concepts in robust design and analysis, and demonstrate how robustness often changes our perspective when contrasted with simulation optimization approaches. After defining basic terminology, we present several numerical examples with discussions of how to apply these techniques in qualitative, quantitative, and optimization contexts. Evaluating responses using loss functions can yield solutions and results that are substantially different from those based solely on expected values. Benefits in engineering practice include that robust solutions are advantageous in moving from new product development to production, in focusing decision makers on controllable aspects of their problem, and in facilitating communication between the various stakeholders. Robust solutions are designed to yield consistently good performance even in the face of uncertainty and uncontrollable factors by incorporating those aspects of the system into the problem formulation.

1 INTRODUCTION

What is robust design? It is a system optimization and improvement process that springs from the view that a system should not be evaluated on the basis of mean performance alone. In addition to exhibiting an acceptable mean performance, a “good” system must be relatively insensitive to uncontrollable sources of variation present in the system’s environment. The goal of robust design is to lead to better decisions by leveraging several benefits:

- robust design yields insights into what drives system performance;
- it focuses the decision-making process on factors that are controllable in practice;
- it identifies levels and consistency of performance based on those controllable factors;
- robust configurations are more likely to yield better engineering implementations;
- those real-world implementations have in many cases achieved greater reliability and performance at lower cost.

The robust design approach originated in quality planning and engineering product design activities (Taguchi 1987). Taguchi found that it was often more costly to control causes of manufacturing variation than to make a process insensitive to these variations, and through the use of simple experimental designs and *loss functions* it was often possible to greatly improve product performance by “building in” the quality. Taguchi’s philosophy and strategy were widely praised in both the applied statistics and manufacturing communities (Pignatiello Jr. and Ramberg 1991), but many of the methods and tactics he advocated were controversial (Box 1988; Nair et al. 1992). The approach described in this paper (see also Sanchez

978-1-7281-9499-8/20/$31.00 ©2020 IEEE

et al. (1998) and Ramberg et al. (1991)) combines Taguchi's strategy with response-surface metamodeling techniques for investigating robustness of simulation responses to quantitative factors, and illustrates how Taguchi's strategy can be incorporated into ranking and selection procedures for investigating robustness for qualitative factors. The additional insights that can be gained make a robust approach particularly beneficial when analyzing simulations of complex systems.

In the simulation context, robust design can be viewed from two slightly different perspectives. One view is that simulation is used primarily as a surrogate for a real system, because of the cost, time, and risks required to make and observe changes in a real system; a second view is that robust design is an integral part of the simulation process.

From the first perspective, the total time required to perform an experiment using robust design is greatly reduced, but the designs and analyses used are the same as those that would be applied to a physical system if cost and time permitted. Applications have included the product designers' uses of computer models for experimentation in place of physical prototypes, particularly in the semiconductor industry (Sacks et al. 1989; Welch et al. 1990). These experiments have typically involved Monte Carlo simulation, although clearly robustness can be used as a criteria for evaluating discrete-event simulation systems as well. Those who use simulation to study systems primarily because of the difficulty of experimenting on the real system may realize the benefits of improved performance and decreased cost cited by many manufacturers if they decide to evaluate performance in terms of robustness.

Applying robust design principles to simulation experiments is discussed in Sanchez (2000). A more detailed discussion and examples appear in Kleijnen et al. (2005), where *identifying robust systems and processes* is considered one of three primary goals of simulation experiments. However, both of these earlier references are dated—we would now recommend the use of more capable designs which did not exist when those papers were written. Others also advocate the use of response surface metamodeling in conjunction with the robust design philosophy to identify robust systems, including Kleijnen et al. (2005), Dellino et al. (2009), Dellino et al. (2010), and Kleijnen (2017).

The second perspective encompasses the process of building as well as analyzing a simulation model. A simulation model is constructed assuming a variety of system inputs (e.g., distributional forms and characteristics, simplifying assumptions, level of detail) which are unlikely to be completely accurate. Model verification and validation are important issues, as is simulation sensitivity analysis. From this perspective, one can view robust design as a process of simulation optimization, where the "best" answer is not overly sensitive to small changes in the system inputs. Kleijnen (2017) calls this "robust optimization." If robust configurations are identified, then the actual results are more likely to conform to the anticipated results after implementation. This approach can also lead to model simplification or reduced cost if the model is found to be insensitive to distributional assumptions or wide ranges of parameterizations—this could negate the need for detailed modeling of subsystems and corresponding data collection.

Section 2 contains an overview of key concepts in robustness. In Section 3 we discuss robust analysis and optimization within two settings—one involving only qualitative factors, another suitable for quantitative factors. A simple queuing model provides numerical illustrations. We conclude with a few thoughts about bringing together the robustness and optimization mindsets.

2 KEY CONCEPTS IN ROBUSTNESS

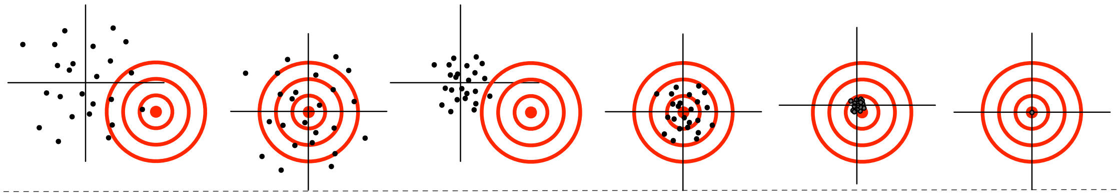
2.1 Accuracy and Precision

The English language can be quite ambiguous, so technical fields often refine the definitions of particular words to avoid miscommunication. In statistics:

Accuracy refers to the location of the expected value of the distribution of outcomes relative to some intended target. For instance, if our target is to find the true weight of an object, a scale which yields readings with a distribution that has that true weight as its expected value is deemed to be an accurate scale even if individual readings might vary by a large amount.

Precision refers to the spread of the distribution of outcomes. If the scale gives readings that are tightly clustered, it is deemed to be a precise scale.

Figure 1 illustrates many possible combinations of accuracy and precision for somebody shooting at a target. Subplots (a) and (c) have shots with low accuracy, because their center of mass (indicated by the cross-hairs) is not at the center of the target. Subplots (b) and (d) have high accuracy—the center of mass is on-target, regardless of the spread. Subplots (a) and (b) have low precision with a large amount of spread. Subplots (c) and (d) have moderate precision due to their reduced spread.



(a) low accuracy low precision (b) high accuracy low precision (c) low accuracy moderate precision (d) high accuracy moderate precision (e) high accuracy high precision (f) ideal accuracy ideal precision

Figure 1: Visual representations of accuracy vs. precision.

If those subplots in Figure 1(a)–(d) represent the only four alternatives available, we would clearly prefer (d)—each of the other three are worse in terms of at least one of the accuracy or precision criteria. However, it is less clear which of those other three options would be second-best. If the goal is to have a high chance of hitting anywhere in the target area, then the low-precision, high-accuracy alternative in (b) would place second, followed by the low-precision low-accuracy alternative in (a), with the worst alternative being the moderate-precision, low-accuracy alternative in (c). However, if it is easy to calibrate the weapon (that is, to adjust the crosshairs) then (c) can quickly be turned into a good option.

All of the options in Figure 1(a)–(d) could be improved. Figure 1(e) has very high (though not perfect) accuracy and is much more precise. Figure 1(f) is the ideal—it is perfectly accurate and perfectly precise.

2.2 Inferring Causality

If observational data reveals a correlation between two variables X and Y, ground truth is one of the following four basic cases: (i) changes in X cause changes in Y; (ii) changes in Y cause changes in X; (iii) changes in X and Y are both caused by changes in other potentially unknown or unobservable factors; or (iv) this is a spurious correlation. One drawback of observational data is that we have no real way of determining which of these cases applies.

The simplest way of establishing cause-and-effect is via an experiment. Design of experiments (DOE) is a well established field (see, e.g., Montgomery 2017 or Ryan 2007). Three important concepts in DOE are *control*, *randomization*, and *replication*. For real-world experiments, we exercise control over the situation by deciding which inputs X are of interest. We also decide how to control for everything else that is not of interest, perhaps by holding it constant or using a control group for comparison purposes. Randomization is used to guard against hidden or uncontrollable sources of bias. For example, if we are measuring the miles per gallon of different vehicles by using a single driver, randomizing the order in which they are driven will remove any systematic bias due to fatigue or practice effects. With replication we collect multiple observations to assess the magnitude of the variability associated with Y, so we can construct confidence intervals or conduct significance tests.

In simulation experiments, the analyst has total control, and we use these concepts in different ways. Potential factors in simulation experiments include the inputs and distributional parameters of a simulation model, whether or not they are controllable in the real world. The analyst can also control the random number seeds and streams. Thus, unlike a physical experiment, the results from a simulation experiment are perfectly repeatable, and randomization is not needed to guard against hidden or uncontrollable sources of bias. Replication means we get multiple experimental units (runs or batches) to gain a sense of the magnitude of the variability associated with Y. While homogeneous (i.e., constant) variance is commonly assumed for physical experiments, heterogeneous (i.e., non-constant) variance is pervasive in stochastic simulation. Consequently, we should not view response variability as merely a nuisance for estimating means or other output statistics, but as an important characteristic of the simulation's behavior.

One of the first things an experimenter or tester must do to design a good experiment is identify the experiment's factors. In DOE parlance, *factors* are the input (or independent) variables that potentially have some impact on *responses* (i.e., experimental outputs). In general, an experiment might have many factors, each of which might assume a variety of values, called *levels* of the factor. A primary goal of many DOEs is to identify which of the factors play a significant role in determining the responses, and which do not and can thus be dropped from further consideration, greatly reducing the experimental effort and simplifying the task of interpreting the results. Also, for the important factors we would like to identify the nature of the impact on the responses (e.g., increasing, linear, quadratic), and whether the levels of some factors influence the effects that other factors have—a phenomenon known as *factor interactions*.

Factors can be inputs to the simulation, such as the number of servers (n) and the arrival and service rates (λ, μ) for an M/M/1 queue. Alternatively, the inputs may be functions of the factors. For example, n, μ, and the traffic intensity ρ would be equally valid as factors which can be converted to n, μ, and λ if those are the inputs to our M/M/1 simulation. Using the second parameterization rather than the first would help us avoid scenarios where varying μ and λ independently could inadvertently lead to unstable queues. Each factor has different levels: n may be an integer ranging from 1 to 10, ρ may be varied over the interval [0.70, 0.95], etc. A key concept to remember is that factors are changed in deliberate (structured) ways when conducting a designed experiment. In this way, a simulation analyst will be able to uncover cause-and-effect relationships within the context of the simulation model.

2.3 Characterizing Factors: Decision, Noise, and Artificial

In systems where stochastic variation is present, the response exhibits random fluctuation or variation. In order to achieve systems or products for which the variation around the target value is low, several steps are necessary. First, one must identify factors in the system which are anticipated to affect the system response. The factors are classified as *decision factors*, *noise factors*, or *artificial factors*. Notationally, we will denote $X_1, \ldots, X_k$ as the k decision factors of interest, $W_1, \ldots, W_w$ as the w noise factors, and $A_1, \ldots, A_a$ as the a artificial factors (if any).

Decision factors are those which are (or will be) controllable in the real-world setting being modeled by the simulation. Noise factors are those which are not easily controllable, or are controllable only at great expense in the real-world setting. Noise factors include intrinsic sources of variation within the real-world system (e.g., within a manufacturing plant) as well as exogenous factors (such as customer or supplier characteristics). Finally, artificial factors are specific to simulation modeling, such as the initial state of the system, the warm-up period (truncation point), termination conditions (run duration), or random number stream(s) (seed values, whether antithetic RVs are used, etc.). Consider, for example, a simulation of search-and-rescue operations after a natural disaster. The decision factors might include choices of the communication systems, the number of people and desired skillsets of the rescue team, and equipping and resupply policies. Noise factors might include weather conditions, the number and location of those in need of rescue, and the skill levels of the currently available emergency medical technicians.

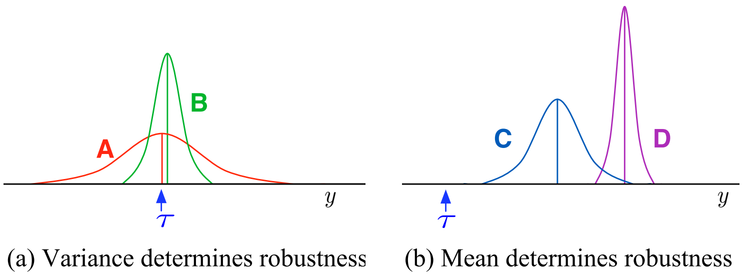
The distinction between decision factors and noise factors is often recognized in simulation experiments, and can be leveraged to advantage when studying the system. The classification is important in several ways. It is necessary for determining system robustness, and also presents an opportunity for reducing the number of runs required by concentrating sampling efforts on assessing decision factor effects, since these are the only ones we can control in the real-world system. The additional layer of control made possible by the artificial factors can also be exploited in the experimental design (Schruben et al. 1992). This is not new—it is the basis of many variance reallocation techniques. However, manipulating artificial factors is not required, and will not be discussed further in this paper.

2.4 Responses, Targets, and Goals

The simplest form of experiment to discuss involves a single categorical factor, where the purpose of the simulation study may be to select one or more alternatives from a list of potential alternatives. This is the basis of ranking and selection (R&S) procedures (Kim 2013), which are a method of discrete simulation optimization. Suppose we are considering which piece of equipment to buy and four vendors have placed bids for contracting, providing us with $\{A,B,C,D\}$ as the set of alternatives. Each of these has different service time and reliability characteristics, and we can simulate usage of the alternative systems to measure some output of interest. We evaluate the alternatives using the following concepts:

- **Response** – the output of interest.
- **Target** – an ideal value for the response, denoted as τ.
- **Goal** – whether we wish the response to be as large as possible, as small as possible, or as close to the target value as possible.

We start by considering cases where we want to get as close to the target as possible. Figure 2(a) shows the true (but unknown) distributions for systems A and B. The mean response for A is exactly on target, so it is best in terms of accuracy. If our goal was to choose the most accurate system, A would be the ideal choice. However, system B is a close contender in terms of the mean and has greater precision—it is much more likely to yield values close to τ because of its lower variance. Based on this example, we might even say that from a robustness perspective, the means do not justify the ends.



(a) Variance determines robustness (b) Mean determines robustness

Figure 2: Robustness of qualitatively different systems.

Figure 2(b) shows a different type of configuration that might arise. Both C and D have mean performance above the target value, but option D has a much higher likelihood of being farther from the target when we take precision into account. Hence C is the more robust choice, despite having a larger variance. The implication of Figures 2(a) and (b) is that robustness is not solely determined by either the accuracy or the precision of the outcomes relative to the target. Tradeoffs may be necessary.

2.5 Loss Functions

The discussion in Section 2.4 was qualitative, but in situations where robustness may involve tradeoffs between the means and variances of different alternatives, we need a more formal means of quantifying robustness.

2.5.1 Generic Loss Functions

Robust analysis is based on the use of a *loss function*. Loss functions should be chosen to assess the degree of risk associated with having outcomes that deviate significantly from the specified target. Risk should be non-negative, so loss functions are monotonically non-decreasing as the magnitude of deviations from target increases. However, a loss function can be asymmetric about the target. For example, the loss of revenue or risk to human lives may be higher by undershooting the target than by overshooting, or vice-versa.

2.5.2 Quadratic Loss

One particular loss function that is easy to evaluate is *quadratic loss*. It takes the form

$$\ell_{Y_x} = c(Y_x - \tau)^2,$$

where Y_x is the observed outcome based on input x, τ is the target value, and c is a scaling constant that is often used to adjust the loss to cost. When c is set to 1, it is referred to as *scaled loss*.

If we add and subtract $E[Y_x]$ inside the squared term, group terms appropriately, and take expectation, it quickly falls out that the expected scaled loss is

$$E[\ell_{Y_x}] = \sigma^2_{Y_x} + (E[Y_x] - \tau)^2,$$

i.e., expected loss can be decomposed into the sum of the variance of Y_x and the square of how far the expected value of Y_x is from the desired target. This result offers a nice interpretation—low loss configurations are those that are consistently (as measured by variance) close to the target (as measured by squared deviations of the mean from the target). Getting the lowest expected loss may involve a tradeoff where we will accept small expected deviations from the target if they are associated with sufficient improvement in consistency of the outcomes, or vice-versa. Note that this implies that quadratic loss is only relevant when the variance is non-homogeneous, since with homogeneous variance the low loss configuration will be solely determined by how close the mean is to the target.

3 Analysis Techniques

Sanchez (2020) compares and contrasts three broad mindsets with which analysts may approach simulation studies: optimization, understanding, and robustness. She lays out several ways in which a robust mindset can be used by simulation professionals. First, for essentially no additional effort, the analyst can step back from the problem and see if it is possible to identify an ideal target value for the response, and for which some type of goal (the bigger the better, the smaller the better, or the closer to target the better) is appropriate. If so, it can be straightforward to apply simulation optimization techniques to seek to minimize expected loss rather than the mean response. We will demonstrate this in Section 3.1.1. Second, for a little additional effort, the analyst can use designed experiments for noise factors to better assess the losses associated with the systems being modeled. We illustrate this for a two-stage optimization procedure in Section 3.1.2. Finally, if many of the decision factors are (or can be cast as) quantitative, then a large-scale designed experiment may be appropriate not just in order to identify robust solutions, but also to reveal the key drivers of simulation performance. We discuss this more in Section 3.2.

3.1 Robust Analysis and Optimization with Qualitative Factors

A common case of categorical selection occurs when assessing the suitability of goods, services, processes, or policies from several alternatives. The response may be quantitative, but the alternatives are limited to a fixed set of choices that may differ in qualitative as well as quantitative dimensions. The decision maker does not have the option of specifying an ideal choice, but must choose amongst a fixed number of available options. This is a classic ranking and selection problem.

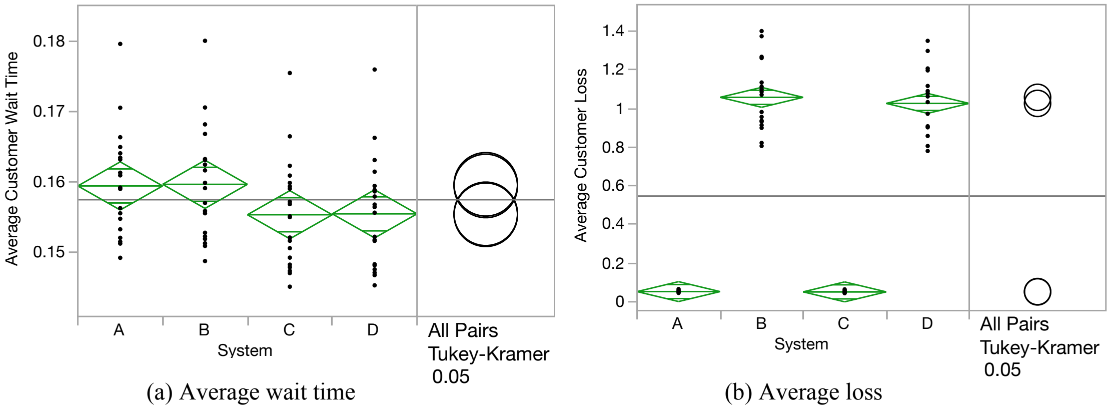
3.1.1 Qualitative Example: M/M/k: FIFO vs LIFO

Our first example involves simple M/M/k queues with arrival rate 114 per hour. Suppose we have four alternatives:

- A: a single server FIFO with service rate $\mu = 120$ per hour;
- B: a single server LIFO with service rate $\mu = 120$ per hour;
- C: a FIFO queue with two servers, each with service rate $\mu = 60$ per hour; and
- D: a LIFO queue with two servers, each with service rate $\mu = 60$ per hour.

For each of the four alternatives, we stop running the model at the first point in time that there are no customers waiting in line after 1,000,000 customers have been served. This removes the bias that would be introduced by ignoring LIFO customers who are "bottled up" in the queue when the nominal stopping condition of a million customers served is reached, by continuing operations until all of those customers have been removed from the queue. As a result the number of customers observed is a random variable, but we minimize the potential for termination bias by using common random numbers for each replicate of the set of all four alternatives. In other words, systems A–D all have the exact same set of customers, arriving at the same times, and having the same transactions and their associated service times. The simulation reports the average, standard deviation, and maximum of the customer wait times, as well as the average quadratic loss. We make 20 replications of the experiment. Altogether, this is a total of $2^2 \times 20 = 80$ simulation runs representing 80 million customers.

The analytical results obtained from queueing theory are that customers have an expected delay in queue of $0.158\overline{33}$ hours (9m30s) in the line with one server, vs. 0.1543 hours (9m15s) when there are two servers, regardless of the queue discipline. Figure 3(a) provides boxplots of the average wait times from the 20 replications for each alternative. Note that even though the individual estimates obtained via simulation are each based on one million customers, there is still variation in the actual values. From this figure, the empirical results (indicated by the center of the diamond in each of the means-diamonds plots) agree with theory for both the single fast server and the two slower server scenarios. Having overlapping circles for the all-pairs Tukey-Kramer test indicates there are no statistically significant differences among the means of the four alternatives (p-value$= 0.13$).



(a) Average wait time

(b) Average loss

Figure 3: Performance of four queuing systems, from the customer's perspective.

Contrast those results with Figure 3(b), which shows a means-diamonds plot where the response is the average customer's quadratic loss based on a target value $\tau = 0$ (ideally, customers would like no delay in queue). These results are radically different. The all-pairs Tukey-Kramer test shows that systems A and C (the FIFO queues) are indistinguishable from each other, as are systems B and D (the LIFO queues). The differences between these two pairs of systems are both statistically and practically significant (p-value < 0.0001, with roughly a 20-fold difference associated with the queue discipline). We conclude that the best system in terms of robustness is C, the two-server, FIFO system.

The major finding in this example is that robust analysis leads us to a very different result than what analysts would conclude if they were focused solely on mean performance. The source of the difference can be seen in Figure 4, which shows that with the LIFO system many customers receive very quick service, but others can be "bottled up" for long periods of time. Such results, although well known from renewal theory, are not evident when means are the response being studied. Studying the four alternatives using robust analysis is more likely to yield a decision that customers perceive as equitable.

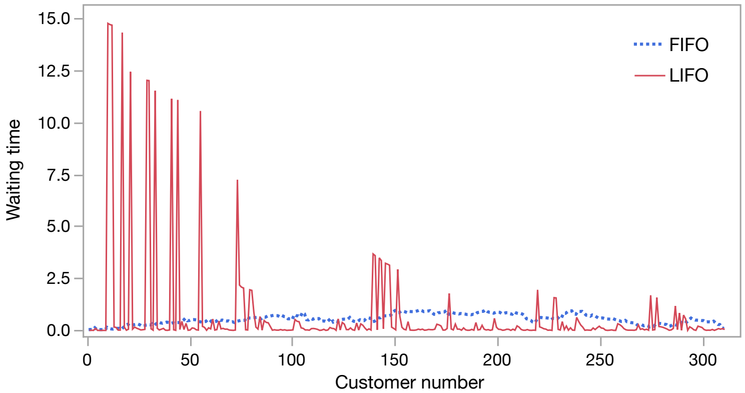


Figure 4: Partial results of a single set of runs with a common random number, under FIFO and LIFO queue disciplines. The queues do not empty out between the 9th and 310th customer to arrive.

3.1.2 NSGS: A Representative Case for Loss-based Selection Procedures

In this section, we illustrate how a robust approach can be incorporated into a well-known selection procedure, and lead to different recommendations. We create two variants of the two-stage procedure of Nelson et al. (2001); we call this the NSGS procedure using the initials of the authors' last names (Nelson, Swann, Goldsman, and Song). We have chosen this procedure because it is relatively simple to use; stopping after the first stage can be viewed as a subset selection technique; and carrying through the second stage leads to an "optimal" choice in terms of selecting the single "best" alternative. We define notation and provide a brief description of the NSGS procedure and our two variants in Figure 5.

We remark that the original procedure has the goal of selecting one or more alternatives with *large* sample means ("the bigger, the better"). However, since low wait times and low losses are both desirable and nonnegative, we simply reverse the signs in order to use the procedure as a minimization technique. The "observations" (i.e., the Y's) will vary depending on which variant of the procedure we use.

NSGS: $Y_{im} = -V_{im}$, and we will seek alternatives associated with high $E[Y_i]$, where m denotes the observation index.

NSGS-R1: $Y_{im} = -\ell(V_{im}, \tau = 0) = -c\left(\overline{V}_{im}^2 + S_{V_{im}}^2\right)$. We seek alternatives associated with high $E[Y_i]$. These are the low loss systems given the target value $\tau = 0$. There is no noise factor *per se*, but each specific $S_{V_{im}}^2$ represents the intrinsic noise in the response for system i.

NSGS-R2: We explicitly specify and control noise factors according to some noise factor design $\mathbf{W}$. The noise factors can be a mix of qualitative and quantitative factors. V_{imd} denotes the m^{th} observation of alternative i using design point d from the noise factor design $\mathbf{W}$, yielding $Y_{im} = \frac{1}{n_w}\sum_{d=1}^{n_w} \ell\left(V_{imd}, \tau\right)$. This loss characterization includes both intrinsic and extrinsic sources of noise.

The first variant, which we call NSGS-R1, simply changes the response measure in order to select an alternative with *lowest intrinsic loss*, where loss reflects the standard deviation as well as the mean of the response of interest. For the second variant (NSGS-R2), we incorporate designed experiments for the noise factors in our problem in order to select an alternative with *lowest overall (intrinsic and extrinsic) loss*. By so doing, we explicitly seek alternatives that are robust to uncontrollable sources of uncertainty

```
NSGS PROCEDURE AND ROBUST VARIANTS NSGS-R1, NSGS-R2

Notation
    $k \geq 2$: the number of alternatives to be evaluated
    $Y_{im}$: the $m$th observation on alternative $i$; this can be:
        NSGS: an aggregate measure from a run or batch,
        NSGS-R1: an aggregate measure from a run or batch, or
        NSGS-R2: an aggregate measure from a set of noise factor runs or batches
    $\alpha_0 > 0$, $\alpha_1 > 0$: if the desired PCS for Stage 1 is $(1-\alpha_0)$, the desired PCS for Stage 2 is $(1-\alpha_0)$,
        then the overall desired PCS is $1-(\alpha_0+\alpha_1)$
    $\delta \geq 0$: the minimum difference in the $\mu$'s that is of practical interest, we must have $\delta > 0$ for Stage 2
    $n_0 \geq 2$: the initial sample size for each alternative
    $\alpha_1 > 0$: desired probability of correct selection for Stage 2
    $t = t_{\beta,n_0-1}$: critical $t$-value with $1-\beta$ in the upper tail, where $\beta = (1-\alpha_0)^{\frac{1}{k-1}}$
    $h = h(1-\alpha_1, n_0, k)$: Rinott's constant (used only in Stage 2)
    $\tau$, $\ell(\cdot,\tau)$: target value and loss function (used in NSGS-R1 and NSGS-R2)
    $\mathbf{W}$: experimental design for the noise factors $W_1,\ldots,W_w$ (used in NSGS-R2)
    $n_w$: number of design points in $\mathbf{W}$ (used in NSGS-R2)

Assumptions
    The $Y_{im}$ are i.i.d. $N(\mu_i, \sigma_i^2)$, and "bigger is better"

Stage 1
    Run the simulations to collect $n_0$ observations from each alternative
        NSGS: Sample $Y_{im}$, $i = 1,\ldots,k$; $m = 1,\ldots,n_0$, where each $Y_{im}$ is the response of interest for that
            simulation run or batch
        NSGS-R1: Sample $Y_{im}$, $i = 1,\ldots,k$; $m = 1,\ldots,n_0$, where each $Y_{im}$ is the negative loss associated with
            the response of interest for that simulation run or batch
        NSGS-R2: Sample $Y_{im}$, $i = 1,\ldots,k$; $m = 1,\ldots,n_0$. Each $Y_{im}$ is the negative average loss over the noise space.
            Let $V_{imd}$ denote the $m$th observation (run or batch) on alternative $i$ using design point $d$ from
            the noise factor design $\mathbf{W}$. Then $Y_{im} = -\frac{1}{n_w}\sum_{d=1}^{n_w} \ell(V_{imd}, \tau)$
    Compute the sample means and variances $\overline{Y}_i$ and $S_i^2$ for $i = 1,\ldots,k$.
    Compute $\Delta_{ij} = t\left(\frac{S_i^2}{n_0} + \frac{S_j^2}{n_0}\right)^{1/2} \quad \forall i,j$
    Set $I = \{i : 1 \leq i \leq k \text{ and } \overline{Y}_i \geq \overline{Y}_j + (\Delta_{ij} - \delta)^+, \forall j \neq i\}$, where $z^+ = \max\{0, z\}$.
        $I$ contains those alternatives whose sample means are not statistically inferior to the best of the rest.
        The size of $I$ is random.
    Return $I$.

Stage 2
    If $I$ contains a single index:
        Return $I$.
    Else:
        Compute $N_i = \left\lceil \max\left\{n_0, \left(\frac{hS_i}{\delta}\right)^2\right\}\right\rceil \qquad \forall i \in I$
        Sample $N_i - n_0$ additional observations from alternative $i$ for all $i \in I$.
        Compute overall sample means $\overline{Y}_i = \frac{1}{N_i}\sum_{m=1}^{N_i} Y_{im}$ for all $i \in I$.
        Return the system $i^* \in I$ associated with the largest $\overline{Y}_i$.
```

Figure 5: Two-stage selection procedures for robust optimization.

in our simulation model, which increases the likelihood that the selected alternative will be suitable for real-world implementation. This can be viewed as a way of mitigating the impact of input uncertainty (Song and Nelson 2019), rather than seeking to quantify input uncertainty and provide a variety of conditional choices. Unlike input uncertainty quantification methods, these robust selection methods can be applied even when data are not available, e.g., to decide which of several prospective facility layouts to adopt.

The NSGS procedure and its variants can be used in two ways. An analyst seeking a subset of good choices may find it convenient to set $\delta = 0$ and consider that a "Correct Selection" has been made if subset I contains the true best systems at the end of Stage 1. See Eckman and Henderson (2018) for more on selecting good solutions. The procedure has a *Probability of Correct Selection* (PCS) guarantee of PCS $\geq 1-\alpha_0$, regardless of the configuration of the true underlying means. The size of the Stage 1 subset I is random, and it will depend on the characteristics of the systems being studied, as well as the initial sample sizes and the desired PCS levels. If there is a clear best alternative then the subset will tend to be small. If there are many viable contenders, then the subset will tend to be large. Subset selection procedures are often viewed as screening procedures to eliminate clearly inferior alternatives where the final decision may be based in part on considerations not included in the simulation model, such as the cost or ease of implementing the chosen alternative. Continuing on to Stage 2 turns the subset selection procedure into an indifference zone procedure that selects a single alternative as "best." Note that the indifference zone value δ must be strictly greater than zero in order to conduct the second stage analysis. Nelson et al. (2001) describe several nice properties of this procedure. For example, with high confidence $(1-\alpha)$ we can simultaneously claim that those systems not in I are less than δ better than the true best; that the best system is chosen if the difference between the top two systems is $\geq \delta$; and that the selected system will be within δ of the actual best system in the subset I.

3.1.3 Numerical Example

We return to the queueing example of Section 3.1.1, but now view it as a selection problem. We use the initial formulation for NSGS, include intrinsic variability using NSGS-R1, and then extend it to consider both intrinsic and extrinsic variability using NSGS-R2 by considering a variety of potential service distributions. The latter makes the systems under consideration M/G/k rather than M/M/k. M/G/k systems are an important class of systems for real-world modeling, but unlike M/M/k systems, they are not amenable to closed-form analytic solution (Gupta et al. 2010). However, they are just as easy to simulate as the M/M/k systems.

It is challenging to investigate distribution shapes without confounding the distributional choices with variations in the mean and variance. To deal with this issue, we have derived parameterizations to match the first two moments while yielding strictly nonnegative values for three classes of distributions with very different shapes: shifted exponentials, triangles, and uniforms. We include unshifted exponential distributions, which have greater variability given the same mean, since this is a commonly-used choice for analytical modeling. The distributional parameter calculations appear below, all based on specifying a common mean service time θ.

$$\begin{aligned}
\text{sexp:}\quad & \theta \cdot \big(1-(1/\sqrt{3})\big) + \text{Exponential}\big(\textbf{rate: } \sqrt{3}/\theta\big) \\
\text{tri:}\quad & \text{Triangle}(\textbf{min: } 0,\ \textbf{mode: } 6\theta/(9+\sqrt{45}),\ \textbf{max: } 3\theta - \text{mode}) \\
\text{unif:}\quad & \text{Uniform}(\textbf{min: } 0,\ \textbf{max: } 2\theta) \\
\text{exp:}\quad & \text{Exponential}(\textbf{rate: } 1/\theta)
\end{aligned}$$

The response of interest is the customer wait time, so we choose a target value $\tau = 0$. Each observational unit m on alternative i is derived from a single (long) simulation run, from which we will gather either a single output measure (V_{im}, defined as the average customer wait time for that run) or two output measures (both V_{im} and either the standard deviation or the variance of customer wait time for that run).

We have $k = 4$ alternatives. Suppose we specify $\alpha_0 = \alpha_1 = 0.05$, and take $n_0 = 5$ first-stage observations from each system. For extrinsic noise factors, we vary the service distributions between the three with matching moments described above, which we view as plausible models for an unknown service distribution. Numerical results for a single test case are provided in Table 1. If $\delta = 0$ (i.e., the analyst is solely interested in a subset selection procedure), the NSGS procedure is unable to eliminate any of the contenders after the first stage, but both robust variants quickly reduce the contenders to A and C (the FIFO queues).

Table 1: Selection procedure examples.

	Procedure		
	NSGS	NSGS-R1	NSGS-R2
Stage 1 $\overline{Y}_i$ (S_i)	A: -0.1613 (0.0109)	-0.0545 (0.0072)	-0.0236 (0.0015)
	B: -0.1615 (0.0109)	-1.0926 (0.1868)	-0.4665 (0.0393)
	C: -0.1572 (0.0109)	-0.0531 (0.0070)	-0.0231 (0.0015)
	D: -0.1574 (0.0111)	-1.0612 (0.1815)	-0.4533 (0.0414)
Subset after Stage 1 when $\delta = 0$	A, B, C, D	A, C	A, C
δ for Stage 2	0.00833[1]	0.00174[2]	0.00174[2]
Subset after Stage 1 for given δ	A, B, C, D	A, C	A, C
Stage 2 observations needed	28, 28, 28, 30	323, 313	11, 10
δ for Stage 2	0.00278[3]	0.00056[4]	0.00056[4]
Subset after Stage 1 for given δ	A, B, C, D	A, C	A, C
Stage 2 observations needed	292, 290, 290, 302	3016, 3012	140, 135

[1] 30s change in wait time. [2] Based on change in wait time from 6m to 6m30s.

[3] 10s change in wait time. [4] Based on change in wait time from 6m to 6m10s.

To go through the second stage, the analyst must specify an indifference value $\delta > 0$. We chose two scenarios: $\delta = 30$s and $\delta = 10$s. Since the mean wait times vary by roughly 15s, a decision maker concerned only with the means would be indifferent among the four systems when $\delta = 30$s but be interested in selecting C when $\delta = 10$s. Specifying an indifference value δ for the robust procedures is not as straightforward. We chose to calculate the difference in loss that would be incurred if a customer waited for 6m30s (or 6m10s) instead of 6m for NSGS-1 and NSGS-2. Consequently, a decision maker making assessments for robustness would be indifferent between A and C at the larger value of δ, but interested in selecting C at the smaller. Table 1 shows the additional observations needed in each case. Each procedure requires roughly an order of magnitude more second-stage sampling for $\delta = 10$s than for $\delta = 30$s. NSGS-R1 requires substantially more second-stage sampling than the other two variants. NSGS-R2 has the lowest second-stage sampling requirements, even when taking into consideration that each NSGS-R2 "observation" is based on three noise runs. Note that second-stage sampling requirements are problem specific. In general, any of the procedures could be associated with the lowest sampling requirement, and any of the three might halt after Stage 1. If we had run the selection procedures only on options $\{$B,C,D$\}$, then the second-stage sampling requirement for NSGS would be 64 (for $\delta = 30$s) or 694 (for $\delta = 10$s), but both NSGS-R1 and NSGS-R2 would select C as best after the first stage. If all options have equal variance, we do no harm by using a robust variant and have the added satisfaction of knowing that our selected alternative is robust.

Now consider the quality of the selections made, supposing that a quadratic loss does, in fact, reflect the customers' actual loss function. System C (two slower FIFO servers) has both the lowest mean and the lowest variance of customer wait time. The loss for system A (one faster FIFO server) is only 3% higher. In contrast, the losses for systems B and D are over 20 times as large as those for the FIFO queues, with B slightly worse than D. This illustrates the fundamental problem of "optimizing" a stochastic system based solely on mean performance. An analyst using NSGS would be as likely to select A as B, and as likely to select C as D, so with probability 0.5 they will incorrectly select a LIFO system as best and incur a large (but unanticipated) penalty after implemention.

3.2 Robust Analysis and Optimization with Quantitative Factors

We have focused so far on robust analysis and optimization for situations where the alternatives are qualitatively different. If one or more of the factors are quantitative, then different approaches may be warranted or more informative. A designed experiment can be formulated to vary the factors simultaneously in a data farming approach, where we systematically "grow" (generate) data from the simulation model by leveraging designed experiments. For more on this, see Sanchez and Sanchez (2017).

3.2.1 Simulation Metamodels

Simulation models attempt to model systems as computer programs to capture behaviors of interest. Analysts are often interested in characterizing those behaviors as functions of the inputs to the simulation. For stochastic simulation models this requires statistical modeling. The statistical model is based on the simulation, which is itself a model, so the result is commonly referred to as a *metamodel*—a model of a model—within the simulation community. Designed experiments are advantageous for identifying and constructing causal metamodels, i.e., statistical models of the input/output relationships. Space-filling designs are particularly useful because of their flexibility for fitting many different types of metamodels, and because of the lack-of-fit diagnostic information they can provide.

Simulation metamodels can take many forms, but popular choices include multiple regression, partition trees and forests, or kriging. The goal is generally to facilitate understanding of the input/output relationships of the system or to provide analytic tractability for selection or optimization techniques. Metamodels can be constructed for experiments involving mixtures of continuous, discrete-valued, and qualitative factors. However, metamodel terms involving quantitative factors reveal the nature of the underlying relationships, rather than simply revealing whether different systems have different responses.

With a quadratic loss function we have two choices: (i) fit metamodels of loss directly, or (ii) fit separate metamodels for the mean and the variability. We often find it convenient to fit the standard deviation as our measure of variability, since it is on the same scale as the mean, but other options such as variance or log(variance) are possible (Myers et al. 1992). One reason for fitting separate metamodels—especially for experiments involving large numbers of decision factors—is to enhance understanding by identifying which factors, interactions, or non-linear terms are the key causal drivers of average performance, and which are the key drivers of the variability. In practice, it is often easier to adjust the mean than to reduce the variability. Concrete examples of how separate mean/variance metamodels can be leveraged will be illustrated during the presentation.

When metamodels are used to seek optimality and quantitative factors are involved, the recommended solutions usually do not correspond to design points that have been directly observed. In such cases the analyst should conduct confirmation runs at the suggested configurations. This will either confirm the suitability of the recommendation, or indicate that there is a lack-of-fit requiring additional exploration.

3.2.2 Numerical Example: FIFO vs. LIFO M/M/k Queues Revisited

Consider again the example in Section 3.1.1. The observations made by qualitative analysis there are verified by fitting a metamodel with two main effects and one interaction effect. The initial (full) metamodel has low explanatory power due to its very high variability, with an $R^2 = 0.071$, but the number of servers is statistically significant with p-value $= 0.0183$. Neither the queue discipline nor the interaction term are significant, both having p-values > 0.9. Dropping the two non-significant terms results does not change the R^2 to three significant digits. The resulting metamodel yields a confidence interval of $[0.157, 0.162]$ for one server, and $[0.153, 0.158]$ for two. Note that both CI's contain their respective analytical means. The results are statistically significant even though a 15 second difference relative to a 9.5 minute delay is not likely to be considered practically significant by customers. Depending on context, for decision makers the cumulative effect of the small difference in delay may be important to the system as a whole.

We also constructed a metamodel with loss as the response. We used a log transformation of the loss to stabilize the nonhomogeneous variance clearly visible in the residuals from a preliminary fit. The resulting metamodel shows queue discipline as the only statistically significant predictor for loss (p-value < 0.0001), and yields an $R^2 = 0.993$. This result is both statistically and practically significant. We conclude that system A and C, the two FIFO systems, completely dominate the LIFO systems in terms of robustness.

3.2.3 Numerical Example: Impact of Service Distribution on M/G/k

Consider a decision maker faced with purchasing new servers for a computing cluster. Modern servers are multi-core machines, and a large part of the decision-making process is whether to purchase slower processors with more cores, or faster processors with fewer cores. The cores are modeled as servers in an M/G/k queueing model, with uncertainty about the service distribution discussed in Section 3.1.3. We also want the decision to be robust across the span of possible job arrival rates (which are not under our control) and cumulative service rates (which are determined by choice of processors). Our goal here is to seek understanding of what is important in the model itself, to determine whether detailed and expensive data collection is justified. Consequently, our second quantitative example expands on the first to include a selection of probability distributions for the service process.

We created a designed experiment to study different service distributions as described above, a range of currently available numbers of cores for the servers, three different sets of job sizes, and LIFO vs FIFO job scheduling options, in order to gain confidence that our results would be consistent across ranges of arrival and service rates. Rather than vary the arrival and service rate independently—which could lead to unstable queues—we formulated the design in terms of traffic intensity and arrival rate, from which we can derive the implicit service rate.

For the continuous factors we used a fully-orthogonal second-order frequency-based design with 4 stacks (Sanchez and Sanchez 2017), which we crossed with the categorical factors and replicated 10 times. The resulting 31,040 simulation runs were completed in about 5 hours on a 16 core laptop.

We used multiple regression to model quadratic loss as the response, after settling on a log transformation for variance stabilizing purposes. We created indicator (0/1) variables for three of the service distribution choices (‘exp’/‘sexp’/‘tri’), implicitly making ‘unif’ the default for comparison purposes. The initial model included quadratic terms for all continuous factors, and two-way interactions for all pairs of factors, and yielded a fit with $R^2 = 0.992$. Given that the dominant factor had a t-ratio with a magnitude in excess of 1300, we chose to eliminate terms with $|t| < 16$ as not being practically significant even if they are statistically significant. (Recall that statistical significance is strongly determined by degrees of freedom, and we had over 31,000 observations.) All terms retained in the model had $|t| > 90$, and the final model lost virtually no explanatory power with $R^2 = 0.990$.

The results are summarized in Figure 6. Traffic intensity and arrival rate (hence service rate) are major drivers of loss measured for delays in queue. However, after taking those into account the queue discipline plays a major role, with LIFO systems having substantially higher loss. As the number of cores increases, quadratic loss decreases by small amounts, presumably because more servers provide a stabilizing effect on variability when the load is distributed. Finally, and significant for our modeling effort, the only distribution that matters is the unshifted exponential—recall that it has the same mean as the other distribution choices, but greater variability. All other distribution indicators dropped out of the model, which indicates that across the service time variations that we explored, the service distribution’s variability seems to have more impact than its shape. These results highlight the distinction between testing the impact of a distribution’s parameters versus its shape. Modelers who use unshifted exponential service distributions may make assessments that are overly optimistic without robust analysis, and overly risk-averse with it because both the mean and variance are fully determined by the rate. Given that the exponential is often implausible due to its unrealistic mode of zero, the implication is that the modeler should use a robust design approach to explore service distributions with a variety of plausible shapes, means, and variances.

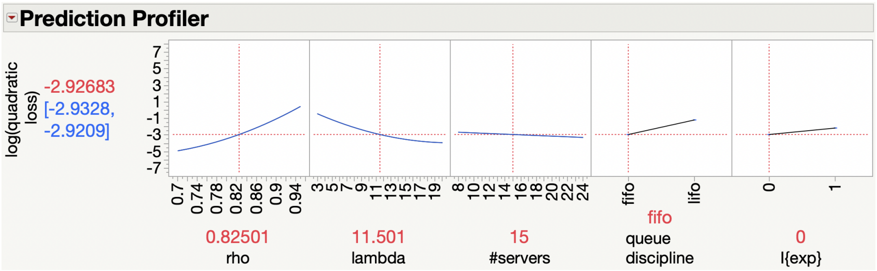


Figure 6: Prediction profiler shows the shape of the response surface relative to each of the factors. This is an accurate portrayal since interaction effects, while present, were minimal in their impact.

4 Concluding Thoughts

This paper is intended to encourage a mutual dialogue between two distinct simulation communities which have been active along seemingly divergent paths: those using and studying ranking and selection or other simulation optimization approaches, and those using and studying designed experiments and response surface metamodeling. A key difference that robust analysis brings is its view of variability as an intrinsic part of a system that must explicitly be assessed during simulation optimization efforts, rather than as a nuisance that can increase the computational effort required to compare alternatives in terms of their mean responses. We have presented examples drawn from selection and metamodel-based optimization, but there is room to incorporate a robust view into other simulation optimization approaches as well.

Robust approaches—specifically, via the use of designed experiments—are also beneficial in the early stages of simulation studies. We assert that applying designed experimentation and robust analysis at the initial model formulation stage may help inform the analysts and interested parties about their actual need (or lack thereof) to incorporate detailed subcomponent modeling and perform expensive data collection. In many cases, there is no question about the structural form of the system being studied, but rather only about the specific parameterizations or choices of distributions. The mantra of keeping a model "as simple as possible, but no simpler" is useful—and robust design can help us determine how simple that should be.

ACKNOWLEDGMENTS

DoD Distribution Statement: Approved for public release; distribution is unlimited. The views expressed in this document are those of the authors and do not necessarily reflect the official policy or position of the DoD or the U.S. Government.

REFERENCES

Box, G. E. P. 1988. "Signal-to-Noise Ratios, Performance Criteria, and Transformations (with Discussions)". *Technometrics* 30(1):1–40.

Dellino, G., J. P. C. Kleijnen, and C. Meloni. 2009. "Robust Simulation Optimization Using Metamodels.". In *Proceedings of the 2009 Winter Simulation Conference*, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 540–550. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Dellino, G., J. P. C. Kleijnen, and C. Meloni. 2010. "Parametric and Distribution-free Bootstrapping in Robust Simulation Optimization". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 1283–1294. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Eckman, D. J., and S. G. Henderson. 2018. "Guarantees on the Probability of Good Selection". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Gupta, V., M. Harchol-Balter, J. G. Dai, and B. Zwart. 2010. "On the Inapproximability of M/G/K: Why Two Moments of Job Size Distribution are Not Enough". *Queueing Systems* 64(1):5–48.

Kim, S.-H. 2013. "Statistical Ranking and Selection". In *Encyclopedia of Operations Research and Management Science*, edited by S. I. Gass and M. C. Fu, 1459–1469. Boston, Massachusetts: Springer US.

Kleijnen, J. P. C. 2017. “Design and Analysis of Simulation Experiments: A Tutorial”. In *Advances in Modeling and Simulation*, edited by A. Tolk, J. Fowler, G. Shao, and E. Yücesan, 135–158. Cham, Switzerland: Springer International Publishing AG.

Kleijnen, J. P. C., S. M. Sanchez, T. W. Lucas, and T. M. Cioppa. 2005. “A User’s Guide to the Brave New World of Designing Simulation Experiments”. *INFORMS Journal on Computing* 17(3): 263–289.

Montgomery, D. C. 2017. *Design and Analysis of Experiments*. 9th ed. Hoboken, New Jersey: Wiley.

Myers, R. H., A. I. Khuri, and G. Vining. 1992. “Response Surface Alternatives to the Taguchi Robust Parameter Design Approach”. *The American Statistician* 46(2):131–139.

Nair, V. N., B. Abraham, J. MacKay, G. Box, R. N. Kacker, T. J. Lorenzen, J. M. Lucas, R. H. Myers, G. G. Vining, J. A. Nelder, M. S. Phadke, J. Sacks, W. J. Welch, A. C. Shoemaker, K. L. Tsui, S. Taguchi, and C. F. J. Wu. 1992. “Taguchi’s Parameter Design: A Panel Discussion”. *Technometrics* 34(2):127–161.

Nelson, B. L., J. Swann, D. Goldsman, and W. Song. 2001. “Simple Procedures for Selecting the Best Simulated System When the Number of Alternatives is Large”. *Operations Research* 49(6):950–963.

Pignatiello Jr., J. J., and J. S. Ramberg. 1991. “Top Ten Triumphs and Tragedies of Genichi Taguchi.”. *Quality Engineering* 4(2):211–225.

Ramberg, J. S., S. M. Sanchez, P. J. Sanchez, and L. J. Hollick. 1991. “Designing Simulation Experiments: Taguchi Methods and Response Surface Metamodels”. In *Proceedings of the 1991 Winter Simulation Conference*, edited by B. L. Nelson, W. D. Kelton, and G. M. Clark, 167–176. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Ryan, T. P. 2007. *Modern Experimental Design*. Hoboken, New Jersey: Wiley.

Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn. 1989. “Design and Analysis of Computer Experiments (includes comments and rejoinder)”. *Statistical Science* 4:409–435.

Sanchez, S. M. 2000. “Robust Design: Seeking the Best of all Possible Worlds”. In *Proceedings of the 2000 Winter Simulation Conference*, edited by J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 69–76. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sanchez, S. M. 2020. “Data Farming: Methods for the Present, Opportunities for the Future”. *ACM Transactions on Modeling and Computer Simulation—Special Issue: Toward an Ecosystem of Models and Data* (Forthcoming).

Sanchez, S. M., and P. J. Sanchez. 2017. “Better Big Data via Data Farming Experiments”. In *Advances in Modeling and Simulation: Seminal Research from 50 Years of Winter Simulation Conferences*, edited by A. Tolk, J. Fowler, G. Shao, and E. Yücesan, 159–179. Cham, Switzerland: Springer International Publishing.

Sanchez, S. M., P. J. Sanchez, and J. S. Ramberg. 1998. “A Simulation Framework for Robust System Design”. In *Concurrent Design of Products, Manufacturing Processes and Systems*, edited by B. Wang, Chapter 12, 279–314. New York: Gordon and Breach.

Schruben, L. W., S. M. Sanchez, P. J. Sanchez, and V. A. Czitrom. 1992. “Variance Reallocation in Taguchi’s Robust Design Framework”. In *Proceedings of the 1992 Winter Simulation Conference*, edited by J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, 548–556. Institute of Electrical and Electronics Engineers, Inc.

Song, E., and B. L. Nelson. 2019. “Input–Output Uncertainty Comparisons for Discrete Optimization via Simulation”. *Operations Research* 67(2):562–576.

Taguchi, G. 1987. *System of Experimental Design*, Volume 1 and 2. White Plains, New York: UNIPUB/Krauss International.

Welch, W. J., T. K. Yu, S. M. Kang, and J. Sacks. 1990. “Computer Experiments for Quality Control by Robust Design”. *Journal of Quality Technology* 22:15–22.

AUTHOR BIOGRAPHIES

SUSAN M. SANCHEZ is a Distinguished Professor in the Operations Research Department at the Naval Postgraduate School, and Co-Director of the Simulation Experiments & Efficient Design (SEED) Center for Data Farming. She also holds a joint appointment in the Graduate School of Defense Management. She has a B.S. in Industrial & Operations Engineering from the University of Michigan, and a Ph.D. in Operations Research from Cornell. She has been an active member of the simulation community for many years, and has been recognized as a Titan of Simulation and an INFORMS Fellow. Her web page is http://faculty.nps.edu/smsanche/ and her email is ssanchez@nps.edu.

PAUL J. SANCHEZ is a Research Associate Professor in the Operations Research Department at the Naval Postgraduate School, and a member of the Simulation Experiments & Efficient Design (SEED) Center for Data Farming. He has an SB in Economics from MIT, and MS and PhD degrees in Operations Research from Cornell University. His research interests include design of experiments, simulation output analysis, and object-oriented modeling. He actually enjoys programming. His web page is http://faculty.nps.edu/pjsanche/ and his email is pjsanche@nps.edu.