

## THE INITIAL TRANSIENT IN STEADY-STATE POINT ESTIMATION: CONTEXTS, A BIBLIOGRAPHY, THE MSE CRITERION, AND THE MSER STATISTIC

Raghu Pasupathy

Department of Industrial and Systems Engineering  
Virginia Tech  
Blacksburg, VA 24061 USA

Bruce Schmeiser

School of Industrial Engineering  
Purdue University  
West Lafayette, IN 47907 USA

### ABSTRACT

The initial transient is an unavoidable issue when estimating parameters of steady-state distributions. We discuss contexts and factors that affect how the initial transient is handled, provide a bibliography (from the system simulation literature), discuss criteria for evaluating initial-transient algorithms, arguing for focusing on the mean squared error (mse). We discuss the MSER statistic, showing that it is asymptotically proportional to the mse and therefore a good foundation for initial-transient algorithms. We suggest two new algorithms (MSER-LLM and MSER-LLM2) for using the MSER statistic and compare them, based on empirical results for M/M/1 and AR(1) data processes, to the original MSER algorithm (MSER-GM).

### 1 INTRODUCTION

We consider a problem that arises in determining steady-state behavior of stochastic systems. Whether the analysis is via closed-form analysis, a deterministic numerical method, or Monte Carlo simulation, analysis must distinguish between *transient* results, which depend upon the initial state of the system, and *steady-state* results, which do not.

Initial-transient problems arise in Monte Carlo simulation experiments designed to estimate the unknown value of a scalar steady-state performance measure  $\theta$ . Multiple performance measures might be, and often are, of interest; any algorithm then can be applied to each scalar performance measure. All such problems are provided time-series data from a given simulation oracle. One replication yields data  $Y_0, Y_1, \dots, Y_N$ . Any two observations, say  $Y_i$  and  $Y_j$ , are not identically distributed because the initial observation  $Y_0$  is not from the steady-state distribution. The output data are possibly autocorrelated because of the state-change logic of the system (e.g., traffic, factory, finance). The sample size  $N$  might be random, such as when the data are based on numbers of transactions in a fixed amount of simulated time. In *multiple-replication* variations of the problem, the data-generation process is repeated, usually independently distributed and always identically distributed.

When using Monte Carlo simulation to estimate a steady-state performance measure  $\theta$ , a classic problem is what to do about the early data that reflect the transient effect of not starting in steady state. The problem is that of choosing the point estimator,  $\hat{\Theta}$ , for  $\theta$ , given the time-series data from the simulation experiment. Typically the form of the point estimator matches the form of the performance measure; for example, if  $\theta$  is a mean, then  $\hat{\Theta}$  is a sample average. The choice of estimator then reduces to the weight to be given to each of the data points, with early (transient) data receiving less weight and later (close to steady state) data receiving more weight. Mostly for simplicity, the weights are often zero for the early data and a positive constant for later data; that is, the early data are deleted and the later data are assumed to be from the steady-state distribution.

Consistent with the view that early data are to be deleted (or at least discounted) is the common-sense idea that the initial state should be chosen to be typical of the steady-state distribution. The steady-state distribution, however, is unknown, because if it were known there is no problem associated with determining the value of  $\theta$ . Therefore, choosing a typical value is not trivially easy. And, often, choosing an atypical value

is much simpler for the practitioner, such as starting with an empty factory. Therefore, the use of a *warm-up period* is common. Determining an appropriate length of the warm-up period is at the heart of this tutorial.

This paper is organized as follows. In Section 2 we discuss four application areas in which initial-transient issues arise and discuss five factors that affect how a practitioner might want to deal with the initial transient. In Section 3 we provide a bibliography of initial-transient publications, with emphasis on literature from the stochastic system simulation community. In Section 4 we discuss performance measures for comparing initial-transient algorithms, with an emphasis on the mean squared error (mse) of the resulting point estimator. In Section 5 we state the static single-replication initial-transient problem and provide Monte Carlo results for the performance of the (magical) algorithm of mse-optimal constant deletion (OCD).

In the remainder of the paper we discuss, and present new results for, the Marginal Standard Error Rule (MSER). Section 6 is a review the MSER statistic; in addition we show that the MSER expected value is asymptotically proportional to the mse; the implication is that, for large sample sizes, the mse-optimal deletion amount can be estimated by minimizing the MSER statistic. In Section 7, we define three algorithms based on the MSER statistic. The first is MSER-GM, an existing algorithm that deletes to the global minimization over the first half of the data; the second is MSER-LLM, deletion to the left-most local minimum; the third is MSER-LLM2, deletion to the left-most local minimum of the local minima. A subsection shows Monte Carlo results comparing these three algorithms to OCD. The final subsection discusses why we do not consider MSER-5, an existing algorithm that pre-whitens the data by taking non-overlapping batch averages of size five. Section 8 contains a summary, contributions, and future research.

Our intention is not to provide an annotated bibliography, briefly discussing every paper related to the initial transient. Rather, we provide some overview and opinion, as well as a reasonably complete bibliography, before focusing on the mse criterion and MSER, both the statistic and associated algorithms.

## 2 CONTEXTS

We categorize contexts of the initial-transient issue in two aspects: first, in Section 2.1, is applications; second, in Section 2.2, is factors. Research and practice communities tend to be based on applications, but equivalent problems arise in various applications because the factor settings are the same.

### 2.1 Applications

The initial-transient issue arises in (at least) four broad-based application areas. In all four, one or more parameters of the steady-state distribution is statistically estimated. Much of the WSC community experiences the initial transient in applications of discrete-event system simulation (e.g., [Law and Kelton 2000](#), [Banks et al. 2010](#), [Bratley, Fox, and Schrage 1987](#), [Leemis and Park 2006](#)). The Bayesian community statistics faces the initial transient in Markov Chain Monte Carlo (MCMC), where the steady-state distribution is the Bayesian posterior distribution (references). The third application, seldom discussed, is in designing stochastic algorithms whose purpose is to estimate optimal solutions or equation roots. The steady-state distribution is degenerate at the optimal solution, a point  $\theta$ ; the data are the solutions visited by the algorithm, and the problem is to decide how to estimate  $\theta$  given that both bias and variance are asymptotically zero. Experiments with humans and animals is a fourth application; here the subjects (e.g., college sophomores and monkeys) undergo task training before data collection begins.

The first two applications are mathematically equivalent, differing only in the simulation model: a factory or traffic intersection in system simulation and the Markov Chain based on the prior and likelihood functions in MCMC. The third application, stochastic algorithms, differs in that the steady-state distribution is degenerate, which suggests that the best point estimator is the last observation; because algorithms stop before variance goes to zero, however, averaging recent observations is sometimes better ([Polyak and Juditsky 1992](#)). Experimentation with live subjects, the fourth application, differs from the first three in that the quality of the data does not necessarily monotonically improve. Live subjects eventually become tired, bored, or irritated, leading them (unintentionally or not) to provide misleading data. In addition, the amount of live-subject training often needs to be decided before the experiment begins, unlike the first three applications where the data can be save to be analyzed after the experiment ends.

### 2.2 Factors

We now assume one of the first three application areas. For any application area, the context of the initial transient depends upon five factors.

First, is the experiment static or dynamic? In a static environment, the time series of data is stored for later analysis; the deletion amount can be determined in light of the entire data sequence, including knowledge

of the length of the sequence. In a dynamic environment, the time to begin data collection is determined “on the fly,” with each observation either ignored or used to update a statistical accumulator, either way lost to future calculations. The dynamic environment is popular in the early literature, while the static environment has become more common as data storage has become faster and cheaper.

Second, is there a single replication or are there multiple replications? If there are multiple replications, is the deletion amount required to be the same for each? If the deletion amounts differ, are the results from the multiple replications required to be combined symmetrically?

Third, what is the form of the performance measure  $\theta$  to be estimated? In general,  $\theta$  is a vector of constants. The components of  $\theta$  can be of many types: means (including probabilities), variances, standard deviations, covariances, correlations, quantiles, and any other property of a steady-state distribution. Even for the same data set, the deletion amount might depend upon the component type. Many, probably most, algorithms for determining the amount of data to delete assume that the performance measure is a mean.

Fourth, is human interaction available? In this, and most, contexts, we think that automatic algorithms, with no human in the loop, are desirable. Nevertheless, some published initial-transient algorithms require human judgment, typically to indicate where the initial transient disappears.

Fifth, is computing effort important? Are computationally intensive algorithms feasible or should the truncation decision be made in (close to)  $O(n)$  time, where  $n$  is the sample size?

### 3 A BIBLIOGRAPHY

Simulation researchers have long been drawn to the initial-transient issue. The literature includes papers covering most (all?) of the combinations of the five factors of the previous section. Some references, loosely grouped, follow.

- Early discussion: Conway, Johnson, and Maxwell (1959), Conway (1963), Emshoff and Sisson (1970), Fishman and Kiviat (1967), Gordon (1969), Fishman (1972), Fishman (1973), Cheng (1976), Lavenberg et al. (1981);
- Theses: Morisaku (1976), Wilson (1977), Kimbler (1987), Murray (1988), McClarnon (1990), Delaney (1995), Spratt (1998), Ghorbani (2004);
- Surveys: Wilson and Pritsker (1978b), Kelton (1980), Pawlikowski (1990), Nelson (1992), White and Robinson (2010);
- Evaluation: Gafarian, Ancker, and Morisaku (1976), Wilson and Pritsker (1977), Gafarian, Ancker, and Morisaku (1977), Gafarian, Ancker, and Morisaku (1978), Wilson and Pritsker (1978a), Cash et al. (1992);
- Comparisons: Kelton and Law (1984), Kimbler and Knight (1987), Ma and Kochhar (1993), Spratt (1998), White, Cobb, and Spratt (2000), Linton and Harmonosky (2002), Sandikçi and Sabuncuoğlu (2006);
- Random start: Kelton and Law (1983), Kelton (1989), Deligönül (1997);
- Random deletion: Glynn and Iglehart (1987), Awad and Glynn (2006);
- Graphical: Welch (1982), Welch (1983);
- Multiple replications: Cheng (1976), Glynn and Heidelberger (1991), Bause and Eickhoff (2002), Bause and Eickhoff (2003);
- Replicate, longrun: Glynn (1987), Whitt (1991), Alexopoulos and Goldsman (2004);
- Chaos theory: Lee and Oh (1994);
- Control charts: Robinson (2002), Rossetti, Li, and Qu (2005), Robinson (2007);
- Kalman filter: Gallagher, K. W. Bauer, and Maybeck (1996);
- Frequency domain: Morrice, Schruben, and Jacobson (1990);
- Time series: Fishman (1971), Richards (1983), Mackulak et al. (2002), Sheth-Voss, Willemain, and Haddock (2005);
- Confidence intervals: Tafazzoli, Steiger, and Wilson (2011);
- AR(1): Fishman (1972), Snell and Schruben (1982), Snell and Schruben (1985), White and Franklin (2010);
- Testing: Schruben (1982), Schruben, Singh, and Tierney (1983), Vassilacopoulos (1989), Yücesan (1993), Goldsman, Schruben, and Swain (1994), Ockerman and Goldsman (1999);
- Detection: Asmussen, Glynn, and Thorisson (1992);
- Variance reduction: Nelson (1990);
- Asymptotics: Glynn (1984), Glynn and Iglehart (1988), Glynn and Whitt (1992), Glynn (1995), Glynn (2005);
- Other methods: Jackway and deSilva (1992);
- Bias-reducing estimators: Hsieh, Iglehart, and Glynn (2004), Awad and Glynn (2007);
- Dynamic: Adlakha and Fishman (1982), Lee, Kyung, and Jung (2009);
- Arbitrary performance measures: Eickhoff, McNickle, and Pawlikowski (2005b),

Eickhoff, McNickle, and Pawlikowski (2007);  
 Quantiles: Eickhoff, McNickle, and Pawlikowski (2005a);  
 Run-length control: Heidelberg and Welch (1983);  
 Multivariate: Schruben (1981);  
 Impact on variance estimators: Ockerman and Goldsman (2008);  
 Ill-stated problem: Schmeiser (2001).

#### 4 ALGORITHM CRITERIA

Schmeiser (2001) uses the initial-transient as an example of an ill-posed problem. In most published papers, the authors do not specify how initial-transient algorithms are to be compared. When a criterion is specified, it is often point-estimator bias; indeed, reference is often made to the “initial-bias problem.” (Recall that bias is  $\text{bias}(\hat{\Theta}, \theta) = E(\hat{\Theta} - \theta)$ , with an unbiased estimator having zero bias.) Deleting data to make bias close to zero is not the problem, however, as illustrated when  $\theta$  is a mean and  $\hat{\Theta}$  is a sample average; in this case deleting all but the last observation would be optimal.

Early algorithms were common-sense heuristics, with no criterion stated. An example is to delete observations until the current observation is neither the minimum nor maximum of the remaining data.

In the late 1970s, Gafarian, Ancker, and Morisaku (1976), Gafarian, Ancker, and Morisaku (1977), and Gafarian, Ancker, and Morisaku (1978) at the University of Southern California were the first to systematically compare initial-transient algorithms. They define the optimal deletion point, say  $k^*$ , and then compare, via Monte Carlo experiments, the (random) amount of data, say  $K$ , deleted by each algorithm. Their definition of  $k^*$  depends on when the expected value of the point estimator is within a specified fraction of the steady-state mean; therefore it is dependent both upon an application having a positive mean (as occurs with queueing systems) and an arbitrarily specified fraction. Their statistical criteria for whether an algorithm is good include the mean and variance of the number of observations deleted. Their use of the mean is reasonable, but as seen below (in Table 2) their use of the variance is not; the ability to delete different amounts of data based on the realization can improve algorithm performance. At best, how well an algorithm estimates a perceived good deletion amount is an indirect criterion; we argue that the focus should be directly on the quality of the point estimator.

Soon after, Wilson (1977) and Wilson and Pritsker (1977) thought carefully about the initial transient, including both the choice of initial condition and the problem of deleting initial data. Wilson and Pritsker (1978b) argue that the focus should be on point-estimator performance, such as mse, rather than on perceived ideal truncation amount. Wilson and Pritsker (1978a) suggest an evaluation procedure for both initial conditions and deletion algorithms; their procedure is elegant, but confounds the comparison by using the criterion of confidence-interval coverage.

We prefer the logic of Fishman (1972) and Snell and Schruben (1985), who think of the initial-transient problem as straight-forward point estimation. They separate the quality of the point estimator from the quality of any statement (such as a confidence interval or an estimated standard error) about the quality of the point estimator. Specifically, they follow the statistical tradition of measuring point-estimator quality with the mean squared error,

$$\text{mse}(\hat{\Theta}, \theta) = E[(\hat{\Theta} - \theta)^2] = \text{bias}^2(\hat{\Theta}, \theta) + \text{var}(\hat{\Theta}),$$

squared bias plus squared standard error. Throughout the rest of this paper, we define mse as the primary criterion for comparing and evaluating initial-transient algorithms.

Although we discuss them only in passing, the other usual secondary criteria apply. A deletion rule should require little computing effort. A rule should be useful to a wide class of applications, both in the type of point estimators (e.g., means, standard deviations, coefficient of variation, quantiles) and the types of data processes (e.g., traffic, factory, finance). A rule should be automatic, in the sense that human judgment is not required, either to set algorithm parameters or to judge (subjectively) the “look” of the output data. A rule should always return a deletion amount; whether the sample size  $n$  is adequate is not part of our initial-transient problem.

#### 5 THE STATIC SINGLE-REPLICATION PROBLEM

In this section, we define the static single-replication initial-transient problem, which is the only problem considered in the rest of this paper. In Section 5.1 we define the problem, in Section 5.2 we discuss the associated mse-optimal constant deletion amount,  $k^*$ , and show Monte Carlo results for the mse associated with deleting  $k^*$  observations. These mse results serve as a baseline for implementable algorithms.

### 5.1 Problem Statement

As considered here, the *static single-replication initial-transient problem* assumes that a single vector of output data  $y_1, y_2, \dots, y_n$  is available as stored values, with the sample size  $n$  known and constant. (The sample size  $N$  might have been random, such as when the data are based on numbers of transactions in a fixed amount of simulated time. For our problem, however, the replication has occurred and sample size is the constant  $n$ .) The problem is to decide the value of  $k$ , the amount of the initial data  $y_1, y_2, \dots, y_k$  to delete, in the calculation of the modified point estimate  $\hat{\Theta}_{n,k}$ . For example, the corresponding sample mean is  $\bar{Y}_{n,k} = \sum_{i=k+1}^n Y_i / (n-k)$ . Just as  $y_1, y_2, \dots, y_n$  is a realization of  $Y_1, Y_2, \dots, Y_n$ , the amount of deleted data  $k$  is a realization of  $K$ , a random variable (defined by the deletion algorithm) that is a function of the output data. Statistically, an algorithm for defining  $K$  in terms of  $Y_1, Y_2, \dots, Y_n$  is good to the extent that it provides a small value of  $\text{mse}(\hat{\Theta}_{n,k}, \theta)$ . A benchmark algorithm, available to researchers but not to practitioners, is to set  $K$  to  $k^* = \text{argmin}_k \text{mse}(\hat{\Theta}_{n,k}, \theta)$ . In discussing statistical performance, we refer to both the mse and the *relative mse* (rmse), defined as  $E_K[\text{mse}(\hat{\Theta}_{n,K}, \theta)] / \text{mse}(\hat{\Theta}_{n,k^*}, \theta)$ .

Despite not being independent and identically distributed (iid), the output data are useful in the sense that they arise from a process whose steady-state distribution has  $\theta$  as a property. Therefore, the data can be useful in the computation of some point estimator  $\hat{\Theta}_n$ . We assume that the data are useful, in the sense of Assumption 1.

**Assumption 1.** *There exists a finite constant  $c$  such that*

$$\lim_{n \rightarrow \infty} n \text{mse}(\hat{\Theta}_n, \theta) = c < \infty. \quad (1)$$

Commonly  $\theta$  is an expected value and the point estimator  $\hat{\Theta}_n$  is the sample mean  $\bar{Y}_n = \sum_{i=1}^n w_i Y_i$ , with weights  $w_i = 1/n$ . The assumed limiting behavior in Equation 1 means that infinite computing effort, as reflected by  $n$ , yields the value of  $\theta$  with zero mse, regardless of initial conditions, varying distributions, and autocorrelations. In practice, of course, computing effort is limited and sampling error causes  $\hat{\Theta}_n$  to differ from  $\theta$ .

### 5.2 Optimal Constant Deletion

To evaluate and compare deletion algorithms, we use the mse of the resulting estimator,  $\text{mse}(\hat{\Theta}_{n,K}, \theta)$ . Here  $K$  is the random number of deleted observations, calculated using the single-replication observations  $Y_0, Y_1, \dots, Y_n$ . In addition to the specific deletion algorithm, the distribution of  $K$  depends on the distribution of the initial observation  $Y_0$ , the data process, and the sample size  $n$ .

As a benchmark for evaluating and comparing algorithms, we compute  $\text{mse}(\hat{\Theta}_{n,k^*}, \theta)$ . In practice, the value of  $k^*$  is unknown, but a researcher comparing algorithms can estimate  $k^*$  with a Monte Carlo experiment that computes  $\hat{\Theta}_{n,k}$  for  $k = 1, 2, \dots, n-1$  for each of  $r$  independent replications of  $Y_0, Y_1, \dots, Y_n$  and identifies  $k^*$  with the value of  $k$  with the smallest  $r^{-1} \sum_{i=1}^r (\hat{\Theta}_{n,k} - \theta)^2$ .

### 5.3 AR(1) Properties

Because many properties are tractable, we base our analysis on first-order autoregressive (AR(1)) data. Here  $Y_i = \theta + \alpha(Y_{i-1} - \theta) + \varepsilon_i$ . Steady state occurs if and only if  $|\alpha| < 1$ . If the errors  $\varepsilon_i$  are independent and normal with zero mean and variance  $\sigma^2$ , then at steady state the mean is  $E(Y) = \theta$  and the variance is  $\sigma_Y^2 = \sigma^2 / (1 - \alpha^2)$ . The initial bias is the value of  $Y_0 - \theta$ , but the process is in steady state only if  $Y_0$  is normally distributed with mean  $\theta$  and variance  $\sigma_Y^2$ . The lag- $h$  autocorrelation is  $\alpha^h$  and the sum of all autocorrelations is  $\gamma_0 = (1 + \alpha) / (1 - \alpha)$ .

Fishman (1972) derives the bias, variance, and mse of the sample average conditional on starting in a specific state  $Y_0$ . We correct here his Equation 12 for the variance of the sample average. (A similar error arises in the mse in his Equation 14.)

$$\text{var}(\bar{Y}_{n,k}) = \frac{\sigma_Y^2}{n-k} \left\{ 1 - \frac{\alpha^{2(k+1)} - \alpha^{2(n+1)}}{(n-k)(1-\alpha^2)} + \frac{2\alpha}{1-\alpha} \left[ \frac{n-k-1}{n-k} + \frac{\alpha^{2(k+1)} - \alpha^{2n}}{(n-k)(1-\alpha^2)} + \frac{\alpha^{n-k} - \alpha + \alpha^{n+k+1} - \alpha^{2n}}{(n-k)(1-\alpha)} \right] \right\}.$$

The sample size  $n$  has negligible effect when  $\alpha^n$  is small. Snell and Schruben (1985) show the asymptotic mse-optimal deletion amount for estimating the mean  $\theta$ . Let  $b = (Y_0 - \theta) / \sigma_Y$ , the initial bias in units of the standard

deviation of  $Y$ . Then  $k^* = 0$  if  $b^2 < 1 + 1/\alpha^2$ ; otherwise,  $k^* = \lfloor (-\ln(b^2 - 1))/(\ln(\alpha^2)) \rfloor$ , where  $\lfloor \cdot \rfloor$  is the floor function. (Notice the type-setting error in the  $k^*$  formula in [Snell and Schruben 1985](#).)

#### 5.4 Performance of Optimal Constant Deletion

As a benchmark, we now consider the performance of mse-optimal constant deletion (OCD) for AR(1) data processes when estimating the steady-state mean. In Table 1, three factors are varied: the (standardized) initial bias  $b = (Y_0 - \theta)/\sigma_Y = Y_0$ , the sample size  $n$ , and the sum of all autocorrelations,  $\gamma_0$ . Each cell has three entries: the OCD deletion amount  $k^*$ , the corresponding value of  $n \text{mse}(\bar{Y}_{n,k^*})$ , and the corresponding value with no deletion,  $n \text{mse}(\bar{Y}_{n,0})$ . For convenience, we set the steady-state mean to  $\theta = 0$ . For clarity, we set the steady-state variance to  $\sigma_Y = n/\gamma_0$ , so that asymptotically (in  $n$ ) the mse cell entries are one. The tables entries are a combination of the Snell-Schruben asymptotic approximation to  $k^*$  and its associated error and Monte Carlo results.

As a second test-bed process, we created an analogous table for M/M/1 queue waiting time. Because of lack of space, we don't include the table, but the message is the same as from Table 1. More generally, the type of process affects performance less directly than the three fundamental factors  $n$ ,  $b$ , and  $\gamma_0$ . We now discuss the AR(1) results of Table 1.

First consider independent data, which correspond to  $\alpha = 0$  and  $\gamma_0 = 1$ . Because the initial bias has no effect on future data,  $k^* = 0$ : deleting no data is mse-optimal. Then  $\bar{Y}_{n,0}$  is unbiased and  $n \text{mse}(\bar{Y}_{n,k^*}) = \text{nvar}(\bar{Y}_{n,0}) = 1$ .

Now consider dependent data. Large initial bias, of course, is where data deletion is most helpful. All three entries, the optimal deletion  $k^*$ , the resulting mse  $n \text{mse}(\bar{Y}_{n,k^*})$ , and the untruncated mse increase with large initial absolute bias  $|Y_0|$ . The role of autocorrelation (as measured by either  $\alpha$  and  $\gamma_0$ ) is explained by the limiting result  $\text{nvar}(\bar{Y}) = \gamma_0 \sigma_Y^2$ ; that is,  $n/\gamma_0$  is the sample size of independent observations that is equivalent to a sample size of  $n$  dependent observations ([Fishman 1972](#)). For example,  $n = 10000$  dependent observations with  $\gamma_0 = 199$  should behave much like  $n = 50$  independent observations. This asymptotic explanation fails, of course, for table entries where effective sample size,  $n/\gamma_0$ , is small. Where the asymptotic explanation holds, however, the three factors of  $n$ ,  $b$ , and  $\gamma_0$  are reduced to the two factors of effective sample size  $n/\gamma_0$  and scaled bias  $b = (Y_0 - \theta)/\sigma_Y$ . Typically, the performance measure is a mean and the effective sample size is reasonably long, so these two factors are fundamental to consider when comparing initial-transient algorithms.

Table 1, and its analog for any other data process, is our benchmark for evaluating initial-transient algorithms. Since mse-optimal constant deletion uses the value of  $k^*$ , which is unknown in practice, creating an initial-transient algorithm that performs better than OCD is unlikely, although not impossible (as mentioned in [White and Minnox \(1994\)](#), p. 217). How could an implementable algorithm perform better? Because OCD is required to delete the same amount of data from every realization. An implementable algorithm can (and will) delete less data when randomness provides a time series with less apparent initial bias. Indeed, in Section 7.2 we show a few factor settings where MSER-based algorithms provide a percent or two smaller mse than OCD.

## 6 THE MSER STATISTIC

[McClarnon \(1990\)](#), in an M.S. thesis advised by Pres White, first advocated the use of the MSER statistic in dealing with the initial transient. In her thesis, Mary McClarnon used the phrase Confidence Maximization Rule (CMR), [White and Minnox \(1994\)](#) use no phrase but select a truncation point that “minimizes the width of the confidence interval about the truncated sample mean, [White \(1995\)](#) and [White \(1997\)](#) use the phrase Marginal Confidence Rule (MCR), and [White, Cobb, and Spratt \(2000\)](#) through [White and Franklin \(2010\)](#) use Marginal Standard Error Rules (MSER). We use MSER, the more-recent phrase.

### 6.1 Definition

MSER has come to refer to a family of initial-transient algorithms based on what we will refer to as the MSER statistic, defined as

$$MSER(n, k) = \frac{S_{n,k}^2}{n - k}$$

and  $S_{n,k}^2 = \sum_{i=1}^n (Y_i - \bar{Y}_{n,k})^2 / (n - k)$ , the sample variance using  $n - k$  rather than  $n - k - 1$ . That is, if the data were iid, the  $MSER(n, k)$  statistic would be the estimated variance of the sample mean of the remaining  $n - k$  observations after the first  $k$  observations were deleted.

As discussed in Section 7, several theses and papers have advocated using the MSER statistic by computing it for  $k = 1, 2, \dots, \kappa$  and deleting initial observations to the value of  $k$  that provides the minimal  $MSER(n, k)$  value. The recent consensus upper limit is  $\kappa = n/2$ ; that is, delete no more than half of the data.

Table 1: Performance of mse-optimal constant deletion,  $k^*$ , when estimating the mean  $\theta$  of standard AR(1) data. The initial bias is  $|b| = |Y_0| \sigma_Y$ . The sum of correlations is  $\gamma_0 = (1 + \alpha)/(1 - \alpha)$ . The cell entries are  $k^*$ ,  $n \text{mse}(\bar{Y}_{n,k^*})$ , and  $n \text{mse}(\bar{Y}_{n,0})$ . Monte Carlo sampling error is negligible, but the low-order digits are not necessarily correct.

$n$	$\gamma_0$	$(\alpha)$	Initial Bias: $b = Y_0/\sigma_Y$				
			0	1	10	100	1000
100	.053	(-.9)	0	0	19	42	63
			1.05	1.09	1.5	2.2	3.8
			1.05	1.09	5.3	424	42300
	1	(0)	0	0	0	0	0
			1	1	1	1	1
			1	1	1	1	1
	19	(.9)	0	0	22	44	66
			0.86	0.90	1.2	1.6	2.5
			0.86	0.90	5.1	427	42630
	199	(.99)	0	0	99	99	99
			0.17	0.37	7.2	674	67326
			0.17	0.37	20	1980	197900
1000	.053	(-.9)	0	0	22	42	65
			1.00	1.00	1.03	1.05	1.08
			1.00	1.00	1.42	43	4232
	1	(0)	0	0	0	0	0
			1	1	1	1	1
			1	1	1	1	1
	19	(.9)	0	0	21	43	65
			0.98	0.99	1.02	1.04	1.06
			0.98	0.99	1.4	44	4264
	199	(.99)	0	0	235	469	711
			0.85	0.89	1.20	1.66	2.64
			0.85	0.89	5.76	493	49246
10000	.005	(-.99)	0	0	230	460	690
			1.00	1.01	1.04	1.06	1.09
			1.00	1.01	1.50	50	5000
	.053	(-.9)	0	0	20	43	65
			1.00	1.00	1.00	1.01	1.02
			1.00	1.00	1.05	5.1	420
	1	(0)	0	0	0	0	0
			1	1	1	1	1
			1	1	1	1	1
	19	(.9)	0	0	20	42	63
			1.00	1.00	1.00	1.00	1.00
			1.00	1.00	1.04	5.19	427
199	(.99)	0	0	226	460	685	
		0.99	1.00	1.02	1.04	1.07	
		0.99	1.00	1.5	50	4900	
1999	(.999)	0	0	2400	4800	7100	
		0.83	0.87	1.2	1.6	2.5	
		0.83	0.87	5.7	500	50000	

Since McClarnon (1990), several authors have emphasized, or at least mentioned, that minimizing the MSER statistic was a substitute for minimizing the mse. But as seen in Table 1, the benchmark performance of OCD depends on autocorrelation. So how can algorithms that ignore autocorrelation be mse-competitive?

As pointed out by Franklin and White (2008), MSER has not been widely embraced. We suspect that part of the confusion is the claim that minimizing MSER is closely related to minimizing the width of the resulting confidence interval on the steady-state mean. Because the width of an appropriate confidence interval for the mean  $\theta$  is proportional to  $\gamma_0$ , ignoring autocorrelations and minimizing confidence-interval width easily causes confusion.

## 6.2 MSER and Mean Squared Error

That MSER can provide good mse while ignoring autocorrelations depends directly on the effective-sample-size result:  $n\text{var}(\bar{Y}(n,k)) = \gamma_0 \sigma_Y^2$ . When there is no bias, the expected value of  $\text{MSER}(n,k)$  is  $\sigma_Y^2/(n-k)$ , since it doesn't "see" the autocorrelation factor  $\gamma_0$ . For large effective sample size,  $(n-k)/\gamma_0$ , the effect of autocorrelation is essentially constant, so the location  $k$  of the minimum value of  $\text{MSER}(n,k)$  is asymptotically unaffected by autocorrelation. Therefore, the naive-looking and simple-to-compute MSER statistic, with its inability to consider autocorrelations, can work well.

We make the previous paragraph's argument more specific with two lemmas and a theorem. Consider a data process  $Y_i$  with steady-state mean  $\theta$ . Let  $b_i = E(Y_i) - \theta$ ,  $\sigma_i^2 = \text{var}(Y_i)$ , and  $\rho_{i,j} = \text{corr}(Y_i, Y_j)$ . The first lemma expresses mse in terms of the individual biases, variances, and correlations. The second lemma expresses the expected value of the MSER statistic in terms of the biases and variances. The theorem says that their asymptotic ratio is the sum of autocorrelations,  $\gamma_0$ .

**Lemma 1.**  $\text{mse}[\bar{Y}_{n,k}, \theta] = E[(\bar{Y}_{n,k} - \theta)^2] = (n-k)^{-2} \sum_{k+1}^n \sum_{j=k+1}^n (b_i b_j + \sigma_i \sigma_j \rho_{i,j})$

**Lemma 2.**  $E[\text{MSER}(n,k)] = (n-k)^{-2} [\sum_{k+1}^n (b_i^2 + \sigma_i^2) - (n-k)\text{mse}[\bar{Y}_{n,k}, \theta]]$

Compare the results of the two lemmas. The mse is a function of all combinations of bias cross products and correlations between observations; the expected value of the MSER statistic is not. The only structure assumed about the various biases, variances, and correlations is that implied indirectly by Assumption 1: that the data are useful for estimating the performance measure  $\theta$ . Theorem 1 says that MSER, despite ignoring the bias cross products and the correlations, is asymptotically tied to the mse.

**Theorem 1.** For every data process having steady-state with mean  $\theta$ , lag- $h$  autocorrelations  $\rho_h$ , and finite sum of autocorrelations  $\gamma_0 = 1 + 2 \sum_{h=1}^{\infty} \rho_h$ ,

$$\lim_{n \rightarrow \infty} \frac{\text{mse}[\bar{Y}_{n,k}, \theta]}{E[\text{MSER}(n,k)]} = \gamma_0,$$

for every  $k = 0, 1, 2, \dots$

White and Robinson (2009) suggest that "MSER works well because it minimizes an approximation to the mean-squared error in the estimated steady-state mean." Theorem 1 corrects the suggestion to say that MSER tends to be proportional to mse. Because they are asymptotically proportional, their minima tend to lie close to the same deletion point  $k$ .

## 7 MSER AGLORITHMS

Since McClarnon (1990), various papers and theses (White and Minnox 1994, Rossetti, Delaney, and White 1995, White 1995, White 1997, Spratt 1998, White, Cobb, and Spratt 2000, Franklin and White 2008, Hoad, Robinson, and Davies 2008, White and Robinson 2009 and White and Franklin 2010) have analyzed and advocated initial-transient algorithms based on the MSER statistic.

### 7.1 Algorithm Design

Given time-series data,  $Y_1, Y_2, \dots, Y_n$ , all MSER-algorithm variations compute the  $\text{MSER}(n,k)$  statistic for every  $k$  in some set  $\mathcal{K} \subset \{0, 1, \dots, n-1\}$  and then delete data to some point  $\hat{k}^*$  for which  $\text{MSER}(n,k)$  is small. The four design decisions are (1) whether, and how, to preprocess the data, (2) the choice of the set  $\mathcal{K}$ , and (3) the exact form of the MSER statistic, and (4) the selection of the deletion amount,  $\hat{k}^*$ .

The first design decision is data preprocessing. The prominent example is prewhitening by creating averages of non-overlapping batches of size  $m$  (especially  $m = 5$ , as advocated by Spratt (1998) and White, Cobb, and Spratt (2000), and recent papers) via the algorithm referred to as MSER-5. The choice of  $m = 5$  is claimed to perform well, especially compared to no prewhitening,  $m = 1$ . Any one-size-fits-all preprocessing, however, leads to a contradiction: if the preprocessed data are better in some sense, then why not preprocess the preprocessed data? If preprocessing is to make sense, then its form needs to be based on an analysis of the given data  $Y_1, Y_2, \dots, Y_n$ . Another point is that the use of non-overlapping batches is suboptimal to using overlapping batches, which leave no orphaned observations at the end of the data series, which cause no graininess in the analysis, and which still requires only  $O(n)$  computation.

The second design decision is the choice of  $\mathcal{K}$ , the set of  $k$  values to calculate  $\text{MSER}(n,k)$ . All values of  $k$  can be considered while maintaining  $O(n)$  computation, so choosing  $\mathcal{K} = \{0, 1, 2, \dots, n-1\}$  is not unreasonable. Because of the positive correlation between MSER values, computing time could be saved by considering no more

than, say, 100  $k$  values spread over the set  $\{0, 1, 2, \dots, n-1\}$ . A common suggestion in the recent literature is to choose  $\mathcal{K} = \{0, 1, 2, \dots, \lfloor n/2 \rfloor\}$ , so that no more than half of the data can be deleted.

The third design decision is the form of the MSER statistic. One could argue about whether to divide by  $n-k$  or  $n-k-1$ .

The fourth decision, how to use the observed MSER curve to select the deletion amount, is the most interesting. The curve is composed of the calculated values of  $\text{MSER}(n, k)$  for all  $k \in \mathcal{K}$ . Small values are good.

We denote by MSER-GM the algorithm that selects the global minimum, the original form of MSER algorithms. Because the  $\text{MSER}(n, k)$  statistic has larger sampling error when  $k$  is close to  $n$ , the global minimum in  $\{0, 1, 2, \dots, n-1\}$  is sometimes erroneously close to  $n$ , which is why MSER-GM is often restricted to  $\{0, 1, 2, \dots, \lfloor n/2 \rfloor\}$ . We denote by MSER-LLM a new algorithm. It does no preprocessing, it considers  $\mathcal{K} = \{0, 1, 2, \dots, n-1\}$ , it divides by  $n-k-1$ , and it selects  $\hat{k}^*$  to be the location of the left-most local minimum of the MSER curve. We denote by MSER-LLM2 another new algorithm. It is identical to MSER-LLM, except that it selects  $\hat{k}^*$  to be the location of the left-most local minimum of the local minima of the MSER curve. (Yes, you might want to draw an example.)

The definitions of the three algorithms imply that for every realization  $\hat{k}_{LLM}^* \leq \hat{k}_{LLM2}^* \leq \hat{k}_{GM}^*$ , except when MSER-GM truncates the deletion at half the data because of its choice of  $\mathcal{K}$ . Because they are designed to delete less data than MSER-GM, both MSER-LLM and MSER-LLM2 can safely consider all deletion points. In particular, if the initial-transient effect remains strong throughout the data set, MSER-LLM and MSER-LLM2 can delete all but the last observation.

## 7.2 Performance Comparison

Table 2, analogous to Table 1, shows relative-mse performance for each of the three MSER algorithms. When the effective sample size,  $n/\gamma_0$  is large, all three MSER algorithms provide mse within a percent of that provided by mse-optimal constant deletion (OCD), with MSER-LLM and MSER-LLM2 performing a bit better. When the effective sample size is small and the initial bias is large, such as the lower right cell with  $n/\gamma_0 = 5$ , MSER-LLM and MSER-LLM2 perform much better than MSER-GM but with more than double the mse provided by OCD.

The mse-performance for negative autocorrelations and small effective sample sizes is only sporadic good. As an example of surprisingly good, consider  $n = 100$ ,  $\alpha = -0.9$ , for all levels of initial bias; MSER-LLM performs better than OCD (although by a negligible amount). Extrapolating these negative-autocorrelation results seems risky. (Performance of the three MSER algorithms is similar for M/M/1 time-in-queue.)

## 8 SUMMARY, CONCLUSIONS, AND FUTURE RESEARCH

In summary, the initial transient is ever present in statistical experiments intended to estimate steady-state performance measures. Of the four application areas, we focus on the literature of steady-state system simulation, for which we provide a bibliography. We argue for using mean squared error as the primary statistical criterion for comparing deletion algorithms. We show that the MSER statistic is asymptotically proportional to the sample mean's mse, so that deleting based on the MSER statistic is reasonable, even for autocorrelated processes. Finally, we compare the mse obtained from three variations, of which MSER-LLM and MSER-LLM2 are new, of deletion algorithms based on the MSER statistic; for large effective sample sizes, they yield mse values within a percent of mse-optimal constant deletion. We conclude that, if mse is the goal and if the performance measure is a mean, then the MSER statistic is a solid foundation for initial-transient algorithms. Both MSER-LLM and MSER-LLM2 perform better than MSER-GM across sample sizes, sum of autocorrelations, and initial bias, but all three perform well when the effective sample size is large.

Creating algorithms based on the MSER statistic is easy. For preprocessing, a fixed number of nonoverlapping batches (rather than a fixed batch size) makes sense (Schmeiser 1982); using overlapping batches, rather than nonoverlapping batches, makes sense (Meketon and Schmeiser 1984). Multiple batch sizes could be used, selecting deletion amounts that are indicated by multiple batch sizes. Linear combinations of the results from multiple algorithms could be used. But, given that the three existing MSER algorithms do so well compared to OCD for large effective sample sizes, any substantial performance gain would be when the effective sample size is small.

## ACKNOWLEDGMENTS

We thank William Franklin, Kathryn Hoad, Stewart Robinson and Pres White for discussions about MSER, which they have advocated for some years. We also acknowledge financial support from the Office of Naval Research, via contract N00014-08-1-0066, which supported the second author's visit to Virginia Tech during spring 2010.

## REFERENCES

Adlakha, V. G., and G. S. Fishman. 1982. Starting and stopping rules for simulations using a priori information. *European Journal of Operational Research* 10:379-394.

Table 2: Performance of MSER–GM, MSER–LLM, and MSER–LLM2 for AR(1) data. The three entries in each cell are the relative (compared to k\*) mse values; that is, a value of one means that the MSER algorithm provides an mse equal to OCD. MSER–GM entries with an asterisk indicate cases for which the truncation limitation (of no more than half the data) affected MSER–GM performance.

$n$	$\gamma_0$	$(\alpha)$	Initial Bias: $b = Y_0/\gamma_0$					
			0	1	10	100	1000	
100	0.053	(-.9)	1.11	1.12	1.07	1.02*	5.00*	GM
			0.9997	0.998	0.98	0.97	0.97	LLM
			1.06	1.07	1.10	1.24	1.42	LLM2
	1	(0)	1.04	1.04	1.04	1.04	1.04	GM
			1.01	1.01	1.01	1.01	1.01	LLM
			1.02	1.02	1.02	1.02	1.02	LLM2
	19	(.9)	1.40	1.33	1.19	1.08*	8.40*	GM
			1.01	1.12	1.15	1.19	1.37	LLM
			1.18	1.21	1.25	1.28	1.29	LLM2
	199	(.99)	1.67	1.66	1.67*	1.67*	1.71*	GM
			1.15	1.97	1.06	1.01	1.01	LLM
			1.48	1.67	1.02	1.01	1.01	LLM2
1000	0.053	(-.9)	1.015	1.013	1.010*	1.015	1.016	GM
			1.000	0.999	1.000	1.002	1.000	LLM
			1.004	1.003	1.001	1.008	1.004	LLM2
	1.0	(0)	1.005	1.005	1.005	1.005	1.005	GM
			1.001	1.001	1.001	1.001	1.001	LLM
			1.005	1.005	1.005	1.005	1.005	LLM2
	19	(.9)	1.024	1.023	1.040	1.036	1.073	GM
			1.000	1.006	1.016	1.015	1.016	LLM
			1.003	1.009	1.022	1.008	1.018	LLM2
	199	(.99)	1.44	1.22	1.09*	2.43*	85*	GM
			1.00	1.15	1.22	1.37	1.95	LLM
			1.02	1.13	1.19	1.29	1.67	LLM2
10000	0.0050	(-.99)	1.009	1.028	1.033	1.029	1.046	GM
			1.007	1.010	1.012	1.009	1.016	LLM
			1.006	1.013	1.006	1.013	1.016	LLM2
	0.053	(-.9)	1.000	1.000	1.006	1.004	1.006	GM
			1.001	1.000	1.000	1.002	1.002	LLM
			1.002	0.999	1.004	1.003	1.004	LLM2
	1	(0)	1.00	1.00	1.00	1.00	1.00	GM
			1.00	1.00	1.00	1.00	1.00	LLM
			1.00	1.00	1.00	1.00	1.00	LLM2
	19	(.9)	1.003	1.008	1.005	1.006	1.004	GM
			1.000	1.001	1.002	1.002	1.002	LLM
			1.001	1.000	1.002	1.003	1.001	LLM2
199	(.99)	1.056	1.064	1.068	1.055	1.043	GM	
		1.000	1.014	1.015	1.028	1.013	LLM	
		1.002	1.008	1.019	1.030	1.017	LLM2	
1999	(.999)	1.36	1.15	1.11*	13.8*	456*	GM	
		1.00	1.20	1.27	1.66	2.70	LLM	
		1.03	1.18	1.25	1.59	2.52	LLM2	

Alexopoulos, C., and D. Goldsman. 2004. To batch or not to batch. *ACM Transactions on Modeling and Computer Simulation* 14:76–114.

Asmussen, S., P. W. Glynn, and H. Thorisson. 1992. Stationarity detection in the initial transient problem. *ACM Transactions on Modeling and Computer Simulation* 2:130–157.

Awad, H. P., and P. W. Glynn. 2006. On an initial transient deletion rule with rigorous theoretical support. In *Proceedings of the 2006 Winter Simulation Conference*, ed. L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, 186–191. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

- Awad, H. P., and P. W. Glynn. 2007. On the theoretical limit of low-bias steady-state estimators. *ACM Transactions on Modeling and Computer Simulation* 17:1–30.
- Banks, J., J. S. Carson, B. L. Nelson, and D. M. Nicol. 2010. *Discrete-event system simulation*. 5th ed. Upper Saddle River, New Jersey: Prentice-Hall, Inc.
- Bause, F., and M. Eickhoff. 2002. Initial transient period detection using parallel replications. In *Proceedings of the 14th European Simulation Symposium*, 85–92.
- Bause, F., and M. Eickhoff. 2003. Truncation point estimation using multiple replications in parallel. In *Proceedings of the 2003 Winter Simulation Conference*, ed. S. Chick, P. Sánchez, D. Ferrin, and D. J. Morrice, 414–421. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Bratley, P., B. Fox, and L. Schrage. 1987. *A guide to simulation*. 2nd ed. New York: Springer Verlag.
- Cash, C. R., D. G. Dippold, J. M. Long, B. L. Nelson, and W. P. Pollard. 1992. Evaluation of tests for initial-condition bias. In *Proceedings of the 1992 Winter Simulation Conference*, ed. J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, 577–585. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Cheng, R. C. H. 1976. A note on the effect of initial conditions on a simulation run. *Operational Research Quarterly* 27:467–470.
- Conway, R. W. 1963. Some tactical problems in digital simulation. *Management Science* 10:47–61.
- Conway, R. W., B. M. Johnson, and W. L. Maxwell. 1959. Some problems of digital systems simulation. *Management Science* 6:92–110.
- Delaney, P. J. 1995. *Control of initialization bias in queueing simulations using queueing approximations*. M.S. thesis, School of Engineering and Applied Science, University of Virginia, Charlottesville, Virginia. Available from University of Virginia Library, lib-lend@virginia.edu, 434–982–3094.
- Deligönül, Z. S. 1997. Antithetic bias reduction for discrete-event simulations. *Journal of the Operational Research Society* 38:431–437.
- Eickhoff, M., D. C. McNickle, and K. Pawlikowski. 2005a. Depiction of transient performance measures using quantile estimation. In *Proceedings of the 19th European Conference on Modeling and Simulation*, 358–363. Riga, Latvia.
- Eickhoff, M., D. C. McNickle, and K. Pawlikowski. 2005b. Efficient truncation point estimation for arbitrary performance measures. In *Proceedings of the Third International Industrial Simulation Conference*, 5–12. Berlin: ISC'2005.
- Eickhoff, M., D. C. McNickle, and K. Pawlikowski. 2007. Detecting the duration of initial transient in steady state simulation of arbitrary performance measures. In *Proceedings of the 19th European Conference on Modeling and Simulation*, ???–??? Nantes, France: Valuetools '07.
- Emshoff, J. R., and R. L. Sisson. 1970. *Design and use of computer simulation models*. MacMillan.
- Fishman, G. S. 1971. Estimating sample size in computer simulation experiments. *Management Science* 18:21–38.
- Fishman, G. S. 1972. Bias considerations in simulation experiments. *Operations Research* 20:785–790.
- Fishman, G. S. 1973. *Concepts and methods in discrete event digital simulation*. New York: John Wiley & Sons.
- Fishman, G. S., and P. Kiviat. 1967. The analysis of simulation generated time series. *Management Science* 13:525–557.
- Franklin, W. W., and K. P. White. 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*, ed. S. J. Mason, R. R. Hill, L. M onch, O. Rose, T. Jefferson, and J. W. Fowler, 541–546. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Gafarian, A. V., C. J. Ancker, and T. Morisaku. 1976. The problem of the initial transient in digital computer simulation. In *Proceedings of the 1976 Winter Simulation Conference*, ed. H. J. Highland, T. J. Schriber, and R. G. Sargent, 49–51. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Gafarian, A. V., C. J. Ancker, and T. Morisaku. 1977. The problem of the initial transient with respect to mean value in digital computer simulation and the evaluation of some proposed solutions. Technical Report 77–1, Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, California.
- Gafarian, A. V., C. J. Ancker, and T. Morisaku. 1978. Evaluation of commonly used rules for detecting 'steady state' in computer simulation. *Naval Research Logistics Quarterly* 25:511–529.
- Gallagher, M. A., J. K. W. Bauer, and P. S. Maybeck. 1996. Initial data truncation for univariate output of discrete-event simulations using the kalman filter. *Management Science* 42:559–575.
- Ghorbani, B. 2004. *The issue of initial transient in sequential steady-state simulation*. M.S. thesis, Department of Computer Science, University of Canterbury. referenced in Eickhoff, McNickle and Pawlikowski 2005.
- Glynn, P. W. 1984. Some asymptotic formulas for markov chains with applications to simulation. *Journal of Statistical Computation and Simulation* 19:97–112.
- Glynn, P. W. 1987. Limit theorems for the method of replication. *Stochastic Models* 4:343–350.
- Glynn, P. W. 1995. Some new results on the initial transient problem. In *Proceedings of the 1995 Winter Simulation Conference*, ed. C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman, 165–170. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Glynn, P. W. 2005. Initial transient problem for steady-state output analysis. In *Proceedings of the 2005 Winter Simulation Conference*, ed. M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 739–740. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

- Glynn, P. W., and P. Heidelberger. 1991. Analysis of initial transient deletion for replicated steady-state simulations. *Operations Research Letters* 8:437–443.
- Glynn, P. W., and D. L. Iglehart. 1987. A new initial bias deletion rule. In *Proceedings of the 1987 Winter Simulation Conference*, ed. A. Thesen, H. Grant, and W. D. Kelton, 318–319. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Glynn, P. W., and D. L. Iglehart. 1988. Conditions under which a markov chain converges to its steady-state in finite time. *Probability in the Engineering and Informational Sciences* 2:377–382.
- Glynn, P. W., and W. Whitt. 1992. The asymptotic efficiency of simulation estimators. *Operations Research* 40:505–520.
- Goldman, D., L. W. Schruben, and J. J. Swain. 1994. Tests for transient means in simulated time series. *Naval Research Logistics* 41:171–187.
- Gordon, G. 1969. *System simulation*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Heidelberger, P., and P. D. Welch. 1983. Simulation run length control in the presence of an initial transient. *Operations Research* 31:1109–1144.
- Hoad, K., S. Robinson, and R. Davies. 2008. Automating warm-up length estimation. In *Proceedings of the 2008 Winter Simulation Conference*, ed. S. J. Mason, R. R. Hill, L. M. Onch, O. Rose, T. Jefferson, and J. W. Fowler, 532–540. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Hsieh, M.-H., D. L. Iglehart, and P. W. Glynn. 2004. Empirical performance of bias-reducing estimators for regenerative steady-state simulations. *ACM Transactions on Modeling and Computer Simulation* 14:325–343.
- Jackway, P. T., and B. M. deSilva. 1992. A methodology for initialization bias reduction in computer simulation output. *Asia-Pacific Journal of Operational Research* 9:87–100.
- Kelton, W. D. 1980. The startup problem in discrete-event simulation. Technical Report 80–1, Department Of Industrial Engineering, University Of Wisconsin, Madison, Wisconsin.
- Kelton, W. D. 1989. Random initialization methods in simulation. *IIE Transactions* 21:355–367.
- 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.
- Kelton, W. D., and A. M. Law. 1984. An analytical evaluation of alternative strategies in steady-state simulation. *Operations Research* 32:169–184.
- Kimble, D. L. 1987. *Applications of exponential smoothing to the initial transient problem in digital simulation*. M.S. thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.
- Kimble, D. L., and B. D. Knight. 1987. A survey of current methods for the elimination of initialization bias in digital simulation. In *Proceedings of the 20th Annual Simulation Symposium on Simulation*, Volume 20, 133–152. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Lavenberg, S., G. Fishman, W. Kelton, L. Schruben, and P. Welch. 1981. The initial transient in steady state simulation. In *Proceedings of the 13th Winter Simulation Conference*, ed. T. I. Ören, 113–120. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Law, A. M., and W. D. Kelton. 2000. *Simulation modeling & analysis*. 3rd ed. New York: McGraw-Hill, Inc.
- Lee, Y.-H., K.-H. Kyung, and C.-S. Jung. 2009. On-line determination of steady state in simulation outputs. *Computers and Industrial Engineering* 33:805–808.
- Lee, Y.-H., and H.-S. Oh. 1994. Detecting truncation point in steady-state simulation using chaos theory. In *Proceedings of the 1994 Winter Simulation Conference*, ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila, 353–360. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Leemis, L. M., and S. K. Park. 2006. *Discrete-event simulation: A first course*. Upper Saddle River, New Jersey: Prentice Hall.
- Linton, J. R., and C. M. Harmonosky. 2002. A comparison of selective initialization bias elimination methods. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 1951–1957. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Ma, X., and A. K. Kochhar. 1993. A comparison study of two tests for detecting initialization bias in simulation output. *Simulation* 61:94–101.
- Mackulak, G. T., S. Park, J. W. Fowler, S. E. Leach, and J. B. Keats. 2002. An effective truncation heuristic for bias reduction in simulation output. *Simulation* 78:643–654.
- McClarnon, M. A. 1990. *Detection of steady state in discrete event dynamic systems: An analysis of heuristics*. M.S. thesis, School of Engineering and Applied Science, University of Virginia, Charlottesville, Virginia. Available from University of Virginia Library, lib-lend@virginia.edu, 434–982–3094.
- Meketon, M. S., and B. W. Schmeiser. 1984. Overlapping batch means: Something for nothing? In *Proceedings of the 1984 Winter Simulation Conference*, 227–230. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Morisaku, T. 1976. *Techniques for data truncation in digital computer simulation*. Ph.D. thesis, Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, California.

- Morrice, D., L. W. Schruben, and S. Jacobson. 1990. Initial transient effects in the frequency domain. In *Proceedings of the 1990 Winter Simulation Conference*, 357–359. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Murray, J. R. 1988. *Stochastic initialization in steady-state simulations*. Ph.D. thesis, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan.
- Nelson, B. L. 1990. Variance reduction in the presence of initial-condition bias. *IIE Transactions* 22:340–350.
- Nelson, B. L. 1992. Statistical analysis of simulation results. In *Handbook of Industrial Engineering* (2nd ed.), ed. G. Salvendy. New York: Wiley.
- Ockerman, D. H., and D. Goldsman. 1999. Student t-tests and compound tests to detect transients in simulated time series. *European Journal of Operational Research* 116:681–691.
- Ockerman, D. H., and D. Goldsman. 2008. The impact of transients on simulation variance estimators. In *Proceedings of the 1997 Winter Simulation Conference*, ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson, 234–239. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Pawlikowski, K. 1990. Steady-state simulation of queueing processes: A survey of problems and solutions. *ACM Computing Surveys* 22:123–170.
- Polyak, B., and A. B. Juditsky. 1992. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization* 30:838–855.
- Richards, J. E. 1983. *Minimizing run length in simulations subject to initial transients*. Ph.D. thesis, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.
- Robinson, S. 2002. A statistical process control approach for estimating the warm-up period. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 439–446. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Robinson, S. 2007. A statistical process control approach to selecting a warm-up period for a discrete-event simulation. *European Journal of Operational Research* 176:332–346.
- Rossetti, M. D., P. J. Delaney, and K. P. White. 1995. Generalizing the half-width minimization heuristic for mitigating initialization bias. In *International Conference on Systems, Man, and Cybernetics*, 212–216. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Rossetti, M. D., Z. Li, and P. Qu. 2005. Exploring exponentially weighted moving average control charts to determine the warm-up period. In *Proceedings of the 2005 Winter Simulation Conference*, ed. M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 771–780. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Sandikçi, B., and I. Sabuncuoğlu. 2006. Analysis of the behavior of the transient period in non-terminating simulations. *European Journal of Operational Research* 173:252–267.
- Schmeiser, B. W. 1982. Batch size effects in the analysis of simulation output. *Operations Research* 30:556–568.
- Schmeiser, B. W. 2001. Some myths and common errors in simulation experiments. In *Proceedings of the 2001 Winter Simulation Conference*, 39–46. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Schruben, L. W. 1981. Control of initialization bias in multivariate simulation response. *Communications of the ACM* 24:246–252.
- Schruben, L. W. 1982. Detecting initialization bias in simulation output. *Operations Research* 30:569–590.
- Schruben, L. W., H. Singh, and L. Tierney. 1983. Optimal tests for initialization bias in simulation output. *Operations Research* 31:1167–1178.
- Sheth-Voss, P. A., T. R. Willemain, and J. Haddock. 2005. Estimating the steady-state mean from short transient simulations. *European Journal of Operational Research* 162:403–417.
- Snell, M., and L. W. Schruben. 1985. Weighting simulation data to reduce initialization effects. *IIE Transactions* 17:354–363.
- Snell, M., and L. W. Schruben. April 1979, 1982. Weighting simulation data to reduce initialization effects. Technical Report 395, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York.
- Spratt, S. C. 1998. *An evaluation of contemporary heuristics for the startup problem*. M.S. thesis, School of Engineering and Applied Science, University of Virginia, Charlottesville, Virginia. Available from University of Virginia Library, lib-lend@virginia.edu, 434–982–3094.
- Tafazzoli, A., N. M. Steiger, and J. R. Wilson. 2011. N-Skart: A nonsequential skewness- and autoregression-adjusted batch-means procedure for simulation analysis. *IEEE Transactions on Automatic Control* 56. forthcoming.
- Vassilacopoulos, G. 1989. Testing for initialization bias in simulation output. *Simulation* 52:151–153.
- Welch, P. D. 1982. A graphical approach to the initial transient problem in steady state simulations. In *Proceedings of the 10th IMACS World Conference on Systems Simulation, and Scientific Computation*, ed. M. Ruschitzka, 219–221. Montreal: North Holland.
- Welch, P. D. 1983. The statistical analysis of simulation results. In *The Computer Performance Modeling Handbook*, ed. S. S. Lavenberg, 268–328. New York: Academic Press.
- White, K. P. 1995. A simple rule for mitigating initialization bias in simulation output: Comparative results. In *International Conference on Systems, Man, and Cybernetics*, 206–211. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

- White, K. P. 1997. An effective truncation heuristic for bias reduction in simulation output. *Simulation* 69:323–334.
- White, K. P., 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*, ed. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 755–760. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- White, K. P., and W. W. Franklin. 2010. Parameteric expression for MSER with geometrically decaying bias. In *Proceedings of the 2010 Winter Simulation Conference*, ed. B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc. forthcoming.
- White, K. P., and M. A. Minnox. 1994. Minimizing initialization bias in simulation output using a simple heuristic. In *International Conference on Systems, Man, and Cybernetics*, 215–220. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- White, K. P., and S. Robinson. 2009. The problem of the initial transient (again), or why MSER works. In *Proceedings of the 2000 Inform Simulation Society Research Workshop*, ed. L. H. Lee, M. E. Kuhl, J. W. Fowler, and S. Robinson, 90–95. Institute for Operations Research and the Management Sciences.
- White, K. P., and S. Robinson. 2010. The initialization problem in discrete-event simulation. In *Encyclopedia of Operations Research and Management Science*, ed. J. Cochran. New York: John Wiley & Sons. forthcoming.
- Whitt, W. 1991. The efficiency of one long run versus independent replications in steady-state simulation. *Management Science* 77:645–666.
- Wilson, J. R. 1977. *A procedure for evaluating startup policies in simulation experiments*. M.S. thesis, School of Industrial Engineering, Purdue University, West Lafayette, Indiana.
- Wilson, J. R., and A. A. B. Pritsker. 1977. A procedure for evaluating startup policies in simulation experiments. In *Proceedings of the 1977 Summer Simulation Computer Conference*, 83–89. Chicago, Illinois.
- Wilson, J. R., and A. A. B. Pritsker. 1978a. Evaluation of startup policies in simulation experiments. *Simulation* 31:79–89.
- Wilson, J. R., and A. A. B. Pritsker. 1978b. A survey of research on the simulation startup problem. *Simulation* 31:55–58.
- Yücesan, E. 1993. Randomization tests for initialization bias in simulation output. *Naval Research Logistics Quarterly* 40:643–663.

#### AUTHOR BIOGRAPHIES

**RAGHU PASUPATHY** is an assistant professor in the Industrial and Systems Engineering Department at Virginia Tech. His research interests lie broadly in Monte Carlo methods with a specific focus on simulation optimization and stochastic root finding. He is a member of INFORMS, IIE, and ASA, and serves as an Associate Editor for ACM TOMACS and INFORMS Journal on Computing. His e-mail address is [pasupath@vt.edu](mailto:pasupath@vt.edu) and his web page is <https://filebox.vt.edu/users/pasupath/pasupath.htm>.

**BRUCE SCHMEISER** is Professor of Industrial Engineering at Purdue University. His research interests center on Monte Carlo simulation experimentation, including methods for input modeling, random-variate generation, variance reduction, output analysis, and simulation-based stochastic root finding and optimization. He served on the WSC Board from 1982 until 2000, was WSC Board Chair in 2000 and 1983 WSC Program Chair. He is a fellow of Inform and IIE. His e-mail address is [bruce@purdue.edu](mailto:bruce@purdue.edu) and his web page is [cobweb.ecn.purdue.edu/~bruce/](http://cobweb.ecn.purdue.edu/~bruce/).