Proceedings of the 2004 Winter Simulation Conference
R .G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, eds.

EVALUATION OF METHODS USED TO DETECT WARM-UP PERIOD IN STEADY STATE SIMULATION

Prasad S. Mahajan
Ricki G. Ingalls

School of Industrial Engineering and Management
322 Engineering North
Oklahoma State University
Stillwater, OK 74078, U.S.A.

ABSTRACT

This paper reviews the performance of various methods used to detect the warm up length in steady state discrete event simulation. An evaluation procedure is used to compare the methods. The methods are applied to the output generated by a simple job shop model. The performance of the methods is tested at different levels of utilizations. Various measures of goodness are used to assess the effectiveness of the methods.

1 INTRODUCTION

The goal of steady state simulation studies is to estimate the long run characteristics of the system. In order to estimate the long run performance measures, the simulation model is run for a certain period of time. Most of the simulation models are started empty and idle. Almost every time, these conditions differ from the steady state condition. Due to this, the simulation model takes some time to reach steady state. During this time period the model is said to be in transient state. The observations collected during this time period may affect the accuracy of the estimates of the performance measure, if the transient state lasts for a relatively long time. This problem is called as the *initialization bias* or the *start-up* problem in simulation literature. One of the ways to overcome this problem is to run the simulation model for a time period L called as the warm-up length. Reset all statistics after time L and start recording observations for a period of m - L, where m is the run length. Another difficulty here is in predicting the warm-up length?

Over the years many authors have come up with various methods that determine the warm up length. These methods can be broadly classified into four groups:

I. Graphical
II. Statistical
III. Heuristic
IV. Initialization Bias

The statistical methods can be further classified into basic and advanced. Advanced methods involve time series analysis. Robinson (2002) categorizes time series methods and other complex methods in a separate group called advanced methods.

The goal of this research is to compare some of the methods from groups I, II and III based on their performance. Initialization Bias tests are not considered. The performance will be judged using five measures of goodness. This research will provide answers to following: Which methods work under which conditions and which methods fail? If some methods work well for a particular condition, which one of them is the most effective? Are there any modifications, which when applied to the methods will improve their performance?

Similar research was carried out by Wilson and Pritsker (1978), Gafarian, Ancker, and Morisaku (1978) and Cash *et al*., (1992).

The remaining paper is organized in four sections. Section 2 gives a brief explanation of the methods used, Section 3 explains the experimental model, run conditions, measures of goodness used, procedure and run conditions. Results are presented in Section 4 and conclusion is Section 5.

2 A BRIEF REVIEW OF THE METHODS

This paper compares the performance of the six methods. Table 1 below shows the methods and the respective group. In the proceeding sections, we give a brief explanation of the six methods. For a detailed explanation, refer to Mahajan (2004) or respective references.

2.1 Welch's Method

This is the simplest and most general technique used for determining the warm-up length. It is a graphical technique that requires multiple replications. The Welch's

Table 1: Warm-Up Length Methods

Method	Reference	Group
Welch's Method	Welch (1983)	Graphical
SPC Method	Robinson (2002)	Graphical
Randomization Test	Yücesan (1993)	Statistical
Conway Rule	Gafarian, Ancker, and Morisaku (1978)	Heuristic
Crossing of the Means Rule	Gafarian, Ancker, and Morisaku (1978)	Heuristic
Marginal Standard Error Rule -5	White, Cobb, and-Spratt (2000)	Heuristic

method is explained in the following steps as given in Law and Kelton (2000) :

1. Make n replications of the simulation each with run length m. Let Y_{ji} be the i^{th} observation from the j^{th} replication. Thus i takes values from 1 to m and j from 1 to n.
2. Calculate the ensemble averages over the replications. These will be $\overline{Y}_i$'s where

$$\overline{Y}_i = \sum_{j=1}^{n} \frac{Y_{ji}}{n} \text{ for } i = 1, 2, ..., m \tag{1}$$

3. Define a moving average $\overline{Y}_i(w)$ to smooth out the high frequency oscillations in $\overline{Y}_1, \overline{Y}_2, ..., \overline{Y}_m$. w is the window and is a positive integer. w is less than or equal to $m/4$. $\overline{Y}_i(w)$ is as follows:

$$\overline{Y}_i(w) = \frac{\sum_{s=-w}^{w} \overline{Y}_{i+s}}{2w+1} \tag{2}$$
$$\text{for } i = w+1, w+2, ..., m-w$$

$$\overline{Y}_i(w) = \frac{\sum_{s=-(i-1)}^{i-1} \overline{Y}_{i+s}}{2i-1} \quad \text{for} \quad i = 1, 2, 3, ..., w \tag{3}$$

4. Plot $\overline{Y}_i(w)$, for $i = 1, 2, ..., m-w$ and choose L to be that value of i beyond which $\overline{Y}_i(w)$ appears to be converged. See Welch (1983) for an aid in determining convergence.

2.2 SPC Method

Robinson (2004) explains this method in four steps. This method requires multiple replications. Let Y_{ij} denote the observation from each replication, where i is the observation number and j is the replication number. Calculate the ensemble averages, that is similar to equation 1.

Ensemble averages $\{\overline{Y}_i : i = 1,2,..., m\}$ are batched in b batches of size k. The batch means are represented as $\overline{\overline{Y}}_x$, $x = 1,2,..., b$. The batch size is so selected that the resulting batch means pass the Anderson Darling test for normality and Von Neumann test for correlation. Minimum 20 batch means are recommended. For more information refer to Robinson (2002).

The resulting time series after batching is represented as:

$$Y_{(k)} = \{\overline{\overline{Y}}_1(k),..., \overline{\overline{Y}}_b(k)\} \tag{4}$$

A control chart is generated for the above time series. The estimates of population mean (μ) and standard deviation (σ) are calculated from the last half of the series $y(k)$. After calculating the mean and the standard deviation, the control limits are calculated using the formula:

$$CL = \hat{\mu} \pm z\hat{\sigma}/\sqrt{b/2}, \text{z}=1,2 \text{ and } 3 \tag{5}$$

A control chart is constructed showing the three control limits, the mean ($\hat{\mu}$) and the time series $Y_{(k)}$. When the rules below are true, it can be said that the time series is out-of-control. Steady state is reached when the process is in-control and remains in-control. Following are the rules to identify if the process is out-of-control:

- A point plots outside a 3-sigma control limit.
- Two out of three consecutive points plot outside a 2-sigma control limit
- Four out of five consecutive points plot outside a 1-sigma control limit
- Eight consecutive points plot on one side of the mean
- Initial points all plot to one side of the mean (as per expected bias)

2.3 Randomization Test

Yücesan (1993) presented a method to detect the initialization bias. It is based on randomization tests. Yücesan formulated the problem of initialization bias in a hypothesis testing framework concerning the mean of the process. Randomization tests are applied to test the null hypothesis that mean of the process is unchanged throughout the run. The advantage of using this method is that, no assumptions, like that of normality are required. The null hypothesis for this method is that there is no initialization bias.

The steps needed to perform this test are summarized below:

1. Run the simulation for a length of time m hours. Let Y_i be the i^{th} observation from the simulation output which is run for m hours.
2. Obtain an output time series $Y_1, Y_2, ..., Y_m$.
3. Batch the data into b batches of length k.
4. Obtain b batch means $\overline{Y}_1, \overline{Y}_2, \overline{Y}_3, ..., \overline{Y}_b$.
5. Partition the batch means into two groups. For the first iteration the first group must include the first batch mean and the second group should contain remaining b-1 batch means.
6. For each iteration, the grand means of the two groups are compared. If the difference between the two grand means is significantly* different from zero, the null hypothesis is rejected.
   **To access the significance a distribution of difference is required. Since it is unknown, randomization is used.*
   By using randomization, an empirical distribution is obtained and the original observed difference is seen far in the tail.
7. If the hypothesis is rejected, the groups are rearranged; second batch is added to the first group and the second group will contain (b-2) batch means and step 6 is repeated.
8. If hypothesis is accepted then the group 2 data is the steady state simulation output.

2.4 Conway Rule

Conway (1963) suggested the following rule to truncate the initial data in order to reduce bias. "Truncate a series of measurements until the first of the series is neither the maximum nor the minimum of the remaining set."

This is done for a few pilot runs to decide upon a stabilization period. After this is done, the period is deleted from the result of each run.

Following algorithm is constructed using the steps given in Gafarian, Ancker, and Morisaku (1978):

1. Decide n and m the number of exploratory replications and the length of the exploratory replications.
2. Compute y_{jr}^{+} and y_{jr}^{-} using following formulae:

$$y_{jr}^{+} = \max(y_{jl} : l = r, ..., m) \; j = 1, ...n \tag{6}$$

$$y_{jr}^{-} = \min(y_{jl} : l = r, ..., m) \; j = 1, ...n \tag{7}$$

3. For r = 1, 2, …, m determine t_j such that $t_{j = \min_r} \{Y_{jr}^{-} < Y_{jr} < Y_{jr}^{+}\}$ occurs for the first time.
4. Estimate of the truncation point t^* is given by $\max\{t_1, t_2, t_3, \ldots, t_n\}$

2.5 Crossing of the Means Rule

This rule is stated in Fishman (1973). This rules states that:

> Compute the running cumulative mean as data are generated. Count the number of crossings of the mean, looking backwards to the beginning. If the number of crossings reaches a pre-specified value, which means you have reached the truncation point.

The pre-specified value depends on the user. Following algorithm is based on steps given in Gafarian, Ancker, and Morisaku (1978):

1. Generate the simulation output $\{Y_1, Y_2, \ldots, Y_m\}$ Define:

$$w_j = \begin{cases} 1, \text{ if } Y_j > \overline{Y}_m, \; Y_{j+1} < \overline{Y}_m \\ \quad \text{or } Y_j < \overline{Y}_m, \; Y_{j+1} > \overline{Y}_m \\ 0, \text{ otherwise} \end{cases} \tag{8}$$

where j = 1, 2, …, m-1

$$\overline{Y}_m = \frac{1}{m} \sum_{j=1}^{m} Y_j \tag{9}$$

2. The number of times the series crosses the mean is given by

$$\Omega_m = \sum_{j=1}^{m-1} w_j \tag{10}$$

3. Calculate $\Omega_1, \Omega_2, ..., \Omega_l$ such that at l the number of crossings is equal to the pre-specified number.

2.6 Marginal Standard Error Rule -5

For this method, the author defines m as batch size, n as the run length and b as the number of batches. MSER Rule (White (1997) states that for a finite stochastic process $\{Y_i(j): i$=1, 2, ..., $n\}$ the optimal truncation point is given by

$$d_j^* = \arg\min_{n > d(j) \geq 0} \left[\frac{z_{\alpha/2}\, s(d(j))}{\sqrt{n(j) - d(j)}} \right] \tag{11}$$

where $z_{\alpha/2}$ is the value of the unit normal distribution associated with a 100(1 - α) percent confidence. For a fixed confidence level, $z_{\alpha/2}$ is a constant. The expression then, can be written as:

$$d_j^* = \underset{n > d_j \geq 0}{\arg\min} \left[\frac{1}{(n(j) - d(j))^2} \sum_{i=d+1}^{n} (Y_{i(j)} - \bar{Y}(n,d)_{(j)})^2 \right] \tag{12}$$

For a given output sequence *d(j)** is determined by solving the unconstrained minimization problem defined by the above equation. The MSER heuristic is applied to the raw data where as the MSER-*m* rule is applied to *b* batch means where *b* = run length/batch size (*m*). For MSER-5 rule (Spratt 1998), the batch size is 5.

Codes for Welch's Method, Conway Rule, Crossing of the Means Rule and Randomization Test can be found in Mahajan (2004).

3 EXPERIMENTAL DETAILS

In this section we describe the experimental model, performance measures used to evaluate the method, model parameters and run conditions and the procedure.

3.1 Experimental Model

The methods are tested on simple job shop models. There are five cells, C_1, C_2, C_3, C_4 and C_5. Each cell has different number of machines (resources). There are three customer classes (three different types of parts). The overall arrival rate is Poisson (λ) hours and the service times for servers are exponential with means $\mu_{i,j}$ hours (*i* (*customer class*) = 1, 2, 3 and *j* (*cells*) = 1, 2, 3, 4, 5). The model is built and run in Arena 5 simulation software. The arriving parts are split into three types (*customer classes*), namely, type A, type B and type C with probabilities 0.5, 0.3 and 0.2 respectively. The parts get processed in the cells in different service times. After being processed the parts exit the system.

The methods are applied to the model with different levels of utilization.

Type I Model has a high level of utilization. The average utilization of all resources is close to 90%. The individual utilizations may vary from 80% - 95%.

Type II Model has a moderate level of utilization . The average utilization of all resources is close to 70%. The individual utilizations may vary from 65% to 80%.

Type III Model has a low level of utilization. The average utilization of all resources is close to 50%. The individual utilizations may vary from 45% - 65%.

3.2 Performance Measures to Evaluate the Methods

1. Mean square error (MSE)

$$\text{MSE} = \{\text{E}(\theta - \hat{\theta})^2\} + \text{Var}(\hat{\theta}) \tag{13}$$

where θ = Actual mean of simulation output data.

Last paragraph in Section 3.2 shows the procedure used to estimate the value of θ.

$\hat{\theta}$ = Mean of the remaining data left after deletion. If *L* is the warm up length and m is the run length, then :

$$\hat{\theta} = \frac{\sum_{i=L+1}^{m} Y_i}{m - L} \tag{14}$$

m is the run length, *L* is the warm up length and Y_i, for *i*= 1 to *m* is the value of the performance measure at *i* time units. A good method will yield a low value of MSE.

2. Variance

By deleting data, the variance of the point estimator is likely to increase. Thus it is necessary to assess the quality of deleted data based on the variance of $\hat{\theta}$.

3. Percentage change in mean square error

This calculates the percentage change in the initial mean square error (with no data deleted) after data deletion has been applied. The simulation output will have a MSE before data deletion has been applied. Let this be denoted by MSE_{ini}. After applying the warm up length and deleting *L* data points, MSE is calculated again. Let this be denoted by *MSE*

$$\begin{aligned} &\%\text{ change in MSE} = \\ &\quad 100 * (MSE_{fin} - MSE_{ini}) / (MSE_{ini}) \end{aligned} \tag{15}$$

A good method will always give a negative value of percentage change in MSE.

4. Percentage change in variance

It is the percentage change in the initial variance after data deletion has been applied. Initial variance is the variance of the simulation output data with no warm-up. Let this be denoted by V_{ini}. Final variance is the variance of the data left after applying the warm-up length.

$$\begin{aligned} &\%\text{ change in variance} = \\ &\quad 100 * (V_{fin} - V_{ini}) / (V_{ini}) \end{aligned} \tag{16}$$

5. Cost

   Cost is calculated in terms of average computer time. This is the sum of computer time required to collect data and computer time required to perform the method. This is measured in seconds.

Estimating the actual value of θ :

The models are run for a very long time in the range of 10^6 - 10^9 hours. Values of the performance measure, say θ_i for these long runs are observed. The values of θ_i for each long run are within 1% of each other. The average of θ_i's for all long runs is assumed to be a true estimate of θ for the given model.

For verification, the value of θ is also calculated theoretically using a queuing software "RAQS". For model type II and type III, the results obtained from RAQS and from the averages of long runs are within 1%. For type I model, the results are within 3.5%. This verifies the results for all the models.

3.3 Model Parameters

Run Conditions: System is started empty and idle. Initial run length is 1000 hours.

Performance measure: Number in system (Inventory) measured at the end of each hour. This performance measure is chosen arbitrarily. Results may vary for different performance measures.

3.4 Procedure

In this section we explain in detail the procedure used to obtain values for the measures of goodness. The methods are applied to type I, type II and type III models. Applying a method to a model constitutes one experiment.

Initially all the models are run for 1000 hours. Variance and MSE for this data are calculated against the actual value of Number in System.

For methods that require multiple replications, the model is run for 5 replications. The measures of goodness, time, MSE and variance are then calculated after applying the warm-up length.

Methods that do not need multiple replications, are applied to each of the 5 replications independently to get 5 different warm-up lengths. This will give 5 values of time, MSE and variance. The final values of time, MSE and variance are calculated as averages of the 5 values obtained from the five replications independently.

Based on the values of MSE and variance after applying the warm-up length prescribed by the respective method, the % change in MSE and % change in variance are calculated against the initial MSE and variance without data deletion.

Some of the methods demand for more data if a test for autocorrelation and/or normality fails. For such methods, more data is generated. The maximum run length required for any method when implemented on any model is noted. All models are run again with a run length equal to the maximum. By doing this all the methods are evaluated against a constant run length. Thus a total of 180 experiments are performed.

4 RESULTS

For each method, we have chosen certain specifications which are kept constant for all the experiments. These are mentioned in Section 4.1. Results are presented in Section 4.2 and 4.3. A method is said to perform well if it reduces both the MSE and variance and is also computationally efficient.

4.1 Implementation Specifications

4.1.1 Welch's Method

Welch's Method requires three parameters to be specified. The run length *m*, the number of replications *n* and the window size *w*. Law and Kelton (2000) suggest taking the minimum value of *w* that so that graph appears smooth. For all the models value of *w* = 10 was enough to obtain a smooth graph. Welch's procedure is a graphical procedure. The user decides the warm-up length by observing a graph of averaged values against time. Thus there is subjectivity involved in this procedure. To minimize this subjectivity, twenty five users were asked to observe the graphical output and give their values for the warm-up length. Finally, the warm-up length was calculated as the average of 25 values.

4.1.2 Conway Rule

This rule also requires multiple replications. The number of replications is set to 5.

4.1.3 Statistical Process Control Method

This method requires the data to be batched with a batch size such that the data is approximately normally distributed and has negligible serial autocorrelation. Initial batch size is kept 1. The batch size is doubled if either the test for autocorrelation or normality fails. The number of batches is at least 20. If the number of batches falls below 20, then the test demands more data. For this research, if the test demands more data, the run length is incremented by 200 and the test is re-started.

4.1.4 Crossing the Means Rule

This rule requires the user to decide the value of number of crossings. Gafarian, Ancker, and Morisaku (1978) used a value of three. The same value is used here.

4.1.5 MSER-5 Rule

This rule uses a batch size of 5. It needs only one replication. Refer to Spratt (1998) for more details.

Tables 2 and 3 show the results of experimental runs.

4.2 Results for Run Length = 1000 Hours

The results for 1000 hours run length are shown in Table 2. For Type I model, it can be seen that none of the methods work. Randomization Test and the SPC method need more data. But given more data they reduce both the MSE and variance. MSER-5 rule works well for type II model. For Type III model, MSER-5 Rule and SPC method work well. Randomization test reduces both the MSE and variance, however it takes more time to perform (750 seconds). For Type I model, the SPC method requires 5200 hours run length to be able to perform. All the experiments are run again for 5200 hours.

4.3 Results for Run Length = 5200 Hours

Table 3 shows the results when the models are run for 5200 hours. For Type I model, the SPC method, Welch's Method and the Randomization Test perform well. For model Type II the MSER-5 Rule and Crossing the Means Rule performs well. However the Randomization Test

Table 2: Results for Run Length = 1000 Hours

Method	Average Run length (hours)	Final MSE	Final Variance	Average Computing Time (seconds)	% Change in MSE	%Change in variance
Model Type I						
Welch (1)	1000	213.3846	1.8777	8	-33.42	13.33
SPC (1) (3)	5200	72.3188	0.2984	27	-77.44	-81.99
Conway (1)	1000	332.6243	1.6383	11	3.78	-1.12
MSER-5 (2)	1000	395.7777	1.9550	5	23.49	17.99
Randomization (2) (4)	1480	259.8652	1.1773	21	-18.92	-28.95
Crossing the means (2)	1000	266.0326	1.7685	22	-16.99	6.74
Initial (before truncation)	1000	320.4976	1.6569	5	0.00	0.00
Model Type II						
Welch (1)	1000	1.27	0.067	8.5	-34.67	5.51
SPC (1)	1000	2.01	0.065	13.2	3.33	2.00
Conway (1)	1000	2.02	0.063	12.1	3.70	-0.75
MSER-5 (2)	1000	1.41	0.062	5	-27.80	-2.99
Randomization (2)	1000	2.01	0.064	10.6	3.15	0.41
Crossing the means (2)	1000	1.48	0.065	7	-23.78	1.93
Initial (before truncation)	1000	1.95	0.064	5	0.00	0.00
Model Type III						
Welch (1)	1000	0.070	0.021	11	-7.59	8.52
SPC (1)	1000	0.062	0.026	31	-36.25	-4.88
Conway (1)	1000	0.060	0.019	15	0.64	-7.78
MSER-5 (2)	1000	0.061	0.019	9	-0.02	-5.64
Randomization (2)	1000	0.060	0.019	750	-0.24	-7.21
Crossing the means (2)	1000	0.069	0.020	25	2.18	6.04
Initial (before truncation)	1000	0.065	0.019	8	0.00	0.00

(1) These methods need 5 replications, hence the warm-up lengths are same for all 5 runs
(2) These methods need single replication. They are applied 5 times to 5 runs.
(3) The Modified SPC Method does not work for run length of 1000. It demands an increase in run length.
(4) The Randomization Test does work for run length of 1000. It demands an increase in run length.

Table 3: Results for Run Length = 5200 Hours

Method	Average Run length (hours)	Final MSE	Final Variance	Average Computing Time (seconds)	% Change in MSE	%Change in variance
Model Type I						
Welch (1)	5200	70.6416	0.2982	33	-2.32	-0.05
SPC (1)	5200	53.7927	0.3349	33	-83.22	-79.79
Conway (1)	5200	72.3384	0.2977	54	0.03	-0.24
MSER-5 (2)	5200	51.6606	0.2907	16	-28.57	-2.59
Randomization (2)	5200	44.0937	0.2997	31	-39.03	0.44
Crossing the means (2)	5200	67.1267	0.2993	39	-7.18	0.31
Initial (before truncation)	5200	72.3188	0.2984	15	0.00	0.00
Model Type II						
Welch (1)	5200	0.0954	0.0112	18	4.83	0.26
SPC (1)	5200	0.1115	0.0116	39	22.52	3.72
Conway (1)	5200	0.0942	0.0111	21	3.48	-0.17
MSER-5 (2)	5200	0.0798	0.0111	15	-12.35	-0.74
Randomization (2)	5200	0.0957	0.0111	145	5.19	0.01
Crossing the means (2)	5200	0.0906	0.0111	29	-0.43	-0.07
Initial (before truncation)	5200	0.0910	0.0111	15	0.00	0.00
Model Type III						
Welch (1)	5200	0.0175	0.00394	19	6.01	1.25
SPC (1)	5200	0.0250	0.00979	32	51.45	151.47
Conway (1)	5200	0.0165	0.00389	16	-0.31	-0.13
MSER-5 (2)	5200	0.0167	0.00390	15	0.78	0.01
Randomization (2)	5200	0.0165	0.00389	1795	-0.05	-0.09
Crossing the means (2)	5200	0.0166	0.00391	27	0.60	0.48
Initial (before truncation)	5200	0.0165	0.00389	14	0.00	0.00

(1) These methods need 5 replications, hence the warm-up lengths are same for all 5 runs
(2) These methods need single replication. They are applied 5 times to 5 runs.

takes longest time to perform. For Type III model, the initial MSE and variance are very low. There is less scope for reduction in MSE and variance. A method is said to perform well if it does not increase the MSE and variance by more than 1% or it reduces both the MSE and variance. We see that except for the Welch's Method and the SPC Method, all other methods perform well for this type of model. However, the Randomization Test needs extremely long time to perform. This is due to a large number of iterations.

5 CONCLUSION

The evaluation procedure presented in this paper is easy to implement on various methods and can be used to test the performance of the same. The measures of goodness used clearly indicate the quality of the methods.

From the results we conclude that there is no method which works well for all types of models. Some methods work well for low utilized systems, where as some work for longer run lengths as opposed to smaller run lengths. The Randomization Test works well for highly utilized system if additional data that it requires can be made available. Same applies to SPC Method. For Type III models, the Randomization Test takes a lot of time due to a large number of iterations. In such situations partial Randomization Tests are recommended. Results with partial randomization test show that the performance with not only the time but also with MSE and variance is improved. Following modifications can improve the performance of Randomization Test:

- Minimum number of batches should be 20.
- Approximate randomization tests are recommended for low utilization models where the expected run time is very high.

The SPC Method works well for Type I systems. But for medium and low utilized systems, it tends to give a very long warm-up length and removes vital observations. The performance is worse for low utilized systems with longer run lengths. The Welch's Method didn't work well for most of the experiments. It is highly subjective. Different results may be observed for different set of users. The Crossing the Means Rule seems to work well for very long run lengths on Type II and Type III models. The Conway Rule is very aggressive. It works well only for low utilized systems. The MSER-5 Rule works well for longer run lengths. For shorter run lengths it works well for low and moderately utilized models. This rule is highly efficient. Table 4 presents the recommended methods depending on the system used.

Table 4: Recommended Methods for Different Types of Systems

Highly utilized system	
Name of the method	**Recommendation**
SPC Method	Use long run length
Randomization Test	Only if more data can be generated
MSER-5 Rule	Use long run length
Moderately utilized system	
MSER-5 Rule	Use long run length
Low utilized system	
Randomization Test	None
Conway Rule	None
MSER-5 Rule	Use long run length

6 FUTURE RESEARCH

For this research a particular system performance measure was chosen (number in system measured at the end of every hour). The results may vary for different measures of performance. Also, the complexity of the system was kept constant. The methods need to be tested by varying the system performance measures and the complexity of the system. Test for Initialization Bias were not included in this research. Future research may deal with developing deletion strategies for the Initialization Bias Tests and assessing their performance.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Jerry Banks, Dr. David Goldsman, Dr. Barry Nelson, Dr. Bruce Schmeiser and Dr. Lee Schruben for their vital and helpful discussions on this research.

REFERENCES

Cash, C.R. Nelson, B. L., Long, J. M., Dippold, D. G., and Pollard W.P., 1992. *"Evaluation of Tests for initial-condition bias"* Proceedings of the 1992 Winter simulation conference 1992, pp 577-585

Conway R.W. 1963. *" Some Tactical Problems in Digital Simulation"* Management Science, Volume 10, Issue 1, (Oct., 1963), 47-61

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

Gafarian, A.V., C.J Ancker, and T. Morisaku. 1978. *"Evaluation of commonly used rules for detecting steady-state in computer simulation"*. Naval Research Logistics Quarterly 25: 511-529.

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

Mahajan, P.S. 2004, *" Analysis of methods to detect warm-up length in Steady State Simulation"* Master's Thesis, Oklahoma State University, School of Industrial Engineering and Management

Robinson Stewart 2002 *"A Statistical Process Control Approach For Estimating the Warm-up Period"* Proceedings of the 2002 Winter Simulation Conference E Yücesan, C.-H. Chen, J. L. Snowdon, and J.M.Charnes, Eds.

Robinson Stewart 2004 *"A Statistical Process Control Approach to select a warm-up period for discrete event simulation"* Unpublished paper currently in review

Spratt, S. C. 1998. *"Heuristic for startup problem"* M.S. thesis, Department of Systems Engineering, University of Virginia

Welch, P. 1983. *"The statistical analysis of simulation results."* In The computer modeling handbook, ed. S. Lavenberg, 268-328. New York: Academic Press.

White K. Preston Jr. 1997 *"An effective Truncation Heuristic for Bias Reduction in* Simulation Output". Simulation 69 6 323 –334

White K. P. , M.J. Cobb, and S. C. Spratt (2000) "*A comparison of five steady state truncation heuristics for simulation"* Proceedings of 2000 winter simulation conference J.A Joines R. P. Barton, K.Kang and P.A. Fishwick, eds.

Wilson J.R. and A.B. Pritsker. 1978. *"A survey of research on the simulation startup problem"*. Simulation 31: 55-59

Wilson J.R. and A.B. Pritsker. *"Evaluation of startup Policies in Simulation Experiments"* Simulation. Volume 31 no. 3 September 1978

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

Arena 5 Simulation Software Developed by Rockwell Software Inc.

RAQS is a Queuing Software developed by the Center for Computer Integrated Manufacturing at Oklahoma State University Stillwater

AUTHOR BIOGRAPHIES

PRASAD S. MAHAJAN is a Master's student in the School of Industrial Engineering and Management at Oklahoma State University. He has a B.S. in Mechanical Engineering from University of Pune, Pune, India. He is a member of INFORMS and APICS. He is presently working with Transplace as a Logistics Engineer. His email address is <mahajanprasad@yahoo.com>.

RICKI G. INGALLS is Associate Professor and Director of the Center for Engineering Logistics and Distribution (CELDi) in the School of Industrial Engineering and Management at Oklahoma State University. He is also the Co-Editor of the *Proceedings of the 2004 Winter Simulation Conference*. He joined OSU in 2000 after 16 years in industry with Compaq, SEMATECH, General Electric, and Motorola. He has a B.S. in Mathematics from East Texas Baptist College (1982), a M.S. in Industrial Engineering from Texas A&M University (1984) and a Ph.D. in Management Science from the University of Texas at Austin (1999). His research interests include the supply chain design issues and the development and application of qualitative discrete-event simulation. He is a member of IIE. His e-mail address is <ingalls@okstate.edu>.