Proceedings of the 1990 Winter Simulation Conference
Osman Balci, Randall P. Sadowski, Richard E. Nance (eds.)

DESIGN AND ANALYSIS OF SIMULATION EXPERIMENTS FOR MANUFACTURING APPLICATIONS

Averill M. Law

Simulation Modeling and Analysis Company
440 N. Alvernon Way, Suite 103
Tucson, Arizona 85711

ABSTRACT

One of the most important but often neglected aspects of a simulation study is the proper statistical design and analysis of simulation experiments. In this paper we give a state-of-the-art and practical discussion of this subject in the context of manufacturing systems. The presentation is based on the existing simulation literature as well as our experiences in applying these techniques in numerous manufacturing projects.

1. INTRODUCTION

In many simulation studies a great deal of time and money is spent on model development and "programming," but little effort is made to design appropriate simulation runs or to analyze correctly the resulting output data. Since random samples from the input probability distributions (e.g., for machine operating or repair times) "drive" a simulation model through time, basic simulation output data (e.g., hourly throughputs in a factory) or an estimated performance measure computed from then (e.g., average hourly throughput from the entire run) are also random. *Thus, a simulation model only produces a statistical estimate of the (true) performance measure, not the measure itself.*

In order for a simulation estimate to be statistically precise (have a small variance) and free of bias (have a mean equal to the desired performance measure), the analyst must specify appropriate choices for the following:

- Length of each simulation run.
- Number of independent simulation runs.
- Initial conditions for each simulation run (e.g., all machines idle and no parts present).
- Length of the warmup period, if one is appropriate (see Section 5).

Then the overall estimate of the performance measure is computed from the output data beyond the warmup period in all runs.

We now describe more precisely the random nature of simulation output. Let Y_1, Y_2, ... be an output stochastic process from a *single* simulation run. For example, Y_i might be the throughput (number of parts produced) in the ith hour for a manufacturing system or the time in system (total cycle time) of the ith part to be completed for such a system. The Y_i's are random variables that will, in general, be neither independent nor identically distributed. Thus, most of the formulas of classical statistics that assume independence (see Section 1.1), do not *directly* apply to the analysis of simulation output data.

Our goal in this paper is to discuss methods for statistical analysis of simulation output data in the context of manufacturing systems, and to present the material with a practical focus that should be accessible to a simulation practitioner with a basic knowledge of probability and statistics. (See Law and Kelton [1991] for a more comprehensive treatment of output analysis.) After providing a review of statistics for independent data in Section 1.1, we discuss in Sections 2 and 3 types of simulations with regard to output analysis as well as measures of performance or parameters for each type. Finally, Sections 4 through 6 show how to get a point estimator and confidence interval for each type of parameter. Because of space limitations, we will, however, restrict our attention to only mean system performance.

1.1 Review of Statistics for Independent Data

We now provide a brief review of probability and statistics, which will be useful for the remainder of this paper. Suppose that $X_1, X_2, \ldots, X_n$ are independent, identically distributed (IID) random variables (observations) with mean μ and variance σ^2 and that our primary objective is to estimate μ; the estimation of σ^2 is of secondary interest. Then the *sample mean*

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

is an unbiased point estimator of μ. Similarly, the *sample variance*

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

is an unbiased estimator of σ^2.

The difficulty with using $\bar{X}(n)$ as an estimator of μ without any additional information is that we have no way of assessing how close $\bar{X}(n)$ is to μ. Because $\bar{X}(n)$ is a random variable with variance $\mathrm{Var}[\bar{X}(n)] = \sigma^2/n$, on one experiment $\bar{X}(n)$ may be close to μ while on another $\bar{X}(n)$ may differ from μ by a large amount. The usual way to assess the precision of $\bar{X}(n)$ as an estimator of μ is to construct a confidence interval for μ. In particular, an approximate $100(1 - \alpha)$ percent $(0 < \alpha < 1)$ confidence interval for μ is given by

$$\bar{X}(n) \pm t_{n-1,1-\alpha/2}\sqrt{\frac{S^2(n)}{n}} \tag{3}$$

where $t_{n-1,1-\alpha/2}$ is the upper $1 - \alpha/2$ critical point for the t distribution with n - 1 degrees of freedom.

We will see in Sections 4 through 6 how the above results can be *indirectly* applied to the analysis of simulation output data.

2. TRANSIENT AND STEADY-STATE BEHAVIOR OF A STOCHASTIC PROCESS

Consider the output stochastic process $Y_1, Y_2, \ldots$. Let $F_i(y|I) = P(Y_i \leq y|I)$ for $i = 1, 2, \ldots$, where y is a real number and I represents the initial conditions used to start the simulation at time 0. [The conditional probability $P(Y_i \leq y|I)$ is the probability that the event $\{Y_i \leq y\}$ occurs *given* the initial conditions I.] For a manufacturing system, I might specify the number of jobs present, and whether each machine is busy or idle, at time 0. We call $F_i(y|I)$ the *transient distribution* of the output process at (discrete) time i for initial conditions I. Note that $F_i(y|I)$ will, in general, be different for each value of i and each set of initial conditions I. The density functions for the transient distributions corresponding to the random variables Y_{i_1}, Y_{i_2}, Y_{i_3}, and Y_{i_4} are shown in Figure 1 for a particular set of initial conditions I and increasing time indices i_1, i_2, i_3, and i_4, where it is assumed that these random variables have density functions. A density specifies how the corresponding random variable can vary from one replication to another.

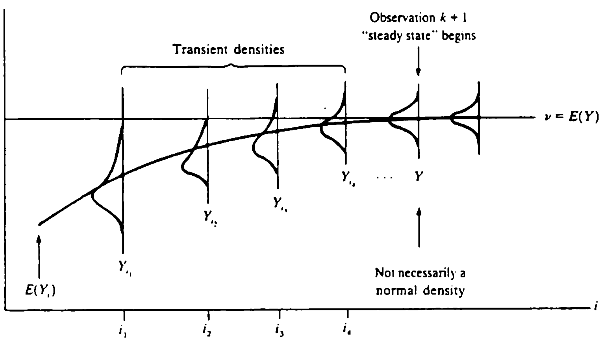

Figure 1. Transient and Steady-State Density Functions for a Particular Stochastic Process Y_1, Y_2, ... and Initial Conditions I

For fixed y and I, the probabilities $F_1(y|I)$, $F_2(y|I)$, ... are just a sequence of numbers. If $F_i(y|I) \to F(y)$ as $i \to \infty$ for all y and for any initial conditions I, then $F(y)$ is called the *steady-state distribution* of the output process Y_1, Y_2, ... Strictly speaking, the steady-state distribution $F(y)$ is only obtained in the limit as $i \to \infty$. In practice, however, there will often be a finite time index, say, $k + 1$, such that the distributions from this point on will be approximately the same as each other; "steady state" is figuratively said to start at time $k + 1$ as shown in Figure 1. Note that steady state does *not* mean that the random variables Y_{k+1}, Y_{k+2}, ... will all take on the same value in a particular simulation run; rather, it means that they will all have approximately the same *distribution*. Furthermore, these random variables will not be independent, but will approximately constitute a covariance-stationary stochastic process (see Law and Kelton [1991, p. 280]). Note that the steady-state distribution $F(y)$ does not depend on the initial conditions I; however, the rate of convergence of the transient distributions $F_i(y|I)$ to $F(y)$ does.

3. TYPES OF SIMULATIONS WITH REGARD TO OUTPUT ANALYSIS

The options available in designing and analyzing simulation experiments depend on the type of simulation at hand, as depicted in Figure 2. Simulations may be either terminating or nonterminating, depending on whether there is an obvious way for determining run length. Furthermore, measures of performance or parameters for nonterminating simulations may be of several types, as shown in the figure. These concepts are defined more precisely below.

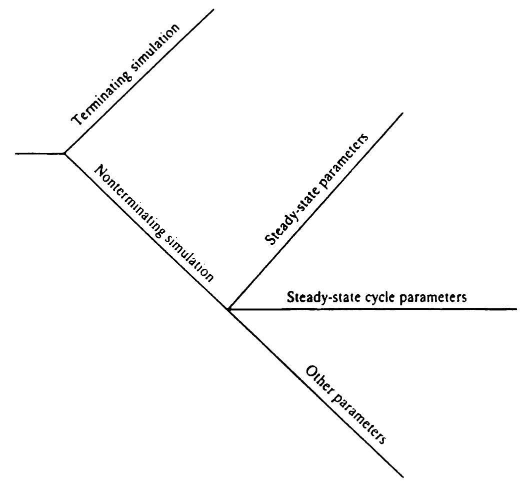

Figure 2. Types of Simulations with Regard to Output Analysis

A *terminating simulation* is one for which there is a "natural" event E which specifies the length of each run (replication). Since different runs use independent random numbers and the same initialization rule, this implies that comparable random variables from the different runs are IID (see Section 4). The event E often occurs at a time point beyond which no useful information is obtained or at a time point when the system is "cleaned out." It is specified before any runs are made and the time of occurrence of E for a particular run may be a random variable.

Example 1. An aerospace manufacturer receives a contract to produce 100 airplanes, which must be delivered within eighteen months. The company would like to simulate various manufacturing configurations to see which one can meet the delivery deadline at least cost. In this case E = {100 airplanes have been completed}.

Example 2. Consider a manufacturing system for food products. A production schedule is issued, the system produces product for 13 days, and then the system is completely cleaned out on the fourteenth day. Then a new production schedule is issued and the 2-week cycle is repeated, etc. In this case E = {thirteen days of production have been completed}.

Example 3. Consider a manufacturing company which operates 16 hours a day (2 shifts) with work-in-process carrying over from one day to the next. Would this qualify as a terminating simulation with E = {16 hours of simulated time have elapsed}? No, since this manufacturing operation is essentially a continuous process, with the ending conditions for one day being the initial conditions for the next day.

A *nonterminating simulation* is one for which there is no natural event E to specify the length of a run. A measure of performance for such a simulation is said to be a *steady-state parameter* if it is a characteristic of the steady-state distribution of some output stochastic process Y_1, Y_2, ... In Figure 1, if the random variable Y has the steady-state distribution, then we might be interested in estimating the steady-state mean $\upsilon = E(Y)$.

Example 4. Consider a company which is going to build a new manufacturing system and would like to determine the long-run (steady-state) mean hourly throughput of their system after it has been running long enough for the workers to know their jobs and for mechanical difficulties to have been worked out. Assume that:

(a) The system will operate 16 hours a day for 5 days a week.

(b) There is negligible loss of production at the end of one shift or at the beginning of the next shift.

(c) There are no breaks (e.g., lunch) which shut down production at specified times each day.

Let N_i be the number of parts manufactured in the ith hour. If the stochastic process N_1, N_2, ... has a steady-state distribution with corresponding random variable N, then we are interested in estimating the mean $\upsilon = E(N)$.

Consider an output process Y_1, Y_2, ... for a nonterminating simulation which does not have a steady-state distribution. Suppose that we divide the time axis into equal-sized, contiguous time intervals called *cycles*. (For example, in a manufacturing system a cycle might be an 8-hour shift.) Let Y_i^C be a random variable defined on the ith cycle, and assume that Y_1^C, Y_2^C, ... are comparable. Suppose that the process Y_1^C, Y_2^C, ... has a steady-state distribution F^C and that $Y^C \sim F^C$. Then a measure of performance is said to be a *steady-state cycle parameter* if it is a characteristic of Y^C such as the mean $\upsilon^C = E(Y^C)$. Thus, a steady-state cycle parameter is just a steady-state parameter of the appropriate cycle process Y_1^C, Y_2^C, ...

Example 5. Suppose for the manufacturing system in Example 4 that there is a half-hour lunch break at the beginning of the fifth hour in each 8-hour shift. Then the process of hourly throughputs N_1, N_2, ... has no steady-state distribution. Let N_i^C be the average hourly throughput in the ith 8-hour shift (cycle). Then we might be interested in estimating the steady-state expected average hourly throughput over a cycle, $\upsilon^C = E(N^C)$, which is a steady-state cycle parameter.

For a nonterminating simulation, suppose that the stochastic process Y_1, Y_2, ... does not have a steady-state distribution, and that there is no appropriate cycle definition such that the corresponding process Y_1^C, Y_2^C, ... has a steady-state distribution. This can occur, for example, if the input parameters for the model continue to change over time. In these cases, however, there will typically be a fixed amount of data describing how these parameters change over time. This provides, in effect, a terminating event E for the simulation and, thus, the analysis techniques for terminating simulations in Section 4 are appropriate. This is why we don't treat this situation as a separate case later in this paper. Measures of performance or parameters for such simulations usually change over time and are included in the category "Other parameters" in Figure 2.

Example 6. Consider a manufacturing system for microcomputers consisting of an assembly line and a test area. There is a 3-month build schedule available from marketing that describes the types and numbers of computers it is desired to produce each week. The schedule changes from week to week because of changing sales and the introduction of new computers. In this case, weekly or monthly throughputs do not have steady-state distributions. We therefore perform a terminating simulation of length 3 months and estimate the actual mean throughput for each week.

4. STATISTICAL ANALYSIS FOR TERMINATING SIMULATIONS

Suppose we make n independent replications of a terminating simulation, where each replication is terminated by the event E and each replication is begun with the "same" initial conditions (see Law and Kelton [1991, pp. 543-544]). The independence of replications is accomplished by using different random numbers for each replication. Let X_j be a random variable defined on the jth replication for $j = 1, 2, ..., n$; it is assumed that the X_j's are comparable random variables for different replications. Then the X_j's are IID random variables. For Example 1, X_j might be the time to produce 100 airplanes on the jth replication. In the case of Example 2, X_j might be the number of cases of food products produced in a cycle on the jth replication.

Suppose that we would like to obtain a point estimate and confidence interval for the mean $\mu = E(X)$, where X is a random variable defined on a replication as described above. Make n independent replications of the simulation and let $X_1, X_2, ..., X_n$ be the resulting IID random variables. Then by substituting the X_j's into (1) and (3) we get that $\bar{X}(n)$ is an unbiased estimator for μ, and an approximate $100(1 - \alpha)$ percent confidence interval for μ is given by

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

where the sample variance $S^2(n)$ is given by Eq. (2). If we increase the number of replications from n to $4n$, the half-length of the confidence interval (i.e., the quantity that is added to and subtracted from $\bar{X}(n)$ to get the confidence-interval endpoints) will decrease by a factor of approximately 2.

5. STATISTICAL ANALYSIS FOR STEADY-STATE PARAMETERS

Let Y_1, Y_2, ... be an output stochastic process from a single run of a nonterminating simulation. Suppose that $P(Y_i \leq y) = F_i(y) \to F(y) = P(Y \leq y)$ as $i \to \infty$, where Y is the steady-state random variable of interest with distribution function F. (We have suppressed the dependence of F_i on the initial conditions I.) Suppose that we want to estimate the steady-state mean $\upsilon = E(Y)$. One difficulty in estimating υ is that the distribution function of Y_i (for $i = 1, 2, ...$) is different from F, since it will generally not be possible to choose I to be representative of "steady-state behavior." In particular, this causes the sample mean $\bar{Y}(m)$ of the observations $Y_1, Y_2, ..., Y_m$ to be a biased estimator of υ for all finite values of m, that is, $E[\bar{Y}(m)] \neq \upsilon$. The problem we have just described is called the *problem of the initial transient* in the simulation literature.

The technique most often suggested for dealing with this problem is called *warming up the model* or *initial-data deletion*. The idea is to delete some number of observations from the beginning of a run and to use only the remaining observations to estimate υ. For example, given the observations $Y_1, Y_2, ..., Y_m$, it is often suggested to use

$$\bar{Y}(m,l) = \frac{\sum_{i=l+1}^{m} Y_i}{m-l}$$

($1 \leq l \leq m - 1$) rather than $\bar{Y}(m)$ as an estimator of υ. In general, one would expect $\bar{Y}(m,l)$ to be less biased than $\bar{Y}(m)$, since the observations near the "beginning" of the simulation are not very representative of steady-state behavior due to the choice of initial conditions.

The question naturally arises as to how to choose the *warmup period* (or deletion amount) l. We would like to pick l (and m) such that $E[\bar{Y}(m,l)] \approx \upsilon$. (The symbol "$\approx$" means approximately equal.) If l and m are chosen too small, then $E[\bar{Y}(m,l)]$ may be significantly different from υ. On the other hand, if l is chosen larger than necessary, then $\bar{Y}(m,l)$ will probably have an unnecessarily large variance. The simplest and most general technique for determining l is a graphical procedure due to Welch [1983]. Its specific goal is to determine a time index l such that $E(Y_i) \approx \upsilon$ for $i > l$, where l is the warmup period.

Welch's procedure is based on making n independent replications of the simulation and employing the following four steps (see Law and Kelton [1991, pp. 545-550] for details):

1. Make n replications of the simulation ($n \geq 5$) each of length m (m large). Let Y_{ji} be the ith observation from the jth replication ($j = 1, 2, \ldots, n$; $i = 1, 2, \ldots, m$).

2. Let $\bar{Y}_i = \Sigma_{j=1}^{n} Y_{ji}/n$ for $i = 1, 2, \ldots, m$. The averaged process $\bar{Y}_1, \bar{Y}_2, \ldots$ has means $E(\bar{Y}_i) = E(Y_i)$ and variances $\mathrm{Var}(\bar{Y}_i) = \mathrm{Var}(Y_i)/n$. Thus, the averaged process has the same transient means as the original process, but its plot is only $(1/n)$th as variable.

3. To smooth out the high frequency oscillations in $\bar{Y}_1, \bar{Y}_2, \ldots$ (but leave the low frequency oscillations or long-run trend of interest), we further define the moving average $\bar{Y}_i(w)$ (where w is the *window* and is a positive integer) as the simple average of the $2w + 1$ observations $\bar{Y}_{i-w}, \ldots, \bar{Y}_i, \ldots, \bar{Y}_{i+w}$ for $i = w + 1, \ldots, m - w$. (For $i = 1, \ldots, w$, see Law and Kelton [1991, p. 546] for the appropriate formula.)

4. Plot $\bar{Y}_i(w)$ for $i = 1, 2, \ldots, m - w$ and choose l to be that value of i beyond which $\bar{Y}_1(w), \bar{Y}_2(w), \ldots$ appears to have converged.

Example 7. A small factory consists of a machining center and inspection station in series, as shown in Figure 3. Unfinished parts arrive to the factory with exponential interarrival times having a mean of 1 minute. Processing times at the machine are uniform on the interval [0.65, 0.70] minute and subsequent inspection times at the inspection station are uniform on the interval [0.75, 0.80] minute. Ninety percent of inspected parts are "good" and are sent to shipping; ten percent of the parts are "bad" and are sent back to the machine for rework. (Both queues are assumed to have infinite capacity.) The machining center is subject to randomly occurring breakdowns. In particular, a new (or freshly repaired) machine will break down after an exponential amount of *calendar* time with a mean of 6 hours. Repair times are uniform on the interval [8, 12] minutes. Assume that the factory is initially empty and idle.

Consider the stochastic process $N_1, N_2, \ldots$, where N_i is the number of parts produced in the ith hour. Suppose that we want to determine the warmup period l so that we can eventually estimate the steady-state mean hourly throughput $\upsilon = E(N)$ (see Example 8). We made $n = 10$ independent replications of the simulation each of length $m = 160$ hours (or 20 days). In Figure 4 we plot the moving average $\bar{N}_i(w)$ with a window of $w = 30$, from which we chose a warmup period of $l = 24$ hours.

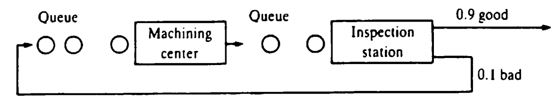

Figure 3. Small Factory Consisting of a Machining Center and an Inspection Station

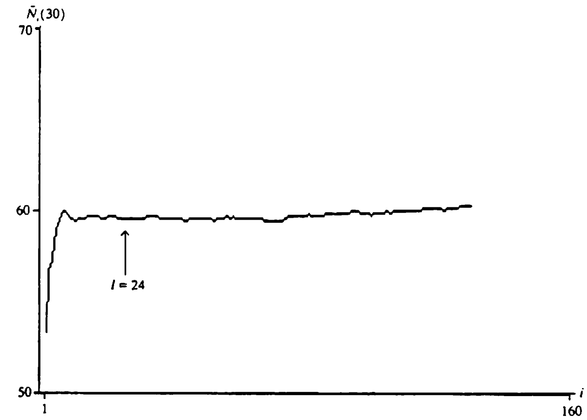

Figure 4. Moving Average for Hourly Throughputs

5.1 Replication/Deletion Approach for Confidence-Interval Construction

We now present the *replication/deletion approach* for obtaining a point estimate and confidence interval for the steady-state mean $\upsilon = E(Y)$. The analysis is similar to that for terminating simulations except that now only those observations beyond the warmup period l in each replication are used to form the estimates. Specifically, suppose that we make n' new replications (production runs) of the simulation each of length m' observations, where m' is much larger than the warmup period l determined by Welch's graphical method (see above). Let Y_{ji} be as defined before and let X_j be given by

$$X_j = \frac{\sum_{i=l+1}^{m'} Y_{ji}}{m'-l} \qquad \text{for } j = 1, 2, \ldots, n'$$

(Note that X_j uses only those observations from the jth replication corresponding to "steady state." Then the X_j's are IID random variables with $E(X_j) \approx \upsilon$, $\bar{X}(n')$ is an approximately unbiased point estimator for υ, and an approximate $100(1 - \alpha)$ percent confidence interval for υ is given by

$$\bar{X}(n') \pm t_{n'-1,1-\alpha/2} \sqrt{\frac{S^2(n')}{n'}} \qquad (4)$$

where $\bar{X}(n')$ and $S^2(n')$ are computed from Eqs. (1) and (2), respectively.

In some situations, it should be possible to use the initial n pilot runs of length m observations to both determine l and to construct a confidence interval. In particular if m is significantly larger than the selected value of the warmup period l, then it is probably safe to use the "initial" runs for both purposes. Since Welch's graphical method is only approximate, a "small" number of observations beyond the warmup period l might contain significant bias relative to υ. However, if m is much larger than l, these biased observations will have little effect on the overall quality (i.e., lack of bias) of X_j (based on $m - l$ observations) or $\bar{X}(n)$. Strictly speaking, however, it is more statistically correct to base the replication/deletion approach on two independent sets of replications.

Example 8. For the manufacturing system of Example 7, suppose that we would like to obtain a point estimate and 90 percent confidence interval for the steady-state mean hourly throughput $\upsilon = E(N)$. From the $n = 10$ replications of length $m = 160$ hours used there, we specified a warmup period of $l = 24$ hours. Since $m = 160$ is much larger than $l = 24$, we will use these same replications to construct a confidence interval. Let

$$X_j = \frac{\sum_{i=25}^{160} N_{ji}}{136} \qquad \text{for } j = 1, 2, \ldots, 10$$

Then a point estimate and 90 percent confidence interval for υ are given by

$$\hat{\upsilon} = \bar{X}(10) = 59.97$$

and

$$\bar{X}(10) \pm t_{9,0.95}\sqrt{\frac{0.62}{10}} = 59.97 \pm 0.46$$

Thus, in the long run we would expect the small factory to produce an average of about 60 parts per hour.

The half-length of the replication/deletion confidence interval given by (4) depends on the variance of X_j, $\mathrm{Var}(X_j)$, which will be unknown when the first n replications are made. Therefore, if we make a fixed number of replications of the simulation, the resulting confidence-interval half-length may or may not be small enough for a particular purpose. We know, however, that the half-length can be decreased by a factor of approximately 2 by making four times as many replications.

6. STATISTICAL ANALYSIS FOR STEADY-STATE CYCLE PARAMETERS

Suppose that the output process $Y_1, Y_2, \ldots$ does not have a steady-state distribution. Assume, on the other hand, that there is an appropriate cycle definition so that the process $Y_1^C, Y_2^C, \ldots$ has a steady-state distribution F^C, where Y_i^C is the random variable defined on the ith cycle (see Section 3). If $Y^C \sim F^C$, then suppose we are interested in estimating the mean $\upsilon^C = E(Y^C)$. Clearly, estimating a steady-state cycle parameter is just a special case of estimating a steady-state parameter, so the techniques of Section 5 apply, except to the *cycle* random variables Y_i^C rather than to the original Y_i's. Thus, we can use Welch's method to get a warmup period and then apply the replication/deletion approach to obtain a point estimate and confidence interval for υ^C. See Law and Kelton [1991, pp. 565-568] for an example.

REFERENCES

Law, A.M. and W.D. Kelton (1991), *Simulation Modeling and Analysis, Second Edition*, McGraw-Hill, New York.

Welch, P.D. (1983), "The Statistical Analysis of Simulation Results," *The Computer Performance Modeling Handbook*, S.S. Lavenberg, ed., pp. 268-328, Academic Press, New York.