Detection and Significance of the Initial Transient Period in Quantitative Steady-State Simulation

D. McNickle, K. Pawlikowski and C. Stacey *

University of Canterbury,
Christchurch, New Zealand

Abstract

Numerous techniques have been proposed for detecting the length of the initial transient in steady-state simulations of queueing-type processes, for example when studying performance of telecommunication networks and protocols. Unfortunately, these techniques behave quite badly when applied in practical simulation experiments, especially when simulating heavily loaded systems. We assess the performance of two of the most successful methods for estimating the length of the initial transient period, based on statistical tests developed for testing stationarity of time series proposed by Schruben et al [16, 15] and Yücesan [18]. The performance of these methods is compared with the theoretical relaxation time in a class of M/E_k/1 and M/H_k/1 queueing systems. We also study the influence of data collected during the initial transient phase on the quality of the final simulation results. Two methods of simulation output data are considered: the method of Spectral Analysis, as proposed by Heidelberger and Welch [6], and a sequential version of Independent Replications.

Introduction

In stochastic discrete-event simulation executed for studying steady-state performance of a system, the initial data in the primary data stream are not representative of the steady-state behaviour of the system because of bias introduced during the *initial transient* or *warm-up phase* characterising the initial system's behaviour. The initial transient phase is simply the period during which the system settles into its *steady-state* behaviour. The most commonly used technique to remove the "bias of initialisation" introduced into the output data by the initial transient phase is referred to as *data truncation*. The biased data is simply discarded and the remaining unbiased data analysed. In the method of Independent Replications every one of the multiple replications within a simulation experiment contains bias introduced by its initial transient phase. There are problems with the coverage performance of single replication methods (Batch Means, Spectral Analysis, etc). As well as this, recent results suggest that the difference in the theoretical efficiency between the methods based on single and multiple replication is not as great as previously thought [17]. Therefore we decided to seek a way to apply the method of Independent Replications in a fully automatic procedure for output data analysis in steady-state simulation. The reported research project investigated using the length of the initial transient period to determine the length of each replication in an Independent Replications simulation experiment. This makes accurate detection of the end of the initial transient phase even more important than is normally the case. Normally, it is sufficient to remove enough data such that the bias introduced by the initial transient phase is removed, so a conservative approach that removes more than enough is acceptable. However, when the intent is to implement the method of Independent Replications sequentially, automatically stopping the simulation when the precision of estimates reaches the required precision, or when one tries to make use of data collected during the initial transient to determine the replication length, it becomes important to determine the length of the initial transient phase as accurately as possible.

A number of different methods to assess the length of the transient period in a steady-state simulation has been proposed. For analytically tractable systems it may be possible to calculate an estimate of the expected length of the initial transient phase. However, for most systems of interest this is not possible. One possibility is to find an upper bound of the length of the initial transient phase for a class of systems. This method is not very reliable as different systems and even the same system with a different sequence of random numbers may vary widely in how long they remain in a transient state.

Rather than simply deciding a priori how long the transient period of a simulation will be we may use a heuristic based on the observed simulation output data to estimate the length of the initial transient phase. Such heuristics are surveyed in [11]. Our experience has shown that no heuristic can be considered relatively robust and those that we have studied have frequently produced inaccurate results. In quantitative stochastic simulation more powerful techniques should be applied. Special stationarity tests, used to test if a simulated process has reached steady-state, may be applied to test whether the data contains significant bias due to the initial transient phase. Two methods that apply such tests in an iterative manner to obtain estimates of the length of the initial transient phase are detailed in Section 2.

1 Analytical Measures of the Initial Transient Phase

Once we have some method or procedure that is designed to estimate the length of the initial transient phase we need to assess its accuracy in some way. As noted above there exist analytical methods by which we can calculate an estimate of the length of the initial transient phase for some analytically tractable systems. Two such measures have been studied: a relaxation time, with a formula which is valid for single server queueing systems (GI/G/1), see Section 1.1, and the exact expected waiting time of the nth customer, which can be calculated exactly for M/E$_k$/1 systems using Heathcote and Winer's formula [5] and following [9], see Section 1.2.

1.1 Relaxation Time

It has been shown that the rate at which mean queue lengths or mean delays tend to their steady-state is, after some period of time, dominated by a term of the form $\exp(-t/\tau_r)$, where τ_r is called the *relaxation time* of the queue [11].

It has been postulated that the initial transient phase is over after the time $t_\beta = -\tau_r \ln \beta$, where β is the permissible relative residue of the initial state, $0 < \beta < 1$. Assuming $\beta \leq 0.02$, at $t = 4\tau_r$ we find that the queue characteristics, such as mean time spent in the queue for a customer, are within 2% of their steady-state values. In other words, output data collected from that point of time should be biased by the initial state by less than 2% [11].

Odoni and Roth [10] obtained an approximation to the relaxation time of the mean queue length in Markovian queueing systems based on the coefficients of variation of the interarrival and service times, τ_r. As noted, the time taken for a system to fall within 2% of steady-state is $4\tau_r$ time units. Jackway and de Silva [7], who presented a slightly modified version of the Schruben test, convert the relaxation time estimate to the number of observed service completions needed to reach steady-state. Their approximation involves dividing the required relaxation time, $4\tau_r$, by $1/\mu$ (the mean service time). For the $M/M/1$ queue $C_A^2 = C_S^2 = 1$ and cancelling μ reduces to a formula for the length of the initial transient phase, τ_n, in $M/M/1$ queueing systems. It should be noted that this approximation assumes that the server is busy continuously.

$$\tau_n = \frac{8}{2.8(1-\sqrt{\rho})^2}, \tag{1}$$

where ρ is the system load and τ_n is the number of service completions.

It is noted by Odoni and Roth [10] that this relaxation time estimate of the length of the initial transient phase is a *gross lower bound*. This stipulation is reinforced by Anderson [1] who suggests that the time for a real system to approach steady-state is larger than the relaxation time. Anderson claimed that theoretically systems converge to their steady-state at the rate $\varepsilon(e^{-\lambda T})$, while empirical convergence to steady-state of the first kind (mean value) occurs at the rate $\varepsilon(\frac{1}{T})$, and empirical convergence to steady-state of the second kind (variance) occurs at the rate $\varepsilon(\frac{1}{\sqrt{T}})$ [1].

The translation from relaxation in time to relaxation in observed service completions is an approximation at best since it assumes that the server will always be busy. The greater the loading or the utilisation of the system the more likely this assumption is to hold. The inaccuracy introduced by this approximation is acceptable as it will produce slightly longer relaxation periods in terms of simulated time, erring on the side of safety. This is because of the assumption that the server will always be busy.

1.2 Expected Waiting Time

A possible alternative to the relaxation time as an analytic measure of the length of the initial transient length could involve the number of observations required for some parameter of the system (at a given loading) to become sufficiently close to its steady-state mean value. This estimate could be obtained by studying, for example, the *expected average delay of the first n customers*. The mean waiting time of the nth customer in an M/M/1 system when started from an empty-and-idle state is given as:

$$EW_n = \sum_{j=1}^{n} j^{-1} ES_j^+, \tag{2}$$

where S_j is the jth partial sum of random variables independently and identically distributed as the difference between a service time and an interarrival time, and

$$\begin{aligned} ES_n^+ &= \int_0^\infty y dP(S_n \leq y) \quad (3) \\ &= \frac{1}{\mu}\left(\frac{\rho}{1+\rho}\right)^n \frac{1}{(n-1)!}\sum_{j=0}^{n-1}\left(\frac{1}{1+\rho}\right)^j \frac{n+j-1)!}{j!}(n-j), (n>1) \end{aligned}$$

see Heathcote and Winer [5]. In particular

$$EW_1 = \frac{\rho}{\mu(1+\rho)} \tag{4}$$

$$EW_\infty = \frac{\rho}{\mu(1-\rho)} \tag{5}$$

Similar formulae are given in [5] for the M/D/1 and D/M/1 queueing systems.

With these formulae we are able to calculate the point at which the actual expected value of a parameter comes within a given tolerance of the steady-state value for the first time. This point may then be used as an estimate of the length of the initial transient phase for that parameter. Figure 1 shows

a figure goes here

Figure 1: Expected waiting time of the nth customer in an M/M/1 queue; $\rho = 0.9$.

a figure goes here

Figure 2: Comparison of estimators of the length of the initial transient phase for the M/M/1 queue.

how the expected waiting time of the nth customer tends towards the steady-state value, EW_∞, for an M/M/1 queue at a loading of 0.9.

Figure 2 shows a comparison between the relaxation time estimate (measured by the number of service completions) and various versions of the expected waiting time estimate. Each curve for the expected waiting time estimate is determined by assuming different values of relative "closeness" of EW_n to EW_∞ for finding

$$n_0 = \min_n \{n : \frac{EW_\infty - EW_n}{EW_\infty} \leq \varepsilon\} \tag{6}$$

for $\varepsilon = 0.01, 0.05, 0.1$. From this graph it may be seen that when the expected waiting time is required to be with 0.5% of the steady-state value it is very close to the curve given by the relaxation time which specifies a greater tolerance of 2%. It may also be seen that as the tolerance is lowered to 0.1% the curve moves higher than the relaxation time. This shows that if we want to estimate the length of the initial transient phase more safely we should require the initial transient detector to produce values greater than that provided by the relaxation estimate.

Unfortunately, although the convergence of the expected waiting time to steady-state gives an intuitively more understandable basis for estimating the length of the initial transient phase it is analytically tractable for M/E$_k$/1 systems only. The relaxation time estimate is known for any single server queueing system of GI/G/1 type. Because of this, for the rest of this study the relaxation time estimate, translated into the number of service completions, is used as the reference when evaluating the quality of various initial transient phase detectors.

2 Statistical Tests

As mentioned, the length of the initial transient phase has traditionally been determined using different heuristic rules. For the purposes of this study the heuristic labelled R5 in [11] is used:

> *the initial transient period is over after n_0 observations $x_1, x_2, x_3, \ldots, x_{n_0}$ crosses the mean $\bar{X}(n_0)$ k times,* where $\bar{X}(n_0) = \frac{1}{n_0}\sum_1^{n_0} x_i$.

This rule is sensitive to the value of k, the number of crossings of the mean required. Too large a value will usually lead to an overestimated value of n_0, regardless of system's utilisation, while too small value of k can result in an underestimated value of n_0 in more heavily loaded systems. Results of previous studies [11] have supported the selection of $k = 25$, recommended in [4].

More precise measures of the length of the initial transient could be obtained by using statistical tests invented to test the stationarity of data sequences. A few such techniques for determining the length of the initial transient phase have been developed, of which two are studied in the following

Sections. Each operates in a hypothesis testing framework, formally testing the null hypothesis that *there is no initialisation bias in the output mean* against the alternate hypothesis that initialisation bias in the output exists.

2.1 Schruben Test

The first stationarity test was proposed by Schruben in [14] and improved on by Schruben et al. in [16]. It is used to test the hypothesis that a sufficient number of initial transient data have been (or have not been) discarded based on a standardised time series. The statistic,

$$T = \frac{45}{n_t^{1.5} n_v^{0.5} \hat{\sigma}[\bar{X}(n_v)]} \sum_{k=1} n_t k (1 - \frac{k}{n_t})[\bar{X}(n_t) - \bar{X}(k)] \quad (7)$$

and Var(T) are calculated from the most recent observations collected and compared with the corresponding value from the Student-t distribution. If the sequence of tested data cannot be considered as stationary, it is discarded and the next sequence of observations tested. This process is repeated until the test determines that the system has reached steady-state or some predefined upper limit on the simulation length is reached.

a figure goes here

Figure 3: Sequential analysis of the length of the initial transient phase.

The Schruben test requires foreknowledge of the steady-state variance of the system which is not normally available when the test is applied because the system is still in its initial transient phase. The Schruben test solves this problem by estimating the steady-state variance over the latter portion of the collected data. This is done on the assumption that this latter portion of data is more representative of the steady-state behaviour of the system, thus giving a better estimate of the steady-state variance. The data used for calculating the test statistic and an estimate of the steady-state variance are contained in the test and variance windows, respectively. The relative sizes and positions of the test and variance windows used by the method are variable and affect the performance of the test. The effectiveness of the test is strongly dependent on the effectiveness of the variance estimator used [11, 16].

In the original version of the Schruben test it is assumed that the data that is tested for stationarity is the most recently collected data. This assumption creates the problem mentioned above where an estimate of the steady-state variance of the system is required whilst still in the transient phase. As has been observed, the performance of the Schruben test is highly dependent on the variance estimator used [11, 16], therefore to improve performance of the test as a whole the accuracy of the variance estimation process should be improved.

With this idea in mind a modification proposed is to relax the constraint that the test is unable to look into the "future". It is noted that relaxation of this constraint introduces the need for the simulation system to be able to store and reuse the data that is collected while determining the length of the initial transient phase, but not discarded as part of the initial transient phase. The need to store all data increases the memory resources required by the procedure when long initial transient phases are encountered. This new version of the Schruben test uses two separate windows. The leading window, called the variance window, contains data from which the steady-state variance of the means is calculated. The trailing window, called the test window, contains the data over which the Schruben test is performed.

a figure goes here

Figure 4: Relative size and position of the test and variance windows in the original version of the Schruben test.

a figure goes here

Figure 5: Relative size and position of the test and variance windows in the modified version of the Schruben test.

Upon experimentation it was found that increasing the distance into the future that the test was able to look resulted in no significant increase in performance. This appeared to be due to the small size of the windows of data being tested, usually equivalent to 200 observations. An initial sensitivity analysis of the Schruben test to determine the significance of the various parameters on the performance of the test showed that the test as originally specified in [11] was most sensitive to the size of the test and variance windows. Although increasing the window sizes increased the power of the test it also decreased the accuracy of estimates, since the amount of data discarded each time the hypothesis is not satisfied is proportional to the window size. This means that in a lightly loaded system an excessive amount of data could be discarded if the window sizes are optimised for heavily loaded systems. The results of these findings were incorporated into our modification of the Schruben test, by making the windows sizes proportional to the system load.

It was also noted during experimentation that the Spectral Analysis method used to estimate the variance of collected data as previously implemented in SAM [13] and AKAROA [12] assumed that only 200 data points would be used. In this application the size of the variance window is variable. So, Fishman's autoregressive method [3] was used to estimate the variance of the observations within the variance window, as also used by Schruben in [14] and Jackway and de Silva [7] in their version of the Schruben test.

Simulation results show that using a window size that is 3000 times the system loading, i.e., 3000ρ, and placing the variance window ten times the window size into the "future" gave the best results, as shown in Section 3. This version of the Schruben test is referred to as version 2, while the version as proposed in [11] is referred to as version 1.

The version of the Schruben test procedure described in [11] included a mechanism that was intended to allow the test window size to increase as the variance present in the system increased. The mechanism set the size of the test window on the basis of the maximum of a default value and a portion of the length given as a first estimate by the heuristic used. Unfortunately, this mechanism turned out not to be particularly effective since the heuristic is insensitive to the increased variance present in the system at high loadings, as we describe in Section 3.

2.2 Yücesan Test

The test proposed by Yücesan [18] for detecting the length of the initial transient phase is a randomisation test. The only characteristic of the simulated system that is of interest to the test is the commonality of batch means within the data. The requirements enforced by the assumption that a test statistic comes from a standard distribution are deemed unnecessarily restricting. This assumption may cause the null hypothesis to be rejected on the grounds that one of the "extra" requirements does not hold while the characteristic of interest does in fact satisfy the hypothesis. A randomisation procedure is used to avoid the introduction of unnecessary constraints, such as those present in the Schruben test.

The aim of randomisation tests is to test the characteristic of interest, in this case the means of batches of data, without introducing any other constraints to the test procedure. The randomisation test is used to approximate the distribution of the test statistic by *shuffling* or permuting the test data. The significance of the actual test statistic for the unshuffled data is then assessed relative to the empirically generated distribution obtained through shuffling rather than from a specified distribution.

The test statistic used by Yücesan is based on the means of batches within the data, rather than a sequence of partial sums in a standardised time series as in the Schruben test. Once the observations have been collected into b batches of size m the means of these batches are used as secondary data points.

$$X_i = \sum_{m*(i-1)+1}^{m*i} x_j \tag{8}$$

where X_i is the mean of the ith batch and x_j is the jth primary data point. Using the batch means rather than the primary data stream in the test makes the Yücesan procedure computationally much faster than methods based on the Schruben tests which have to recalculate the test statistic from the primary data stream every time. Once the batch means have been calculated they are split into two groups, G_1 and G_2. The first group initially contains the first batch mean and the second group the other $b-1$ batch means. The means within each group are then averaged to create *group* means,

$$X_{G_k} = \sum X_i \tag{9}$$

where X_i is in group G_k. These two group means are then compared and the absolute difference forms the *actual* test statistic,

$$T = |X_{G_1} - X_{G_2}|. \tag{10}$$

The order of all b means (X_i) is then randomised and split in the same way to compute a *pseudo* statistic, T'. This pseudo statistic is calculated for N_S different randomised orderings of the batches. The calculated significance level, sl, of the test is calculated from the number of times the pseudo statistic, T', is greater than or equal to the actual statistic, T.

$$sl = \frac{nge + 1}{N_S + 1} \tag{11}$$

where nge is a count of the number of times T' exceeds T. The addition of one protects against division by zero and nil results. If the calculated significance level, sl, has reached the desired value the procedure is finished and the initial transient phase is said to be finished at the end of the first group of unshuffled batches. If the significance has not been reached then the first batch mean from the original ordering of the second group is transferred to the first group (so the first group contains 2 means while the second contains $b-2$ means) and the whole process repeated.

This is summarised as follows:

Step 1 Collect b sufficiently uncorrelated batches and calculate b batch means.

Step 2 Split batch means into two groups, G_1 and G_2.

Step 3 Calculate actual test statistic, T, following (10

Step 4 Randomise the order of the batch means and calculate the pseudo-statistic, T'.

Step 5 Increase nge if $T' \geq T$.

Step 6 Repeat steps 4 and 5 N_S times.

Step 7 Calculate significance level of the test, sl.

If the desired significance level has not been reached swap one batch from G_2 to G_1 and repeat from Step 3. If there is only one batch in G_2, either the user can be warned and a cut-off made at the end of the collected data or more data can be collected and the whole process repeated. If the desired significance has been reached then the truncation point is said to be at the end of the batches in G_1 in their original order.

a figure goes here

Figure 6: Data batches as used in the Yücesan test.

In the original version of the Yücesan test [18], when the batch size is selected a check is made to ensure that the batches are sufficiently large that any significant serial correlation between the batch means has been removed. If the serial correlation is too high (ie., 0.5 or more) then the batch size is increased and the batch means recalculated. The original version of the Yücesan test used $b = 30$ batches with an initial batch size $m = 500$ observations, which is doubled each time the correlation between the batch means is too high. This technique was found to lead to batches in the order of 8000 observations when applied in heavily loaded systems. Having such large batch sizes made the test very imprecise in its estimate of the length of the initial transient phase which is undesirable in the given task domain. As an alternative to this, a modified implementation of the Yücesan test that we investigated, uses 50 batches of 100 observations, increasing each batch by 100 observations when inter-batch mean correlation is too high. These batch sizes and numbers were used based upon the performance of automated versions of the batch means method [11].

Simulation results showed that the Yücesan test, described in [18], did not always give particularly “safe” estimates of the length of the initial transient phase. However, as may be seen in Figure 7, it performed better than either version of the Schruben test. As discussed earlier, a detection method should ideally give estimates above the relaxation estimate. Yücesan [18] mentioned that the significance of the test statistic should increase monotonically with the number of batches contained in the first group in the case where initialisation bias is present. Following this, if we wish to increase the number of batches placed into the first group before the test’s target significance is reached we simply increase the target significance. Results showed that selecting a target significance of 50% gave much safer results. In using a hypothesis testing framework to determine the length of the initial transient phase we wish to be very certain of the acceptance of the hypothesis, hence the relatively high level of significance required.

In Figure 7 the original version of the Yücesan test is referred to as Version 1, while the modified test is referred to as Version 2.

3 Comparison of Initial Transient Detectors

The performance of each of the detectors of the length of the initial transient phase was evaluated in a variety of queueing systems over a range of system loads. The results obtained for the M/M/1, M/D/1, M/E_2/1 and M/H_k/1 queueing systems are shown in Figures 7, 8, 9 and 10 respectively, each point is

a figure goes here

Figure 7: Performance of initial transient detectors for the M/M/1 queue. Coefficient of variation, $C = 1$.

a figure goes here

Figure 8: Performance of initial transient detectors for the M/D/1 queue. Coefficient of variation, $C = 0$.

the average of 500 experiments. To obtain safer results the initial transient detector should produce results above the relaxation time (shown as a continuous line). The results from using the heuristic (25 crossings) alone, without applying a stationarity test, are represented with a line with diamonds on it. As may be seen, the modification of the Schruben test (Version 2) offers better performance than the original (Version 1) but does not prevent it from "saturating" in a high load situation. On the other hand, both the original version of the Yücesan test (Version 1) and its modification (Version 2) follow the trend given by the relaxation estimate. Version 2 increases the variance of the results, but generally increases their "safety". The figures show that in all queueing systems studied the relative ordering in performance of the tests was the same, regardless of the system under study, and that none of the tests were able to avoid some degree of saturation at high loads in a system with a high degree of variation, (eg., the M/H_k/1 system).

4 Significance of the initial transient phase

Traditionally in steady-state simulation, data collected during the initial transient phase of a simulation experiment has been discarded in an effort to eliminate the bias that it introduces. This practice, called data-truncation, designed to reduce bias, may increase the mean squared error of the estimate or the variance of the data [9]. The question arises as to the significance of the bias introduced by the initial transient phase, and if it warrants the removal of the transient phase data. Table 1 shows the mean total simulation lengths of experiments in which the total time spent by a customer in an M/M/1 queueing system was estimated, with and without deletion of the initial transient phase, using Independent Replications. Also shown is the relative participation of the initial transient period in the total length of these simulations. The length of the initial transient phase is determined using the theoretical value given by the relaxation time estimate, in Section 1.1. The estimates are calculated over 500 experiments, with each replication using a sample of 5000 observations. Results of a similar investigation were reported in [8].

Table 1 suggests that the participation of the initial transient phase in the total simulation is considerable, especially in heavily loaded systems. This is confirmed in Figure 11 which shows the effect of the initial transient period on the quality of final results, measured by the coverage of final

a figure goes here

Figure 9: Performance of initial transient detectors for the M/E_2/1 queue. Coefficient of variation, $C = 0.5$.

a figure goes here

Figure 10: Performance of initial transient detectors for the M/H$_k$/1 queue. Coefficient of variation, $C = 10$.

	With deletion		Without deletion		Relative length
ρ	Replication	Total	Replication	Total	of transient (%)
0.2	5009	44510	5000	44980	0.2
0.4	5021	48071	5000	47990	0.4
0.6	5056	65970	5000	64770	1.1
0.8	5256	192843	5000	180940	4.9
0.9	6084	763670	5000	612505	17.8

Table 1: Mean replication lengths and total simulation length when analysing the time spent in the system by a customer in an M/M/1 queueing system, with and without deletion of the initial transient phase, when using Independent Replications. Also shown is the relative length of the initial transient period. Estimates are over 500 experiments, with each replication using a sample of 5000 observations.

results obtained using a traditional implementation of Independent Replications, where the length of each replication is set by the experimenter before simulation begins. Estimates are over 500 experiments, with each replication using a sample of 5000 observations to estimate the final result and the number of replications increased until an estimate of least 5% relative precision was obtained.

In each case the length of the initial transient phase and thus the relative length of the initial transient phase with respect to the total simulation increases as the system load, ρ, increases. As the fixed length of replications used increases the stability and quality of results produced by the method increases. This trend is matched by a corresponding increase in the total simulation length as the replication length increases.

These results are compared with similar results for the single replication Spectral Analysis method proposed by Heidelberger and Welch [6]. This method is based on analysis of the spectral density function of the simulation output data and in this experiment estimates were required to be of at least 5% relative precision.

Table 2 suggests that the influence of the initial transient phase on the method of Spectral Analysis is very small. Figure 12 shows the effect of the initial transient period on the quality of final results as measured by the coverage of results when using Spectral Analysis. The results show that although the initial transient phase may appear to have a very small participation in the output from a simulation experiment it has a significant effect on the quality of the steady-state results produced by the method and thus it should be removed.

a figure goes here

Figure 11: Coverage analysis of an M/M/1 queue using Independent Replications in its traditional version, using replications of 5000 observations.

ρ	Mean Total Simulation Length		Relative length of transient (%)
	With deletion	Without deletion	
0.2	6629	5074	4.4
0.4	13465	11567	4.5
0.6	34639	32547	4.6
0.8	172045	166436	0.5
0.9	700449	677995	0.25

Table 2: Mean total simulation length when analysing the time spent by a customer in an M/M/1 queueing system, with and without deletion of the initial transient phase, when using Spectral Analysis. Also shown is the relative length of the initial transient period.

a figure goes here

Figure 12: Coverage analysis of an M/M/1 queue using Spectral Analysis. Estimates are over 500 experiments.

5 Conclusions

The aim of this research was to produce a fully automated simulation output analysis method, that would be statistically valid, and able to be used by the novice user, based on independent replications. With this task in mind a number of problems which have traditionally stood in the way of developing such a method were investigated. These problems include accurate and statistically sound detection of the length of the initial transient phase within a simulation experiment and automated selection of the replication length for Independent Replications.

The investigation of the initial transient phase detection problem included a comparative study of two of candidate methods. The methods originally proposed by Schruben and Yücesan were implemented and their performance investigated through the use of simulation studies in a variety of queueing systems. From this investigation an initial transient detector based on the Yücesan method was developed that performs well in a wide variety of systems in a completely automated and system independent manner.

References

[1] ANDERSON, K. Some Steady-state Problems in Simulation. In *The 18th Simulation Symposium* (Mar. 1985), A. Miller, Ed., pp. 79–93.

[2] CASH, C. R., .NELSON, B. L., DIPPOLD, D. G., LONG, J. M., AND POLLARD, W. P. Evaluation of Tests for Initial-Condition Bias. *Proc. of the Winter Sim. Conference* (1992), 577–585.

[3] FISHMAN, G. S. *Principles of Discrete Event Simulation*. John Wiley & Sons, 1978.

[4] GAFARIAN, A. V. Evaluation of Commonly Used Rules for Detecting "Steady State" in Computer Simulation. *Naval Res. Logistics Quarterly* (1978), 511–529.

[5] HEATHCOTE, C. R., AND WINER, P. An Approximation for the Moments of Waiting Times. *Operations Res., 17* (Jan. 1969), 175–186.

[6] HEIDELBERGER, P., AND WELCH, P. D. A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations. *Comunications of the ACM, 24*, 4 (Apr. 1981), 233–245.

[7] JACKWAY, P. T., AND DESILVA, B. M. A Methodology for Initialisation Bias Reduction in Computer Simulation Output. *Asia-Pacific J. of Operational Res., 9* (1992), 87–100.

[8] KELTON, W. D., AND LAW, A. M. An Analytical Evaluation of Alternative Stragies in Steady-State Simulation. *Operations Res., 32*, 1 (Jan. 1984), 169–184.

[9] LAW, A. M. Statistical Analysis of Simulation Output Data. *Operations Res., 31*, 6 (Nov. 1983), 983–1029.

[10] ODONI, A. R., AND ROTH, E. An Empirical Investigation of the Transient Behaviour of Stationary Queuing Systems. *Operations Res., 31* (1983), 432–455.

[11] PAWLIKOWSKI, K. Steady State Simulation of Queuing Processes: Survey of Problems and Solutions. *ACM Computing Surveys* (June 1990), 123–170.

[12] PAWLIKOWSKI, K., MCNICKLE, D., AND YAU, V. Final Report on R&D Investigations of Experimental Control for Simulations, for AOTC Research Laboratories, Melbourne. Tech. Rep. COSC 02/93, Department of Computer Science, University of Canterbury, Feb. 1992.

[13] PAWLIKOWSKI, K., AND YAU, V. Methodology of Stochastic Simulation for Performance Evaluation of Data Communication Networks. Tech. Rep. COSC 03/93, Department of Computer Science, University of Canterbury, March 1993.

[14] SCHRUBEN, L. Detecting Initialization Bias in Simulation Output. *Operations Res., 30*, 2 (May 1982), 569–590.

[15] SCHRUBEN, L. Confidence Interval Estimation Using Standardised Time Series. *Operations Res., 31*, 6 (Nov. 1983), 1090–1145.

[16] SCHRUBEN, L., SINGH, H., AND TIERNEY, L. Optimal Tests for Initialization Bias in Simulation Output. *Operations Res., 31*, 6 (Nov. 1983), 1167–1178.

[17] WHITT, W. The Efficiency of One Long Run Versus Independent Replications in Steady-State Simulation. *Management Sci., 37*, 6 (June 1991), 645–665.

[18] YÜCESAN, E. Randomisation Tests for Initialisation Bias in Simulation Output. *Naval Res. Logistics, 40* (1993), 643–663.