TRANSIENT DELETION AND THE QUALITY OF SEQUENTIAL STEADY-STATE SIMULATION

Don McNickle
Department of Management
University of Canterbury
Christchurch, New Zealand
Email:Don.McNickle@canterbury.ac.nz

Gregory C. Ewing and Krzysztof Pawlikowski
Computer Science and Software Engineering
University of Canterbury
Christchurch, New Zealand
Email: Krys.Pawlikowski@canterbury.ac.nz

ABSTRACT

In discrete event steady-state simulation, deletion of data from the initial transient phase of the simulation is usually recommended in order to reduce the bias of the final estimates. Various heuristics and tests have been proposed to aid with this. The plummeting cost of simulation, combined with uncertainties about the overall reliability of the estimation of the transient period, suggest revisiting the notion that deletion is essential, especially for longer simulations. We consider this in a sequential simulation framework.

1. INTRODUCTION

A standard part of simulation methodology for discrete event steady-state simulation is that data from the initial transient phase of simulation should be deleted in order to reduce the bias of the final estimates. The assumption usually underlying this is that the expected value of the process being simulated may be changing over the transient phase, and thus including these data biases the results or increases the variance. Various heuristic methods for selecting the number of observations to delete, and for testing if the system is adequately close to “steady state” have been proposed. A survey of these up to 1990 can be found in Pawlikowski (1990). The controversy over the virtues of deletion at that time is also described in Section 2 of that paper. Notable transient methods proposed since then include Yücesan (1993), Jackway and deSilva (1992), Goldsman, Schruben and Swain (1994). However two comparative studies (McNickle, Pawlikowski and Stacey, 1993; Ghorbani 2004) have failed to find a substantial improvement in overall reliability over a method based on Schruben’s test (Schruben 1982, and Schruben Singh and Tierney 1983). In fact some of the alternatives that have been proposed do not appear to perform well at all. Details of our sequential implementation of a Schruben-based method can be found in Section 3 of Pawlikowski (1990).

A recent survey of transient deletion methods (Ghorbani 2004) showed quite mixed results, with some methods giving quite inconsistent results in terms of the length of transient removed. Some methods showed a strong tendency to select an empty-and-idle state as the end of the transient period, producing potentially very long transient periods that in fact were no use at all since simulations are often started from that state. Even the method we use, a combination of a heuristic and a test, turns out to be predominantly driven by the heuristic, with steady state being identified at the end of the 25 crossings of the running mean. Given these results we ask if it is worth persisting with transient deletion methods, or should the reduction in the cost of computing simply be used for longer runs or runs to more accurate results (smaller values of relative precision)? Alternatively does the effect of ignoring the transient problem disappear within today’s acceptable run times?

There are a number of problems with the evaluation of transient deletion methods. This has largely been carried out on simple queueing models (since it can only be done for systems for which we know the steady-state result), and further, much of the evaluation has been against a different, albeit related measure of convergence to steady-state: the relaxation time. In addition studies of methods for dealing with the initial transient have concentrated on estimators of means, with the result that their performance on other statistics: higher moments, quantiles, remains uncertain (one exception to this is Lee, Pawlikowski and McNickle, 2000). Finally few methods have been tested against different initial starting conditions. So the overall reliability of many proposed methods remains an open question. Here for simplicity we will also restrict ourselves to queueing models, means, and initiate simulation from the "empty-and-idle" state.

With the steep decline in the cost of computing, the availability of large-scale computing resources via networks and the web, and simulation software that can carry out replications in parallel, such as Akaroa2, (Ewing, Pawlikowski and McNickle 1999) it is now possible to collect very large samples of simulation output data for steady-state simulation problems in acceptable time and at acceptable cost. Given this, is it true that the influence of the initial state of the simulated system could be expected to be quite limited, since the initial transient may now form a very small fraction of the total run? And given the uncertainty about the overall performance of some of the deletion methods has the balance shifted back in favour of not deleting observations?

Proceedings 21st European Conference on Modelling and Simulation
Ivan Zelinka, Zuzana Oplatková, Alessandra Orsoni ©ECMS 2007
ISBN 978-0-9553018-2-7 / ISBN 978-0-9553018-3-4 (CD)

2. COVERAGE ANALYSIS

The measure we will use for estimating the effect of the initial transient is the coverage of the estimated confidence intervals for the final result. That is, if supposedly 95% (for example) confidence intervals of a specified relative precision are being used as the stopping criterion for sequential simulation, we use independent replications to measure the fraction of estimated confidence intervals which actually contain the true value of the parameter of interest. Any bias due to the initial transient could be expected to show in reduced actual coverage. We follow the coverage estimation methodology described in Pawlikowski, McNickle and Ewing (1998). There we argued that:

R1. Coverage of the final results from sequential simulation should be analyzed sequentially, i.e. analysis of coverage should be stopped when the relative precision (the relative half-width of the confidence interval) of the estimated coverage reaches a specified level, say 5%.

R2. An estimate of coverage has to be calculated from a representative sample of data, so the coverage estimation starts only after a minimum number (say 200) “bad” confidence intervals have been recorded.

R1 arises from the fact that sequential analysis is the only practical way of guaranteeing results of a pre-specified precision when standard errors are unknown, and hence should be used in most steady-state simulations. Thus coverage studies, as meta-simulations, should also reflect this practice. Practically also, for this study, sequential analysis was essential as producing just one of the estimates of coverage for this study could involve up to 15,000 independent replications, each using thousands of observations and requiring days of elapsed time as a background task.

Sequential analysis does, however, have the problem that some of the simulation experiments may stop after an abnormally small number of observations have been collected, because, by chance, the stopping criterion is temporarily satisfied. In Pawlikowski, McNickle and Ewing. (1998) and Ewing, McNickle and Pawlikowski (2002) our objective was to compare methods of confidence interval estimation such as Spectral Analysis and Batch Means in a sequential (and multiple processor) framework. So simulation runs that were abnormally short were discarded by removing those more than one standard deviation shorter than the average length to remove this extraneous source of variation. This step does not seem appropriate here, because the initial transient might have some influence on the occurrence of abnormally short runs.

Lee, Pawlikowski and McNickle (1999) show that in practice there are some practical heuristics that can guard against runs that are too short. However in this paper the effect of discarding the initial transient data is considered without discarding abnormally short simulations or applying these heuristics. Thus the coverage results appear worse than they would be in practice.

3. METHODOLOGY

The experiments were run using the Akaroa2 simulation package, using it in its single-processor mode. The automated method based on Schruben’s test as referenced above was used to determine the length of the initial transient period. This sequential version of the Schruben test uses a heuristic to decide on an estimated length of initial transient period. This heuristic was proposed by Gafarian, Anker and Morisaku (1978) and is described in detail in Pawlikowski (1990). Using this heuristic, the length of initial transient period is taken to be over when the sequence has crossed its running mean 25 times. Schruben’s test is then used to test for stationarity. If the null hypothesis of stationarity is rejected, a larger potential transient period is considered (Pawlikowski, 1990).

Sequential Spectral Analysis, a modification of the method proposed by Heidelberger and Welch (1981) and specified in Pawlikowski (1990), was used to estimate the confidence interval width. We have found that this method gives accurate confidence intervals, especially for highly correlated data, such as waiting times in highly loaded queues (Ewing, McNickle and Pawlikowski 2002, McNickle, Pawlikowski, and Ewing, 2004).

Experiments were conducted for response times in a representative range of queueing models: M/M/1, M/D/1 and M/H_2/1 and a simple computer network model as shown in Figure 1. For the M/H_2/1 model the square of the coefficient of variation is set to 5.

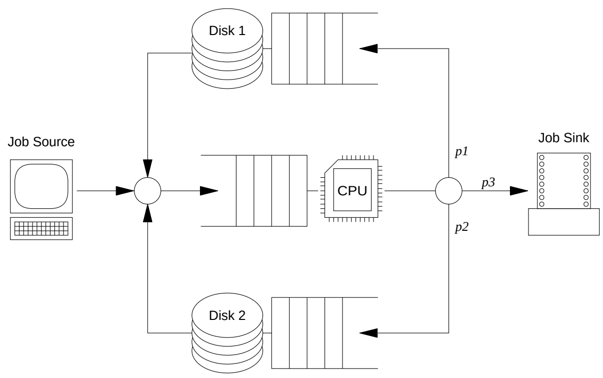
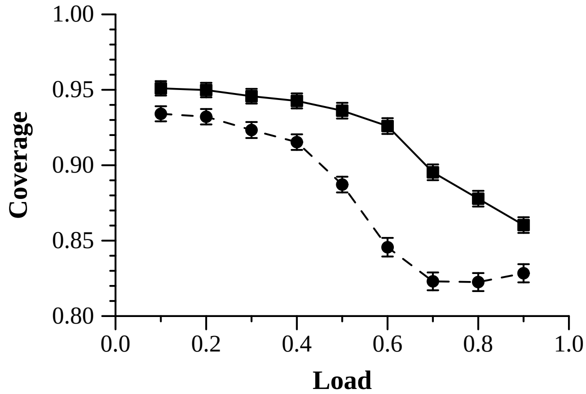


Figure 1: The computer network model

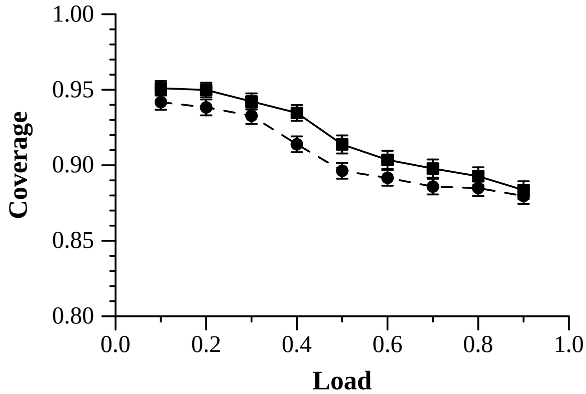
For the computer network model the mean CPU service time = 6, the mean service time for each disk = 14, p_1 = p_2 = 0.4, all distributions are negative exponential, and the source rate is set to give traffic intensities at the CPU ranging from 0.1 to 0.9.

4. RESULTS

These graphs plot the average coverage (together with 95% confidence intervals) from simulating the response times in the models specified, for transient deletion (solid line) or no deletion (dashed line). So they show the actual coverage that was achieved when the required coverage was set to that of a 95% (0.95) confidence interval having a relative width of either (a) 10% or (b) 5%

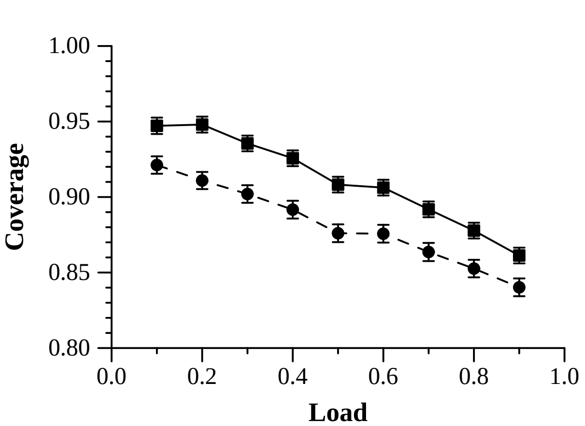


(a) Relative Precision = 10%

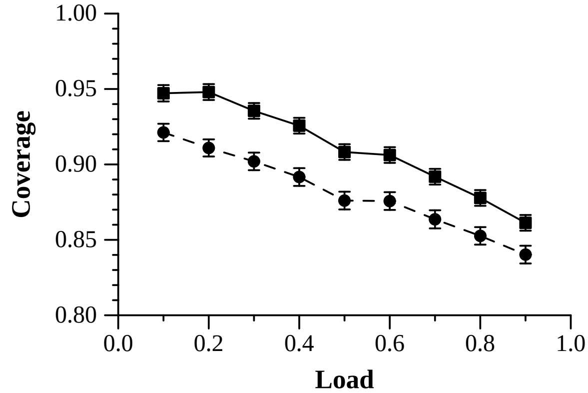


(b) Relative Precision = 5%

Figure 2: Effect of transient deletion M/D/1

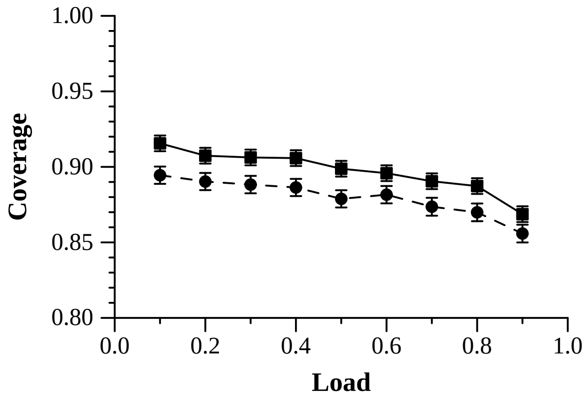


(a) Relative Precision = 10%

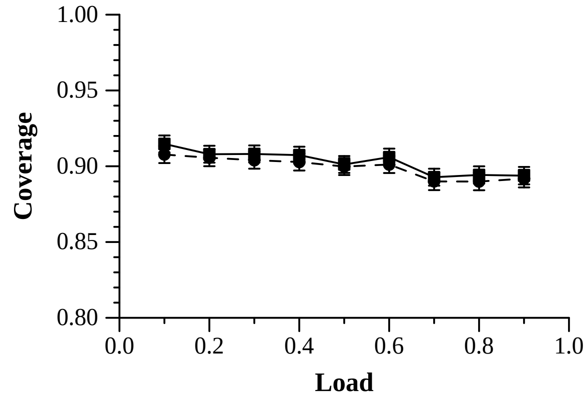


(b) Relative Precision = 5%

Figure 3: Effect of transient deletion M/M/1

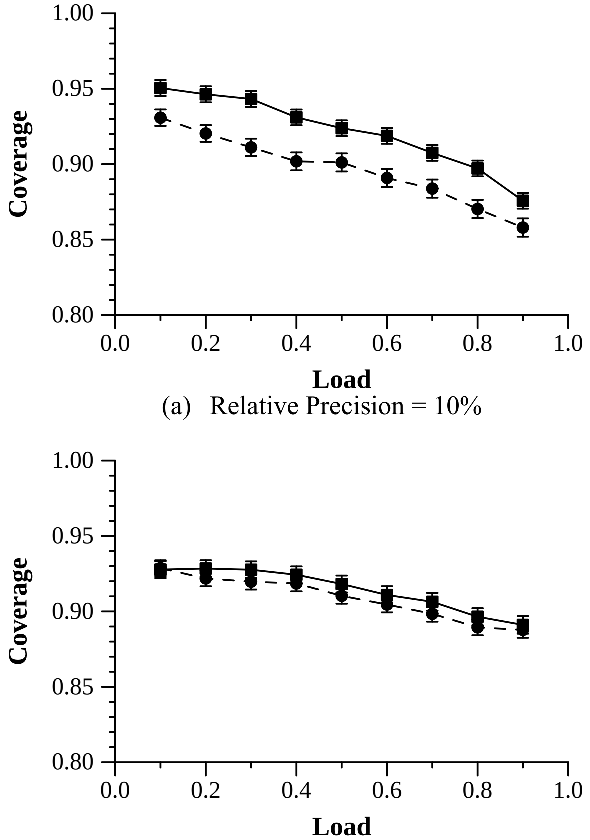


(a) Relative Precision = 10%



(b) Relative Precision = 5%

Figure 4: Effect of transient deletion $M/H_2/1$



(a) Relative Precision = 10%

(b) Relative Precision = 5%

Figure 5: Effect of transient deletion CPU Queue in Computer Network

5. DISCUSSION

From the graphs it can be seen that deletion of the initial transient does appear to produce some effect on coverage for models in which the run lengths are short, but that the effect reduces with: the variability of the model, and the accuracy of the precision (aside from initial effects due to the analysis method the 5% runs are about four times as long as those for 10% precision), to the point where in a single experiment the reduction in coverage would no longer be deemed to be statistically significant.

Thus for $M/H_2/1$ at 5% precision the effects are very small. Similarly for the CPU queue in the computer network model the 5% relative precision runs are also almost indistinguishable from those in which a transient period has been deleted. This is presumably due to the usually positive autocorrelation in the input process produced by the feedback of jobs (McNickle, 1984) resulting in long run lengths. In all cases the no-deletion coverage increases towards that produced when an initial transient period is deleted, as the run length increases.

It might be thought that unless the transient period is deleted, substantially longer runs are needed to achieve similar precision. But this turns out not to be the case. We give only two tables to save space since all the results were similar. Table 1 gives the average numbers of observations recorded firstly after the deletion of a transient period (column 3) and then if no transient period is deleted (column 4), for various values of the traffic intensity (ρ) for the M/M/1 model stopped at a relative precision of 5%. Thus it corresponds to the results shown in Figure 3(b). Table 2 gives similar values for the $M/H_2/1$ queue, corresponding to the results shown in Figure 4(b).

Table 1. Average Run Lengths for M/M/1 with a Relative Precision of 5%

ρ	No. of obs. deleted	No. of obs. after deletion	No. of obs. without deletion
0.1	259	3194	2936
0.2	265	4440	4139
0.3	274	6400	5965
0.4	286	9542	8963
0.5	302	14945	14089
0.6	328	24938	23564
0.7	373	48158	45434
0.8	466	115418	110470
0.9	713	492805	488304

Table 2. Average Run Lengths for $M/H_2/1$ with a Relative Precision of 5%

ρ	No. of obs. deleted	No. of obs. after deletion	No. of obs. without deletion
0.1	295	23256	22058
0.2	322	39233	37624
0.3	356	58712	55976
0.4	398	82419	78938
0.5	460	116171	111961
0.6	553	172622	167787
0.7	711	277253	270914
0.8	1032	556721	548065
0.9	1743	1934901	1920458

Rather than requiring more observations, we note that in fact the average run lengths without deletion are uniformly shorter than those when deletion is used. We conjecture that this may be because while the results measured during the transient period are biased, they are also of low variance, leading to, on the average, shorter runs.

We further note that the shorter run lengths alone are sufficient to account for the reduction in coverage between the deletion and the no-deletion results.

6. CONCLUSIONS

While the use of a reliable method of transient deletion such as that based on Schruben's test or Goldsman, Schruben, and Swain (1994) can still be recommended, the gains in reduced bias or variance appear to be modest. For situations outside those for which transient deletion methods have been validated it appears to be equally possible to rely on high-precision (i.e. small) confidence intervals in order to guarantee the accuracy of the final results. Aside from the increased run lengths needed to produce this higher precision we did not find any evidence that dealing with the transient problem by means of specifying higher precision for the results will require the collection of substantially more observations.

REFERENCES

Ewing, G., K. Pawlikowski and D. McNickle. 1999 Akaroa-2: Exploiting network computing by distributed stochastic simulation *13th European Simulation Multiconference*, Warsaw, June 1999, SCSC: 175-81.

Ewing, G., D. McNickle and K. Pawlikowski. 2002. Spectral analysis for confidence interval estimation under multiple replications in parallel, *Proceedings 14th European Simulation Symposium,* Dresden October 2002, 52-61.

Gafarian, A. V., Ancker, C. J., and Morisaku, T. 1978. Evaluation of Commonly Used Rules for Detecting "Steady State" in Computer Simulation. *Naval Research Logistics Quarterly*, 78, pp. 511-529.

Ghorbani, B. 2004, *Initial Transient Phase of Steady State Simulation: Methods of Its Length Detection and Their Implementation in Akaroa2*, M.Com. thesis, Computer Science and Software Engineering, University of Canterbury

Goldsman, D., L. Schruben, and J. Swain, 1994. Tests for Transient Means in Simulated Time Series. *Naval Research Logistics*, **41**, 171 – 187.

Heidelberger, P. and P. D. Welch, 1981, A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM*, 24, no. 4 (April) 1981, 233-245.

Jackway, P. T., and deSilva, B. M. 1992. A Methodology for Initialisation Bias Reduction in Computer Simulation Output. *Asia-Pacific Journal of Operational Research*. 9, 87-100.

Lee, J.-S. R., K. Pawlikowski and D. McNickle. 1999. Sequential steady-state simulation: rules of thumb for improving the accuracy of the final results *Proc. ESS99 (Erlangen, Germany) Oct 26-28 1999.*

Lee, J.-S. R., K. Pawlikowski and D. McNickle. 2000 Initial Transient Period Detection for Steady-State Quantile Estimation, *Proc. Int. Summer Computer Simulation Conf. (SCSC'2000*, Vancouver, Canada, July 16-20, 2000), Int. Society for Computer Simulation, 2000, S213.1-6

McNickle, D., 1984 Autocorrelations in the input and output processes in simple Jackson networks, *New Zealand Operational Research*, 12(2).

McNickle, D., Pawlikowski, K. and Ewing, G. (2004) Refining spectral analysis for confidence interval estimation in sequential simulation. Budapest, Hungary: *Proceedings of the ESS 2004*, 99-103, Oct 2004.

McNickle, D., K. Pawlikowski and C. Stacey, 1993. Detection and significance of the initial transient period in quantitative steady-state simulation, (with) *Proceedings of the Eighth Australian Teletraffic Research Seminar, RMIT Melbourne, 6-8 December 1993*, 193-202.

Pawlikowski, K., D. McNickle and G. Ewing. 1998 Coverage of confidence intervals from sequential steady-state simulation, Simulation *Practice and Theory, 6*, (1998), 255-267.

Pawlikowski, K, 1990, Steady state simulation of queueing processes: a survey of problems and solutions, *ACM Computing Surveys*, 22, June 1990, 123-170

Schruben, L., Singh, H., and Tierney, L. 1983. Optimal Tests for Initialisation Bias in Simulation Output. *Operations Research*, 31(6), 1167-1178.

Schruben, L. 1982. Detecting Initialisation Bias in Simulation Output. *Operations Research*, 30(3), 569-590.

Yücesan, E. 1993. Randomisation Tests for Initialisation Bias in Simulation Output. *Naval Research Logistics,* 40, 643-663.

BIOGRAPHIES

DON MCNICKLE is an Associate Professor of Management Science in the Department of Management at the University of Canterbury. His research interests include queueing theory; networks of queues and statistical aspects of stochastic simulation. He is a member of INFORMS. His email address is <don.mcnickle@canterbury.ac.nz>.

GREG EWING is a research associate in the Department of Computer Science and Software Engineering at Canterbury; where received a Ph.D. His research interests include simulation; distributed systems; programming languages, 3D graphics and graphical user interfaces. He has made contributions to the Python programming language; and has recently been nominated for membership of the Python Software Foundation. His email address is <greg.ewing@canterbury.ac.nz>

KRZYSZTOF PAWLIKOWSKI is a Professor of Computer Science at the University of Canterbury. His research interests include quantitative stochastic simulation; and performance modelling of telecommunication networks. He received a PhD in Computer Engineering from the Technical University of Gdansk; Poland. He is a Senior Member of IEEE and a member of SCS and ACM. His email address is <krys.pawlikowski@canterbury.ac.nz> and his web page is <http://www.cosc.canterbury. ac.nz/krys/>.

ACKNOWLEDGEMENT

This research was supported in part, by a grant from the College of Business and Economics, University of Canterbury