Proceedings of the 1997 Winter Simulation Conference
ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson

THE IMPACT OF TRANSIENTS ON SIMULATION VARIANCE ESTIMATORS

Daniel H. Ockerman

Retek Information Systems
7 Piedmont Center, Suite 501
Atlanta, GA 30305, U.S.A.

David Goldsman

School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332, U.S.A.

ABSTRACT

Given a stationary simulation process with unknown mean m, interest frequently lies in, and various methods exist for, developing estimates and confidence intervals for m. Typically, the sample mean is used as the point estimate for m. It is also useful to estimate the variance parameter, s^2, a measure of sample mean's precision. While there are many methods for estimating the variance parameter for such processes, they usually assume that the process has reached steady state before data collection begins. If this is not the case, then transient behavior can have a significant impact on the estimates of m and s^2.

We present empirical evidence which suggests that transient behavior distorts some variance estimators much more than others. Specifically we consider batch-means estimators and standardized time series based L_p-norm estimators; and we show that the batch-means estimators appear to be significantly less robust to bias.

1 INTRODUCTION

Common issues in simulation output data analysis include estimation of steady-state parameters such as the mean, construction of confidence intervals for these parameters, and selection among various alternative systems. Given stationary simulation output $Y_1, Y_2, \ldots, Y_n$, with unknown mean m, the first step in output data analysis is to compute the sample mean,

$$\bar{Y}_n \equiv n^{-1} \sum_{k=1}^{n} Y_k,$$

as an unbiased point estimator for m. Additionally, as interest frequently lies in obtaining confidence intervals for m, an estimate is usually required of the *variance parameter*,

$$s^2 \equiv \lim_{n\to\infty} nVar(\bar{Y}_n).$$

Unfortunately, while there are many methods for estimating the variance of such processes (Alexopoulos and Seila 1996, Law and Kelton 1991, Pawlikowski 1990, Wilson 1984, Law 1983, or most simulation texts), they all assume that the process has reached steady state before data collection begins (i.e., that $Y_1, Y_2, \ldots, Y_n$ is stationary). If this is not the case, then transient behavior can have a significant impact on estimates of s^2. As it is not a trivial problem to ascertain whether or not a process is stationary, a number of formal initialization bias tests have been developed that try to determine if a process contains an initial transient. This paper discusses the impact of initialization bias on variance estimation, and how initialization bias tests can be used prior to variance estimation to help avoid obtaining bad estimates.

Section 2 provides a brief introduction to the two types of variance estimators that we studied (batch-means and standardized time series L_p-norm estimators), while Section 3 gives some motivational background on initialization bias tests. Section 4 presents experimental results from Monte Carlo simulations for several stochastic processes and bias functions which suggest that batch-means variance estimators are less robust to initialization bias (i.e., they produce potentially highly biased variance estimates in the presence of initialization bias) than are L_p-norm variance estimators. Section 5 expounds upon these issues and provides conclusions and recommendations for future research.

2 BACKGROUND

We evaluate and compare the impact of initialization bias on batch-means and L_p-norm variance estimators. As batch-means estimation is probably the most commonly used variance estimation technique, and is described in most simulation texts, we provide limited background on this technique. On the other hand, L_p-norm variance estimators, as developed in Tokol, et al. (1997), are relatively new and so we give a more comprehensive review.

When defining both of these estimators, we assume that the simulation output $Y_1, Y_2, \ldots, Y_n$ has been divided into b adjacent, nonoverlapping batches, each of length

m (we also assume that $n = mb$); thus, batch i ($1 \le i \le b$) consists of the observations $Y_{(i-1)m+1}, \ldots, Y_{im}$. We introduce the notation

$$\bar{Y}_l \equiv l^{-1} \sum_{k=1}^{l} Y_k \ ,$$

and

$$\bar{Y}_{i,j} \equiv j^{-1} \sum_{k=1}^{j} Y_{(i-1)m+k} \ ,$$

for $1 \le l \le n$, $1 \le i \le b$, and $1 \le j \le m$. Note that $\bar{Y}_{i,m}$ is the ith batch mean.

2.1 Batch-Means Variance Estimators

The batch-means estimator, as described almost everywhere, is defined as

$$V_{BM,b}(m) = \frac{m \sum_{i=1}^{b} \left(\bar{Y}_n - \bar{Y}_{i,m} \right)^2}{b-1} \ .$$

For a fixed $b \ge 2$, as $m \to \infty$, it is known that

$$V_{BM,b}(m) \xrightarrow{\mathcal{D}} \frac{s^2 c_{b-1}^2}{b-1} \ ,$$

where "$\xrightarrow{\mathcal{D}}$" denotes convergence in distribution (see Billingsley 1968) and c_{b-1}^2 represents a random variable having a c^2-distribution with b-1 degrees of freedom. Hence, under uniform integrability (see Chung 1974), $E\left[V_{BM,b}(m)\right] \to s^2$, and one can use $V_{BM,b}(m)$ to estimate s^2.

2.2 L_p-norm Variance Estimators

L_p-norm variance estimators are based upon standardized time series (STS) techniques. The STS from batch i is defined as

$$T_{i,m}(t) \equiv \frac{\lfloor mt \rfloor (\bar{Y}_{i,m} - \bar{Y}_{i,\lfloor mt \rfloor})}{s\sqrt{m}} \quad \text{for } 0 \le t \le 1,$$

where $\lfloor \cdot \rfloor$ is the floor function. Under any of several possible sets of relatively mild conditions (see Glynn and Iglehart 1990 or Schruben 1983),

$$\left(\sqrt{m}\left(\bar{Y}_{i,m} - m \right), s\, T_{i,m} \right) \xrightarrow{\mathcal{D}} \left(s\, W(1), s\, B \right),$$

where $W(t)$ [$B(t)$] is a standard Brownian motion [bridge] process (cf. Billingsley 1968).

The L_p-norm variance estimates are based on certain functionals of the STS and the convergence of these functionals to those of a Brownian bridge process. Let

$$L_{p,k,i}(m) \equiv s^2 \left| \int_0^1 \operatorname{sgn}(T_{i,m}^k(t)) \left| T_{i,m}(t) \right|^p dt \right|^{2/p}$$

for $p > 0$ and $k = 0,1$ be the L_p norms of the time series from batch i, where

$$\operatorname{sgn}(x) \equiv \begin{cases} +1 & \text{if } x \ge 0 \\ -1 & \text{if } x < 0 \end{cases}.$$

(Note that when $p = \infty$, this reduces to the square of the common sup-norm functional.) Since

$$s\, T_{i,m} \xrightarrow{\mathcal{D}} s\, B,$$

the continuous mapping theorem (see Theorem 5.1 of Billingsley 1968) with mapping

$$h(x) = \left| \int_0^1 \operatorname{sgn}\left(x^k\right) |x|^p\, dt \right|^{2/p},$$

can be applied to $\{s\, T_{i,m}\}$ to show that

$$L_{p,k,i}(m) \xrightarrow{\mathcal{D}} L_{p,k} \ ,$$

where

$$L_{p,k} \equiv s^2 \left| \int_0^1 \operatorname{sgn}(B^k(t)) \left| B(t) \right|^p dt \right|^{2/p}.$$

Let us define

$$\bar{L}_{p,k,b}(m) \equiv \frac{1}{b} \sum_{i=1}^{b} L_{p,k,i}(m) \ .$$

If we assume that the $L_{p,k,i}$'s are independent and that $\{h(s\, T_{i,m})\}$ are uniformly integrable, then we have that as $m \to \infty$,

$$E\left[\bar{L}_{p,k,b}(m)\right] \to E\left[L_{p,k}\right] \equiv s^2 c_{p,k} \ ,$$

where the $c_{p,k}$ values, as generated in Tokol, et al. (1997), are provided in Table 1. Thus,

$$V_{p,k,b}(m) \equiv \bar{L}_{p,k,b}(m) / c_{p,k}$$

is an asymptotically unbiased estimate of the variance parameter of the original time series. The performance of (and a much more detailed background on) L_p-norm estimators for certain (p,k) pairs is studied in Goldsman, et al. (1997). In particular, the (p,k) = (1,1) case corresponds to Schruben's (1983) area estimator, and the (p,k) = (2,0) case is the Cramér-von Mises estimator of Goldsman, Kang, and Seila (1997).

Table 1: L_p-Norm $c_{p,k}$ Values

p	$c_{p,0}$	$c_{p,1}$
1	0.11667	0.08333
2	0.16667	0.1432
3	0.2084	0.1912
5	0.2754	0.2645
10	0.3848	0.3799
80	0.6807	0.6805
∞	0.82247	

3 INITIALIZATION BIAS TESTS

When constructing either of the variance estimators discussed in Section 2, one assumes that the simulation output does not contain a "significant" initial transient. Although transient behavior is not always easy to detect in simulation output, it can seriously impact the center and the length (and hence the validity) of confidence intervals for m and s^2. Most of the methods for dealing with this dilemma (Ockerman and Goldsman 1996, Goldsman, Schruben, and Swain 1994, Chance 1993, Wilson and Pritsker 1978) involve either specifying initial conditions, providing truncation rules, and/or testing for initialization bias.

In some cases, enough is understood about a system to be able to start the simulation in a condition that is representative of steady state. If possible, this is the desired approach. Otherwise, the simulation is usually started in an arbitrary condition and the output data from an initial time period are discarded (i.e., not used in the output analysis process). This initial time period should extend until the simulation has reached a condition that is representative of steady state; but it is not always obvious how long a simulation must run before it reaches steady state. Rules for determining what portion of the initial data to throw out are called truncation rules (Glynn and Inglehart 1987, Snell and Schruben 1985, Heidelberger and Welch 1983, Kelton and Law 1983).

After we believe that the transient portion of the data has been removed, and the remaining process is deemed stationary, then an initialization bias test can be performed to determine if, statistically, this is indeed the case. These initialization bias tests are typically hypothesis tests with null hypothesis, H_0: *no initialization bias present*, and alternative hypothesis, H_1: *initialization bias present* (Ockerman and Goldsman 1996, Goldsman, Schruben, and Swain 1994, Vassilacopoulos 1989, Schruben, Singh, and Tierney 1983, Schruben 1982).

Certain initialization bias tests are much more powerful at detecting bias than others; unfortunately they also tend to have an increased false alarm rate. It is not clear which test should be applied in any given situation (Cash, et al. 1992). We show in Section 4 that the choice of which initialization bias test to use in a given context should perhaps be linked to the choice of the estimator that will subsequently be used to estimate the variance.

4 MONTE CARLO EXPERIMENTS

How concerned do we have to be about detecting initialization bias? Just how skewed are the various variance estimators by initialization bias? We present results for the batch-means estimator and for the two "extreme" members of the class of L_p-norm estimators (i.e., $V_{1,1,b}(m)$ and $V_{\infty,0,b}(m)$).

Tables 2-4 are based on 10,000 Monte Carlo runs and provide a quantitative feel for the magnitude of the sensitivity of the batch-means and L_p-norm estimators to slight nonstationarities. The tables give estimates of the expected values of the three variance parameter estimators for three stochastic processes, five run lengths (from 64 to 16384), and three batching schemes (4, 8, and 16 batches).

The three stationary processes considered are:

A step process [STEP], Y_j, $j = 1,2,\ldots$, where the Y_j's are independent and identically distributed (i.i.d.) ±1, each with probability 0.5. For this process $s^2 = 1$.

A first-order autoregressive process [AR(1)], $Y_{j+1} = fY_j + Z_{j+1}$, $j = 1,2,\ldots$, where the Z_j's are i.i.d. $\mathcal{N}(0,1-f^2)$ with $-1 < f < 1$. (We consider f = 0.9, in which case s^2 = 19.)

A first-order moving average process [MA(1)], $Y_{j+1} = fZ_j + Z_{j+1}$, $j = 1,2,\ldots$, where the Z_j's are i.i.d. $\mathcal{N}(0,1)$. (We consider f = -0.9, so that s^2 = 0.01.)

For each of the stochastic processes, we consider both a stationary and a nonstationary version. The stationary version is generated by initializing the process from the steady-state distribution. The nonstationary version is formed by adding a slight transient-mean function similar to that in Figure 1 to the stationary process on a point-by-point basis, i.e., $y_i = x_i + e_i$ where $\{x_i\}$ is the stationary stochastic process and $\{e_i\}$ is the bias function, where $e_{i+1} = e_i(1-10/n)$ and e_1 is chosen in such a way that we can compare results among the three stochastic processes. Specifically, to standardize the amount of added bias (in order to regulate the bias levels), we use the artificial measure

$$k_n = \sum_{i=1}^{n} \frac{\left(\bar{e} - e_i\right)^2}{n-1},$$

where $\bar{e}$ is the mean of the bias terms and e_1 is adjusted so that k_n is a fixed multiple of the process variance.

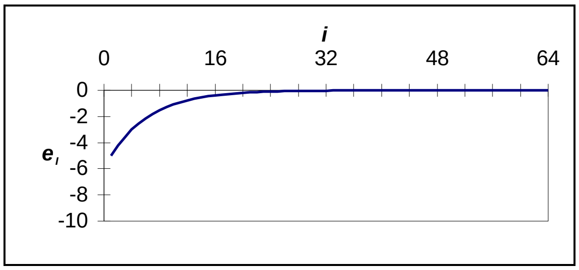

Figure 1: Sample Transient Mean Function

The top entries in each of the tables represent the estimated expected values of the variance estimates based on the nonstationary processes, while the bottom entries (in parentheses) represent the expected values of the variance estimates obtained from the stationary ones. Any significant differences between the top and bottom entries in the tables represent the impact of the initial transient.

Table 2: Impact of Initialization Bias on Variance Estimates for a STEP Process*

b	Estimator Type	*n*=64	*n* =256	*n* =1024	*n* =4096	*n* =16384
4	L_p-Norm (p,k)=(1,1)	1.00 (0.99)	1.01 (0.99)	1.06 (0.99)	1.25 (1.00)	1.98 (1.00)
	L_p-Norm ($p=\infty$)	0.75 (0.75)	0.88 (0.88)	0.96 (0.93)	1.06 (0.97)	1.31 (0.98)
	Batch Means	1.02 (1.01)	1.03 (1.00)	1.14 (1.00)	1.54 (1.01)	3.15 (0.99)
8	L_p-Norm (p,k)=(1,1)	0.99 (0.99)	1.00 (1.00)	1.01 (1.00)	1.04 (1.00)	1.18 (1.00)
	L_p-Norm ($p=\infty$)	0.63 (0.63)	0.83 (0.83)	0.92 (0.91)	0.97 (0.96)	1.04 (0.98)
	Batch Means	1.01 (1.01)	1.02 (1.00)	1.08 (1.00)	1.34 (1.00)	2.32 (1.00)
16	L_p-Norm (p,k)=(1,1)	0.94 (0.94)	1.00 (1.00)	1.00 (1.00)	1.01 (1.00)	1.03 (1.00)
	L_p-Norm ($p=\infty$)	0.48 (0.48)	0.75 (0.75)	0.88 (0.88)	0.94 (0.94)	0.98 (0.97)
	Batch Means	1.01 (1.01)	1.01 (1.00)	1.04 (1.00)	1.17 (1.00)	1.69 (1.00)

* Top (bottom) entries for each estimator represent the sample mean of the variance estimates for a nonstationary (stationary) STEP process. The true variance for the STEP process is 1.0. Significant deviations in the top entries represent the effects of bias.

Table 3: Impact of Initialization Bias on Variance Estimates for an AR(1) Process*

b	Estimator Type	*n*=64	*n* =256	*n* =1024	*n* =4096	*n* =16384
4	L_p-Norm (p,k)=(1,1)	3.3 (2.8)	12.8 (11.3)	23.1 (16.9)	43.5 (18.3)	118.0 (18.8)
	L_p-Norm ($p=\infty$)	1.2 (1.0)	5.9 (5.5)	13.2 (11.4)	22.7 (15.1)	44.8 (17.0)
	Batch Means	8.5 (7.6)	18.9 (15.6)	32.1 (18.4)	72.6 (18.6)	234.0 (19.0)
8	L_p-Norm (p,k)=(1,1)	1.1 (1.0)	6.9 (6.6)	16.0 (14.8)	22.7 (18.1)	37.4 (18.6)
	L_p-Norm ($p=\infty$)	0.4 (0.3)	2.8 (2.7)	9.0 (8.6)	15.0 (13.6)	21.9 (16.2)
	Batch Means	5.3 (4.7)	15.0 (12.9)	25.9 (17.5)	51.9 (18.5)	151.0 (18.9)
16	L_p-Norm (p,k)=(1,1)	0.3 (0.3)	2.9 (2.8)	11.5 (11.3)	17.7 (17.0)	21.1 (18.4)
	L_p-Norm ($p=\infty$)	0.1 (0.1)	1.1 (1.1)	5.5 (5.5)	11.6 (11.4)	16.0 (15.1)
	Batch Means	2.9 (2.7)	10.4 (9.3)	20.4 (16.0)	35.7 (18.2)	87.8 (18.7)

* Top (bottom) entries for each estimator represent the sample mean of the variance estimates for a nonstationary (stationary) AR(1) process. The true variance for the AR(1) process is 19.0.

The sensitivity of the batch-means estimator is clear. For example, compare the 16-batch variance estimators for the AR(1) process with length 16384 from Table 3. The true variance of this AR(1) process is 19, and ideally the expected value of each of the three variance estimators should be reasonably close to this figure. The batch-means estimator provides the best estimates when no initialization bias is present (i.e., the average estimate from batch means is 18.7, while the other estimators provide average estimates of 15.1 and 18.4). However, when bias is introduced, the batch means average estimate is skewed to 87.7, far beyond the other averages of 21.1 and 16.0.

Table 4: Impact of Initialization Bias on Variance Estimates for an MA(1) Process*

b	Estimator Type	*n*=64	*n* =256	*n* =1024	*n* =4096	*n* =16384
4	L_p-Norm (p,k)=(1,1)	0.344 (0.343)	0.096 (0.094)	0.037 (0.031)	0.040 (0.015)	0.109 (0.011)
	L_p-Norm ($p=\infty$)	0.467 (0.467)	0.183 (0.182)	0.073 (0.070)	0.042 (0.031)	0.050 (0.019)
	Batch Means	0.152 (0.151)	0.049 (0.045)	0.032 (0.019)	0.066 (0.012)	0.226 (0.011)
8	L_p-Norm (p,k)=(1,1)	0.673 (0.673)	0.180 (0.180)	0.053 (0.052)	0.025 (0.021)	0.031 (0.013)
	L_p-Norm ($p=\infty$)	0.689 (0.689)	0.297 (0.297)	0.113 (0.112)	0.047 (0.045)	0.030 (0.023)
	Batch Means	0.261 (0.261)	0.075 (0.073)	0.034 (0.026)	0.047 (0.014)	0.144 (0.011)
16	L_p-Norm (p,k)=(1,1)	1.280 (1.280)	0.345 (0.345)	0.095 (0.094)	0.032 (0.031)	0.018 (0.015)
	L_p-Norm ($p=\infty$)	0.891 (0.891)	0.468 (0.468)	0.183 (0.183)	0.070 (0.070)	0.033 (0.031)
	Batch Means	0.487 (0.486)	0.132 (0.130)	0.044 (0.040)	0.035 (0.018)	0.081 (0.012)

* Top (bottom) entries for each estimator represent the sample mean of the variance estimates for a nonstationary (stationary) MA(1) process. The true variance for the MA(1) process is 0.01.

For the STEP and the MA(1) process, the results are similar. Thus, for these types of processes and bias functions, it would appear that the batch-means variance estimators, while perhaps more effective for truly stationary processes (note the convergence rates to the proper values in the tables), are indeed less robust to transients than the L_p-norm variance estimators.

5 DISCUSSION AND CONCLUSIONS

For these examples, we see that the batch-means estimator is more sensitive to bias than are the L_p-norm estimators. It would seem that this is true because the batch-means estimators use information *between* batches (where this type of bias tends to manifest itself dramatically), while the L_p-norm estimators only use information *internal* to batches (where the bias is somewhat hidden). Clearly, we could construct bias functions with high frequency components, such that the bias would be hidden when looking across batches and significant when looking inside batches. However, in most real-world applications, the bias tends to be slow moving and positively correlated as in our examples.

Thus, in practice, the impact of initialization bias on variance estimates can be significant. When removing an initial transient from simulation output, it is important to keep in mind the eventual variance estimator that will be used. If this variance estimator is less robust to transients, then one should precede variance estimation with a fairly powerful initialization bias test. If one does not want to be concerned with removing the initial transient, a more robust variance estimator may be chosen.

We demonstrated that the notion of robustness of a variance estimator can, in some sense, be characterized. Further research will be directed toward formalizing this process so that estimator robustness can be used as a criterion for evaluating various variance estimators.

ACKNOWLEDGMENT

David Goldsman's work was supported by National Science Foundation Grant DMI-96-22269.

REFERENCES

Alexopoulos, C. and A. F. Seila. 1996. Implementing the batch means method in simulation experiments. In *Proceedings of the 1996 Winter Simulation Conference*, 214-221.

Billingsley, P. 1968. *Convergence of Probability Measures*. New York: John Wiley & Sons.

Cash, C. R., B. L. Nelson, J. M. Long, D. G. Dippold, and W. P. Pollard. 1992. Evaluation of tests for initial-condition bias. In *Proceedings of the 1992 Winter Simulation Conference*, 577-585.

Chance, F. 1993. A historical review of the initial transient problem in discrete event simulation literature. Technical Report. School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York.

Chung, K. L. 1974. *A Course in Probability Theory*. New York: Academic Press.

Glynn, P. W. and D. L. Iglehart. 1987. A new initial bias deletion rule. In *Proceedings of the 1987 Winter Simulation Conference*, 318-319.

Glynn, P. W. and D. L. Iglehart. 1990. Simulation output analysis using standardized time series. *Mathematics of Operations Research* 15: 1-16.

Goldsman, D., K. Kang, and A. F. Seila. 1997. Cramér-von Mises variance estimators for simulations. To appear in *Operations Research*.

Goldsman, D., M. Koksalan, D. H. Ockerman, J. Picciuto, and G. Tokol. 1997. Standardized time series L_p-norm confidence interval estimators for simulations. Technical Report. School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.

Goldsman, D., L. W. Schruben, and J. J. Swain. 1994. Tests for transient means in simulated time series. *Naval Research Logistics* 41: 171-187.

Heidelberger, P. and P. D. Welch. 1983. Simulation run length control in the presence of an initial transient. *Operations Research* 31: 1109-1145.

Kelton, W. D. and A. M. Law. 1983. A new approach for dealing with the startup problem in discrete event simulation. *Naval Research Logistics Quarterly* 30: 641-658.

Law, A. M. 1983. Statistical analysis of simulation output data. *Operations Research* 31: 983-1029.

Law, A. M. and W. D. Kelton 1991. *Simulation Modeling and Analysis*, 2nd Edition. New York: McGraw-Hill.

Ockerman, D. H. and D. Goldsman. 1996. Tests for transients in simulated time series based on L_p-norm variance estimators. Technical Report. School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.

Pawlikowski, K. 1990. Steady-state simulation of queueing processes: a survey of problems and solutions. *ACM Computing Surveys* 22: 123-170.

Schruben, L. W. 1982. Detecting initialization bias in simulation output. *Operations Research* 30: 569-590.

Schruben, L. W. 1983. Confidence interval estimation using standardized time series. *Operations Research* 31: 1090-1108.

Schruben, L. W., H. Singh, and L. Tierney. 1983. Optimal tests for initialization bias in simulation output. *Operations Research* 31: 1167-1178.

Snell, M. and L. W. Schruben. 1985. Weighting simulation data to reduce initialization effects. *IIE Transactions* 17(4): 354-363.

Tokol, G., D. Goldsman, D. H. Ockerman, and J. J. Swain. 1997. Standardized time series L_p-norm variance estimators for simulations. To appear in *Management Science*.

Vassilacopoulos, G. 1989. Testing for initialization bias in simulation output. *Simulation* 52(4): 151-153.

Wilson, J. R. (1984). Statistical aspects of simulation. *Operational Research* 921-937.

Wilson, J. R. and A. A. B. Pritsker. 1978. A survey of research on the simulation start-up problem. *Simulation* 31: 55-58.

AUTHOR BIOGRAPHIES

DAVID GOLDSMAN is an associate professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. He received his Ph.D. from the School of Operations Research and Industrial Engineering at Cornell University in 1984. His research interests include simulation output analysis and ranking and selection techniques. He is an active participant in the Winter Simulation Conference, having been the Program Chair in 1995.

DAN OCKERMAN is a staff scientist at Retek Information Systems in Atlanta. He received his Ph.D. from the School of Industrial and Systems Engineering at the Georgia Institute of Technology in 1996. His research interests include simulation output analysis, forecasting, and neural networks.