Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

SCREENING SIMULATED SYSTEMS FOR OPTIMIZATION

Jinbo Zhao
David J. Eckman

Industrial and Systems Engineering
Texas A&M University
3131 TAMU
College Station, TX 77843, USA

Javier Gatica

Instituto de Ingeniería Matemática y Computacional
Pontificia Universidad Católica de Chile
Avda. Vicuña Mackenna 4860
Santiago, 7820436, CHILE

ABSTRACT

Screening procedures for ranking and selection have received less attention than selection procedures, yet they serve as a cheap and powerful tool for decision making under uncertainty. Research on screening procedures has been less active in recent years, just as the advent of parallel computing has dramatically reshaped how selection procedures are designed and implemented. As a result, screening procedures used in modern practice continue to largely operate offline on fixed data. In this tutorial, we provide an overview of screening procedures with the goal of clarifying the current state of research and laying out opportunities for future development. We discuss several guarantees delivered by screening procedures and their role in different decision-making settings and investigate their impact on screening power and sampling efficiency in numerical experiments. We also study the implementation of screening procedures in parallel computing environments and how they can be combined with selection procedures.

1 INTRODUCTION

Ranking and selection (R&S) refers to a class of problems featuring a finite number of simulated alternatives where a decision-maker wishes to identify one or more alternatives having good expected performance. Each alternative represents a specific scenario of a simulation model, henceforth referred to as a system. In the R&S problem, the expected performance of a system can only be observed with error via simulation. For a recent review of R&S research, see Hong et al. (2021).

In practical R&S problems with large numbers of systems, many systems have poor performance. One would ideally like to eliminate these inferior systems from contention without needing to spend significant computational effort simulating them. This task is commonly referred to as screening or subset selection. In this tutorial, we use the term screening so as to not create confusion with the topic of subset selection arising in statistical learning. Screening has a long history dating back to the seminal work of Gupta (1965), who developed the first screening procedure under the assumptions of normally distributed outputs with known equal variance and a common sample size across systems. In the typical R&S setting, all systems are simulated to some degree, thus devoting some simulation effort to poor systems is unavoidable. Other methods exist that can screen unsimulated systems while delivering similar statistical guarantees to those we study here, e.g., plausible screening methods (Eckman et al. 2022), but to do so they rely on known or assumed structure relating the expected performances of different systems. In contrast, R&S procedures typically treat systems categorically, meaning any relationship between the expected performances of systems and the values of the decision variables describing them is ignored.

Screening procedures involve running replications of the simulation model associated with each system and performing some statistical comparison of the estimated expected performances before finally returning a subset of systems. The returned subset can be used for a variety of purposes. For instance, the decision maker

979-8-3503-6966-3/23/$31.00 ©2023 IEEE

may inspect the surviving systems and ultimately choose one of them based on secondary considerations. This approach might be taken when making an expensive, one-time decision. In this setting, one might alternatively use a selection procedure, which recommends a single system. Another application is to run a screening procedure to reduce the number of systems under consideration before running a more expensive selection procedure in a second stage (Nelson et al. 2001). In other situations, such as when different solutions may be adopted in different scenarios, the decision maker may truly desire to have an assortment of high-quality systems from which to choose.

Some of the earliest screening procedures consisted of a single stage of sampling, where sample sizes are fixed in advance (Gupta 1965). Later procedures were designed to conduct sampling in two stages (Rinott 1978; Dudewicz and Dalal 1975) or fully sequentially (Pei et al. 2022), in which case the total sample size may not be known in advance. See Gupta and Panchapakesan (1985) for an overview of early developments in the field. One way to categorize screening procedures is based on whether systems' sample sizes are fixed up front or determined adaptively. For screening procedures in which the sample size is fixed up front, sampling is typically conducted in a single stage. The ability to control the total sample size in advance is especially appealing to a user with a limited simulation budget. Adaptive screening procedures, on the other hand, guide the sampling process and can thereby avoid excessively sampling clearly inferior systems. For this reason, adaptive screening procedures might be expected to return a smaller subset than those with pre-specified sample sizes if given the same sampling budget. The total sample size of an adaptive screening procedure may or may not be fixed up front, depending on the type of statistical guarantee the procedure seeks to deliver.

The goal of this advanced tutorial is to bring a fresh perspective to an old problem. This tutorial is not intended to be a comprehensive review of screening procedures to date, but rather a collection of ideas—some borrowed, some new—that extend or generalize screening procedures or apply them in new ways. We believe this revisit of screening will be of broad interest to the simulation community, researchers and practitioners alike. The rest of the tutorial is organized as follows: In Section 2, we introduce the problem, present several frequentist guarantees of screening procedures, and discuss their relative merits. In Section 3, we examine three screening procedures and present some numerical experiments that shed light on how well they perform in a variety of situations. Section 4 investigates how screening methods can be run in parallel computing environments and studies the anticipated efficiency gains of several divide-and-conquer schemes. Section 5 explores the sampling efficiency of procedures that combine screening and selection and how their statistical guarantees can be paired to deliver an overall guarantee on the selected system. We conclude in Section 6 and point out some opportunities for future research.

2 SCREENING AND SCREENING GUARANTEES

A system refers to a design or configuration of the simulation model that is being studied. Suppose we have k systems under consideration, and let their collective expected performance be represented by an unknown vector $\mu = (\mu_1, \mu_2, ..., \mu_k)$, i.e., μ_i is the expected performance of System i for $i = 1, 2, \ldots, k$. We assume without loss of generality that $\mu_1 \leq \mu_2 \leq ... \leq \mu_k$ and that larger expected performance is better. Simulating an arbitrary System i produces independent and identically distributed (i.i.d.) outputs $X_{i1}, X_{i2}, ..., X_{in_i}$ where n_i is the number of replications taken at System i. The numbers of replications from each system can differ, and $N = \sum_{i=1}^{k} n_i$ denotes the total sample size. We assume that outputs from different systems are independent, though some R&S procedures can benefit from inducing dependence in the outputs from different systems by using common random numbers (CRN)—doing so typically requires a common sample size across systems. We make the following standard assumption about the outputs:

Assumption 1 For each System i, $i = 1, 2, ..., k$, the outputs $X_{i1}, X_{i2}, ..., X_{in_i}$ from system i are normally distributed with unknown mean μ_i and unknown variance σ_i^2.

In the R&S literature, Assumption 1 is adopted for a variety of reasons: First, it is useful for providing finite-sample statistical guarantees for R&S procedures. Second, many common performance metrics such

as profit, service level and average waiting time can be viewed as sums or averages of many terms and may therefore be approximately normally distributed by the Central Limit Theorem. Lastly, even when the outputs are not approximately normally distributed, batching outputs from multiple replications can produce quantities that are, provided the batch size is large enough.

In this tutorial, we assume that the decision maker regards systems whose expected performances are within $\delta \geq 0$ of the best as "good." The case $\delta = 0$ corresponds to the well-studied correct-selection setting. Let $\mathcal{G} = \{i\colon \mu_i \geq \mu_k - \delta\}$ denote the set of indices of δ-optimal systems, i.e., good systems. The set $\mathcal{G}$ may contain multiple systems, even when $\delta = 0$ if there are multiple systems tied for the best. Because μ is unknown, we cannot identify $\mathcal{G}$ with certainty given only finite sampling. Screening procedures address this challenge by returning a subset $\mathcal{S} \subseteq \{1, 2, ..., k\}$ that is related to $\mathcal{G}$ at some confidence level $1-\alpha$. The subset $\mathcal{S}$ is constructed based on the simulation outputs, often through sufficient statistics like the sample mean $\widehat{\mu}_i = n_i^{-1} \sum_{\ell=1}^{n_i} X_{i\ell}$ and sample variance $\widehat{\sigma}_i^2 = (n_i - 1)^{-1} \sum_{\ell=1}^{n_i} (X_{i\ell} - \widehat{\mu}_i)^2$, for $i = 1, 2, ..., k$.

For screening procedures, the confidence level, $1 - \alpha$, total sample size, N, and returned subset size, $|\mathcal{S}|$, form an impossible triangle. Just as no food can be simultaneously healthy, cheap and tasty, no procedure can simultaneously control $1 - \alpha$, N and $|\mathcal{S}|$; only two of the three can be specified by the user, with the other left free. Enforcing $|\mathcal{S}| = 1$ reduces the problem to that of selecting a single system and shows that for selection procedures there is a tension between the confidence level and the sample size. The stricter the user's demands on the procedure, i.e., higher $1 - \alpha$, lower N, or lower $|\mathcal{S}|$, the more the uncontrolled aspect of the procedure suffers. The screening procedures of Gupta (1965) and Nelson et al. (2001) are examples of those that fix $1 - \alpha$ and N and allow $|\mathcal{S}|$ to vary. We focus on procedures of this type in this tutorial. When $1 - \alpha$ and $|\mathcal{S}| = m > 1$ are fixed, the problem is referred to as restricted subset selection (Koenig and Law 1985; Sullivan and Wilson 1989). Fixed-confidence selection procedures fall into the category of fixing $1 - \alpha$ and $|\mathcal{S}| = 1$ and take as many replications as necessary to deliver a guarantee at the specified confidence level (Kim and Nelson 2001). Although we are not aware of any screening procedures that fix N and $|\mathcal{S}| = m > 1$ but allow $1 - \alpha$ to vary, procedures that do so can be easily developed from the Bayesian perspective (Eckman et al. 2020). On the other hand, for fixed N and $|\mathcal{S}| = 1$, the fixed-budget knockout-tournament (FBKT) procedure of Hong et al. (2022) provides a lower bound on the probability of selecting the best system.

2.1 Fixed-Confidence Frequentist Guarantees

In the frequentist treatment of the R&S problem, the joint probability distribution of the outputs of the systems is considered fixed, but unknown, and a procedure's statistical guarantee is with respect to repetitions of the sampling experiment on the same problem instance. Consequently, the vector of expected performances, μ, and the set of good systems, $\mathcal{G}$, are fixed. In contrast, the returned subset, $\mathcal{S}$, is random.

We study four frequentist guarantees, all of which feature a confidence level $1 - \alpha$ specified by the decision maker. While the names we give to these guarantees may be new, the guarantees themselves are not. Figure 1 helps to visualize how the guarantees differ.

Definition 1 A screening procedure delivers the *Set-wise Probability of Good Selection (Set-wise PGS) guarantee* if for any $\mu \in \mathbb{R}^k$, $\mathbb{P}(\mathcal{G} \subseteq \mathcal{S}) \geq 1 - \alpha$.

Definition 2 A screening procedure delivers the *System-wise Probability of Good Selection (System-wise PGS) guarantee* if for any $\mu \in \mathbb{R}^k$, $\mathbb{P}(i \in \mathcal{S}) \geq 1 - \alpha$ for all $i \in \mathcal{G}$.

Definition 3 A screening procedure delivers the *Expected False Elimination Rate (EFER) guarantee* if for any $\mu \in \mathbb{R}^k$, $\mathbb{E}[|\mathcal{S} \cap \mathcal{G}|]/|\mathcal{G}| \geq 1 - \alpha$.

Definition 4 A screening procedure delivers the *Probability of Good Inclusion (PGI) guarantee* if for any $\mu \in \mathbb{R}^k$, $\mathbb{P}(|\mathcal{S} \cap \mathcal{G}| \geq 1) \geq 1 - \alpha$.

The four guarantees form a chain of implications, with the set-wise PGS guarantee being the strongest and the PGI guarantee being the weakest, as described in Proposition 1. When there is only one good system, i.e., $|\mathcal{G}| = 1$, all four guarantees are equivalent.

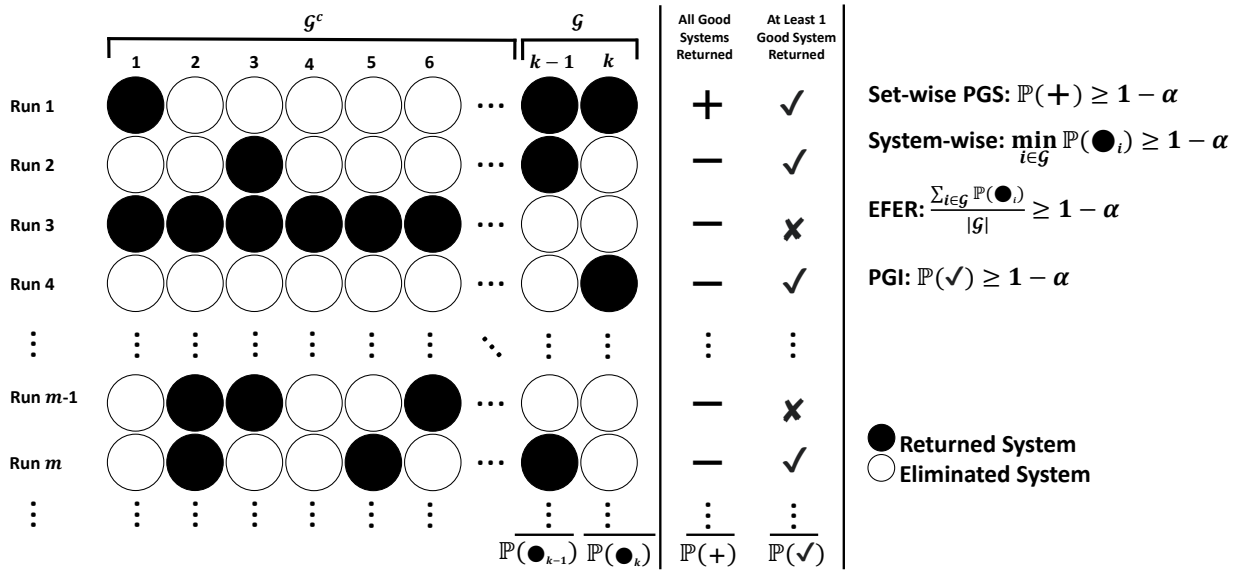


Figure 1: Illustration of four fixed-confidence frequentist guarantees for screening on a problem instance with k systems, two of which are good. $\mathbb{P}(\bullet_i)$ denotes the probability that System i is returned.

Proposition 1 For any fixed confidence level $1-\alpha \in (0,1)$, if a screening procedure delivers the set-wise PGS guarantee, then it also delivers the system-wise PGS guarantee; if a screening procedure delivers the system-wise PGS guarantee, then it also delivers the EFER guarantee; if a screening procedure delivers the EFER guarantee, then it also delivers the PGI guarantee.

The four guarantees serve different purposes and a decision maker's preferences regarding them may depend on the situation. The set-wise PGS guarantee states that with high probability *all* good systems are returned, which may be desirable when screening is used for population selection within an evolutionary algorithm (Lam 1986; Hedlund and Mollaghasemi 2001). The system-wise PGS guarantee is more appropriate if the decision maker wants to ensure that each good system has a good chance of being returned. The EFER guarantee is inspired by a homonymous guarantee in Pei et al. (2022) that controls the proportion of good systems that are incorrectly screened out. Here, we express the guarantee in terms of a complementary inequality, so as to make it more closely resemble the others. The EFER guarantee ensures that a large proportion of the good systems are returned on average and may be satisfying to a more tolerant decision maker. If the decision maker were to combine a screening procedure and a selection procedure with the goal of selecting a good system, then the PGI guarantee may be sufficient because it asserts that, with high probability, at least one good system survives the first stage (Trawinski 1969).

2.2 Other Guarantees

The aforementioned guarantees aim to control the Type I error rate—the probability that good systems are accidentally screened out. Few screening procedures deliver guarantees controlling the Type II error rate—the probability that bad systems survive screening. The ability to control the prevalence of bad systems in the returned subset may be appealing to a risk-averse decision maker. An example of a guarantee that controls the Type II error rate is to ensure that with high probability, only good systems are returned, i.e., for any $\mu \in \mathbb{R}^k$, $\mathbb{P}(\mathcal{S} \subseteq \mathcal{G}) \geq 1-\alpha$ (Desu 1970). Another, weaker guarantee is that for any $\mu \in \mathbb{R}^k$, $\mathbb{P}(i \notin \mathcal{S}) \geq 1-\alpha$ for all $i \notin \mathcal{G}$, meaning that for each bad system, the probability it will be eliminated is sufficiently high. For guarantees controlling the Type II error rate, a serious concern is that with nontrivial probability the returned subset may be empty. Another guarantee for screening procedures is to control the average optimality gap of the returned systems, referred to as the expected opportunity cost (Gao and

Chen 2015a). When the decision maker is not demanding optimality, other definitions of *good* systems include being a top-m system (Gao and Chen 2015b) or having an expected performance greater than $\min\{\mu_k, \mu^\dagger\}$, where $\mu^\dagger$ is standard specified by the decision maker (Pei et al. 2022).

3 EXAMPLES OF SCREENING PROCEDURES

In this section, we describe several screening procedures that deliver the set-wise PGS, system-wise PGS, and EFER guarantees. As for the PGI guarantee, it can be delivered by prematurely terminating any sequential selection procedure that delivers the PGS guarantee, i.e., one that selects a δ-optimal system with high probability. One such example is the Envelope procedure of Ma and Henderson (2017). Because the probability of eliminating all good systems is guaranteed to be below α at any time during the run, the probability of at least one good system remaining in the surviving subset is above $1-\alpha$. For the $\delta = 0$ case, indifference-zone-free procedures with a probability of correct selection (PCS) guarantee can also be terminated prematurely (Fan et al. 2016; Wang et al. 2023). This alternative use of sequential selection procedures offers flexibility in terms of the stopping criteria. One could choose to terminate such procedures upon exhausting a specified budget of wall-clock time, CPU time, or total sample size, or when the subset of systems still in contention shrinks to some desired size. This observation motivates the need for further study of the *rate* at which procedures eliminate systems and how that rate changes over time.

3.1 Screen-to-the-Best

We start with the Screen-to-the-Best (STTB) procedure of Nelson et al. (2001), which is implemented in commercial simulation software, including Simio and Arena. STTB performs pairwise comparisons, taking turns comparing each system against all the others and deciding whether it should be eliminated, i.e., screened out. Systems that are not eliminated comprise the returned subset. We present a modified version of STTB that returns the subset

$$\mathcal{S}^{\text{STTB}} = \left\{ i \colon \widehat{\mu}_i + \delta \geq \widehat{\mu}_j - t_{\beta,\nu}\sqrt{\frac{\widehat{\sigma}_i^2}{n} + \frac{\widehat{\sigma}_j^2}{n}} \text{ for all } j \neq i \right\}, \tag{1}$$

where $t_{\beta,\nu}$ is the β quantile of the Student's t-distribution with ν degrees of freedom. For $\beta = (1-\alpha)^{1/(k-1)}$ and $\nu = n-1$, STTB delivers the system-wise PGS guarantee. Algorithm 1 shows pseudocode for how to construct the subset $\mathcal{S}^{\text{STTB}}$ given the output data. We present STTB with a common sample size, n, for ease of exposition; the procedure has been extended to handle unequal sample sizes (Boesel et al. 2003).

Algorithm 1: Screen-to-the-Best (STTB)

```
S = {1, 2, ..., k}
for i = 1, 2, ..., k do
    for j = 1, 2, ..., k do
        if  μ̂_i + δ < μ̂_j − t_{β,ν} sqrt(σ̂_i^2/n + σ̂_j^2/n)  then
            S = S \ {i} and break
return S
```

In the original version of STTB, the δ term in (1) appeared on the right-hand side of the inequality, paired with the $t_{\beta,\nu}\sqrt{\widehat{\sigma}_i^2/n + \widehat{\sigma}_j^2/n}$ term with a positive-part operator applied. We propose this modified version because it preserves a certain transitive-elimination property when $\delta > 0$. The original procedure possesses this property when $\delta = 0$ and is purported to when $\delta > 0$, but there are known counter-examples. Transitive elimination refers to the property that for any Systems i, j, and ℓ, if System i eliminates System

j, and System j eliminates System ℓ, then System i eliminates System ℓ. This property implies that System i would eliminate all systems that System j would eliminate, which allows a screening procedure to be implemented in parallel using a divide-and-conquer scheme, as will be discussed further in Section 4.

3.2 Decoupled Screen-to-the-Best

We next introduce a new variation of STTB called Decoupled STTB (DSTTB) that returns the subset

$$\mathcal{S}^{\mathrm{DSTTB}} = \left\{ i \colon \widehat{\mu}_i + t_{\beta,\nu}\sqrt{\frac{\widehat{\sigma}_i^2}{n}} + \delta \geq \widehat{\mu}_j - t_{\beta,\nu}\sqrt{\frac{\widehat{\sigma}_j^2}{n}} \text{ for all } j \neq i \right\}. \tag{2}$$

DSTTB delivers the system-wise PGS guarantee when $\beta = (1-\alpha)^{1/k}$ and $\nu = n-1$ and delivers the set-wise PGS guarantee when $\beta = (1+(1-\alpha)^{1/k})/2$ and $\nu = n-1$. We again choose to show the case of common sample sizes, but an extension to unequal sample sizes is straightforward. DSTTB's decoupling of the square-root term in (1) results in less screening power compared to STTB, as implied by Proposition 2.

Proposition 2 For any $\mu \in \mathbb{R}^k$, $\alpha \in (0,1)$ and $\delta \geq 0$, when DSSTB is set up to deliver the system-wise PGS guarantee, $\mathbb{P}(\mathcal{S}^{\mathrm{STTB}} \subseteq \mathcal{S}^{\mathrm{DSTTB}}) = 1$.

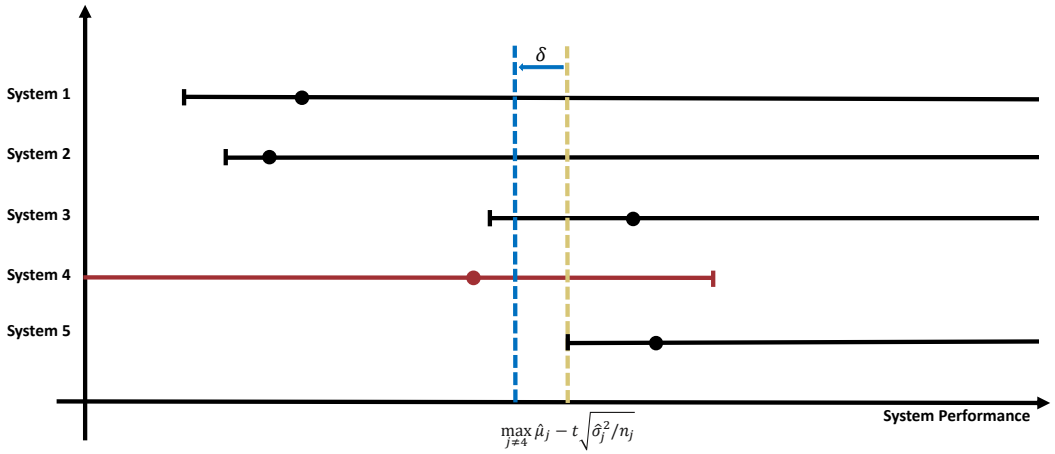
DSTTB constructs either a one-sided or two-sided $(1-\alpha)^{1/k}$ confidence interval for each system. When systems are simulated independently, the probability that all k systems' confidence intervals contain their expected performances is then $1-\alpha$. When screening a given system, we optimistically estimate its expected performance with an upper confidence bound and pessimistically estimate the expected performances of all other systems with their lower confidence bounds. As illustrated in Figure 2, a system is returned if its upper confidence bound plus δ is greater than the maximum lower confidence bound of the other systems. The set-wise PGS guarantee follows from using the same fixed two-sided confidence intervals for screening all systems, and the system-wise PGS guarantee follows from using one-sided confidence bounds whose orientations depend on whichever system is being screened. Algorithm 2 shows how to efficiently construct $\mathcal{S}^{\mathrm{DSTTB}}$ for either guarantee by computing a benchmark T for making comparisons.

Algorithm 2: Decoupled Screen-to-the-Best (DSTTB)

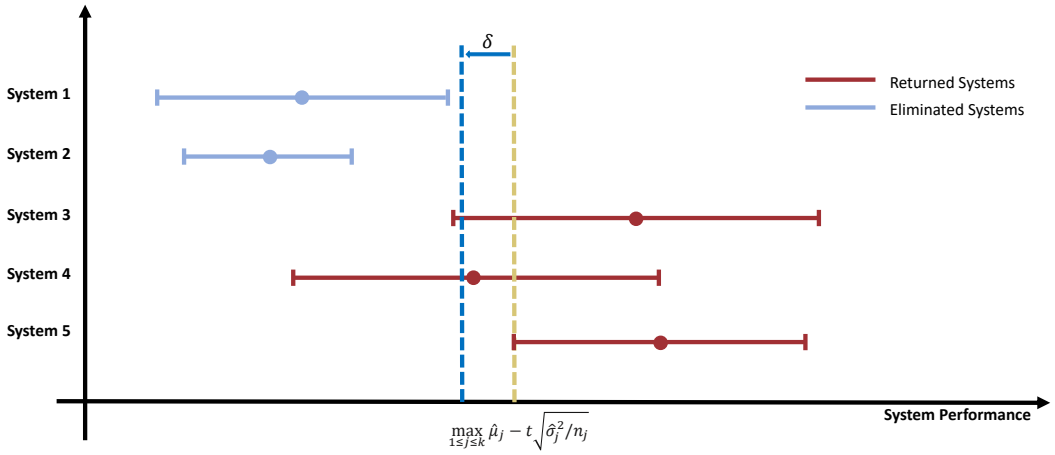
$\mathcal{S} = \{1, 2, \ldots, k\}$
for $i = 1, 2, \ldots, k$ **do**
  $l_i = \widehat{\mu}_i - t_{\beta,\nu}\sqrt{\frac{\widehat{\sigma}_i^2}{n}}$
  $u_i = \widehat{\mu}_i + t_{\beta,\nu}\sqrt{\frac{\widehat{\sigma}_i^2}{n}}$
$T = \max_{i\in\{1,2,\ldots,k\}} l_i$
for $i = 1, 2, \ldots, k$ **do**
  if $u_i + \delta < T$ **then**
    $\mathcal{S} = \mathcal{S} \setminus \{i\}$
return $\mathcal{S}$

3.3 Bi-PASS

The last procedure we exhibit is the bisection Parallel Adaptive Survivor Selection (bi-PASS) of Pei et al. (2022). Unlike STTB and DSTTB, which fix the sample sizes in advance, the bi-PASS procedure takes replications sequentially from systems still under consideration and continuously eliminates systems. Systems are eliminated if their estimated expected performance falls too far below an adaptive standard,



(a) System-wise PGS guarantee (Screening System 4)



(b) Set-wise PGS guarantee

Figure 2: The confidence intervals used by DSTTB when delivering (a) the system-wise PGS guarantee or (b) the set-wise PGS guarantee on a problem instance with five systems.

which is adopted to overcome the lack of knowledge of the desired standard by which systems are deemed good. Algorithm 3 shows how the bi-PASS procedure updates the adaptive standard $\widehat{\mu}^*$ and uses it to screen out inferior systems for the setting where $\delta = 0$. The procedure uses a boundary function $g(\cdot)$ to control how far a system's estimated expected performance is allowed to deviate from the adaptive standard.

The adaptive standard is designed to converge to the desired standard as the sample size increases. For example, when the adaptive standard is the average of the sample means of the surviving systems, it will converge to μ_k, the expected performance of the best system, thus delivering the EFER guarantee with $\delta = 0$. To the best of our knowledge, bi-PASS is the only screening procedure designed to deliver the EFER guarantee for $\delta = 0$, specifically when run with a fixed total sample size. It might be possible to modify the bi-PASS procedure to handle the setting where $\delta > 0$ by shifting $\widehat{\mu}^*$ down by δ.

3.4 Numerical Comparisons

We next compare the empirical performance of the STTB, DSTTB, and bi-PASS procedures. We devise problem instances having expected performances described by $\mu_i = F^{-1}(i/k)$ for $i = 1, 2, \ldots, k$, where F^{-1} is the inverse cumulative distribution function (cdf) of a certain Beta distribution. To investigate how

Algorithm 3: Bisection Parallel Adaptive Survivor Selection (bi-PASS)

$\mathcal{S} = \{1, 2, \ldots, k\}$
for $i \in \mathcal{S}$ **do**
  $n_i = n_0$ and simulate each system n_0 times
Compute the initial estimated standard $\widehat{\mu}^* = |\mathcal{S}|^{-1} \sum_{i \in \mathcal{S}} \widehat{\mu}_i$ and total sample size $N = \sum_{i \in \mathcal{S}} n_i$
while $|\mathcal{S}| > 1$ and $N < N_{max}$ **do**
  for $i \in \mathcal{S}$ **do**
    if $(\widehat{\mu}_i - \widehat{\mu}^*) n_i / \widehat{\sigma}_i^2 \leq -g(n_i/\widehat{\sigma}_i^2)$ **then**
      $\mathcal{S} = \mathcal{S} \setminus \{i\}$ and $\widehat{\mu}^* = |\mathcal{S}|^{-1} \sum_{i \in \mathcal{S}} \widehat{\mu}_i$
    else
      Simulate System i and update n_i, $\widehat{\mu}_i$, $\widehat{\sigma}_i^2$,N, and $\widehat{\mu}^*$
return $\mathcal{S}$

conservative DSTTB is relative to STTB, we consider a problem instance with $k = 1000$ systems and we use the Beta($a = 4, b = 2$) distribution for which there are many near-optimal systems. STTB and two versions of DSTTB are run with the β and ν values set to deliver PGS guarantees (where $\delta = 0$) at a confidence level $1 - \alpha = 0.95$. We assume a common, unknown variance of $\sigma^2 = 0.5$ and allow the procedures to use pooled-variance estimators; this simplification is not believed to affect the procedures' relative performance. Figure 3 shows the subset size based on 40 macroreplications given different sampling budgets. The two versions of DSTTB have similar screening power, but STTB eliminates significantly more systems for all total sample sizes. The same relationships were observed for another problem instance with many inferior systems, but the differences in subset sizes were smaller.

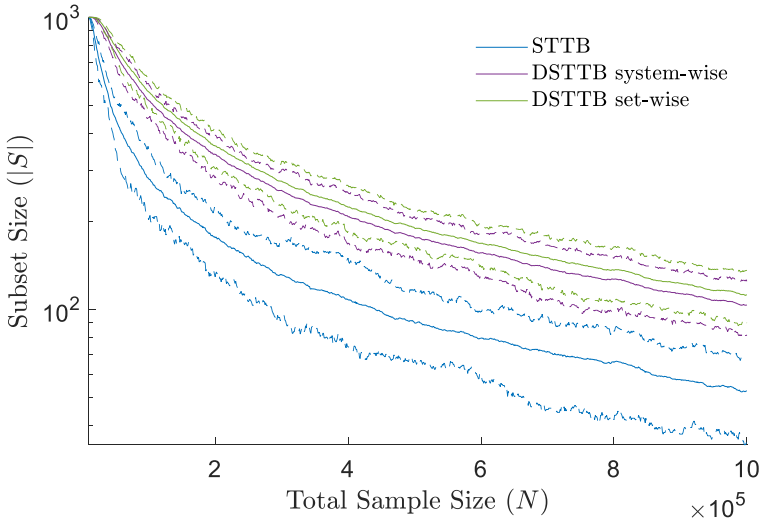
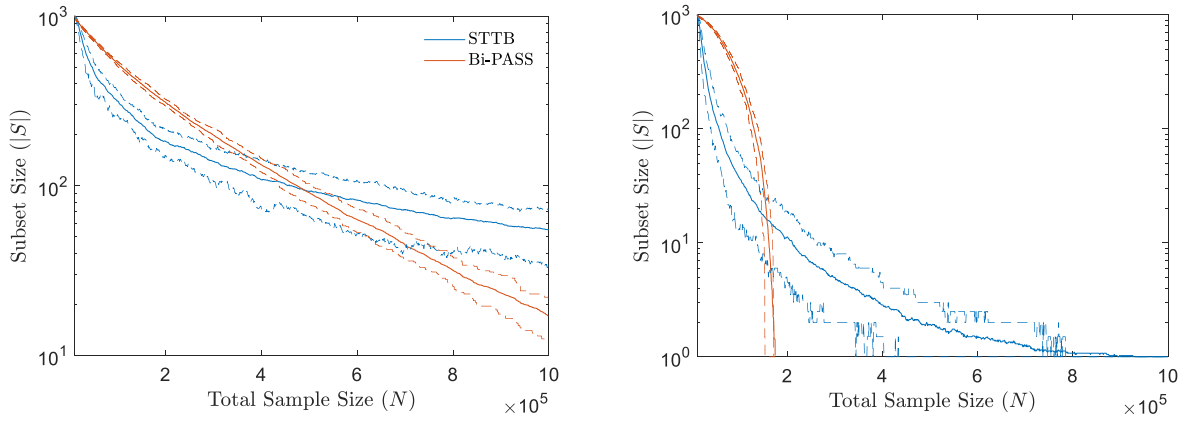


Figure 3: Mean and 10th/90th percentiles (dashed) of returned subset size for STTB and DSTTB on a problem instance with 1000 systems with expected performances from the Beta($a = 4, b = 2$) distribution.

We next compare STTB and bi-PASS. To make for a fair comparison, we test the two procedures in a correct-selection setting where $\delta = 0$ and there is a unique optimal system. In this setting, the system-wise PCS guarantee delivered by STTB and the EFER guarantee delivered by bi-PASS are equivalent.

In the first experiment, we compare how many systems the two procedures screen out when run on the same problem instance for a range of sampling budgets. We test two problem instances, each with $k = 1000$ systems. In the first problem instance, we again use the Beta($a = 4, b = 2$) distribution. In the second



(a) Beta($a = 4, b = 2$) (b) Beta($a = 2, b = 4$)

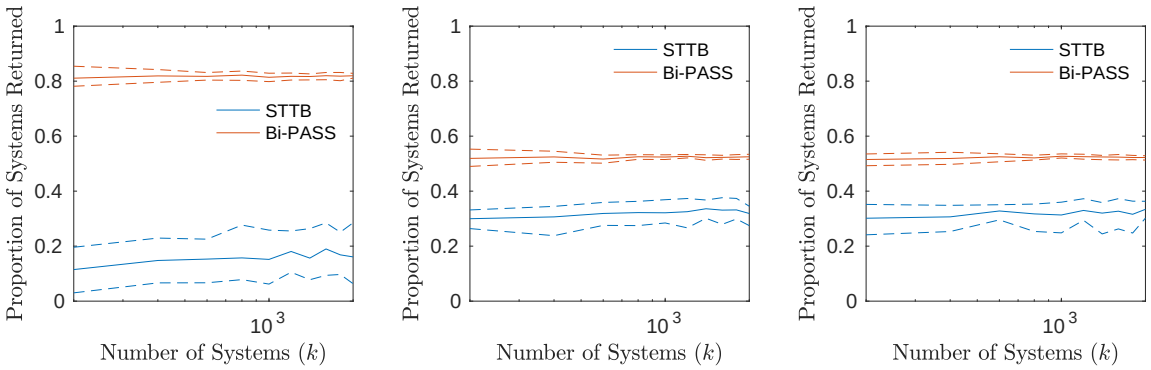
Figure 4: Mean and 10th/90th percentiles (dashed) of subset size for STTB and bi-PASS on two fixed problem instances when varying the total sample size.

problem instance, we use the Beta($a = 2, b = 4$) distribution for which most systems have poor expected performance. Both procedures are run with a confidence level $1-\alpha = 0.95$. For bi-PASS, we choose to use the boundary function $g(n_i/\widehat{\sigma}^2) = \sqrt{(5 + \log(n_i/\widehat{\sigma}^2 + 1))(n_i/\widehat{\sigma}^2 + 1)}$ to achieve the specified confidence. The bi-PASS procedure initially takes $n_0 = 10$ replications of all systems and is run with a total budget of $N = 1{,}000{,}000$ replications; its set of surviving systems is tracked over time. In contrast, STTB is run with different budgets and allocates replications equally across systems, obtaining N/k replications from each system. We perform 40 macroreplications of each procedure, each time recording the returned subset. Figures 4(a) and 4(b) show the size of the subsets returned by each procedure. In both problem instances, STTB screens out more systems than bi-PASS for small total sample sizes, but bi-PASS eventually becomes more powerful than STTB, presumably due to its ability to sample adaptively. For the problem instance with many near-optimal systems, the size of the subset returned by bi-PASS shrinks at an almost linear rate, while for the problem instance with many inferior systems, the rate is superlinear. The percentile bands in the two plots show that the size of the subset returned by STTB is more variable.

In a second experiment, we analyze the limiting performance of STTB and bi-PASS as the number of systems increases. We consider the two problem instances from the previous experiment as well as the Beta($a = 1, b = 1$) distribution, i.e., the Uniform(0, 1) distribution. We test problems with $k = 200$ to $k = 2000$ systems and let the total sample size scale linearly with the number of systems, setting $N = 40k$. As before, we perform 40 macroreplications of each procedure, with the same parameters as before, and record the returned subsets. Figure 5 shows the proportion of returned systems for the two procedures in the three problem instances. In all three problem settings, we see that STTB screens out a greater proportion of systems than bi-PASS. This observation is likely specific to the per-system sample size of 40; for larger total sample sizes, Figures 4(a) and 4(b) suggest the ordering will be reversed. The most interesting finding from Figure 5 is that the proportion of systems returned by both procedures is largely independent of the number of systems. This phenomenon might be explained by how the problem instances involve "filling in" a fixed distribution for the expected performances as the number of systems increases.

4 SCREENING IN PARALLEL COMPUTING ENVIRONMENTS

Modern R&S procedures are used to solve simulation-optimization problems featuring large numbers of systems, as can arise when systems represent combinations of integer-ordered or categorical decision variables. The ability of R&S methods to tackle such large-scale problems has increased significantly in recent years, due to the development of procedures that exploit parallel computing. Problems that were once



(a) Beta($a = 2, b = 4$) (b) Uniform$(0, 1)$ (c) Beta($a = 4, b = 2$)

Figure 5: Mean and 10th/90th percentiles (dashed) of the proportion of systems returned by STTB and bi-PASS when run with a total sample size of $N = 40k$ on three problem instances.

deemed impossible to solve now show promise. In this setting, procedures that obtain samples adaptively are of special interest, as they can identify and eliminate clearly inferior systems over time and thereby save considerable sampling effort. Even so, preserving the statistical validity of the guarantees of R&S procedures that compare systems simulated on different processors remains a formidable challenge.

Recent research on R&S has concerned either how to parallelize existing selection procedures or how to develop new schemes for comparing systems in parallel environments (Hunter and Nelson 2017). A computational issue that arises in parallel environments is that under the common master-worker framework, the master processor can be a bottleneck if overwhelmed by communication tasks (Luo et al. 2015). This can be overcome by employing a divide-and-conquer scheme in which certain tasks are performed on worker processors before being performed again on the master processor (Ni et al. 2017). When the distributed task is screening, systems are first simulated and screened on the worker processors, and then the results (the identities and outputs of the returned systems) are sent to the master processor for another round of screening, without additional replications being taken. Under this setup, fewer systems' outputs are reported to the master processor, reducing the amount of coordination and the overall wall-clock time.

When designing a R&S procedure to identify optimal or near-optimal systems, it is natural to compare every system to every other system. However, pairwise comparisons require immense coordination across processors. To reduce this need for synchronization, bi-PASS instead compares individual systems to an adaptive standard maintained on the master processor and updated based on the results returned by the worker processors (Pei et al. 2022).

Relatedly, Zhong et al. (2022) observed that in the elimination check within the KN procedure (Kim and Nelson 2001) and Paulson's procedure (Paulson 1964)—when CRN are not used—one can decouple a certain term $\widehat{\sigma}_{ij}^2$ into $\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2$ and thereby avoid pairwise comparisons. This reduces the time complexity of the two procedures from $O(k^2 \log k)$ and $O(k^2)$ to $O(k \log k)$ and $O(k)$, respectively. The DSTTB procedure is inspired by this observation and applies a similar decoupling to the pairwise comparisons performed in STTB. Whereas the time complexity of constructing $\mathcal{S}^{\mathrm{STTB}}$ is $O(k^2)$, that of $\mathcal{S}^{\mathrm{DSTTB}}$ is $O(k)$, regardless of the guarantee being delivered, as can be seen by the two (unnested) `for` loops in Algorithm 2. Taken together, this time complexity comparison and the nested relationship established in Proposition 2 suggest that DSTTB might be useful for pre-screening systems before constructing $\mathcal{S}^{\mathrm{STTB}}$ from the surviving systems. Moreover, both STTB and DSTTB have the aforementioned transitive-elimination property, which implies that they can be parallelized using a divide-and-conquer approach without any loss in screening power. Thus, any combination of DSTTB and STTB concluding with STTB—whether performed on a single processor or on multiple processors via a divide and conquer scheme—will ultimately return $\mathcal{S}^{\mathrm{STTB}}$.

We investigate the potential wall-clock-time savings of pre-screening with DSTTB in various divide-and-conquer schemes. We run 20 macroreplications of each procedure on the Beta($a = 2, b = 4$) problem instance in which most systems are inferior and there are $k = 100{,}000$ systems. We set $n = 20$ in all cases and record the wall-clock time spent constructing $\mathcal{S}^{\text{STTB}}$; the time spent running the simulation replications is not counted. Table 1 reports the wall-clock times of the schemes assuming different numbers of processors. The experiments are actually conducted on a single processor, so the communication costs present in a real parallel computing environment are not reflected in the reported times. In Table 1, we see that in a single-processor setting, using DSTTB to pre-screen can result in a two-fold speed-up for this large-scale problem. The divide-and-conquer strategy also has a tremendous effect on speeding up screening, with the divide-and-conquer version of STTB, without using DSTTB, being the fastest. Using DSTTB to pre-screen on both the worker processor and the master processor is the next fastest.

Table 1: Mean and 10th/90th percentiles of wall-clock time (in seconds) spent screening by combinations of the STTB and DSTTB procedures given different numbers of processors (p) on a problem with $k = 100{,}000$ systems. DSTTB + STTB indicates that STTB is used to screen the systems returned by DSTTB.

On Workers	On Master	$p = 10$	$p = 20$	$p = 50$	$p = 100$
–	STTB	38.32 [35.71, 39.96]			
–	DSTTB + STTB	19.62 [18.07, 21.21]			
STTB	STTB	2.55 [2.16, 2.86]	2.64 [2.57, 2.74]	2.65 [2.56, 2.73]	2.84 [2.73, 3.02]
DSTTB	STTB	3.97 [3.81, 4.19]	3.99 [3.84, 4.11]	4.14 [4.01, 4.25]	4.20 [4.12, 4.30]
DSTTB	DSTTB + STTB	5.02 [4.44, 5.55]	4.75 [4.09, 5.28]	4.97 [4.74, 5.35]	4.76 [4.39, 5.22]
DSTTB + STTB	DSTTB + STTB	3.26 [3.07, 3.44]	3.14 [2.94, 3.29]	3.24 [3.14, 3.35]	3.50 [3.31, 3.65]

5 PAIRING SCREENING AND SELECTION PROCEDURES

Screening procedures can prove useful even when a decision maker's objective is to ultimately select a single system. A screening procedure can simulate all systems and winnow out non-competitive systems in a first stage, before a selection procedure simulates the surviving systems and identifies a single, most promising system. The screening procedure is intended to cheaply shrink the set of contending systems prior to running a more sampling-intensive selection procedure, thereby reducing the total sample size required to deliver a fixed-confidence guarantee. An example of this two-stage approach is the NSGS procedure of Nelson et al. (2001) which pairs STTB with the selection procedure of Rinott (1978).

Selection procedures offer frequentist guarantees similar to those of screening procedures. For example, some selection procedures guarantee to select a δ-optimal system with probability exceeding $1-\alpha$, referred to generically as the PGS guarantee. This guarantee can be thought of as the PGI guarantee with the restriction that $|\mathcal{S}| = 1$. The PGS guarantee for selection procedures pairs well with the PGI guarantee for screening procedures. In particular, if a first-stage screening procedure delivers the PGI guarantee for some $\delta_1 > 0$ and a second-stage selection procedure delivers the PGS guarantee for some $\delta_2 > 0$, then, with high probability, the selected system has an expected performance within $\delta = \delta_1 + \delta_2$ of the optimal, μ_k. As we will soon see, the confidence level associated with the assertion that a δ-optimal system is selected depends on whether the selection procedure reuses the outputs collected during the screening stage.

Reusing data from the screening stage would seem to improve the efficiency of a combined procedure by giving the selection procedure a head start with its sampling. However, the reused data, e.g., the sample means, are not independent of the identities of the surviving systems. That is, systems that survive screening are more likely to have estimated performances that are better than their true expected performances. Nelson et al. (2001) show that for the NSGS procedure, first-stage data can be reused if the allowable error, α, is split across the two stages. The combined procedure entails running STTB with $\delta_1 = 0$ and $1 - \alpha_1$, followed by Rinott's procedure with $\delta_2 = \delta$ and $1 - \alpha_2$. The system ultimately selected is guaranteed to be δ-optimal with probability exceeding $1 - \alpha_1 - \alpha_2$. To establish this guarantee, the constant used by Rinott's procedure must be based on the original number of systems, k, and not the number of surviving systems. This results in a larger value of the constant and thus larger sample sizes taken in the second stage, undercutting the efficiency gains anticipated from reusing the first-stage data. For certain problem instances, this trade-off may be worthwhile. Nevertheless, the proof of this additive decomposition lemma relies on a sample-path-wise coupling argument that is too restrictive to be expected to hold for selection procedures that sample systems adaptively. Wilson (2001) shows that for the NSGS procedure, a slightly less conservative, multiplicative decomposition is possible, in which the overall confidence is $(1-\alpha_1)(1-\alpha_2)$. The basis for this result is that the event that the best system survives screening and the event that the best system is ultimately selected when it survives to the second stage are positively correlated. This relationship makes intuitive sense and, if it could be established for other combinations of screening and selection procedures, may present more opportunities for reusing first-stage data without issue.

If the first-stage data are not reused, the second-stage selection procedure can be run as if it were seeing a problem instance consisting of only the returned systems. Assuming the screening and selection procedures are run independently, the resulting confidence level is $(1 - \alpha_1)(1 - \alpha_2)$. There may be some loss in efficiency from throwing away the first-stage data, as the second-stage procedure must now obtain fresh replications from each surviving system. Still, if the first-stage screening procedure greatly reduces the number of systems under consideration, this additional sampling may be negligible compared to the larger sample sizes that a second-stage selection procedure (with reuse) would have otherwise required.

The approach of throwing away the first-stage data can be applied to any combination of screening and selection procedures. We experiment with a simple procedure that combines STTB (with $\delta_1 = 0$) with Rinott's procedure (with $\delta_2 = \delta$) without reusing first-stage data. Overall, the procedure selects a δ-optimal system with high probability. For this combined procedure, the decision maker must choose the initial per-system sample size, n_0, of STTB. Increasing n_0 increases the number of samples obtained in the screening stage, but by more precisely estimating the expected performances of the systems, the screening procedure can return a smaller subset of systems, thereby reducing the number of replications the second-stage selection procedure will take. Thus, there is a trade-off between the first- and second-stage sample sizes. A natural question is then how best to choose n_0 so as to minimize the total sample size.

In our numerical experiment, we consider a fixed problem instance with $k = 100$ systems where the expected performances are $\mu_i = i$ for $i = 1, 2, \ldots, k$. For simplicity, we fix $\alpha_1 = \alpha_2 = 1 - \sqrt{0.95}$ and $\delta_2 = \delta = 5.5$ so that 5 of the 100 systems are δ-optimal. We run 100 macroreplications of the procedure at different combinations of n_0 between 5 and 15 and σ^2 between 30 and 35. On each macroreplication, we record the total sample size of the combined procedure and the size of the subset returned by the first-stage screening procedure. Heat maps in Figure 6 show the average value of these quantities at different values of n_0 and σ^2. Figure 6(a) shows that for small values of n_0, the total sample size increases rapidly, with most of it attributed to Rinott's procedure. Meanwhile, for larger values of n_0, more systems are screened out and the total sample size increases almost linearly. We can also see from Figure 6(a) that for this problem instance, a very small first-stage sample size (fewer than 10) can optimize the trade-off between the first- and second-stage sample sizes. The location of the minimizing n_0 appears to be insensitive to the variance σ^2, at least within the tested range. Figure 6(b) shows that the size of subset returned by the first-stage procedure increases as n_0 increases, as expected, and it is very insensitive to σ^2. At the optimal values of n_0 in Figure 6(a), we see that STTB screens out about 80% of the systems.

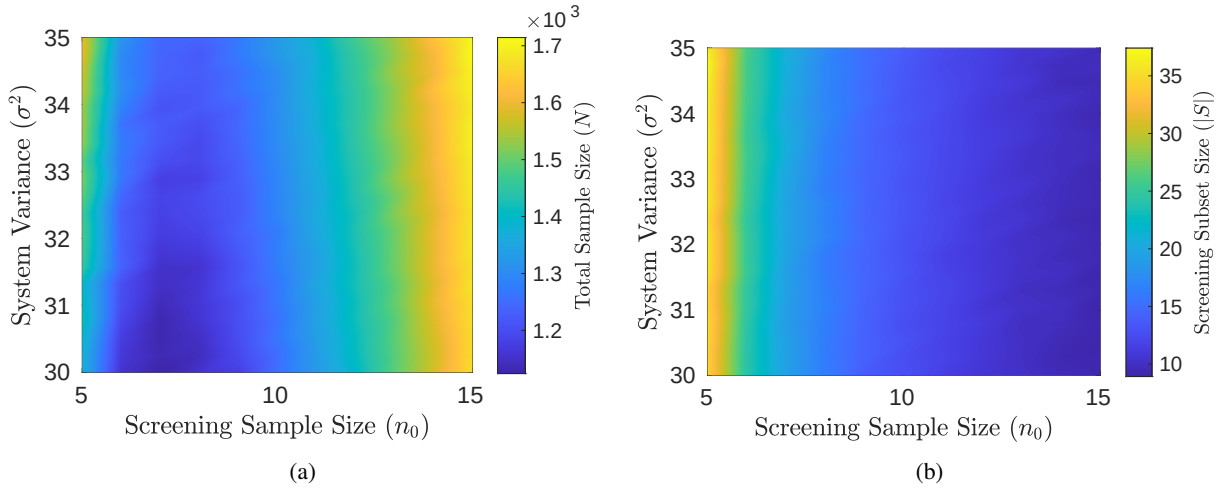


Figure 6: Heat maps of (a) the average total sample size and (b) the average size of the subset returned by STTB with respect to the sampling variance σ^2 and the initial sample size n_0 for a combined procedure of STTB (with $\delta = 0$) paired with Rinott's selection procedure on a problem instance with $k = 100$ systems.

More generally, a decision maker must also choose how to split the overall desired tolerance δ and error α so that $\delta_1 + \delta_2 = \delta$ and $(1-\alpha_1)(1-\alpha_2) = 1-\alpha$. Absent other information about the problem instance, one might naturally choose to split these quantities evenly, setting $\delta_1 = \delta_2 = \delta/2$ and $\alpha_1 = \alpha_2 = 1 - \sqrt{1-\alpha}$. Setting $\delta_1 = 0$ forces STTB to ensure that the best system survives with probability exceeding $1 - \alpha_1$, whereas setting $\alpha_1 = 0$ forces the screening procedure to always return all systems. On the other hand, the terms δ_2 and α_2 cannot be set too small without causing the sample size of the second-stage screening procedure to increase dramatically as it strives to deliver a more exacting PGS guarantee.

To study this question experimentally, we again pair the STTB and Rinott procedures, but enforce $\delta_1 + \delta_2 = \delta$ and $(1 - \alpha_1)(1 - \alpha_2) = 1 - \alpha$ where $\delta = 5.5$ and $1 - \alpha = 0.95$. The problem instance is the same as in the previous experiment, but with common variance $\sigma^2 = 30$ and initial sample size $n_0 = 10$. Figure 7 shows a heat map of the average total sample size for values of α_1 ranging from 0.005 to 0.03 and δ_1 ranging from 0 to 2.75. It can be seen from the figure that the combined procedure becomes more efficient as δ_1 shrinks to zero. This finding is somewhat surprising, as it suggests that no tolerance should be given to the first-stage screening procedure; instead, we should require that it protect the optimal solution with high probability. For any fixed δ_1, we can see the trade-off between assigning too much or too little confidence to the first-stage. For the optimal choice of $\delta_1 = 0$, the optimal splitting of α gives a higher confidence level to the first-stage procedure. Both observations indicate that most of the slack in the combined procedure, in terms of greater tolerance and allowable error, should be given to the second-stage selection procedure.

6 CONCLUSIONS

Although the bulk of recent research in the field of R&S has focused on the design and parallelization of selection procedures, screening procedures continue to be used extensively by simulation practitioners when tackling large-scale optimization problems. This tutorial highlighted a few screening procedures, discussing their differences in designs, guarantees, and empirical performance under different regimes. Further advances in the design and analysis of screening procedures will be needed to ensure they continue to serve as a cheap and effective means of identifying good candidate systems.

On the design side, although many existing screening procedures can be naively parallelized using a divide-and-conquer approach, few allow for fully sequential sampling in parallel settings, with bi-PASS being the notable exception. There are ample opportunities to build off of this adaptive-standard

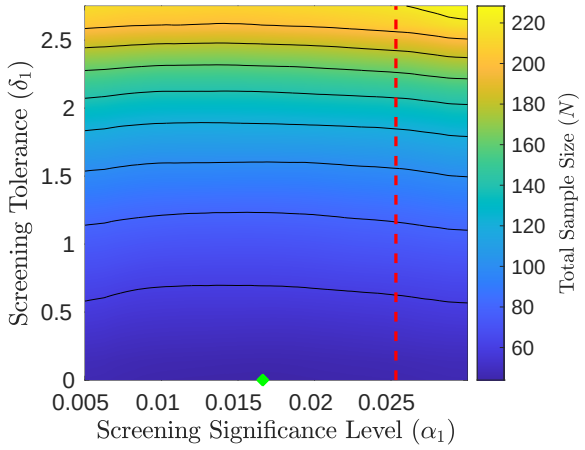


Figure 7: Heat map of the average total sample size for a combined procedure of STTB (with δ_1 and $1-\alpha_1$) and Rinott's selection procedure (with δ_2 and $1-\alpha_2$) on a problem instance with $k = 100$ systems. The red line represents evenly splitting α_1 and α_2; the green diamond indicates the optimal α_1 when $\delta_1 = 0$.

framework to handle alternative definitions of "good" systems. Other directions for developing altogether new screening frameworks include Bayesian and bootstrapping methods. Bayesian screening procedures would offer considerable flexibility in being terminated either when a sampling budget is exhausted or when the posterior probability of some good-selection event exceeds a threshold. Bootstrapping procedures also offer flexible stopping conditions while avoiding the assumption that outputs are normally distributed. Both approaches face nontrivial computational challenges for problems with many systems. In terms of analysis, more techniques are needed for studying the rate at which sequential screening procedures eliminate non-competitive systems. A measure of potential interest is the probability that a system with a given optimality gap is still in contention after expending a given budget.

ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation Grant CMMI-2206972. We also thank the reviewers for suggestions that helped to improve the organization of the tutorial.

REFERENCES

Boesel, J., B. L. Nelson, and S.-H. Kim. 2003. "Using Ranking and Selection to 'Clean Up' After Simulation Optimization". *Operations Research* 51(5):814–825.

Desu, M. M. 1970. "A Selection Problem". *The Annals of Mathematical Statistics* 41(5):1596–1603.

Dudewicz, E. J., and S. R. Dalal. 1975. "Allocation of Observations in Ranking and Selection with Unequal Variances". *Sankhyā: The Indian Journal of Statistics, Series B* 37(1):28–78.

Eckman, D. J., M. Plumlee, and B. L. Nelson. 2020. "Revisiting Subset Selection". In *Proceedings of the 2020 Winter Simulation Conference*, edited by K.-H. G. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, 2972–2983. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Eckman, D. J., M. Plumlee, and B. L. Nelson. 2022. "Plausible Screening Using Functional Properties for Simulations With Large Solution Spaces". *Operations Research* 70(6):3473–3489.

Fan, W., L. J. Hong, and B. L. Nelson. 2016. "Indifference-Zone-Free Selection of the Best". *Operations Research* 64(6):1499–1514.

Gao, S., and W. Chen. 2015a. "Efficient Subset Selection for the Expected Opportunity Cost". *Automatica* 59:19–26.

Gao, S., and W. Chen. 2015b. "A Note on the Subset Selection for Simulation Optimization". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 3768–3776. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Gupta, S. S. 1965. "On Some Multiple Decision (Selection and Ranking) Rules". *Technometrics* 7(2):225–245.

Gupta, S. S., and S. Panchapakesan. 1985. "Subset Selection Procedures: Review and Assessment". *American Journal of Mathematical and Management Sciences* 5(3-4):235–311.

Hedlund, H. E., and M. Mollaghasemi. 2001. "A Genetic Algorithm and an Indifference-Zone Ranking and Selection Framework for Simulation Optimization". In *Proceedings of the 2001 Winter Simulation Conference*, edited by B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, 417–421. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Hong, L. J., W. Fan, and J. Luo. 2021. "Review on Ranking and Selection: A New Perspective". *Frontiers of Engineering Management* 8(3):321–343.

Hong, L. J., G. Jiang, and Y. Zhong. 2022. "Solving Large-Scale Fixed-Budget Ranking and Selection Problems". *INFORMS Journal on Computing* 34(6):2930–2949.

Hunter, S. R., and B. L. Nelson. 2017. "Parallel Ranking and Selection". In *Advances in Modeling and Simulation*, edited by A. Tolk, J. Fowler, G. Shao, and E. Yücesan, 249–275. New York: Springer.

Kim, S.-H., and B. L. Nelson. 2001. "A Fully Sequential Procedure for Indifference-Zone Selection in Simulation". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 11(3):251–273.

Koenig, L. W., and A. M. Law. 1985. "A Procedure for Selecting a Subset of Size m Containing the l Best of k Independent Normal Populations, With Applications to Simulation". *Communications in Statistics–Simulation and Computation* 14(3):719–734.

Lam, K. 1986. "A New Procedure for Selecting Good Populations". *Biometrika* 73(1):201–206.

Luo, J., L. J. Hong, B. L. Nelson, and Y. Wu. 2015. "Fully Sequential Procedures for Large-Scale Ranking-And-Selection Problems in Parallel Computing Environments". *Operations Research* 63(5):1177–1194.

Ma, S., and S. G. Henderson. 2017. "An Efficient Fully Sequential Selection Procedure Guaranteeing Probably Approximately Correct Selection". In *Proceedings of the 2017 Winter Simulation Conference*, edited by V. W. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. H. Page, 2225–2236. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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.

Ni, E. C., D. F. Ciocan, S. G. Henderson, and S. R. Hunter. 2017. "Efficient Ranking and Selection in Parallel Computing Environments". *Operations Research* 65(3):821–836.

Paulson, E. 1964. "A Sequential Procedure for Selecting the Population with the Largest Mean from k Normal Populations". *The Annals of Mathematical Statistics* 35:174–180.

Pei, L., B. L. Nelson, and S. R. Hunter. 2022. "Parallel Adaptive Survivor Selection". *Operations Research*. Articles in Advance.

Rinott, Y. 1978. "On Two-Stage Selection Procedures and Related Probability-Inequalities". *Communications in Statistics* A7:799–811.

Sullivan, D. W., and J. R. Wilson. 1989. "Restricted Subset Selection Procedures for Simulation". *Operations Research* 37(1):52–71.

Trawinski, B. 1969. "Asymptotic Approximation to the Expected Size of a Selected Subset". *Biometrika* 56(1):207–213.

Wang, W., H. Wan, and X. Chen. 2023. "Bonferroni-Free and Indifference-Zone-Flexible Sequential Elimination Procedures for Ranking and Selection". *Operations Research*. Articles in Advance.

Wilson, J. R. 2001. "A Multiplicative Decomposition Property of the Screening-and-Selection Procedures of Nelson et al.". *Operations Research* 49(6):964–966.

Zhong, Y., S. Liu, J. Luo, and L. J. Hong. 2022. "Speeding Up Paulson's Procedure for Large-Scale Problems Using Parallel Computing". *INFORMS Journal on Computing* 34(1):586–606.

AUTHOR BIOGRAPHIES

JINBO ZHAO is a Ph.D. student in the Wm Michael Barnes '64 Department of Industrial and Systems Engineering at Texas A&M University. His research interests are simulation optimization and multiple criteria decision making. His e-mail address is jinbozhao@tamu.edu.

JAVIER GATICA is a Mathematical Engineering undergraduate student from Pontificia Universidad Católica de Chile specializing in numerical analysis and optimization. His research interests are scientific computing and optimization methods. His e-mail address is javier.gatica@uc.cl.

DAVID J. ECKMAN is an Assistant Professor in the Wm Michael Barnes '64 Department of Industrial and Systems Engineering at Texas A&M University. His research interests deal with optimization and output analysis for stochastic simulation models. He is a co-creator of SimOpt, a testbed of simulation optimization problems and solvers. His e-mail address is eckman@tamu.edu.