Department of Computer Science
University of Canterbury
Christchurch , New Zealand

STEADY STATE SIMULATION OF QUEUEING PROCESSES : A Survey of Problems and Solutions

K. Pawlikowski

TECHNICAL REPORT COSC 01/88

Section 3 of this report includes a revised version of Technical Report COSC 04/87

STEADY STATE SIMULATION OF QUEUEING PROCESSES : A Survey of Problems and Solutions .

K. Pawlikowski

SUMMARY . For years computer-based stochastic simulation has been a commonly used tool in the performance evaluation of various systems. Unfortunately results of some studies can be given little credibility, since they are presented neglecting their random nature. On the other hand, a variety of techniques for collection and statistical analysis of simulation data have been proposed that vary in efficiency and sometimes required a deeper knowledge of the statistical methods involved. Practitioners still await a possibly general automatic technique that could be applied in stochastic simulation by users having little knowledge or interest in the statistical analysis of output data.

In this report we discuss the main factors that can affect the accuracy of stochastic simulations designed to give insight into the steady state behaviour of queueing processes. The most promising methods for achieving suitably accurate simulation results are also surveyed.

Key Words: stochastic discrete-event simulation, steady-state simulation, statistical analysis of simulation output data, length of initial transient period, warm-up period in stochastic simulation, simulation output data collection and analysis, sequential analysis of confidence intervals .

1. INTRODUCTION

Computer-based stochastic simulation, traditionally regarded as a 'last resort' tool (if analytical methods fail), has become a valid and probably one of the most commonly used methods of performance evaluation. This popularity is due to the continuing development of more powerful and cheaper computers, as well as significant achievements in software engineering. One can observe a trend towards integrating simulation methodology with concepts and methods of artificial intelligence; see for example [ARTI88]. Various user-friendly simulation packages offer visual interactive capabilities; traditional discrete-event simulation modeling is more and more frequently supported by object- and logic-oriented programming; see for example [BELL87], [NIEL86], [OKEE86], [OREN87], [REDD87], [RUIZ87] and [STAI88] . All these developments offer users increasingly powerful and versatile techniques for performance evaluation, leading towards automatic, knowledge-based simulation packages. Simulation programming techniques and languages are discussed in numerous publications, including textbooks by Bulgren [BULG82], Kreutzer [KREU86], Law and Kelton [LAWK82], and Payne [PAYN82].

Regardless of how advanced the programming methodology applied to simulation modeling is, experiments in which events are controlled by random numbers produce results that are nothing more than statistical samples. Despite this fact, from time to time various simulation studies have been reported primarly as programming exercises. Their authors, after putting much intellectual effort and time into building models, and then writing and running programs, have very little or no interest in a proper analysis of the simulation results. It is true that "*...the purpose of modeling is insight, not numbers...*" (Hamming, [HAMM62]) but proper insight can only be obtained from correctly analyzed numbers. In the stochastic simulation of queueing systems "*...computer runs yield a mass of data but this mass may turn into a mess...*", and then "*... instead of an expensive simulation model, a toss of the coin had better be used...*" (Kleijnen, [KLEI79]), if the random nature of the results is ignored. Statistical inference is an absolute necessity in the situation where the same (correct) program produces different (but correct) output data from each run.# Simulation applied to modelling and performance analysis has been compared to the surgical scalpel, [SHNN81], that *..."in the right hand can accomplish tremendous good, but it must be used with great care and by someone who knows what they are doing."*

The simplest objective of simulation studies is the estimation of the mean of an observed process, which is done by calculating the average of an ensemble of its observations $x_1, x_2, ..., x_n$, i.e., by

$$\bar{X}(n)=\sum_{i=1}^{n}\frac{x_i}{n} \qquad (1)$$

In an example given in [LAW83] , see also [LAWK82, p.287], estimates of the mean delay of the first 25 customers, from simulation studies of a M/M/1 queue, range from 0.4 to 6.4 , whilst the exact theoretical mean value of this delay is 2.1 (utilization 90%, the system initially empty). For another illustration of randomness of simulation results see e.g. [WELC83, Sec.6.1].

The accuracy of the estimator $\bar{X}(n)$ of an unknown average μ_x can be assessed by the probability

$$P\left(|\bar{X}(n) - \mu_x| < \Delta_x \right) = 1 - \alpha \ , \qquad (2)$$

or, equivalently, by

$$P\left(\mu_x - \Delta_x \le \bar{X}(n) \le \mu_x + \Delta_x \right) = 1 - \alpha \ , \qquad (2a)$$

where Δ_x is *the half-width of the confidence interval* for the estimator and $(1-\alpha)$ is the confidence level, $0<\alpha<1$. Thus, if the width $2\Delta_x$ of the confidence interval is found for an assumed confidence level of $(1-\alpha)$, then roughly speaking, if the simulation experiment were repeated a number of times, the estimate $\bar{X}(n)$ would fall in the interval $(\mu_x - \Delta_x, \mu_x + \Delta_x)$ in $100(1-\alpha)\%$ of cases, and in $100\alpha\%$ of cases it would not. It is well known that if observations $x_1, x_2, ..., x_n$ can be regarded as realizations of independent and normally distributed random variables, then the $100(1-\alpha)\%$ confidence interval for an unknown mean μ_x is given by

$$\bar{X}(n) \pm t_{n-1,1-\alpha/2}\, \hat{\sigma}[\bar{X}(n)] \ , \qquad (3)$$

where

$$\hat{\sigma}^2[\bar{X}(n)] = \sum_{i=1}^{n} \{x_i - \bar{X}(n)\}^2 / n(n-1) \qquad (4)$$

is the (unbiased) estimator[¥] of the variance of $\bar{X}(n)$, and $t_{n-1,1-\alpha/2}$, for $0<\alpha<1$, is the upper $(1-\alpha/2)$ critical point obtained from the t-distribution with $(n-1)$ degrees of freedom[#]. The formula (3) can be also applied if the observations are realizations of *independent and identically distributed* (i.i.d.) *random variables* , since then, according to the central limit theorem (see e.g. [TRIV82, Sec.4.7]), the distribution of the variable $\bar{X}(n)$ tends to the normal distribution as the number of collected observations, n, increases. In practice the formula (3) gives a good approximation for $n>100$. Results obtained from Eqs. (1) and (3) are called *point* and *interval estimates*, respectively. Both of them are very important: the former characterizes the system analyzed, while the later states the accuracy of these characteristics.

If the assumption of independent and identical distributions for the observations x_1, x_2, ..., x_n is invalid then we have to consider some modifications to the above estimators. This raises the problem of measuring the quality of estimators. There are three common measures of estimator effectiveness :

- *the bias* , which measures the systematic deviation of the estimator from the true value of the estimated parameter, for example in the case of $\bar{X}(n)$:

¥ Estimators are distinguished in the text from the parameters they estimate by placing a ^ over the parameter's symbol.

For $n>30$ the t-distribution can be replaced by the standard normal distribution, and $t_{n-1,1-\alpha/2}$ by $z_{1-\alpha/2}$, which is the upper $(1-\alpha/2)$ critical point obtained from the standard normal distribution .

$$\text{Bias}\,[\bar{X}(n)] = E[\bar{X}(n) - \mu_x] \qquad (5)$$

- *the variance* which measures the mean (squared) deviation of the estimator from its mean value, i.e.,

$$\sigma^2[\bar{X}(n)] = E\{\bar{X}(n) - E[\bar{X}(n)]\}^2 \qquad (6)$$

and
- *the mean square error* (MSE) of the estimator, defined as

$$MSE[\bar{X}(n)] = E\{[\bar{X}(n) - \mu_x]^2\} \qquad (7)$$

Note that from these definitions

$$MSE[\bar{X}(n)] = \{\text{Bias}\,[\bar{X}(n)]\}^2 + \sigma^2[\bar{X}(n)] \qquad (8)$$

The main analytical problem encountered in the analysis of simulation results is that they are usually highly correlated, and thus do not satisfy the precondition of statistical independence. If the sequence of observations $x_1, x_2, ..., x_n$ is considered to be autocorrelated, and stationary, then the variance of $\bar{X}(n)$ is given by the following formula:

$$\sigma^2[\bar{X}(n)] = \left[R(0) + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) R(k) \right] / n \qquad (9)$$

where

$$R(k) = E\,[(x_i - \mu_x)(x_{i-k} - \mu_x)] \qquad , 0 \le k \le n-1, \qquad (10)$$

is the autocovariance of order k (the lag k component of the autocorrelation function# {R(k)}) of the original observations. Note that Eq.(9) can be reduced to R(0)/n if and only if the observations are uncorrelated. Neglecting the statistical correlation of the observed variables $x_1, x_2, ..., x_n$ is equivalent to removing all the components except R(0) from Eq.(9). Such an approximation is usually unacceptable. For example, in an M/M/1 queueing system with 90% utilization , the variance of the mean queue length calculated according to Eq. (9) is 367 times greater than that from Eq. (4), [BLOM67]; see [LAWK82, p.146] for another example. Any variance analysis disregarding correlations among the observations would lead to either an excessively optimistic confidence interval for μ_x , in case of positively correlated observations, or to an excessively pessimistic confidence interval for μ_x , in the case of negatively correlated observations, see Eq. (3). A positive correlation between observations is typical in simple queueing systems without feedback connections, and it is stronger for a higher system utilization; see e.g. [DALE68] for correlation analysis of the M/M/1 queue.

Generally the variance analysis of correlated processes , and the analysis of their autocorrelation functions in particular, is a complex statistical problem and therefore creates a major problem in the statistical analysis of simulation output data. In *terminating* (or *finite-horizon*) *simulation*, used for studying the behaviour of systems

Independence from the index i in Eq.(10) is due to the assumed stationarity of the analysed processes.

during specified intervals of time, the above problem can be overcome simply by making a number of independent replications of the simulation experiment, since then the means of individual observations collected during different simulation runs can be regarded as a sequence of independent (secondary) output data, and Eq.(4) can be applied. Exhaustive discussions on the statistical analysis of output data from terminating simulation can be found for example in [KLEI79], [LAW80], and [LAWK82, Sec.8.5].

In this report we discuss *steady state (infinite horizon) simulation,* aimed to give insight into the behaviour of queueing processes after a (very) long period of time. The methodology for this kind of simulation study is much more complicated. After 'launching" a queueing process is initially in a nonstationary phase (warm-up period), and then, if the process is stable, it moves asymptotically towards a steady state (statistical equilibrium). Since observations gathered during the initial transient period do not characterize the steady state, a natural idea is to discard all such observations before further analysis. This requires an estimation of the effective length of the initial transient period. Ignoring the existence of this period can lead to a significant bias of the final results. On the other hand, the removal of any observations increases the variance of estimates, which in turn can increase the value of the mean-square error; see [FISH72], [TURN77], [WILS78] and [DONN81]. Thus a decision whether to delete or not to delete initial observations depends on the assumed criterion of goodness of the estimators . This also affects methods used to collect observations, which are discussed in the next section. These and other aspects of *the problem of initialization* are presented broadly in Section 3.

A number of analytical techniques has been proposed to overcome the theoretical problems which arise from the correlated nature of observations collected during steady state simulation. They are distinguished by the way they estimate the variance of observed processes, which is needed for determining the width of the confidence intervals. Usually they impose special requirements on how the output data from simulation experiments should be collected and preprocessed. All these *methods of data collection and analysis* proposed for steady state simulation can be divided into two groups:

- methods which attempt to weaken or even remove statistical dependencies among observations, and
- methods which attempt to take into consideration the actual correlations among observations.

The former group is represented by:

- the method of replications,
- the method of batch means, and
- the regenerative method,

while the latter one contains:

- the method based on spectral analysis,
- the method based on standardized time series, and
- the method based on autoregressive model.

Any statistical analysis of simulation output data involves approximations that bias the final estimates. The robustness of the methods listed above is usually measured by the *coverage of confidence intervals*, defined as the frequency with which confidence intervals contain the true parameter, at a particular confidence level, [SCRU80]. Thus coverage analysis can be applied only to systems with theoretically well known behaviour. After a series of independent simulation runs one can determine the fraction of experiments which had confidence intervals covering the true mean value of the estimated parameter. In more detailed studies one can estimate confidence intervals for the coverage (see e.g. [LAVE77]) and conclude that a given method of data collection and analysis produces valid $100(1-\alpha)\%$ confidence intervals (for, say, the mean delay) if the upper bound of the confidence intervals for the coverage is at least $(1-\alpha)$. Otherwise, confidence intervals for the estimated parameter should be regarded as invalid and the method should be assessed as inaccurate (at least for a given class of simulated processes). A few additional measures for the effectiveness of methods used for data collection and analysis were proposed in [SCRI81]. The weakest point of such analyses is that there is no theoretical basis for extrapolating results found for simple, analytically tractable systems to more complex systems, which are the real subjects of simulation studies, [FOX78]. However, ... a better approach has not been proposed yet.

Even if we were able to collect independent and identically distributed output data from simulation runs, we cannot be fully protected from erroneous conclusions because of inherent variations of simulation output data caused by the pseudo-random nature of input data; see [PIDD84, Sec.8.4.2] for a more detailed discussion. To neutralize this undesired effect we can use *a variance reduction technique* (VRT), which can reduce the variance of recorded results without affecting their mean value; see for example [KLEI74, Ch.3], [LAWK82, Ch.11], [BRAT83, Sec.2] and [FROS88] for a survey of variance reduction techniques. Unfortunately, despite the fact that these techniques have been extensively studied theoretically, they have found limited practical application, as they are either strongly model-dependent or difficult to implement in simulation studies of even moderately complex systems. The simplest VRT in the context of queueing processes consists in direct estimation of the mean time-in-queue (without service time), since this estimator has the smallest variance, see [CARS80]. Other parameters, like the mean time-in-the-system or the mean queue length, can be obtained indirectly by adding the assumed mean service time, or from Little's formula, when this estimate is known.#

The methods for data collection and analysis are surveyed in Section 2. Note that any of these methods can be used either in its *fixed-sample-size version* or in its *sequential version*. In the former case, statistical analysis is performed once, at the end of the simulation experiment, when a predetermined number of observations, assumed to be sufficient to get results of a required accuracy, has been collected. The present survey of methods for data collection and analysis, used in steady state simulation, focusses on their sequential versions, in which the length of simulation is increased sequentially, from one checkpoint to the next , until a prespecified accuracy of the point estimators is obtained. Such procedures, automatically controlling the length of simulation experiments, are very desirable in user friendly simulation

The efficiency of this VRT has been proved for G/G/c queueing systems, see [CARS80], but we can expect that it is efficient for more complex queueing systems as well. As much as a 90% variance reduction due to this technique has been reported, although generally the reduction tends to 0 as the system utilization tends to 100%.

packages. Sequential statistical analysis is also more efficient , since usually it is very difficult to determine *a priori* the length of simulation needed by a fixed-size procedure that would be sufficient to obtain a required width of confidence intervals at the assumed level of confidence; c.f. [LAW82] and [LAWK84].

The criterion for stopping the simulation is usually *the relative (half) width of the confidence interval* at a given confidence level (1-α) , i.e., the ratio

$$\varepsilon = \Delta_x / \bar{X}(n) , \qquad (11)$$

$0<\varepsilon<1$; c.f. Eq. (2). This is also called *the relative precision of the confidence interval.* The simulation experiment is stopped at the first checkpoint for which $\varepsilon \le \varepsilon_{max}$, where ε_{max} is the required limit relative precision of the results at the 100(1-α)% confidence level, $0< \varepsilon_{max} < 1$. Note that if

$$1- \alpha \le P [| \bar{X}(n) - \mu_x | \le \varepsilon | \bar{X}(n) |] , \qquad (12)$$

then, for $\mu_x \neq 0$,

$$P [| \bar{X}(n) - \mu_x | \le \varepsilon | \bar{X}(n) |] = P[| \bar{X}(n) - \mu_x | \le \varepsilon | \bar{X}(n) - \mu_x + \mu_x |] \le$$
$$\le P[| \bar{X}(n) - \mu_x | \le \varepsilon | \bar{X}(n) - \mu_x | + \varepsilon | \mu_x |]$$

and finally

$$1- \alpha \le P[| \bar{X}(n) - \mu_x | \le \varepsilon | \bar{X}(n) |] \le P[(| \bar{X}(n) - \mu_x | / | \mu_x |) \le \varepsilon /(1-\varepsilon)] \qquad (13)$$

where

$$| \bar{X}(n) - \mu_x | / | \mu_x | \qquad (14)$$

is called *the relative error of the confidence interval* .

Sequential reasoning about the statistical accuracy of results is particularly advisable if higher precision is required, although there is a danger that when precision requirements are increased the resulting confidence intervals have a greater chance of not containing the true value of parameters (*the narrower the confidence interval, the worse the coverage* effect). One can expect, too, that lower coverage is more probable in the case of negatively correlated observations, for which it is more likely to get an underestimated value of their variance. Higher accuracy requirements can also unacceptably lengthen simulation runs controlled by a sequential procedure. In this context any variance reduction technique can be regarded as a technique for speeding-up the simulation, since any decrease in the value of the variance decreases the width of the resulting confidence intervals, and a specified accuracy can be met more quickly.

The sequential approach is regarded by some as the only alternative to steady state simulation, see [BRAT83, p.101]. Many consider that it is possible to devise a procedure that would fully automatically conduct data collection and analysis, using a

sequential rule for assessing the accuracy of estimates; however, a fully acceptable solution has not yet been invented. Relatively few simulation packages offer some degree of automation of statistical analysis, c.f. [CATA87]. For example, sequential procedures, automated to some extent and based on independent replications, regenerative and spectral methods of data collection and analysis, are implemented in RESQ (Research Queueing Package), [SAUE82] and [SAUE84], and its network-oriented extension PET (Performance Evaluation Tool), [BHAR84]. A method of batch means is incorporated in SIMSCRIT II.5, [MILS87], and its specialized variations, such as Network II.5 and COMNET II.5. Partial automation of data analysis is also offered in QNAP2 (Queueing Network Analysis Package Version 2), see [POTI84] and [POTI86]. The effectiveness of proposed solutions generally depends on the level of a priori knowledge of the system's behaviour. A fully automatic procedure that could be used in stochastic simulation studies of a broader class of systems by users having a little knowledge or interest in the statistical analysis of output data is a matter for the future, although some doubt that it is possible at all. In Section 4 two exemples of sequential procedures are presented in more detail, together with an automatic procedure for determining the effective length of the initial transient period.

This report is not addressed to statisticians. We try to avoid the strict mathematical formulation of the problems considered, and only basic statistical terminology is used. Interested readers are referred to the references for more details.

2. METHODS OF DATA COLLECTION AND ANALYSIS

During the last 25 years of discussion on the methodology of statistical analysis of output data from steady-state simulation, initiated by Conway's paper on "Some Tactical Problems in Digital Simulation", [CONW63], a variety of methods for data collection and analysis has been proposed to circumvent the nonstationarity of simulated queueing processes (especially the initial nonstationarity caused by the existence of the initial transient period) and autocorrelation of events (correlations among collected observations). As has been mentioned, these methods either try to weaken (or remove) autocorrelations among observations, or to exploit the correlated nature of observations in analysis of variance needed for determining confidence intervals for the estimated parameters.

In ***the method of replications*** , adopted from terminating simulation, the autocorrelated nature of the original output data, i.e. correlation among collected observations, is overcome in a conceptually simple way: the simulation is repeated a number of times, each time using a different, independent sequence of random numbers, and the average value of observations collected during each run is computed. These means are used in further statistical analysis as secondary, evidently independent and identically distributed, output data.

Thus, if m observations are collected during any replication, the sequence of primary observations from k_b replications $(x_{11}, x_{12}, ..., x_{1m})$, $(x_{21}, x_{22}, ..., x_{2m})$, ..., $(x_{k_b1}, x_{k_b2}, ..., x_{k_bm})$ is replaced by the sequence of their means $\bar{X}_1(m), \bar{X}_2(m), ..., \bar{X}_{k_b}(m)$, where

$$\bar{X}_i(m) = \frac{1}{m} \sum_{j=1}^{m} x_{ij} \qquad (15)$$

which are used to obtain the point and interval estimates of the process. Namely, adopting Eqs. (1)-(4), in the case of k_b replications, each of which is of length m, we get the estimator of the mean μ_x as

$$\bar{\bar{X}}(k_b,m) = \frac{1}{k_b} \sum_{i=1}^{k_b} \bar{X}_i(m) \quad , \qquad (16)$$

and the 100(1-α)% confidence interval of μ_x is approximated by

$$\bar{\bar{X}}(k_b,m) \pm t_{k_b-1,1-\alpha/2} \; \hat{\sigma}[\bar{\bar{X}}(k_b,m)] \quad , \qquad (17)$$

where

$$\hat{\sigma}^2[\bar{\bar{X}}(k_b,m)] = \sum_{i=1}^{k_b} \{\bar{X}_i(m) - \bar{\bar{X}}(k_b,m)\}^2 / k_b(k_b-1) \qquad (18)$$

is the estimator of the variance of $\bar{\bar{X}}(k_b,m)$ and $t_{k_b-1,1-\alpha/2}$, for $0<\alpha<1$, is the upper (1-α/2)

critical point from the t-distribution with k_b-1 degrees of freedom#.

There exists a discrepancy of opinions on the effectiveness of this method, comparing it with other methods of data collection and analysis, all of which are based on a single (longer) run of the simulation experiment. Arguments defending the method of replications are provided by the results of [TURN77], [LAVE81, p.114] and [KELT84] which reveal that better accuracy of the estimator $\bar{\bar{X}}(k_b,m)$, measured by its MSE (see Eq.(7)), can be achieved if the simulation is run a few times, than if it is run one time only. But Cheng [CHEN76] argues that such a policy cannot be always correct ; see also [MADA76]. On the other hand, the method of replications appears to be much more sensitive to the nonstationarity of observations collected during the initial transient period than methods based on single simulation runs, since any new replication begins with a new warm-up period. If the bias of the estimator $\bar{\bar{X}}(k_b,m)$ is our main concern, then data collected during the initial transient period should be discarded, see Section 3, and in Eq.(16) $\bar{X}_i(m)$ should be replaced by

$$\bar{X}_i(m-n_o) = \frac{1}{m-n_o} \sum_{j=n_o+1}^{m} x_{ij} \qquad (19)$$

where n_o is the number of observations discarded from each replication¥. Thus the total number of initial observations discarded from k_b replications would be about k_b times larger than in corresponding single run methods. In the sequential version of the method new replications are generated until the required accuracy is reached. It was found that proper estimation of the length of the initial transient period can significantly improve the final coverage of confidence intervals obtained by the method of replications. There is a trade-off between the number of replications and their length for achieving a required accuracy of estimators. Fishman suggests choosing $m \geq 100$, to secure normality of the replication means; see [FISH78, p.122]. Results of [LAW77] and [KELT84] show that it is better to keep replications longer than to make more replications, since it will usually improve the final coverage too.

All other methods of data collection and analysis have been developed for obtaining steady state estimators from single simulation runs rather than from multiple replications. In ***the method of batch means*** the recorded sequence of original observations $x_1, x_2, x_3, \ldots,$ is divided into a series of nonoverlapping batches $(x_{11}, x_{12}, \ldots, x_{1m}), (x_{21}, x_{22}, \ldots, x_{2m}), \ldots,$ of size m, and batch means $\bar{X}_1(m), \bar{X}_2(m), \ldots,$ corresponding to the means over replications from Eq. (15), are next used as (secondary) output data in statistical analysis of the simulation results. The point and interval estimators of the mean are given by Eqs. (16)-(18), with k_b meaning now the number of batches and m being the batch size. This approach is based on the assumption that observations more separated in time are less correlated, so for sufficiently long batches of observations, batch means should be (almost) uncorrelated. If the bias of the estimator $\bar{\bar{X}}(k_b,m)$ is our main concern, then again the effective length of the initial transient period should be determined, see Section 3, and the first n_o observations collected during this period should be deleted. Thus the division of observations into batches should begin with setting $x_{11} = x_{n_o+1}$.

See the footnote on p.4.
¥ This number is generally different for different replications.

Selection of a batch size that ensures uncorrelated batch means appears to be the main problem associated with the method of batch means, besides the problem of selecting a suitable length of the initial transient period. A natural solution is to estimate correlation between batch means starting from an initial batch size m_1, and, if the correlation cannot be ignored, increase the batch size and repeat the test. At this stage, the method of batch means in its sequential version requires two procedures: the first sequentially testing for an acceptable batch size, and the second, sequentially testing the accuracy of estimators.

Correlation between the means of batches of size m can be measured by estimators of the autocorrelation coefficients

$$\hat{r}(k,m) = \hat{R}(k,m) / \hat{R}(0,m) \qquad (20)$$

where

$$\hat{R}(k,m) = \frac{1}{k_b-k} \sum_{i=k+1}^{k_b} [\bar{X}_i(m)-\bar{\bar{X}}(k_b,m)][\bar{X}_{i-k}(m)-\bar{\bar{X}}(k_b,m)] \qquad (21)$$

is the estimator of autocovariance of lag k, k= 0, 1, 2,...., in the sequence of batch means $\bar{X}_1(m), \bar{X}_2(m),..., \bar{X}_{k_b}(m)$. The definitions of $\bar{X}_i(m)$ and $\bar{\bar{X}}(k_b,m)$ are given in Eqs. (15) and (16), assuming that m is the batch size, and k_b is the number of batches.

The sequence of batch means can be regarded as non-autocorrelated when all $\hat{r}(k,m)$, k = 1, 2, ... assume small magnitudes, say, if they are less than 0.05. One can also determine the threshold for neglecting the autocorrelations in a statistical way, by testing their values at an assumed level of significance; see [ADAM83] and [WELC83, p.306]. The main analytical problem is caused by the fact that $\hat{r}(k,m)$'s of higher order are less reliable since they are calculated from fewer data points#. The higher the lag of an autocovariance, the fewer the observations which are available to estimate this autocovariance within a batch. Usually it is suggested to consider autocovariances of the lag not greater than 25% of the sample size ([BOXJ70, p.33]) or even not greater than 8-10% (c.f., [GEIS64]). Law and Carson [LAWC78] have proposed a procedure for selecting the batch size for processes with autocovariances monotonically decreasing with the value of the lag; see also [LAWK82]. In such a case only the lag 1 autocorrelation has to be taken into account . In this procedure three types of behaviour of $\hat{r}(1,m)$ as a function of m are distinguished. In the same class of processes Fishman has proposed testing against autocorrelation using von Neumann's statistic of batch means. One version of his procedure can be applied to processes with positive values of $\hat{r}(1,m)$, which decrease monotonically with m, [FISH78a], while another one includes cases when $\hat{r}(1,m)$ is a function oscillating in a damped harmonic fashion, assuming both positive and negative values [FISH78, p.240].

One might suppose that discarding some observations between consecutive batches should be an effective way for attaining an additional decrease of correlation between

The variance of the estimator of $\hat{R}(k,m)$ is reduced if the factor $1/(k_b-k)$ in Eq.(21) is replaced by $1/k_b$. But this variance reduction is followed by an increase of the bias of the estimator; see [PARZ61].

batch means. Solomon [SOLO83, p.200] proposed a more extreme action: to retain only every ν^{th} observation, $\nu > 1$, and to discard all observations between selected ones. This is equivalent to using (separated) batches of size m=1. In the example considered in [SOLO83, p.99] an interval of length $\nu = 25$ was selected by applying the Spearman rank correlation test. No results on effectiveness of this approach are given, but discarding 96% of collected observations seems to be quite wasteful. In fact, as shown by Conway [CONW63], the benefit of introducing the separation intervals is doubtful: it increases the variance of estimates and creates the problem of selecting the length for such intervals. Thus separating intervals are rarely used in practice.

Procedures for selecting the batch size employ different statistical techniques to find the batch size m^* for which correlations among batch means can be neglected, and hence they usually select different batch sizes. It has been reported that some of these procedures can consequently lead to interval estimates with very poor coverage, probably caused by accepting batch sizes which are too small . For example, the above-mentioned Fishman's procedures can select batches of as few as 8 observations. Law [LAW83] refers to simulation studies of M/M/1 queues in which the method of batch means with the procedure proposed in [LAWC78] was used. Using $k_b = 10$ batches of size m=32, for system utilisation $\rho = 0.9$, and 500 repeated simulation experiments, the achieved coverage of the nominal 90% confidence intervals was only 63%. For these reasons, Kleijnen et al. [KLEI82] suggested the use of a modified Fishman's procedure accepting batches at least 100 observations long, while Welch [WELC83, p.307] recommends constructing batches at least 5 times larger than the size m^* given by a test against autocorrelation, provided that at least 10 such batches can be recorded.

Schmeiser [SCHM82] analysed theoretically the trade-off between the number of batches, the batch size and the coverage of confidence intervals. These results suggest that the number of batches used in the analysis of confidence intervals should usually not be greater than 30, and $10 \leq k_b \leq 30$ is reasonable for most simulation experiments, regardless of the simulation run length. For better coverage, it is much more important to use longer batches than a greater number. Although the problem of selecting a proper batch size has not yet been fully satisfactorily solved, the method of batch means generally behaves better than the method of replications, see [LAW77], and can be regarded as a candidate for implementation in packages offering automated analysis of simulation output data.

In ***the regenerative method*** observations are also grouped into batches, but the batches are of random length, determined by successive instants of time at which the simulated process starts afresh (in the probabilistic sense), i.e. at which its future state transitions do not depend on the past. In the theory of regenerative processes, see for example [CINL75], which gives theoretical support for this method, such instants of time are called *regeneration points*. The special nature of the process behaviour after each regeneration point - its fresh "re-birth" - causes batches of observations collected during different *regenerative cycles* (i.e. within periods of time bounded by consecutive regenerative points) to be statistically independent and identically distributed. So are the means of these batches. For example, the regenerative points in the behaviour of simple single-server queueing systems are clearly the time instants at which newly arriving customers find the system empty and idle. From any such

moment on, no event from the past influences the future evolution of the system. More examples are given, for example, in [WELC83, p.317]. Note that usually a few, or even infinitely many, regenerative points of different nature can be distinguished in the behaviour of a system.

As a consequence of the identical distributions of output data collected within consecutive regenerative cycles, the problem of initialization vanishes if a simulation experiment commences from a selected regeneration point. The regenerative method was first suggested by Cox and Smith [COXS61, p.136], and then independently developed by Fishman ([FISH73a], [FISH74]), and by Crane and Inglehart ([CRAN74], [CRAN75]). Because of the random length of batches, these methods require special estimators, usually in the form of a ratio of two variables. In particular, if observations $x_1, x_2, ..., x_n$ are collected during N consecutive regenerative cycles, then the mean μ_x of the observed process is estimated by

$$\bar{X}(N) = \bar{Y}(N) \; / \bar{T}(N) \tag{22}$$

where

$$\bar{Y}(N) = \sum_{i=1}^{N} Y_i \, /N \tag{23}$$

$$\bar{T}(N) = \sum_{i=1}^{N} T_i \, /N \; . \tag{24}$$

In the above formulae,

$$T_i = n_{i+1} + 1 - n_i \tag{25}$$

is the length of the ith regenerative cycle, or equivalently the number of observations collected during the cycle i, n_i is the serial number of an observation collected at the ith regeneration point, and

$$Y_i = \sum_{j=n_i}^{n_{i+1}-1} x_j \tag{26}$$

Thus Y_i is the sum of observations collected during the ith regenerative cycle. If sufficiently many regenerative cycles is recorded then the $100(1-\alpha)$ % confidence interval of unknown parameter μ_x is bounded by

$$\bar{X}(N) \pm z_{1-\alpha/2} \frac{s}{\bar{T}(N)\sqrt{N}} \tag{27}$$

where $z_{1-\alpha/2}$ is, as usual, the upper $(1-\alpha/2)$ critical point from the standard normal distribution, and

$$s^2 = s_Y^2 - 2 \bar{X}(N) \, s_{YT} + [\bar{X}(N)]^2 \, s_T^2 \tag{28}$$

$$s_Y^2 = \sum_{i=1}^{N} \{Y_i - \bar{Y}(N)\}^2 / (N-1) \qquad (29)$$

$$s_{YT} = \sum_{i=1}^{N} \{Y_i - \bar{Y}(N)\} \{T_i - \bar{T}(N)\} / (N-1) \qquad (30)$$

$$s_T^2 = \sum_{i=1}^{N} \{T_i - \bar{T}(N)\}^2 / (N-1) \quad . \qquad (31)$$

It can be shown that $\bar{X}(N)$ given by Eq. (22) is a biased estimator of μ_x (we approximate the mean value of the ratio of two variables by the ratio of their mean values, see Eq.(22), which generally is not correct), although it is a consistent estimator, which means that $\bar{X}(N)$ tends to μ_x with probability 1 as $N \rightarrow \infty$. Additionally, the asymptotic normality of the ratio estimator $\bar{X}(N)$, on which the formula given by Eq.(27) is based, is questionable even for relatively large N. Thus these methods eliminate the bias of initialization but introduce new sources of systematic errors , caused by special forms of estimators. Some efforts have been made to obtain less biased estimators than those of Eqs. (22) - (25). Less biased estimators of μ_x have been proposed in [FISH77] (Tin's estimator) , [IGLE75] (the 'jackknife" estimator) and [MINH87]. Comparative studies reported in [IGLE75], [IGLE78], [GUNT80] and [LAWK82a] show that using the jackknife approach for the mean and variance estimation can significantly improve the accuracy of the estimates, although some question the generality of these results [BRAT83, p.92]. In some reported cases, especially if a small number of regenerative cycles is recorded, the performance of the regenerative method appears to be poor indeed, worse than that of the method of batch means, see [LAWK82a], [LAWK84].

A very effective sequential, regenerative procedure for output data analysis has been proposed by Fishman [FISH77]. Because of reservations about the appropriateness of the assumption of the approximate normality of $\bar{X}(N)$, the procedure is equipped with a statistical test for normality of the collected data (the Shapiro-Wilk test, see [SHAP65] or [BRAT83, App. A]. This normality test requires grouping output data (means over observations collected during consecutive regenerative cycles) into fixed size batches. Fishman [FISH77] proposed using batches containing data collected during at least 100 cycles and increasing the size of batches if the normality test fails. Results presented in [LAWK82a] show that this method, although rather more complicated numerically because of testing for normality, produces more accurate results in comparison with both a sequential "plain" regenerative method, proposed by Lavenberg and Saver [LAVE77], and a sequential method of batch means proposed by Law and Carson [LAWC79]; see discussion in [LAWC79]. A sophisticated modification of the regenerative method was also proposed by Heidelberger and Lewis [HEID81b], who suggest interactive intervention by users in the process of data collection and analysis, for achieving better accuracy. The regenerative method of data collection and analysis requires regenerative points to be well chosen to ensure that sufficient data can be collected for statistical analysis. To satisfy the last requirement a few approximations to the method have been proposed; see [CRAN75a], [CRAN77], [HEID79] and [GUNT80]. Gunter and Wolff proposed

replacing single regenerative states by sets of states and defining (almost) regenerative cycles, bounded by entries of the simulated process to such sets of states rather than to a single regeneration state as in the original method. Such modification can lead to even better accuracy of results than that obtained by the original (accurate) regenerative method, at least in the cases reported in [GUNT80]. But users must still, of course, select a proper set of (almost) regenerative states, which can sometimes involve substantial preparatory work. This method certainly deserves to be more thoroughly compared with others.

Any variant of the regenerative method offers very attractive solution to the main "tactical" problems of stochastic simulation, but it requires a deeper a priori knowledge of the simulated processes. Continuous repeated checking for regeneration conditions can increase the time of simulation experiments. The random length of regenerative cycles makes the control of the accuracy of results more difficult, since stopping the simulation at a non-regenerative point can cause a substantial additional bias [MEKE82].

As has been said, some methods of data collection and analysis specially exploit the correlated nature of observations when the variance, needed for the analysis of confidence intervals, is estimated. The simplest, but usually heavily biased, estimator of the variance $\sigma^2[\bar{X}(n)]$ can be obtained directly from Eq. (9). Namely,

$$\hat{\sigma}^2[\bar{X}(n)] = \frac{1}{n}[\hat{R}(0) + 2\sum_{k=1}^{n-1}(1 - \frac{k}{n})\hat{R}(k)] \qquad (32)$$

where

$$\hat{R}(k) = \frac{1}{n-k}\sum_{i=k+1}^{n}\{x_i - \bar{X}(n)\}\{x_{i-k} - \bar{X}(n)\} \qquad (33)$$

for $0 \le k \le n-1$.

This estimator can be improved by discarding $\hat{R}(k)$'s of higher order, since, as has been mentioned on p.12, they are less reliable when calculated from fewer data points. Usually it is assumed that the largest lag of autocorrelations included in Eq. (32) should equal 8-10% of the sample size; see [GEIS64]. Thus the number of components in Eq.(32) should be correspondingly decreased. The methods that we will present now offer further improvements of the estimator $\hat{\sigma}^2[\bar{X}(n)]$. All of them can be applied to a single simulation run and require the analyzed process to be a stationary one; thus they should be supported by an efficient procedure detecting the effective length of the initial transient period so that initial data may be discarded if necessary.

In ***the spectral method*** the analysis of a recorded sequence of observations is shifted into the frequency domain by applying a Fourier transformation to the autocorrelation function $\{R(k)\}$, $k = 0, 1, 2, \ldots$, yielding the *spectral density function*

$$p_x(f) = R(0) + 2\sum_{j=1}^{\infty} R(j)\cos(2\Pi fj) \ , \qquad (34)$$

for $-\infty \leq f \leq +\infty$; see e.g. [BRIL81] or [JENK68].

Note that because of the randomness of the collected observations, the spectral density function is a random function too¥. Comparing Eq. (34) with Eq. (9) one can see that, for sufficiently large n,

$$\sigma^2[\bar{X}(n)] \cong p_x(0) / n \quad . \tag{35}$$

Thus the estimator of $\sigma^2[\bar{X}(n)]$ can be obtained from an estimator $p_x(f)$ at f=0. Several techniques have been proposed for obtaining good estimators of the spectral density function $p_x(f)$. Most of them follow classical techniques of spectral estimation, based on the concept of *spectral windows* (special weighting functions introduced for lowering the final bias of the estimators); see e.g. [JENK68], [FISH73], [FISH78] and [MARK81]. The best results were obtained by applying the Tukey-Hanning window, [JENK68], [LAWK84]. Using this approach, the confidence interval can be determined by assuming that the normalized variable

$$(\bar{X}(n) - \mu_x) / \hat{\sigma}[\bar{X}(n)] \quad , \tag{36}$$

where

$$\hat{\sigma}^2[\bar{X}(n)] = \hat{p}_x(0) / n \quad , \tag{37}$$

has the t-distribution with κ_s degrees of freedom, where κ_s depends on the ratio of n/k_{max}, and k_{max} is the value of the upper lag considered in the autocorrelation function $\{R(k)\}$, i.e., $0 \leq k \leq k_{max}$; see [FISH73], also [BRAT83, p.97]. This approach can sometimes produce quite accurate final results, see [LAWK84], but it cannot be regarded as a good candidate for a more user friendly implementation because of its rather sophisticated nature. In particular there is no definitive method for choosing the parameter κ_s, c.f. [FISH78, p.265], [BRAT83, p.97] and [LAWK84].

The usefulness of spectral windows in reducing the bias of the estimate $p_x(0)$ has been questioned in [DUKE78], [WAHB80], [HEID81] and [HEID81a]. The last three papers propose estimating $p_x(0)$ by using the *periodogram* of the sequence $x_1, x_2, \ldots, x_n$. The periodogram $\{\Pi_x(j/n)\}$, $j = 0, 1, \ldots$, is a function of the discrete Fourier transforms $\{A_x(j)\}$ of the observations $x_1, x_2, \ldots, x_n$, namely

$$\Pi_x(j/n) = |A_x(j)|^2 / n \tag{38}$$

and

$$A_x(j) = \sum_{s=1}^{n} x_s \exp[-2\pi i(s-1)j/n] \quad , \tag{39}$$

where# $i^2 = -1$. It can be shown that for $0<j<n/2$

$$p_x(j/n) \approx E\,[\Pi_x(j/n)] \quad . \tag{40}$$

¥ Some point out that applications of the spectral method in analysis of simulation output results are natural only in the case of experiments in which, following the basic assumption of discrete Fourier transformation, observations are collected at equally spaced time intervals, [BRAT83, p.96].

The symbol i has this special meaning only in Eq.(39).

To find an unbiased estimate of $p_x(0)$ the periodogram is transformed into a smoother function, namely into the logarithm of the averaged periodogram

$$L_x(f_j) = \log \{ [\Pi_x((2j-1)/n) + \Pi_x(2j/n)] / 2 \} \qquad (41)$$

for $f_j = (4j-1)/n$. Next, this smoother function (but still not the smoothest one) is approximated by a polynomial to get its value at zero; see Appendix, and [HEID81] or [HEID81a] for details. Despite a number of approximations involved, the method produces quite accurate results, in particular in terms of coverage. Heidelberger and Welch proposed a sequential method of spectral analysis that uses a constant number of (aggregated) output data points instead of a growing number of individual observations, since, as they show, both individual observations and their batch means (of arbitrary size) can be used in the variance analysis. Namely, if n observations are grouped into b batches of m observations each, then for $n = bm$

$$\hat{p}_x(0) / n = \hat{p}_{\bar{X}(m)}(0) / b \qquad (42)$$

where, for $-\infty \le f \le +\infty$,

$$\hat{p}_{\bar{X}(m)}(f) = R(0,m) + 2 \sum_{j=1}^{\infty} R(j,m)\cos(2\Pi jf) \qquad (43)$$

is the spectral density function of the autocorrelation function $\{R(k,m)\}$ (k=0, 1, 2, ...,) of the batch means, see definition (21). This insensitivity of the method to batching the observations allows the batch size to be increased dynamically (starting from m=1), keeping in memory only a limited number of the batch means. A special batching / rebatching procedure is presented in [HEID81] and [HEID81a]. It appears to be an efficient way of limiting the required memory space. A modified version of this method is presented in Section 4.

The method of standardized time series was originally proposed by Schruben [SCRU83a]. According to this approach, after having discarded the initial n_o observations representing the nonstationary warm-up period, a sequence of collected observations $x_{n_o+1}, ..x_{n_o+2}, ..., x_{n_o+n}$ is divided into b batches $(x_{11}, x_{12}, ..., x_{1m})$, $(x_{21}, x_{22}, ..., x_{2m})$, ..., $(x_{b1}, x_{b2}, ..., x_{bm})$ of size m, $b \ge 1$, $x_{11} = x_{n_o+1}$. The batch i is then transformed into a standardized sequence $\{T_i(t)\}$, for $t = 0, 1/m, 2/m, ..., m/m=1$; which has a mean of zero and variance equal to 1. In this sequence:

$$T_i(t) = (mt) [\bar{X}_i(m) - \bar{X}_i(mt)] / \hat{\sigma}[\bar{X}_i(m)] \sqrt{m} \qquad (44)$$

with $T_i(0)$ defined to be zero, and

$$\bar{X}_i(mt) = \frac{1}{mt} \sum_{j=1}^{mt} x_{i,mt} \qquad (45)$$

for $t = 1/m, 2/m, ..., m/m=1$. Thus $\bar{X}_i(k)$, $1 \le k \le m$, is the cumulative average of the first k observations in the ith batch, and $\hat{\sigma}^2[\bar{X}_i(m)]$ is the variance estimator , which should take into account the correlated nature of the original observations, c.f. Eq. (9).

It can be proved that in the limit, as $m \to \infty$, the sequence $\{T_i(t)\}$, $0 \le t \le 1$, becomes the (standard) Wiener process, known also as the Brownian bridge, with independent increments; see [BILL68]. This fact is employed to get b (secondary) asymptotically independent output data, one from each batch, that are then used in two procedures proposed by Schruben for determining confidence intervals; see [SCRU83a]. A sequential version of one of them has been described in [DUER86]. The authors understressed the numerical simplicity of their proposal, despite the sophisticated statistical techniques that were applied . No results of studies comparing this method with other methods of data collection and analysis have been published yet.

Another approach for estimating the variance of correlated observations collected during a single simulation run is applied in ***the autoregressive method***, developed by Fishman [FISH71], [FISH73] and [FISH78]. Again it is assumed that after having decided about observations gathered during the initial output period, the recorded sequence of n observations represents a stationary process.

Let these observations be numbered starting from x_1. The main assumption of this method is that the sequence of originally correlated observations $x_1, x_2, ..., x_n$ possesses an *autoregressive representation* $y_1, y_2, ..., y_n$ of order q, where

$$y_i = \sum_{k=0}^{q} c_k (x_{i-k} - \mu_x) \qquad (46)$$

(for i = q+1, q+2,...,n) are i.i.d. variables with

$$E[y_i] = 0 \qquad (47)$$

$$E[\bar{Y}(n)] = \sum_{i=1}^{n} y_i / n \cong C\,[\bar{X}(n) - \mu_x] \qquad (48)$$

$$\sigma^2[\bar{Y}(n)] \cong C^2\ \sigma^2[\bar{X}(n)] \qquad (49)$$

where $C = c_0+c_1+...+c_q$; $c_0 = 1$. $\sigma^2[\bar{Y}(n)]$, as the variance of the mean of i.i.d. variables, can easily be estimated using Eq.(4), provided the coefficients q, $c_1, c_2, ..., c_q$ are known. The correct auto-regressive order q can be determined by examining the convergance of the distribution of a test statistic to an F distribution# (see [HANN70, p.336], [BRAT83]), or to a χ^2 distribution ([HANN70, p.336], [FISH78, p.251]). Having selected q, the estimates of the coefficients of $c_1, c_2, ..., c_q$ can be found from a set of k linear equations of the form

$$\sum_{i=1}^{q} \hat{c}_i\ \hat{R}(k-i) = -\hat{R}(k)\ , \qquad (50)$$

for k = 1,2,..., q, where $\hat{R}(k)$ in the estimator of the lag k autocovariance of the original sequence $x_1, x_2, ..., x_n$, see Eq. (10) and the discussion after Eq. (33). Next, having determined $\hat{\sigma}^2[\bar{Y}(n)]$, one can easily find the estimator of the variance of $\bar{X}(n)$, since from Eq. (49):

The F distribution, where F stands for Fisher, is also known as the variance-ratio distribution , the Fisher distribution or the Snedecor distribution.

$$\hat{\sigma}^2[\bar{X}(n)] = \hat{\sigma}^2[\bar{Y}(n)] / C^2 \quad . \qquad (51)$$

Finally, the confidence interval for $\bar{X}(n)$ can be determined assuming that $\sqrt{n}\,(\bar{X}(n) - \mu_x) / \hat{\sigma}[\bar{X}(n)]$ is governed by the t-distribution with

$$\kappa_r = n\,C / 2 \sum_{j=0}^{q} (q - 2j)\hat{c}_j \qquad (52)$$

degrees of freedom; see arguments given in [FISH78, p.252]. Thus the resulting confidence interval is determined by

$$\bar{X}(n) \pm t_{\kappa_r,1-\alpha/2}\ \hat{\sigma}[\bar{X}(n)] \qquad (53)$$

where $t_{\kappa_r,1-\alpha/2}$ is the upper $1-\alpha/2$ critical point obtained from the t-distribution with κ_r degrees of freedom.

The main restriction of the method seems to be the required existence of an autoregressive representation of the simulated process. Results of empirical studies of the method's efficiency published in [FISH71] were not very encouraging, since frequently the final coverage was below 80%. However these results were achieved in short simulation runs. Andrews and Schriber, in their studies of the autoregressive method reported in [ANDR78], [SCRI79] and [SCRI81], observed a significant variability of the average half-widths of confidence intervals produced by the considered variants of the method. Law and Kelton [LAWK84], after comparative studies of different fixed-size methods of data analysis, found also that the autoregressive approach does not offer better results than other, computationally simpler methods of data analysis. And, in contrast to both the method of batch means and the method of spectral analysis, the improvement of the final coverage when increasing the number of collected observations was very slow. Continuous execution of the test for determining the autoregressive order q and solving the sets of equations for determining the coefficients $c_1, c_2, \ldots, c_q$ could be time consuming in a sequential version, especially if longer sequences of observations have to be collected. Taking all these reservations into account, it is unlikely that the autoregressive method will find broader applications in user friendly simulation packages.

3. THE PROBLEM OF INITIALIZATION

It is well known that just after initialization any queueing process is in a transient phase, during which its (stochastic) characteristics vary with time. This is caused by the fact that , like any (stochastic) dynamic system, every queueing system or network initially "moves" along a nonstationary trajectory. After a period of time the system approaches its statistical equilibrium on a stationary trajectory if the system is stable, or remains permanently on a non-stationary trajectory if the system is unstable. Note that in practice only queueing systems with infinite populations of customers and unlimited queue capacities can never enter a stationary trajectory and this happens if the average request for service is greater# than the average supply of service, i.e., if

$$\lambda > c\mu_s \qquad (54)$$

where λ is the mean arrival rate, $1/\mu_s$ is the mean service time , and c is the number of service facilities. In such a case the systems become congested and then deadlocked as a result of queues increasing in length with time. On the other hand, queueing systems with limited queue capacities always reach an (inner) statistical equilibrium, even if the system's load expressed by the traffic intensity $\rho = \lambda/c\mu_s$ is much greater than 1. In such a case internally stationary queueing systems are in the non-stationary environment of streams of rejected customers. Of course output data collected during transient periods do not characterise steady state behaviour of simulated systems and so they can cause quite significant deviation of the final "steady state" results from their true values. Although it seems quite natural that the deletion of untypical initial observations should result in better steady state estimators, the problem *" to delete or not to delete"* is a perennial dilemma of stochastic simulation practice. Each of these two alternatives has its advocates. The answer depends on the assumed measure of goodness and the resource limitations of simulation experiments (the maximum possible number of recorded observations). The influence that the initial transient data can have on the final results is a function of the strength of the autocorrelation of collected observations. With no restrictions imposed on the length of the simulation run, this influence can be arbitrarily weakened by running the simulating program sufficiently longer. But in most practical situations simulation experiments are more or less restricted in time, and that time can be more or less effectively used to calculate estimators. If the initial output data are retained, the bias of the point estimator $\bar{X}(n)$ is greater than if they were deleted.

Contrary opinions on the usefulness of deletion are caused by the fact that it increases the variance of the point estimator (see [FISH72] and [FISH73, Sec.10.3], [TURN77]), and, in effect, can increase its MSE, see Eq. (8). Let us note that an increase of the variance can be compensated for by applying one of the variance reduction techniques. Deletion of initial observations seems to be justified if the variance of the estimator is smaller than the squared bias, and/or if observations are strongly correlated (the initial conditions have a longer effect on the evolution of the system in time).

#The D/D/c queueing systems are stable also if $\lambda = c\,\mu_s$.

On the other hand, Blomqvist [BLOM70] showed that for long run simulations of GI/G/1 queueing systems the minimum MSE of the mean delay usually occurs for the truncation point $n_o = 0$, which supports the thesis that no initial observations need to be deleted. Results of experiments conducted by Turnquist [TURN77], and Wilson and Pritsker [WILS78a] provide the same argument.

The usefulness/uselessness of data deletion depends also on methods used for data collection and analysis. Independent replications give much more "contaminated" data than methods of data collection based on single runs, since each replication begins with a new initial transient period. In consequence data deletion seems to be more crucial for plurality of transient periods than for just one transient period in one long run. In an example discussed in [KELT83] , the estimator of mean delay in an M/M/1 queue obtained from replications of 500 observations, without initial deletions, was biased -43.2%, for ρ=0.95. In the case of higher accuracy requirements, a significant bias of estimators will normally increase their chances of being outside the theoretical confidence intervals, thus it will decrease the coverage of confidence intervals. Law and Kelton ([LAW83] and [KELT84]) analyzed the influence of initial data deletion on the coverage of the final results in the case of the method of independent replications and stated a clear improvement of the actual coverage ("effective" confidence levels)# to levels near nominal theoretical values (1-α), without unduly widening confidence intervals, especially if replications were not too long and/or not too many observations were deleted. In methods of data analysis based on single runs and an assumption that the observed process is (covariance) stationary deletion of data from the initial nonstationary period improves approximate stationarity of the remaining process.

The nature of the convergence of simulated processes to steady state depends on many factors; the initial conditions of simulation are one of them. Conway [CONW63] advised a careful selection of starting states (typical ones for steady state of the simulated process) to shorten the duration of the initial transient phase. Since then many trials have been undertaken to determine the optimal initial conditions, in the sense that they would cause the weakest influence of the transient phase on the steady state results, but ambiguous conclusions have been reached. Madansky [MADA76] proved that the MSE of the mean queue length in simulation studies of M/M/1 queueing systems (without data deletion) can reach its minimum value if they are initialized as empty-and-idle, i.e., in their modal states. Wilson and Pritsker [WILS78a], having examined a slightly broader class of queueing processes, concluded that the optimal (in the MSE sense) initial state is the most likely state in statistical equilibrium (the mode of the steady state distribution), if it differs from the empty-and-idle state. Moreover they found that a judicious selection of initial conditions can be more effective than the deletion of initial data. Similar conclusions were reached also by Donnelly and Shannon [DONN78] after a more methodical investigation.

On the other hand Kelton and Law [KELT84], and Kelton [KELT85a], investigating queueing systems with exponential and Erlangian distributions of interarrival and service times, discovered that the shortest transient periods occur if the simulated processes start from states slightly larger than their steady state means, for example

For the reasons mentioned in the introduction, in steady state simulations the real confidence levels of estimators usually differ from their theoretical (1-α) values, assumed in the formula (2).

in the M/M/1 queue the mean delay reaches its steady state in the shortest time (with accuracy $\pm$ 0.01, ρ = 0.9) if the initial queue length is 15, while the steady-state mean equals 9. These results have no theoretical explanations yet, but they clearly indicate that it is better to start from the initial state equal to the steady-state mean rather than from the mode of the distribution. However it was also shown that starting from states greater than the mean usually gives a much longer transitive period than starting from states the same distance from the mean, but below it. Thus, because in real situations the steady-state mean is unknown, it is much safer to initialise systems as empty-and-idle, particularly if the bias of an estimator concerns us more than its MSE .

Having decided to discard data collected during transient periods we face the next problem : how long such periods last. In simulation practice we can encounter both very short initial transient effects and transient effects that are spread over tens of thousands of observations, see e.g. [HEID83].

THE DURATION OF THE INITIAL TRANSIENT PERIOD. The problem of determining the duration of the initial transient period in simulation runs appears to be quite complicated . Conway [CONW63] suggested that

R1. *in a time series of observations* $x_1, x_2, ..., x_n$ *the initial transient lasts until the first of the series is neither the maximum nor minimum of the rest.*

This rule of thumb, associating the beginning of steady state with the occurrence of the first "typical" observation, appears to give a very poor approximation of the duration of the initial transient. As was shown in [GAFA78], using this rule we can significantly overestimate the length of the initial transient for small ρ and underestimate it for high ρ , c.f. also [WILS78a].

The duration of the initial transient period is also analyzed in the queueing theory. It has been shown that the rate at which queues tend to their steady state is, after some period of time, dominated by a term of the form exp (-t /τ), where τ is called the *relaxation time* of the queue . We could assume that

R2. *the initial transient period is over after the time* $t_\beta = -\tau \ln \beta$, *where* β *is the permissible relative residue of the initial state ,* $0 < \beta < 1$,

since at $t = t_\beta$ the queue characteristics retain only about 100β % of their initial value, and output data collected from that point of time on should be biased by initial states by less than 100β %. The analysis of relaxation times was initiated by Morse [MORS55], who considered the correlation function of the M/M/1 queue length. Cohen [COHE69] analyzed transient distributions of queue lengths and determined the relaxation time for GI/G/1 queueing systems. These appear to be from 9 to 2 times greater than Morse's results for M/M/1 systems, as ρ changes from 0.1 to 1.0. This diversity of results has stimulated search for approximate formulae for relaxation times, such as Newell's result for GI/G/1 under heavy traffic queues, [NEWE71]. Probably the most accurate results for Markovian queueing systems have been obtained by

Odoni and Roth [ODON83]. Having studied various Markovian systems, they determined the relaxation time to be

$$\tau = (C_A^2 + C_S^2 / 2.8 \mu_s (1 - \sqrt{\rho})^2 \qquad (55)$$

where C_A^2 and C_S^2 are the coefficients of variation for the interarrival and service times, respectively, and $1/\mu_s$ is the mean service time. This result shows clearly that more heavily loaded systems tend more slowly to their statistical equilibrium .

Relaxation times have also been analysed theoretically in some simple queueing networks; for example Blanc [BLAN84] analyzed the relaxation time in an open network of K service centers with a Poisson arrival stream, an unlimited number of servers at each center, general distribution of service times and a homogenous transition matrix. He showed that the relaxation time in such a network has an upper limit, namely

$$\tau \leq K / \mu_s \ . \qquad (56)$$

The equality occurs for the tandem connection of queueing systems. Also a conjectural relaxation time for Jacksonian queueing networks with K single server centers has been proposed. For more complex queueing networks the relaxation times have not yet been theoretically determined. But the usefulness of even known formulae for relaxation times can be questioned in simulation studies.They can be used only as first approximations of the duration of simulated initial transients, since it has been shown that estimators of the mean values from simulation tend to their steady state more slowly than exponentially; e.g. Anderson [ANDE85] showed that in queueing systems with limited queue capacities the rate at which the estimator of mean queue length tends to its steady state eventually becomes inversely proportional to time. It has also been shown that the standard deviation of estimators converges even more slowly, namely in inverse proportion to the square root of time, see [FISH68] and [ANDE85]. Both these facts have found application in various heuristic rules proposed for determining the duration of the initial transient period.

Studying the convergence of a moving average of output data to determine a possible end of the initial transient period is attributable to Gordon [GORD69, p.285]#. The simplest way would be to find an instant of time at which the running mean $\bar{X}(n)$ (see Eq.(1)) approaches a constant level with a given accuracy δ, $\delta>0$. Thus , we can assume that

R3. *in a time series of observations $x_1, x_2, ..., x_i, ...$, the initial transient period is over after n_o observations if k consecutive values of the running mean $\bar{X}(i)$ recorded after the observation n_o differ less then 100δ % from $\bar{X}(n_o+k)$, i.e., for all i, $n_o < i \leq n_o + k$,*

$$| \bar{X}(n_o+k) - \bar{X}(i) | / | \bar{X}(n_o+k) | < \delta \ . \qquad (57)$$

Solomon [SOLO83, p.195] attributed this approach to Emshoff and Sisson , referring to [EMSH70, Sec. 8.2.].

The stabilization of $\bar{X}(n)$ should be tested over a sufficiently long sequence of observations, so the parameter k should be large (in the statistical sense), i.e. $k \geq 30$. The above rule has two weaknesses. Firstly, as has been indicated already by Conway [CONN63], accumulative statistics such as running means usually stabilize very slowly with time, and so usually give overestimated values of n_o . Additionally, fluctuations of the running mean $\bar{X}(n)$, calculated over data collected during a single simulation run, can be significant for a longer time . For these reasons the above rule and its various modifications are usually employed with the method of replications, and the inequality of rule R3 is applied to the running mean after having additionally smoothed it by averaging over replications. Despite this the resulting length of the initial transient period is usually still overestimated, c.f. [GAFA78], [WILK78] and [ROTH85]. Welch proposed a special technique for smoothing running means that uses the concept of a "moving window", within which mean values over replications are additionally averaged, producing a smoother (but still highly correlated) process [WELC83, p.293]. The effectiveness of this technique has not been studied yet .

Another rule of thumb can be based on the supposition that in steady state typical observations are evenly dispersed around the mean value. For example, Fishman [FISH73, p.275] proposed that

R4. *the initial transient period is over after* n_o *observations if the time series* $x_1, x_2, ..., x_{n_o}$ *crosses the mean* $\bar{X}(n_o)$ *k times.*

This rule is sensitive to the value of k , see [GAFA78] . Too large a value will usually lead to an overestimated value of n_o regardless of system's utilization, while too small a value can result in an underestimated n_o in more heavily loaded systems. In [GAFA78] k = 25 was recommended for M/M/1/∞ queueing systems, while in [WILS78a] k =7 was chosen for the M/M/1/15 system. The system-dependent selection of the parameter k in the rule R4 seems to be too arduous for potential users .

Solomon [SOLO83] proposed# applying the χ^2 (chi-square) goodness-of-fit test to select a time from which the number of observations below and above the running mean are equal (in the statistical sense). According to this test, the sequence of observations should be partitioned into batches of at least m_o=10 observations each (Solomon selected m_o=30) and then one can conclude

R5. *in a time series of observations* $x_1, x_2, ..., x_n$ *the initial transient is over after* n_o *observations, if the* χ^2 *goodness-of-fit test confirms that in the batch of observations* $x_{n_o+1}, x_{n_o+2}, ..., x_{n_o+m_o}$ *following the observation* n_o *the numbers of observations above and below the running mean* $\bar{X}(n_o)$ *are about the same.*

This rule seems to be quite simple and independent of any system-related parameter. No results are available on its effectiveness and relationship to other criteria.

Solomon attributed this approach to Emshoff and Sisson.

For reducing the fluctuations of analyzed sequences, and saving memory space if long transient periods are expected, batches of individual observations can be replaced by their mean values. Let us note that such batches must be introduced before the procedures for selecting the size of uncorrelated batch means, discussed in Section 2, can be applied. There are no established rules in this case for selecting the batch size. If a statistical test is used to help in deciding about the length of the initial transient, then we should follow the requirements of the test or use statistically large batches, which usually means taking $m_o>30$. Otherwise the only recommendation is to select a batch size which gives the desired data reduction while retaining the stabilizing trend of the original sequence. After the batch size m_o is selected the sequence of batch means $\bar{X}_1(m_o), \bar{X}_2(m_o), ...$, can be tested in a similar way to the sequence of the original observations. For example, Wilson and Pritsker [WILS78] formulated the following rule which they attributed to Schriber [SCRI74]# :

R6. *in a time series of batch means $\bar{X}_1(m_o), \bar{X}_2(m_o), ...$, the initial transient is over after b_o batches , i.e. , after $n_o = b_o m_o$ observations, if the k most recent batch means all fall within an interval of width δ_1, i.e. , if*

$$| \bar{X}_{b_o-i}(m_o) - \bar{X}_{b_o-j}(m_o) | < \delta_1 , \qquad (58)$$

for $0 < i < k-1$, $0 < j < k-1$.

This rule, like rule R3, is sensitive to the value of the parameter k, which should depend on the variability of the observed process. A small value for k, for example k=2, as was assumed in [WILS78] and [SOLO83, p.196], can lead to an underestimation of n_o, since the difference between averages, having dropped only k times below δ_1, can easily rise again to an unacceptable level , as in an example considered in [SOLO83, p.197].

For further data reduction and additional smoothing of the tested sequence, Kelton and Law [KELT83] proposed applying such a batching technique also to simulation experiments based on independent replications, see Section 2. Namely they batched the sequence of mean observations (means over replications), and then analysed the sequence of the means of these batches. Thus they assumed :

R7. *in a time series $\bar{X}_1(m_o), \bar{X}_2(m_o), ...$, the initial transient period is over after the batch b_o , i.e., after $n_o = m_o b_o$ observations, if the means of batches after the batch b_o can be approximated by a straight line with zero slope.*

This rule can be applied only in the case of monotonic convergence to steady state but , as was proved by Kiefer and Wolfowitz [KIEF55], in any stable, initially empty-and-idle GI/G/c queueing system the mean delay-in-queue grows monotonically in time. The procedure implementing rule R7 appears to be quite effective, if properly used, see [ROTH85]. Kelton and Law proposed testing the slope of the regression line backwards, after collecting an assumed number of observations. The test for zero slope is over a fixed number of batch means and , if zero slope is accepted, the

Solomon [SOLO83, p.195] attributed this approach also to Emshoff and Sisson , and referred to [EMSH70, Sec. 8.2.].

test is repeated over an earlier sequence of batch means, to find whether the initial transient period had expired earlier. Otherwise, if the test fails at the beginning, a new check point is chosen after gathering further output data from the simulation. Note that this requires that the process of collecting new observations in all previously stopped replications be continued again. Because correlations between batch means can still be significant, they are approximated by a straight line using a generalized least squares procedure proposed by Amemiya [AMEM73], which allows for autocorrelation of the analyzed data. For additional saving of memory space, the number of batches could be kept constant, by allowing the size of batches to grow when simulation runs are continued.

Rules R3 - R7 are based on the convergence of the mean of observations to its steady state value. Other criteria have been proposed taking advantage of the fact that queueing processes in steady state are stationary. Namely, because the variance of the mean of observations taken from a (covariance) stationary process is approximately inversely proportional to the number of observations (see [FISH73, p.281] or [GAFA78]), i.e.,

$$\hat{\sigma}^2[\bar{X}(n)] = C_1 / n + o(1/n) \quad , \qquad (59)$$

c.f. Eq.(9), where C_1 is a positive constant, and n is the number of observations, Gordon [GORD69] proposed that ,

R8. *in a time series* $x_1, x_2, ..., x_n$ *the initial transient is over after the observation* n_o *if the graph* $(lg\ n, lg\ \hat{\sigma}[\bar{X}(n)])$, *becomes approximately linear with slope -0.5 from this observation on.*

To smooth variations of the analysed curve, Gordon [GORD69] proposed analyzing the variance of the mean of observations averaged over a number of replications. This rule was analysed in [GAFA78] and [WILS78], using formula (4) to calculate $\hat{\sigma}^2[\bar{X}(n)]$, thus rejecting existing correlations between observations. In this case rule R9 can give a very overestimated values of n_o. No results are known about the effectiveness of this rule if more accurate estimators of $\sigma^2[\bar{X}(n)]$ are applied.

Fishman [FISH71, p.29] proposed equating the variance of the mean of autocorrelated observations with the variance of the mean of a hypothetical sequence of independent observations, to find the number of collected (autocorrelated) observations equivalent, in the above sense, to one independent (hypothetical) observation. After some simplification we get the following rule :

R9. *in a time series of observations* $x_1, x_2, ..., x_n , ...,$ *the initial transient is over after*

$$n_o = 2 \sum_{k=1}^{n-1} (1 - \frac{k}{n}) \hat{R}(k) / \hat{R}(0) \qquad (60)$$

observations, where $\hat{R}(k)$ *is the estimator of the autocorellation of the lag* k, $0 \le k \le n-1$; *see Eq. (33) .*

The sequence of observations collected after the observation n_o should be (approximately) independent of the initial conditions. The autocovariance estimators $\hat{R}(k)$ should be analyzed with caution; see the comments after Eq.(33). Comparing this with the results given for example in [ROTH85], one can state that rule R9 usually gives underestimated values of n_o ; no exhaustive comparisons of this rule with other criteria are available.

The above-mentioned rules proposed for determining the length of the initial transient periods are either quite elaborate and, as such, do not ensure an accurate control of the initialization bias, or they can determine quite precisely the length of the initial transient period, but only for restricted classes of simulated processes and/or by applying quite sophisticated techniques to collect and analyze the output data. This can unnecessarily lengthen the time of simulation experiments, especially if the required accuracy is tested sequentially, c.f. the rule R7. Some of these rules have been implemented as built-in options offered in simulation packages such as GPSS, SLAM and SIMSCRIPT II.5, [LAWK82]. Thus potential users should be aware of their limitations.

A promising approach for detecting the expiration of the initial transient period is offered by statistical stationarity tests, based on the theory of dependent stochastic processes, developed by Billingsley [BILL68]. According to this approach

R.10 *the initial transient data have been removed from a given sequence of observations if the (standardized) sequence determined over the remaining observations behaves in a way consistent with a standard (stationary) stochastic process.*

Schruben et al. ([SCRU80] and [SCRU81]) have proposed transforming the sequence of original observations $x_1, x_2, ..., x_n$ into the standardized sequence $\{T(t)\}$, $t = 0, 1/n, 2/n, ..., n/n=1$; which has a mean of zero and variance equal 1, and

$$T(t) = (nt) [\bar{X}(n) - \bar{X}(nt)] / \hat{\sigma} [\bar{X}(n)] \sqrt{n} \quad , \qquad (61)$$

for $t = 1/n, 2/n, ..., n/n=1$; $T(0) = 0$; and then to test the convergence of the sequence $\{T(t)\}$ to the Brownian bridge process# , i.e., applying the same approach as in the methods of standardized time series discussed in Section 2. Heidelberger and Welch [HEID83] listed a few other standardized sequences that can be used to find statistics for the above rule. Rejection or acceptance of the hypothesis that a given subsequence of observations is stationary, or equivalently, that the initial transient period is not included in the observations, depends on the probability characterizing the scalar value calculated from the considered sequence. Despite the sophisticated theory hidden behind these tests they appear to be quite simple numerically, and can be applied to a wide class of simulated processes. A sequential version of one of the tests proposed by Schruben et al. in [SCRU83] is presented in Section 4.

The main problem with practical implementation of the last rule is that usually a priori knowledge of the variance of the simulated process in steady state is required. To estimate this variance one can use a sequence of observations collected at some distance from an assumed truncation point, assuming that the process is then at

The mathematical model of Brownian motion on the [0,1] interval.

least closer to steady state. One of the tests presented in [SCRU82] does not need the steady state variance to be given (the final inference is drawn from the Fisher distribution¥, with 3 and 3 degrees of freedom); but no results on the robustness of this test are available. To shorten searching for the beginning of the stationary phase, the rule R10 can be preceded by one of the simple rules of thumb to find a rough approximation of the beginning of this phase. For this purpose one can use, for example, the rule R1 or R5.

¥ See the footnote on p. 19.

4. SEQUENTIAL STEADY STATE PROCEDURES: EXAMPLES

In this section we present an implementation of a sequential procedure that uses a statistical test for detecting the length of the initial transient period, and implementations of two sequential procedures for stopping the simulation run when the required relative accuracy is achieved. The former is based on the spectral method proposed by Heidelberger and Welch [HEID83], and the latter is based on the method of the batch means proposed by Adam [ADAM83].

4.1. DETECTING THE LENGTH OF THE INITIAL TRANSIENT PERIOD

As was mentioned in Section 3, the stationarity tests proposed by Schruben et al. ([SCRU82] and [SCRU83]) can be used to test the hypothesis that a sufficient number of initial transient data has been (or has not been) discarded. As in any statistical test, the value of a chosen statistic calculated from the tested sequence of observations is compared with the corresponding value from the standard sequence, and the decision about rejection or acceptance of the hypothesis is taken at an assumed significance level α_t #, $0<\alpha_t<1$. To get a first approximation for the truncation point, n_0^*, we can use one of the heuristic rules R1- R9, presented in Section 3. For example, in simulation studies of satellite communication protocols, [PAWL88], the rule R4 was applied with the parameter k = 25.

The problem encountered during testing a sequence of n_t observations for stationarity is that the steady state estimator for the variance $\sigma^2[\bar{X}(n)]$, and the number κ of degrees of freedom for its distribution¥, has to be known earlier than we know that the process has entered the stationary region. To get a robust estimate of that variance the estimation should be done using only a subsequence of the last n_v observations from the sequence of n_t observations tested for stationarity; these n_v observations are more likely to be already from the stationary region, even if the truncation point of the initial transient period has been initially underestimated. One can for example assume $n_v = 0.5n_t$. Of course both the value of n_v and n_t should be selected after having taken into account the minimum sample sizes required by the method of variance analysis and the stationarity test. Heidelberger and Welch assumed $n_v \geq 100$, [HEID83]. Having assumed this value of n_v we have at least $n_t = 200$ observations stored in a buffer for testing against nonstationarity, which is the size of the sample assumed by Schruben [SCRU82]. Since the number of observations tested for stationarity should be larger if longer transient periods are suspected, one can assume that

$$n_t = \max(\Delta_n, \gamma n_0^*), \qquad (62)$$

where Δ_n is the smallest sample size required by the stationarity test, and γn_0^* is the smallest number of new observations selected to represent a given process, for a given truncation point n_0^* and $0<\gamma<1$.

#The value α_t can be interpreted as the probability of erroneously rejecting the hypothesis that the tested process is stationary.

¥ As usual, the distribution of $\hat{\sigma}^2[\bar{X}(n)]$ is approximated by the χ^2 distribution with κ degrees of freedom.

Thus, after having discarded n_o^* observations, the next n_t observations are collected and the last n_v observations are used to find $\hat{\sigma}^2[\bar{X}(n)]$ and κ. The variance and the degrees of freedom of its χ^2 distribution can be estimated using a few different methods, presented for example in [FISH73, p.289], [SCRU82] and [HEID81]. The last method, based on spectral analysis of the sequence of observations, was used in [PAWL88] and is summarized in Appendix. Having estimated the variance that by assumption represents steady state of the process, we can start testing the first n_t observations for stationarity. If the test accepts the hypothesis that the end of the initial transient has been detected correctly and the process has already entered its stationary region, this stage of analysis of the simulation output results is finished and the program may start to analyze confidence intervals, see Section 4.2. Otherwise transient effects have been detected. Consequently an additional γn_o^* observations from the beginning of the tested sequence is discarded, γn_o^* new observations are collected, the last n_v observations are again used to estimate $\sigma^2[\bar{X}(n_v)]$ and κ, and the first n_t observations in the stored sequence are tested for stationarity. This procedure is continued until the stationarity of the sequence of observations is confirmed or $n_{o,max}$, the upper limit of the number of tested observations, is reached. As the longest acceptable length of the initial transient period one can assume $n_{o,max}=0.5n_{max}$, where n_{max} is the maximum length of the simulation run. If, at any stage of this procedure, the transient phase extends beyond $n_{o,max}$ observations, then it is either an unstable system or the allowed maximum length of the simulation run is too short.

Summing up, the procedure requires the following parameters :

n_{max} : the maximum allowed length of the simulation run measured in the number of recorded observations (to be decided in advance);
$n_{o,max}$: the maximum allowed length of the initial transient period (the default value: $n_{o,max} = 0.5\, n_{max}$);
n_v : the length of the sequence used for estimating the steady state variance (the default value: $n_v = 100$);
Δ_n : the minimum length of the sequence tested for stationarity ($\Delta_n > n_v$; the default value $\Delta_n = 2n_v$);
α_t : the significance level of the stationarity test ($0<\alpha_t<1$; the default value: $\alpha_t = 0.1$)
γ : the "exchange" coefficient, determining the length of a step in sequential testing for stationarity ($\gamma > 0$; the default value: $\gamma = 0.5$);

and it can be described as follows:

procedure *DetectInitialTransient* ;
{ *determine the length of the initial transient period applying the Schruben's test preceded by an heuristic rule of truncation* }

Step 1:
Start the simulation run from the empty-and-idle state ;
apply one of the heuristic deletion rules R1-R9 {*see Section 3*} to determine n_o^* ;
{ n_o^* *is the first approximation of the number of observations to be deleted* }

if (the initial transient period embraces more than $n_{o,max}$ observations) **then**
goto *Step 6*
else $n_o := n_o^*$; discard first n_o observations
endif;
if $\Delta_n \geq \gamma n_o^*$ **then** $n_t := \Delta_n$ **else** $n_t := \gamma n_o^*$ **endif**;
$\Delta_t := n_t$; { *the number of new observations which should be appended to the tested sequence* }

Step 2:
if $n_o + n_t \leq n_{o,max}$ **then**
append Δ_t observations to the tested sequence; {*some of these observations may have been already collected when the heuristic deletion rule was applied*}
goto *Step 3*
else { *if* $n_o + n_t > n_{o,max}$ }
goto *Step 6*
endif;

Step 3 :

Determine the variance $\hat{\sigma}^2[\bar{X}(n_v)]$ and κ, the degrees of freedom of its distribution, using the last n_v collected observations starting from the observation $(n_o+n_t-n_v+1)$; {*e.g. apply the procedure SpectralVarianceAnalysis described in Appendix, assuming m=1, and*

$\bar{X}_s(m) = x_{n_o+n_t-n_v+s}$, *for* $s = 1, 2, ..., n_v$ }

Step 4 : {*the Schruben test for stationarity, [SCRU83, p. 1173]*}
Take all n_t observations, starting from the observation (n_o+1), and calculate the test statistic

$$T = (\sqrt{45} / n_t^{1.5}\ \hat{\sigma}^2[\bar{X}(n_v)]) \sum_{k=1}^{n_t} k(1-k/n_t)\,[\,\bar{X}(n_t) - \bar{X}(k)\,]\,, \quad \text{where} \quad \bar{X}(i) = \sum_{j=n_o+1}^{n_o+i} x_j / i \ ;$$

if (a negative bias of the mean $\bar{X}(n_t)$ is suspected) **then goto** *Step 5*
elsif (a positive bias is suspected) **then** T := -T
else T := | T | { *if a sign of the initial bias is difficult to predict either sign is equally likely* # }
endif;

Step 5 :

if $T \leq t_{\kappa,1-\alpha_t/2}$ **then**

{$t_{\kappa,1-\alpha_t/2}$ *is the upper* $(1-\alpha_t/2)$ *critical point from the t distribution with* κ *degrees of freedom*}
write ('*the initial transient period is not longer than* n_o *observations* ');
start sequential analysis of confidence intervals
{*call e.g. one of procedures presented in Section 4.2*}

else {*if* $T > t_{\kappa,1-\alpha_t/2}$ }

discard first γn_o^* observations from the tested sequence;

The reason for considering bias of an assumed sign is that two-sided tests are usually less powerful than their one-sided correspondents.

$n_o := n_o + \gamma n_o^*$; $\Delta_t := \gamma n_o^*$; **goto** *Step 2*
endif

Step 6 : {*if the initial transient period embraces more than* $n_{o,max}$ *observations*}
stop the simulation run;
write ('*the length of initial period is longer than* $n_{o,max}$ *observations or the simulated process is unstable* ')
end ***DetectInitialTransient*** **.**

The effectiveness of this procedure depends mainly on the effectiveness of the variance estimator $\hat{\sigma}^2[\bar{X}(n)]$.

4.2 SEQUENTIAL TESTING FOR A REQUIRED ACCURACY OF RESULTS

The sequential procedures for stopping a simulation experiment which are presented here require the analyzed sequence of observations to be representative of steady state, so n_o observations representing the initial transient period have to have been discarded beforehand.

The first procedure applies the spectral method of analysis proposed by Heidelberger and Welch ([HEID81], [HEID81a] and [HEID83]), in which confidence intervals are calculated from the sequence of means of batches of size 2^v, $v = 1, 2, ...$; see comments on p.17. The batch means are introduced here not for decreasing autocorrelations among analyzed data, but for reducing the amount of data kept in memory. If 2M is the maximum number of batch means that can be stored in the buffer AnalyzedSequence, then whenever 2M means over batches of m observations are recorded, they are consolidated into M means of batches of size 2m. Subsequent observations are lumped into M successive batches of size 2m, and if more than M such new batches are needed, the rebatching procedure is repeated. Thus the buffer AnalyzedSequence can be implemented simply as a one-dimensional array of size 2M. The accuracy of estimators is measured by the relative precision ε of confidence intervals, defined in Eq. (11), and the simulation is stopped if $\varepsilon \le \varepsilon_{max}$, where ε_{max} is the acceptable maximum value of the relative precision of the final results. The current values of ε are evaluated at consecutive checkpoints ω_k ($k = 1, 2, ...$; $\omega_k \le n_{max}$), i.e., each time that $(\omega_k - \omega_{k-1})$ new observations have been collected, with $\omega_o = n_o$. To limit the number of possible checkpoints one can assume that they are geometrically distributed, i.e., for a given ω_1

$$\omega_{k+1} = \min \{ \lfloor \gamma_a(\omega_k - n_o) \rfloor + n_o, n_{max} \} , \qquad (63)$$

where $k=1,2,...$; $\gamma_a > 1$[#]. To avoid too large a distance between consecutive checkpoints one can assume $\omega_1 = \max(2M, 2n_o)$, c.f. [ASGA88]; although note that for example a value of ω_1 such that

$$\lfloor 0.1(n_{max} - n_o) \rfloor + n_o \le \omega_1 \le \lfloor 0.2(n_{max} - n_o) \rfloor + n_o \qquad (64)$$

$\lfloor x \rfloor$ denotes the "floor function " of x, that gives the greatest integer not greater than x.

was suggested in [HEID83]. The estimator of variance, and the number κ of its degrees of freedom, needed in analysis of confidence intervals, are calculated using the procedure SpectralVarianceAnalysis. It requires at least $n_v \geq 100$ batch means to be available (see Appendix). This approach is described below by a pseudocode procedure which uses the following parameters:

n_o : the number of discarded initial observations (determined by the procedure DetectInitialTransient);
n_{max} : the maximum allowed length of the simulation run, measured in the number of recorded observations ($n_{max} \geq \max(3n_o, n_o + 2M)$; to be decided in advance);
n_v : the length of the sequence used for estimating the steady state variance ($n_v \geq 100$; the default value: $n_v = 100$);
γ_a : the checkpoint incremental coefficient for sequential testing for accuracy ($\gamma_a > 1$; the default value : $\gamma_a = 1.5$);
$(1-\alpha)$: the assumed confidence level of the final results ($0<\alpha<1$; the default value: $\alpha = 0.1$);
ε_{max} : the maximum acceptable value of the relative precision of confidence intervals ($0< \varepsilon_{max}<1$; the default value: $\varepsilon_{max} = 0.1$).

procedure *SpectralAnalysis* ;
{ *sequential analysis of simulation output data based on spectral analysis of the series of collected observations* }
const M = 100; {*the default value of the minimum number of data points for the analysis of confidence intervals; $M \geq n_v \geq 100$; see Section 4.1* }

procedure *Batching* ;
{ *preparation of secondary output data ; transformation of individual observations into the sequence of no more than 2M batch means of sequentially increasing batch size* }

begin {*calculate the batch mean $\bar{X}_j(m)$, and store it as the jth data item in the buffer AnalyzedSequence* }

$\bar{X}_j(m)$:= sum/m;
if j = 2M **then** {*consolidate 2M means of batches of size m into M means of batches of size 2m*}
for s:=1 **to** M **do**

$\bar{X}_s(2m) := 0.5\ (\bar{X}_{2s-1}(m)+\bar{X}_{2s}(m))$
enddo
m := 2m; j := M
endif
j := j+1; sum :=0 {*start to calculate the next batch mean*}
end *Batching* ;

procedure *Estimation* ;
{*sequentially calculate estimates and test their precision until the required precision is reached*}
begin

find $\sigma^2[\bar{X}(n_v)]$, the variance estimator of the sequence $\bar{X}_{j-n_v}(m), \bar{X}_{j-n_v+1}(m), ..., \bar{X}_j(m)$, the last n_v batch means stored in the buffer AnalyzedSequence, and

determine κ, the number of degrees of freedom of $\sigma^2[\bar{X}(n_v)]$; {*apply the procedure SpectralVarianceAnalysis; see Appendix*}
calculate the relative half width of the confidence interval at the confidence level $(1-\alpha)$ for the current checkpoint ω_k :

$\varepsilon = t_{\kappa,1-\alpha/2} \; \hat{\sigma}[\bar{X}(n_v)] / \bar{X}(jm)$,

where $\bar{X}(jm) = \sum_{s=1}^{j} \bar{X}_s(m) / j$

is the current value of the estimated mean after jm observations, and $t_{\kappa,1-\alpha/2}$ is the upper $(1-\alpha/2)$ critical point of the t-distribution with κ degrees of freedom ;
{ *test conditions of stopping the simulation run* }
if $(\varepsilon \le \varepsilon_{max})$ **then** {*print the final results and stop the simulation*}
write('*the required precision of results has been obtained after* ', $n_o + jm$, '*observations recorded*');
StopSimulation := **true**
else{*the required precision has not been reached yet; determine the next , (k+1)st , checkpoint*}
k := k + 1;
ω_k := min ($\lfloor \gamma_a(\omega_{k-1} - n_o) \rfloor + n_o, n_{max}$)
endif
end ***Estimation*** ;

begin { *main procedure* }
m := 1 ; {*the initial batch size*}
k := 1 ; {*the initial checkpoint is after* $\omega_K = \omega_1$ *observations*}
ω_o := n_o; i := 1; {*having discarded* n_o *observations, collect next observations starting from the observation* (ω_o+i) }
ω_1 := max(2M, $2n_o$); { *the default location of the first checkpoint* }
sum := 0; j := 1; {*start calculating the 1st batch mean*}
StopSimulation := **false**; { *a condition of stopping the simulation has not been met yet* }
while (**not** StopSimulation) **do** {*collect and process new* (ω_k - ω_{k-1}) *observations*}
get the observation x_{ω_o+i} ;
sum := sum + x_{ω_o+i} ;
if (i **mod** m = 0) **then** ***Batching*** **endif**;
if (i = ω_k) **then** ***Estimation*** **endif**;
if (**not** StopSimulation) **then**
i := i+1;
if(i>n_{max} -n_o) **then**
write ('*having collected* n_{max} *observations the required precision has not been achieved; increase* n_{max} *or* α *or* ε_{max} ');
StopSimulation := **true**
endif
endif
enddo;
write ('*the final relative precision :* ',100 ε % ,'*the final* $(1-\alpha)100\%$ *confidence*

interval : ', $\bar{X}(jm)[1 \pm \varepsilon]$)
end *SpectralAnalysis* .

The next sequential procedure for simulation output analysis applies a method of analysis of weakly correlated batch means, c.f. Section 2. As in the case of the procedure SpectralAnalysis, it should be preceded by the procedure DetectInitialTransient for discarding an initial nonstationary sequence of n_0 observations. For weakening serial correlations of analyzed output data individual observations are replaced here by the less correlated means of their batches. Thus the problem of direct analysis of confidence intervals from correlated observations is replaced by the problem of determining the batch size m*, such that batch means are (almost) uncorrelated at a given level of significance. Generally, in a sequence of correlated data the autocorrelation coefficients of lag k, k = 1,2,..., (see Eqs. (20) and (21)) are not necessarily decreasing as the lag increases, although all autocorrelation coefficients are zero if the sequence contains uncorrelated observations. For this reason we follow here the test proposed by Adam [ADAM82]: a given batch size can be accepted as the batch size for approximately uncorrelated batch means if all L autocorrelation coefficients of lag k (k = 1, 2, ..., L) are statistically negligible, at a given significance level β_k, $0<\beta_k<1$. The analytical problems encountered during estimation of the autocorrelation coefficients suggest that the number of considered lags should be limited to $L=0.1\ k_{b0}$, where k_{b0} is the number of batch means tested for autocorrelation; see comments on p.12. The autocorrelation coefficients can be better estimated by the so-called jackknife estimators [MILE74], which are usually less biased than the ordinary estimators defined by Eqs. (20) and (21). A jackknife estimator of autocorrelation coefficient of lag k for a sequence of batch means of size m is calculated from the following formula:

$$\hat{\hat{r}}(k,m) = 2\hat{r}(k,m) - [\hat{r}'(k,m) + \hat{r}''(k,m)]/2 \qquad (65)$$

where the estimators on the right hand side of Eq. (64) are the ordinary estimators of autocorrelation coefficients (see Eq. (20) and (21)) but $\hat{r}(k,m)$ is the estimator over all k_{b0} batch means, while $\hat{r}'(k,m)$ and $\hat{r}''(k,m)$ estimate the same coefficient, but only over the first and the second half of the analyzed sequence of k_{b0} batch means, respectively. Let us note that:

(i) to get acceptable estimators of the autocorrelation coefficients, at least 50 batch means should be available, [BOXJ70, p.33]; thus in the case of jackknife estimators one should assume $k_{b0} \geq 100$;

(ii) to ensure approximate normality of batch means: the size of considered batches should not be less than 50, [ADAM83];

(iii) to get an acceptable overall significance level β when testing the value of L autocorrelation coefficients of lag k (k=1, 2, ...,L), each at the significance level β_k, we have to assume

$$\beta < \sum_{k=1}^{L} \beta_k \qquad (66)$$

hence in practice L should not be chosen to be too large. This restriction is, of course, irrelevant if the autocorrelation coefficients decrease monotonically with the value of the lag, since then only $\hat{r}(1,m)$ has to be considered.

To avoid wastefully collecting an excessive number of observations, especially when testing batch sizes, the procedure BatchMeansAnalysis uses two buffers for storing batch means: a buffer called ReferenceSequence is used to store the batch means $\overline{X}_1(m_o), \overline{X}_2(m_o),...$ of a batch size m_o, and a buffer AnalyzedSequence used for storing an assumed number k_{b0} of batch means of batch size $m_s = sm_o$ (s = 1, 2, ...), formed from the batch means kept in the ReferenceSequence for consecutive tests against autocorrelations. Thus, since the number of data items collected in the ReferenceSequence grows in time during a simulation run, a linked list of batch means seems to be a proper data structure for this buffer. The number of data items in the AnalyzedSequence is limited to k_{b0}, so it can be implemented as an ordinary one-dimensional array. By selecting m_o properly, we can secure a sequential increase of tested batch sizes slower than in the batching schemes proposed in [FISH73], [LAWC73] and [ADAM83], which should reduce the resultant simulation run length; also when, to neutralize the observed randomness of the estimators of correlation coefficients, $m^* = m_s$ is selected as the final batch size of weakly correlated means iff the hypothesis of all zero autocorrelation coefficients is accepted in two successive tests, both for the batch size m_{s-1} and m_s.

Having selected the batch size m^* one can sequentially analyse the accuracy of results by calculating confidence intervals from Eq. (17) and (18), which are valid for independent and identically distributed batch means. The sequence of batch means for batch size m^* kept in the buffer ReferenceSequence can have sequentially appended new batch means if more observations are needed to improve the accuracy of results.

Schmeiser [SCME82] showed that using 10 to 30 batch means over longer batches one can usually obtain more accurate results (and a better coverage of the estimators) than using more batch means but over smaller batches. Following these recommendations, when the accuracy test of the estimator from k_{be} batch means stored in ReferenceSequence fails, then these k_{be} batch means are used to form k_{bo}=30 batch means in the AnalyzedSequence buffer for the additional accuracy test. Such test is done before a new batch mean is appended to ReferenceSequence.

The whole method can be summarized by the following pseudocode procedure , which requires the following parameters:

n_o : the number of discarded initial observations (determined by the procedure DetectInitialTransient);
n_{max} : the maximum allowed length of the simulation run, measured as the number of recorded observations ($n_{max} \geq n_o + m_o k_{bo}$; to be determined in advance);
$1-\alpha$: the assumed confidence level of the final results ($0<\alpha<1$; the default value : $\alpha = 0.1$);
ε_{max} : the maximum acceptable value of the relative precision of confidence intervals ($0<\varepsilon_{max}<0.5$; the default value: $\varepsilon_{max} = 0.1$).

The initial values of other parameters are given in the procedure below.

```
procedure BatchMeansAnalysis ;
{ sequential analysis of simulation output data based on analysis of uncorrelated means of batches of
observations }
const m_o = 50; { the default value of the batch size for means stored in ReferenceSequence }
k_bo = 100; {the default value of the number of batch means stored in AnalyzedSequence }

procedure DetermineBatchSize ;
{ determine the batch size for approximately uncorrelated batch means }
begin
    s: = 1; j := 1; sum := 0; {these are the initial values of parameters for determining X̄_j(m_o), the jth
    batch mean of size m_o, at the s^th sequential step }
    while ( (not StopSimulation) and ( not Uncorrelated ) ) do
        { (i-1) observations have been already recorded }
        for v := i to i + k_bo m_o - 1 do
            { collect new k_bo m_o observations for the next k_bo batch means of size m_o to be store in
            ReferenceSequence }
            get the observation x_{n_o + v} ; { the initial n_o observations have been discarded }
            sum := sum + x_{n_o + v} ;

            if (v mod m_o = 0) then { calculate the batch mean X̄_j(m_o); the jth data item in
                ReferenceSequence }

                X̄_j(m_o) = sum / m_o ; sum := 0; j := j+1
            endif { the next batch mean has been determined }
        enddo { the next k_bo batch means have been determined }

        for v := 1 to k_bo do {consolidate sk_bo batch means X̄_1(m_o), X̄_2(mo), ..., from

            ReferenceSequence into k_bo batch means Ȳ_1(sm_o), Ȳ_2(sm_o), ..., in AnalyzedSequence,
            which will be tested for autocorrelation }

                                  s
            Ȳ_v(sm_o) := Σ  X̄_{(v-1)s+r}(m_o) / s
                         r=1
        enddo
        TestCorrelation ;
        if ( not Uncorrelated ) then
            { the batch size has not been found yet, re-initialize the data for the next autocorrelation
            tests }
            i:= i + k_bo m_o ; s:= s+1
        endif
        if ( n_o + i > n_max ) then
            {stop selecting the batch size; the test requires more than n_max observations }
            write (' the batch size of uncorrelated batch means can not be determined;
            increase n_max or β ');
            StopSimulation := true
        endif
    enddo { of search for the batch size of uncorrelated batch means }
end DetermineBatchSize ;
```

procedure ***TestCorrelation*** ;
{ *test significance of autocorrelations between batch means for a given batch size* }
const L = 10; { *the default number of autocorrelation coefficients considered ; if autocorrelation coefficients monotonically decrease with the value of lag then L=1, else L:=0.1* k_{bo} }
β = 0.1; { *the default value of the overall significance level of L tests against autocorrelation* }
begin
Correlation :=0; { *the initial value for testing L correlation coefficients* }
for k := 1 **to** L **do** { *test whether all L autocorrelation coefficients are statistically negligible each at the* β_k *significance level* }
calculate the jacknife estimator $\hat{\hat{r}}(k,sm_o)$ of the autocorrelation coefficient of lag k for the sequence $\bar{Y}_1(sm_o)$, $\bar{Y}_2(sm_o)$, ..., $\bar{Y}_{k_{bo}}(sm_o)$ stored in the AnalyzedSequence; {*apply Eq. (64) for* $m = sm_o$ }
$\beta_k := \beta / L$;
$\hat{\sigma}^2[\hat{\hat{r}}(k,sm_o)]$:= **if** (k=1) **then** $(1/ k_{bo})$
$\left[1+2 \sum_{u=1}^{k-1} \hat{\hat{r}}^2(u,sm_o)\right] / k_{bo}$)
endif {*the variance of the autocorrelation coefficient of lag k according to [BART46]*}
if ($|\hat{\hat{r}}(k,sm_o)| < z_{1-\beta_k/2}\ \hat{\sigma}[\hat{\hat{r}}(k,sm_o)]$)
{ $z_{1-\beta_k/2}$ *is the upper* $(1-\beta_k/2)$ *critical point of the standard normal distribution* }
then $\hat{\hat{r}}(k,sm_o):=0$
endif { *the lag k autocorrelation is statistically negligible at the confidence level* $1-\beta_k$ }
Correlation := Correlation + $|\hat{\hat{r}}(k,sm_o)|$;
enddo
if ((Correlation=0) **and** AcceptableSize) **then** {*accept the current batch size since the previous batch size has already given negligible correlations* }
m^* := sm_o;
Uncorrelated := **true**
{ *the batch size* m^* *of uncorrelated batch means has been selected* }
elsif (Correlation=0) **then** { *start collecting next observations for testing larger batch size* }
AcceptableSize := **true**
{ *autocorrelations for the current batch size are negligible but they were not negligible for the previous batch size , thus the next batch size should be considered* }
endif
end ***TestCorrelation*** ;

procedure ***Estimation*** ;
{*sequentially calculate estimates and test their precision until the required precision is reached* }
begin

calculate the mean $\bar{\bar{X}}(k_{be},m)$ and the relative half width ε of the confidence intervals at the confidence level $(1 - \alpha)$, where#

$\varepsilon := t_{k_{be}-1,1-\alpha/2}\ \hat{\sigma}[\bar{\bar{X}}(k_{be},m)] / \bar{\bar{X}}(k_{be},m)$,
from the whole sequence of k_{be} data items stored in ReferenceSequence; {*apply*

Eq. (16) - (18), for $k_b := k_{be}$, $m := m^*$ and $\bar{X}_j(m) := \bar{X}_j(m^*)$ }

if (($\varepsilon > \varepsilon_{max}$) **and** ($k_{be}$ **mod** 30 = 0)) **then** { *the additional test of accuracy of estimators after consolidation of k_{be} batch means from ReferenceSequence into k_{bo}=30 means of longer batches stored in AnalyzedSequence; see the recommendations in [SCME82]* }

$m := k_{be}/30$;

for j := 1 **to** 30 **do**

$$\bar{Y}_j(mm^*) := \sum_{r=1}^{m} \bar{X}_{(j-1)m+r}(m^*) / m ;$$

calculate the mean $\bar{\bar{X}}(k_{bo}, mm^*)$ and the relative half width ε of the confidence interval at the confidence level $(1 - \alpha)$, where

$\varepsilon := t_{k_{bo}-1,1-\alpha/2}\ \hat{\sigma}[\bar{\bar{X}}(k_{bo}, mm^*)] / \bar{\bar{X}}(k_{bo}, mm^*)$,
from the sequence of 30 data items stored in the AnalyzedSequence ;

{*apply Eq. (16) - (18) for $k_{bo} := 30$, $m := mm^*$, $\bar{X}_j(m) := \bar{Y}_j(mm^*)$; $t_{k,1-\alpha/2}$ is the upper $(1-\alpha/2)$ critical point of the t-distribution with k degrees of freedom* }

enddo

endif

if ($\varepsilon \le \varepsilon_{max}$) { *the required precision has been reached* } **then**

write('*the required precision of results has been obtained after* ', $n_o + k_{be}\ m^*$ '*observations recorded*');

StopSimulation := **true**

elsif ($n_o + (k_{be} + 1)\ m^* \le n_{max}$) **then**

{ *k_{be} batch means of size m^* have been used to estimate a parameter, but the required precision has not been reached yet; collect the next batch of m^* observations and store their mean in the ReferenceSequence* }

sum := 0;

for j := 1 **to** m^* **do**

get the observation $x_{k_{be}m^*+j}$;

sum := sum + $x_{k_{be}m^*+j}$

enddo

$k_{be} := k_{be} + 1$;

$\bar{X}_{k_{be}}(m^*) := sum / m^*$

elsif ($n_o + (k_{be} + 1)\ m^* > n_{max}$) **then**

{ *the required precision has not been reached ; too short the simulation run assumed* }

write('*the required accuracy can not be reached; increase n_{max} or α or ε_{max}*')

StopSimulation := **true** ;

endif

end ***Estimation*** ;

begin { *main procedure* }

StopSimulation := **false**; { *a condition of stopping the simulation has not been met yet* }

AcceptableSize := **false**; {*a batch size for uncorrelated batch means has not been found yet* }

Uncorrelated := **false**;{*the batch size for uncorrelated batch means has not been determined yet* }

$\varepsilon := 1 - \varepsilon_{max}$; {*the initial precision $\varepsilon > \varepsilon_{max}$* }

i := 1 ; { *having discarded n_o observations, collect the next observations starting from the observation (n_o+1)st* }
DetermineBatchSize ; { *if the batch size m^* has been selected then ReferenceSequence contains sk_{bo} batch means of size m^*/s and AnalyzedSequence contains k_{bo} batch means of size m^** }
k_{be} := k_{bo} ; { *the size of ReferenceSequence at the beginning of estimation* }
for j := 1 **to** k_{be} **do** { *prepare data for estimation; consolidate sk_{bo} batch means of size $m_o = m^*/s$ in ReferenceSequence into $k_{be} = k_{bo}$ batch means of size m^** }

$$\bar{X}_j(m^*) := \sum_{r=1}^{s} \bar{X}_{(j-1)s+r}(m_o) / s$$

enddo
while (Uncorrelated **and** (**not** StopSimulation)) **do**
{ *sequentially calculate and test the precision of the estimators until the required precision or the maximum length of simulation run is reached* }
Estimation ;
enddo
write (' *the final relative precision: ',100 ε%', the final 100(1- α)% confidence interval : ',* $\bar{\bar{X}}(k_{be}, m^*)\ [1 \pm \varepsilon]$);

end ***BatchMeansAnalysis*** .

In practice the last procedure employs a sequential search for the batch size of uncorrelated batch means, rather than the sequential stopping rule for the simulation run (sequential testing the precision of estimates), since usually many more batch means have to be tested during the first stage (testing against autocorrelation) than during the second stage (testing the precision of the results). The number of observations recorded, $k_{bo}m^*$, when the mean and the width of confidence intervals are to be calculated for the first time, is much larger than is usually required for obtaining the required level of accuracy, c.f. [SCME83]. From this point of view, the spectral method of analysis is more thrifty. Exhaustive comparative studies of both procedures have not been performed yet, but reported results (see for example [PAWL88]) indicate that the spectral method is the more efficient method both in the sense of the coverage recorded in reference experiments and in the sense of the final simulation run lengths for obtaining the required accuracy of the results.

5. SUMMARY AND GENERAL COMMENTS

We have discussed in detail methods for dealing with the main phenomena encountered in steady state simulation of queueing processes: the inherent initial nonstationarity and the permanent autocorrelation of collected observations. We have emphasised the methods of sequential analysis, having in mind their possible implementation in user-friendly simulation packages, which would produce results automatically. In such a context, methods of analysis based on single simulation runs seem to be more attractive than the methods of independent replications.

Our discussion was limited to simulation experiments performed to estimate a single measure of performance, the sample mean $\bar{X}(n)$ and its confidence interval (Eqs. (1) and (2)), but this methodology can be modified to estimate, for example, quantiles#. The estimation of quantiles using the method of independent replications is discussed in [WELC83, p.295], while an extension of the method of spectral analysis for this purpose is presented in [HEID84]. The same application of the regenerative method is discussed in [IGLE76] and [SEIL82].

The methodology of simultaneous analysis of more than one measure of performance in a single system is discussed for example in [LAWK82, p.308], [RAAT87] and [SCRU81], but inherent theoretical problems with reaching a satisfactory level of precision for estimators strongly limits the number of parameters for which one can determine meaningful confidence intervals. Specific statistical problems are met in *comparative simulation studies of alternative systems*. These are surveyed in [LAWK82, Ch.9]; see also [BALM87] and [FRIE86]. Note that in these applications the bias of estimators caused by an initial transient period is unimportant, so long as the estimators are equally biased.

This report is restricted to the statistical analysis of simulation output data, but practitioners are aware that that is not the only problem that must be overcome in obtaining useful results. Before observations are collected, the processes for which performance is to be investigated have to be properly modelled, and each model should be validated and verified to make the simulation experiments credible. Various aspects of the *validation and verification of simulation models* are discussed, for example, in [BANK84], [BULG82], [LAWK82], and [SHNN81].

Having presented methods of data collection and analysis which are used in stochastic steady state simulation, one can conclude that no definite conclusions can be made about their applicability. The need for more exhaustive comparative studies expressed by Schriber and Andrews in [SCRI81] seems still to be a live issue. In the light of the increasing popularity of simulation experiments conducted on small computers, the problem of accurate estimation of the length of the initial transient period, as a way of shortening simulation runs, remains an important issue as well. The length of simulation runs is also a critical issue in the case of simulation studies of complex systems, which sometimes can be performed only if some techniques for speeding-up the experiments are applied. In this context an important role could be played by the variance reduction techniques, which, by reducing the variance, narrow

---

For a given random variable X the 100 βth quantile (or percentile) x_β is defined to be the value x such that Prob[$X<x$]$\geq\beta$ and Prob [$X>x$] $>1-\beta$, $0<\beta<1$.

confidence intervals and consequently reduce the number of steps needed by sequential procedures for reaching the required accuracy of results. But, as was mentioned in Section 1, practical implementations of variance reduction techniques have been reported infrequently. Among the recently published ones are those in [FROS88], [IZYD84] and [WALR87] .

Performance evaluation studies involving simulation experiments can be accellerated by the *decomposition* of analyzed systems into subsystems, which are modelled separately, applying both simulation and analytical models, mutually interacting if necessary. Various hybrid techniques have been surveyed in [SHNT83]; see also [FROST88] and [KURO88].

The time needed for simulation studies can also be significantly reduced in multiprocessor systems. In such an environment individual processors can be used for running independent replications of one simulation experiment, or, in a more sophisticated solution, logical processes occuring during one simulation run can be executed in parallel by different processors. The main problem encountered in the latter case is the synchronization among processes run on different processors. There exists a danger that the contribution of large synchronization overheads will slow the simulation experiment. Various aspects of *distributed and parallel simulation* are discussed in [MISR86]; see also [KURO88]. Specific analytical problems accompanying distributed simulation are discussed in [HEID85].

Acknowledgements

The author wishes to thank his collegues, Rod Harries and Tim Bell, for helpful comments during preparation of this report, and Mehdi Asgarkhani and Tim Street, for testing the presented sequential procedures.

APPENDIX. Spectral Analysis of Variance.

As was mentioned in Section 2, Heidelberger and Welch ([HEID81], [HEID81a]) proposed calculating the variance of the mean of an autocorrelated sequence of observations from formulae (35) and (42), by obtaining $\hat{p}_{\bar{X}(m)}(0)$ from the smoothed averaged periodogram, see Eqs. (38) ~ (41). The series of approximations they propose can be summarised as follows:

Procedure *SpectralVarianceAnalysis* ;

{ Preconditions:
$\bar{X}_1(m), \bar{X}_2(m), ..., \bar{X}_{n_v}(m)$: a stationary sequence of n_v data points ($m \geq 1$, $n_v \geq 100$, the default value: $n_v = 100$);
n_{ap} : the number of points of the averaged periodogram used to fit it to a polynomial by applying the least squares procedure ($n_{ap} \leq n_v/4$, the default value $n_{ap} = 25$);
d : the degree of the polynomial fitted to the logarithm of the averaged periodogram (the default value d = 2);
C_σ : a normalizing constant, chosen to make $\hat{p}_{\bar{X}(m)}(0)$ approximately unbiased (for the default values of n_{ap} and d: $C_\sigma = 0.882$; see [HEID81a, Table1]) }

Step 1:
Calculate $2n_{ap}$ values of the periodogram of the sequence $\bar{X}_1(m), \bar{X}_2(m), ..., \bar{X}_{n_v}(m)$:

$$\Pi(j/n_v) = \left| \sum_{s=1}^{n_v} \bar{X}_s(m) \exp[-2\pi i(s-1)j/n_v] \right|^2 / n_v \ , \qquad (A.1)$$

{*c.f. Eqs. (38) and (39)*} for $j = 1, 2, ..., 2n_{ap}$, and $i^2 = -1$;

Step 2:
Calculate n_{ap} values of the function $\{L(f_j)\}$, $j = 1, 2, ..., n_{ap}$; where

$$L(f_j) = \log \{ [\Pi((2j-1)/n_v) + \Pi(2j/n_v)] / 2 \} \ ; \qquad (A.2)$$

Step 3:
Apply the least squares extrapolation procedure { *see for example [PRES86, p.509]* } for determining the coefficient a_o in the polynomial

$$g(f) = \sum_{s=0}^{d} a_s f^s \ , \qquad (A.3)$$

fitted to the function $\{L(f_j) + 0.270\}$, $j = 1, 2, ..., n_{ap}$;
{ *the value a_o is an unbiased estimate of log $p_{\bar{X}(m)}(0)$* }

Step 4:
Calculate $\hat{\sigma}^2[\bar{X}(n_v)] = C_\sigma e^{a_o} / n_v$;

determine κ ;

{ *the degrees of freedom* κ *for the estimator* $\hat{\sigma}^2[\bar{X}(n_V)]$, *for given* n_{ap} *and d are given in [HEID81a, Table 1] ; for* $n_V = 100$, $n_{ap} = 25$ *and* $d = 2$: $\kappa = 7$ }

end *SpectralVarianceAnalysis*

References

ADAM83 Adam, N.R. "Achieving a Confidence Intervals for Parameters Estimated by Simulation". Management Sci., 1983, pp. 856-866.

ANDE85 Anderson, K. "Some Steady-State Problems in Simulation". Record of Proceedings, Annual Simulation Symp., Tampa, Florida, IEEE Computer Society,1985, pp. 79-93.

ANDR79 Andrews, R.W., and T.J.Schriber. "Interactive Analysis of Output from GPSS-based Simulations". Proc. 1978 Winter Simulation Conf. ACM, N.Y., 1978, pp.267-278.

ARTI88 "Artificial Intelligence and Simulation: The Diversity of Applications". Proceedings of the SCS Multiconference , San Diego, Feb. 1988.

ASGA88 Asgarkhani, M., and K.Pawlikowski. " Simulation Studies of TDMA-Reservation Protocol for Sattellite Channels". (In preparation)

BALM87 Balmer, D.W. "Polishing the Analysis of the Statistical Output of Comparative Simulation Experiments". Simulation, 1987, pp.123-126.

BANK84 Banks, J., and J.S. Carson. "Discrete Event System Simulation". Prentice-Hall, Englewood Cliffs, N.Y., 1984.

BART46 Bartlett, M.S. "On the Theoretical Specification of the Sampling Properties of Autocorrelated Time Series". J. Roy. Statist. Soc. Ser. B, 1946, pp. 27-41.

BELL87 Bell, P.C., and R.M. O'Kneefe. "Visual Interactive Simulation - History, Recent Developments, and Major Issues". Simulation, 1987, pp.109-116.

BHAR84 Bharath-Kumar, K., and P. Kermani. "Performance Evaluation Tool (PET): An Analysis Tool for Computer Communication Networks". IEEE J. Select Areas Commun., 1984, pp. 229-225.

BILL68 Billingsley, P. "Convergence of Probability Measures". Wiley, N.Y., 1968.

BLAN84 Blanc, J.P.C. "The Transient Behaviour of Networks with Infinite Servers Nodes". Performance '84, ed. E. Gelenbe, North Holland, Amsterdam, 1984, pp. 159-174.

BLOM67 Blomqvist, N. "The Covariance Function of the M/G/1 Queueing System." Skandinavisk Aktuarietidskrift, 1967, pp.157-174.

BLOM70 Blomqvist, N. "On the Transient Behavior of the GI/G/1 Waiting-times". Skandinavisk Aktuarietidskrift, 1970, pp. 118-129.

BOXJ70 Box, G.E.P., and G.M. Jenkins. "Time Series Analysis, Forecasting and Control." Holden-Day, San Francisco,1970.

BRIL81 Brillinger, D.R. "Time Series: Data Analysis and Theory". Holden-Day, San Francisco, 1981.

BRAT83 Bratley, P., B.L. Fox and L.E. Schrage. "A Guide to Simulation". Springer-Verlag, N.Y. 1983.

BULG82 Bulgreu, W.G. "Discrete System Simulation". Prentice-Hall, 1982.

CARS80 Carson, J.S., and A.M. Law. "Conservation Equations and Variance Reduction in Queueing Simulations." Operations Res., 1980, pp. 535-546.

CATA87 "Catalog of Simulation Software". Simulation, 1987, pp. 165-181.

CHEN76 Cheng, R.C.H. "A Note on the Effect of Initial Conditions on a Simulation Run." Operational Research Quart., 1976, pp. 467-470.

CINL75 Cinlar, E. "Introduction to Stochastic Processes." Prentice Hall, 1975.

COHE82 Cohen, J.W. "The Single Server Queue". North Holland, Amsterdam, 1982.

CONW63 Conway, R.W. "Some Tactical Problems in Digital Simulation". Management Sci., 1963, pp.47-61.

COX61 Cox, D.R., and S.W.L. Smith. "Queues." Wiley, N.Y., 1961.

CRAN74 Crane, M. and D.L. Inglehart. "Simulating Stable Stochastic Systems: i. General Multiserver Queues." J. of the ACM, 1974, pp. 103-113.

CRAN75 Crane, M.A., and D.L. Inglehart. "Simulating Stable Stochastic Systems: III. Regenerative Processes and Discrete-Event Simulations." Operations Res.,1975, pp. 33-45.

CRAN75a Crane, M.A., and D.L. Inglehart. "Simulating Stable Stochastic Systems. IV: Approximation Techniques". Management Sci., 1975, pp.1215-1224.

CRAN77 Crane, M.A., and A.J. Lemoine. "An Introduction to the Regenerative Method for Simulation Analysis". Lecture Notes in Control and Information Science, No. 4, A.W. Balakrishnan and M. Thoma (eds), Springer-Verlag, 1977.

DALE68 Daley, D.J. "The Serial Correlation Coefficients of Waiting Times in a Stationary Simple Server Queue". J. Australian Math. Society, 1968, pp. 683-699.

DONN81 Donnelly, J.N. and R.E. Shannon. "Minimum Mean-Squared Error Estimates for Simulation Experiments". Comm. of the ACM, 1981, pp. 253-259.

DUER86 Duersch, R.R., and L.W. Schruben. "An Interactive Run Length Control for Simulations on PC's." Proc. of the 1986 Winter Simulation Conf., IEEE, 1986, pp. 866-870.

DUKE78 Duket, S.D., and A.A.B. Pritsker. "Examination of Simulation Output Using Spectral Methods." Math. Comput. Simul., 1978, pp. 53-60.

EMSH70 Emshoff, J.R., and R.L. Sisson. "Design and Use of Computer Simulation Models". Macmillan, N. Y., 1970.

FISH67 Fishman, G.S. "Problems in the Statistical Analysis of Simulation Experiments: The Comparison of Means and the Length of Sample Records". Comm. of the ACM, 1967, pp. 94-99.

FISH71 Fishman, G.S. "Estimating Sample Size in Computing Simulation Experiments". Management Sci., 1971, pp. 21-38.

FISH72 Fishman, G.S. "Bias Considerations in Simulation Experiments". Operations Res., 1972, pp. 685-790.

FISH73 Fishman, G.S. "Concepts and Methods in Discrete Event Digital Simulation". Wiley , N.Y. 1973.

FISH73a Fishman, G.S. "Statistical Analysis for Queueing Simulation." Management Sci., 1973, pp. 363-369.

FISH74 Fishman, G.S. "Estimation in Multiserver Queueing Simulations." Operations Res., 1974, pp. 72-78.

FISH77 Fishman, G.S. "Achieving Specific Accuracy in Simulation Output Analysis." Comm. of the ACM, 1977, pp.310-315.

FISH78 Fishman, G.S. "Principles of Discrete Event Simulation." Wiley, N. Y., 1978.

FISH78a Fishman, G.S. "Grouping Observations in Digital Simulation." Management Sci., 1978, pp. 510-521.

FRIE86 Friedman, L.W., and H.H. Friedman. "Comparing Simulated Alternatives Using a Distribution-Free Statistic with Blocking by Random Number Stream". Simulation, 1986, pp.68-70.

FROS88 Frost, V.S., W.W. Laure, and K.S. Shanmugan. "Efficient Techniques for Simulation of Computer Communication Networks". IEEE J. on Selected Areas in Comm., 1988, pp.146-157.

GAFA78 Gafarian, A.V., et al. "Evaluation of Commonly Used Rules for Detecting 'Steady State' in Computer Simulation". Naval Research Logistics Quart.,1978, pp. 511-529.

GEIS64 Geisler, M.A. "The Size of Simulation Samples Required to Compute Certain Inventory Characteristics with Stated Precision and Confidence". Management Sci., 1964, pp. 261-271.

GORD69 Gordon, G. "System Simulation". Prentice Hall, Englewood Cliffs, N.J., 1969.

GROS74 Gross D., and C.M. Harris. "Fundamentals of Queueing Theory". Wiley, N. Y., 1974.

GUNT80 Gunther, F.L., and R.W. Wolff. "The Almost Regenerative Method for Stochastic System Simulations." Operations Res., 1980, pp.375-386.

HAMM62 Hamming, R.W. "Numerical Methods for Scientists and Engineers". McGraw-Hill, N. Y., 1962.

HANN70 Hannan, E.J. "Multiple Time Series". Wiley, N.Y., 1970.

HEID79 Heidelberger, P. "A Variance Reduction Technique that Increases Regeneration Frequency". In "Current Issues in Computer Simulation", Academic Press, 1979, pp. 257-269.

HEID81 Heidelberger, P., and P.D. Welch. "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations". Comm. of the ACM, 1981, pp. 233-245.

HEID81a Heidelberger, P., and P.D. Welch. "Adaptive Spectral Methods for Simulation Output Analysis." IBM Res. Develop.,1981, pp. 860-876.

HEID81b Heidelberger, P. and P.A.W. Lewis. "Regression - Adjusted Estimates for Regenerative Simulations, with Graphics". Comm. of the ACM, 1981, pp. 260-273.

HEID83 Heidelberger, P., and P.D. Welch. "Simulation Run Length Control in the Presence of an Initial Transient". Operations Res., 1983, pp. 1109-1144.

HEID84 Heidelberger, P. and P.A.W. Lewis. "Quantile Estimation in Dependent Sequences". Operations Res., 1984, pp. 185-209.

HEID85 Heidelberger, P. "Statistical Analysis of Parallel Simulation". Proc. Winter Simulation Symp., Washington, Dec. 1985, pp. 290-295.

IGLE75 Iglehart, D.L. "Simulating Stable Stochastic Systems. V: Comparison of Ratio Estimators". Naval Research Logistics Quart., 1975, pp. 553-565.

IGLE76 Iglehart, D.L. "Simulating Stable Stochastic Systems VI: Quantile Estimation". J. of the ACM, 1976, pp. 347-360.

IGLE78 Iglehart, D.L. "The Regenerative Method for Simulation Analysis." In: "Current Trends in Programming Methodology", vol. III: Software Modeling, K.M. Chandy and P.T. Yeh (eds), Prentice Hall, N.J., 1978, pp. 52-71.

IZYD84 Izydorczyk, J., J. Mierzwa, and A. Wolisz. "A Control Variable Method for Increasing the Efficiency of Multiple Access Protocol Simulation". In: "Performance of Computer-Communication Systems", H. Rudin and W. Bux (eds), North Holland, 1984, pp. 95-108.

JENK68 Jenkins, G.M., and D.G. Watts. "Spectral Analysis and Its Applications". Holden-Day, San Francisco, 1968.

KELT83 Kelton, W.D., and A.M. Law. "A New Approach for Dealing with the Startup Problem in Discrete Event Simulation". Naval Research Logistics Quart., 1983, pp. 641-658.

KELT84 Kelton, W.D., and A.M. Law. "An Analytical Evaluation of Alternative Strategies in Steady-State Simulation". Operations Res., 1984, pp. 169-184.

KELT85 Kelton, W.D., and A.W. Law, "The Transient Behaviour of the M/M/S Queue,with Implications for Steady-State Simulation". Operations Res., 1985, pp. 378-396.

KELT85a Kelton, W.D. "Transient Exponential-Erlang Queues and Steady-State Simulation". Comm. of the ACM, 1985, pp. 741 - 749.

KIEF55 Kiefer, J., and J. Wolfowitz. "On the Theory of Queues with Many Servers". Tr. of the American Math. Society, 1955, pp. 1-18.

KLEI74 Kleijnen, J.P.C. "Statistical Techniques in Simulation". Marcel Dekker , N.Y. , 1974 , vol. 1.

KLEI79 Kleijnen, J.P.C. "The Role of Statistical Methodology in Simulation". In: "Methodology in Systems Modelling and Simulation", B.P.Zeigler et al. (eds), North-Holland , Amsterdam, 1979.

KLEI82 Kleijnen, J.P.C., R. Van der Ven and B. Saunder. "Testing Independence of Simulation Subruns: A Note on the Power of the von Neumann Test." Eur. J. Operational Res., 1982, pp. 92-93.

KREU86 Kreutzer, W. "System Simulation. Programming Styles and Languages". Addison Wesley , Sydney, 1986.

KURO88 Kurose, J.F., and H.T. Mouftah. "Computer-Aided Modeling, Analysis, and Design of Communication Networks". IEEE J. Select. Areas Commun., 1988, pp. 130-145.

LAVE77 Lavenberg, S.S. and C.H. Sauer. "Sequential Stopping Rules for the Regenerative Method of Simulation". IBM J. Res. Develop., 1977, pp.545-558.

LAVE81 Lavenburg, S.S., et. al. "The Initial Transient in Steady State Simulation (Panel Discussion)." 1981 Winter Simulation Conference Proc., ACM, 1981, pp. 113-120.

LAW77 Law, A.M. "A Confidence Interval in Discrete Event Simulation: A Comparison of Replication and Batch Means." Naval Research Logistics Quart., 1977, pp.667-678.

LAW80 Law, A.M. "Statistical Analysis of the Output Data from Terminating Simulations". Naval Research Logistics Quart., 1980, pp. 131-143.

LAW83 Law, A.M. "Statistical Analysis of Simulation Output Data". Operations Res., 1983, pp. 983-1029.

LAWC79 Law, A.M., and J.C. Carson. "A Sequential Procedure for Determining the Lenght of a Steady State Simulation." Operations Res., 1979, pp. 1011-1025.

LAWK82 Law, A.M., and D.W.Kelton . "Simulation, Modeling and Analysis". McGraw-Hill, N.Y. 1982.

LAWK82a Law, A.M., and D.W.Kelton . "Confidence Intervals for Steady State Simulations. II: A Survey of Sequential Procedures". Management Sci., 1982, pp. 550-562.

LAWK84 Law, A.M., and W.D. Kelton. "Confidence Intervals for Steady State Simulations. I: A Survey of Fixed Sample Size Proceaures." Operations Res., 1984, pp. 1221-1239.

MADA76 Madansky, A. "Optimal Initial Conditions for a Simulation Problem". Operations Res., 1976, pp. 172-577.

MARK81 Marks, N.B. "Further Investigation into Spectral Analysis for Confidence Intervals in Steady State Simulation." Proc. 1981 Winter Simulation Conf., IEEE, 1981, pp. 461-464.

MEKE82 Meketon, M.S., and P. Heidelberger. "A Renewal Theoretic Approach to Bias Reduction on Regenerative Simulations". Management Sci., 1982, pp. 173-181.

MILE74 Miller, R.G. "The Jackknife - A Review". Biometrika, 1974, pp. 1-15.

MILS87 Mills, R. "Statistical Analysis of Steady State Simulation Output Data with SIMSCRIPT II.5." CACI, Los Angeles, 1987.

MINH87 Minh, D.L. "Simulating GI/G/k Queues in Heavy Traffic". Management Sci., 1987, p p. 1192-1199.

MISR86 Misra, J. "Distributed Discrete Event Simulation". ACM Computing Surveys, 1986, pp. 39-66.

MITR 82 Mitrani, I. "Simulation Techniques for Discrete Event Systems". Cambridge Academic Press, 1982.

MORS55 Morse, P.M. "Stochastic Processes of Waiting Lines". Operations Res. , 1955, pp. 255-261.

NEWE71 Newell, G.F. "Applications of Queueing Theory". Chapman & Hall, London, 1971.

NIEL86 Nielsen, N.R. "Expert Systems and Simulation". In: "Computer Networks and Simulation III ". S. Schoemaker (ed.), North-Holland, 1986, pp. 41-52.

ODON83 Odoni, A.R., and E. Roth. "An Empirical Investigation of the Transient Behaviour of Stationary Queueing Systems". Operations Res., 1983, pp. 432-454.

OKEE86 O'Keefe, R.M. "Simulation and Expert Systems - A Taxonomy and Some Examples". Simulation, 1986, pp. 10-16.

OREN87 Ören, T.I., and B.P. Zeigler. "Artificial Intelligence in Modelling and Simulation: Directions to Explore". Simulation, 1987, pp. 131-134.

PARZ61 Parzen, E. "Mathematical Consideration in the Estimation of Spectra". Technometrics, 1961, pp. 167-190.

PAWL88 Pawlikowski, K., and M. Asgarkhani. "Sequential Procedures in Simulation Studies of Satellite Protocols". Proc. ITC'12, North-Holland, 1988 (in printing).

PAYN82 Payne, J.A. "Introduction to Simulation". McGraw-Hill, N.Y., 1982.

PIDD84 Pidd , M. " Computer Simulation in Managment Science ". Wiley , Chichester , 1984.

POTI84 Potier, D. "New User's Introduction to QNAP2". Rapport Technique INRIA, No. 40, 1984.

POTI86 Potier, D. "The Modeling Package QNAP2 and Applications to Computer Networks Simulation." In: "Computer Networks and Simulation III", North Holland, 1986, pp.235-265.

PRES86 Press, W.H., et al. "Numerical Recipes. The Art of Scientific Computing." Cambridge University Press, 1986.

RAAT87 Raatikainen, K.E.E. "Simultaneous Estimation of Several Percentiles". Simulation, 1987, pp. 159-164.

REDD87 Reddy, R. "Epistemology of Knowledge Based Simulation". Simulation, 1987, pp. 162-166.

ROTH85 Roth, E., and A.H. Rutan. "A Relaxation Time Approach for Reducing Initialization Bias in Simulation" . Record of Proceedings, Annual Simulation Symp., Tampa, Florida, IEEE Computer Society,1985, pp. 189-203.

RUIZ87 Ruiz-Mier, S., and J. Talavage. "A Hybrid Paradigm for Modeling of Complex Systems". Simulation, 1987, pp. 135-141.

SAUE82 Sauer, C.H., E.A. MacNair and J.F. Kurose. "The Research Queueing Package: Past, Present and Future." Proc. National Computer Conf., Houston, 1982, pp.273-280.

SAUE84 Sauer, C.H., E.A. MacNair and J.K. Kurose. "Queueing Network Simulations of Computer Communications". IEEE J. Select Areas Commun. 1984, pp. 203-220.

SCME82 Schmeiser, B. "Batch Size Effects in the Analysis of Simulation Output." Operations Res., 1982, pp.556-568.

SCRI74 Schriber, T. "Simulation Using GPSS". Wiley, N.Y., 1974.

SCRI79 Schriber, T.J., and R.W. Andrews. "Interactive Analysis of Simulation Output by the Method of Batch Means." Proc. 1979 Winter Simulation Conference, ACM, 1979, pp. 512-524.

SCRI81 Schriber, T.J., and R.W. Andrews. "A Conceptual Framework for Research in the Analysis of Simulation Output." Comm. of the ACM, 1981, pp. 218-232.

SCRU80 Schruben, L.W. "A Coverage Function for Interval Estimators of Simulation Response." Management Sci., 1980, pp. 18-27.

SCRU81 Schruben, L.W. "Control of Initialization Bias in Multivariate Simulation Response". Comm. of the ACM, 1981, pp. 246-252.

SCRU82 Schruben, L.W. "Detecting Initialisation Bias in Simulation Output". Operations Res., 1982, pp. 569-590.

SCRU83 Schruben, L.W., H. Singh and L. Tierney. "Optimal Tests for Initialisation Bias in Simulation Output". Operations Res., 1983, pp. 1167-1178.

SCRU83a Schruben, L. "Confidence Interval Estimation Using Standardised Time Series." Operations Res., 1983, pp.1090-1108.

SEIL82 Seila, A.F. "A Batching Approach to Quantile Estimation in Regenertive Simulation". Management Sci., 1982, pp. 573-581.

SHNN81 Shannon, R.E. "Test for the Verification and Validation of Computer Simulation Models," Proc. 1981 Winter Simulation Conference, IEEE, 1981, pp. 573-577.

SHNT83 Shanthikumar, J., and R. Sargent. "A Unifying View of Hybrid Simulation/Analytic Models and Modeling". Operations Res., 1983, pp. 1030-1052.

SHAP65 Shapiro, S.S., and M.B. Wilk. "An Analysis of Variance Test for Normality (Complete) Samples". Biometrika, 1965, pp. 591-611.

SOLO83 Solomon, S.L. "Simulation of Waiting-Line Systems". Prentice-Hall, Englewood Cliffs, 1983.

STAI88 Stairmand, M.C., and W.Kreutzer. "POSE: A Process-Oriented Simulation Environment Embedded in Scheme". Simulation, 1988, pp.143-153.

TRIV82 Trivedi, K.S. "Probability and Statistics with Reliability, Queueing, and Computer Science Applications". Prentice-Hall, Englewood Cliffs, 1982.

TURN77 Turnquist, M.A., and J. M. Sussman. "Toward Guidelines for Designing Experiments in Queuing Simulation". Simulation, 1977, pp. 137-144.

WAHB80 Wahba, G. "Automatic Smoothing of the Log Periodogram." J. Amer. Statist. Association, 1980, pp. 122-132.

WALR87 Walrand, J. "Quick Simulation of Queueing Networks". Proc. 2nd Workshop on Appl. Math. and Performance/Reliability Models of Comput./Commun. Systems, Rome, Italy, 1987, pp. 275-286.

WELC83 Welch, P.D. "Statistical Analysis of Simulation Results". Chapter 6 in : "Computer Performance Modelling Handbook", S.S. Lavenberg (ed.), Academic Press, 1983, pp. 287 - 329.

WILS78 Wilson, J.R., and A.A.B. Pritsker. "A Survey of Research on the Simulation Startup Problem". Simulation, 1978, pp. 55-58.

WILS78a Wilson, J.R., and A.A.B. Pritsker. "Evaluation of Startup Policies in Simulation Experiments". Simulation, 1978, pp. 79-89.