AN ABSTRACT OF THE THESIS OF

Lori K. Baxter for the degree of Master of Science in Industrial and Manufacturing Engineering presented on June 4, 1990.

Title: Truncation Rules in Simulation Analysis: Effect of Batch Size, Time Scale and Input Distribution on the Application of Schriber's Rule.

Abstract Approved: Redacted for Privacy

Sabah U. Randhawa

The objective of many simulations is to study the steady-state behavior of a nonterminating system. The initial conditions of the system are often atypical because of the complexity of the system. Simulators often start the simulation with the system empty and idle, and truncate, or delete, some quantity of the initial observations to reduce the initialization bias.

This paper studies the application of Schriber's truncation rule to a queueing model, and the effects of parameter selection. Schriber's rule requires the simulator to select the parameters of batch size, number of batches, and a measure of precision. In addition,

Schriber's rule assumes the output is a time series of discrete observations. Previous studies of Schriber's rule have not considered the effect of variation in the time scale (time between observations).

The performance measures for comparison are the mean squared error and the half-length of the confidence interval. The results indicate that the time scale and batch size are significant parameters, and that the number of batches has little effect on the output. A change in the distribution of service time did not alter the results. In addition, it was determined that multiple replicates should be used in establishing the truncation point instead of a single run, and the simulator should carefully consider the choice of time scale for the output series and the batch size.

Truncation Rules in Simulation Analysis: Effect of Batch Size, Time Scale and Input Distribution on the Application of Schriber's Rule.

by

Lori K. Baxter

A THESIS

Submitted to
Oregon State University

in partial fulfillment of
the requirements for the degree of

Master of Science

Completed June 4, 1990
Commencement June 1991

APPROVED:

Redacted for Privacy

Associate Professor of Industrial and Manufacturing Engineering in charge of major

Redacted for Privacy

Chairman of Department of Industrial and Manufacturing Engineering

Redacted for Privacy

Dean of Graduate School

Date thesis is presented June 4, 1990

TABLE OF CONTENTS

Page

LIST OF FIGURES

<u>Figure</u> <u>Page</u>

LIST OF TABLES

Table | Page

Truncation Rules in Simulation Analysis: Effect of Batch Size, Time Scale and Input Distribution on the Application of Schriber's Rule.

I. Introduction and Background

A simulation is a representation of the operation of a process or system over time. The behavior of the system as it changes over time is studied by developing a simulation study, usually consisting of a simulation model that is excercised under multiple configurations. The simulation model usually includes a set of assumptions about the operation of the system. Once developed and validated, the study can be used to investigate a variety of changes to the system, without changing the system itself. In this way, a simulation study can be used as an analysis tool for predicting the effects of change to an existing system, and it can also be used for proposed systems. The data generated by the simulation model is used to estimate the performance of the system.

A. Types of Simulation Studies

There are two basic types of simulation studies, terminating and nonterminating. A terminating, or transient simulation has a specified duration of time T_E, where E is a specified event or set of events (e.g. a specified time period has elapsed, or a specified number of units have exited the system). The simulation begins at time 0, under well specified initial conditions, often empty and idle.

A nonterminating system runs continuously, or for a very long period of time. A steady-state simulation is a simulation whose objective is to study long run or steady-state behavior of a nonterminating system. The steady-state properties are not influenced by the initial conditions and there is no natural event E to end the simulation. In general, a steady-state analysis is done to determine how a system will respond to a peak load of infinite duration. In this type of simulation, the simulator must decide to stop the simulation after some number of observations have been collected or after some length of time, T_E. The stopping time is thus a design choice and is not determined by the inherent nature of the problem.

B. Performance Measures and Their Estimation

1. Point Estimate

One of the objectives of statistics is to make an inference about a population based on the information contained in a sample. Since the population can be characterized by numerical descriptive measures called parameters, the objective of a statistical investigation may be to make an inference about one or more of the parameters. Estimation of the parameter of interest, or the target parameter, is one method of inference. A single number as an estimate of the target parameter, with the intention that this number be as close to the target parameter as possible, is called a point estimate.

As described above, the simulation study is used as an analysis tool. The data generated by the simulation model is used as a random sample to estimate the population parameters of interest. Suppose the simulation study is of a bank, and the parameters of interest are how many customers are in the bank at time t, and how long each customer, i, must wait before being helped. The time each customer must wait, X_i, can be recorded in the form $\{X_1, X_2, \ldots, X_n\}$. The average time a customer must wait is an ordinary mean. When the output data is in this form, call the parameter of interest θ.

The point estimator of θ based on the data $\{X_1,\ldots,X_n\}$ is defined by

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$\hat{\theta}$ is a sample mean based on a sample of size n.

The average number of customers in the bank at any given time t, can be recorded in the form $\{X(t),\ 0<t<T\}$. In this case, it is important to consider the amount of time that has passed in calculating the average, so this is

a time-weighted mean. Let ϕ be the parameter when the output data is of this form.

The point estimator of ϕ based on the data $\{X(t),\ 0<t<T\}$, where T is the total time elapsed, is defined by

$$\hat{\phi} = \frac{1}{T}\int_0^T X(t)\,dt$$

$\hat{\phi}$ is a time weighted average of X(t) over [0,T]

The estimation of proportions, such as how much of the time a teller is busy in the bank example, is a special case of the estimation of means. The discussion that follows will focus on the ordinary mean, θ, since the concepts for time weighted averages and proportions are similar.

In classical statistics, a point estimator is said to be an unbiased estimator if the expected value of the point estimator is the parameter of interest.

$$E(\hat{\theta}) = \theta$$

In general,

$$b = E(\hat{\theta}) - \theta \tag{1.1}$$

where b is the bias in the point estimator.

If the bias b equals θ, then the point estimator is unbiased. It is desirable to have point estimators that are unbiased, or have as small a bias b as possible, relative to the magnitude of the parameter. In addition, if two unbiased estimators are compared, the one with the smaller variance is preferred. In this way, in repeated sampling, a higher fraction of the values of the estimator will be "close" to the target parameter θ.

The point estimator most commonly used in simulation studies is the sample mean, $\bar{X}(n)$, where n is the sample size, as an estimate of the population mean, μ.

$$\bar{X}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1.2}$$

The sample mean, $\bar{X}(n)$, is an ordinary mean, so it is a special case of θ, and it is an unbiased estimator of the population mean. In the example of the bank, a sample consisting of the amount of time different customers had to wait could be used to estimate the mean length of time a customer must wait before being helped.

It is often of interest to estimate the steady-state characteristics of the system. The steady-state mean is given by

$$\bar{X} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1.3}$$

The value of $\bar{X}$ is independent of the initial conditions. However, the sample mean as an estimator of the population mean, μ, without any other information, is insufficient. A method for interval estimation is also needed.

2. Interval Estimation

Interval estimation is where an interval of possible values intended to enclose the parameter of interest is specified. In general, the two endpoints of the interval are estimated. Ideally, the resulting interval should have two properties; it should contain the target parameter, θ, and the interval should be relatively narrow. Since the endpoints of the interval are functions of the sample, they will vary in a random manner from sample to sample. The amount of the variability in the sample, then, should be relatively unbiased to provide a good interval estimate. The sample variance, S^2, is an unbiased estimator for σ^2, the population variance when the samples are independent and identically distributed (i.i.d.).

$$S^2(n) = \frac{1}{n-1}\sum_{i=1}^{n}[X_i - \bar{X}(n)]^2 \tag{1.4}$$

Since the variance of $\bar{X}$ is given by

$$Var(\bar{X}) = \frac{\sigma^2}{n} \tag{1.5}$$

an unbiased estimator of $Var(\bar{X})$ is given by

$$\hat{Var}[\bar{X}(n)] = \frac{S^2}{n} \tag{1.6}$$

3. Confidence Intervals

An interval estimate is commonly called a confidence interval, and the probability that a confidence interval will enclose the target parameter, θ, is called a confidence coefficient. The confidence coefficient gives the fraction of time, in repeated sampling, that the interval estimate will contain the target parameter, θ. If the X_i's are independent and identically distributed normal random variables, as in classical statistics, then an exact confidence interval, c.i., of 100(1-α) for μ is

$$\bar{X}(n) \pm t_{n-1,1-\frac{\alpha}{2}}\sqrt{\frac{S^2(n)}{n}} \qquad (1.7)$$

where $t_{n-1,\ 1-\alpha/2}$ is the upper $1-\alpha/2$ critical value for a t distribution with n-1 degrees of freedom (d.f), and $\sqrt{S^2(n)/n}$ is the standard error of $\bar{X}(n)$.

The half-length, HL, of the c.i. is used as a measure of absolute precision, which is dependent on the population variance of the X_i's, σ^2.

$$HL = t_{n-1,1-\frac{\alpha}{n}}\sqrt{\frac{S^2(n)}{n}} \qquad (1.8)$$

The mean-square error is a performance measure that combines the bias and the variance. Thus, it encompasses both the accuracy and the precision of the point estimate.

$$MSE = b^2 + \hat{Var}(\bar{X}(n)) \qquad (1.9)$$

C. Output Analysis for Steady-state Simulations

1. Autocorrelation

All the equations developed above are useful only if the observations $X_1, X_2, \ldots, X_n$ are i.i.d. random variables. However, this does not seem to be true for most simulation output (Law and Kelton, 1983). The output is autocorrelated rather than independent and nonstationary rather than identically distributed. For example, if the ith customer arrives and waits a long time, then it is highly likely that the (i+1)st customer will also wait a long time. The output data is likely to be nonstationary because of the difficulty in choosing the initial conditions of the simulation to be representative of the "typical" operation of the system, so the distributions of the output observations change over time.

Suppose the observations $X_1, X_2, \ldots, X_n$ are from a covariance stationary process, i.e. the covariance of lag i is independent of time, with a common finite mean μ and common finite variance σ^2. The sample mean, $\bar{X}$, is still an unbiased point estimator for this process, but the sample variance S^2, is no longer an unbiased estimator of the population variance, σ^2. The variance of the sample mean is given by (Fishman,1973)

$$Var(\bar{X}(n)) = \frac{\sigma^2}{n}\left[1 + 2\sum_{i=1}^{n-1} \frac{1-i}{n}\rho_i\right] \tag{1.10}$$

The correlation between any two observations at lag i (i.e. i observations apart) is given by ρ_i. If the observations are positively correlated, (i.e. $\rho_i > 0$ for $i=1,2,\ldots,n-1$), which is true of the output data of most queueing simulations (Banks and Carson, 1984), the sample variance will have a negative bias: $E[S^2(n)] < \sigma^2$. However, the limit of $E[S^2(n)]$ as n approaches infinity is σ^2. Thus, $S^2(n)$ is asymptotically unbiased. The net effect when ρ_i is positive is an unjustified confidence in the apparent accuracy of the point estimator, thus the actual coverage of a desired 90% c.i. could be significantly less than 90%. This means that if the process is really covariance stationary, and the i.i.d. method is used to estimate the variance of the sample mean, there are two sources of error. The first is the bias in $S^2(n)$ as an estimator of σ^2 and the second is the effect of neglecting the correlation term, ρ_i. When the output is not stationary, even the sample mean, $\bar{X}$, is a biased estimator of the population mean, μ.

2. Initialization Bias

As noted above, the initial conditions for the simulation must be specified. The initial state is often chosen with little knowledge of the system behavior and hence is atypical. Convenient values, such as empty and idle, are often used. The output of the system is strongly influenced by these initial conditions. Wilson and Pritsker (1978b) found that the choice of initial conditions for a run had a greater influence on the accuracy of the results than any other factor. This can cause the collected data to be significantly biased near the start of each run. The most popular technique for reducing this bias is to divide each simulation run into two phases: an initialization phase from time 0, followed by a data collection phase. This allows the model to "warm-up", reaching conditions more similar to steady-state before data collection begins. The data from the initialization phase is discarded, and this method is referred to as truncation. The period of time before steady-state is achieved is called the transient time. However, with this method it is difficult to identify an appropriate truncation point.

3. Truncation

Wilson and Pritsker (1978a) identified three categories for the more common truncation rules: those based on time series analysis, (notably Fishman); those derived from queueing theory models, (Blomqvist, Cheng, Law, and Madansky); and heuristic rules (Conway, Fishman, Gafarian, Gordon, and Schriber). The rules based on time series analysis and queueing theory have limited applicability, even though they are rigorously correct. Although many real-world systems can be described by an autoregressive or queueing theory model, and thus these methods are applicable, several model parameters must be estimated to apply the truncation rules. The estimation process can be time consuming and cumbersome for a complex model. In addition, as the systems become more complex, a single queueing theory model or autoregressive model may no longer fully describe the system. Because of the desire for simplicity, many heuristic truncation rules have been developed.

The first study of the performance of several popular heuristic truncation point selection methods was done by Gafarian et al (1978). The purpose of a truncation rule was defined to be the determination of the minimum time, t*, such that

$$1-\epsilon < \frac{E[X_t]}{\mu_x} < 1+\epsilon \qquad \textit{for all } t > t^* \tag{1.11}$$

where ϵ is an assigned tolerance. This is essentially identifying the point at which the bias has been reduced to an acceptable level. A set of relative criteria was developed for comparing five heuristic truncation rules. The truncation rules were applied to a Markovian model (M/M/1/∞) under various initial conditions. The criteria included accuracy, precision, generality (apply to both autoregressive and queueing theory models), cost (in CPU time), and simplicity. This set of criteria is useful for comparing truncation rules, although it is not clear that the best policy for estimating t* is also the best policy for estimating the steady-state mean, μ_x. The heuristic rules discussed by Gafarian, et al (1978) included rules developed from queueing models and autoregressive models, but none met their set of criteria, typically because of accuracy or cost.

Wilson and Pritsker (1978b) recommended using loss of confidence (standardized confidence intervals) as a performance measure for comparisons of different systems. Two finite-space Markovian models, including a M/M/1/15 queueing model, were used for the analysis. Heuristic truncation rules were applied to a time series output of the model representing the length of the queue. The model was run 1000 times so that the distribution of the truncation point could be studied, since previous studies neglected the randomness of the truncation point. The theoretical bias, variance, and mean square error of the sample mean over a fixed range of truncation points were calculated and tabulated. The tabulated values were then averaged with respect to the empirical truncation point distribution, from the 1000 replications. Those results were used to construct confidence intervals for the steady-state mean, and were standardized for comparison of four different heuristic truncation rules. The two systems studied indicated that the choice of initial conditions affected the performance of the sample mean as an estimator of steady-state mean more than the choice of truncation rules. In these Markovian systems, the steady-state mode was the best choice for initial conditions. The truncation rules were also very sensitive to parameter selection.

Schruben (1982) developed a general approach to testing for initialization bias in the mean of a simulation output series. The procedure recommends the output be grouped into small batches (of size 5) before the tests are applied. The method can be used to detect initialization bias before or after truncation. The assumption is that the point estimate will be improved if there is no initialization bias. However, then a good method would be to discard all but the last observation to minimize the bias (for example, if the output data was a time series). But this increases the variance (since the sample size is now 1) and hence the size of the confidence interval. It is interesting to note that although the performance measure is different, Schruben also recommends using batch means as a method for smoothing the simulation output.

Kelton and Law (1984) developed a procedure based on independent and probabalistically identical replications, deletion of initial data, and a time series regression technique that may be valid in reducing initialization bias. First order autoreggressive models with finite run lengths were considered. The performance measures were the mean absolute deviation and confidence interval performance. The purpose was to find a general method that would perform adequately for many types of models. Thus, general trends were identified and observed. Kelton and Law found that replication and deletion of some initial amount improved the c.i. performance in many cases, and did not severely worsen the results of those models not improved.

Although the effectiveness of truncation rules was questioned in the studies done by Wilson and Gafarian, truncation is still one of the most common methods of reducing initialization (warmup) bias. The performance measures vary from study to study, but so far all investigated have been evaluated in terms of models with theoretical results. The application of heuristic truncation rules generally require discrete observations of the system, but the effect of the time between observations has not been studied.

II. A Study of the Application of Schriber's Rule

A. Schriber's Heuristic Truncation Rule

Schriber (1974) suggests that the approach to steady-state conditions may be monitored by partitioning the observed time series $\{X_t: 1 < t < n\}$ into batches of some fixed size b. The behavior of the batch means can be used to identify stable behavior in the output - thus implying that steady-state has been achieved. A typical initial conditions generally produce extreme values in the set of batch means. As t increases, the batch means become relatively stable, i.e. become less variable. This stability indicates that the batch means were observed outside the transient period. Schriber (1974) uses a detailed example to illustrate this method, but uses inspection rather than an algorithm to identify the appropriate truncation point. Wilson (1977) used a formulation of Schriber's rule where a batch size b, a batch count k, and a tolerance ϵ were specified. The truncation point, d, is set at time n if the k most recent batches all fall within the tolerance ϵ of each other. The mathematical expression is:

$$\max\{|\bar{X}_j(b) - \bar{X}_l(b)| : 1<j,l<k\} < \epsilon \tag{2.1}$$

where $\bar{X}_j(b)$ is the batch mean of the jth batch, and $\bar{X}_l(b)$ is the batch mean of the lth batch.

The minimum truncation point, as seen by inspection, must be at time $n = k*b$. If at that time, the truncation rule is satisfied, then $d = n$.

$$\text{minimum truncation point: } d = k * b \tag{2.2}$$

Otherwise, the oldest batch $\{X_1, \ldots, X_b\}$ is dropped, and the batch mean for the next batch $\{X_{n+1}, \ldots, X_{n+b}\}$ is calculated. The truncation rule is again applied (equation 2.1). If the truncation rule is still not satisfied, the oldest batch is dropped and the steps are repeated. If the rule is satisfied, truncation occurs and $d = n$. The truncation point, d, is sensitive to the selection of parameters k, b and ϵ, as well as the scale of the discrete observations of the output.

Schriber's rule is conceptually appealing, particularly because of other work with batch means as a method to overcome autocorrelation, such as Schruben (1982), Law (1977), and Law and Carson (1979). Wilson's study of Schriber's rule (1977) considered a single selection of k, b, and ϵ but did not consider the effect of the time scale or a service time distribution other than exponential.

1. Performance Measures

In order to determine the effects of parameter selection, time scale and service time distributions, some performance measures are needed. First of all, the goodness of the point estimate must be considered. The mean squared error (MSE) will be used as a measure of the accuracy, since it encompasses both the bias and the variance and the HL will be used as a measure of precision. Second, the number of initial observations where the system is empty and idle will be used to describe the effects of the time scale on the truncation point. Third, the empirical truncation point distributions will be used for comparison of parameter sets to determine if there is any significant difference.

2. Computation of Mean Squared Error

The mean squared error is the sum of the bias squared and the variance (Equation 1.9, page 7). Thus, for each design level,

$$MSE_R = [\overline{X}_R - \mu_x]^2 + V\hat{a}r(\overline{X}_R) \tag{2.3}$$

where R is the number of replicates.

This assumes that all the replicates (R = 1000) are used to calculate a single point estimate. This method should be used if multiple replicates will be used to define the truncation point.

If each of the replicates is considered a trial, and the average MSE is calculated, then the MSE must be calculated for each trial, then averaged. This method is used if the practitioner will base the truncation point on a single run. The mean $\bar{X}_r$ and Variance S_r^2 are outputs of each replicate, r. Then

$$MSE_r = [\bar{X}_r - \mu_x]^2 + S_r^2 \tag{2.4}$$

where

$$\bar{X}_r = \bar{X}_{n-d}$$

or

$$\bar{X}_r = \frac{1}{n-d} \sum_{j=d+1}^{n} X_j \tag{2.5}$$

where n is the sample size and d is the truncation point. then

$$\overline{MSE} = \frac{1}{R1} \sum_{r=1}^{R} MSE_r \tag{2.6}$$

or

$$\overline{MSE} = \bar{b}^2 + \bar{S}^2 \tag{2.7}$$

Since the bias can be either positive or negative, the bias must be squared, then averaged if equation (2.7) is used, because

$$\overline{b^2} > (\bar{b})^2$$

3. Computation of Confidence Interval Precision

The half-length, HL is used as a measure of confidence interval precision, which means a small HL is desirable. The sample size of 1000 replicates is large enough to allow the use of the normal approximation Z_α for t_α. Then, for each design level, equation (1.8) page 7, becomes

$$HL_R = Z_{1-\frac{\alpha}{2}} (s.e.)_R \tag{2.8}$$

where s.e., the standard error, is defined as

$$(s.e.)_R = \sqrt{\frac{V\hat{a}r}{n}} \tag{2.9}$$

For each replicate, equation (1.8) page 7, becomes

$$HL_r = Z_{1-\frac{\alpha}{2}} (s.e.)_r \tag{2.10}$$

where s.e. is

$$(s.e.)_r = \sqrt{\frac{S_r^2}{n}} \tag{2.11}$$

The average HL, $\overline{HL}$ is

$$\overline{HL} = \frac{1}{R} \sum_{r=1}^{R} HL_r \tag{2.12}$$

B. Application to a Single-Server Queue

The simulation study was run on a Hewlett-Packard QS/20, a 386 personal computer. The simulation study was written in SIMSCRIPT II.5, release 2.20. The system is analogous to a bank, which has one teller, and a customer arrival rate and service time. The arrival rate and service rate are independent. In addition, when there are 15 customers in the bank, any potential new customers will leave rather than join the waiting line. The number of customers in the system (in the bank) was recorded at discrete time intervals, as noted above. Subroutines were written as part of the SIMSCRIPT II.5 program to apply Schriber's rule to the output series, and record the output measures defined above (page 16).

1. Exponential Distribution of Service Time

The first model considered is a M/M/1 queue with a traffic intensity of .9, which is a simple, nontrivial queueing system commonly cited in the literature; see Schruben, Singh and Tierney (1983), Schruben (1982), Gafarian, et al (1978), and Kelton and Law (1983). A model with an arrival rate of 4.5/time unit and a service rate of 5.0/time unit with a finite capacity of 15 was used, as discussed by Wilson (1977) in the evaluation of several truncation rules. The theoretical steady-state mean for the number in the system is 5.361 (Wilson, 1978). Since the steady-state mean is known, the goodness of the point estimate can easily be evaluated. A modification of the distribution for the service time will be made so that the effect of distributions can be considered.

2. Model with Weibull Distribution for Service Time

In addition to the M/M/1/15 model, a similar model with a Weibull distribution for service time (M/G/1/15) was used because it allows (with a shape parameter) a skewed distribution to be considered. A shape parameter of alpha = 2 was used, with a scale parameter of beta = .2257. Thus, the mean service time for both models is 0.2, but the variance for the exponential is 0.04 while the variance for the Weibull is 0.0109.

C. Discussion of Parameter Selection

The number in the system is the output measure, and the observations are recorded at discrete time intervals. After 50 observations of the number in the system are recorded, regardless of the time scale, the truncation rule is applied. The truncation point, the mean and variance of the truncated sample, and the number of initial observations with the system empty are recorded. If the truncation point is not reached before 50, the value of 50 is recorded as the truncation point and the recorded mean is the batch mean of the last batch.

The formulation of Schriber's rule by Wilson (1977) was applied to the series output of the two models described above. The parameters of batch size b, number of batches k, and the time between observations are varied to determine the impact on the estimate of the steady-state mean. ϵ (see equation 2.1, page 15) is fixed at 4.03, which for the M/M/1/15 model corresponds to confidence interval of alpha = .25 (Wilson, 1978).

The batch size can affect the minimum truncation point (see equation 2.2, page 16) as well as the estimate of the steady-state mean. The batch size must be large enough so that serial correlation between batch means is reduced and the assumption of independence is valid. Thus, a minimum batch size of 5 was chosen. A large batch size may cause too much initial data to be discarded, which typically increases the variance when the sample size is fixed. The bias is reduced by increasing the truncation point, but at the expense of the variance.

The number of batches, k, is also a factor. When k is 2, the comparison is strictly between the pair. But if $k > 2$, then comparisons are pairwise, which increases the sensitivity of the truncation rule to gradual increases or decreases in the output series. However, it also increases the computational time required, as well as increasing the minimum truncation point (see equation 2.2, page 16). In addition, since the maximum deviation (see equation 2.1, page 15) is not always between adjacent pairs, the output series may appear to still have initialization bias, and so additional batches are needlessly truncated. This increase in the truncation point causes an increase in the variance of the truncated sample.

The observations of the output, in this case the number in the system, are recorded at discrete time intervals. The time scale, T, is a measure of the average number of arrivals between observations. For a given arrival rate, the time scale, T, is directly proportional to the time between observations. For example, suppose the arrival rate is 5/unit time. Then if the time between observations is 1, the time scale, T is 5 (5 arrivals between observations). If the time between observations is .5, the time scale, T is 2.5 (2.5 arrivals between observations). When the time between observations is very short, multiple initial observations with little or no activity can occur. The batch means of the first two batches are compared, and since they are similar, the system appears to be at steady-state. In reality, few actual arrivals have occurred, and significant initialization bias may still exist. Suppose a long time is allowed between observations; then the simulation run time is increased and thus unnecessary cost may be added.

1. Experimental Design

The experimental design for the M/M/1/15 queueing model is a 2 X 2 X 4 design, with parameters b (2 levels), k (2 levels) and T (4 levels) as shown in table 1. A thousand replicates were run at each design level.

Table 1. Parameters and factor levels for the experimental design.

Parameters	Factor Levels
b, batch size	5, 10
k, number of batches	2, 3
T, time scale: average arrivals per observation	1, 3, 4.5, 6.5

Each design level of 1000 replicates uses the same sequence of random numbers and corresponding runs start with the same seed to minimize the variation in the system. This technique, called common random numbers, or sychronization, is used to achieve an ideal degree of blocking for simulation experiments. The inverse transform method (Law and Kelton, 1982) is used to generate the samples from the exponential distribution, which requires a single uniform random number. The Weibull distribution also requires a single random number to generate a sample via the inverse transform method, preserving the sychronization. Separate random number streams are used for arrivals and service time.

III. Analysis of Results

The M/M/1/15 model will be discussed first, and the results will be used for the experimental design for the M/G/1/15 model. The results of the M/G/1/15 model will then be discussed, and finally the two models will be compared.

A. M/M/1/15 Model

A summary of the mean, bias, variance, mean squared error, standard error (s.e.) and half-length (significance level of α = .1) are displayed in table 2 based on using all the replicates for a single point estimate (equations 2.3, page 17, and 2.8, page 19). This is the method of choice if the simulator will be using multiple replications to determine the truncation point. The variance, mean squared error and half-length were minimized with the parameter set of T = 6.5, k = 2, and b = 5, but the bias was minimized with the parameter set T = 6.5, k = 2, and b = 10.

In some cases, the simulator will use a single run to estimate the steady-state mean; then the average performance of the truncation rule should be considered. The average performance for each design level was computed and summarized in table 3, including the mean, average

Table 2. Results for each design level for the M/M/1/15 model using the replicates as observations for a single point estimate. Equations 2.3 and 2.8 are used, with a significance level of .1 for calculating the half-length.

T,k,b	$\bar{X}(n)$	bias	var	mse	s.e.	HL
1,2,5	4.021	-1.340	6.591	8.385	0.081	1.726
1,2,10	4.276	-1.085	8.788	9.965	0.094	1.739
1,3,5	4.154	-1.207	7.862	9.319	0.089	1.734
1,3,10	4.503	-0.858	12.004	12.740	0.110	1.755
3,2,5	5.069	-0.292	5.218	5.304	0.072	1.717
3,2,10	5.202	-0.159	7.327	7.353	0.086	1.731
3,3,5	5.134	-0.228	6.904	6.956	0.083	1.728
3,3,10	5.290	-0.071	10.834	10.839	0.104	1.749
4.5,2,5	5.198	-0.163	3.798	3.824	0.062	1.707
4.5,2,10	5.312	-0.049	5.768	5.770	0.076	1.721
4.5,3,5	5.237	-0.124	5.371	5.387	0.073	1.718
4.5,3,10	5.198	-0.163	7.970	7.997	0.089	1.734
6.5,2,5	5.261	-0.100	2.934	2.944	0.054	1.699
6.5,2,10	5.319	-0.042	4.268	4.270	0.065	1.710
6.5,3,5	5.295	-0.066	4.398	4.403	0.066	1.711
6.5,3,10	5.236	-0.125	6.649	6.664	0.082	1.727

Table 3. Results for each design level for the M/M/1/15 model. Equations 3.1 and 2.12 are used, with a significance level of .1 for calculating the half-length.

T,k,b	$\overline{X}(n)$	$\overline{b}^2$	$\overline{S}^2$	$\overline{S}$	$\overline{MSE}$	s.e.	$\overline{HL}$
1,2,5	4.021	8.385	6.611	2.371	14.996	0.075	1.720
1,2,10	4.276	9.965	5.362	2.123	15.327	0.067	1.712
1,3,5	4.154	9.319	5.930	2.238	15.249	0.071	1.716
1,3,10	4.503	12.740	3.468	1.697	16.208	0.054	1.699
3,2,5	5.069	5.304	12.432	3.383	17.736	0.107	1.752
3,2,10	5.202	7.353	10.859	3.117	18.212	0.099	1.744
3,3,5	5.134	6.956	11.046	3.142	18.002	0.099	1.744
3,3,10	5.290	10.839	7.894	2.601	18.733	0.082	1.727
4.5,2,5	5.198	3.824	14.258	3.667	18.082	0.116	1.761
4.5,2,10	5.312	5.770	12.651	3.401	18.421	0.108	1.753
4.5,3,5	5.237	5.387	12.787	3.418	18.174	0.108	1.753
4.5,3,10	5.198	7.997	9.836	2.938	17.833	0.093	1.738
6.5,2,5	5.261	2.944	15.340	3.838	18.284	0.121	1.766
6.5,2,10	5.319	4.270	14.059	3.633	18.329	0.115	1.760
6.5,3,5	5.295	4.403	13.840	3.593	18.242	0.114	1.759
6.5,3,10	5.236	6.664	11.449	3.214	18.114	0.102	1.747

bias, average variance, average mse, and average HL. The MSE was calculated for each replicate, r, using the theoretical steady-state mean μ_x of 5.361. The variance of the number in the system was recorded for each replication, and so from equation 2.4, page 18, the MSE for the truncated sample is

$$MSE_r = [\bar{X}_r(n-d) - \mu_x]^2 + Var(\bar{X}_r(n-d)) \tag{3.1}$$

where n = 50 (total observations), d is the number of observations that were deleted, and r = 1,..1000.

The average MSE and average HL are minimized at T = 1, k = 2, and b = 5. The average variance is minimized at T = 1, k = 2, and b = 5 while the average squared bias is minimized at T = 6.5, k = 2, b = 10. Notice that these results are dissimilar to table 2, page 26. These are somewhat unexpected results because in general, with a fixed sample size, as more observations are discarded the variance will increase (because the sample size decreases). Wilson (1977) found that for the M/M/1/15 model, as the truncation point increases, the bias decreased and the variance increased. However, his study was based on the theoretical bias and variance of the model given an experimental truncation point. This study considers experimental bias, variance, and truncation point. This will be discussed further at the end of this section (page 42).

1. Initial Observations with the System Empty and Idle

For each replicate, the number of initial observations with the system empty and idle was also recorded. Since the number in the system is observed at discrete time intervals, if an entity arrives, is served, and exits between observations, the system is still considered empty and idle. The observations begin at simulation time 0.00, so there must always be at least 1 observation with the system empty and idle; the mode was 1 observation before the system was busy for all time scales. Figure 1 shows the first fifteen observations for three different time scales (with the same random number stream). For this replicate, when T = 1, there were 6 observations of an empty system, 2 observations when T = 4.5, and only 1 when T = 6.5. In this case, the number of initial observations of an empty system decreases as T increases.

The set of initial observations of an empty system for a given time scale forms an empirical distribution (see table 4). These distributions can then be compared statistically to describe the effects of the time scale.

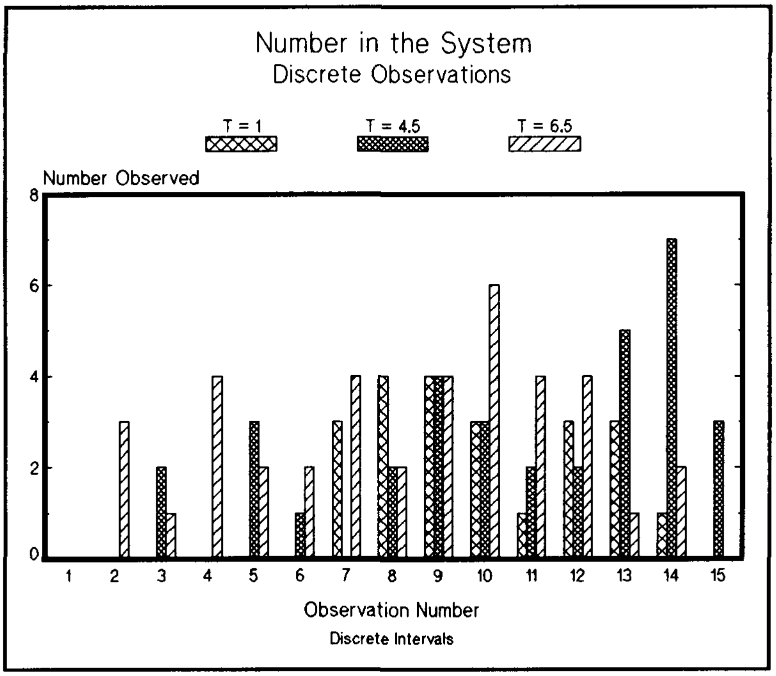

Figure 1. Number in the system. Discrete observations with varying time scales with the same random number stream.

Table 4. Number of initial observations with the system empty and idle for M/M/1/15 model.

T	mean	std dev	max
1	2.186	1.620	16
3	1.551	0.967	8
4.5	1.414	0.770	6
6.5	1.345	0.661	7

Consider the case where T = 1, (average arrivals between observations of 1); the mean is 2.186, but at least once, 16 observations were recorded before the number in the system was more than 0. In general, as the time scale increases, the maximum, mean, and variability all decrease. A graphical representation of the results is shown in figure 2. The shift in the mean and the variability for T = 1 and T = 3 is particularly evident.

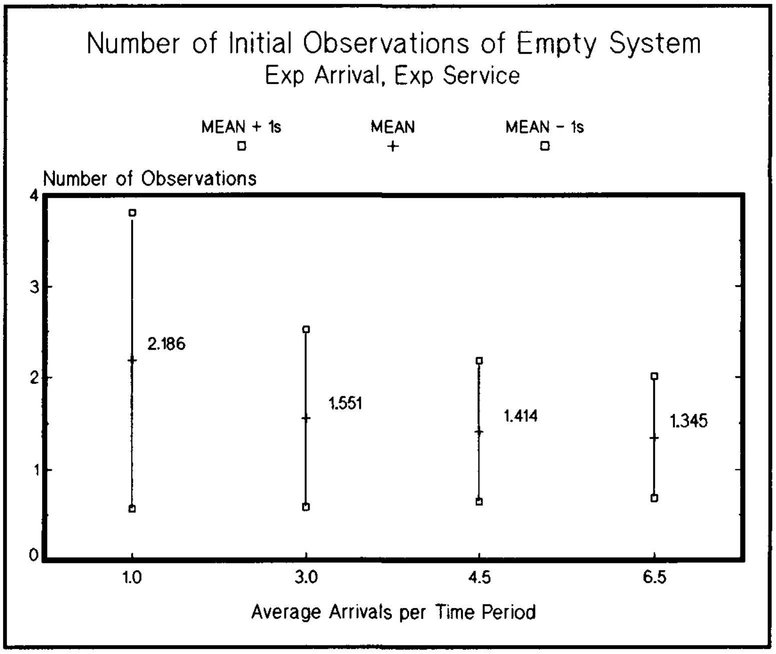

Figure 2. Number of initial observations with the system empty and idle. The mean, and mean +/- 1 standard deviation are shown for different time scales, T.

The number of observations before the system is busy becomes more consistent from replication to replication as T increases (the variance is reduced). So, if very few replications are used to determine the truncation point using Schriber's rule, a longer time scale (more arrivals per observation) will be more likely to yield consistent results. In practice, Schriber's rule will be applied to a limited number of output series, so it is important to have consistent results. Thus, a very short time scale (such as T = 1) should be avoided.

2. Empirical Truncation Distributions

The empirical distributions for the truncation point, d, are generated from the independent replications of the system operation under each parameter set, or design level. Each run provides a random sample from the theoretical distribution of the truncation point, d. The empirical distributions were then compared for the different design levels to determine if there was any significant difference. Table 5 summarizes the frequency distributions for the M/M/1/15 model.

Paired statistical comparisons of the truncation distributions for T = 4.5 and T = 6.5 indicates that for two batches of size five, the means are equal with a significance level of p = 0.66, while for T = 3 and T = 4.5, the means are equal with a significance of p = .0034. The mean and standard deviation typically increase as T, b, or k increase independently.

Representative frequency distributions for the truncation point, d, are given in figures 3 through 7. Figure 3 shows the frequency distribution for a typical design level, T = 4.5, b = 5, k = 2. Here, the majority of the replicates, 76.7%, had a truncation point of 10, which

Table 5. Truncation point distributions for the M/M/1/15 model.

Design Level T,K,B	mean (d)	std dev(d)	median
1,2,5	10.26	1.239	10
1,2,10	21.55	4.326	20
1,3,5	16.73	4.222	15
1,3,10	34.73	7.745	30
3,2,5	11.23	2.974	10
3,2,10	23.21	6.560	20
3,3,5	20.10	8.532	15
3,3,10	37.84	9.177	30
4.5,2,5	11.65	3.425	10
4.5,2,10	23.63	7.482	20
4.5,3,5	21.22	9.191	15
4.5,3,10	38.78	9.413	30
6.5,2,5	11.72	3.627	10
6.5,2,10	23.55	7.138	20
6.5,3,5	21.95	9.838	15
6.5,3,10	38.75	9.232	40

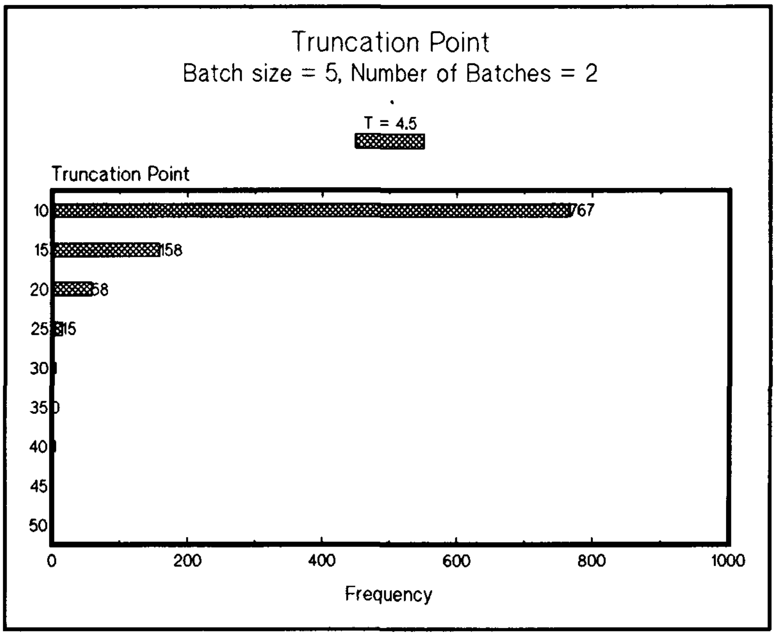

Figure 3. Frequency distribution of the truncation point, d, for the design level T = 4.5, b = 5, k = 2.

from equation 2.2, page 16, is the minimum truncation point for this k and b.

A comparison of the empirical truncation distributions for two design levels, with only the batch size changed is shown in figure 4. The time scale, T, is constant at T = 4.5, and the number of batches, k, is constant at k = 2 for both distributions. The mode in each case is the minimum truncation point (d = k*b). So for a batch size of 5 with 2 batches, the mode is 10 and for a batch size of 10 with 2 batches, the mode is 20. The frequency of occurance decreases as the truncation point, d, increases.

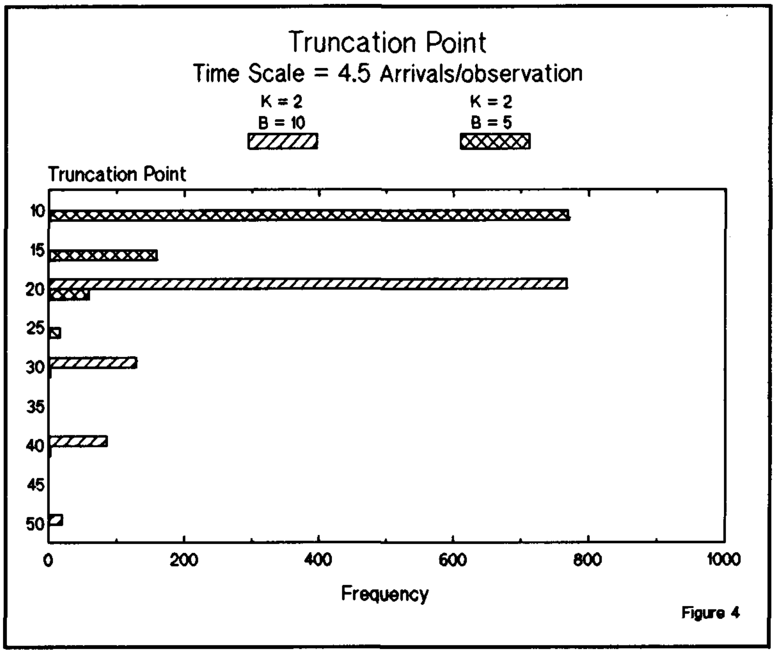

Figure 4. A comparison of the frequency distributions for the truncation point, d, with a constant time scale of T = 4.5 and number of batches of k = 2 for the M/M/1/15 model.

A typical comparison of the effect of the number of batches, k, is shown in figure 5. The time scale T is constant at T = 4.5 and the batch size is constant at b = 5. In both cases, the most common truncation point is the minimum truncation point. When k = 3, and the batch size is 5, the mode is 15; similarly, when k = 2, the mode is 10. The frequency for both decrease as d increases, but the frequency decreases more slowly when k = 3, which is also seen in table 5, page 33; the standard deviation for T = 4.5, b = 2, k = 2 is 3.425, while the standard deviation when k = 3 is 9.191. Pairwise comparisons of the batch means are made when k = 3, which makes the application of Schriber's rule more sensitive to gradual increases or decreases in the output series. (batch means) When the series is gradually increasing, a comparison of two adjacent batch means might yield a difference less than some ϵ while the difference between the first and third batches is more than ϵ. Thus, if the ouput series is gradually increasing, the application of Schriber's rule with k = 2 may indicate steady-state had been achieved while the application with k = 3 may indicate steady-state had not been achieved.

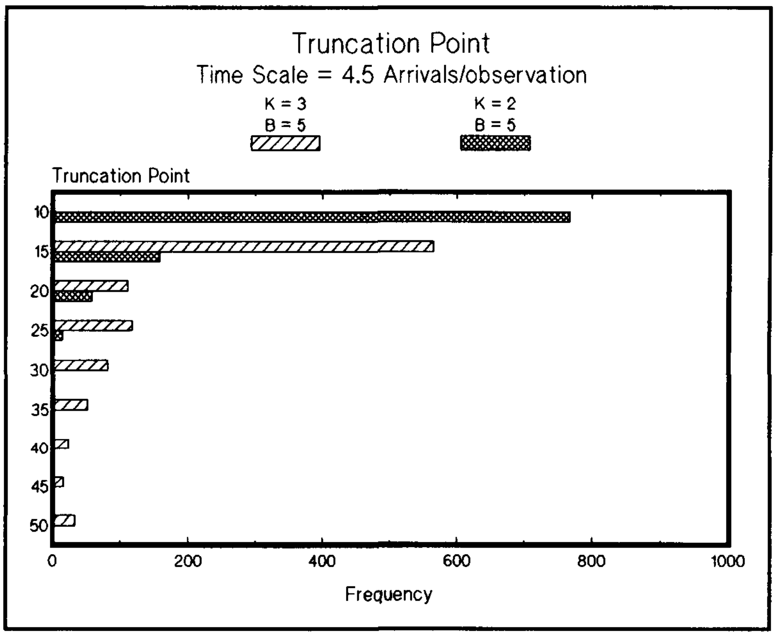

Figure 5. A comparison of the effect of the number of batches on the frequency distribution of the truncation point, d. The time scale is constant at T = 4.5 and the batch size is constant at size 5 (b = 5).

Figure 6 shows a comparison of the frequency distributions when the batch size is 10. It is clear that the frequency of d = 50 is significantly higher than the frequency of d = 40. Recall that if the truncation rule is not satisfied within the initial 50 observations, a value of 50 is recorded. Thus, the output series is terminated even if Schriber's rule is not satisfied and the truncation point is the last possible value. This is the likely explanation for the parameter set k = 3 and b = 10.

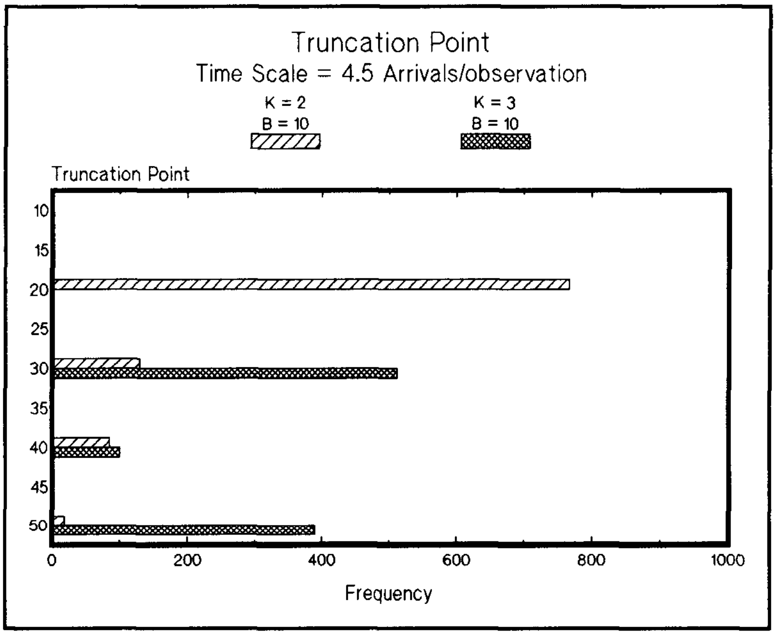

Figure 6. A comparison of the number of batches on the frequency distribution of the truncation point, d. The time scale is constant at T = 4.5, and the batch size is constant at b = 10.

The effect of the changing time scale, T, on the frequency distribution of the truncation point, d, is shown in figure

7. The batch size is 5, and there are two batches, k. The mode is always 10 and the frequency decreases as the truncation point, d increases. However, as the time scale, T, increases from T = 1 to T = 6.5, the frequency of the truncation point d = 10 decreases. With a short time scale, T, as seen in the previous section, a significant number of observations with the system empty and idle can occur; this may incorrectly indicate steady-state has been achieved. The frequency distributions for the truncation point when T = 4.5 and T = 6.5 (with two batches of size 5) are very similar. Recall from the discussion following table 5, page 33, there was not a statistical difference between these distributions.

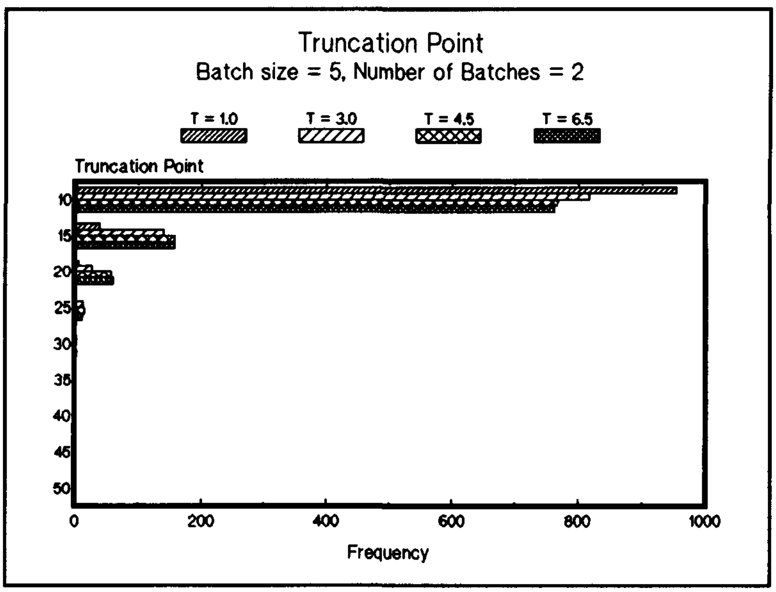

Figure 7. A comparison of the frequency distribution of the truncation point, d, when only the time scale varies. The batch size is constant at size 5 and the number of batches is 2.

In summary, as T, k, or b increase, independently of the other variables, the mean and variance of the truncation distribution increase. The mean was expected to increase as k and b increased because the minimum truncation point also increases (equation 2.2, page 16).

3. ANOVA

There were 16 design levels (see table 1, page 24) with 1000 replicates, hence N is 16,000. The MSE and mean were the dependent variables which were calculated for each replicate using equations 2.4 and 2.5, page 18. The ANOVA showed that for both mean and MSE, the time scale, T was significant as well as the batch size b. The interaction of T and b was significant for the mean but not for the MSE.

Table 6. ANOVA for M/M/1/15 model with mean and MSE as the dependent variable.

Dependent variable: mean; N: 16,000.

Source	Sum-Squares	D.F.	Mean-Square	F-Ratio	P
T	2964.353	3	988.118	148.042	0.000
K	9.330	1	9.330	1.398	0.235
B	58.366	1	58.366	8.744	0.003
T*K	30.909	3	10.303	1.544	0.199
T*B	55.191	3	18.397	2.756	0.040
K*B	1.433	1	1.433	0.215	0.648
T*K*B	10.200	3	3.400	0.509	0.680
error	106686.244	15984	6.675		

Table 6. (Continued)

Dependent variable: MSE; N: 16,000

Source	Sum-Squares	D.F.	Mean-Square	F-Ratio	P
T	22470.794	3	7490.265	82.639	0.000
K	85.207	1	85.207	0.940	0.330
B	363.517	1	363.517	4.011	0.043
T*K	468.844	3	156.281	1.724	0.158
T*B	418.638	3	139.546	1.540	0.200
K*B	0.056	1	0.056	0.001	0.929
T*K*B	238.100	3	79.367	0.876	0.455
error	1448766.823	15984	90.639		

In an effort to consider the effects if the theoretical mean is unknown, and hence estimated, the grand mean, $\bar{\bar{X}}$ was used in place of μ_x in calculations of $\hat{MSE}$ for ANOVA. The results were the same (table 7); the time scale, T, and the batch size were significant and k was not.

Table 7. ANOVA for M/M/1/15 model with $\hat{MSE}$ as the dependent variable.

Dependent Variable: $\hat{MSE}$; N:16,000

Source	Sum-Squares	D.F.	Mean-Square	F-Ratio	P
T	48124.001	3	16041.334	128.516	0.000
K	244.642	1	244.642	1.960	0.158
B	934.691	1	934.691	7.488	0.006
T*K	928.477	3	309.492	2.480	0.058
T*B	1227.250	3	409.083	3.277	0.020
K*B	0.274	1	0.274	0.002	0.915
T*K*B	285.921	3	95.307	0.764	0.517
error	1995118.772	15984	124.820		

The results for the M/M/1/15 model show that the number of batches was optimal at 2. In fact, when the number of batches, k, was 3, the results were misleading and too many initial observations were discarded. In addition, there was no significant difference ($\alpha = .1$) in the results for T = 4.5 and T = 6.5.

4. Results for M/M/1/15 Model

In order to effectively evaluate the results of the experiment, it must be determined if the practitioner will be using a single run or multiple replicates to determine the truncation point. If a single run is to be used, then it is critical to minimize the variance of the truncation point, d. This can be done by considering the empirical distribution of the truncation point and the number of initial observations of an empty and idle system. In this case, the parameter set T = 4.5, b = 5, and k = 2 should be used, even though the average MSE is not minimized. If the theoretical bias and variance for this model are used in conjuction with the empirical truncation distribution (as tabulated and used by Wilson, 1977), the MSE decreases as T increases. Examination of some of the independent replicates showed that in some cases, as the truncation point increased, the bias increased and the variance decreased. The only cause for this is the particular sequence of random numbers used for that replicate. The variability of the average results is a result of the independent realizations of the model. Thus, it is recommended to use multiple replicates to determine the truncation point, d, so that the probability of extreme results is minimized. In addition, the replicates used to define the truncation point should not be used (in general) to estimate the steady-state mean because of the correlation in the output series.

For multiple replicates, in particular when all 1000 replicates were used for the point estimates, the MSE and the HL were minimized with the parameter set T = 6.5, b = 5, and k = 2. When the empirical truncation distributions for T = 4.5 and T = 6.5 were compared, the means for the two distributions were not statistically significantly different. The number of initial observations of the system empty and idle indicated that T = 1 should not be used, while other time scales, T are acceptable. The ANOVA indicated that T and b and their interaction was significant, while k was not significant. Since the variance of the empirical truncation distributions was adversely affected by k, it was determined that k = 2 should be used. Thus, the recommendation for this model is to use multiple replicates with a parameter set of T = 4.5, b = 5, and k = 2.

B. M/G/1/15 Model

The parameter set for the M/G/1/15 model was simplified based on the results for the M/M/1/15 model. Two levels were still used for the batch size, b, but the time scale, T, was reduced to three levels and the number of batches, k, was not varied. The experimental design is summarized in table 8.

Table 8. Parameters and factor levels for experimental design.

Parameters	Factor Levels
b, batch size	5, 10
T, time scale: average arrivals between observations	1, 3, 4.5

The results (mean, bias, variance, MSE, and half-length) for the M/G/1/15 model using all the replicates for a single point estimate (equations 2.3, page 17, and 2.8, page 18) are shown in table 9. The grand mean, $\bar{\bar{X}}$, is used as an estimate for the steady-state mean.

Table 9. The results for each design level for the M/G/1/15 model using the replicates as observations for a single point estimate. Uses the grand mean as an estimate of μ_x, and a significance level of .1 for calculation of HL.

Weibull T,k,b	$\overline{X}(n)$	$\hat{b}^2$	var	MSE	s.e.	HL
1,2,5	14.414	0.207	0.722	0.929	0.027	1.672
1,2,10	14.910	0.002	0.143	0.145	0.012	1.657
3,2,5	14.966	0.009	0.007	0.016	0.003	1.648
3,2,10	14.973	0.011	0.007	0.018	0.003	1.648
4.5,2,5	14.974	0.011	0.003	0.015	0.002	1.647
4.5,2,10	14.976	0.011	0.004	0.016	0.002	1.647

The MSE is minimized for T = 4.5, k = 2, b = 5, but the means are equal between T = 3 and T = 4.5 with a significance of .1. The HL is minimized for T = 4.5, but there is no appreciable difference in the HL for T = 3 and T = 4.5. The mean in all cases is very near 15, which is the maximum allowable in the system. The results indicate that the system becomes congested early and does not clear out. Thus, any of the truncation rules are satisfactory and in fact the estimates are not severely worsened if no truncation occurred.

The average results are also important because the performance of a single application of Schriber's rule may be used by the simulator in estimating the steady-state mean. These results, using the grand mean, $\bar{\bar{X}}$ as an estimate of the steady-state mean and computed with equations 3.1, page 28, and 2.12, page 19, are summarized in table 10.

Table 10. The average results for each design level for the M/G/1/15 model are displayed in this table. The grand mean is used as an estimate of μ_x.

Weibull T,k,b	$\overline{S^2}$	$\overline{S}$	$\overline{MSE}$	s.e.	$\overline{HL}$
1,2,5	3.011	1.144	3.733	0.036	1.681
1,2,10	0.251	0.157	0.394	0.005	1.650
3,2,5	0.086	0.133	0.093	0.004	1.649
3,2,10	0.061	0.089	0.068	0.003	1.648
4.5,2,5	0.064	0.112	0.067	0.004	1.649
4.5,2,10	0.060	0.095	0.065	0.003	1.648

The average MSE and average variance are minimized at T = 4.5 with a batch size of 10. The average HL is highest for T = 1 with a batch size of 5; the HL for all other parameter sets is essentially the same. In this case, the average variance, average MSE, and average HL are all worst for T = 1, b = 5; all other applications of the truncation rule are satisfactory.

1. Initial Observations with the System Empty and Idle

The frequency distributions for the number of initial observations with the system empty and idle are compared in Table 11. The maximum, mean, and standard deviation are considerably larger when T = 1, (average arrivals between observations of 1); the variance for T = 1 is more than five times larger than the variance for T = 3. Paired t tests show the distributions do not have equal means ($p = .000$ for T = 3 and T = 4.5).

Table 11. Number of initial observations with the system empty and idle for M/G/1/15 model.

T	mean	std dev	max
1	2.105	1.402	14
3	1.238	0.542	5
4.5	1.115	0.349	3

A graphical representation of the results is shown in figure 8. The downward shift in the mean and the decrease in the variability between T = 1 and T = 3 is evident.

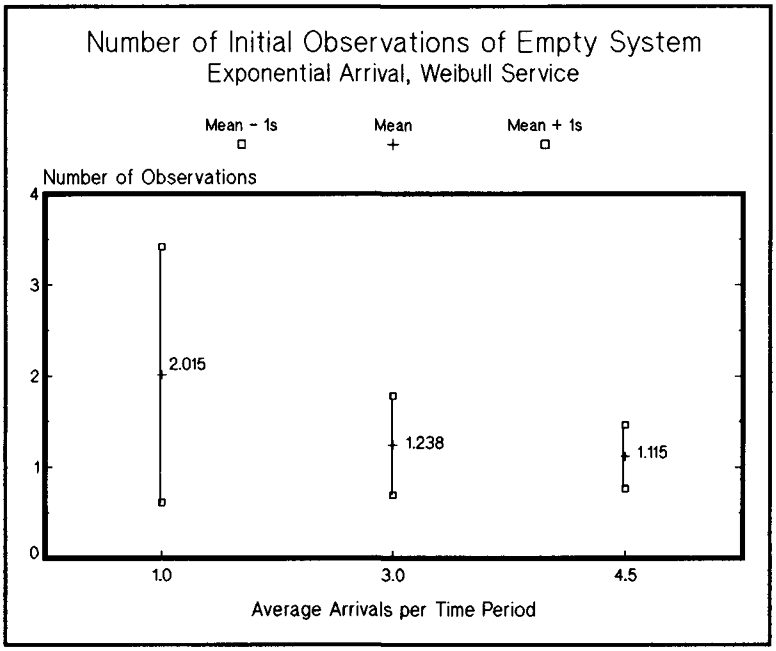

Figure 8. Number of initial observations with the system empty and idle. The mean, and mean +/- 1 standard deviation are shown for different time scales.

2. Empirical Truncation Distributions

Table 12 summarizes the frequency distributions for the truncation point, d, for the M/G/1/15 model. Paired t tests for equal means for these distributions with a significance of .1 showed that the distributions do not have equal means.

When the batch size was 10, the mean truncation point decreased as T increased, which was expected, but the variance was minimized when T = 3. When T = 3, the mean truncation point was 28.26, which is very near a possible truncation point of 30. For T = 4.5, the truncation point, d, decreased to 23.85, but the nearest truncation points

Table 12. Truncation point distributions for the M/G/1/15 model.

T,k,b	mean(d)	std dev (d)	median
1,2,5	14.85	4.503	15
1,2,10	31.55	4.391	30
3,2,5	15.26	1.353	15
3,2,10	28.26	3.868	30
4.5,2,5	14.96	0.669	15
4.5,2,10	23.85	4.926	20

(since b = 10) for this application of Schriber's rule are 20 or 30. Thus, the variance is clearly increased for T = 4.5. When the batch size is 5, the mean truncation point is maximized when T = 3, but the variance decreases as T increases. The mean truncation point for both T = 3 and T = 4.5 is very near 15, a possible truncation point. This system is highly congested, and stablizes after the system is full. Thus, either enough time or enough entities must be processed to stabilize. Consequently, the truncation point is near 15, and as T increases, more independent replications have a truncation point of 15. Thus, the variance is minimized when T = 4.5 and b = 5. From the preceding section, the highest average variance and average MSE occurred for T = 1, b = 5, which also has the highest variance for the truncation point distribution.

A graphical comparison of the effect of T when the batch size is constant at b = 5 (and k = 2) is shown in figure 9.

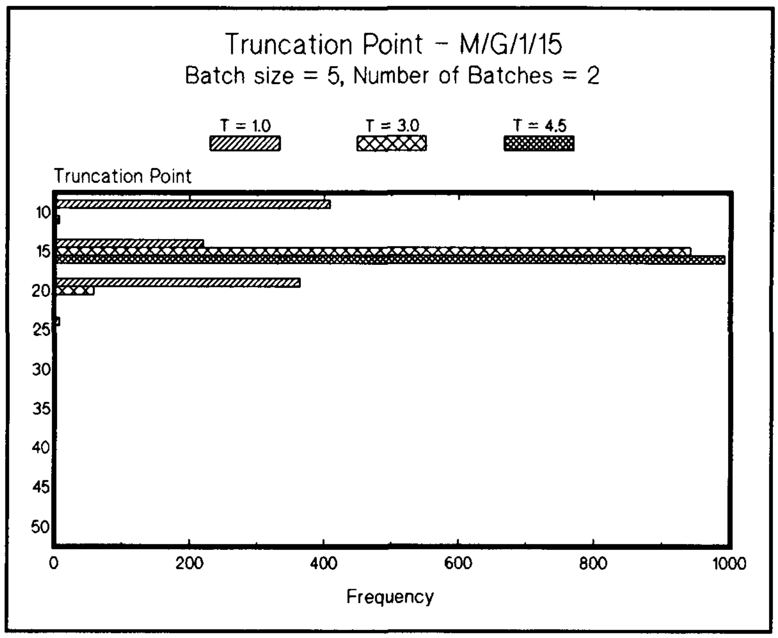

Figure 9. A comparison of the frequency distributions with a batch size of 5 and the time scale varies for the M/G/1/15 model.

The shapes of the frequency distributions are quite different as the time scale changes. The mode is at the minimum truncation point only for T = 1. The mode for both T = 3 and T = 4.5 is at 15; both also have a low variance (see Table 12, page 48).

3. ANOVA

The ANOVA (table 13) using the Weibull distribution for the length of service indicated that T and b are significant for both the dependent variables MSE and mean calculated with equations 2.4 and 2.5, page 18. There were 6 design levels with 1000 independent observations each, so N = 6000. The MSE was calculated using the grand mean, $\bar{\bar{X}}$ as an estimate of the steady-state mean.

Table 13. ANOVA for M/G/1/15 model.

Dependent variable: mean; N = 6000.

Source	Sum-Squares	D.F.	Mean-Square	F-Ratio	P
T	128.382	2	64.191	434.532	0.000
B	42.467	1	42.467	287.474	0.000
T*B	80.626	2	40.313	272.894	0.000
error	885.462	5994	0.148		

Dependent variable: MSE; N = 6000.

Source	Sum-Squares	D.F.	Mean-Square	F-Ratio	P
T	5281.305	2	2640.652	280.792	0.000
B	1888.351	1	1888.351	200.796	0.000
T*B	3684.443	2	1842.222	195.891	0.000
error	56369.442	5994	9.404		

4. Results for M/G/1/15

This model, because there are fewer very short service times, is highly congested. Hence, the steady-state number in the system is very near the upper limit allowed in the system (15). In this model, the results when all the replicates are used for a single point estimate (Table 9, page 44) and the average results (Table 10, page 45), are similar. The results for the number of intitial observations of an empty and idle system indicate that $T = 1$ should not be used, and that either $T = 3$ or $T = 4.5$ is acceptable. The results from the empirical truncation point distributions show that the variability is increased when the batches are of size 10. Thus, batches of size 5 should be used. The ANOVA showed that T and b as well as the interaction between T and b are significant. For this model, a parameter set of $T = 4.5$, $b = 5$, and $k = 2$ is recommended. This is consistent with the results for the M/M/1/15 model in the previous section, page 41.

C. Comparison of Results

When multiple replicates are used, both the M/M/1/15 and M/G/1/15 models show that as the time scale, T, increases, the MSE and the HL decrease and are thus better point estimates. However, while the M/M/1/15 model also shows an improved point estimate when the batches are of size 5 instead of size 10, the M/G/1/15 model shows no difference.

The number of initial observations with the system empty and idle were compared for both models. As seen in figure 10, page 52, each had the maximum number of initial observations of an idle system for $T = 1$, and these decreased as T increased. In general, the M/M/1/15 model had higher means, variances, and maximums than the M/G/1/15 model. Paired t tests with a significance of .1 for equal means between the two models for each time scale were rejected in every case.

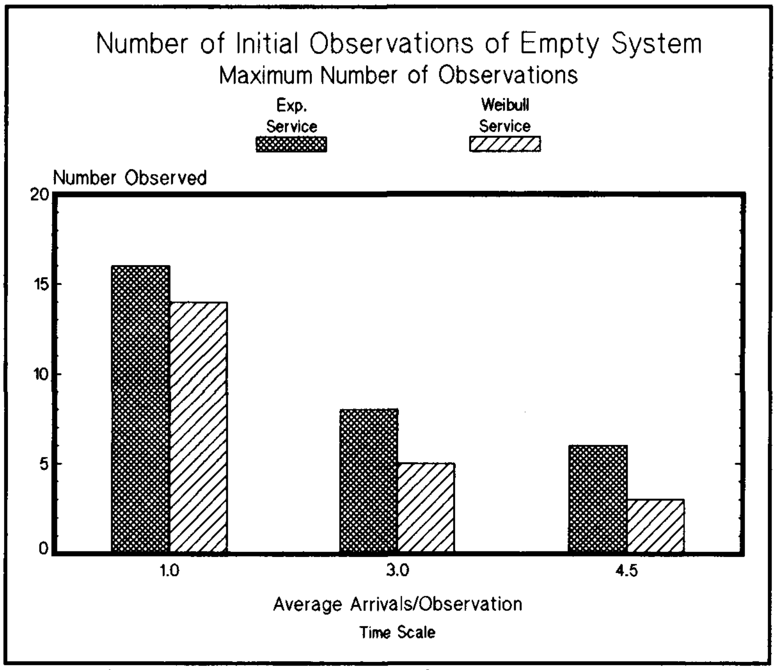

Figure 10. Comparison of the results for the number of initial observations with the system empty and idle for the M/M/1/15 and the M/G/1/15 model (effect of service time distribution).

A typical comparison of the truncation distributions for the M/M/1/15 and M/G/1/15 models are shown in figure 11. The time scale is T = 4.5 average arrivals/observation, the batch size is 5, and the number of batches, k is 2. Notice that the M/G/1/15 model does not have a mode at d = 10; instead it is at d = 15. While the modes are not similar, the variability of the distributions increase when the batch size increases from 5 to 10. The variability of the truncation distribution decreased as T increased for the M/G/1/15 model with a batch size of 5, while for the M/M/1/15 model with batches of size 5, the variability of the truncation distribution increased as T increased. This may be because for the M/M/1/15 model, $\bar{X}_d$ increases away from a possible truncation point of 10 as T increases, while for the M/G/1/15 model, $\bar{X}_d$ remains close to 15, which is a possible truncation point. In general, the empirical truncation distributions for the M/G/1/15 model had a much lower variance than for the M/M/1/15 models.

The ANOVA for both models showed that T, b and the interaction of T and b were significant for the dependent variables of mean and MSE.

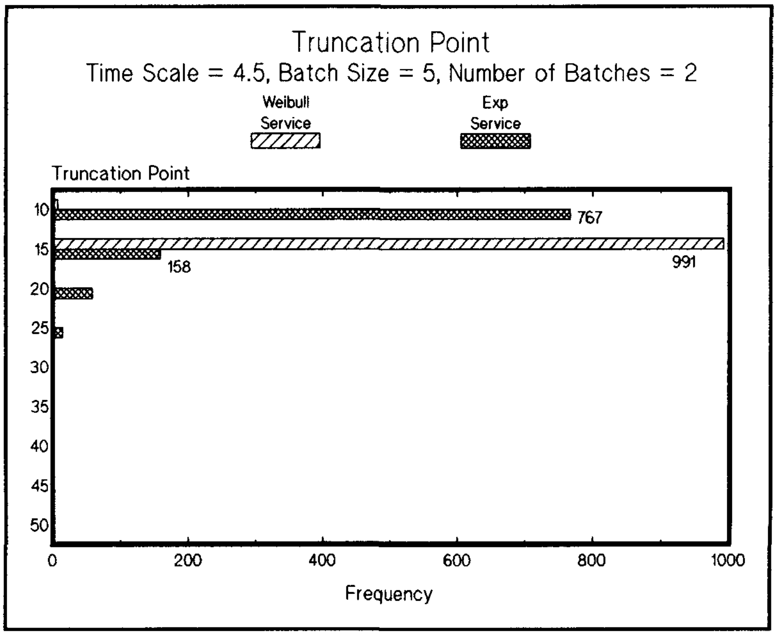

Figure 11. A comparison of the truncation distributions for the M/M/1/15 and the M/G/1/15 model when T = 4.5, the batches are of size 5, and the number of batches is 2.

IV. Conclusions

A. Conclusions and Recommendations

The performance measures were the goodness of the point estimate and the consistency of the results considering the randomness of the simulation. Although previous studies recommended starting the models at steady-state conditions, the models were started with the system empty and idle because this is most convenient and will be likely in practice. The effect of parameter selection for Schriber's rule was considered for two different models, the M/M/1/15 model with an arrival rate of 4.5 and service rate of 5 and the M/G/1/15 model with the same arrival and service rate. The parameters varied were the time scale, T, the batch size, b, and the number of batches, k. In queueing theory, the time scale does not have an effect on the transient time, but it does have an effect in the simulation studies.

The ANOVA results indicate that the time scale, T, and the batch size, b, and their interaction are significant for the dependent variables of mean and MSE. The empirical truncation distributions and the number of initial observations of an empty and idle system were also affected by the time scale. In general, the time scale needs to be selected so that enough activity occurs between observations to give meaningful results. When the time scale, T, was 1 (an average of 1 arrival per time period), the results were very poor. The results for the M/G/1/15 model were adequate for T = 3, but this was not sufficient for the M/M/1/15 model.

A side issue that emerged during this study was a need for multiple replicates to determine the truncation point. The M/M/1/15 model was particularly vulnerable. The particular sequence of random numbers would cause the variance to decrease and the bias to increase as the truncation point increased, which is the converse of the theoretical results (Wilson, 1977). If for the same model, Wilson's tabulated theoretical values for bias and variance are used, the theoretical results are achieved (variance increases and bias decreases and truncation point increases). When all the replicates were used for the point estimate, as in Table 2 (page 23) and Table 9 (page 41), the results were as expected: an increase in variance and a decrease in bias. The use of multiple replicates was recommended by Welch (1983) in evaluating the behavior of a random sequence because the random variable, such as the truncation point d, has a probability distribution associated with it.

The practical implications are that a single run should not be used to determine the truncation point, particularly when different sequences of random numbers will be used for the model. In addition, the time scale for observations should be carefully considered, in particular to ensure the time scale is not too short. The change in distributions did not have significantly negative results: improper selection of the distributions will likely affect the congestion of the system and the output results more than the truncation point.

B. Future Research

Schriber's rule was used to evaluate two specific queueing theory models, one with a known steady-state mean number in the system. The number of observations was limited to 50 (based on Wilson's results), this may not have been enough observations particularly with batches of size 10 and 3 batches. Certainly in some cases the truncation rule was not satisfied and hence assumed to be 50 with the mean assumed to be the mean of the last batch.

One of the conclusions was that multiple replicates should be used. A topic for future research would be how many replicates are sufficient, and how this is affected by the time scale and the truncation rule used.

Schriber's rule is certainly not the only truncation rule that can be used and its performance should be compared with other heuristic rules, such as Welch's moving average rule (Welch, 1983), Kelton and Law's (1984) time series regression technique, and Schruben's (1982) test for initialization bias. In particular, now that many simulations can be run on personal computers that are reasonably fast, the obstacle of requiring too much computational time is becoming less of a factor. In addition, many simulation languages allow easy manipulation of the output sequences or allow exporting to spreadsheets and statistical packages for manipulation. Hence, the error of discarding too much data is less of a problem than not discarding enough, assuming that the remaining set is sufficiently large.

The behavior of this truncation rule, or the effect of the selection of the time scale and batch size for any truncation rule, when used for a complex system is a final area for future research. A complex system may not behave like a queueing system and may have a shorter or longer transient time. It is much more difficult to identify the steady-state conditions for start-up, and so it will likely start with an initially empty and idle system. Hence, at least some initial transient will exist.

Bibliography

Banks, J. and J.S. Carson. 1984. "Discrete-Event System Simulation" Prentice-Hall

Fishman, G.S. 1973. "Concepts and Methods in Discrete Event Digital Simulation" Wiley, New York

Gafarian, A.V. et al. 1978. 'Evaluation of Commonly used rules for Detecting Steady-State in Computer Simulation' Naval Research Logistical Quarterly v25 pp511-529

Kelton, W.D. and A.M. Law. 1983. 'A New Approach for Dealing with the Startup Problem in Discrete Event Simulation' Naval Research Logistical Quarterly v30 pp641-58

Kelton, W.D. and A.M. Law. 1984. 'An Analytical Evaluation of Alternative Strategies in Steady-State Simulation' Operations Research v32 pp169-184

Kelton, W.D. and A.M. Law. 1985. 'The Transient Behavior of the M/M/s Queue with Implications for Steady-State Simulation' Operations Research v33 n2

Law, A.M. 1983. 'Statistical Analysis of Simulation Output Data' Operations Research v31 pp983-1029

Law, A.M. 1977. 'Confidence intervals in Discrete Event Simulation: A Comparison of Replication and Batch Means' Naval Research Logistical Quarterly v27 pp667-678

Law, A.M., and J.S. Carson. 1979. 'A Sequential Procedure for Determining the Length of a Steady-State Simulation' Operations Research v27 1011-25

Law, A.M. and W.D. Kelton. 1982. "Simulation Modeling and Analysis" McGraw-Hill, New York

Law, A.M. and W.D. Kelton. 1984. 'Confidence Intervals for Steady-State Simulations - A Survey of fixed sample size procedures' Operations Research v32 pp1221-39

Schriber, T. 1974. "Simulation using GPSS" John Wiley, New York

Schruben, L., H. Singh, and L Tierney. 1983. 'Optimal Tests for Initialization Bias in Simulation Output' Operations Research v31 p1167-78

Schruben, L.W. 1982. 'Detecting Initialization Bias in Simulation Output' Operations Research v30 pp569-590

Welch, P.D. 1983 'The Statistical Analysis of Simulation Results' "The Computer Performance Modeling Handbook" The Academic Press pp267-329

Wilson, J.R. 1977. 'A Procedure for Evaluating, Startup Policiies in Simulation Experiments' Unpublished thesis, Purdue University.

Wilson, J.R. and A.A.B. Pritsker. 1978. 'A Survey of Research on the Simulation Start-up Problem' Simulation v31 pp55-58

Wilson, J.R. and A.A.B. Pritsker. 1978. 'Evaluation of Startup Policies in Simulation Experiments' Simulation v31 pp79-89