*Proceedings of the 2005 Winter Simulation Conference*
*M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds.*

# NEW DEVELOPMENTS IN RANKING AND SELECTION: AN EMPIRICAL COMPARISON OF THE THREE MAIN APPROACHES

Jürgen Branke

Institut AIFB
Universität Karlsruhe (TH)
D-76128 Karlsruhe, GERMANY

Stephen E. Chick

INSEAD
Technology Management Area
Boulevard de Constance
F-77305 Fontainebleau, FRANCE

Christian Schmidt

Institut AIFB
Universität Karlsruhe (TH)
D-76128 Karlsruhe, GERMANY

## ABSTRACT

Selection procedures are used in many applications to select the best of a finite set of alternatives, as in discrete optimization with simulation. There are a wide variety of procedures, which begs the question of which selection procedure to select. This paper (a) summarizes the main structural approaches to deriving selection procedures, (b) describes an innovative empirical testbed, and (c) summarizes results from work in progress that provides the most exhaustive assessment of selection procedures to date. The most efficient and easiest to control procedures allocate samples with a Bayesian model for uncertainty about the means, and use a new expected opportunity cost-based stopping rule.

## 1 INTRODUCTION

Ranking and selection procedures seek to identify the best of a finite set of alternatives, where best is determined with respect to the largest sampling mean, and the mean is inferred through statistical sampling. Procedures are used in commercial simulation products like ARENA (Kelton et al. 1998) and in combination with optimization tools like evolutionary algorithms or discrete optimization via simulation (Boesel et al. 2003, Branke and Schmidt 2004), among other areas.

There are three main approaches to the selection problem: the indifference zone (IZ, Kim and Nelson 2005), the expected value of information procedure (VIP, Chick and Inoue 2001), and the optimal computing budget allocation (OCBA, Chen 1996) approaches. The approaches are distinguished by their assumptions about how the evidence for correct selection is described. Recent *WSC Proceedings* describe many developments for the three main approaches, including many variations for the sampling assumptions, approximations, stopping rules and parameters that combine to define a procedure.

Few papers present a thorough assessment of how those variations compare with each other. Special cases of the VIP outperform specific IZ and OCBA procedures (in a comparison of two-stage procedures), and specific sequential VIP and OCBA procedures are more efficient than two-stage procedures (Inoue et al. 1999). He et al. (2005) derived an OCBA-type procedure, $\mathcal{OCBA}_{LL}$, that uses an expected opportunity cost (EOC) loss function inspired by the VIP approach. They showed that the original $\mathcal{OCBA}$ procedure, the new $\mathcal{OCBA}_{LL}$ and the VIP-based $\mathcal{LL}$ performed better than some other procedures in several empirical tests. Branke et al. (2005) provides the most exhaustive comparison of a wide variety of procedures (some new, some old), with new stopping rules that improve the performance of both VIP and OCBA procedures, tested over a large battery of selection problem instances.

This paper summarizes some findings from work in progress (Branke et al. 2005), and includes some observations that arose during our study but are not included in that paper. The goal is to understand the strengths and weaknesses of each approach. The focus here is on ranking and selection, but the results are intended to find techniques for approaching very large numbers of different system designs.

For each of the three main approaches, we selected "state of the art", highly sequential procedures (the IZ

procedure $\mathcal{KN}{+}{+}$ of Goldsman et al. 2002; the $\mathcal{LL}$ and 0-1 of Chick and Inoue 2001; the $\mathcal{OCBA}$ of Chen et al. 2005 and the $\mathcal{OCBA}_{LL}$ of He et al. 2005), in conjunction with new and old allocation and stopping rules. We assess:

- Efficiency: The mean evidence for correct selection as a function of the mean number of samples.
- Controllability: The ease of setting a procedure's parameters to achieve a targeted evidence level.
- Robustness: The sensitivity of a procedure's effectiveness to the underlying problem structure.

We focus on jointly independent and normally distributed simulation output with unknown variances.

The results indicate that a Bayesian EOC-based stopping rule is the most controllable and robust of the stopping rules we considered, and is typically the most efficient. It is certainly more efficient than stopping rules used in the original formulations of both the VIP and OCBA. The $\mathcal{KN}{+}{+}$ can be more efficient in some special cases, but it is typically somewhat less efficient and it appears to be very difficult to control. Probability of good selection rules can be more efficient when problem instances are sampled randomly, but are also difficult to control. Among all tested procedures, the $\mathcal{LL}$, $\mathcal{OCBA}$ and $\mathcal{OCBA}_{LL}$, modified with new stopping rules, are the most effective. The 0-1 and Equal allocations are the least effective.

## 2 ASSUMPTIONS, NOTATION, PROCEDURES

The best of $k$ simulated systems is to be identified, where "best" means the largest output mean. Let $X_{ij}$ be a random variable whose realization $x_{ij}$ is the output of the $j$-th simulation replication of system $i$, for $i = 1, \ldots, k$ and $j = 1, 2, \ldots$. Let $w_i$ and $\sigma_i^2$ be the unknown mean and variance of system $i$, and let $w_{[1]} \leq w_{[2]} \leq \ldots \leq w_{[k]}$ be the ordered means. The ordering $[\cdot]$ is unknown, and system $[k]$ is to be identified with simulation. Vectors are written in boldface, such as $\mathbf{w} = (w_1, \ldots, w_k)$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_k^2)$. The procedures considered below are derived from the assumption that simulation output is independent and normally distributed, *conditional* on $w_i$ and $\sigma_i^2$, for $i = 1, \ldots, k$,

$$\{X_{ij} : j = 1, 2, \ldots\} \overset{i.i.d.}{\sim} \texttt{Normal}\left(w_i, \sigma_i^2\right).$$

A problem instance ("configuration") is denoted by

$$\boldsymbol{\chi} = (\mathbf{w}, \boldsymbol{\sigma}^2).$$

Although the normality assumption is not always valid, it is often possible to batch outputs so that normality is approximately satisfied. Let $n_i$ be the number of replications for system $i$ run so far. Let $\bar{x}_i = \sum_{j=1}^{n_i} x_{ij}/n_i$ be the sample mean and $\hat{\sigma}_i^2 = \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2/(n_i - 1)$ be the sample variance. Let $\bar{x}_{(1)} \leq \bar{x}_{(2)} \leq \ldots \leq \bar{x}_{(k)}$ be the ordering of the sample means. The quantities $n_i$, $\bar{x}_i, \hat{\sigma}_i^2$ and $(i)$ may change as more replications are observed. Each selection procedure generates estimates $\hat{w}_i$ of $w_i$, for $i = 1, \ldots, k$. In procedures studied here, $\hat{w}_i = \bar{x}_i$, and a correct selection is when the selected system, system $\mathfrak{D}$, has the same mean as the best system, $[k]$. Usually $\mathfrak{D} = (k)$ is selected as best.

The Student $t$ distribution with mean $\mu$, precision $\kappa$, and $\nu$ degrees of freedom is denoted $\texttt{St}(\mu, \kappa, \nu)$. The variance is $\kappa^{-1}\nu/(\nu - 2)$ if $\nu > 2$. The difference $Z_i - Z_j$ of independent $t$ random variables $Z_\ell \sim \texttt{St}(\mu_\ell, \kappa_\ell, \nu_\ell)$ is approximated below by a $t$ distribution with mean $\mu_i - \mu_j$, scale $(\kappa_i^{-1} + \kappa_j^{-1})^{-1}$, and the Welch (1938) approximation for the degrees of freedom $\nu_{ij}$. Let $\Phi_\nu()$ be the cdf of the standard $t$ distribution ($\mu = 0$, $\kappa = 1$) and $\phi_\nu()$ be the pdf.

### 2.1 Evidence for Correct Selection

The procedures in Sections 2.2 to 2.4 below each run an initial stage of sampling, then allocate additional replications sequentially until the evidence for correct selection is sufficient. Loss functions are used here to measure selection quality. The zero-one loss function, $\mathcal{L}_{0-1}(\mathfrak{D}, \mathbf{w}) = \mathbb{1}\left\{w_{\mathfrak{D}} \neq w_{[k]}\right\}$, equals 1 if the best system is not correctly selected, and is 0 otherwise. The opportunity cost $\mathcal{L}_{oc}(\mathfrak{D}, \mathbf{w}) = w_{[k]} - w_{\mathfrak{D}}$ is 0 if the best system is correctly selected, and is otherwise the difference between the best and selected systems. The opportunity cost makes more sense in business applications.

The IZ procedures take a frequentist perspective. The frequentist probability of correct selection ($\text{PCS}_{\text{IZ}}$) is the probability that the system selected as best (system $\mathfrak{D}$) is the system with the highest mean (system $[k]$), conditional on the problem instance. The probability is with respect to the simulation output $X_{ij}$ that determines $\mathfrak{D}$,

$$\text{PCS}_{\text{IZ}}(\boldsymbol{\chi}) \overset{\text{def}}{=} 1 - E\left[\mathcal{L}_{0-1}(\mathfrak{D}, \mathbf{w}) \,|\, \boldsymbol{\chi}\right]. \tag{1}$$

Indifference zone procedures attempt to guarantee a lower bound on $\text{PCS}_{\text{IZ}}$, subject to the indifference-zone constraint that the best system is at least $\delta^*$ better than the others,

$$\text{PCS}_{\text{IZ}}(\boldsymbol{\chi}) \geq 1 - \alpha^*, \text{for all } \boldsymbol{\chi} \text{ s.t. } w_{[k]} \geq w_{[k-1]} + \delta^*. \tag{2}$$

The frequentist EOC (Chick and Wu 2005) is

$$\text{EOC}_{\text{IZ}}(\boldsymbol{\chi}) \overset{\text{def}}{=} E\left[\mathcal{L}_{oc}(\mathfrak{D}, \mathbf{w}) \,|\, \boldsymbol{\chi}\right]. \tag{3}$$

Bayesian approaches (VIP, OCBA) use the posterior distribution of the unknown means to measure the quality

of a selection. Given the data $\mathcal{E}$ seen so far, the quantities

$$\begin{aligned} \text{PCS}_{Bayes} &\stackrel{\text{def}}{=} 1 - E\left[\mathcal{L}_{0-1}(\mathfrak{D}, \mathbf{W}) \,|\, \mathcal{E}\right] \\ \text{EOC}_{Bayes} &\stackrel{\text{def}}{=} E\left[\mathcal{L}_{oc}(\mathfrak{D}, \mathbf{W}) \,|\, \mathcal{E}\right], \end{aligned} \tag{4}$$

measure selection quality, the expectation taken over both $\mathfrak{D}$ and the posterior distribution of $\mathbf{W}$. Assuming a non-informative prior distribution for the unknown mean and variance, the posterior marginal distribution for the unknown means $W_i$ given $n_i > 2$ samples is $\texttt{St}\left(\bar{x}_i, n_i/\hat{\sigma}_i^2, \nu_i\right)$ where $\nu_i = n_i - 1$ (Chick and Inoue 2001). Each of the Bayesian procedures (VIP and OCBA) select the system with the best posterior mean after sampling stops, $\mathfrak{D} = (k)$.

Approximations in the form of bounds on the above losses are useful to improve the speed of computing an allocation. *Slepian's* inequality states the posterior evidence that system $(k)$ is best satisfies

$$\text{PCS}_{Bayes} \geq \prod_{j:(j)\neq(k)} \Pr\left(W_{(k)} > W_{(j)} \,|\, \mathcal{E}\right). \tag{5}$$

The r.h.s. of Inequality (5) is approximately (Welch)

$$\text{PCS}_{Slep} = \prod_{j:(j)\neq(k)} \Phi_{\nu_{(j)(k)}}(d^*_{jk}). \tag{6}$$

if $d^*_{jk}$ is a normalized distance for systems $(j)$ and $(k)$,

$$\begin{aligned} d^*_{jk} &= d_{(j)(k)}\lambda_{jk}^{1/2}, \qquad (7) \\ d_{(j)(k)} = (\bar{x}_{(k)} - \bar{x}_{(j)}) \quad &\text{and} \quad \lambda_{jk}^{-1} = \left(\frac{\hat{\sigma}^2_{(j)}}{n_{(j)}} + \frac{\hat{\sigma}^2_{(k)}}{n_{(k)}}\right). \end{aligned}$$

The term $\text{EOC}_{Bayes}$ may be expensive to compute if $k > 2$. Summing the losses from $(k-1)$ pairwise comparisons between the current best and each other system gives an easily computed upper bound (Chick and Inoue 2001, Chick and Inoue 2002). Let $f_{(j)(k)}(\cdot)$ be the posterior pdf for the difference $W_{(j)} - W_{(k)}$ given all data $\mathcal{E}$ (approximately $\texttt{St}\left(-d_{(j)(k)}, \lambda_{jk}, \nu_{(j)(k)}\right)$ distributed), and set $\Psi_\nu[s] = \int_{u=s}^{\infty}(u-s)\phi_\nu(u)du = \frac{\nu+s^2}{\nu-1}\phi_\nu(s) - s\Phi_\nu(-s)$. Then $\text{EOC}_{Bayes} \leq \text{EOC}_{Bonf}$, where

$$\begin{aligned} \text{EOC}_{Bonf} &= \sum_{j:(j)\neq(k)} \int_{w=0}^{\infty} w\, f_{(j)(k)}(w)\, dw \\ &\approx \sum_{j:(j)\neq(k)} \lambda_{jk}^{-1/2}\Psi_{\nu_{(j)(k)}}\left[d^*_{jk}\right]. \end{aligned} \tag{8}$$

Some IZ procedures satisfy frequentist probability of good selection ($\text{PGS}_{\text{IZ},\delta^*} \geq 1-\alpha^*$, for selections within $\delta^*$ of the best) guarantees (Nelson and Banerjee 2001). We propose the following PCS-related measure for VIP and OCBA stopping rules to incorporate $\delta^*$ to stop sampling if all competitors for the best are "good enough",

$$\text{PGS}_{Slep,\delta^*} = \prod_{j:(j)\neq(k)} \Phi_{\nu_{(j)(k)}}(\lambda_{jk}^{1/2}(\delta^* + d_{(j)(k)})).$$

Chen and Kelton (2005) used $\max$ instead of $+$,

$$\text{PCS}_{Slep,\delta^*} = \prod_{j:(j)\neq(k)} \Phi_{\nu_{(j)(k)}}(\lambda_{jk}^{1/2}\max\{\delta^*, d_{(j)(k)}\}).$$

The VIP and OCBA will use these **stopping rules** below:

1. Sequential ($\mathcal{S}$): Repeat sampling if $\sum_{i=1}^{k} n_i < B$ for a given total budget $B$.
2. Repeat if $\text{PCS}_{Slep,\delta^*} < 1-\alpha^*$ for a given $\delta^*, \alpha^*$.
3. Repeat if $\text{PGS}_{Slep,\delta^*} < 1-\alpha^*$ for a given $\delta^*, \alpha^*$.
4. Repeat if $\text{EOC}_{Bonf} > \beta^*$, for an EOC target $\beta^*$.

We use $\text{PCS}_{Slep}$ to denote $\text{PCS}_{Slep,0}$. The IZ requires $\delta^* > 0$, but we allow $\delta^* = 0$ for the VIP and OCBA to allow for a pure PCS-based stopping condition. All previously published sequential VIP and OCBA work appears to have used the $\mathcal{S}$ stopping rule, but the other stopping rules will be shown to improve the efficiency of both approaches. Let $\text{PICS} = 1 - \text{PCS}$ and $\text{PBS}_{\delta^*} = 1 - \text{PGS}_{\delta^*}$ measure evidence for the probability of *incorrect* and *bad* selections.

### 2.2 Indifference Zone (IZ) Procedure

The IZ approach (Kim and Nelson 2005) seeks to guarantee $\text{PCS}_{\text{IZ}} \geq 1-\alpha^*$, whenever the best system is at least $\delta^*$ better than the other systems. Early IZ procedures were statistically conservative in the sense of excess sampling except with very particular configurations of the means. The $\mathcal{KN}$ family of procedures improves sampling efficiency over a broad set of configurations (Kim and Nelson 2001). While a PCS guarantee in the sense of Equation (2) was not proven, an asymptotic guarantee as $\delta^* \to 0$ was shown. One member of the family, $\mathcal{KN}{+}{+}$ (Goldsman et al. 2002), might be considered to be the state of the art for the IZ approach. That procedure can handle correlation. Here we specialize Procedure $\mathcal{KN}{+}{+}$ for *independent* replications. The procedure screens out some systems as runs are made, and each non-eliminated system is simulated the same number of times.

**Procedure $\mathcal{KN}{+}{+}$ (independent samples)**

1. Specify a confidence level $1-\alpha^*$, an indifference-zone parameter $\delta^* > 0$, a first-stage sample size $n_0 > 2$ per system, and a number $\xi$ of samples per noneliminated system per subsequent stage.

2. Initialize the set of noneliminated systems, $I \leftarrow \{1, \ldots, k\}$, set $n \leftarrow 0, \tau \leftarrow n_0$.
3. WHILE $|I| > 1$ DO another stage:
   (a) Observe $\tau$ additional samples from system $i$, independent of all other samples, for all $i \in I$. Set $n \leftarrow n + \tau$. Set $\tau \leftarrow \xi$.
   (b) Update: For all $i \in I$, set $\bar{x}_i \leftarrow \sum_{j=1}^{n} x_{ij}/n$ and $\hat{\sigma}_i^2 \leftarrow \sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2/(n-1)$. Set $\eta \leftarrow \frac{1}{2}\left\{[2(1-(1-\alpha^*)^{1/(k-1)})]^{-2/(n-1)} - 1\right\}$ and $h^2 \leftarrow 2\eta(n-1)$.
   (c) Screen: For all $i, j \in I$, $i > j$, set $d_{ij} \leftarrow \bar{x}_j - \bar{x}_i$ and $\epsilon_{ij} \leftarrow \max\left\{0, \frac{\delta^*}{2n}\left(\frac{h^2(\hat{\sigma}_i^2 + \hat{\sigma}_j^2)}{\delta^{*2}} - n\right)\right\}$. If $d_{ij} > \epsilon_{ij}$ then $I \leftarrow I \backslash \{i\}$. If $d_{ij} < -\epsilon_{ij}$ then $I \leftarrow I \backslash \{j\}$.
4. Select the remaining system ($\mathfrak{D}$) as best.

### 2.3 Value of Information Procedure (VIP)

Two VIPs in Chick and Inoue (2001) allocate samples to each alternative to maximize the expected value of information (EVI) subject to a sampling budget constraint. **Procedures 0-1**($\mathcal{S}$) and $\mathcal{LL}(\mathcal{S})$ are sequential variations of those procedures that improve $\text{PCS}_{Bayes}$ and $\text{EOC}_{Bayes}$, respectively. Allocations were derived with asymptotic approximations to the EVI. They allocate $\tau$ replications per stage until a total of $B$ replications are run. That stopping rule allows for full control of the number of replications. This section examines stopping rules that afford more efficiency and a more direct comparison with IZ procedures.

**Procedure 0-1.**

1. Specify a first-stage sample size $n_0 > 2$, a number of samples $\tau > 0$ to allocate per subsequent stage, and **stopping rule** parameters (Section 2.1).
2. Take independent replications $X_{i1}, \ldots, X_{in_0}$, and initialize the number of replications $n_i \leftarrow n_0$ run so far for each system, $i = 1, \ldots, k$.
3. Determine the sample statistics $\bar{x}_i \leftarrow \sum_{j=1}^{n_i} x_{ij}/n_i$ and $\hat{\sigma}_i^2 \leftarrow \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2/(n_i - 1)$, and the sample mean ordering, so that $\bar{x}_{(1)} \leq \ldots \leq \bar{x}_{(k)}$.
4. WHILE stopping rule not satisfied DO:
   (a) Initialize the set of systems considered for additional replications, $\mathcal{S} \leftarrow \{1, \ldots, k\}$.
   (b) For each $(i)$ in $\mathcal{S}\backslash\{(k)\}$: If $(k) \in \mathcal{S}$ then set $\lambda_{ik}^{-1} \leftarrow \hat{\sigma}_{(i)}^2/n_{(i)} + \hat{\sigma}_{(k)}^2/n_{(k)}$, and set $\nu_{(i)(k)}$ with Welch's approximation. If $(k) \notin \mathcal{S}$ then set $\lambda_{ik} \leftarrow n_{(i)}/\hat{\sigma}_{(i)}^2$ and $\nu_{(i)(k)} \leftarrow n_{(i)} - 1$.
   (c) Tentatively allocate $\tau_{(i)}$ replications to systems $(i) \in \mathcal{S}$ (set $\tau_{(j)} \leftarrow 0$ for $(j) \notin \mathcal{S}$):

$$\tau_{(i)} \leftarrow \frac{\tau + \sum_{j \in \mathcal{S}} n_j}{\sum_{j \in \mathcal{S}} \left(\frac{\hat{\sigma}_j^2 \gamma_j}{\hat{\sigma}_{(i)}^2 \gamma_{(i)}}\right)^{1/2}} - n_{(i)}, \text{ where}$$

$$\gamma_{(i)} \leftarrow \begin{cases} \lambda_{ik} d_{ik}^* \phi_{\nu_{(i)(k)}}(d_{ik}^*) & \text{for } (i) \neq (k) \\ \sum_{(j) \in \mathcal{S}\backslash\{(k)\}} \gamma_{(j)} & \text{for } (i) = (k). \end{cases}$$

   (d) IF any $\tau_{(i)} < 0$ THEN remove $(i)$ from $\mathcal{S}$ for all $(i)$ with $\tau_{(i)} \leq 0$; go to Step 4b ELSE round the $\tau_i$ so $\sum_{i=1}^{k} \tau_i = \tau$; go to Step 4e.
   (e) Run $\tau_i$ additional independent replications for system $i$, for $i = 1, \ldots, k$. Update $n_i \leftarrow n_i + \tau_i$; the sample statistics $\bar{x}_i \leftarrow \sum_{j=1}^{n_i} x_{ij}/n_i$ and $\hat{\sigma}_i^2 \leftarrow \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2/(n_i - 1)$, and the sample mean ordering, $\bar{x}_{(1)} \leq \ldots \leq \bar{x}_{(k)}$.
5. Select system $\mathfrak{D} = (k)$ as best.

Step 4b uses the Welch approximation, and the formulas in Step 4c are derived in Chick and Inoue (2001) from optimality conditions to improve a Bonferroni-like bound on the EVI for asymptotically large $\tau$. Step 4 requires the selection of a stopping rule. The resulting procedures are named 0-1($\mathcal{S}$), 0-1($\text{PCS}_{Slep,\delta^*}$), 0-1($\text{PGS}_{Slep,\delta^*}$), 0-1($\text{EOC}_{Bonf}$), with the stopping rule in parentheses.

Procedure $\mathcal{LL}$ (for linear loss) is a variant of 0-1 where sampling allocations seek to minimize $\text{EOC}_{Bonf}$. This procedure can also use any of the stopping rules.

**Procedure $\mathcal{LL}$.** Same as 0-1, except set $\gamma_{(i)}$ in Step 4c to

$$\gamma_{(i)} \leftarrow \begin{cases} \lambda_{ik}^{1/2} \frac{\nu_{(i)(k)} + \lambda_{ik}(d_{ik}^*)^2}{\nu_{(i)(k)} - 1} \phi_{\nu_{(i)(k)}}(d_{ik}^*) & \text{for } (i) \neq (k) \\ \sum_{(j) \in \mathcal{S}\backslash\{(k)\}} \gamma_{(j)} & \text{for } (i) = (k). \end{cases}$$

### 2.4 OCBA Procedures

The OCBA (Chen 1996, Chen et al. 2005) assumes that if $\tau$ replications are allocated for system $i$, but none are allocated for the others, then the variance scales accordingly,

$$\begin{aligned} \tilde{W}_i &\sim \texttt{St}\left(\bar{x}_i, (n_i + \tau)/\hat{\sigma}_i^2, n_i - 1 + \tau\right) \\ \tilde{W}_j &\sim \texttt{St}\left(\bar{x}_j, n_j/\hat{\sigma}_j^2, n_j - 1\right) \qquad \text{for } j \neq i. \end{aligned}$$

The usual OCBA assumes normal distributions to approximate the effect, but we use $t$ distributions, for consistency with a Bayesian assumption for the unknown variance. Chen et al. (2005) found no notable difference in performance when comparing a normal versus $t$ distribution for $\tilde{W}_i$. Allocating an additional $\tau$ replications to system $i$, but

no replications to the others, leads to an *estimated approximate probability of correct selection* (EAPCS) evaluated with respect to $\tilde{\mathbf{W}} = (\tilde{W}_1, \ldots, \tilde{W}_k)$,

$$\text{EAPCS}_i = \prod_{j:(j)\neq(k)} \Pr\left(\tilde{W}_{(j)} < \tilde{W}_{(k)} \,|\, \mathcal{E}\right) \quad (9)$$

$$\approx \prod_{j:(j)\neq(k)} (1 - \Phi_{\tilde{\nu}_{(j)(k)}}(\tilde{\lambda}_{jk}^{1/2} d_{(j)(k)})) \quad (10)$$

$$\tilde{\lambda}_{jk} = \left(\frac{\hat{\sigma}^2_{(k)}}{\tilde{n}_{(k)}} + \frac{\hat{\sigma}^2_{(j)}}{\tilde{n}_{(j)}}\right)^{-1} \quad (11)$$

$$\tilde{n}_{(\ell)} = n_{(\ell)} + \tau \mathbb{1}\{(\ell) = i\}.$$

The OCBA uses these approximations to sequentially allocate samples at each stage to systems that most increase $\text{EAPCS}_i - \text{PCS}_{Slep}$. An innovation for the OCBA is the use of the stopping rules from Section 2.3.

**Procedure** $\mathcal{OCBA}$**.**

1. Specify a first-stage sample size $n_0 > 2$, a number $q$ of systems to simulate per stage, a sampling increment $\tau > 0$ to allocate per subsequent stage, and stopping rule parameters.
2. Do steps 2-3 of Procedure 0-1.
3. WHILE stopping rule not satisfied DO:
   (a) Compute $\text{EAPCS}_i$ for $i = 1, \ldots, k$.
   (b) Set $\tau_i \leftarrow \tau/q$ for the $q$ systems with largest $\text{EAPCS}_i - \text{PCS}_{Slep}$, set $\tau_j \leftarrow 0$ for the others.
   (c) Take $\tau_i$ additional observations for system $i$.
   (d) For all $i$ with $\tau_i > 0$, update $n_i \leftarrow n_i + \tau_i$, the sample statistics $\bar{x}_i \leftarrow \sum_{j=1}^{n_i} x_{ij}/n_i$, $\hat{\sigma}_i^2 \leftarrow \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2/(n_i - 1)$, and order statistics, so that $\bar{x}_{(1)} \leq \ldots \leq \bar{x}_{(k)}$.
4. Select system $\mathfrak{D} = (k)$ as best.

He et al. (2005) proposed an OCBA variation that accounts for the expected opportunity cost. Define AEOC to be the approximation to $\text{EOC}_{Bonf}$ in the right hand side of Equation (8). The OCBA-like approximation for EOC (cf. Equation (9)) with respect to $\tilde{\mathbf{W}}$ is

$$\text{EEOCS}_i = \sum_{j:(j)\neq(k)} \tilde{\lambda}_{jk}^{-1/2} \Psi_{\tilde{\nu}_{(j)(k)}}\left[\tilde{\lambda}_{jk}^{1/2} d_{(j)(k)}\right]. \quad (12)$$

**Procedure** $\mathcal{OCBA}_{LL}$ allocates replications to systems that maximize the improvement in expected opportunity cost (linear loss), $\text{AEOC} - \text{EEOCS}_i$ in Step 3b.

We consider two other variations on the allocations that generalize the idea of $\text{EAPCS}_i$ to account for $\delta^*$ using the approximations at the end of Section 2.1

**Procedure** $\mathcal{OCBA}_{\delta^*}$ allocates replications to systems that maximize the improvement in $\text{EAPGS}_{i,\delta^*} - \text{PGS}_{Slep,\delta^*}$ in Step 3b, with $\delta^* > 0$.

**Procedure** $\mathcal{OCBA}_{max,\delta^*}$ allocates replications to systems that maximize $\text{EAPCS}_{i,\delta^*} - \text{PCS}_{Slep,\delta^*}$ in Step 3b, with $\delta^* > 0$ (cf. Chen and Kelton 2005).

Each procedure can use any of the stopping rules. We implemented a fully sequential OCBA ($q = \tau = 1$).

## 3 NUMERICAL TESTS

Table 1 summarizes the procedures that we evaluated. The naming convention is the type of allocation followed by the stopping rule in parenthesis. There are no EOC analogs to $\mathcal{OCBA}_{\delta^*}$ and $\mathcal{OCBA}_{max,\delta^*}$ because EOC already accounts for the size of the difference between the best and second best in the allocations. Procedure Equal allocates the same number of replications to each system. Branke et al. (2005) differs in that it derives another set of VIP allocations and tests more configurations, but does not assess the $\mathcal{OCBA}_{max,\delta^*}$ allocation or the $\text{PCS}_{Slep,\delta^*}$ stopping rule.

### 3.1 Evaluation Criteria

Theory that compares the different approaches is hard to develop due to the differing assumptions and approximations of each. We turn here to empirical and practical perspectives.

The efficiency of a procedure is a frequentist measure of evidence for correct selection ($\text{PCS}_{\text{iz}}$, $\text{PGS}_{\text{iz},\delta^*}$ and $\text{EOC}_{\text{iz}}$) introduced in Section 2.1, as a function of the average number of replications $E[N]$. For each problem instance and sampling allocation, the stopping rule parameters *implicitly* define *efficiency curves* in the $(E[N], \log(1 - \text{PCS}_{\text{iz}}))$ plane. Efficiency curves for $\text{EOC}_{\text{iz}}$ and $\text{PGS}_{\text{iz},\delta^*}$ are defined similarly. "More efficient" procedures have curves that are below those of other procedures.

Efficiency curves ignore the question of how to set a procedure's parameters to achieve a particular $\text{PCS}_{\text{iz}}$ or $\text{EOC}_{\text{iz}}$. As a practical matter, some deviation may occur between a stopping rule target and the actual value achieved. The deviation between the desired and realized performance is measured with *target curves* that plot $(\log \alpha^*, \log(1 - \text{PCS}_{\text{iz}}))$ for PCS-based targets $1 - \alpha^*$, and $(\log \beta^*, \log \text{EOC}_{\text{iz}})$ for EOC targets $\beta^*$. Curves that typically follow the line $y = x$ for a broad class of problems indicate that a procedure is "controllable". If the curves depend strongly on the problem instance or $\delta^*$, it is hard to obtain the desired level of evidence for correct selection without additional knowledge of the problem structure.

Procedures that are both efficient and controllable over a broad range of problem instances (robust) are desirable.

Figures were estimated by running $10^5$ macro-replications for each combination of problem instance, sampling allocation, and stopping rule parameter value.

Table 1: Procedures Tested in Addition to Procedures $\mathcal{KN}{+}{+}$ and Equal.

| | VIP | OCBA | OCBA for $\text{PGS}_{Slep,\delta^*}$ | OCBA for $\text{PCS}_{Slep,\delta^*}$ |
|---|---|---|---|---|
| PCS-based allocations | 0-1($\mathcal{S}$) | $\mathcal{OCBA}(\mathcal{S})$ | $\mathcal{OCBA}_{\delta^*}(\mathcal{S})$ | $\mathcal{OCBA}_{max,\delta^*}(\mathcal{S})$ |
| | 0-1($\text{PCS}_{Slep,\delta^*}$) | $\mathcal{OCBA}(\text{PCS}_{Slep,\delta^*})$ | $\mathcal{OCBA}_{\delta^*}(\text{PCS}_{Slep,\delta^*})$ | $\mathcal{OCBA}_{max,\delta^*}(\text{PCS}_{Slep})$ |
| | 0-1($\text{PGS}_{Slep,\delta^*}$) | $\mathcal{OCBA}(\text{PGS}_{Slep,\delta^*})$ | $\mathcal{OCBA}_{\delta^*}(\text{PGS}_{Slep,\delta^*})$ | $\mathcal{OCBA}_{max,\delta^*}(\text{PGS}_{Slep})$ |
| | 0-1($\text{EOC}_{Bonf}$) | $\mathcal{OCBA}(\text{EOC}_{Bonf})$ | $\mathcal{OCBA}_{\delta^*}(\text{EOC}_{Bonf})$ | $\mathcal{OCBA}_{max,\delta^*}(\text{EOC}_{Bonf})$ |
| EOC-based allocations | $\mathcal{LL}(\mathcal{S})$ | $\mathcal{OCBA}_{LL}(\mathcal{S})$ | | |
| | $\mathcal{LL}(\text{PCS}_{Slep,\delta^*})$ | $\mathcal{OCBA}_{LL}(\text{PCS}_{Slep,\delta^*})$ | | |
| | $\mathcal{LL}(\text{PGS}_{Slep,\delta^*})$ | $\mathcal{OCBA}_{LL}(\text{PGS}_{Slep,\delta^*})$ | | |
| | $\mathcal{LL}(\text{EOC}_{Bonf})$ | $\mathcal{OCBA}_{LL}(\text{EOC}_{Bonf})$ | | |

We used the Gnu Scientific Libary (gsl) for calculating cdfs and for the Mersenne twister RNG (Matsumoto and Nishimura 1998, with 2002 revised seeding), and FILIB++ (Lerch et al. 2001) for interval arithmetic. The interval arithmetic discerned whether the maximum EVI or $\text{EAPCS}_i$ were numerically unique (due to numerical stability issues for the OCBA family and other VIP procedures that were tested but not reported here).

### 3.2 Configurations

In a *slippage configuration (SC)* the means of all systems except the best are tied for second best. We use the parameters $\delta, \rho$ to describe the configurations of the independent outputs with $\texttt{Normal}\left(w_i, \sigma_i^2\right)$ distribution,

$$\begin{aligned} X_{1j} &\sim \texttt{Normal}\left(0, \sigma_1^2\right) \\ X_{ij} &\sim \texttt{Normal}\left(-\delta, \sigma_1^2/\rho\right) \text{ for } i = 2, \ldots, k \\ \delta^* &= \gamma\delta. \end{aligned}$$

All systems have the same variance if $\rho = 1$. The best system has the largest variance if $\rho > 1$. We set $\sigma_1^2 = 2\rho/(1+\rho)$ so that $\text{Var}[X_{1j} - X_{ij}]$ is constant for all $\rho > 0$. The parameter $\gamma$ allows the indifference zone parameter $\delta^*$ to differ from the difference in means $\delta$.

In a *monotone decreasing means (MDM)* configuration the means are equally spaced. Again $\rho$ controls the variances, $\gamma$ relates $\delta^*$ to the difference in means, and independent outputs have a $\texttt{Normal}\left(w_i, \sigma_i^2\right)$ distribution,

$$\begin{aligned} X_{ij} &\sim \texttt{Normal}\left(-(i-1)\delta, 2\rho^{2-i}/(1+\rho)\right) \\ \delta^* &= \gamma\delta. \end{aligned}$$

*Random problem instances* (RPI) are more realistic in the sense that problems faced in practice typically are not the SC or MDM configuration. The RPI experiment here samples configurations $\boldsymbol{\chi}$ from normal-inverse gamma family. If $S \sim \texttt{InvGamma}\left(\alpha, \beta\right)$, then $E[S] = \beta/(\alpha - 1)$ and $S^{-1} \sim \texttt{Gamma}\left(\alpha, \beta\right)$ with $E[S^{-1}] = \alpha\beta^{-1}$ and $\text{Var}[S^{-1}] = \alpha\beta^{-2}$. A random $\boldsymbol{\chi}$ is generated by sampling the $\sigma_i^2$ independently, then sampling the $W_i$ conditionally independent, given $\sigma_i^2$,

$$\begin{aligned} p(\sigma_i^2) &\sim \texttt{InvGamma}\left(\alpha, \beta\right) \\ p(W_i \,|\, \sigma_i^2) &\sim \texttt{Normal}\left(\mu_0, \sigma_i^2/\eta\right). \end{aligned} \tag{13}$$

Increasing $\eta$ makes the means more similar. We set $\beta = \alpha - 1 > 0$ to standardize the mean of the variances to be 1. Increasing $\alpha$ reduces the variability in the variances. The noninformative prior distributions used to derive VIP and OCBA procedures correspond to $\eta \to 0$, so there is a mismatch in the sampling distribution of $\boldsymbol{\chi}$ and the prior distributions assumed by the VIP and OCBA.

For the SC and MDM, we tested *many* combinations of $n_0$, number of systems $k$, spacings of the means, and degrees of heterogeneity in the variances. For the RPI we tested $k = 2, 5, 10; \eta = .707, 1, 1.414, 2; \alpha = 2.5, 100$.

## 4 RESULTS

The results below summarize work to date for Branke et al. (2005), which will present a much more thorough discussion and broader set of experiments. In addition, we compare $\text{PCS}_{Slep,\delta^*}$ with $\text{PGS}_{Slep,\delta^*}$, the OCBA's use of $t$ vs. Gaussian distributions, and an alternative to Welch's approximation. Additional subscripts refer to specific parameter values (e.g. $\mathcal{KN}{+}{+}_{\delta^*}$ specifies $\delta^*$). Graphs below use $n_0 = 6$.

For $k = 2$ systems and equal variance, the Equal allocation is optimal from both Bayesian and frequentist perspectives (e.g. Gupta and Miescke 1994). Figure 1 compares different stopping rules on SC or MDM (which are equivalent for $k = 2$) with Equal. The $\text{EOC}_{Bonf}$ stopping rule is more efficient than the $\text{PCS}_{Slep}$ stopping rule, which is more efficient than the $\mathcal{S}$ stopping rule, an order that could be observed for all SC and MDM configurations (also for $k > 2$, or when $k = 2$ and the variances are unequal, and for $\text{PCS}_{\text{iz}}$ efficiency as well as $\text{EOC}_{\text{iz}}$ efficiency). As $\mathcal{KN}{+}{+}$ also samples equally for $k = 2$, the efficiency of its stopping rule can be directly compared with the other stopping rules on the scenario of Figure 1. For low levels of

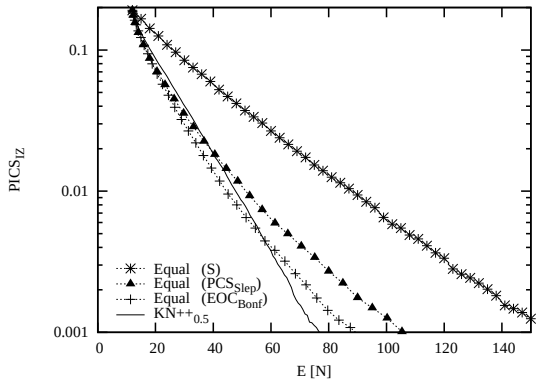


Figure 1: Performance of Equal Allocation and $\mathcal{KN}{+}{+}$ with $\delta^* = \delta$ (SC, $k = 2$, $\delta = 0.5$, $\rho = 1$).

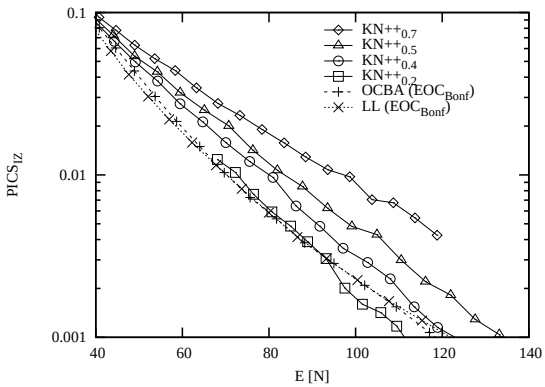


Figure 2: Efficiency of Procedure $\mathcal{KN}{+}{+}$ Better for Smaller $\delta^*$ (MDM, $k = 5$, $\delta = 0.5$, $\rho = 1$).

evidence for correct selection, $\mathcal{KN}{+}{+}$ was more efficient than the Bayesian EOC$_{Bonf}$ stopping rule, but EOC$_{Bonf}$ is more competitive for larger levels of PICS$_{\text{iz}}$. In general, $\mathcal{KN}{+}{+}$ tended to do better than in the graph for SC with close means ($\delta < 0.5$) and was worse than EOC$_{Bonf}$ for all PICS$_{\text{iz}} > 10^{-3}$ in tests with $\delta > 2^{-1/2}$.

For $k = 2$ and equal variance, allocation procedures other than Equal naturally perform slightly worse. However, as configurations diverge from this special case ($\rho \neq 1, k > 2$) the relative efficiency of $\mathcal{KN}{+}{+}$ and Equal allocation became worse than that for $\mathcal{LL}$(EOC$_{Bonf}$), $\mathcal{OCBA}_{LL}$(EOC$_{Bonf}$) and $\mathcal{OCBA}$(EOC$_{Bonf}$) (assuming that $\delta^*$ was set to the difference between the best and second best). Figure 2 shows a typical phenomenon seen for a variety of SC or MDM configurations. The efficiency curve of $\mathcal{KN}{+}{+}$ with $\delta^* = \delta$ is worse than for e.g. $\mathcal{OCBA}(EOC_{Bonf})$ or $\mathcal{LL}(EOC_{Bonf})$, however it becomes competitive as $\delta^*$ is decreased. Unfortunately, this phenomenon depends on the problem configuration (it does not hold for RPI, for example), and the correspondence between the desired and obtained PICS varies widely depending upon the relation of the difference $\delta$ between the best two systems, and the $\delta^*$ selected for the procedure. Figure 3 shows that as $\delta^*$ gets smaller, $\mathcal{KN}{+}{+}$ samples much more than necessary to obtain a given desired level of evidence (curve below the diagonal on the target plot). This makes it difficult to set $\alpha^*$ to actually achieve a desired PICS$_{\text{iz}}$ with $\mathcal{KN}{+}{+}$, as the target curves are highly sensitive to the underlying (and typically unknown) configuration. One samples much more than necessary if $\delta^* \ll \delta$.

For the RPI configurations, it was necessary to choose $\delta^* > 0$ for the PGS$_{Slep,\delta^*}$ and PCS$_{Slep,\delta^*}$ stopping rules because there was a reasonable probability that the two best systems had very similar means, in which case $\delta^* = 0$ resulted in excessive sampling. Therefore $\delta^* = 0$ is to be avoided in practice. The EOC$_{Bonf}$ rule does not suffer from that problem, and it replaces the difficulty of specifying

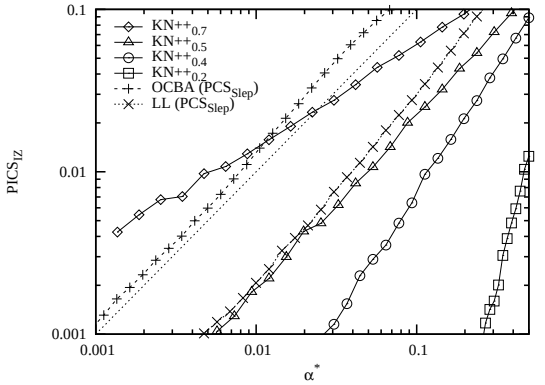


Figure 3: Sensitivity of $\mathcal{KN}{+}{+}$ w.r.t. $\delta^*$ and $\alpha^*$ (MDM, $k = 5$, $\delta = 0.5$, $\rho = 1$).

two parameters, $\delta^*, \alpha^*$, with one parameter, $\beta^*$. EOC$_{Bonf}$ gave excellent control over the actual EOC$_{\text{iz}}$ received for the RPI. For MDM, the target plot for EOC$_{Bonf}$ tended to be parallel to the desired $y = x$, but was shifted high or low depending upon $\delta$, whereas PGS$_{Slep,\delta^*}$ could have different slopes for different $\delta^*$, not unlike $\mathcal{KN}{+}{+}$ in Figure 3

Figure 4 compares different stopping rules in combination with Equal allocation based on PGS$_{\text{iz},\delta^*}$ efficiency. The effect is quite dramatic, with PGS$_{Slep,\delta^*}$ stopping rule and appropriate $\delta^*$ performing best. Note that while this is not surprising for PGS$_{\text{iz},\delta^*}$ efficiency, also for EOC$_{\text{iz}}$ efficiency there seems to exist a setting for $\delta^*$ such that PGS$_{Slep,\delta^*}$ outperforms the EOC$_{Bonf}$ stopping rule (Figure 5). Whether that finding is of practical use remains to be seen, as it is not yet clear how to set $\delta^*$, and PGS$_{Slep,\delta^*}$ under-delivers EOC relative to $\beta^* = \delta^*\alpha^*$ (Figure 6).

Another interesting fact to note in Figure 4 and Figure 5 is that the line for Equal allocation and Budget stopping rule is curved, while it is straight for all SC and MDM configurations. While this might at first sight appear to be inconsistent with a hypothesis of exponential convergence

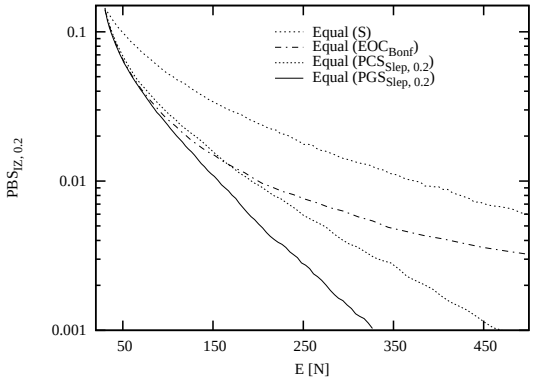


Figure 4: $\text{PBS}_{\text{iz},\delta^*} = 1 - \text{PGS}_{\text{iz},\delta^*}$ Efficiency for Equal Allocation (RPI, $k = 5$, $\eta = 1$, $\alpha = 2.5$).

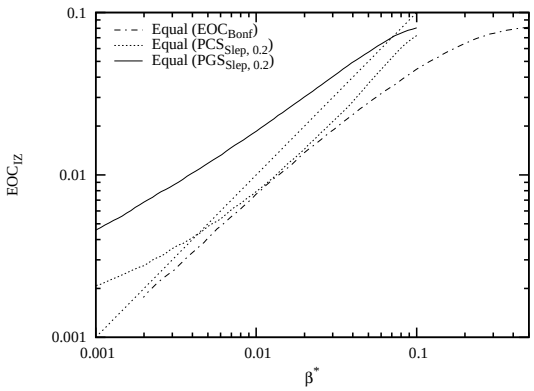


Figure 6: Target Plot for EOC (RPI, $k = 5$, $\eta = 1$, $\alpha = 2.5$).

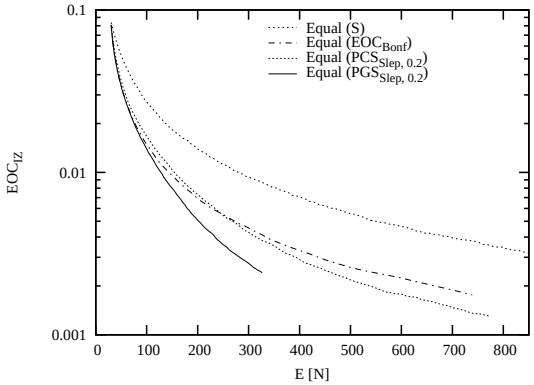


Figure 5: $\text{EOC}_{\text{iz}}$ Efficiency for Equal Allocation (RPI, $k = 5$, $\eta = 1$, $\alpha = 2.5$).

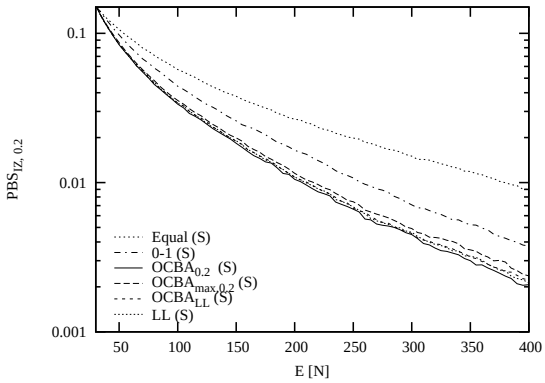


Figure 7: Different Allocations for Budget Stopping Rule (RPI, $k = 5$, $\eta = 1$, $\alpha = 100$).

for ordinal comparisons, those convergence results are typically for a fixed configuration. For RPI, we observed that the curvature was largely due to a long tail associated with a large number of samples for some very "hard" configurations (the means of the best two systems are very close, especially with large variances).

While the stopping rule has a very large influence on efficiency, the $\mathcal{LL}$, $\mathcal{OCBA}_{LL}$ and $\mathcal{OCBA}$ were more or less equivalent, with the first two usually being somewhat better, with 0-1 worse (it was derived with more approximations, and it is hard to improve PCS for two very close competitors in the RPI) and Equal worst. A typical plot is shown in Figure 7 for the $\mathcal{S}$ stopping rule (which may be needed if a simulation project has a strict time constraint).

Figure 8 compares three selection procedures with flexible stopping rules, Equal$(\text{PGS}_{Slep,\delta^*})$, $\mathcal{KN}{+}{+}$, and $\mathcal{OCBA}_{\delta^*}(\text{PGS}_{Slep,\delta^*})$ as representative for the Bayesian procedures. As is typical for the RPI problems tested, $\mathcal{OCBA}_{\delta^*}$ outperforms $\mathcal{KN}{+}{+}$ not only in terms of efficiency, but also with respect to meeting the target (Figure 9).

Figure 10 compares $\mathcal{OCBA}_{\delta^*}$ and $\mathcal{OCBA}_{max,\delta^*}$ with both the $\text{PGS}_{Slep,\delta^*}$ and $PCS_{Slep,\delta^*}$ stopping rules. The result is typical, namely that $\mathcal{OCBA}_{\delta^*}$ is the better allocation and $\text{PGS}_{Slep,\delta^*}$ is the better stopping rule.

We now turn to two implementation issues. Chen et al. (2005) wrote that the efficiency of $\mathcal{OCBA}(\mathcal{S})$ was not significantly different whether a $t$ or a normal distribution is used for $\text{EAPCS}_i$ (by substituting in the sample variance for the unknown actual variance into a normal distribution version of $\text{EAPCS}_i$), but did not publish results. Figure 11 confirms those claims and generalizes to other stopping rules. A normal distribution in the allocation is denoted $\mathcal{OCBA}_{Gaussian}$. On the other hand, using a normal distribution for the stopping rule ($\text{PCS}_{Slep,Gaussian}$) *does* degrade performance. The probable cause is that absolute values are important for stopping, but for allocation, relative values for different systems are compared.

A refined estimator of the degrees of freedom that gave good CI coverage for queueing experiments with small numbers of observations (Wilson and Pritsker 1984) didn't improve upon Welch's approximation for the SC in Fig-

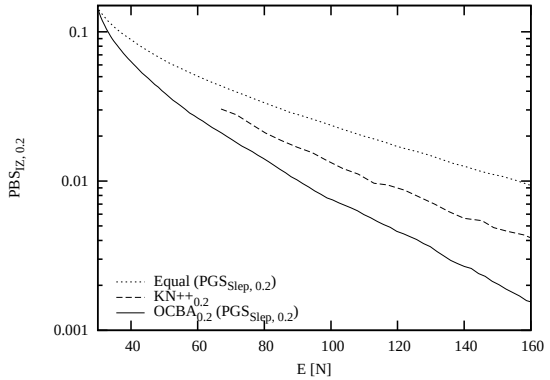


Figure 8: Efficiency of Flexible Procedures (RPI, $k = 5$, $\eta = 1$, $\alpha = 2.5$).

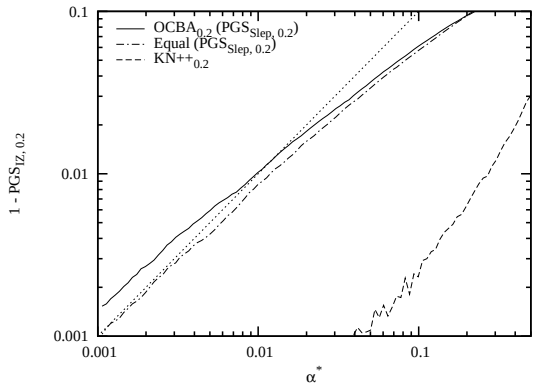


Figure 9: Target Plot for Flexible Procedures (RPI, $k = 5$, $\eta = 1$, $\alpha = 2.5$).

ure 12. The associated target plot gave a small (statistically significant) decrease in $\text{PCS}_{\text{iz}}$ for W&P relative to Welch.

## 5 DISCUSSION

For a fixed budget constraint on the number of samples, Procedures $\mathcal{LL}(\mathcal{S})$, $\mathcal{OCBA}_{LL}(\mathcal{S})$ and $\mathcal{OCBA}(\mathcal{S})$ were most efficient. Among flexible stopping rules, the $EOC_{Bonf}$ stopping rule was the most controllable for reaching a desired level of evidence for correct selection over a broad range of problems (for RPI the control was very precise), and were often the most efficient (for SC, MDM, RPI tested), especially with $\mathcal{LL}$, $OCBA_{LL}$ and $OCBA_{\delta^*}$. The $\text{PGS}_{Slep,\delta^*}$ stopping rule for RPI instances can be more efficient, but is not as controllable. $\mathcal{KN}{+}{+}$ was more efficient than the original OCBA and VIP proposals, but was less efficient than $\mathcal{LL}(\text{EOC}_{Bonf})$ and $\mathcal{OCBA}_{LL}(\text{EOC}_{Bonf})$, except for special configurations and mostly for low PICS. A concern for practical usage of $\mathcal{KN}{+}{+}$ is its sensitivity to the indifference zone parameter for efficiency and moreso for controlling $\text{PCS}_{\text{iz}}$ or $\text{PGS}_{\text{iz},\delta^*}$.

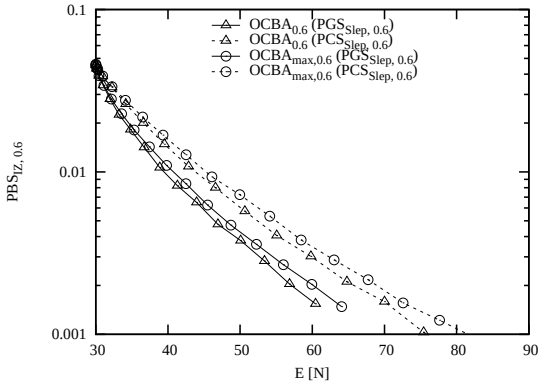


Figure 10: Different Ways to Use $\delta^*$ (RPI, $k = 5$, $\eta = 1$, $\alpha = 100$).

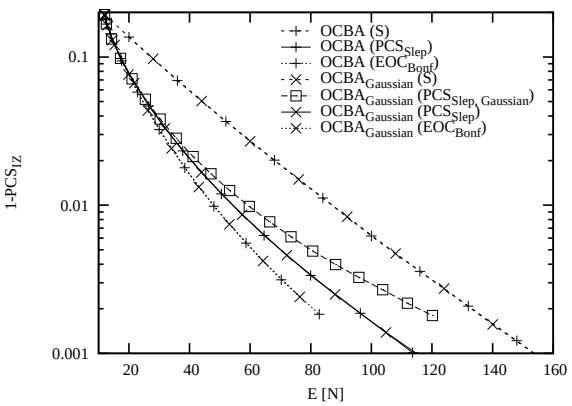


Figure 11: Allocation with Normal Approximation as Efficient as $t$, But Normal for Stopping Rule is Less Efficient (SC, $k = 2$, $\delta = 0.5$, $\rho = 1$).

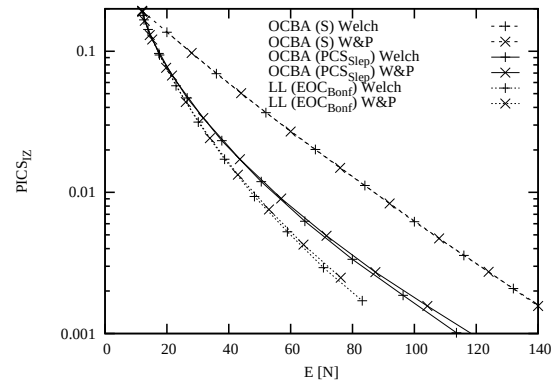


Figure 12: Wilson and Pritsker's (W&P) Degree of Freedom Correction was Not More Efficient Than Welch's (SC, $k = 2$, $\delta = 0.5$, $\rho = 1$).

We did not test the effect of autocorrelation from steady-state simulations, but do not see why batching would affect one procedure differently than another. Future work includes extensions to common random numbers (CRN) and integra-

tion of efficient selection procedures into optimization tools that handle combinatorially large numbers of alternatives.

## REFERENCES


Boesel, J., B. L. Nelson, and S.-H. Kim. 2003. Using ranking and selection to 'clean up' after simulation optimization. *Operations Research* 51:814–825.

Branke, J., S. E. Chick, and C. Schmidt. 2005. Selecting a selection procedure. *working paper*.

Branke, J., and C. Schmidt. 2004. Sequential sampling in noisy environments. In *Parallel Problem Solving from Nature*, ed. X. Yao et al., Volume 3242 of *LNCS*, 202–211: Springer.

Chen, C.-H. 1996. A lower bound for the correct subset-selection probability and its application to discrete event simulations. *IEEE Transactions on Automatic Control* 41 (8): 1227–1231.

Chen, C.-H., E. Yücesan, L. Dai, and H. Chen. 2005. Efficient computation of optimal budget allocation for discrete event simulation experiment. *IIE Transactions*:to appear.

Chen, E. J., and W. D. Kelton. 2005. Sequential selection procedures: Using sample means to improve efficiency. *European Journal of Operational Research* 166:133–153.

Chick, S. E., and K. Inoue. 2001. New two-stage and sequential procedures for selecting the best simulated system. *Operations Research* 49 (5): 732–743.

Chick, S. E., and K. Inoue. 2002. Corrigendum: New selection procedures. *Operations Research* 50 (3): 566.

Chick, S. E., and Y. Wu. 2005. Selection procedures with frequentist expected opportunity cost bounds. *Operations Research*:to appear.

Goldsman, D., S.-H. Kim, W. S. Marshall, and B. L. Nelson. 2002. Ranking and selection for steady-state simulation: Procedures and perspectives. *INFORMS Journal on Computing* 14 (1): 2–19.

Gupta, S. S., and K. J. Miescke. 1994. Bayesian look ahead one stage sampling allocations for selecting the largest normal mean. *Statistical Papers* 35:169–177.

He, D., S. E. Chick, and C.-H. Chen. 2005. The opportunity cost and OCBA selection procedures in ordinal optimization. *submitted*.

Inoue, K., S. E. Chick, and C.-H. Chen. 1999. An empirical evaluation of several methods to select the best system. *ACM TOMACS* 9 (4): 381–407.

Kelton, W. D., R. P. Sadowski, and D. A. Sadowski. 1998. *Simulation with Arena*. Boston: McGraw-Hill.

Kim, S.-H., and B. L. Nelson. 2001. A fully sequential procedure for indifference-zone selection in simulation. *ACM TOMACS* 11:251–273.

Kim, S.-H., and B. L. Nelson. 2005. Selecting the best system. In *Handbook in Operations Research and Management Science: Simulation*, ed. S. G. Henderson and B. L. Nelson. Elsevier.

Lerch, M., G. Tischler, J. W. von Gudenberg, W. Hofschuster, and W. Kraemer. 2001. The interval library filib++ 2.0 - design, features and sample programs. Preprint 2001/4, University of Wuppertal.

Matsumoto, M., and T. Nishimura. 1998. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM TOMACS* 8 (1): 3–30.

Nelson, B. L., and S. Banerjee. 2001. Selecting a good system: Procedures and inference. *IIE Transactions* 33 (3): 149–166.

Welch, B. L. 1938. The significance of the difference between two means when the population variances are unequal. *Biometrika* 25:350–362.

Wilson, J. R., and A. A. B. Pritsker. 1984. Experimental evaluation of variance reduction techniques for queueing simulation using generalized concomitant variables. *Management Science* 30:1459–1472.


## AUTHOR BIOGRAPHIES


**JÜRGEN BRANKE** is a Research Associate of the Institute for Applied Computer Science and Formal Description Methods (AIFB) at the University of Karlsruhe, Germany. His main research area is nature-inspired optimization, with a special focus on optimization in the presence of uncertainties, including noisy or dynamically changing environments. Other research interests include multi-objective optimization, parallelization, and agent-based modelling. His email address is <jbr@aifb.uni-karlsruhe.de>, and his web page is page is <www.aifb.uni-karlsruhe.de/∼jbr/>.

**STEPHEN E. CHICK** is an Associate Professor of Technology and Operations Management at INSEAD. He has worked in the automotive and software sectors prior to joining academia, and now teaches operations with applications in manufacturing and services, particularly the health care sector. He enjoys Bayesian statistics, stochastic models, and simulation. His email address is <stephen.chick@insead.edu>, and his web page is page is <faculty.insead.edu/chick/>.

**CHRISTIAN SCHMIDT** is a Research Assistent and PhD candidate of the Institute for Applied Computer Science and Formal Description Methods (AIFB) at the University of Karlsruhe, Germany. His main research interest is simulation-based optimization using nature-inspired heuristics. He has applied nature-inspired optimization techniques to a number of logistics problems as part of industry projects. His email address is <csc@aifb.uni-karlsruhe.de>.