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

# A STATISTICAL PROCESS CONTROL APPROACH FOR ESTIMATING THE WARM-UP PERIOD

Stewart Robinson

Operational Research and Systems Group
Warwick Business School
University of Warwick
Coventry, CV4 7AL, UNITED KINGDOM

## ABSTRACT

One means for dealing with initialization bias in simulation experiments is to implement a warm-up period. This requires the correct estimation of the initial transient. A new method for determining the warm-up period, based upon the principles of statistical process control (SPC), is described. The method is tested on empirical data from a simulation model that has been used in a real-life study. In comparing the results to those from two commonly used warm-up methods, it appears that the SPC method performs well. The strengths and weaknesses of the approach are discussed.

## 1 INTRODUCTION

In estimating the steady-state parameters of a simulation model, the correct removal of any initialization bias is of vital importance. There are two methods open to the simulation modeller. First, the starting condition of the model can be set such that there is no bias in the output data. This requires the correct setting of the starting condition. Second, the model can be run for a warm-up period and the data are then deleted from that period (the initial transient). This, of course, requires the correct estimation of the warm-up period. It is this latter approach upon which this paper focuses.

The purpose of this paper is to describe a new method for estimating the warm-up period of a steady-state simulation. The method is based upon the principles of statistical process control (SPC) and is known as the SPC method. Before describing this method there is a brief review of the literature on estimating the warm-up period. The SPC method is applied to the output data from a simulation model used in a real-life study. The results are compared with those from two other commonly used methods. There is a discussion on the strengths and weaknesses of the proposed approach.

## 2 METHODS FOR ESTIMATING THE WARM-UP PERIOD

Various methods have been proposed for identifying the initial transient in simulation models. The warm-up methods that have been devised can be categorised under five headings:

- *Graphical methods*: these involve the visual inspection of time-series of the output data.
- *Heuristics approaches*: these apply simple rules, with few underlying assumptions.
- *Statistical methods*: these rely upon the principles of statistics for determining the warm-up period.
- *Initialization bias tests*: these identify whether there is any initialization bias in the data and, therefore, they are not strictly methods for identifying the warm-up period, but they can be used in combination with warm-up methods to determine whether they are working effectively.
- *Hybrid methods*: these involve a combination of graphical or heuristic methods with an initialization bias test.

There is no attempt to describe these methods here, but simply to categorise and list them (Table1). It is obvious that many methods have been suggested over a period of more than 30 years, albeit that there does not seem to have been a concentrated period of research into this critical issue in simulation modelling.

In looking at the literature and particularly case studies, it is apparent that few of these methods are used in practice. Indeed, the most popular methods appear to be simple time-series inspection and Welch's method. Anecdotal evidence suggests that time-series inspection is the main method used by simulation practitioners, other than guessing! Further to this, many of the methods listed in Table 1 have not been thoroughly tested.

Table 1: Warm-up Methods

| Category | Method | Reference |
|---|---|---|
| Graphical methods | Time-series inspection | Robinson [1994] |
| | Ensemble average plots | Banks et al. [1996] |
| | Cumulative mean rule | Gordon [1969] |
| | Deleting the cumulative mean rule | Banks et al. [1996] |
| | Cusum plots | Nelson [1992] |
| | Welch's method | Welch [1983] |
| | Variance plots | Gordon [1969] |
| Heuristics approaches | Schriber's rule | Pawlikowski [1990] |
| | Conway rule | Gafarian et al. [1978] |
| | Modified Conway rule | Gafarian et al. [1978] |
| | Crossing of the mean rule | Fishman [1973] |
| | Autocorrelation estimator rule | Fishman [1971] |
| | Marginal confidence rule | White [1997] |
| | Goodness of fit | Pawlikowski [1990] |
| | Relaxation heuristics | Pawlikowski [1990] |
| Statistical methods | Kelton and Law regression method | Kelton and Law [1983] |
| | Randomisation tests | Yucesan [1993] |
| Initialization bias tests | Schruben's maximum test | Schruben [1982] |
| | Schruben's modified test | Nelson [1992] |
| | Optimal test | Schruben et al. [1983] |
| | Rank test | Vassilacopoulos [1989] |
| | The new maximum test | Goldsman et al. [1994] |
| | Batch means test | Goldsman et al. [1994] |
| | Area test | Goldsman et al. [1994] |
| Advanced methods | Pawlikowski's sequential method | Pawlikowski [1990] |
| | Scale invariant truncation point method | Jackway and deSilva [1992] |

There is certainly room for further research into estimating the warm-up period, both in terms of testing the existing methods and in developing new approaches that can be implemented by simulation practitioners. One such method is now described.

## 3 THE SPC METHOD

In SPC two forms of variation are identified. "Common" or "inherent" causes are sources of variation that are unavoidable. In manufacturing systems these are caused by small variations in, for instance, the materials being processed, the environment (e.g. heat and light) and the performance of the staff. In these circumstances the process is assumed to be varying according to some fixed distribution about a constant mean. Meanwhile "special" or "assignable" causes of variation can be corrected or eliminated, for instance, a tool may be worn or staff may be poorly trained. In these circumstances the process ceases to vary according to some fixed distribution about a constant mean. A process that is subject only to common causes of variation is said to be "in-control". When special causes of variation occur, the process is said to be "out of control". The purpose of SPC is to determine when a process is going out of control, so the process may be corrected.

There is a close relationship between the concepts of SPC and those of transience and steady-state in simulation output analysis. A model that is in steady-state varies about a constant mean according to some fixed distribution, while in transience it does not. As such, a simulation that is in steady-state could be considered to be "in-control", while during a transient phase it could be considered to be "out of control". It seems possible, therefore, that the methods employed in SPC could be used to detect when a model is in steady-state, and when it is not. The method presented here employs SPC in this way, with a particular view to identifying the initial transient for a simulation with steady-state parameters. Unlike the implementation of SPC in real-life processes, there is no sense of intervening in a model run in order to return the model to a steady-state, it is assumed to be a self-controlling process. The method is now described by a series of stages that are based on SPC theory.

### 3.1 Stage 1: Perform Replications and Collect Output Data

First, time-series data on a key output statistic need to be collected. Typically this would be throughput for manufacturing systems and customer waiting time in service systems. The output data should be collected over a series of replications in order to provide a number of samples for each data point. The observation interval should be relatively short; for typical manufacturing and service system models a one hour interval is probably appropriate.

Standard SPC approaches assume that the data are normally distributed. This, of course, may not be the case for simulation output data. By performing a series of replications, however, it can be assumed (via the central limit theorem) that the sample mean at each time interval tends towards normality, unless the data are highly skewed. The larger the sample size the stronger this assumption. Therefore, the more replications that can be performed the better. It is recommended that between five and ten are performed as a minimum.

The length of each replication also needs to be determined. It is recommended that the run length is at least four times longer than the estimated length of the initial transient and that at least 60 data points are collected in each replication. The rationale for this is given in stage 3 of the method. An initial estimate of the length of the initial transient could be obtained by simply inspecting a time-series of the output data to look for where the model appears to be in steady-state.

On completion of this stage a time-series of the key output data for each replication should have been collected, represented by $Y_{ji}$, where $i$ is the observation interval and $j$ is the replication number. From this the sample means at each observation interval can be calculated as follows:

$$\overline{Y}_i = \frac{\sum_{j=1}^{n} Y_{ji}}{n} \tag{1}$$

where $n$ is the number of replications performed. As a result the vector $y(m)$, representing a time-series of sample means, is formed:

$$y(m) = \left(\overline{Y}_1, \overline{Y}_2, \ldots, \overline{Y}_m\right) \tag{2}$$

where $m$ is the total number of observations made in each replication.

### 3.2 Stage 2: Test that the Output Data Meet the Assumptions of SPC

SPC relies on two key assumptions: that the data are normally distributed and that the data are not correlated. In stage 2 these two assumptions need to be tested for the data that have been collected.

Simulation output data are typically autocorrelated. The batch means method is one approach for dealing with this autocorrelation (Law and Kelton, 2000). By combining observations into batches, the batch means become approximately uncorrelated as the batch size ($k$) increases. The time-series from Equation (2) is combined into a series of batches to generate the vector $y(k)$ as follows:

$$y(k) = \left( \frac{\sum_{i=1}^{k} \overline{Y}_i}{k}, \frac{\sum_{i=k+1}^{2k} \overline{Y}_i}{k}, \ldots, \frac{\sum_{i=(b-1)k+1}^{bk} \overline{Y}_i}{k} \right) \tag{3}$$

which is represented as:

$$y(k) = (y(1), y(2), \ldots, y(b)) \tag{4}$$

where $b$ is the number of batches of length $k$ that can be calculated from the time-series, such that $m=bk$. A popular approach for determining $k$ is Fishman's procedure (Fishman, 1978). Starting with a batch size of 1 the value of $k$ is increased until the lag 1 autocorrelation is close to zero, that is, less than 0.1. Note that by working with mean values from a series of replications the autocorrelation will be lower than for the time-series from the individual replications.

The data should also be tested for normality at each value of $k$. A typical approach is the Kolmogorov-Smirnov test (Law and Kelton, 2000). It should be noted that as $k$ increases so the batch means tend towards normality (Law and Kelton, 2000). SPC approaches are available for data that are not normally distributed, these are, however, beyond the scope of this paper.

### 3.3 Stage 3: Construct a Control Chart for the Batch Means Data

Once a value of k has been found for which the assumption of low autocorrelation and normality are met, a control chart can be constructed for the data in Equation (4). Note this should use the minimum value of $k$ that passes the tests for autocorrelation and normality. The use of batch means control charts is described by Runger and Willemain (1996).

The population mean ($\mu$) and standard deviation ($\sigma$) of the data in $y(k)$ is estimated from the data in the second half of the time-series. In other words, if the total number of batches ($b$) in the series is 100, then the mean and standard deviation are calculated for the data in the range $b$=51 to $b$=100. Note that if $b$ were 99 then the range $b$=51 to $b$=99 would be used. This is a procedure used in other warm-up methods, for instance, Schruben et al. (1983). It is for this reason that it is recommended that at least 60 data points are collected for each replication, so the mean and standard deviation are estimated from a sample size of at least 30. Obviously the number of observations needs to be multiplied by $k$, in order to ensure that there are 60 data points in the batch means time-series. It is recommended that the run length is at least four times the estimated length of the initial transient, so the mean and standard deviation are calculated on steady-state data.

The standard deviation is estimated as follows:

$$\hat{\sigma} = \sqrt{\frac{1}{\left\lfloor \frac{b}{2} \right\rfloor} \sum_{l=b-\left\lfloor \frac{b}{2} \right\rfloor+1}^{b} s_l^2} \tag{5}$$

where $s_l^2$ is the standard deviation about the individual means in *y(k)*, and is calculated from the individual observations obtained from each replication. That is, if 10 replications have been performed, $s_1^2$ would be the standard deviation of the 10 data points in the mean *y(1)*.

From these data two sets of control limits are calculated. The first is the warning limits:

$$WL = \hat{\mu} \pm 1.96\hat{\sigma} / \sqrt{n} \tag{6}$$

and the second is the action limits:

$$AL = \hat{\mu} \pm 3.09\hat{\sigma} / \sqrt{n} \ . \tag{7}$$

The value of $z$ is selected so it is expected that 1 in a 1000 samples will fall outside of the action limits (3.09), while 1 in 40 will fall outside the warning limits (1.96). A control chart is then constructed showing the mean ($\hat{\mu}$), the upper and lower action and warning limits, and the time-series data *y(k)*.

### 3.4 Stage 4: Identify the Initial Transient

The final stage is to view the plot of the control chart. During the initial transient the time-series data *y(k)* will be "out of control". Once the model reaches steady-state, the data will be "in-control". Bissell (1994) identifies a number of rules for determining whether a control chart is showing data that are out of control:

- The time-series data violates an action limit (note that occasional violations, 1 in 1000, are to be expected for in-control data, and can be explained by Type I errors).
- Two consecutive values violate either the upper or the lower warning limit.
- Frequent values in relatively close succession that violate a warning limit.
- Persistent trend in the time-series data.
- A run of seven or more values on either side of the mean ($\hat{\mu}$).
- Excessive zig-zagging, with few points near to the mean ($\hat{\mu}$), and many near to the control limits.
- The warm-up period for the model can be selected by identifying the point at which the time-series data is in-control and remains in-control.

If a steady-state cannot be found this could mean one of two things. Firstly, the model has not been run for long enough to reach a steady-state. This can simply be addressed by returning to stage 1 and running the model for longer. Secondly, the model does not ever reach a steady-state and that the output should be classified as transient.

A summary of the SPC method is provided in Figure 1.

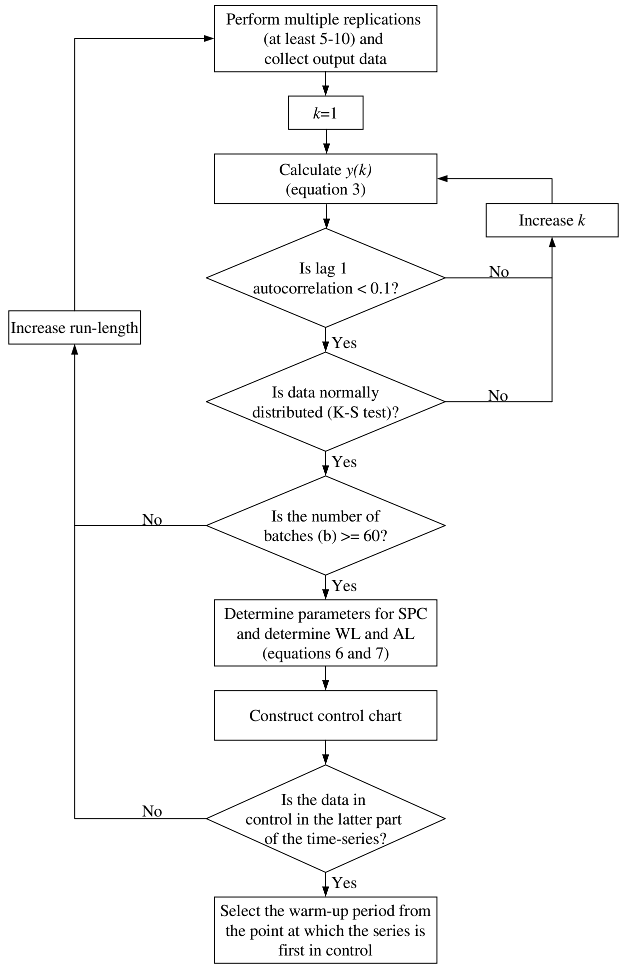


Figure 1: Summary of the SPC Method

## 4 EXAMPLE APPLICATION OF THE SPC METHOD

In order to demonstrate the use of the SPC method it is now applied to a model that has been used in a real-life simulation study. The model represents an engine block machining line. The production facility consists of 19 operations that are performed in series with some having multiple machines. Each operation is separated by a buffer area. Machine failures and tool-changes are the primary source of stochasticity. Throughput is the main measure of system performance. Five replications were performed with a run length of 1,000 hours. Statistics were collected on hourly throughput. Fig-

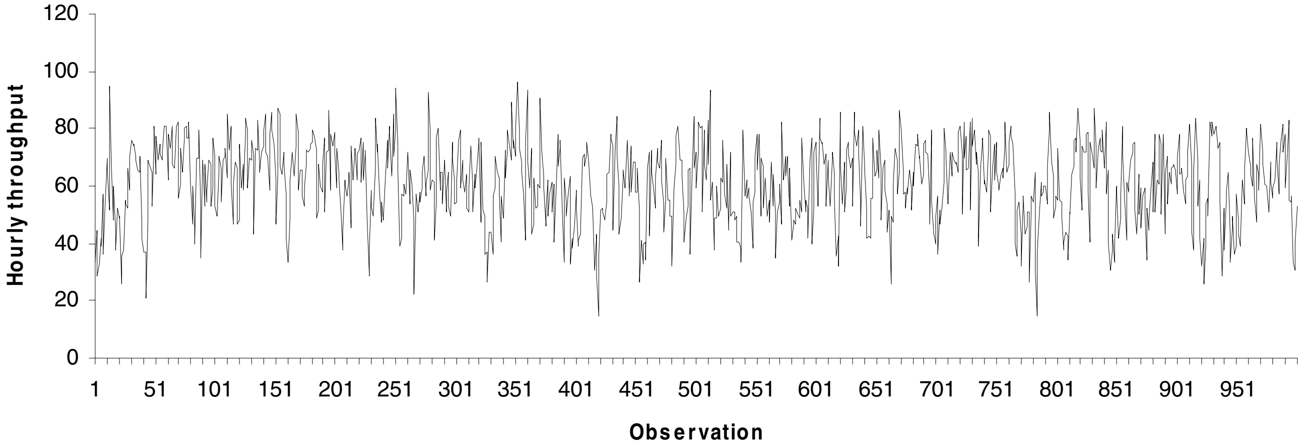


Figure 2: Engine Block Machining Line: Hourly Throughput (Mean of 5 Replications)

ure 2 shows the time-series of hourly throughput, calculated as the mean of the five replications.

The initial transient for the model's output is not known. Therefore, the performance of the SPC method is tested in two ways. First, by comparing the result with other methods. Here time-series inspection and Welch's method are used for comparison, on the grounds that they appear to be in most frequent use. Second, an initialization bias test is applied to the results from the three warm-up methods. This determines whether the selected warm-up period is sufficiently long to remove any initialization bias. Schruben at al's (1983) optimal test is used for this purpose.

### 4.1 Result from the SPC Method

The Kolmogorov-Smirnov test shows that the data are normally distributed even at a batch size ($k$) of 1 (Table2). The data are, however, quite highly autocorrelated. It is not until the batch size is increased to 8 that the autocorrelation criterion is met. Based on this batch size the control chart parameters are determined (Table3) and the chart is constructed (Figure 3).

Observation 1 and 3 are both below the lower warning limit, suggesting a warm-up period of at least 3 observations. Further inspection reveals, however, that observations 16 to 23 are all above the mean, breaking the rules for the chart to be in-control. Indeed, the majority of the observations between 7 and 25 are above the mean. This suggests that there may be some positive bias in this part of the time-series. It is concluded that a warm-up period of 25 observations, or 200 (25$k$) hours, is required.

### 4.2 Result from the Time-series Inspection Method

A detailed inspection of the time-series in Figure 2 shows that the first 9 observations are all below the apparent mean level. A warm-up period of 9 hours is therefore estimated.

Table 2: Autocorrelation and Normality Tests for Engine Block Machining Line Model

| Batch size ($k$) | Lag 1 auto-correlation | Test for normality: K-S value | Test for normality: Results for K-S ($\alpha$=0.05) |
|---|---|---|---|
| 1 | 0.49 | 0.045 | Reject |
| 2 | 0.46 | 0.055 | Accept |
| 4 | 0.35 | 0.078 | Accept |
| * 8 | 0.02 | 0.097 | Accept |

Table 3: SPC Parameters for Engine Block Machining Line Model (k = 8)

| **Mean** | **Standard deviation** | **Warning limits** | **Actions limits** |
|---|---|---|---|
| 60.61 | 18.76 | 60.61 ± 16.45 | 60.61 ± 25.93 |

### 4.3 Result from Welch's Method

A window size of 50 is required to obtain a reasonably smooth line (Figure 4). Obviously a larger window size would provide a smoother line, but the window has not been increased since this would lead to a more conservative estimate of the warm-up period. Based on the time-series presented in Figure 4 a warm-up period of 300 hours is estimated. A shorter period of nearer 100 hours might be selected, but there is some evidence of positive bias between observations 100 and 300 (the moving average appears to be higher).

### 4.4 Comparison of Results

The results for the estimated warm-up period for the engine block machining line are summarised in Table4, along with the results of the initialization bias test. It is immediately obvious that the time-series inspection method does not detect the positive bias in the early part of the time-

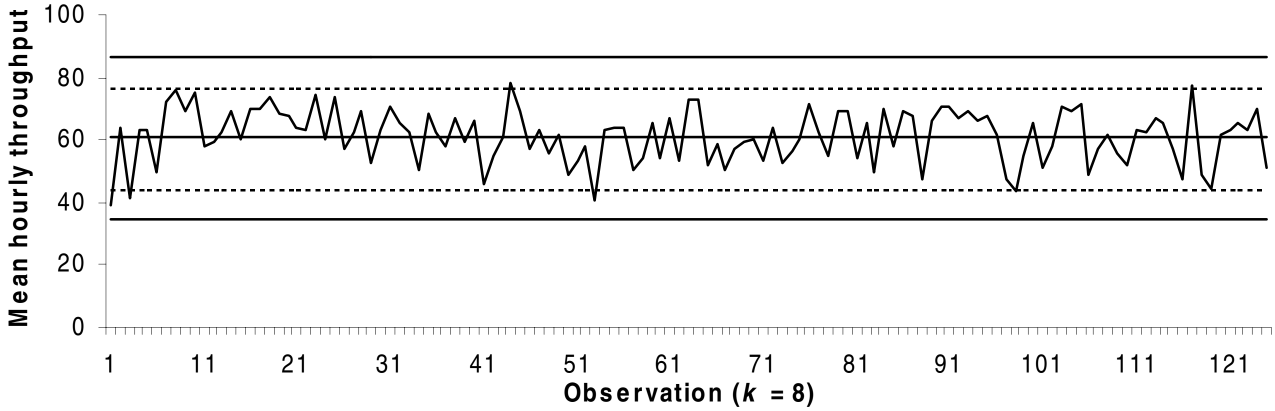


Figure 3: Control Chart for Engine Block Machining Line Model

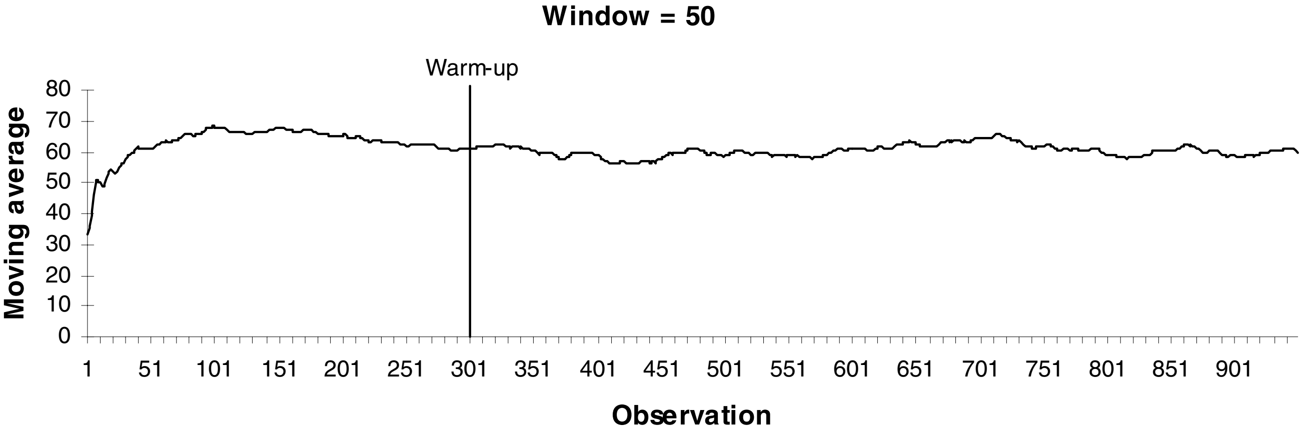


Figure 4: Welch's Method for Engine Block Machining Line Model

Table 4: Comparison of Results for Engine Block Machining Line Model

| | | Initialization bias Optimal test (α=0.05) | |
|---|---|---|---|
| **Method** | **Estimated warm-up period** | **Negative bias** | **Positive bias** |
| SPC method | 200 | No | No |
| Time-series inspection | 9 | No | Yes |
| Welch's method | 300 | No | No |

series. Both the SPC method and Welch's method identify and delete this bias, although it should be noted that the SPC method does so with more clarity. Whereas the control chart clearly shows the breaking of rules for being in-control, in Welch's method the bias is only detected by a subtle rise in the moving average. Had a warm-up period of 100 hours been selected from Welch's method the positive bias would still have been present. Further to this, Welch's method is much more conservative than the SPC method with a further 100 observations being deleted. This is not surprising since it is generally acknowledged that methods based on cumulative statistics are more conservative (Gafarian et al., 1978; Wilson and Pritsker, 1978; Pawlikowski, 1990; Roth, 1994).

The exact reason for the positive bias at the start of the run is unclear, but it is believed to be a consequence of the assumptions made in initializing the machine breakdowns. At the start of the run it is assumed that all the machines are halfway through the first time-between failure that is sampled. This may bias the results. As such a number of breakdowns may need to occur before the failures are shuffled sufficiently to lose this bias.

## 5 DISCUSSION

The example described above demonstrates that the SPC method can be an effective approach for identifying the warm-up period. It is apparent that in this case it provides a less conservative estimate of the warm-up period than does Welch's method. It is also apparent that it provides more guidance than simply inspecting a time-series, which underestimates the amount of warm-up required.

The SPC approach has various strengths. First, it is based upon visual inspection of a time-series of the data.

This has the advantage that the modeller is able to identify any unusual patterns in the simulation output, particularly if there are outliers or the model does not reach a true steady-state. Cumulative statistics approaches, such as Welch's method, can obscure this information. Second, it provides clear rules for determining when a model is in steady-state, albeit that these require some interpretation due to the probabilistic nature of statistical analysis. Neither time-series inspection nor Welch's method have such clear rules for determining when a model is in steady-sate. Indeed, one of the problems with Welch's method is deciding upon an appropriate window size. Finally, SPC is a familiar technique to many operations staff; these being among the main users of simulation modelling.

There are some weaknesses in the approach. The assumptions are quite rigorous, particularly regarding normality and independence in the data. In the example above, these issues have been adequately addressed using the batch means method. It is possible, however, that for some simulation output this may not be sufficient. Indeed, tests on data from M/M/1 queuing models with a high traffic intensity have shown the need for large batch sizes to reduce the autocorrelation. SPC approaches for dealing with correlated and non-normally distributed data could be employed in such circumstances (e.g. MacCarthy and Wasusri (2001)), but this adds to the complexity of the technique. Another problem is that much output data beyond the initial transient may be required to determine the warm-up period, both in terms of multiple replications and run-length. That said, these data are likely to be needed anyway in assessing the performance of the system under investigation.

## 6 CONCLUSION

This paper describes a new approach for determining the warm-up period for a discrete event simulation model. The SPC method is applied to a model used in a real-life simulation study and the result is compared to the results from the simple time-series inspection method and Welch's method. The strengths and weaknesses of the SPC approach are discussed.

Further work is required, particularly in testing the method on additional data sets and in comparing it with a wider range of approaches. Consideration may need to be given to adapting the method when the output data are very highly correlated and/or non-normally distributed.

## ACKNOWLEDGMENTS

The author wishes to thank Agisilaos Ioannou for his contribution to this work. Portions of this paper are based on S. Robinson (2002). Selecting a Warm-up Period: A Statistical Process Control Approach. Operational Research Society, Simulation Workshop, March 2002. Operational Research Society, Birmingham, UK.

## REFERENCES

Banks, J., J.S. Carson, and B.L. Nelson. 1996. *Discrete-Event System Simulation.* 2nd ed. Upper Saddle River, NJ: Prentice-Hall.

Bissell, D. 1994. *Statistical Methods for SPC and TQM*. London: Chapman & Hall.

Fishman, G.S. 1971. Estimating sample size in computer simulation experiments. *Management Science* 18:21-37.

Fishman, G.S. 1973. *Concepts and Methods in Discrete Event Digital Simulation*. New York: Wiley.

Fishman, G.S. 1978. Grouping observations in digital simulation. *Management Science* 24:510-521.

Gafarain, 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.

Goldsman, D., L.W. Schruben, and J.J. Swain. 1994. Tests for transient means in simulation time series. *Naval Research Logistics Quarterly* 41:171-187.

Gordon, G. 1969. *System Simulation*. Upper Saddle River, NJ: Prentice-Hall.

Jackway, P.T., and B.S. DeSilva. 1992. A methodology for initialization bias reduction in computer simulation output. *Asia-Pacific Journal of Operational Research* 9: 87-100.

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.

MacCarthy, B.L., and T. Wasusri. 2001. Statistical process control for monitoring scheduling performance – addressing the problem of correlated data. *Journal of the Operational Research Society* 52:810-820.

Nelson, B.L. 1992. Statistical analysis of simulation results. In *The handbook of industrial engineering, 2nd ed,* ed. G. Salvendy. New York: Wiley.

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

Robinson, S. 1994. *Successful simulation: a practical approach to simulation projects*. Maidenhead, UK: McGraw-Hill.

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.

Runger, R.C., and T.R. Willemain. 1996. Batch-means control charts for autocorrelated data. *IIE Transactions* 28:483-487.

Schruben, L.W. 1982. Detecting initialization bias in simulation output. *Operations Research* 30:569-590.

Schruben, L.W., H. Singh, and L. Tierney. 1983. Optimal tests for initialization bias in simulation output. *Operations Research* 31:1167-1178.

Vassilacopoulos, G. 1989. Testing for initialization bias in simulation output. *Simulation* 52 (4):151-153.

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

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

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

## AUTHOR BIOGRAPHY

**STEWART ROBINSON** is a senior lecturer in operational research at Warwick Business School in the UK. He holds a BSc and PhD in Management Science from Lancaster University. Previously employed in simulation consultancy, he supported the use of simulation in companies throughout Europe and the rest of the world. His research interests are in finding ways to improve the use of operational research, and specifically simulation, within industry. He is author/co-author of two books on simulation. Current work involves an investigation into the use of artificial intelligence to represent a human decision-maker in simulation models, the analysis of complex output behaviours from simulations, developing an understanding of quality in relation to operational research and simulation studies, and understanding the practice of simulation modelling. Email and web address `<stewart.robinson@warwick.ac.uk>` and `<www.btinternet.com/~stewart.robinson1/SLRHOME.HTM>`