Proceedings of the 2009 INFORMS Simulation Society Research Workshop
L.H. Lee, M.E. Kuhl, J.W. Fowler, and S. Robinson, eds
.

THE PROBLEM OF THE INITIAL TRANSIENT (AGAIN), OR WHY MSER WORKS

K. Preston White, Jr.

Department of Systems and Information Engineering
University of Virginia
Charlottesville, VA 22904-4747

Stewart Robinson

Warwick Business School
University of Warwick
Coventry, CV4 7AL, U.K.

ABSTRACT

In a comprehensive study of methods for dealing with the problem of the initial transient, Hoad *et al.* (2008) determined that the MSER (White, 1997) was an efficient and effective truncation rule appropriate for automation. In this paper, we suggest that the MSER works well because it minimizes an approximation to the mean-squared error in the estimated steady-state mean. Using the example of an M/M/1 queue, we provide a clear statement of the problem in both the time and frequency domains, distinguishing between the biasing effects of initialization and autocorrelation. We also demonstrate that, as a result of autocorrelation, the objective of minimizing initialization bias is *not* exclusively a matter of determining the most representative initial condition. This observation further argues against the replication/deletion approach to output analysis.

1 INTRODUCTION

The problem of the *initial transient*, also known as the *start-up* or *warm-up* or *initialization-bias problem*, arises in estimating the limiting distribution of a non-terminating stochastic process from output generated by one or more replications of a discrete-event simulation. Because the steady-state operating regime does not offer natural boundary conditions for starting or stopping simulation runs, initial conditions typically are chosen for convenience and terminating conditions are sought which provide satisfactory point and interval estimates. It is well known that arbitrary initialization can introduce bias in estimators of the limiting statistics. In this paper we illustrate the problem, using the example of an M/M/1 queue, and focus on the behavior of the MSER procedure for removing the bias.

2 A SIMPLE EXAMPLE

The fundamental issues can be elucidated using the straightforward example of an M/M/1 queue for which we are interested in estimating a confidence interval for the mean number in system from a simulation experiment. The output of a single simulation run is a realization of the discrete-state, continuous-parameter stochastic process $\{Y_i;\ i=0, 1, 2, \ldots, n\}$, where $Y_i = Y(t_i)$ is the i^{th} sequential observation from a simulation with run length of n observations and initial condition $Y_0=y_0$. The sample mean and large-sample variance are

$$\bar{Y}(n \mid y_0) = \sum_{i=0}^{k} \omega_i Y_i$$

$$S^2(n \mid y_0) = \sum_{i=0}^{k} \omega_i Y_i^2 - \bar{Y}(n \mid y_0)^2$$

where the weight $\omega_i=1/(n+1)$ and $k=n$ for variables which are tallied (such as customer waiting times), or $\omega_i=(t_{i+1}-t_i)/(t_n-t_0)$ and $k=n-1$ for variables which are time-averaged (such as the number in system). Since the process generating these observations is ergodic, these are consistent estimators for the steady-state mean θ and variance σ^2, independently of the choice of initial condition. However, because the observations are sequentially correlated, for finite n and arbitrary $y_0 \neq \theta$, both estimators are biased. Thus there is bias in both the location (accuracy) and width (precision) of the estimated confidence interval on the mean.

This is illustrated in the time domain in Figure 1. Each panel shows the sequence $\{y_i;\ i=0,1,2, \ldots, n\}$, the number in system for an M/M/1 queue as a function of the observation number for a single simulation run, where observations are taken for each change in state. The mean arrival rate is $\lambda=1$ and the terminating event is 10,000 departures, resulting in a run length $t_n \cong 10{,}000$ with $n \cong 20{,}000$ arrival and departure events. Also shown is the corresponding cumulative sample mean, the sequence $\{\bar{y}(i \mid y_0), i = 0,1,\ldots,n\}$. The queues in panels (a) and (b) have traffic intensities of $\rho=0.7$ and these queues are initialized respectively at $y_0=90$ (a rare event, with $\Pr(Y_i \geq 90)=6.86\times10^{-15}$) and $y_0=4$ (the approximate steady-state mean). The queues in panels (c) and (d) have a traffic intensities of $\rho=0.9$, initialized respectively at

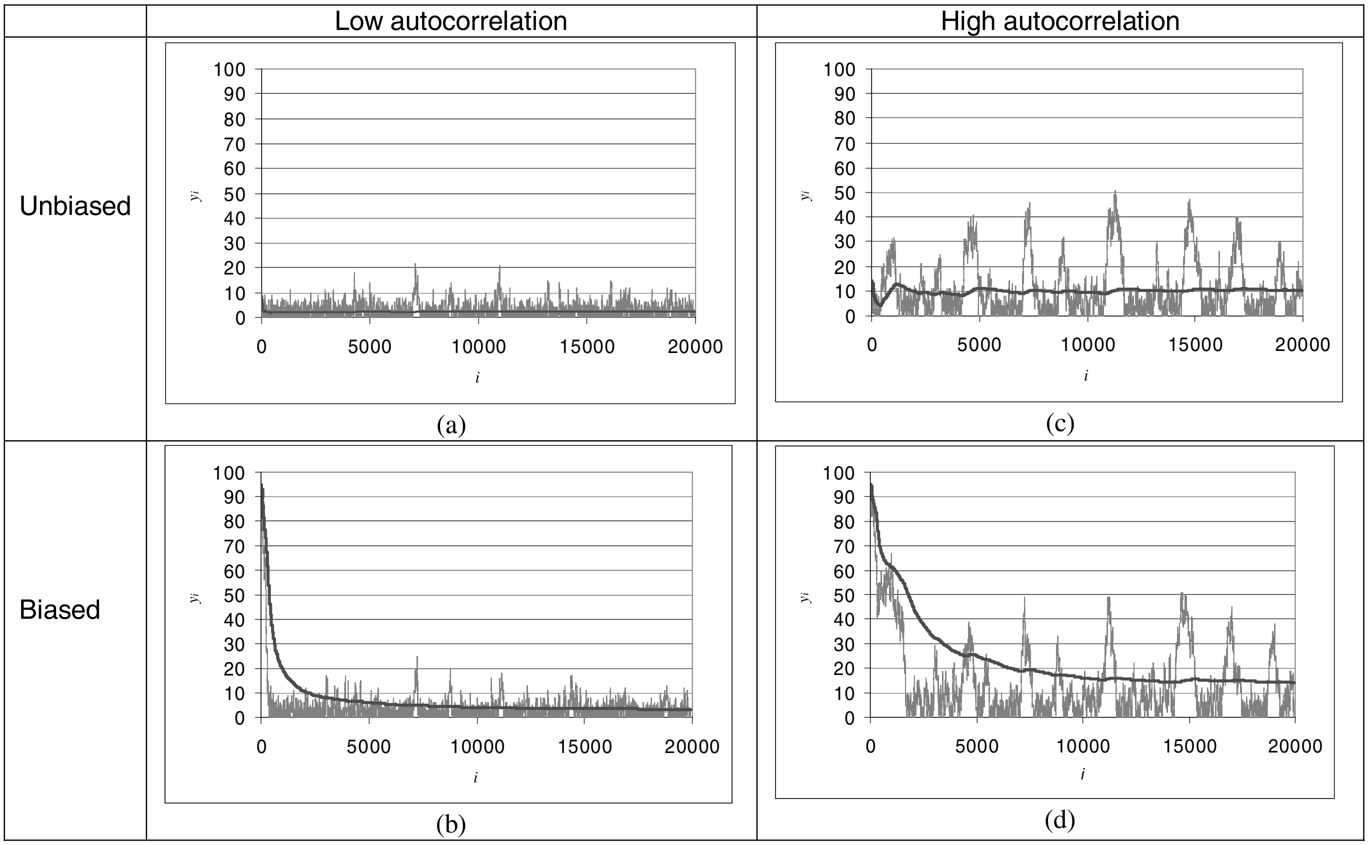

Figure 1. The time-domain effects of initial conditions and autocorrelation on the running estimate of mean number in system for two M/M/1 queues differing in traffic intensity.

y_0=90 (Pr(Y_i>90)=6.86×10^{-5}) and y_0=9 (the steady-state mean). The bias introduced by the rare initial conditions is clear for both queues; the severity of this bias clearly is greater for the queue with higher traffic intensity (greater autocorrelation).

These effects are perhaps even more apparent in the frequency domain, as shown in Figure 2. The stationary-state for the queue is distributed geometrically, Y_{ss}~Geom(1-ρ), with mean θ=ρ/(1-ρ) and variance $\sigma^2=\rho/(1-\rho)^2$. Panels (a) and (c) compare the histograms of the estimated distributions for both the unbiased and biased initial conditions (for the entire run) to the probability mass function (pmf) of this theoretical distribution. For clarity, these comparisons are repeated in panels (b) and (d), where the steady-state probabilities are now plotted on a logarithmic scale. This transforms the geometric pmf into a straight line and exaggerates the differences between the estimated and theoretical distributions at lower probabilities.

This example illustrates the fundamental issue. Every state is *positive recurrent,* Pr(Y_i=y)>0 for all $y \geq 0$ at steady state, and therefore the states themselves are *not* transient in the usual sense. For both the unbiased and biased cases, however, there is a transient period associated with the *sample mean*, $\overline{Y}(i, y_0)$, during which a sufficiently large number of observations must be collected for the sampling distribution to approximate the true stationary distribution. For the biased case, there also is a transient period associated with the system *state*, Y_i, during which the state decreases from an improbable artificial initial value to subsequent values more representative of the steady-state operating regime. *Both* transients are prolonged by increasing autocorrelation in the values of the state. The duration of the mean transient is considerably longer than that of the state transient, because of the additional autocorrelation induced by calculating the cumulative mean.

With these observations, it is apparent that there is some minimum run length required to reduce the sampling error to an acceptable level for estimation of any statistic. This minimum run length increases with the degree of autocorrelation and is an inescapable artifact of the system under study. This minimum run length also increases with the increasing rarity of the initial value of the state. *The problem of the initial transient, therefore, is to reduce the run length required to yield accuracy and precision in the sample statistics comparable to that in the unbiased case; or, alternatively, for a given computing budget, to improve*

.

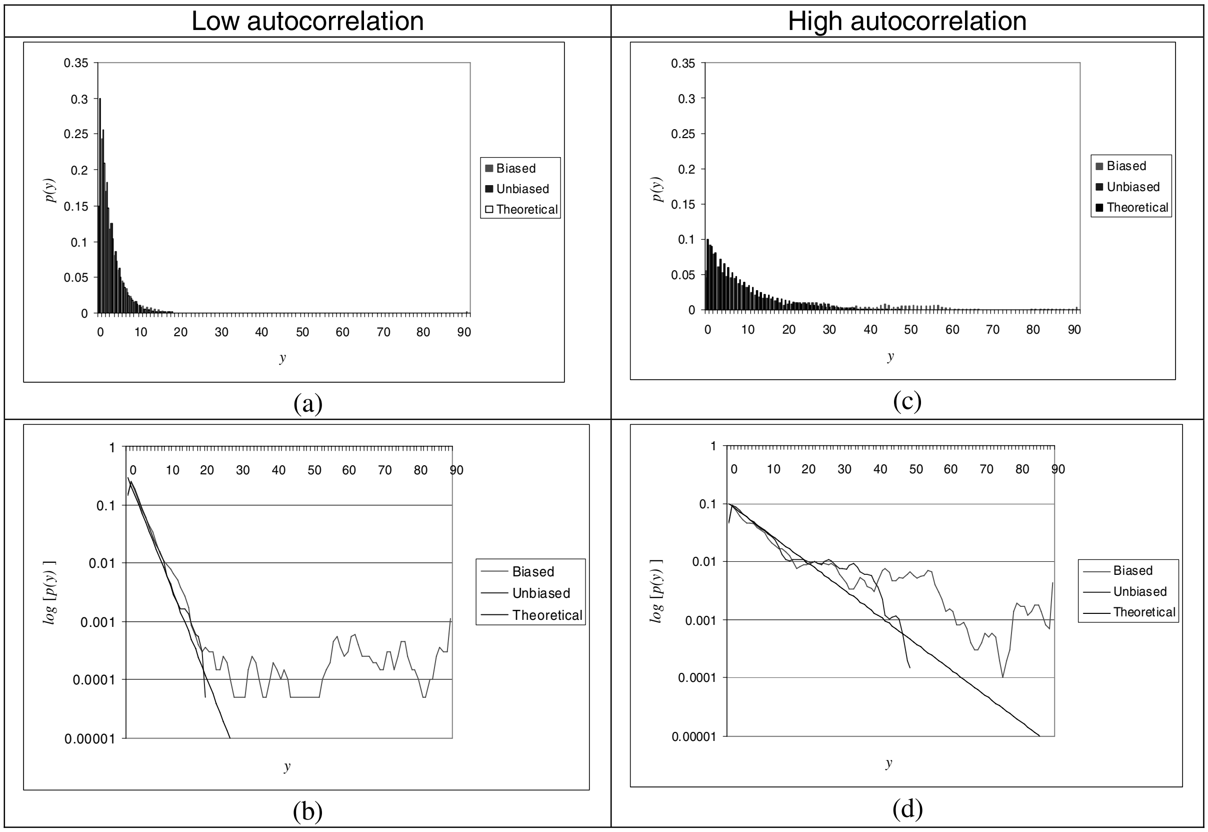

Figure 2. The frequency-domain effects of initial conditions and autocorrelation on the running estimate of mean number in system for two M/M/1 queues differing in traffic intensity.

the accuracy and precision in the sample statistics by removing the contribution of the state transient.

3 THE MSER

Heuristics are the oldest non-graphical approaches to determining the length of the warm-up period. In many instances, these are based on the calculation of what amounts to a visual cue (such as the state trajectory crossing the mean trajectory), which is easy to understand and equally easy to compute. In a recent study, Hoad *et al.* (2008) investigated a wide range of approaches to resolving the problem, including heuristics, graphical procedures, initialization bias tests, statistical methods, and hybrid approaches. Based on criteria which included accuracy, robustness, simplicity, potential for automation, number of estimated parameters, and computation speed, they selected the MSER-5 heuristic as the most suitable for automation.

The original MSER (McClarnon, 1990; White, 1997) determines the length of the warm-up period by solving the following optimization problem:

$$d^* = \arg\min_{n>>d\geq 0}\left[MSER(n,d \mid y_0)\right]$$

where

$$MSER(n,d \mid y_0) = \frac{1}{(n-d)^2}\sum_{i=d+1}^{n}\left(Y_i - \overline{Y}(n,d \mid y_0)\right)^2$$

The intuition is that this choice minimizes the width of marginal confidence interval on the mean. While this interval is calculated using the sample variance of the reserved sequence (the sequence remaining after truncation) as an estimator for the variance, Franklin and White (2008) have shown that substituting an unbiased estimator is less efficient and does not improve the performance of the heuristic. Because the state transient is much shorter than the mean transient, truncation at d^* eliminates initialization bias and approximately minimizes the mean-squared error in the estimate of the mean.

MSER-*m* (Spratt, 1998; White, *et al.*, 2000) is an improvement in MSER, which replaces the output series $\{Y_i;\ i=0,1,\ \ldots,\ n\}$ with the series of batch means $\{Z_j,\ j=0,1,\ldots,\lfloor n/m\rfloor\}$, where

$$Z_j = (1/m)\sum_{p=1}^{m} Y_{m(j-1)+p}$$

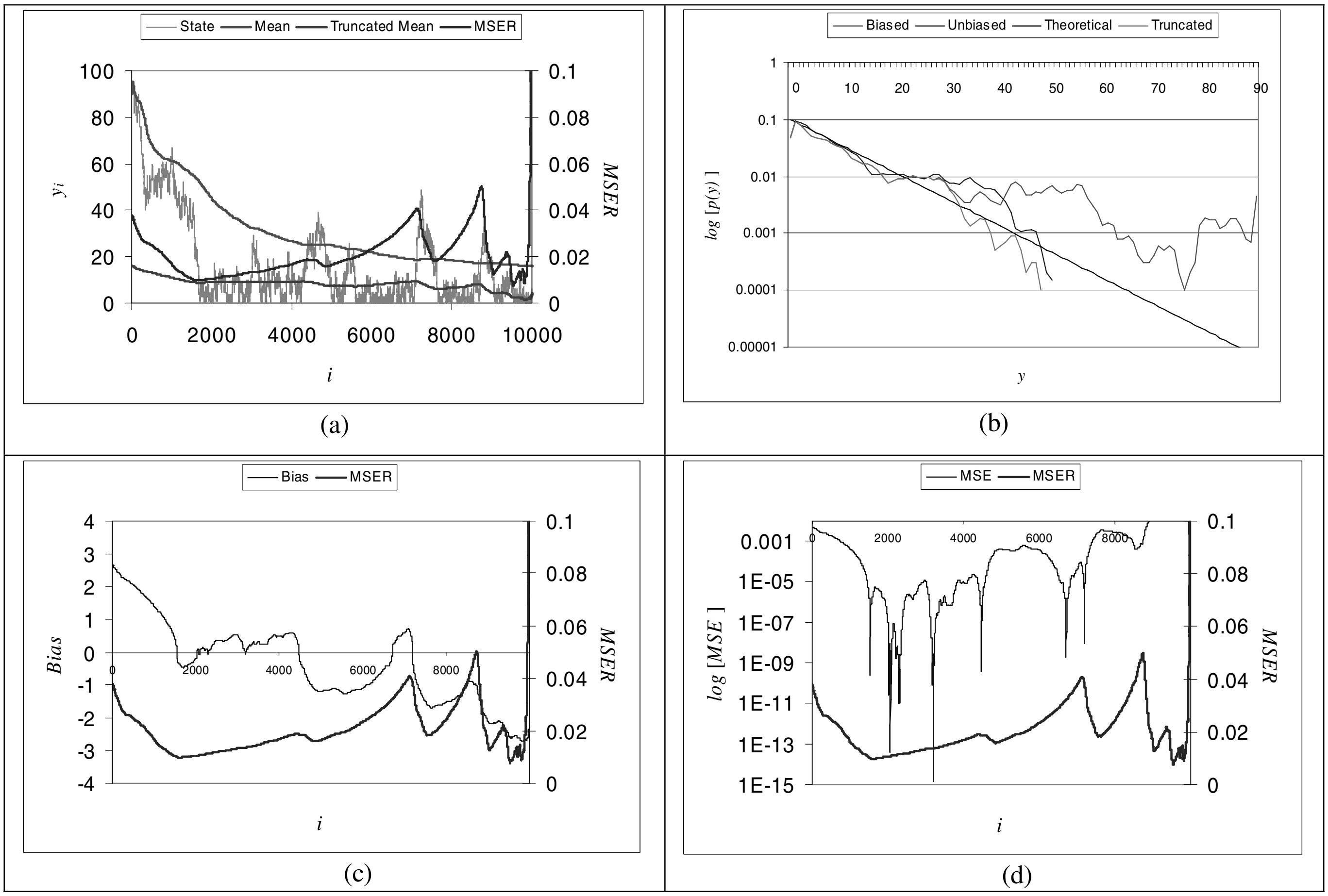

Figure 3. Results of the application of MSER-5 for mean number in system for an M/M/1 queue with traffic intensity ρ=0.9.

The length of the batch is somewhat arbitrary, but the value of m=5 has shown to yield superior results over a wide range of test problems.

Figure 3 illustrates the application of MSER-5 to the M/M/1 queue with traffic intensity of ρ=0.9 with n≅10,000 events. In addition to the output sequence and cumulative mean, panel (a) also shows the truncated mean $\bar{Y}(n,d \mid y_0)$ and $MSER(n,d|y_0)$ statistic as functions of the truncation point $d=i$. The minimum value of the statistic over the initial component of the run obtains at d^*=1605 (t_{d^*} =781.8), with Y_{d^*+1}=18 as the initial value of the reserved sequence and $Y(n,d^*+1) \mid y_0) = 8.871$ as the truncated point estimate of the mean.

In addition to the biased, unbiased, and theoretical log-densities, panel (b) shows the log-density of the reserved sequence for the MSER-5 truncation point. Clearly the state-transient has been removed. Equally clearly, there remains a substantial mean-transient, which causes a *low* estimate of the mean for this run. This is true for the unbiased and biased sequences, as well, and is not an artifact of initialization. Given the high autocorrelation, a significantly longer run length is required to provide reasonable precision in the estimate.

Panels (c) and (d) show the bias and logarithm of the mean-squared error in the estimate as a function of the truncation point, respectively. The absolute value of the bias and the mean-squared error achieve a first local minimum at d=1540. (t_d =755.2), with Y_{d^*+1}=38 and $Y(n,d+1) \mid y_0) = 8.998$. Note that MSER over-truncates by only 65 observations (0.00065%) in this example. *Note, further, that the longest unbiased sequence is not achieved not by "correct" re-initialization at the steady-state mean, but rather by reference to the minimum bias and MSE.*

4 HOW MANY REPLICATIONS?

There are two standard frameworks for steady-state output analysis. *Replication/deletion* uses m independent replications, deleting the mean transient from each. Its advantage is that the mean estimates from each replication are uncorrelated and easily computed. Its disadvantage is inefficiency. The mean transient is repeated in each replication at a cost of $d \times m$ deleted observations, if the same warm-up period is used for each replication (which is typically the case). This is compounded by the need for a conservative value for d, which is sufficiently large to capture the transient for each replication. Moreover, as we have illustrated, selecting the most representative initial condition for the truncated sequence is not necessarily optimal in the MSE sense.

The alternative to replication/deletion is a *single, long run*, with the advantage that the transient need be identified and deleted only once. The disadvantage is the need to construct an unbiased estimator for the variance which corrects for autocorrelation of the observations. This is most often accomplished by the method of *batch means*. The reserved time-series is divided into some number (typically 20-40) sequential batches, the mean is computed for each batch, and the variance of the batch means is used to estimate the confidence interval on the mean. This approach is effective if the batches are sufficiently large to result in uncorrelated batch means. While many authors express a strong preference for replication/deletion approach, the method of batch means can be vastly more efficient (Alexopoulos and Goldsman, 2004).

We illustrate this observation by applying MSER directly to one long run of the example M/M/1 simulation, given a computing budget of $n \cong 100{,}000$ observations. We previously employed this same budget in the determination of a warm-up period applying Welch's procedure (Welch, 1983; Law, 2006; Robinson, 2006), with ten independent replications and an averaging window of 1500 observations. MSER-5 recommends a truncation point of $d^*=3350$, approximately half that determined previously for the Welch moving-averaged sequence. Applying batch means to the reserved sequence with 20 batches, the estimated 95% confidence interval on the mean is $\bar{Y}(n,d \mid y_0) = 9.5061 \pm 1.774$.

Note that in applying Welch's procedure, the entire budget was expended in developing a clear sense of the initial transient, in order to determine a suitable initial condition for further output analysis. In contrast, using this same budget to generate a single, long replication, and applying MSER-5 to the result, yields reasonable point and interval estimates of the mean without additional computation. In our research, this economy has been consistent across all instances of a wide range of test problems.

5 CONCLUSIONS

In this paper, we provide a clear statement of the initialization problem, distinguishing between the biasing effects of initialization and autocorrelation. We demonstrate that, as a result of autocorrelation, the objective of minimizing initialization bias is *not* exclsuviely a matter of determining the most representative initial condition. We suggest that the MSER works well because it minimizes an approximation to the mean-squared error in the estimated steady-state mean. This demonstration further argues against the replication/deletion approach to output analysis.

ACKNOWLEDGMENTS

This paper is adapted from the authors' contribution to the Wiley *Encyclopedia of Operations Research.*

REFERENCES

Alexopoulos, C., and D. Goldsman. 2004. To batch or not to batch? *ACM Transactions on Modeling and Computer Simulation* 14(1)76-114.

Franklin, W.W, and White, K.P., Jr. (2008) "Stationarity tests and MSER-5: Exploring the intuition behind-mean-squared-error reduction in detecting and correcting initialization bias", In *Proceedings of the 2008 Winter Simulation Conference*, S. J. Mason, R. Hill, L. Moench, and O. Rose, eds., Miami, FL, December, 541-546

Hoad, K., S. Robinson, S., and R. Davies. 2008. Automating warm-up length estimation. In *Proceedings of the 2008 Winter Simulation Conference*, 532-540.

Law, A. M., 2006. *Simulation modeling and analysis.* 4th ed., New York: McGraw-Hill.

McClarnon, M. A. (1990) *Detection of Steady-State in Discrete-Event Dynamic Systems: An Analysis of Heuristics,* M.S. Thesis, Department of Systems Engineering, University of Virginia.

Robinson, S. 2004. *Simulation. The practice of Model Development and Use*. England: John Wiley & Sons Ltd.

Spratt, S. C. 1998. *An evaluation of contemporary heuristics for the startup problem.* M. S. thesis, Faculty of the School of Engineering and Applied Science, University of Virginia.

Welch, P.D. 1983. The statistical analysis of simulation results. In *The Computer Performance Modeling Handbook*, ed., S. S. Lavenberg, Academic Press, New York.

White, K. P., Jr. (1997) "An effective heuristic for bias reduction in simulation output," *Simulation*, **69**(6)323-

White K. P., Jr., M. J. Cobb, and S. C. Spratt. 2000. A comparison of five steady-state truncation heuristics for simulation. In *Proceedings of the 2000 Winter Simulation Conference*, 755-760.

AUTHOR BIOGRAPHIES

K. PRESTON WHITE, JR., is Professor of Systems and Information Engineering at the University of Virginia. His e-mail address is <kpwhite@virginia.edu>.

STEWART ROBINSON is Professor of Operational Research and Associate Dean of Specialist Masters Programmes in the Operational Research and Management Sciences Group in the Warwick Business School at the University of Warwick. His e-mail address is <stewart.robinson@warwick.ac.uk>.