

A TRANSIENT MEANS RANKING AND SELECTION PROCEDURE WITH SEQUENTIAL SAMPLING CONSTRAINTS

Douglas J. Morrice

Red McCombs School of Business
1 University Station, B6500
The University of Texas at Austin
Austin, TX 78712, USA

Mark W. Brantley
Chun-Hung Chen

SEOR Department
4400 University Drive, MS 4A6
George Mason University
Fairfax, VA 22030, USA

ABSTRACT

We develop a Ranking and Selection procedure for selecting the best configuration based on a transient mean performance measure. The procedure extends the OCBA approach to systems whose means are a function of some other variable such as time. In particular, we characterize this as a prediction problem and imbed a regression model in the OCBA procedure. In this paper, we analyze a problem with sequential sampling constraints for each configuration and offer a heuristic to use a polynomial regression model when variance reduction is possible.

1 INTRODUCTION

Morrice et al. (2008) develop a ranking and selection procedure for selecting the best configuration based on a transient mean performance measure. The procedure uses a regression-based approach to extend the Optimal Computing Budget Allocation (OCBA) method (Chen et al. 2000 and 2008) to systems whose means are changing according to a linear functional relationship.

In this paper, we extend the Morrice et al. (2008) procedure to address a special but important case with non-linear functional relationships and sequential sampling constraints. More specifically, we assume that transient mean is a function of a discrete index such as observation number or discretely sampled time. Due to the sequential sampling constraints inherent in such indices, the optimal sample location is not the focus of this approach. Instead, we focus on developing a procedure that utilizes all possible simulation observations to reduce the variance of the each transient mean evaluated at a particular index value.

For example, consider a simulated queuing configuration that is initialized as empty and idle. Suppose that one is interested in the waiting time for the M -th customer, $W(M)$, before the system has achieved steady state. A naïve way to estimate the expected value of $W(M)$ is to do replications of the simulation and simply calculate the sample average of $W(M)$ across replications. A more sophisticated way is to fit a model to the transient observations generated by simulation prior to observation n and predict the expected value of $W(M)$. By using as many observations at possible, one makes more efficient use of information or, equivalently, reduces the variance of the statistics used to estimate the transient mean performance measure. Figure 1 illustrates the idea for a generic performance measure and for multiple configurations.

We take a regression approach to fitting the transient. The approach is complicated by non-linear behavior, lack of normality, and statistical dependency in the data. To address the issue of non-linearity, we use a polynomial regression approach (Draper and Smith, 1981, Section 5.1). We deal with non-normality by making macro-replications (Goldsmann et al. 1991) and tackle statistical dependency by using simulation observations spaced far enough apart to in the data series to effectively eliminate serial correlation.

The resultant OCBA procedure incorporates the regression results for configurations in which a variance reduction is possible. Regular OCBA and a naïve approach is used for configurations in which a variance reduction is not possible in order to save the regression analysis overhead.

The rest of the paper is organized in the following manner. Section 2 provides the problem statement and Section 3 covers the details of the regression analysis. In Section 4, we provide a heuristic procedure for the extended OCBA formulation. Section 5 contains an example to illustrate the variance reduction potential of this approach. Section 6 provides our conclusions and future research directions.

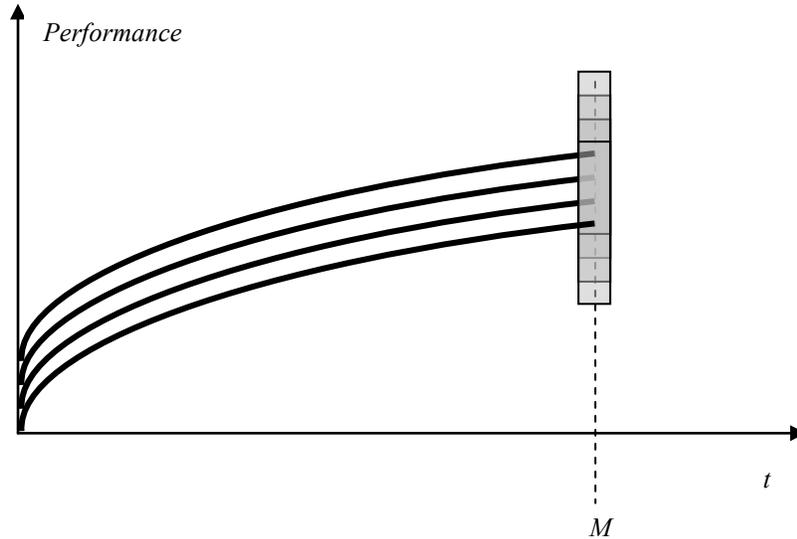


Figure 1: The mean performance of multiple configurations up to time M

2 PROBLEM STATEMENT

Let x represent the observation index and y represent the expected performance measure of interest. The variable x can assume the values $1, 2, \dots, M$. This paper explores a problem with the principle goal of selecting the best of K alternative configurations at index value M . Without loss of generality, we consider the minimization problem shown below where the “best” configuration is the one with smallest expected performance measure, i.e.,

$$\min_j y_j(M); j = 1, \dots, K .$$

The expected performance measure $y_j(x)$ must be estimated via simulation with noise where the simulation output $f_j(x_i)$ is defined as

$$f_j(x) = y_j(x) + \varepsilon; \quad i = 1, \dots, K, \text{ where } \varepsilon \sim N(0, \sigma_j^2) . \tag{1}$$

Note: we assume that the error is normal and independent for the development of the theory. The heuristic procedure in Section 4 accounts for non-normal and statistically dependent data.

In this paper, we assume that the expectation of the unknown underlying function for each configuration is polynomial (or can be closely approximated by a polynomial) on the prescribed domain, i.e.,

$$y_j(x) = \beta_{0j} + \beta_{1j}x + \beta_{2j}x^2 + \dots + \beta_{nj}x^n . \tag{2}$$

Our problem is to select the configuration associated with the smallest mean performance measure from among the K configurations within the constraint of a computing budget with only T simulation replications. We designate the configuration with the smallest estimated mean performance measure at M as $\hat{y}_b(M)$ so that $b = \arg \min_j \hat{y}_j(M)$. Given the uncertainty of the estimate of the underlying function, b is a random variable and we define Correct Selection as the event where b is indeed the best configuration.

We define N_j as the number of simulation replications conducted for configuration j . Since the simulation is expensive and the computing budget is restricted, we seek to develop an allocation rule for each N_j in order to provide as much information as possible for the identification of the best configuration. Our goal then is to determine the optimal allocations to the configurations that maximize the probability that we correctly select the best configuration (PCS). This Optimal Computing Budget Allocation (OCBA) problem is reflected in (3) below.

$$\begin{aligned} \max_{N_j} \quad & PCS = P\{y_b(M) \leq y_j(M) \forall j\} \\ \text{s.t.} \quad & \sum_{j=1}^K N_j = T \end{aligned} \tag{3}$$

The constraint $\sum_{j=1}^K N_j = T$ denotes the total computational cost and implicitly assumes that the simulation execution times for one sample are constant across the configurations and that each simulation replication is of length M .

The OCBA problem statement in (3) simply represents standard OCBA if

$$\hat{y}_j(M) = \frac{\sum_{l=1}^{N_j} f_{jl}(M)}{N_j}$$

where $f_{jl}(M)$ is the simulation output for replication l of configuration j ($j=1, 2, \dots, K$) at index value M . However, if (2) holds, then least squares regression can be used to estimate $y_j(M)$. The additional overhead associated with a regression analysis is justified only if it results in a substantial variance reduction in the statistic $\hat{y}_j(M)$.

3 REGRESSION ANALYSIS

For configuration j , let β_j represent an $(n+1) \times 1$ vector of parameters from (2), let \mathbf{X}_j be an $M \times (n+1)$ matrix with rows of the form $[1 \ x \ \dots \ x^n]$ for each index value, and \mathbf{F}_j be an $M \times 1$ vector with a simulation output observation (from (1)) corresponding to each row in \mathbf{X}_j . Using the Bayesian regression framework detailed in Morrice et al (2008) with a non-informative prior, the posterior distribution of β_j is given by

$$\tilde{\beta}_j \sim N[(\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_j^T \mathbf{F}_j, \sigma_j^2 (\mathbf{X}_j^T \mathbf{X}_j)^{-1}] \tag{4}$$

The non-informative prior distribution represents the fact that we have no information about the unknown parameters and places emphasis upon the information we collect with our simulation runs.

Since $\tilde{y}_j(M)$ is a linear combination of the elements $\tilde{\beta}_j$, this means that $\tilde{y}_j(M)$ has a Gaussian distribution of the form

$$\tilde{y}_j(M) \sim N[\mathbf{X}_M (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_j^T \mathbf{F}_j, \sigma_j^2 \mathbf{X}_M^T (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_M] \tag{5}$$

where $\mathbf{X}_M^T = [1 \ M \ \dots \ M^n]$. After the simulation is performed, the mean and variance of the performance measure for configuration j are estimated by $\mathbf{X}_M (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_j^T \mathbf{F}_j$ and $s_j^2 \mathbf{X}_M^T (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_M$, respectively, where s_j^2 is the mean square residual from the regression analysis.

Proposition 1 *Let*

$$S_p(M) = \sum_{i=1}^p \sum_{j=0}^{i-1} (-1)^j (i-j)^p \binom{M+p-i+1}{M-i} \binom{p+1}{j} \tag{6}$$

for $p = 1, 2, \dots, 2n$. For expression (2), each $\mathbf{X}_j^T \mathbf{X}_j$ has the form

$$\begin{bmatrix} M & S_1(M) & S_2(M) & \cdots & S_n(M) \\ S_1(M) & S_2(M) & S_3(M) & \cdots & S_{n+1}(M) \\ S_2(M) & S_3(M) & S_4(M) & \cdots & S_{n+2}(M) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ S_n(M) & S_{n+1}(M) & S_{n+2}(M) & \cdots & S_{2n}(M) \end{bmatrix} \quad (7)$$

Proof Since x assumes the values $1, 2, \dots, M$, then

$$\mathbf{X}_j = \begin{bmatrix} 1 & 1 & 1^2 & \cdots & 1^n \\ 1 & 2 & 2^2 & \cdots & 2^n \\ 1 & 3 & 3^2 & \cdots & 3^n \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & M & M^2 & \cdots & M^n \end{bmatrix} \quad (8)$$

Hence, each element of $\mathbf{X}_j^T \mathbf{X}_j$ is a power sum $S_p(M) = \sum_{k=1}^M k^p$ which can be rewritten as (6) (Wolfram MathWorld 2009).

Note The $\mathbf{X}_j^T \mathbf{X}_j$ matrix has elements which are power sums regardless of whether or not all terms in the polynomial in (2) are included in the final regression model. Hence, a similar result to Proposition 1 holds for all such models.

Using Proposition 1 and *Mathematica* (Wolfram 1999), we can derive expressions for the prediction variance factor $\mathbf{X}_M^T (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_M$ as a function of M . Tables 1 and 2 contain results for all models up to order $n = 5$. Table 1 describes the 31 possible models and assigns them a model number. The latter is used to denote the model in Tables 2 and 3.

To gain a better understanding of the expressions in Tables 2 and 3, we plot $\mathbf{X}_M^T (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_M$ versus M in Figures 1 and 2 for the models of the form $\mathbf{y}(\mathbf{x}) = \beta_0 + \sum_{i=1}^n \beta_i \mathbf{x}^i$ and the form $\mathbf{y}(\mathbf{x}) = \beta_0 + \beta_n \mathbf{x}^n$ respectively. These models illustrate general principles common to all 31 models. In particular,

1. All 31 models have a similar shape to those in Figures 1 and 2, decreasing monotonically from the value one as M increases. In other words, for large enough M , the prediction variance factor provides a variance reduction.
2. Figure 1 illustrates that as more terms are added to the polynomial regression model, $\mathbf{X}_M^T (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_M$ increases for any given value of M or, equivalently, the variance reduction diminishes as more terms are added to the regression model.
3. Figure 2 illustrates that for an equivalent number of terms in the model (in this case, one term in addition to the constant), $\mathbf{X}_M^T (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_M$ increases with the degree of the polynomial for any given value of M .

Items (2) and (3) support the principle of parsimony which will be used in the heuristic that we develop in Section 4.

Proposition 2 Suppose R replications (or macro-replications, if necessary) are made. Then the prediction variance factor equals $\frac{\mathbf{X}_M^T (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_M}{R}$.

Proof With R replications, each column of the matrix \mathbf{X}_j is just the columns in (8) repeated R times. Therefore each term in $\mathbf{X}_j^T \mathbf{X}_j$ is multiplied by R and the result follows from the inversion of the matrix.

Proposition 2 implies that the regression approach gets the same variance reduction as the naïve approach simply by using replication (i.e., the variance of the prediction decreases by $\frac{1}{R}$). Hence, the variance reduction from indicated by the prediction variance factor $\mathbf{X}_M^T (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_M$ when the regression approach is used is over and above what is possible when the naïve approach is used.

Table 1: All possible polynomial regression models up to order $n = 5$

Model #	Model
1	$\beta_{0j} + \beta_{1j}x$
2	$\beta_{0j} + \beta_{2j}x^2$
3	$\beta_{0j} + \beta_{1j}x + \beta_{2j}x^2$
4	$\beta_{0j} + \beta_{3j}x^3$
5	$\beta_{0j} + \beta_{1j}x + \beta_{3j}x^3$
6	$\beta_{0j} + \beta_{2j}x^2 + \beta_{3j}x^3$
7	$\beta_{0j} + \beta_{1j}x + \beta_{2j}x^2 + \beta_{3j}x^3$
8	$\beta_{0j} + \beta_{4j}x^4$
9	$\beta_{0j} + \beta_{1j}x + \beta_{4j}x^4$
10	$\beta_{0j} + \beta_{2j}x^2 + \beta_{4j}x^4$
11	$\beta_{0j} + \beta_{3j}x^3 + \beta_{4j}x^4$
12	$\beta_{0j} + \beta_{1j}x + \beta_{2j}x^2 + \beta_{4j}x^4$
13	$\beta_{0j} + \beta_{1j}x + \beta_{3j}x^3 + \beta_{4j}x^4$
14	$\beta_{0j} + \beta_{2j}x^2 + \beta_{3j}x^3 + \beta_{4j}x^4$
15	$\beta_{0j} + \beta_{1j}x + \beta_{2j}x^2 + \beta_{3j}x^3 + \beta_{4j}x^4$
16	$\beta_{0j} + \beta_{5j}x^5$
17	$\beta_{0j} + \beta_{1j}x + \beta_{5j}x^5$
18	$\beta_{0j} + \beta_{2j}x^2 + \beta_{5j}x^5$
19	$\beta_{0j} + \beta_{3j}x^3 + \beta_{5j}x^5$
20	$\beta_{0j} + \beta_{4j}x^4 + \beta_{5j}x^5$
21	$\beta_{0j} + \beta_{1j}x + \beta_{2j}x^2 + \beta_{5j}x^5$
22	$\beta_{0j} + \beta_{1j}x + \beta_{3j}x^3 + \beta_{5j}x^5$
23	$\beta_{0j} + \beta_{1j}x + \beta_{4j}x^4 + \beta_{5j}x^5$
24	$\beta_{0j} + \beta_{2j}x^2 + \beta_{3j}x^3 + \beta_{5j}x^5$
25	$\beta_{0j} + \beta_{2j}x^2 + \beta_{4j}x^4 + \beta_{5j}x^5$
26	$\beta_{0j} + \beta_{3j}x^3 + \beta_{4j}x^4 + \beta_{5j}x^5$
27	$\beta_{0j} + \beta_{1j}x + \beta_{2j}x^2 + \beta_{3j}x^3 + \beta_{5j}x^5$
28	$\beta_{0j} + \beta_{1j}x + \beta_{2j}x^2 + \beta_{4j}x^4 + \beta_{5j}x^5$
29	$\beta_{0j} + \beta_{1j}x + \beta_{3j}x^3 + \beta_{4j}x^4 + \beta_{5j}x^5$
30	$\beta_{0j} + \beta_{2j}x^2 + \beta_{3j}x^3 + \beta_{4j}x^4 + \beta_{5j}x^5$
31	$\beta_{0j} + \beta_{1j}x + \beta_{2j}x^2 + \beta_{3j}x^3 + \beta_{4j}x^4 + \beta_{5j}x^5$

Table 2: Expressions for $\mathbf{X}_M^T (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_M$ for all possible polynomial regression models up to order $n = 4$

Model #	$\mathbf{X}_M^T (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_M$
1	$\frac{-2+4M}{M(1+M)}$
2	$\frac{6(1+M+M^2+16M^3)}{M(11+8M)(1+3M+2M^2)}$
3	$\frac{3(2-3M+3M^2)}{M(2+3M+M^2)}$
4	$\frac{8(-1-M+6M^2+6M^3+6M^4+27M^5)}{M(1+M)(-8+69M^2+84M^3+27M^4)}$
5	$\frac{2(-12+32M+102M^2-23M^3+99M^4)}{M(1+M)(13+9M)(2+5M+2M^2)}$
6	$\frac{6(2+3M+9M^2+42M^3+32M^4+15M^5+13M^6)}{M(1+M)(2+M)(11+30M+42M^2+27M^3+6M^4)}$
7	$\frac{8(-3+7M-3M^2+2M^3)}{M(1+M)(2+M)(3+M)}$
8	$\frac{10(3+3M-17M^2-17M^3+19M^4+19M^5+64M^6)}{M(1+3M+2M^2)(31-62M-65M^2+100M^3+119M^4+32M^5)}$
9	$\frac{30-45M-179M^2+264M^3+304M^4+24M^5+208M^6}{M(1+M)(2+M)(-7-6M+44M^2+16M^3)}$
10	$\frac{30(36+54M+173M^2+1335M^3+852M^4+348M^5+256M^6)}{M(1+3M+2M^2)(6+7M+2M^2)(411+479M+128M^2)}$
11	$\frac{120+180M-260M^2-480M^3-530M^4-3075M^5-1037M^6-3462M^7+1872M^8+777M^9+459M^{10}}{M(1+M)(2+M)(62-262M-480M^2-167M^3+402M^4+474M^5+192M^6+27M^7)}$
12	$\frac{6(60-70M+27M^2+862M^3-123M^4+192M^5)}{M(1+M)(2+M)(3+M)(1+2M)(47+32M)}$
13	$\frac{10(-12+28M+58M^2+83M^3+133M^4+123M^5+13M^6+18M^7)}{M(1+M)(2+M)(3+M)(21+50M+65M^2+40M^3+9M^4)}$
14	$\frac{6(240+440M+1410M^2+9535M^3+9270M^4+8255M^5+5901M^6+2306M^7+531M^8+176M^9)}{M(1+M)(2+M)(3+M)(1+2M)(548+1004M+891M^2+528M^3+177M^4+24M^5)}$
15	$\frac{5(24-50M+55M^2-10M^3+5M^4)}{M(1+M)(2+M)(3+M)(4+M)}$

Table 3: Expressions for $\mathbf{X}_M^T (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_M$ for all possible polynomial regression models of order $n = 5$

Model #	$\mathbf{X}_M^T (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_M$
16	$\frac{24 (-5 - 5M + 28M^2 + 28M^3 - 38M^4 - 17M^5 + 17M^6 + 17M^7 + 17M^8 + 50M^9)}{M(1+M)(-120 + 683M^2 - 1011M^3 - 132M^4 + 804M^5 + 528M^6 + 100M^7)}$
17	$\frac{6 (468 - 768M - 1146M^2 + 1017M^3 + 1003M^4 + 156M^5 + 500M^6)}{M(1+M)(6 + 7M + 2M^2)(-267 + 101M + 312M^2 + 100M^3)}$
18	$\frac{3 (40 + 60M - 44M^2 + 504M^3 + 2M^4 - 3609M^5 - 1719M^6 + 2466M^7 + 1456M^8 + 627M^9 + 425M^{10})}{M(1+M)(2+M)(110 + 300M - 214M^2 - 1440M^3 - 1087M^4 + 666M^5 + 1178M^6 + 516M^7 + 75M^8)}$
19	$\frac{2 (-864 - 1296M + 9744M^2 + 15264M^3 - 10892M^4 + 23280M^5 - 29170M^6 - 278415M^7 - 119325M^8 + 98550M^9 + 51800M^{10} + 21225M^{11} + 11875M^{12})}{M(1+M)(2+5M+2M^2)(-864 + 1728M + 11100M^2 - 2040M^3 - 30935M^4 - 23975M^5 + 4450M^6 + 11950M^7 + 4825M^8 + 625M^9)}$
20	$\frac{6 (150 + 225M - 241M^2 - 474M^3 + 376M^4 + 801M^5 + 495M^6 + 2592M^7 + 32M^8 - 5928M^9 - 2552M^{10} + 1056M^{11} + 536M^{12} + 216M^{13} + 112M^{14})}{M(1+M)(2+M)(465 - 1206M^2 - 450M^3 + 3362M^4 + 3834M^5 - 2192M^6 - 5721M^7 - 2640M^8 + 660M^9 + 952M^{10} + 300M^{11} + 32M^{12})}$
21	$\frac{6 (-156 + 204M + 354M^2 - 1025M^3 + 801M^4 + 1351M^5 - 39M^6 + 250M^7)}{M(1+M)(2+M)(3+M)(-41 - 66M + 203M^2 + 264M^3 + 75M^4)}$
22	$\frac{2 (-540 + 1656M + 7026M^2 + 3785M^3 + 14250M^4 + 9119M^5 + 1224M^6 + 1100M^7)}{M(1+M)(3+M)(3+2M)(2+5M+2M^2)(87 + 97M + 25M^2)}$
23	$\frac{24 (-585 + 765M + 1983M^2 - 1098M^3 - 3204M^4 + 626M^5 - 180M^6 + 102M^7 + 5424M^8 + 2544M^9 + 432M^{10} + 256M^{11})}{M(1+M)(2+M)(3+M)(1257 + 672M - 1612M^2 - 3960M^3 - 1068M^4 + 4608M^5 + 4992M^6 + 1920M^7 + 256M^8)}$
24	$\frac{24 (-288 - 528M - 138M^2 - 4093M^3 + 540M^4 + 27590M^5 + 30690M^6 + 30525M^7 + 25275M^8 + 9025M^9 + 2025M^{10} + 625M^{11})}{M(1+M)(2+M)(3+M)(-2112 - 5760M + 1880M^2 + 22440M^3 + 39825M^4 + 38100M^5 + 20550M^6 + 5700M^7 + 625M^8)}$
25	$\frac{6 (1200 + 2200M + 6522M^2 + 46707M^3 + 42162M^4 + 6757M^5 - 1536M^6 + 29584M^7 + 31404M^8 + 44734M^9 + 42057M^{10} + 13642M^{11} + 2967M^{12} + 832M^{13})}{M(1+M)(2+M)(3+M)(2740 + 5020M + 3019M^2 - 1418M^3 - 1496M^4 + 3766M^5 + 6148M^6 + 3604M^7 + 957M^8 + 96M^9)}$
26	$\frac{40 (-864 - 1584M + 5346M^2 + 11241M^3 + 19032M^4 + 71792M^5 + 5602M^6 + 11537M^7 + 714M^8 + 2009M^9 + 1434M^{10} + 4879M^{11} + 4635M^{12} + 1360M^{13} + 285M^{14} + 70M^{15})}{M(1+M)(2+M)(3+M)(-5952 + 69880M^2 + 129240M^3 + 91477M^4 + 19200M^5 - 14410M^6 - 6420M^7 + 10265M^8 + 12300M^9 + 5600M^{10} + 1200M^{11} + 100M^{12})}$
27	$\frac{6 (-240 + 676M + 1300M^2 - 910M^3 + 2915M^4 - 246M^5 + 225M^6)}{M(1+M)(2+M)(3+M)(4+M)(1+2M)(37 + 25M)}$
28	$\frac{3 (1560 - 1650M + 6327M^2 + 9206M^3 + 12461M^4 + 5720M^5 + 7688M^6 + 104M^7 + 464M^8)}{M(1+M)(2+M)(3+M)(4+M)(139 + 294M + 356M^2 + 210M^3 + 48M^4)}$
29	$\frac{2 (-3600 + 10140M + 45516M^2 + 39950M^3 + 119045M^4 + 86790M^5 + 62463M^6 + 27600M^7 + 11880M^8 + 1120M^9 + 496M^{10})}{M(1+M)(2+M)(3+M)(4+M)(1+2M)(425 + 725M + 608M^2 + 344M^3 + 112M^4 + 16M^5)}$
30	$\frac{15 (2304 + 4800M + 20160M^2 + 106520M^3 + 143444M^4 + 148950M^5 + 138185M^6 + 69810M^7 + 29961M^8 + 9950M^9 + 2275M^{10} + 310M^{11} + 55M^{12})}{M(1+M)(2+M)(3+M)(4+M)(3288 + 12600M + 23360M^2 + 23520M^3 + 14095M^4 + 5700M^5 + 1650M^6 + 300M^7 + 25M^8)}$
31	$\frac{6 (-120 + 314M - 225M^2 + 160M^3 - 15M^4 + 6M^5)}{M(1+M)(2+M)(3+M)(4+M)(5+M)}$

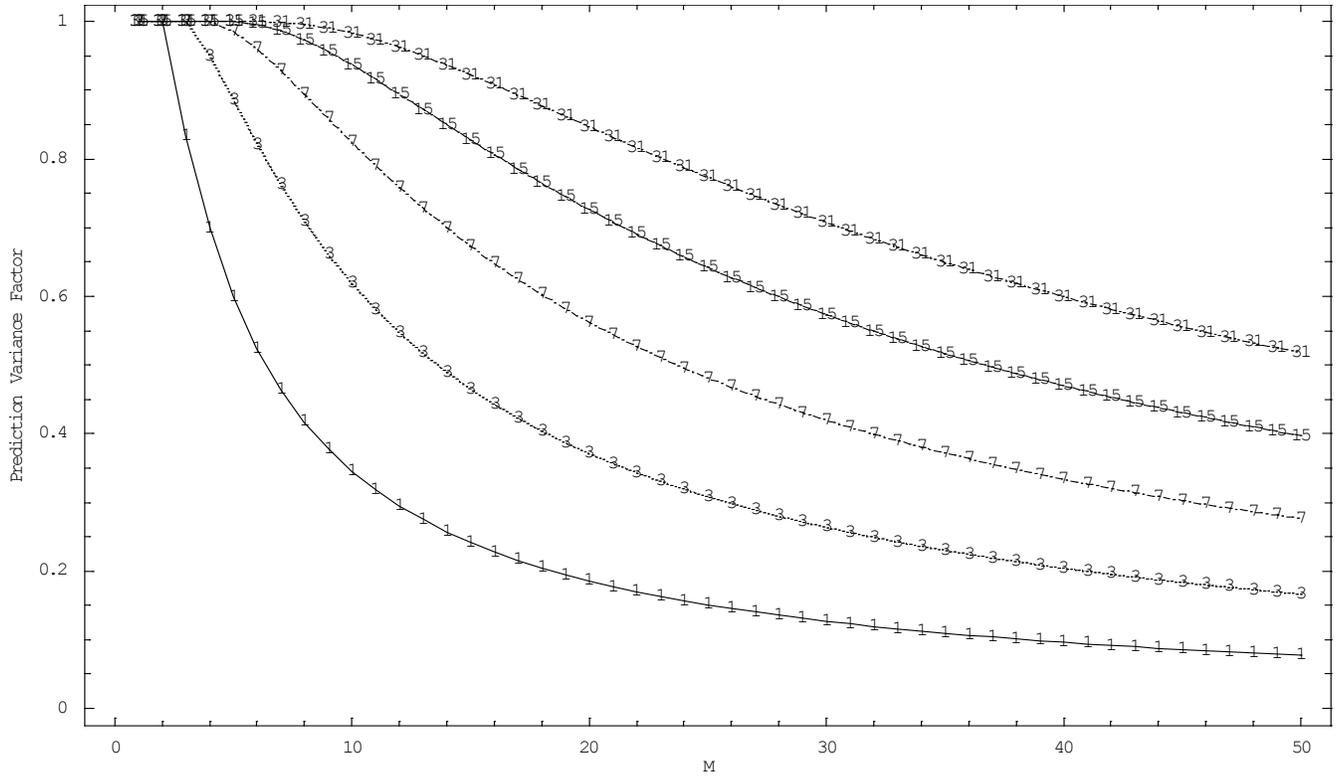


Figure 1: Prediction Variance Factor (i.e., $\mathbf{X}_M^T (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_M$) versus M for Models 1, 3, 7, 15, and 31

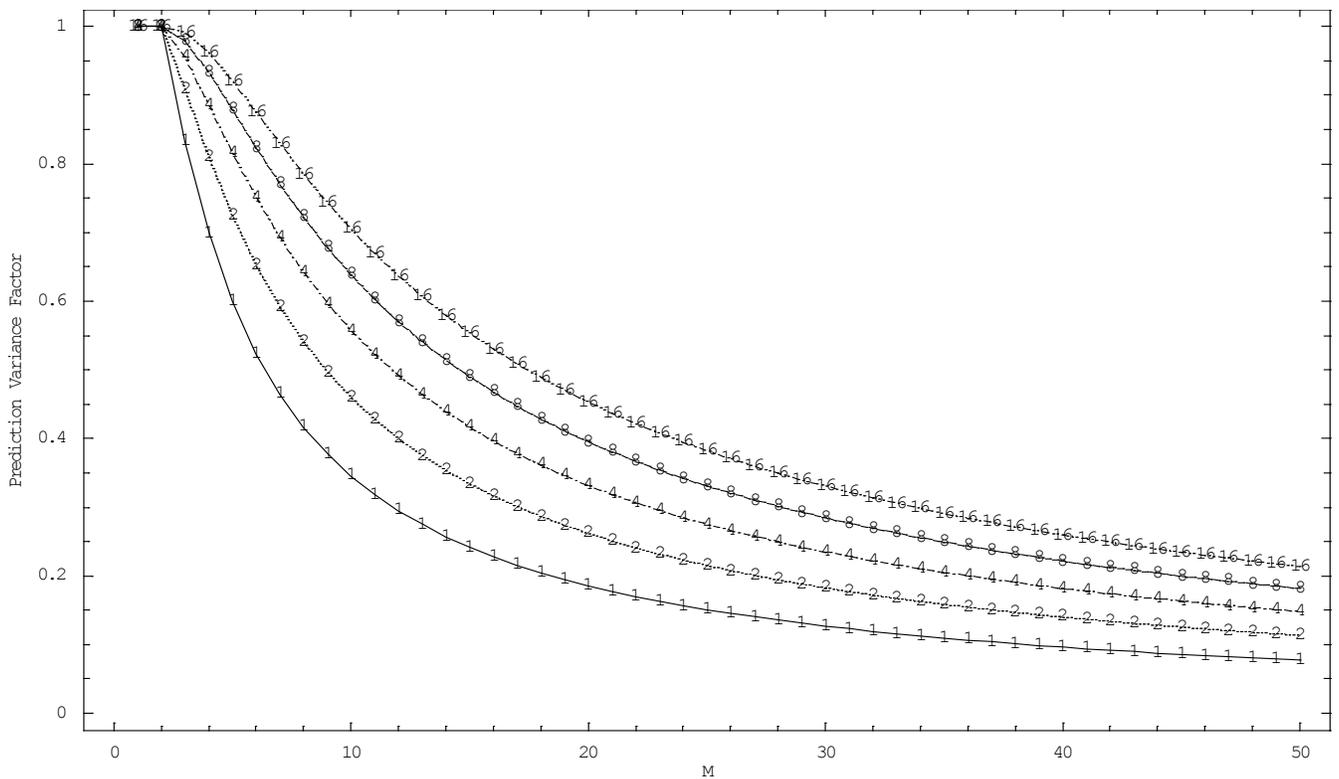


Figure 2: Prediction Variance Factor (i.e., $\mathbf{X}_M^T (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_M$) versus M for Models 1, 2, 4, 8, 16

4 AN HEURISTIC PROCEDURE

The heuristic has two phases. Phase 1 establishes whether or not the regression approach provides variance reduction for each configuration. For any configuration on which regression does not provide variance reduction, it is rejected in favor of the naïve approach in order to save the overhead associated with the regression analysis.

Phase 2 conducts the regular OCBA analysis using either the naïve approach results or the regression results for a given configuration depending on the outcome of Phase 1.

Heuristic

Phase 1:

For each configuration

- 1) Generate N_0 initial replications
- 2) Average each observation across replications
- 3) Fit a regression model to the replication averages
- 4) Check for normality of the residuals
- 5) If residuals fail normality test, make an additional replication and go to 2) else
- 6) Check for serial correlation in the residuals
- 7) If the residuals fail serial correlation test, starting at last observation, discard the number of spacing observations indicated by the largest significant lag and go to step 6) else
- 8) Index the remaining observations $1, 2, \dots, m$ where $m \leq M$
- 9) Fit a regression model to the m observations
- 10) If there is variance reduction based on the number of observations and the regression model from step 9), use regression modeling approach, else use naïve approach

End for.

Phase 2:

Conduct regular OCBA (Chen et al 2000) to determine the number of macro-replications for all configurations. For the configurations in which the regression modeling approach is used, the form of regression model and observation spacing identified in Phase 1 are used. For multiple macro-replications, use Proposition 2 when regression is used.

In step 7 of Phase 1, we make a phi-mixing assumption (Billingsley 1968, p.166) that the serial correlation decreases as the spacing between observations in the data series increases. By the construction of Step 7, Step 8 represents a simple shift and scaling of the index of the remaining observations which does not change the analysis of variance results for the regression model in (2). It enables us to use the results in Propositions 1 and 2.

5 AN EXAMPLE

Since our procedure is a direct extension of regular OCBA, this example is designed to illustrate how the procedure is implemented and the *potential* for variance reduction when the regression approach is used. Additional analysis and testing to rigorously establish variance reduction remains the topic of future research. We simulate an M/M/1 queue with traffic intensity 0.8 for the first 100 customers waiting times when the model is initialized empty and idle. Hence, the average waiting times as a function of customer number form an initial transient as the model “warms up” (Kelton and Law 1985, Law and Kelton, 2000). We chose this example because we can compare our results to the analytical results from Kelton and Law (1985) which show that the expected delay for the 100th customer in this queuing model is 3.81 to two decimal places of accuracy.

We run through Phase 1 of the Heuristic by setting $N_0 = 20$ and averaging the observations across the 20 replications for each customer. Using the add-in *StatTools 5.0 for Excel* (Palisade Corp. 2008) to do the statistical analysis, yields a regression model with linear, quadratic, cubic, and fifth order terms (adjusted $R^2 = 0.8987$) when both forward and stepwise regression procedures are used. A Chi-Square test of the residuals for normality does not reject the null hypothesis of normality at the five percent level.

A serial correlation test of the residuals detects significant lag one and lag two correlations. Starting at observation 100 and working back through the data series, two observations are thinned from the data set between each remaining data point (hence the remaining data points are 100, 97, 94, ..., 7, 4, 1). Forward and stepwise regression on this new thinned data set produces a full quadratic model (adjusted $R^2 = 0.8527$). The resultant serial correlation test on the residuals identifies a significant lag one correlation. Hence, thinning is conducted again leaving the following 17 data points 100, 94, 88, ..., 16, 10, 4

in the data set. As on the last step, forward and stepwise regression yields a full quadratic model (adjusted $R^2 = 0.8111$). The residuals from this regression show no significant correlation.

The observation indices are shifted and rescaled as 1, 2, ..., 17. The regression yields the estimates: $\mathbf{X}_M(\mathbf{X}_j^T\mathbf{X}_j)^{-1}\mathbf{X}_j^T\mathbf{F}_j = 3.94$ and $s_j^2 = 0.284$. From Proposition 1 and Table 2 (model 7), $\mathbf{X}_M^T(\mathbf{X}_j^T\mathbf{X}_j)^{-1}\mathbf{X}_M = 0.422$. Hence, the prediction variance is 0.120 and the prediction standard deviation is 0.346.

Employing the naïve approach of simply averaging the waiting times across the 20 replications for customer 100 yields a prediction average is 3.671 with standard deviation 0.720. Hence, the regression approach yields a standard deviation that is less than half of the standard deviation from the naïve approach.

6 CONCLUSIONS

We have presented a Ranking and Selection procedure for selecting the best configuration based on a nonlinear transient mean performance measure with sequential sampling constraints. We have offered a heuristic to use results from a polynomial regression model when variance reduction is possible. There are certainly ways to improve upon these results. Most notably, an examination of Tables 2 and 3 show that there is some structure to the results and that it may be possible to derive a general form for the results.

REFERENCES

- Billingsley, P. 1968. *Convergence of Probability Measures*. New York: John Wiley.
- Chen, C. H., J. Lin, E. Yücesan, S. E. Chick. 2000. Simulation Budget Allocation for Further Enhancing the Efficiency of Ordinal Optimization. *Journal of DEEDS* 10(3): 251-270.
- Chen, C. H., D. He, M. C. Fu, and L. H. Lee. 2008. Efficient simulation budget allocation for selecting an optimal subset. *INFORMS Journal on Computing* 20(4): 579-595.
- Draper, N. R. and H. Smith. 1981. *Applied Regression Analysis*. 2nd ed. New York: John Wiley & Sons.
- Goldsman, D. M., B. L. Nelson, B. W. Schmeiser. 1991. Methods for Selecting the Best System. In *Proceedings of the 1991 Winter Simulation Conference*, ed. B. L. Nelson, W. D. Kelton, and G. M. Clark, 177-186. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.
- Kelton, W. D., and A. M. Law. 1985. Transient Behavior of the M/M/s Queue. *Operations Research* 33(2): 378-396.
- Law, A. M., and W. D. Kelton. 2000. *Simulation Modeling and Analysis*. 3rd ed. New York: McGraw-Hill, Inc.
- Morrice, D. J., M. W. Brantley, C. H. Chen. 2008. An Efficient Ranking and Selection Procedure for a Linear Transient Mean Performance Measure. In *Proceedings of the 2008 Winter Simulation Conference*, ed. S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, 290-296. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.
- Palisade Corporation. 2008. *Guide to Using StatTools, Statistics Add-in for Microsoft Excel*. Version 5. Ithaca, New York: Palisade Corporation.
- Wolfram MathWorld 2009. Available via www.mathworld.wolfram.com/PowerSum.html [accessed May 14, 2009].
- Wolfram, S. 1999. *The Mathematica Book*. Cambridge: Wolfram Media.

AUTHOR BIOGRAPHIES

DOUGLAS J. MORRICE is a Professor in Operations Management at The University of Texas at Austin. He is also Chair of the Information, Risk, and Operations Management Department, and Director of the University of Texas Supply Chain Management Center of Excellence. He has an ORIE Ph.D. from Cornell University. His research interests include simulation design, modeling, and analysis, and supply chain risk management. Dr. Morrice was Co-Editor of the *Proceedings of the 1996 Winter Simulation Conference*, and 2003 Winter Simulation Conference Program Chair. He is currently serving as a representative for the INFORMS Simulation Society on the Winter Simulation Conference Board of Directors. His email address is morrice@mail.utexas.edu.

MARK W. BRANTLEY is a doctoral candidate in the Department of Systems Engineering and Operations Research at George Mason University. He received his BS degree in Mathematical Sciences at the United States Military Academy and

received MS degrees in Applied Mathematics and Operations Research from Rensselaer Polytechnic Institute. His research interests include simulation and optimization. His e-mail address is mbrant11@gmu.edu.

CHUN-HUNG CHEN is a Professor of Systems Engineering and Operations Research at George Mason University. He received his Ph.D. from Harvard University in 1994. His research interests are mainly in development of very efficient methodology for simulation and optimization and its applications. Dr. Chen has served as Co-Editor of the Proceedings of the 2002 Winter Simulation Conference and Program Co-Chair for 2007 Informs Simulation Society Workshop. He is currently an associate editor of IEEE Transactions on Automatic Control, area editor of Journal of Simulation Modeling Practice and Theory, associate editor of International Journal of Simulation and Process Modeling, and simulation department editor for IIE Transactions. His email address is cchen9@gmu.edu.