Journal of the Operational Research Society (2002) 53, 559–573 0160-5682/02 $15.00

www.palgrave-journals.com/jors

Detecting shifts in the mean of a simulation output process

S Robinson[1]*, RJ Brooks[2] and CD Lewis[3]

[1]*Warwick Business School, University of Warwick, Coventry, UK;* [2]*Lancaster University, Lancaster, UK; and* [3]*Aston Business School, Aston University, Birmingham, UK*

The application of the correct simulation output analysis technique requires a knowledge of the model's behaviour. Traditionally, only two types of model behaviour are discussed for discrete event simulations: transient and steady state. In this paper, a third type of behaviour, denoted shifting steady state, is considered in which the model passes through successive periods of different steady states. A heuristic technique for identifying the shifts in the mean of a time series is applied to the output data from simulation models with known shifting steady-state behaviour in order to test its effectiveness in detecting the shifts. The heuristic performs well, indicating that it may be a valuable additional technique for output analysis.

Journal of the Operational Research Society (2002) **53**, 559–573. DOI: 10.1057/palgrave/jors/2601323

Keywords: discrete event simulation; steady state; shift of mean; Manhattan diagram; output analysis

Introduction

In order to analyse the output of a discrete event simulation model correctly it is important that the type of behaviour of the model is identified. This is vital for gaining a better understanding of the system being studied. Typically, the output of simulation models is classified as being either transient or steady state. This paper considers another type of behaviour that may commonly occur in practice, denoted shifting steady state, in which the model passes through successive steady-state periods.

In some cases, the pattern of behaviour of a simulation model may be identified by inspecting the input distributions for shifts in the input data. This will not always be effective, however, especially where there are complex interactions within the model or between the various input data. An alternative approach is to inspect time series of the simulation's output. Identifying a pattern in simulation output data can be problematic, however, since the random variation is often large compared to the pattern. This variation can be reduced by averaging the output across a number of replications and by calculating a moving average of the resulting values.[1] However, it will often be infeasible to perform enough replications to create a sufficiently smooth output series to reveal complex behaviour patterns. The approach taken in this paper is to apply a heuristic technique called the 'shifting mean' heuristic, devised by Lewis and Yeomans,[2] to identify shifts in the mean of a time series of simulation output. One reason for applying the heuristic in this situation is that it was specifically designed to find shifts in the mean of a data series with a low signal to noise ratio.

The paper starts by discussing the output characteristics of simulation models. The shifting steady-state model is then defined and the shifting mean heuristic is described as well as its application to three models with known output characteristics. The results of the findings are discussed and further tests performed to identify groupings of the steady-state periods. The paper concludes by applying the heuristic to a real model with unknown output characteristics and by discussing some of the limitations of the technique.

Output characteristics of simulation models

Most techniques for simulation output analysis address one of the following two scenarios. The first is a transient terminating model with a natural end-point event at which the simulation ceases. In this scenario, the period tends to be treated as a whole with the output variable often being an average for the period (such as average throughput or average queue length). For fixed initial conditions, repeated replications produce a set of independent values of the output variable which can then be analysed with straightforward statistical techniques.[3]

The second scenario consists of a non-terminating model that reaches steady state. Here, the techniques are generally based on the assumption that it is only the steady-state behaviour that is of interest. Methods are required both to

*Correspondence: S Robinson, Operational Research and Systems Group, Warwick Business School, University of Warwick, Coventry CV4 7AL, UK. E-mail: stewart.robinson@warwick.ac.uk

identify the length of the warm-up period before the system is sufficiently close to steady state (eg Welch[1]) and to analyse the output in steady state. The behaviour of the model in steady state can be studied either with multiple replications, as for the terminating simulation, or with a single long run. Independence of observations cannot be assumed when performing long runs. As a result, more complex statistical procedures must be employed, for instance, batch means, regenerative (renewal) method and standardised time-series method.[1,3–5] The user also has to decide the length of run required to achieve the desired level of confidence in the results.[6]

For both scenarios, the output is often compared for different experimental conditions. Methods of experimental design, including variance reduction, can be used effectively in performing comparisons,[7] and a number of statistical techniques have been suggested for drawing statistical inference.[3,8,9]

A simulation output variable can follow any trajectory however and, in practice, many models do not simply converge to a single steady state. To just treat such behaviour as transient will often miss patterns in the output variable that would facilitate the output analysis and, more importantly, would greatly increase the understanding gained of the nature of the system.

In particular, many simulation models incorporate only discrete changes in conditions, which will tend to result in discrete changes in output behaviour. For example, service models often include customer arrival rates that change each hour or each day. A manufacturing simulation may alter the resources available in different work periods (day versus night, for example) while assuming constant conditions within each period. Model parameters, such as machine cycle times, commonly change when the type of part being produced changes. Within the periods of constant conditions such models may tend towards a steady state so that, in each period, the model moves towards a new steady state. Clearly, complex behaviour, including chaos, can also occur in constant conditions (eg certain recursive functions such as the logistic equation[10]). However, if a model does move between different steady states, the identification of the points at which the change in behaviour occurs would greatly help the output analysis. Law and Kelton[3] appear to be the only authors to analyse this type of behaviour at all. However, they only consider the very limited case of repeating cycles (such as the pattern of work behaviour for each day) in which an output variable measured for the whole cycle (such as average daily throughput) tends towards a steady state, but the output variable measured over a different period (such as hourly throughput, including work breaks, etc) does not. Law and Kelton refer to this as steady-state cycle conditions. If neither this condition nor steady-state conditions apply, they simply refer to the model as having ‘other parameters’ and suggest that it should be treated as transient.

For relatively simple models, examination of the input distributions may indicate the periods of constant conditions on which the output analysis could be based. However, it must be recognised that the output distribution from a model is the result of a complex set of interactions between the input distributions. Therefore, it is not always possible to determine the output patterns that are likely to result from the variations in the inputs. For example, consider a production line that makes several components using a complex production schedule with some machines being dedicated to particular components and others processing more than one component. Each component takes a different path through the line and, in addition, the machine times for different components on the common machines are different. Random machine breakdowns would create randomness in the ordering of the components in the system, making it difficult to identify the periods in which the average throughput is in steady state simply from the inputs to the model. More generally, there may be a delay before a change in a particular input variable has an effect on the output variable of interest, again making it difficult to predict the point at which the behaviour changes. In such cases, a better approach would be to try and identify patterns that exist within the output data.

The next section formally defines a shifting steady-state model after which the shifting mean heuristic is described. Following this the heuristic is applied to an example model.

Definition of a shifting steady-state model

A model is in steady state if the probability distribution of its output is constant over time. Following Law and Kelton,[3] steady state can be defined formally as follows. For given initial conditions I, the cumulative probability distribution function, at time t, of the real-valued output variable x_t is given by

$$F_{t\backslash I}(z) = \Pr\left(x_t \leqslant z | I\right) \quad \text{for all } z \in \Re \tag{1}$$

The output variable exhibits steady-state behaviour if the cumulative probability function converges weakly[11] to a limit function $G(z)$ (the steady-state distribution) as $t \to \infty$, and it converges to the same function for all initial conditions, ie

$$F_{t\backslash I}(z) \xrightarrow{t\to\infty} G(z) \quad \text{for all } z \in \Re \text{ at which } G \text{ is continuous and for all } I \tag{2}$$

This definition only applies in the limit as $t \to \infty$. However, in practice, the interest in analysing simulation output is to identify the periods in which the output distribution is approximately constant for all initial conditions and the model is then described as being in steady state within such periods.[3]

A shifting steady-state model is one in which the output variable tends towards successive steady states. This means that the output can be divided into consecutive periods $P_1, P_2, \ldots, P_i, \ldots$, with the distribution of the output variable tending towards the steady-state distribution G_i in period P_i; that is, if the conditions in the period were maintained over a sufficiently long period of time, Equation (2) would hold approximately. It is assumed that, in each period, the model can be considered to be in steady state by the end of the period. The start and end of the steady-state periods will usually occur at a fixed point in time for each run of the model, but it is also possible for other conditions to determine the change in behaviour such as, for a production line model, the number of units of output produced. There can also be a cyclic pattern so that there exists a positive integer N such that

$$G_{i+jN} = G_i \quad \text{for } i = 1, 2, \ldots, N \text{ and for } j = 0, 1, 2, \ldots \tag{3}$$

Such a model is denoted as a cyclic shifting steady-state model with a cycle length of N periods. It should be noted that this is a quite different situation to Law and Kelton's[3] steady-state cycle conditions described in the previous section.

The shifting mean heuristic

The shifting mean heuristic search procedure is described in detail in Lewis and Yeomans.[2] Here a summary is provided. The heuristic is designed for a time series of n observations $x_1, x_2, \ldots, x_n$ drawn from a process with several shifts in the mean, and with the values between successive shifts being a constant mean value plus an independent random noise variate. The variance of the random noise is assumed to be constant throughout the process. Lewis and Yeomans[2] suggest a test to indicate that the series may not be suitable for the heuristic, which is discussed later.

The heuristic procedure uses CUSUMs (cumulative sums of errors) to search for shifts in the mean. The first stage is to search for the largest shift in mean using the CUSUM, S_r, based on the overall mean of the n observations, $\bar{x}$. The values of S_r are given by

$$S_r = \sum_{t=1}^{r} \left(x_t - \bar{x}\right) = S_{r-1} + (x_r - \bar{x}) \quad r = 1, 2, \ldots, n \tag{4}$$

where S_0 is defined to be equal to zero. Note also that $S_n = 0$. The occurrence of the first shift is determined by the heuristic to be the point at which $|S_r|$ is a maximum for $r = 1$ to n. Tests for the significance of the shift should then be applied and those suggested by Lewis and Yeomans[2] are discussed in the next section. If the shift is significant then the time series is split into two at the point at which the shift occurs and the heuristic repeated on both parts. Lewis and Yeomans recommend that the suitability of each part for the application of the heuristic should also be tested.

Denoting the first shift point as n_1, CUSUMs $S_{1,r}$, and $S_{2,r}$ are calculated on each part of the time series as follows:

$$S_{1,r} = S_{1,r-1} + (x_r - \bar{x}_1) \quad r = 1, 2, \ldots, n_1 \tag{5}$$

$$S_{2,r} = S_{2,r-1} + (x_r - \bar{x}_2) \quad r = n_1 + 1, n_1 + 2, \ldots, n \tag{6}$$

where $\bar{x}_1$ and $\bar{x}_2$ are the means of the two parts of the series $x_1, \ldots, x_{n_1}$ and $x_{n_1+1}, \ldots, x_n$, respectively, and $S_{1,0} = S_{2,n_1} = 0$. The second shift determined by the heuristic occurs at the point at which $|S_{1,r}|$ or $|S_{2,r}|$ is a maximum for $r = 1$ to n. The nature of the CUSUMs means that the values at the end of each part of the series, S_{1,n_1} and $S_{2,n}$, are both zero. In particular, since $S_{1,n_1} = 0$, the first shift is automatically removed from the search for the second and subsequent shifts.

As before, the significance of the shift should be tested. If significant, the series is again split at the point of the new shift. If each part still appears suitable for the heuristic, the procedure is repeated on the three parts of the series. Repeated applications of the heuristic will further divide the series until either the significance or suitability tests fail.

Lewis and Yeomans[2] tested the heuristic on two BS 5703[12] series and found that it successfully identified the shifts previously determined by other methods.

Testing the significance of shifts

At each stage, the heuristic splits a sub-process into two. If the data in the two parts can be assumed to be independent values taken from normal distributions with the same variance, Lewis and Yeomans[2] suggest using the usual t-value statistic for samples with equal variances to assess the significance of the difference in the means. This is given by

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \quad \text{with } n_1 + n_2 - 2 \text{ degrees of freedom} \tag{7}$$

where

$$S_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{8}$$

and where s_1^2 and s_2^2 are the standard deviations of the two parts. If equal variances cannot be assumed then the standard t-value statistic for unequal variances could be used.

However, it is unclear what the critical value of the statistic should be. In fact, the CUSUM, S_i, on a process of n values can also be written as

$$S_i = \frac{i(n-i)}{n}\left(\bar{x}_{i-} - \bar{x}_{i+}\right) \tag{9}$$

where $\bar{x}_{i-}$ and $\bar{x}_{i+}$ are the means of $x_1, \ldots, x_i$ and $x_{i+1}, \ldots, x_n$ respectively. Therefore, specifying the shift as the point at which the CUSUM is a maximum will tend to maximise the difference in means over the possible shift points 1 to $n-1$, which in turn will tend to maximise the t value. There is also an interaction between the stages, with CUSUM values for a given stage depending on the split points determined at previous stages.

In order to counteract dependencies due to the way the heuristic splits the data, Lewis and Yeomans[2] suggest using critical values of $\pm t_{\alpha/(2\sqrt{n_1+n_2-2})}$ rather than $\pm t_{\alpha/2}$ for a desired significance level of α, although no statistical justification is given for this. However, they also recognise that expert knowledge of the series should be applied at this point when assessing the importance of each shift, and they report both probabilities in their analyses of the two test series.

The one-way ANOVA followed by a multiple comparison test is also suggested by Lewis and Yeomans[2] as an alternative test for the significance of the shifts, along with two non-parametric tests, the medians test (again with adjustment of the critical values) and the Kruskal–Wallis one-way ANOVA.

Applicability of the heuristic

Lewis and Yeomans[2] identify four data processes without shifts in the mean that may still produce significant shifts when using the heuristic; namely the random walk, random walk with shift, stochastic mean and stochastic trend. In each case, the lag-one autocorrelation (calculated using the standard autocorrelation formula) of the complete data series is likely to be high. However, a moderate or low value would be expected for shifting mean data, unless the random noise is small compared to the magnitude of the shifts. Therefore, Lewis and Yeomans suggest that an autocorrelation exceeding 0.7 should be taken as an indication that the series may be one of these four processes rather than a shifting mean process. Further tests could then be applied to investigate whether one of the four models does apply.

A process consisting of a constant mean plus random noise (ie with no shifts) should give a very low lag-one autocorrelation. Consequently, a value of less than 0.3 is suggested by Lewis and Yeomans[2] as also indicating that the data may not be a shifting mean process. This is less important, as any shifts detected from a constant mean model should show as insignificant under the significance tests.

The output characteristics of a shifting steady-state simulation model may differ from the assumptions of the shifting mean heuristic in two ways. Firstly, the mean of the output variable will not normally shift immediately from one steady state to the next. Rather, there will be a transient period before the model is in steady state. Secondly, the output variables of a simulation model are sometimes highly autocorrelated.

A transient phase should not hinder the heuristic in detecting the shift in steady state. The shift point found is likely to be within the transient phase when the output process is moving from one mean to the other. This is because the absolute value of the CUSUM will tend to continue to increase until the trend of the output series crosses the mean value for the whole period being split. The precise point found, of course, also depends upon the values of the random noise for the particular runs. This means that the model may move away from the steady state at the very end of the identified steady-state periods, and it may be necessary to apply warm-up methods to both the beginning and end of each period if carrying out detailed analysis on the steady-state period.

A further effect of the transient phase is that the heuristic may subsequently detect one or more additional erroneous shifts of lesser statistical significance in the transient phase. Such periods will almost certainly have a mean between the means of the preceding and following periods and will have a short length. In simulation output analysis, only steady-state periods of a reasonable length are likely to be of interest and so it will usually be appropriate to treat any short periods found by the heuristic as transient.

A high correlation between successive output values can occur for some simulation output variables, such as the length of a queue. In such a case, the output within the steady-state periods does not consist of a mean plus random noise but, instead, behaves as a random walk, probably with a tendency to move towards a mean value. A high correlation within a steady-state period could lead to additional erroneous shifts being produced by the heuristic and so the heuristic should not be used. The autocorrelation test on the data series, combined with the judgement of the modeller as to the nature of the output variable, should be applied to identify such cases. In this situation, the correlation problem may be overcome by changing the output variable to be the average of successive groups of values (as with the batch means method for steady-state output analysis) to create values with a low inter-dependency. This would slightly blur the shift point, however. After application of the heuristic, the correlations within each of the identified steady-state periods can also be calculated. These should be very low for a shifting steady-state process.

Example: the heuristic applied to a manufacturing line model

The heuristic is now applied to the output of a model with known output characteristics, to test its performance. Obviously, the heuristic would normally be applied to a situation where the output characteristics are not known; such an example, based upon a real model, is provided in the next section. First the simulation model is described followed by the results of the analysis.

The manufacturing line model

The test model represents a manufacturing flow line consisting of 30 operations performed in series. Each machine is connected to the previous operation by an accumulating conveyor with a capacity of five parts, which gives some buffering during machine stoppages. The time for parts to traverse a conveyor is one minute. Parts are loaded onto pallets at the first operation and are removed from their pallet and shipped at the thirtieth operation. The empty pallets then return to the first operation via an accumulating conveyor with a capacity of five pallets. There are a total of 100 pallets in the system.

Data for each operation are given in Table 1. Five products are manufactured on the flow line (known as products A, B, C, D and E). The data in Table 1 show the cycle time for product A. Product B requires 20% more time than product A on each machine, product C 30%, product D 40% and product E 50%. During each work period (8 h) a specific product is made for the whole of that period, although at the start of the work period the products in progress from the previous period still have to be completed. The flow line works continuously without any shift breaks.

Table 1 Data for example model of a manufacturing flow line

Operation	*Cycle time (min): product A*	*Mean time between failure (min)*	*Mean repair time (min)*
Op01	2.1	504.0	40.4
Op02	1.9	590.3	42.2
Op03	2.2	534.1	32.1
Op04	2.1	610.8	28.9
Op05	2.0	629.5	29.6
Op06	2.2	478.1	32.3
Op07	2.3	495.7	31.7
Op08	2.2	618.5	37.9
Op09	1.9	514.4	37.9
Op10	2.3	476.4	30.6
Op11	1.9	442.3	42.1
Op12	2.2	603.0	37.1
Op13	1.7	586.9	30.8
Op14	2.0	465.2	34.7
Op15	2.1	469.2	27.0
Op16	2.2	442.8	21.1
Op17	1.8	631.0	38.5
Op18	1.8	445.6	23.5
Op19	2.2	411.3	43.7
Op20	1.8	547.9	38.4
Op21	2.2	445.1	37.3
Op22	2.2	393.9	31.4
Op23	1.9	484.7	35.1
Op24	2.1	517.0	20.2
Op25	1.9	587.2	33.1
Op26	2.0	426.3	36.1
Op27	1.7	578.3	29.2
Op28	2.3	452.5	32.3
Op29	1.7	568.2	31.2
Op30	2.2	536.8	37.1

Variability is present in the model due to the machine breakdowns. The mean time between failure for each operation is shown in Table 1, as is the mean repair time. A busy time model is used for the time between failure,[3] which is sampled from a negative exponential distribution. The repair time for each operation is sampled from an Erlang distribution with $k = 3$.

Three versions of the model were used for experimentation:

- *Model 1:* product A is made continuously.
- *Model 2:* products are made for 10 work periods each in a rolling sequence of A, B, C, D and E.
- *Model 3:* products are made for 10 work periods each in a pseudo-random sequence as follows: A, C, D, B, A, E, B, D, A, A, D, E, C, A, B.

The first model represents steady-state output, the second a shifting steady state with a repetitive cycle (cyclic shifting steady state), and the third a shifting steady state without a repetitive cycle.

Model results

Each version of the model was run for a period of 151 work periods, the first work period being for initial warm-up, and for five replications. Figure 1 shows the time series of mean throughput by work period averaged across the five replications obtained from each model (excluding the warm-up period). For any stochastic simulation model it is essential that several replications are carried out. The shifting mean heuristic should be applied to the output values averaged across the replications. The averaging process will reduce the random noise in the output data series, and the more replications that are carried out the easier it will be for the heuristic to identify the shifts in mean.

Applying the shifting mean heuristic

As previously explained, it is recommended that the lag-1 autocorrelation is calculated for the data to assess the applicability of the heuristic. Table 2 shows the values for the time-series in Figure 1. What is evident is that model 1 has a very low lag-1 autocorrelation. This is unsurprising since there are no shifts in the mean for this model. The other models have relatively high lag-1 autocorrelations that are at the limits of what is recommended. However, the averaging of the 5 replications increases the signal to noise ratio which also tends to increase the autocorrelations. Indeed, the lag-1 autocorrelations for the individual replications are all much lower, between about 0.2–0.4. Knowledge of the model can also be applied. The nature of this model is that the variability (the breakdowns) varies over a much shorter time scale than the time scale of each output value, indicating that autocorrelation within the steady-state

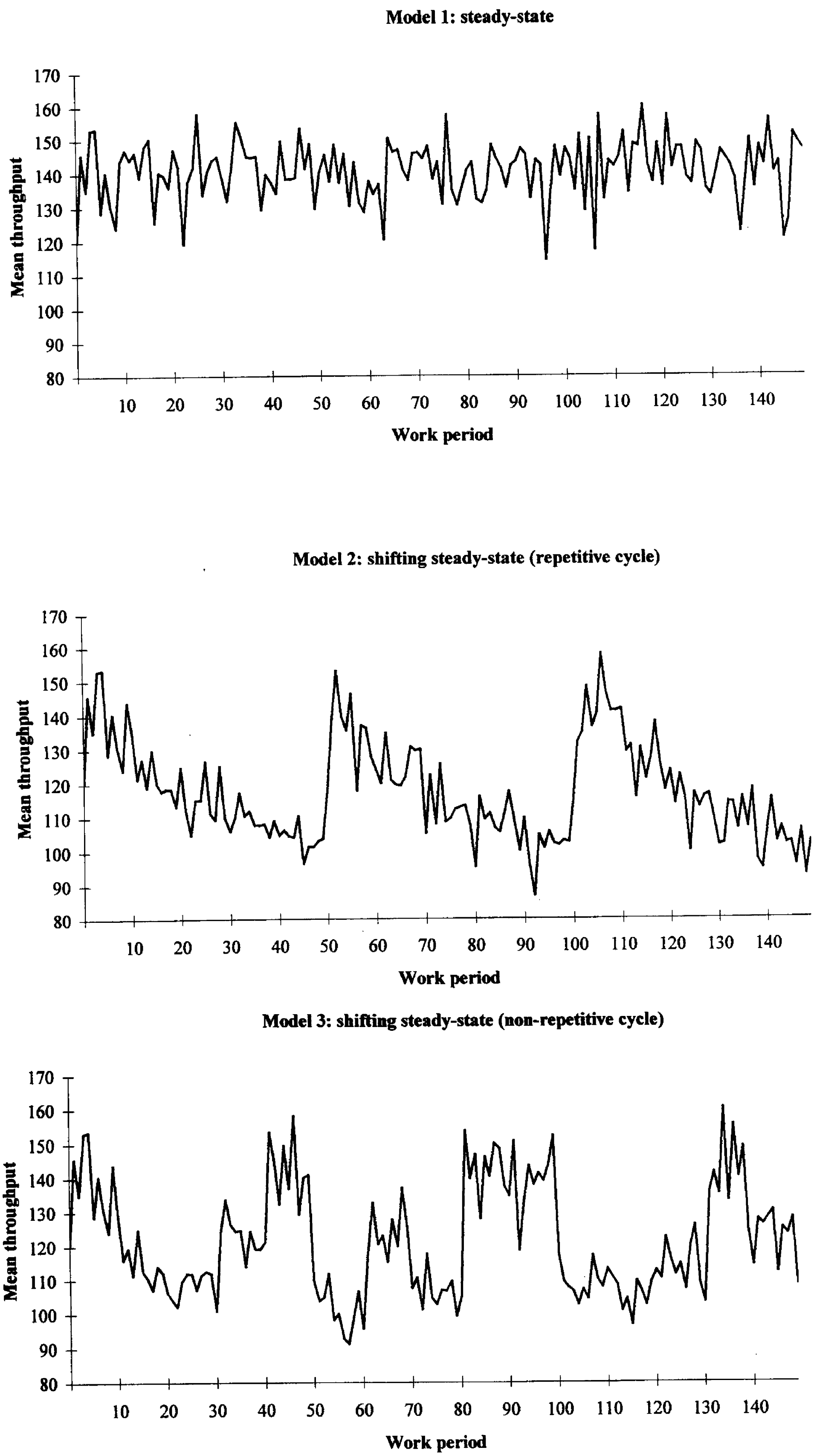


Figure 1 Mean throughput per work period across five replications.

Table 2 Lag-1 autocorrelations for each model

Model	*Lag-1 autocorrelation*
Model 1: steady state	−0.072
Model 2: shifting steady state (repetitive cycle)	0.738
Model 3: shifting steady state (non-repetitive cycle)	0.684

periods would be expected to be low, as required for the use of the heuristic.

Following Lewis and Yeomans,[2] both the standard and the adjusted critical values of the t statistic (next section) are used for assessing the significance of each shift detected by the heuristic. Tables 3–5 show the results obtained from each iteration of the heuristic with the three models. For each iteration the following is reported:

Table 3 Heuristic results for model 1: steady state

Iteration	*Shift point*(n_1)	*Mean 1*	*Mean 2*	*t value*	*Significance level for standard critical value*	*Significance level for adjusted critical value*
1	107	140.13	142.93	−1.797	0.074	0.904
2	125	145.74	140.90	1.841	0.073	0.466
3	56	140.89	139.29	0.958	0.340	3.489
Terminated						

Table 4 Heuristic results for model 2: shifting steady state (repetitive cycle)

Iteration	*Shift point*(n_1)	*Mean 1*	*Mean 2*	*t value*	*Significance level using standard critical value*	*Significance level using adjusted critical value*
1	123	119.77	107.19	4.339	0.000	0.000
2	101	117.00	132.51	−4.925	0.000	0.000
3	70	121.14	107.65	5.064	0.000	0.000
4	15	133.76	117.70	4.493	0.000	0.000
5	51	110.85	130.68	−8.540	0.000	0.000
6	111	141.94	124.65	5.656	0.000	0.000
7	35	114.97	105.70	4.985	0.000	0.000
8	89	110.94	102.43	3.442	0.002	0.009
9	138	111.25	102.12	3.781	0.001	0.004
10	56	142.60	126.43	4.574	0.000	0.001
11	7	139.46	128.78	2.015	0.065	0.234
12	29	116.67	111.00	2.038	0.020	0.057
13	119	127.23	119.50	2.085	0.021	0.201
14	79	113.16	108.94	1.430	0.171	0.705
Terminated						

Table 5 Heuristic results for model 3: shifting steady state (non repetitive cycle)

Iteration	*Shift point* (n_1)	*Mean 1*	*Mean 2*	*t value*	*Significance level using standard critical value*	*Significance level using adjusted critical value*
1	81	119.06	124.64	−2.111	0.036	0.444
2	100	141.51	118.23	6.562	0.000	0.000
3	50	124.76	109.88	4.689	0.000	0.000
4	131	109.96	131.73	−7.621	0.000	0.000
5	41	120.77	142.93	−4.774	0.000	0.000
6	11	136.62	114.95	6.865	0.000	0.000
7	139	144.13	122.71	5.363	0.000	0.000
8	61	101.25	114.62	−3.697	0.001	0.005
9	70	124.44	106.58	6.559	0.000	0.000
10	31	110.87	123.12	−5.818	0.000	0.000
11	120	107.75	113.98	−2.891	0.007	0.039
12	20	114.18	108.16	2.833	0.011	0.047
13	54	107.55	97.66	3.339	0.009	0.026
14	5	141.44	132.60	1.329	0.216	0.649
Terminated						

- the point at which a potential shift is identified
- the value of the mean on either side of the shift
- the significance levels (*P*-values) at which the difference in the means would be accepted as being significant for the two critical values.

A 5% significance level was used as a basis for assessing the significance of the shifts. To avoid premature termination of the heuristic, it was stopped only when the latest shift gave a significance level of greater than 10% for both critical values.

For the steady-state model 1 (Table 3), none of the shifts detected are significant at the 5% level using either critical value. As a result, it is concluded that there are no shifts in the time series of the output of model 1. Note that the significance level for the third iteration is greater than 1 and is clearly outside the meaningful range. Here it is interpreted as meaning the detected shift is absolutely insignificant. The reporting of this value casts some doubt on the meaning and interpretation of the adjusted critical value described by Lewis and Yeomans.[2]

For model 2 (shifting steady state with a repetitive cycle), the first ten iterations detect shifts in the mean that are significant at the 5% level using both critical values (Table 4). At the eleventh iteration, the shift would certainly be rejected using the adjusted critical value, although it is a borderline decision if the standard critical value is used. The following shift is certainly significant using the standard critical value, which may lend some weight to the argument for accepting the shift detected at iteration 11. It is not until iteration 14 that a clearly insignificant shift for both critical values is encountered. Depending on which critical value is adopted, either 13 or 10 significant shifts in the mean are detected.

Table 5 shows the results for model 3, the non-repetitive cycle shifting steady-state model. What is particularly notable here is that using the adjusted critical value no significant shifts in the mean would be detected, since iteration 1 requires a 45% significance level. All other shifts, however, are significant using both critical values through to the fourteenth iteration. Because this time series moves randomly between steady-state periods, it is particularly difficult to identify an initial significant shift. Once the series has been split into smaller components, however, detection of significant shifts becomes easier. This has important implications for the application of the heuristic. Certainly the heuristic should not be stopped as soon as the adjusted critical value is insignificant. The results also raise some doubt over the suitability of the adjusted critical value suggested by Lewis and Yeomans.[2]

Since there appear to be problems with using the adjusted critical value, the standard value has been used as the basis for determining significant shifts. Shift 11 of model 2 has also been accepted even though its significance level slightly exceeds 5%. The effect of using the standard critical value rather than the adjusted one is that four extra shifts are accepted, namely numbers 11, 12 and 13 for model 2, and number 1 for model 3.

Figure 2 shows 'Manhattan' diagrams of the accepted shifts for each of the time series. These show the means of the actual data between the shift points. What is immediately apparent is that, correctly, no shifts in the mean are detected for model 1; the series purely represents variation around a constant mean, or steady-state. However, for the other series significant shifts in the mean of the data are found.

Comparison between actual and detected shifts in the mean

In order to determine whether the heuristic is correctly detecting the shifts in the mean, a comparison is made with the actual shifts that occur in the output data. In this case, it is assumed that these are when there are shifts in the input data, that is, when the product type changes. The first data point after each shift, which gives the throughput for the first eight hours after the product change, will include some output of the previous product but this effect will be small and has been ignored.

In Figure 3 Manhattan diagrams are drawn for both the actual and detected shifts for models 2 and 3. The values shown for both series are the means of the actual output data between the shift points. Obviously there is no need to consider the match for model 1 where no shifts are found. On inspection these diagrams suggest that there is a reasonable match between the actual and detected shifts. Indeed, the average absolute difference between the mean values for the actual and detected shifts across the 150 observations is 3.44 for model 2 and 2.76 for model 3, representing a mean error of around 2.5%.

Inspecting the diagrams in more detail reveals some further issues:

- Some shifts are not detected, for instance, periods 70–89 in model 2 and periods 100–119 in model 3 are both detected as only one mean. These are both shifts where there is only a very small change in the mean and it is therefore not surprising that the heuristic does not detect them. Interestingly, the heuristic does detect a shift in the mean at observation 79 for model 2 at iteration 14 (Table 4), but it is concluded that this shift is not significant.
- When the product changes in the schedule there is often a lag of one period in the detection of a shift in the mean. This is particularly notable in model 3, although it is also apparent in model 2. When the schedule changes product there is a transition period during which the products already in the system feed out so that the first value will be a mixture of the throughput of two products.
- In model 3 one shift (at period 54) is detected which does not genuinely exist. This is probably the result of a mixture of the transient period at the product changeover (period 50–51) and random noise in the output process. It

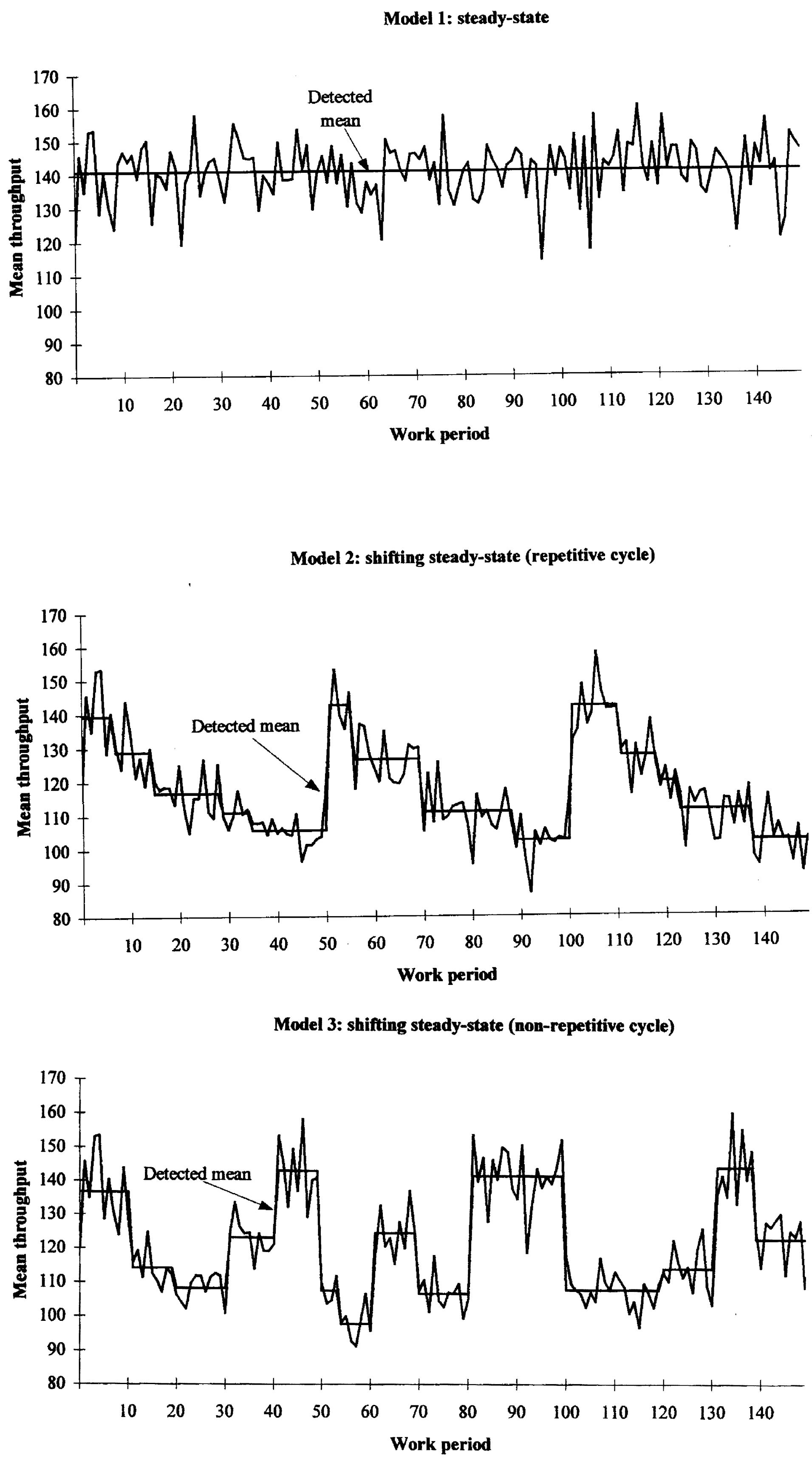
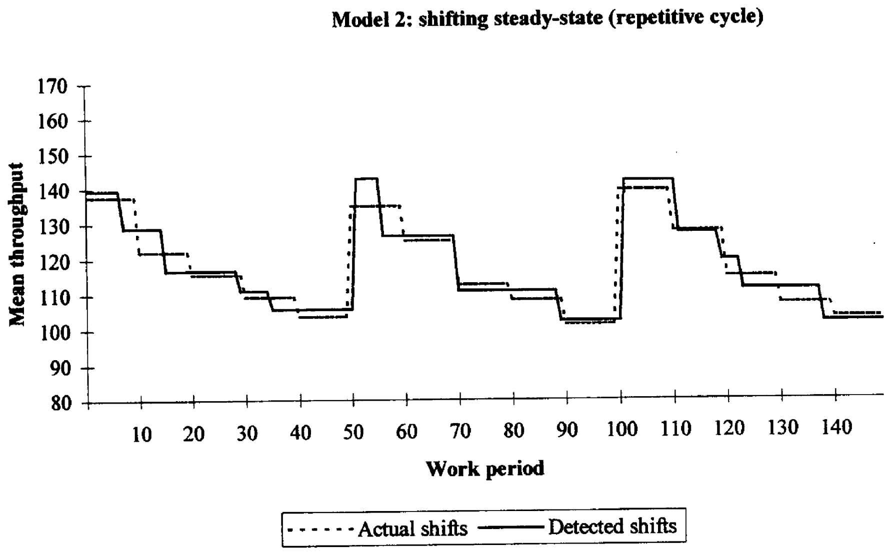


Figure 2 Detected shifts in the time series mean.



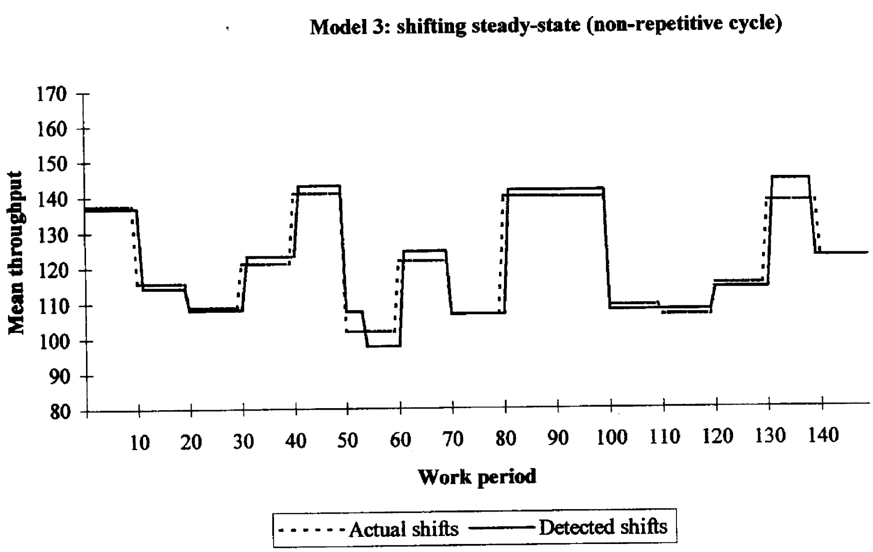


Figure 3 Actual versus detected shifts in the mean.

is notable that this is the last significant shift found for model 3.

- The heuristic performs better with model 3 than model 2, made evident by the greater mismatch between actual and detected shifts in the mean. This can be seen by visual inspection of Figure 3 and the average absolute difference between the actual and detected values. The smaller step changes in model 2 obviously make accurate detection more difficult.

Identifying output characteristics

Visual inspection of Figure 3 certainly suggests that a cycle exists in model 2, but it is less clear whether any cycle exists within model 3. The existence of a cycle depends on the time series returning to the same steady-state output distributions on a cyclical basis. Even if a cycle does not exist, it is useful to identify steady-state periods in which conditions appear to be the same. Therefore, as a next step, a comparison of the detected means is made in order to determine which are similar. Following this, Gantt charts are constructed showing the periods during which the similar means are in operation.

Grouping means

In order to determine which detected means are similar, the Duncan *post hoc* comparison method is applied.[13] This is selected because of its power. As for the heuristic, it assumes independence of the individual values and constant variance of the values from the means.

The results of the *post hoc* test, performed at a significance level of 5% for the subsets, are shown in Table 6. The first column, 'Mean number', refers to the original sequence of the detected means in the time series, ie the means shown on the Manhattan diagrams in order. There is some overlap between the groups of means that are detected. In order to unravel these overlapping groups, the most significant groups are selected first. The residual means are then selected from the other groups, with preference given to the groups with the highest significance. The resulting groups of means are shown in Table 7 along with the product that is actually being made in the model for the majority of the period. If two products are each made for approximately half the period both are shown. For both models clear groupings in the data are found.

For the most part the test does group together the periods in which the same product is actually being made. Periods in which two products are made for about equal time could legitimately be included in either of two groupings. Therefore, the only differences from the actual model behaviour are for model 3 where a period in which product E is made is included in group 2 and the periods in which product A is made are split into two groups. The group 2 period during which E is made is the short time from work period 50 to work period 53 that probably would be treated as transient and ignored.

Inspection of cycles: Gantt charts

As a final step Gantt charts are constructed showing the periods during which similar means are in operation (Figure

Table 6 *Post hoc* comparison of detected means (Duncan's test)

Model 2: shifting steady state (repetitive cycle)

Mean number	*Sample size (n)*	*Means in group 1*	*Means in group 2*	*Means in group 3*	*Means in group 4*	*Means in group 5*	*Means in group 6*
14	12	102.12					
9	12	102.43					
5	16	105.70	105.70				
8	19		110.94	110.94			
4	6		111.00	111.00			
13	15		111.25	111.25			
3	14			116.67	116.67		
12	4				119.50		
7	14					126.43	
11	8					127.23	
2	8					128.78	
1	7						139.46
10	10						141.94
6	5						142.60
Sig.		0.30	0.12	0.11	0.38	0.50	0.36

Model 3: shifting steady state (non-repetitive cycle)

Mean number	*Sample size (n)*	*Means in group 1*	*Means in group 2*	*Means in group 3*	*Means in group 4*	*Means in group 5*	*Means in group 6*
7	7	97.66					
9	11		106.58				
6	4		107.55	107.55			
11	20		107.75	107.75			
3	11		108.16	108.16			
12	11			113.98			
2	9			114.18			
14	11				122.71		
4	10				123.12		
8	9				124.44		
1	11					136.62	
10	19					141.51	141.51
5	9					142.93	142.93
13	8						144.13
Sig.		1.00	0.68	0.08	0.63	0.08	0.47

Table 7 Groups of means selected from *post hoc* analysis along with the products actually being made for the majority of the period

	Means for Model 2		*Means for Model 3*	
Group	*Detected*	*Product*	*Detected*	*Product*
1	102.12	E	97.66	E
	102.43	E		
	105.70	E		
2	110.94	C/D	106.58	D
	111.00	D	107.55	E
	111.25	C/D	107.75	D/E
			108.16	D
3	116.67	C	113.98	C
	119.50	C	114.18	C
4	126.43	B	122.71	B
	127.23	B	123.12	B
	128.78	B	124.44	B
5	139.46	A	136.62	A
	141.94	A		
	142.60	A		
6			141.51	A
			142.93	A
			144.13	A

4). The overall mean of the groups is used for the y values in these charts.

Inspection of these charts fairly clearly shows a cycle of 50 work periods for model 2 with 5 distinct shifts in the mean. The only disruption to this is in the second cycle, where the third mean is not detected. It does not seem unreasonable to conclude that this is a cyclic shifting steady-state model. There is no suggestion that a cycle exists for model 3.

Example: real model of a bottling plant

The heuristic is now applied to a real model with unknown output characteristics. The aim is to provide an illustration of the heuristic in practice and to provide an example of a situation in which there is a shifting steady state. As such, the description below provides a brief outline of the problem situation and the model, as well as a summary of the results obtained from using the heuristic.

The model simulates a soft drinks bottling plant in the UK; the name of the organisation is not given for reasons of

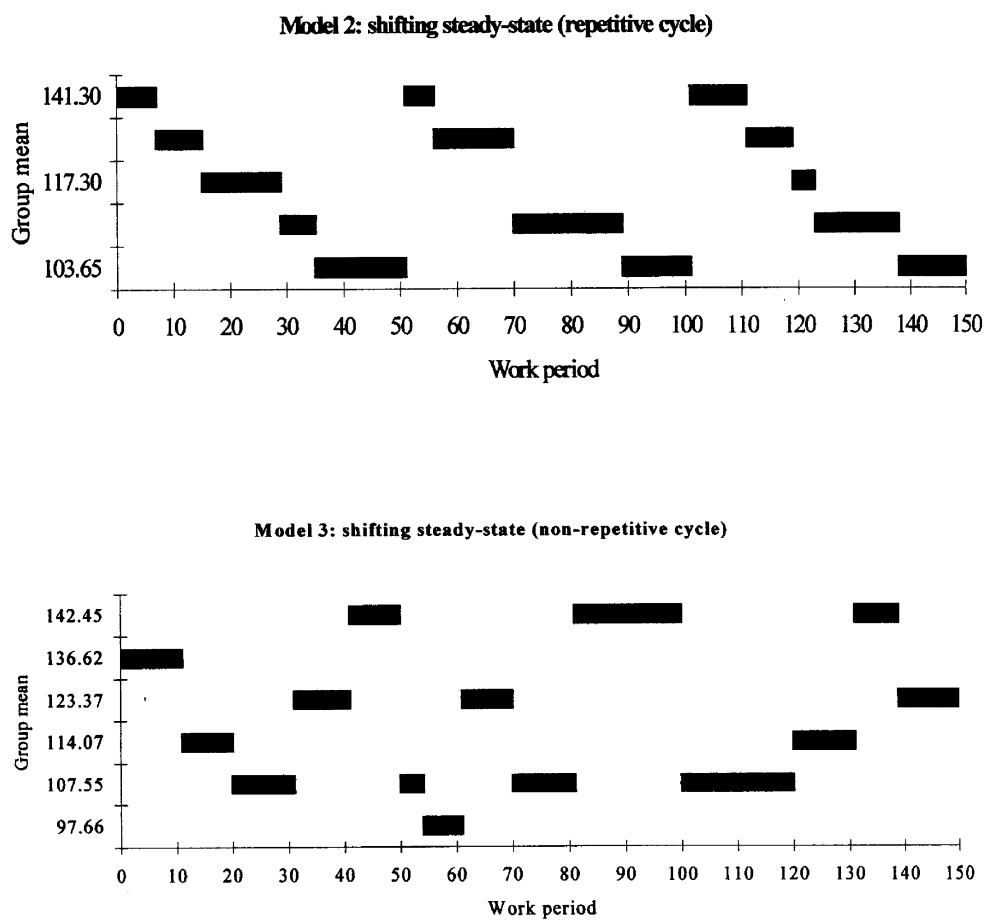


Figure 4 Gantt charts showing pattern of similar means.

confidentiality. The problem faced by the organisation was the loss of production that was largely caused by the frequent long breakdowns of the equipment. The plant consists of five lines, two dedicated to carbonates and three to cordials. Each line consists of five machines (bottle blower, filler, labeller, boxer and palletiser) connected by conveyors. Although the lines are similar in layout, the equipment on each line is not identical. Machines are of different ages and some require greater levels of manual intervention. As a result, the five lines all run at different speeds and the incidence of machine failures varies. Machine failures occur every one to eight days on average. The mean repair time is anything up to eight hours.

Since the plant consists of five lines, each working at different speeds, and each line is effectively on or off (depending on whether a machine is broken down), the overall plant output will change from one state to another, depending on the combination of bottling lines that are in operation. Indeed, with five lines, each in one of two states (on or off), there are 32 (ie 2^5) possible states. Because the machine failures occur at random, the shifts in state also occur at random with a random duration. Using the earlier definition, the model could be described as a shifting steady-state model with a non-repetitive cycle.

One approach to analysing this problem would be to model the five lines separately, although this in itself would involve shifts in the steady state between full and zero production. In order to obtain a full understanding of the bottling plant's operation it is useful, however, to analyse the performance of the complete plant. It is important for the management of such a plant to understand the extent of variability in order to better manage production and customer expectations. There is also some sharing of resources between the five bottling lines which could not be accounted for if separate models were developed.

Figure 5 shows the results from a single run of the bottling line model. The time series records hourly throughput (in pallets of bottles) over 200 h following a suitable warm-up period. Multiple replications have not been performed since the shifts in steady state are random and changing the random number streams would change the position of the shifts. Averaging the results across a series of replications would lose important information about the variability in the system.

It is apparent from an inspection of the time series that the model is not in a permanent steady state, but that it is shifting between various levels of throughput. The shifting mean heuristic has been applied to the time series and the resulting Manhattan diagram is included in Figure 5. The heuristic has been terminated after 21 iterations, after which point most shifts are deemed insignificant. Applying Duncan's *post hoc* comparison method suggests that there are six groups of means in the data. Although this is much less than the 32 predicted it is unsurprising since there are unlikely to be significant differences between all the combinations of throughput. Further to this, the simulation run is relatively short and, as a result, it is unlikely that all 32 combinations have occurred during the period of the run.

Three further issues arise in applying the heuristic in this circumstance. First, because only one replication has been performed the data are very noisy which will inevitably reduce the performance of the heuristic. Second, the time series will include periods of warm-up as the model shifts from one state to another. In other words, it is unlikely that a line is either on or off, but that it takes a period of time to reduce and restore the level of throughput on either side of a machine failure. Third, because machine failures occur at any period in the hour the time-series includes part periods of production for the bottling lines. This will further reduce the performance of the heuristic in identifying the various steady-states. That said, a simple inspection of Figure 5 suggests that the heuristic has provided a reasonable analysis of the time-series data.

Concluding discussion

In this paper the Lewis and Yeomans'[2] shifting mean heuristic is used to identify the timing of shifts from one

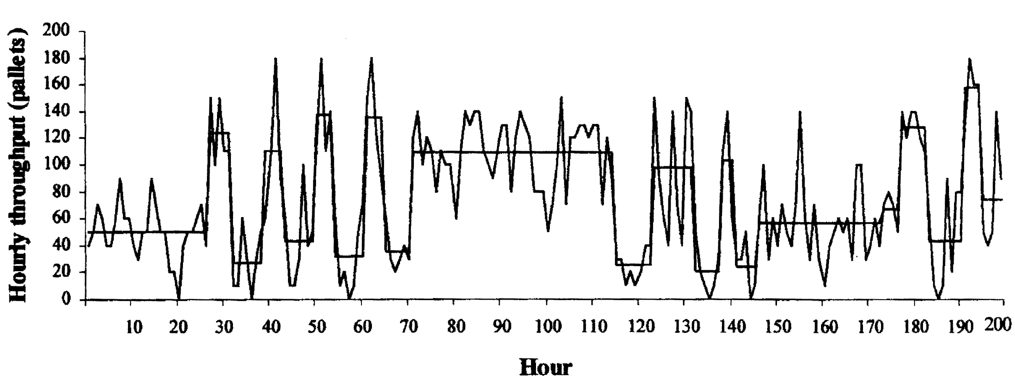


Figure 5 Application of shifting mean heuristic to bottling plant model.

steady state to the next in the output from a simulation model that passes through successive steady-state periods. A *post hoc* comparison method is then used to group similar periods together and test whether the model exhibits cyclic behaviour. The technique provides an additional tool for analysing simulation output that complements existing techniques since it addresses a shifting steady-state model rather than the transient or single steady-state models dealt with by most other methods.

It is important to identify the behaviour of a model in order to apply the correct experimental methodology and output analysis techniques. A model that demonstrates cyclic shifting steady-state behaviour should be run for a multiple of the cycle length, otherwise some bias is introduced into the results. Having identified the cycles, the output could be analysed by identifying the various steady states within the model. Confidence intervals could then be calculated for each steady state using the means for the separate periods.

A general shifting steady-state model without a cycle needs to be considered as transient from the point of view of an output averaged over the length of the run, in that a longer run length changes the overall process mean. However, within the time series of the output data specific steady state means do exist that can be analysed separately. Again, if the steady states are repeated the means can be used for calculating confidence intervals potentially reducing the number of runs required.

In addition to the ability to apply steady-state analysis techniques, the identification of this type of behaviour also gives an important insight into the nature of the model, which should greatly increase the understanding of the system under investigation. The understanding gained is usually a vital part of the overall benefits of the modelling exercise.[14] This should help in the further analysis of the system, including the development of intervention strategies. For example, certain aspects of the model's behaviour may be of particular interest, such as when the system is out of a specified tolerance, and this may only happen within a specific steady-state period. Further examination could then focus on the analysis of the conditions occurring in that period. Similarly, it is likely that the input variables that vary with quite a different pattern to the output have little effect on the variable of interest and so can be ignored.

There are, however, some limitations to the shifting mean heuristic approach described in this paper. Consideration should be given as to whether the assumptions of the heuristic are likely to apply. The heuristic only detects shifts in the means of the time-series data; it is possible that the underlying output distribution is also changing. It cannot, of course, be guaranteed that the heuristic will detect all shifts in the mean of a time series because of the effect of the noise in the data, although the noise can be reduced by increasing the number of replications. It may also detect shifts that are not present, particularly in the transition periods from one steady state to the next. Even if all the shifts are not correctly identified the heuristic can still give a valuable insight into the behaviour of the model. The results have also indicated that the adjusted t-test critical value suggested by Lewis and Yeomans[2] for the significance of the shift may need to be modified. Perhaps a different approach may perform better, such as a test for the statistical significance of the CUSUM value itself

In this paper a model with known output characteristics is used to determine whether the shifting mean heuristic is effective. The heuristic performs well for this model but, as with any heuristic, further usage will improve the strength of the evaluation. The heuristic is also applied to a real model with unknown output characteristics for which its performance appears to be satisfactory. The heuristic needs to be tested further on real models with unknown behaviours, both to test its performance and to investigate the types of output behaviours that real simulations display.

Acknowledgement—We are grateful for the helpful comments of the anonymous reviewers.

This paper is dedicated to Colin Lewis (3 June 1938 to 24 December 2001) who supported me and got me started in my academic career—Stewart Robinson.

References

1 Welch PD (1983). The statistical analysis of simulation results. In: Lavenberg S (ed). *The Computer Performance Modeling Handbook.* Academic Press: New York, pp 268–328.
2 Lewis CD and Yeomans KA (1995). A heuristic search procedure for detecting sudden shifts in stationary time series data. *J Opl Res Soc* **46**: 1471–1480.
3 Law A and Kelton WD (2000). *Simulation Modeling and Analysis*, 3rd edn. McGraw-Hill: New York.
4 Goldsman D and Schruben L (1990). New confidence interval estimators using standardized time series. *Mngt Sci* **36**: 393–397.
5 Kleijnen JPC (1987). *Statistical Tools for Simulation Practitioners.* Marcel Dekker: New York.
6 Robinson S (1995). An heuristic technique for selecting the run-length of non-terminating steady-state simulations. *Simulation* **65**: 170–179.
7 Kelton WD (1997). Statistical analysis of simulation output. In: Andradottir S, Healy KJ, Withers DH and Nelson BL (eds). *Proceedings of the 1997 Winter Simulation Conference* Association of Computing Machinery: New York, pp 23–30.
8 Friedman LW (1986). The analysis of multiple response simulation output data—experiments of comparison. *Comput Opns Res* **13**: 647–652.
9 Balmer DW (1987). Polishing the analysis of the statistical output of comparative simulation experiments. *Simulation* **49**: 123–126.
10 Gleick J (1987). *Chaos: Making a New Science.* Abacus: London.
11 Billingsley P (1995). *Probability and Measure*, 3rd edn. Wiley: New York.
12 BS 5703 Parts 1 and 2 (1980). *Data Analysis and Quality Control Using CUSUM Techniques.* British Standards Institution: London.

13 Winer BJ (1971). *Statistical Principles in Experimental Design*. McGraw-Hill: New York.
14 Brooks RJ and Tobias AM (1996). Choosing the best model: level of detail, complexity and model performance. *Math Comput Model* **24**: 1–14.

*Received January 2001;
accepted November 2001 after two revisions*