Proceedings of the 2002 Winter Simulation Conference
E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds.

A COMPARISON OF SELECTIVE INITIALIZATION BIAS ELIMINATION METHODS

Jennifer R. Linton

Accenture
Supply Chain Practice
161 North Clark Street
Chicago, IL 60601, U.S.A

Catherine M. Harmonosky

Harold and Inge Marcus Depart. of Industrial and
Manufacturing Systems Engineering
Pennsylvania State University
304 Leonhard Building
University Park, PA 16803, U.S.A

ABSTRACT

When simulating a non-terminating system, the issue of initialization bias must be addressed. Many approaches have been developed to remove initialization bias from the output data. This paper provides a comparison of 5 selected methods applied to two slightly different 2-machine flow shop models. The experiment tests for statistical differences between mean and variance of the data used by each method to calculate steady state performance measures. Additionally, for each method, the practicality and ease-of-use for general applicability in larger modeled environments is discussed.

1 INTRODUCTION

Typically, simulation models of any system begin with the model in the empty and idle state. When simulating a terminating system, a system with defined start and finish conditions, this is usually an acceptable starting condition and the output data averages reported for each replication can be assumed to be an accurate representation of the real system. However, this is not true for simulations of non-terminating systems, which run continuously at steady state conditions.

For example, a hospital emergency room runs continuously every day, twenty-four hours a day. An analyst may wish to create a simulation model of the system to see what would happen if the hospital added more beds. When the simulation begins, all of the beds are empty. During the hospital's actual operation, however, this is rarely true. Eventually, in the model the beds will begin to fill up and the model will correctly simulate the on-going steady state hospital conditions. When simulating this, or any, non-terminating system, it is desired that only simulated data based on steady state conditions, which are the true conditions of the physical system, be analyzed. The physical system never initializes, or starts-up, but the simulation model must. As a result of this discrepancy, the data collected by the simulation model will be biased.

The data collected in the very beginning of the simulation, the initialization period or transient period, is typically statistically different from the data when the simulation comes to conditions of steady state. In the transient period, performance measures such as average flowtime, average time in queues, and average entities in the system are often lower than the rest of the simulated time. If this data is kept with the rest of the recorded data, the average for the total run will be biased and not an accurate predication of true steady state performance.

Figure 1 shows a simple simulation of a non-terminating system that was run for 60 time units. Each individual customer's time in the system is plotted on the graph at the time it exited the system. In this simulation, it can clearly be seen that there was a transient period before steady state was achieved. In this transient period of 10-15 time units, the reported values are approximately 15 time units smaller than the subsequent steady state data.

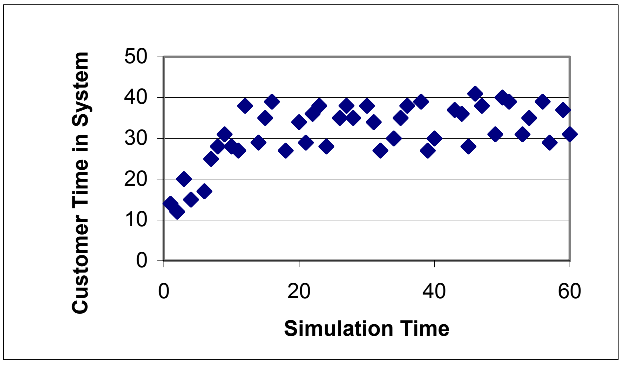

Figure 1: Customer Time in System vs. Simulation Time

Unfortunately, the transient period is not usually as clear and as easy to detect as it is in Figure 1. Therefore, many methods for detecting and dealing with the initialization bias problem have been developed. This paper will examine five such methods. One common method is to simply run the simulation for such a long time that the initial biased data will be statistically dominated by the large amount of subsequent steady state data (Law and Kelton 2000, p200). This method will be referred to in this paper as the Extra Long Replication Method. Other methods are designed to locate the time unit where the transient period ends so that all data points recorded before that point can be truncated from the rest of the data. The specific truncation methods that will be analyzed in this experiment are Welch's method (Welch 1983), the Relaxation Time Heuristic (Roth 1994), Kelton and Law's method (Kelton and Law 1983), and the Marginal Confidence Rule (White 1997).

These methods will all be compared in terms of statistical differences in mean and variance of the output data used to estimate steady state behavior. Also, each method will be evaluated as to its practicality for the analyst, because eliminating initialization bias should ideally be a small percentage of the analyst's time and effort.

2 INITIALIZATION BIAS ELIMINATION METHODS

A significant amount of research has focused on overcoming the initialization bias problem in discrete event simulation (Wilson and Pritsker 1978, Ripley 1988, Law and Kelton 2000), including graphical and mathematical approaches. Some work has focused on detecting initialization bias, such as Yucesan (1993) and Ma and Kochar (1993); but these may be difficult to easily use for choosing a point of truncation to eliminate the bias. Other methods present ways to find an appropriate truncation point, where all output data prior to this point (presumably the non-steady state data) will be deleted from system performance measure calculations. But some truncation point methods involve complex calculations, such as a chaos theory approach by Lee and Oh (1994). Others, such as Rossetti and Delaney (1995), are difficult to apply to the specific models in this experiment.

Since part of the purpose of this paper is to consider methods that are not only effective but also practical for general use, we limited consideration to methods that could be understood in a reasonably small amount of time. Consequently, the five methods described in the next five sections are chosen for this study.

2.1 Extra Long Replication Method

The Extra Long Replication Method, as described earlier, is the only method being tested in this paper that does not designate a point where early data is truncated. Instead, the method calls for running each model replication for such a long length of time that any bias that exists in the beginning of the run is statistically negligible because of the large number of data points taking place after it (Law and Kelton, p200). Theoretically, this method would work best if the transient stage passed quickly. This method is clearly easy to implement since no truncation point calculations must be performed. One must be careful, however, to run the replication for a long enough time to overcome statistical bias. Depending on the simulation model, the appropriate length of time to run the system may be difficult to estimate quickly.

2.2 Welch's Method

Welch's Method (1983) is a graphical, visual method of choosing a truncation point, that involves averaging observations across multiple replication, plotting the moving average of these values, and choosing a truncation point, l, where it appears that the transient period has passed (Law and Kelton, 2000). To begin, simulate n replications, each of length m (where m is large). If Y_{ji} is the ith observation from the jth replication (j = 1, 2, …, n; i = 1, 2, …, m), then let $\bar{Y}_i = \sum_{j=1}^{n} Y_{ji} / n$ for i = 1, 2, …, m (Law and Kelton, 2000). The averaged replications have observations of $\bar{Y}_1$, $\bar{Y}_2$, … $\bar{Y}_m$, means $E(\bar{Y}_i) = E(Y_i)$, and variances $Var(\bar{Y}_i) = Var(Y_i)/n$. To smooth out the high frequency oscillations in the observations, a moving average $\bar{Y}_i(w)$ is taken, where w is the window and is a positive integer such that $w \le \lfloor m/4 \rfloor$. $\bar{Y}_i(w)$ is defined in Equation (1).

$$\bar{Y}_i(w) = \begin{cases} \dfrac{\sum_{s=-w}^{w} \bar{Y}_{i+s}}{2w+1} & \text{if } i = w+1,\ldots,m-w \\[2ex] \dfrac{\sum_{s=-(i-1)}^{i-1} \bar{Y}_{i+s}}{2i-1} & \text{if } i = 1,\ldots w \end{cases} \tag{1}$$

Next, $\bar{Y}_i(w)$ is plotted for i = 1, 2, …, m-w. To choose the truncation point, the analyst finds the value of i where it appears by visual inspection that the $\bar{Y}_i(w)$ values have converged to steady state. This value of i is the truncation point l.

Welch's Method is easy to implement and is not generally time-consuming. It can be applied to any simulation type, and has an advantage of one using his or her own logic and knowledge of the system to choose the truncation point.

2.3 Relaxation Time Heuristic

The Relaxation Time Heuristic (RTH) is a mathematical method intended for simulations with M/M/k queueing systems (Roth 1994). A variable τ_R, the time constant, is calculated based on the characteristics of the system, according to Equation (2):

$$\tau_R = [1.4k\mu\,(1-\rho)^2]^{-1} \qquad (2)$$

where k is the number of servers, μ is the mean service rate, and ρ is the traffic intensity. The time constant represents the point in time where 99% of the initialization bias has dissipated. The truncation point l is then equal to $4\tau_R$. In theory, this will ensure that enough initialization bias will be removed so that any remaining bias will be negligible. Because this method uses system characteristics in Equation (2), the truncation point is independent of the performance measure considered.

The RTH method is also not time-consuming nor is it difficult to implement, especially since it is independent of the performance measure. It is, however, dependent on the model being an M/M/k queueing system, which won't be characteristic of most physical systems being simulated. It also lacks the possible advantage of user input, as the truncation point is chosen before the model is run.

2.4 Kelton and Law's Method

Kelton and Law's Method (1983) is based on the theory that the slope of the output from a non-terminating system at steady state will not be distinguishable from zero. By testing the slope of the time series of a group of replications, starting from the end, one will reach a point where the slope is no longer zero. All data points that occur before this point are then considered to be biased and should be truncated.

To begin, j independent replications of length m_0 points each are created. The next step is to average over the replications to obtain a single time series $\overline{X}_1,\ldots,\overline{X}_m$ where $m = m_0$. Then, the data is divided in half, and the average slope over every pair of points from $\overline{X}_{m/2}$ to $\overline{X}_m$ is calculated and tested to determine whether the average slope is zero. If it is, then m_0 is large enough and does not need to be increased. If not, m_0 should be increased until the slope of the last half of the data is nearly zero.

To find the truncation point, the data set on which the slope test must be performed should be steadily moved toward the beginning of the data until the average slope of all point pairs is no longer zero. After $\overline{X}_{m/2}$ to $\overline{X}_m$ is tested, $\overline{X}_{m/2-5}$ to $\overline{X}_{m-5}$ should be tested. The point where the slope is no longer zero is the truncation point.

Kelton and Law's method has two important disadvantages. Since many slope tests must be performed, the method is time consuming and uses significant computer memory and disc space to store all of the paired slopes.

2.5 Marginal Confidence Rule

The Marginal Confidence Rule (MCR) states that the truncation point, l, should occur at the observation point where the length of the confidence interval about the truncated sample mean is minimized (White 1997).

Given a stochastic sequence of output i of replication j $\{Y_i(j)$: $i = 1, 2, \ldots, n\}$, the optimal truncation point is defined as

$$d(j)^* = \underset{n>d(j)\geq 0}{\arg\min}\left[\frac{z_{\alpha/2}\,s(d(j))}{\sqrt{n(j)-d(j)}}\right] \qquad (3)$$

where $z_{\alpha/2}$ is the value of the unit normal distribution with a 100(1-α) percent confidence interval, j is the replication number, $s(d(j))$ is the sample standard deviation of the reserved sequence (all data following $d(j)$), and $n(j)$ is the total number of observations in replication j. Since the confidence level is fixed, $z_{\alpha/2}$ is a constant, it can then be given a value of 1, since the purpose of using the equation is only to compare all data points to find the minimum.

Although this method is certainly theoretically valid, it is not user-friendly as many calculations are involved for each observation point of each replication, rather than a time series of all the replications averaged together.

3 COMPARISON OF ELIMINATION METHODS

The five initialization bias elimination methods selected for this study are tested using 2 slightly different models and 2 different performance measures, average time in the system and average number of customers in the system. The output data used by each technique to calculate steady state performance measures is tested for equality of variance by Levene's test and tested for equality of mean by a two-sample t-test.

3.1 Description of the Experimental Models

Although initialization bias removal techniques are often evaluated in the literature using a single queue system, most applications of simulation are much larger systems. Therefore, this paper uses a slightly larger system that allows insight into potential practical implementation issues of the methods yet it is small enough to allow for the comparison to be done in a timely manner. In this research, one basic model is used with two different types of arrival and processing distributions. Using two different distributions allows the experiment to test whether or not the methods work both on simulations following the standard queueing model distribution assumptions and on those that violate the assumptions.

The basic model contains two servers in sequence, each with its own infinite queue, with all entities served by server

1 then server 2. Model One retains some standard queueing model characteristics, with exponential inter-arrival and processing distributions at each server. The inter-arrival times between entities entering the system follow an exponential distribution with an average of 8 time units. Server 1's processing time is exponentially distributed with a mean of 6 and server 2's processing time is exponentially distributed with a mean of 7. Making Server 2's average processing time longer causes the queue to build at Server 2. Model Two is the same as Model One except that the distributions are triangular instead of exponential. The inter-arrival time distribution is triangular with a minimum of 6, mode of 8, and maximum of 10. Server 1 and 2's processing time distributions are both triangular with minimum of 4, mode of 6 and maximum of 8 for server 1 and minimum of 5, mode of 7 and maximum of 9 for server 2.

3.2 Statistical Tests

Two statistical tests are used to determine if different methods yield similar steady state output statistics: Levene's test for equality of variance and the two-sample t-test for equality of mean. Minitab 13.31 statistical software is used for both tests.

3.2.1 Levene's Test for Equality of Variance

First, the variance of each replication is compared to the data of the same replication from the other four methods, using Levene's test for equality of variance (Filliben and Heckert 2001). The null hypothesis of the test states that the variances of the two populations are equal. Given a variable Y (observation for a particular performance measure) with sample of size N divided into k subgroups (replications) where N_i is the sample size of the ith subgroup, Levene's test statistic is:

$$W = \frac{(N-k)\sum_{i=1}^{k} N_i(\overline{Z_i.} - \overline{Z..})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{N_i}(Z_{ij} - \overline{Z_i.})^2} \tag{4}$$

where $Z_{ij} = |Y_{ij} - \tilde{Y}_i|$, $i = 1, \ldots, k, j = 1, \ldots, N_i$ and $\tilde{Y}_i$ = median $\{Y_{i1},\ldots,Y_{in}\}$. Levene's test is used to test paired replication data between each of the five methods for time in the system and customers in the system in both models.

3.2.2 Two-Sample t-test for Equality of Means

Once Levene's test indicates the replication pair does not have significant difference in variance, then the pair can be tested to see whether their means are different. Pairs that do have differences in variance do not need to be tested for equality of means because, in this experiment, having differing variances is enough information to declare the results of the truncation methods unequal.

To test for equality of means for two samples, the two-sample t-test is used. The null hypothesis of the test states that the difference between the two means is equal to a variable, δ_0. The test statistic, t, is defined as:

$$t = ((\overline{X}_1 - \overline{X}_2)) - \delta_0) / s \tag{5}$$

where s, since the variances are assumed to be equal, is defined as:

$$s_p{}^2 = \frac{(n_1-1)s_1{}^2 + (n_2-1)s_2{}^2}{n_1+n_2-2} \quad . \tag{6}$$

The test statistic degrees of freedom are (n_1+n_2-2).

Since this test is only conducted on pairs that did not have statistically different variances, any pairs that do not also have statistically different means are considered replication pairs where the two bias elimination methods do not differ significantly in their results. Conversely, pairs that have either differing variances or differing means are pairs where the methods produce different statistically significant results.

3.3 Implementation of Methods for Experimental Models

This section explains how each method was implemented for the two models and the two performance measures. In all cases except for the Extra Long Replication Method, each model was run for 10 replications at 5000 time units. For the Extra Long Replication Method, the run time was tripled to 15000 time units to statistically eliminate bias, as described above.

The Extra Long Replication method needs no modification for model type or performance measure. Welch's method is applied to each combination of model and performance measure, which may yield different truncations points for each combination. Similarly, Kelton and Law's method is also applied to each combination of model and performance measure. This method requires replications to have a standard length based on observation points, which results in different truncation points for each combination. The statistical package Minitab 13.31 was used to analyze the slope. The marginal confidence rule needs no modification. It uses Equation (3) to determine a truncation point for each replication of each model for both performance measures. The maximum of the two points is then used as the actual truncation point.

Because the Relaxation Time Heuristic (RTH) is intended for use in simulations of M/M/k systems, a slight modification is needed here. Since Model One is similar to two M/M/1 models in tandem, calculation of τ_R is done twice, once with each server. For Server 1, λ_1 is equal to 1 divided by the average inter-arrival time, and μ_1 is equal to 1 divided by the average processing time of Server 1. For

Server 2, μ_2 is again equal to 1 divided by the average processing time of that server. According to Ravindran et al. (1987, p 314), the mean throughput rate, R, of a server is equal to the mean arrival rate to that server, λ, assuming the mean arrival rate is constant and every arrival is accepted. Therefore, $R_1 = \lambda_1$. Since Server 2 accepts all entities directly after they exit Server 1, $\lambda_2 = R_1 = \lambda_1$. After calculating τ_{R1} and τ_{R2}, the larger value is selected as τ_R, which is used to choose the truncation point at time $4\tau_R$. This procedure is only performed once for Model One, because all replications have the same system parameter settings and the method is independent of performance measure (see Equation (2)).

Model Two uses triangular distributions, so it cannot be considered a traditional queueing model. However, the RTH is applied to Model Two to test its efficacy on a more commonly used type of simulation model. The same approach for assigning values to variables is used as in Model One, except that the modes of the triangular distributions for the inter-arrival times and processing times are used instead of mean values, as in exponential distributions. The largest value for τ_R is used, which is then multiplied by 4 to obtain the truncation point. As with Model 1, the same truncation point is used for each replication and each performance measure.

4 DISCUSSION OF RESULTS

The actual truncation point determined by each of the 5 methods was different. Table 1 summarizes each truncation point by method, model and performance measure. All methods except the marginal confidence rule specify the same truncation point for all replications of a specific model and specific performance measure. The MCR specifies a truncation point for each replication, therefore the range of truncation points over the 10 replications is reported. The RTH specifies a time-based truncation method for all performance measures, while the other methods may specify number of observations (# obs) for observation-based measures and specify a time (time) for time-persistent measures.

Looking across methods in any given column, the differences in truncation points are apparent. For example, Model 1 with the time in system measure, the Kelton and Law method point is twice the Welch point and for some replications the MCR point is over 3 times the Welch point.

Looking at the same method and performance measure across the 2 models, there is also a difference for 3 methods. Welch, RTH and MCR have larger truncation points for Model 2 compared to Model 1, with differences ranging from approximately 16%-29%. On the other hand, the MCR has virtually the same range of truncation points for time in system, but a decrease of about 14% in the upper limit of the range of truncation points for number of customers in the system.

Table 1: Truncation Points

	Time in the System	**Customers in the System**	**Time in the System**	**Customers in the System**
	Model 1		**Model 2**	
Extra Long Replication	0	0	0	0
Welch	125 # obs	1050 time	161 # obs	1140 time
Relaxation Time Heuristic	1280 time	1280 time	1280 time	1280 time
Kelton and Law	250 # obs	1938 time	314 # obs	2215 time
Marginal Confidence Rule	range: 1-404 # obs	18-3343 time	range: 1-409 # obs	14-2880 time

Although it is visibly evident that the methods choose different truncation points for every replication of both models, it is important to know if the different points create statistical differences in the data that is used to calculate steady state performance measures. Table 2 shows the number of replication pairs between 2 methods that have the assumption of equal variances rejected (# Lev) or the assumption of equal means rejected (# t-test) categorized by model and performance measure. Because the main interest of this paper is knowing if the methods give different performance measure estimates, only the pairs that do not have statistically different variances are tested for equality of mean.

According to Levene's test, The Extra Long Replication Method has statistically different variances in most cases. A statistical difference is found in 145 out of the 160 pairs tested. Out of these 15 with no discernable difference, only 5 scattered replication pairs had no statistical difference in mean as well. Statistically, then, this method does not produce comparable results to any of the other four methods.

Welch's Method and Relaxation Time Heuristic, however, produce almost the same results in both models of this experiment. For Model One, both the variances and means for every replication of Welch's Method cannot be proven different than the matching replication of the Relaxation Time Heuristic. For Model Two, only replications 4 and 6 for the number of customers in the system data and replication 6 in the time in the system data are statistically different in both variance and mean.

Table 2: Number of Replication Pairs Rejected by Levene's Test (# Lev) and Additional Pairs Rejected by t-test (# t-test)

Methods Compared	**Time in System**	**Number in System**	**Time in System**	**Number in System**
	Model 1		**Model 2**	
EL vs. Welch				
# Lev	9	10	8	9
# t-test	1	0	1	0
EL vs. RTH				
# Lev	9	10	8	9
# t-test	0	0	2	0
EL vs. K&L				
# Lev	9	9	9	9
# t-test	0	1	1	1
EL vs. MCR				
# Lev	9	10	10	8
# t-test	1	0	0	2
Welch vs. RTH				
# Lev	0	0	1	1
# t-test	0	0	0	1
Welch vs. K&L				
# Lev	3	6	6	6
# t-test	3	3	3	3
Welch vs. MRC				
# Lev	3	5	4	4
# t-test	1	2	2	3
RTH vs. K&L				
# Lev	5	6	5	3
# t-test	2	3	3	5
RTH vs. MCR				
# Lev	4	3	3	4
# t-test	0	4	3	3
K&L vs. MCR				
# Lev	6	7	5	3
# t-test	1	1	4	5

Welch's Method compared to Kelton and Law's Method in Model One is only statistically similar in both variances and means in 5 out of the 20 replications pairs. Only 2 of the 20 replication pairs of Welch's Method paired with Kelton and Law's Method are not statistically different in Model 2. Since so few of the paired replications of the two methods have statistically similar results, Kelton and Law's Method in general does not produce similar results to Welch's Method.

Comparing Welch's Method and the Marginal Confidence Rule, less than half of the replication pairs between the two methods do not have statistically different results. The replication pairs of Model One have similar results 9 times out of 20. Considering Model Two, only 7 out of the 20 are not significantly different. Thus, Welch's Method and the MCR do not produce similar results.

When comparing the Relaxation Time Heuristic and Kelton and Law's Method, it is clear that the two methods also produce different results. In both Model One and Model Two, 4 out of 20 replication pairs are not statistically different. The Relaxation Time Heuristic can also be considered different than the Marginal Confidence Rule. Nine out of the 20 replication pairs in Model One and 7 out of the 20 pairs in Model Two between the Relaxation Time Heuristic and the Marginal Confidence Rule are not statistically different.

Finally, Kelton and Law's Method appears to produce different results than the Marginal Confidence Rule. Only 5 out of the 20 replication pairs from Model One are similar. Also, only 3 of the 20 pairs from Model Two are similar.

These results of the tests for differences in variances and means show that only Welch's Method and the Relaxation Time Heuristic yield the same results when used to eliminate bias in Models One and Two.

Although Levene's Test and the Two-Sample t-Test can help determine if the bias elimination methods choose different truncation points that produce statistically different estimates of steady state performance measures, they cannot conclude which one gives a better estimate of the true steady state behavior of the system than the others. Further study could be conducted that might include additional experimentation, such as a test to detect bias after the truncation method has been performed so that a confidence interval comparison could be done.

However, based on experience carrying out each method for this research, some comments can be made regarding the practicality of using the methods for simulations of larger, "real-world" applications. The Extra Long Replication Method is certainly the simplest technique, but its effectiveness is very dependent upon having a sufficiently long run length to completely dominate the initialization bias. Ideally, you should spend some time determining where the steady state data may begin (perhaps graphically) to determine a good run length.

The MCR and Kelton and Law's method both require a large number of calculations. Although the MCR is intuitively appealing because it determines a truncation point for each replication, this adds to the computational burden. Similarly, testing for zero slope via Kelton and Law is time consuming and computationally burdensome.

Welch's Method and RTH do yield comparable results. Both are easy to understand and perform. However, because the RTH is based on a M/M/k system, the more the modeled system moves away from emulating a Jacksonian network, it may be more difficult to calculate the RTH truncation point and results might prove more divergent from Welch. Therefore, practically speaking, Welch's method might be a good choice.

5 SUMMARY

Statistical tests prove that only Welch's Method and the Relaxation Time Heuristic produce comparable results with Models One and Two. The other methods, the Extra Long Replication Method, the Marginal Confidence Rule, and Kelton and Law's Method, all result in statistically different data than every other method.

All of the methods are found to be able to adjust to changes in distributions of inter-arrival times and server processing times. Also, the Relaxation Time Heuristic method does not seem to be negatively affected by the modifications that are necessary to apply it to Models One and Two.

Although tests to properly determine whether each method chooses a correct and efficient truncation point would require further research, the steps required to carry out each method may be compared for practicality. The Marginal Confidence Rule and Kelton and Law's Method both are computationally intensive. The Extra Long Replication Method is inconclusive in its results for this experiment. The most practical methods are Welch's Method and the Relaxation Time Heuristic. Because Welch's Method is not based on any assumptions about the type of system being modeled, it may be a good practical choice as an initialization bias removal method.

REFERENCES

Filliben, J. J. and A. Heckert. 2001. Exploratory Data Analysis from *NIST/SEMATECH Engineering Statistics Internet Handbook* [on line]. Available via: `http://www.itl.nist.gov/div898/handbook/eda/eda.htm` [accessed May 15, 2001].

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

Law, A. M., and W. D. Kelton. 2000. *Simulation Modeling and Analysis*. 3rd ed. New York: McGraw-Hill, Inc.

Lee, Y., and H. Oh. 1994. Detecting truncation point in steady-state simulation using chaos theory. In *Proceedings of the 1994 Winter Simulation Conference*, ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila, 353-359. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Ma, X., and A. K. Kochhar. 1993. A Comparison study of two tests for detecting initialization bias in simulation output. *Simulation* 61(2):94-101.

Ravindran, A., D. T. Phillips, and J. J. Solberg. 1987. *Operations Research: Principles and Practice.* New York: John Wiley & Sons.

Ripley, B.D. 1988. Simulation methodology-an introduction for queueing theorists. *Queueing Systems* 3:201-220.

Rossetti, M. D., and P. J. Delaney. 1995. Control of initialization bias in queueing simulations using queueing approximations. In *Proceedings of the 1995 Winter Simulation Conference*, ed. C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman. 322-329. Piscataway, New Jersey: Instutue of Electrical and Electronics Engineers.

Roth, E., 1994. The relaxation time heuristic for the initial transient problem in M/M/k queueing systems. *European Journal of Operational Research* 72:376-386.

Welch, P. D. 1983. The statistical analysis of simulation results. In *Computer Performance Modeling Handbook*, ed. Lavenberg, 267-329.

White, K. P., Jr.. 1997. An effective truncation heuristic for bias reduction in simulation output. *Simulation,* 69(6):323-334.

Wilson, J. R., and A. A. B. Pritsker. 1978. A survey of research on the simulation startup problem. *Simulation* 31(3):55-58.

Yucesan, E. 1993. Randomization tests for initialization bias in simulation output. *Naval Research Logistics* 40:643-663.

AUTHOR BIOGRAPHIES

JENNIFER R. LINTON received her B.S. in Industrial Engineering from Penn State University in 2001. She is a member of Alpha Pi Mu and IIE. Her professional experience includes management engineering and global purchasing analysis. She has just begun her career at Accenture, specializing in supply chain consulting. Her email address is `<jennifer.r.linton@accenture.com>`.

CATHERINE M. HARMONOSKY is an Associate Professor in the Harold and Inge Marcus Department of Industrial and Manufacturing Engineering at Penn State University. She is the Publicity Chair for the 2002 Winter Simulation Conference. Her research and teaching interests are in simulation and production planning and control. She is a member of IIE, SME and SWE. Her email address is `<cmhie@engr.psu.edu>`.