Proceedings of the 1995 Winter Simulation Conference
ed. C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman

ANALYZING MULTIVARIATE OUTPUT

John M. Charnes

School of Business
University of Kansas
Lawrence, Kansas 66045–2003

ABSTRACT

This paper gives an overview of multivariate statistical techniques that can be useful for analyzing discrete-event simulation output, and describes some of the latest directions in research on multivariate output analysis. A general discussion is given of constructing joint confidence regions on the mean vector of multivariate output from independent replications of terminating models. The multivariate batch means method of simultaneous estimation of means from one long run of steady-state simulation models is described. References are also given for autoregressive, spectral analysis and regenerative methods of inference, as well as variance-reduction and sequential techniques.

1 INTRODUCTION

There has been considerable activity recently by researchers on the problem of simultaneously making statistical inferences on more than one output measure of interest in simulation modeling (Seila 1982, 1983, 1984, 1990; Chen and Seila 1987; Chen and Chen 1988; Chen and Cheng 1989; Chen 1991; Munoz 1991; Yang and Nelson 1988, 1992; Charnes 1990, 1991; Charnes and Kelton 1988, 1993; Raatikainen 1993; Gallagher, Bauer and Maybeck 1994). The intent of this paper is to describe some of the latest directions of research in this area. It will attempt to highlight some of the important multivariate statistical techniques that may be found useful in analyzing simulation output.

The methods presented here will be of most interest to those analysts wishing to extract more information from their simulation models. Novice analysts looking for basic information on simulation output analysis should consult simulation textbooks, such as Bratley, Fox and Schrage (1987) or Law and Kelton (1991), or one of the tutorial papers published in previous *Proceedings of the Winter Simulation Conference* and the references therein. This paper is an update of Charnes (1991).

The next section discusses multivariate output from simulation models and contrasts multivariate analysis to univariate. Section 3 discusses terminating simulation models. Section 4 discusses steady-state models. Section 5 concludes the paper and gives references to more advanced techniques of multivariate output analysis.

2 MULTIVARIATE OUTPUT

Most simulation models produce outputs on more than one measure of interest, and these outputs are usually cross-correlated as well as being autocorrelated. If cross correlation of the output measures is important to the simulation analyst, a multivariate technique should be used with the output data generated by the simulation model. Two examples illustrate the usefulness of considering multivariate output from simulation models.

Example 1: Bank Lobby Layout. A bank manager is considering changing the present configuration (Layout 1) of the teller windows in the lobby from one in which both private and corporate customers are served by any of the available tellers, to one in which certain tellers serve only private customers, and certain tellers serve only corporate customers (Layout 2). The two different layouts are illustrated schematically in Figure 1. Corporate customers are represented by the crosses ($\times$) and private customers are represented by the open circles ($\circ$). The bank manager is willing to change the lobby layout if it decreases the time spent waiting in the bank by corporate customers, even if it means the time spent waiting by private customers increases by a small amount. To help make the decision, the manager wants to know the correlation between the average times spent waiting by both types of customers, because she feels that private cus-

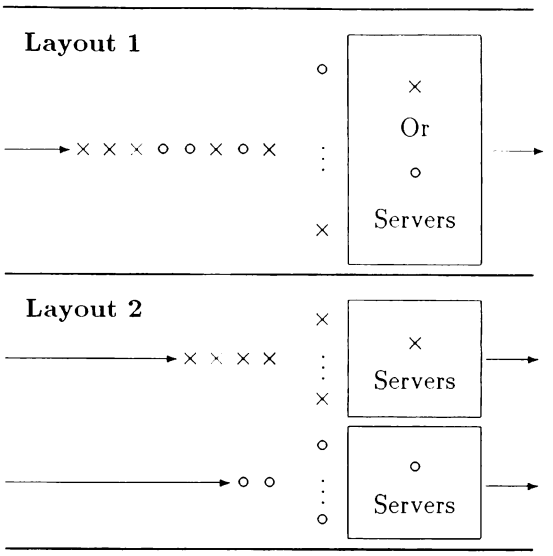

Figure 1: Two Bank Lobby Layouts

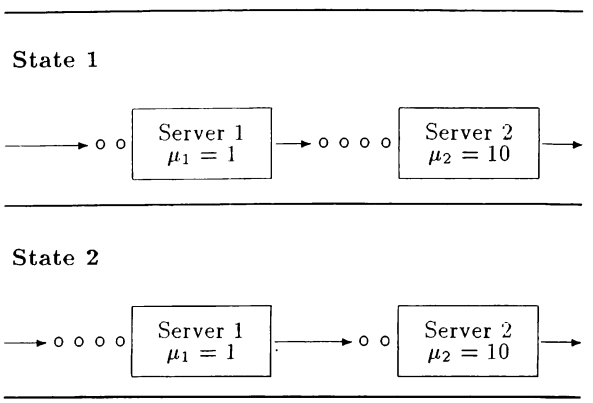

Figure 2: Tandem Queueing System

tomers may be more tolerant of slightly longer delays if they observe corporate customers experiencing long delays when they do.

Example 2 Tandem Queueing System. Two states of a simple tandem queueing system are shown in Figure 2. The customers, depicted as open circles (o), arrive to the system and wait on line, if necessary, to be served individually by Server 1. The customers then proceed to Server 2, and wait on line there, if necessary, to be served individually by Server 2, after which they depart from the system. Server 1 has a mean service rate $\mu_1 = 1$ customer per unit time, while Server 2 has a mean service rate of $\mu_2 = 10$ customers per unit time. If only the total number in system is observed, the two states appear to be identical; in both State 1 and State 2, there are six customers in the system. However, the difference between the two states is quite noticeable to an arriving customer who occupies the last spot on line in Server 1's queue. In State 1, which has only two customers at Server 1 (the slower server), the customer is likely to get through the system much more quickly than in State 2, which has four customers at Server 1. Thus by looking only at univariate output data (such as total number in system), rather than multivariate (such as the 2-dimensional vector of number of customers at each server), a simulation analyst might miss important information about the system that could be useful for making decisions.

For example, if this simple system represented some portion of a factory, and the factory configuration were such that the queues at Server 1 and Server 2 were able to share plant-floor space, the plant manager may well be interested in the correlation between the numbers in queue. Frequent occurrences of the numbers in queue being large simultaneously (indicated by a large positive correlation) could require the allocation of more floor space to the servers' queues.

By using multivariate statistical methods with the data obtained from valid simulation models, decision makers can extract more information from which to make inferences on the processes being modeled. Constructing multivariate confidence regions on the mean vector of the data-generating process is one way to summarize information about each of the univariate processes composing the multivariate process, as well as the correlations among processes. The next two sections describe techniques for constructing confidence regions that could be applied to the two examples given above.

3 TERMINATING MODELS

There are two different types of discrete-event simulation models that call for different basic approaches to experimental design as well as to constructing confidence regions on the mean. In the *terminating* simulation case, where the system being modeled has specific start-up and shut-down times (e.g., the bank described in Example 1, which opens at 9 A.M. and closes at 3 P.M.), the simulation analyst can make independent replications of the model, each representing one complete succession from start up to shut down.

If the simulation analyst calculates point estimates of the parameters of interest from each replication (such as the averages of the time spent in the bank by

the private and the corporate customers), the result will be a sequence of independent and identically distributed (iid) random vectors that can be analyzed using classical multivariate statistical methods. By viewing the output as vectors and using multivariate methods, rather than analyzing the components of the vectors separately with univariate statistical methods, the analyst can get an estimate of the correlations among the vector components that can provide useful information to the decision maker about the process being modeled.

3.1 Joint Confidence Regions

One multivariate technique that can be applied is the construction of a joint confidence region on the mean vector. The procedure is based on Hotelling's T^2 distribution, and is the generalization to higher dimensions of the univariate t-distribution confidence-interval procedure for a single mean. The validity of the procedure rests upon the assumption of multivariate normality of the data. See an introductory multivariate statistics text such as Anderson (1984), Johnson and Wichern (1988), or Morrison (1976) for a fuller discussion of this procedure.

Consider a simulation model that is replicated R times, and that has D measures of interest. The observations are denoted by $\mathbf{X}^{(r)} = \left(X_1^{(r)}, \ldots, X_D^{(r)}\right)'$ ($'$ denotes matrix transposition), where $X_d^{(r)}$ is the value of the dth measure of interest on the rth replication. The measure of interest might be the average cycle time of a specific product, time-average number in a selected queue, or some other point estimator calculable from each replication. The analyst wishes to construct a confidence region on the true mean vector of the parameters $\vec{\mu} = E\left[\mathbf{X}^{(r)}\right] = (\mu_1, \ldots, \mu_D)'$.

To form the confidence region, first find $\overline{\mathbf{X}} = \frac{1}{R}\sum_{r=1}^{R} \mathbf{X}^{(r)} = (\sum_{r=1}^{R} X_1^{(r)}/R, \ldots, \sum_{r=1}^{R} X_D^{(r)}/R)'$. The vector of point estimators $\overline{\mathbf{X}}$ is an unbiased estimator of the mean vector $\vec{\mu}$. An unbiased estimate of the variance-covariance matrix of $\overline{\mathbf{X}}$ is $\mathbf{S} = (1/(R-1))\sum_{r=1}^{R}(\mathbf{X}^{(r)} - \overline{\mathbf{X}})(\mathbf{X}^{(r)} - \overline{\mathbf{X}})'$, and a $100(1-\alpha)\%$ confidence region for the true mean vector of the parameters of interest is given by the set of all vectors $\mathbf{\Theta}$ such that $(\overline{\mathbf{X}} - \mathbf{\Theta})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \mathbf{\Theta}) \leq (D(R-1)F_{\alpha,D,R-D})/(R(R-D))$ where $F_{\alpha,D,R-D}$ is the upper (100α)th percentile of the F distribution with D and $R-D$ degrees of freedom.

With two parameters of interest, the confidence region can be plotted as an ellipse in two-dimensional space. For three parameters, the region is a three-dimensional ellipsoid. For more than three parameters, the region cannot be plotted; however, it is a straightforward calculation to check whether any given vector will be in the confidence region, so that one can easily check for combinations of parameters that are undesirable (such as short corporate customer delays and long private customer delays).

The shape and orientation in parameter space of the ellipsoid depends upon the magnitudes and algebraic signs of the off-diagonal terms of the matrix $\mathbf{S}$. Because it is the relative magnitude of the off-diagonal elements that is important, it is informative to compute the correlation matrix, $\mathbf{C}$, for the mean vector. The correlation matrix is calculated as $c_{ij} = s_{ij}/\sqrt{s_{ii}s_{jj}}$, where s_{ij} is the (i,j)th element of $\mathbf{S}$. The element c_{ij} of $\mathbf{C}$ gives the correlation between point estimator i and point estimator j and thus will be such that $-1 \leq c_{ij} \leq 1$.

Note that the validity of this procedure rests upon the assumption of multivariate normality and independence of the vector observations taken from each replication. Obtaining independent vector observations from replications in simulation modeling is not usually a problem, and averaging over each replication will tend to make the point estimates normally distributed. However, the analyst should be aware that the validity of this procedure rests upon these two assumptions.

3.2 Simultaneous Confidence Intervals

Because joint confidence regions can be difficult to interpret, the analyst might want to construct individual confidence intervals on the mean of each component process in the output vector. Two methods for doing so are Scheffé and Bonferroni intervals.

3.2.1 Scheffé Intervals

Scheffé confidence intervals (also known as Roy-Bose intervals) are the shadows of the ellipsoidal confidence region on the coordinate axes. Using the elements of the matrix $\mathbf{S}$ given above, the Scheffé intervals are given by $\overline{x}_i \pm \sqrt{(D(R-1)F_{\alpha,D,R-D}s_{ii})/(R(R-D))}$ for $i = 1, \ldots, D$, where $\overline{x}_i$ is the ith element of the vector $\overline{\mathbf{X}}$.

These intervals define a rectangular region that circumscribes the ellipsoid given above. Regardless of the dependence structure of the individual estimators, the overall level of confidence that the Scheffé intervals cover their respective means will thus be greater than or equal to $(1-\alpha)$ when the assumptions above hold for the ellipsoidal region.

The Scheffé intervals are easier to interpret than the ellipsoidal region because they yield a range for each component of the output vector. However, because they utilize all the information contained in the variance-covariance matrix, $\mathbf{S}$, they require more

computation than do the confidence intervals resulting from univariate procedures.

3.2.2 Bonferroni Intervals

A naive approach to analyzing simulation output is to use univariate procedures on each output measure without being aware of the limitations of doing so. One must be careful in constructing more than one confidence interval from simulation output with the usual univariate procedures because the overall level of confidence that all intervals with the same nominal confidence level will cover their respective means is less than the nominal confidence level of each interval. The exact amount less is usually difficult to determine and is affected by the dependence structure of the univariate estimators; however, the Bonferroni Inequality yields a simple method of setting the individual confidence levels for a group of univariate procedures in order to obtain a lower bound on the overall level of confidence in the set of inferences.

Let $\Pr(C_i \text{ true}) = 1 - \alpha_i$ for $i = 1, \ldots, D$ where C_i denotes a confidence statement about the mean value of the ith component of the output vector. The Bonferroni Inequality holds that $\Pr(\text{all } C_i \text{ true}) \geq 1 - (\alpha_1 + \alpha_2 + \cdots + \alpha_D)$. This result is often used as follows. If each one of D confidence intervals is constructed at the $1 - \alpha/D$ level, then the overall level of confidence is at least $1 - \alpha$ that all D parameters lie in the D-dimensional box defined by these confidence intervals.

If the vector $\overline{\mathbf{X}}$ and the matrix $\mathbf{S}$ have been calculated as described above, individual Bonferroni Intervals can be constructed on each component mean as follows: $\overline{x}_i \pm t_{\alpha/2D;R-1}\sqrt{s_{ii}/R}$ for $i = 1, \ldots D$ where $t_{\alpha/2D;R-1}$ is the upper $(100\alpha/(2D))$th percentile of the t distribution with $(R-1)$ degrees of freedom.

An advantage of using Bonferroni Intervals rather than Scheffé is their ease of construction. Many commercial simulation software packages provide one or more methods of obtaining confidence intervals using standard univariate procedures. If so, Bonferroni Intervals can be constructed by using one of the available methods with an appropriate choice of the confidence coefficient. The disadvantage is that for large D the intervals may be very wide, and thus not very precise. Further, the same caveats in regard to the independence and normality assumptions given previously for joint confidence regions apply here. The actual lower bound on the coverage probability for the Bonferroni method depends upon the true coverage probabilities of the individual confidence intervals. If the univariate confidence intervals do not obtain their nominal coverage individually, then the Bonferroni Inequality will not ensure that the bound on the nominal coverage given by the inequality will obtain.

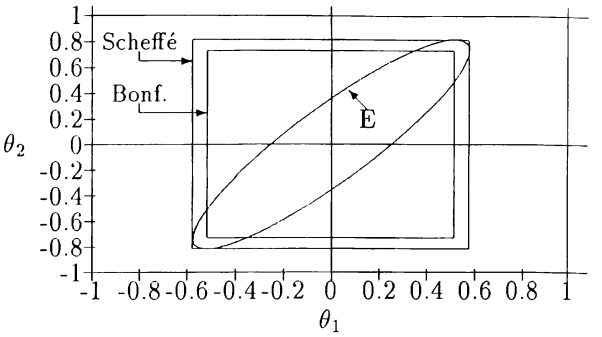

Figure 3: Comparison of Confidence Regions

Figure 3 illustrates the three confidence regions described above. The Scheffé and Bonferroni regions are labeled and the ellipse labeled "E" represents an ellipsoidal region obtained as described in Section 3.1. Similar comparisons have been given by Miller (1981) and Chen (1991).

The three 90% confidence regions depicted were constructed from the variance-covariance matrix $\mathbf{S} = \begin{pmatrix} 1 & 1.27 \\ 1.27 & 2 \end{pmatrix}$ for $R = 22$. Without loss of generality, the sample mean vector was taken to be $\overline{\mathbf{X}} = \mathbf{0}$. The plot shows the typical pattern in that the Scheffé region is larger than the Bonferroni region. It also shows that portions of the ellipsoidal confidence region are not covered by the Bonferroni Intervals.

4 STEADY-STATE MODELS

In a *steady-state* simulation, the system being modeled has no specific "start-up" or "shut-down" times. An example is the simulation of a factory that operates twenty-four hours a day, seven days a week. In cases like this, the simulation analyst is most often interested in estimating steady-state parameters of the model. That is, the analyst assumes that if the model is in operation long enough, it will reach a state of statistical equilibrium, in which the means, cross covariances, and autocovariances (defined below) of the output process will be invariant to the passage of simulated time.

As in the univariate case, if the initial conditions for the simulation are not representative of steady-state, the simulation must be allowed to "warm up" by run-

ning for a suitable length of time to mitigate any bias induced by the non-representative initial conditions. Schruben (1981) and Gallagher, Bauer, and Maybeck (1994) give multivariate methods for deciding when the simulation appears to have reached steady state.

Once the initial transient observations have been identified, they are usually discarded and the remaining observations are analyzed. However, as in the univariate case, the autocorrelation problem makes it more difficult to analyze these data than the serially independent data obtained from terminating models. This section deals with analyzing multivariate output obtained from steady-state models.

4.1 Autocorrelation Function

Assume that the initial transient observations generated by a simulation model have been discarded. In general, if a simulation model produces the stationary sequence of D-dimensional vector observations $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_T\}$, where $\mathbf{X}_t = (X_{1t}, X_{2t}, \ldots, X_{Dt})'$ and $E[\mathbf{X}_t] = \vec{\mu} = (\mu_1, \mu_2, \ldots, \mu_D)'$, the output vectors will not be iid. The dependence among the elements and across time is characterized by the *autocovariance function*, $\mathbf{\Gamma}(h) = E[(\mathbf{X}_t - \vec{\mu})(\mathbf{X}_{t+h} - \vec{\mu})']$, which is a function of only the lag, h, for a stationary sequence. For univariate processes, the autocovariance function is a scalar function of h but for multivariate processes, $\mathbf{\Gamma}(h)$ is a matrix. The autocovariance of the ith component of the vector output sequence is given by the corresponding diagonal element in $\mathbf{\Gamma}(h)$, which is denoted $\gamma_{ii}(h)$. The cross covariances are given by the off-diagonal terms in the autocovariance function, $\gamma_{ij}(h)$ $(i \neq j)$.

In practice, if the observations are simultaneous (all elements of the observation vector are taken at the same point in simulated time), and equally spaced in simulated time, it may be informative to compute the *sample autocorrelation function*, $\mathbf{R}(h)$, which is a normalized version of the sample autocovariance function, $\mathbf{G}(h)$, calculated from the simulation output. The sample autocovariance function is found from the data as $\mathbf{G}(h) = \sum_{t=1}^{T-h}(\mathbf{X}_t - \overline{\mathbf{X}})(\mathbf{X}_{t+h} - \overline{\mathbf{X}})'/(T-h)$ for $h = 0, 1, 2, \ldots, T-1$. For large h, the estimates will be calculated from only a few observations, and thus may be poor; however, much insight can be gained from calculating these matrices for small lags (e.g., $h = 0, 1, 2, 3$). Then the sample autocorrelation function is computed from the elements of the autocovariance function as $r_{ij}(h) = g_{ij}(h)/\sqrt{g_{ii}(0)g_{jj}(0)}$. Because these are correlations, it will be true that $-1 \leq r_{ij}(h) \leq 1 \quad \forall i, j, h$.

The sample autocorrelation function may reveal important information about the dependence structure of the processes being modeled. For example, a model of a factory with ten work centers on which a 10-dimensional vector of numbers at each work center is observed at equally spaced time periods will yield a (10×10) autocorrelation matrix for each lag, h, that will indicate how much a work center downline may be affected by backups at previous work centers. High values of $r_{ij}(0)$ for instance, will tell the analyst that the relative (to the mean) number at work center j will follow closely the relative number at work center i. High values of $r_{ij}(h)$ will indicate that high (low) numbers at work center i will tend to be followed by high (low) numbers at work center j, but not until a lag of h time units later. The matrix autocorrelation function may be worthwhile calculating for only this reason—it gives the analyst more information about the characteristics of the operation.

4.2 Multivariate Batch Means

Consider a stationary process that produces a sequence $\{\mathbf{X}_t\}_{t=1}^{T}$ of D-dimensional vector-valued observations. The analyst wishes to estimate the mean vector $\vec{\mu} = E(\mathbf{X}_t)$ with a joint confidence region. The multivariate batch means (MBM) method calls for dividing the sequence of output vectors into B batches of M (vector) observations each (where $T \geq BM$) and computing the batch-mean vectors as $\mathbf{Y}_b = \sum_{m=1}^{M} \mathbf{X}_{(b-1)M+m}/M$ for $b = 1, \ldots, B$. The B vectors of batch means are then treated as if they are uncorrelated observations from a multivariate normal distribution, and standard multivariate statistical techniques are used to form a confidence region on the mean vector, $\vec{\mu}$, just as for terminating models. Let $\mathbf{S} = \sum_{b=1}^{B}(\mathbf{Y}_b - \overline{\mathbf{X}})(\mathbf{Y}_b - \overline{\mathbf{X}})'/(B-1)$ denote the sample variance-covariance matrix for the batch mean vectors, where the point estimator of $\vec{\mu}$ is the D-dimensional vector $\overline{\mathbf{X}} = \sum_{i=1}^{m} \mathbf{Y}_i/m$. An approximate $100(1-\alpha)\%$ confidence region for $\vec{\mu}$ is then given by the set of all vectors $\mathbf{\Theta}$ such that $(\overline{\mathbf{X}} - \mathbf{\Theta})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \mathbf{\Theta}) \leq (D(B-1)F_{\alpha;D,B-D})/(B(B-D))$. Chen and Seila (1987) proposed the use of the multivariate batch means method at a previous Winter Simulation Conference.

An important step in forming the MBM confidence region is the determination of the number of vector observations per batch, M (equivalently, the number of batches, B). One method of making this determination is to assume that the batch-means process can be sufficiently approximated by the VAR(1) (first-order, vector-autoregressive) model $\mathbf{Y}_b = \mathbf{\Phi}\mathbf{Y}_{b-1} + \epsilon_b$ for $b = 1, \ldots, B$, where $\mathbf{\Phi}$ is a $D \times D$ matrix of autoregression coefficients and the ϵ_b are $D \times 1$ independent and identically distributed vectors of random

errors drawn from the multivariate normal distribution. Then M is chosen such that the null hypothesis of no first-order serial correlation, H_0: $\Phi = \mathbf{0}$, is not rejected.

Anderson (1978) suggests H_0: $\Phi = \mathbf{0}$ can be tested with one of the criteria given in Anderson (1984) for testing the general linear hypothesis. Charnes (1990) compared these criteria, which include the Lawley-Hotelling trace criterion, the Bartlett-Nanda-Pillai criterion, and three slightly different forms of the Wilks likelihood-ratio criterion. Because the batch sizes are chosen upon *not* rejecting H_0, the criteria were compared on their power to detect departures from H_0. The conclusion from the comparison was that there is little difference in the power of each of these statistics. However, Rao's (1951) approximation to the Wilks likelihood-ratio procedure can be recommended because (*i*) it has a degrees-of-freedom correction, which makes it appropriate for a small number of batches containing a large number of vectors, and (*ii*) the critical value for the hypothesis test comes from the F distribution rather than tabulated values in Anderson (1984), which makes it possible to calculate the critical value with a computer algorithm, thus making it amenable to inclusion in a software package that automates simulation output analysis.

Rao's approximation to Wilks's procedure uses $R = [(ks-r)/D^2] \cdot [(1-U^{1/s})/U^{1/s}]$ where the scalar $U = |\mathbf{S}(0) - \mathbf{S}(1)\mathbf{S}^*(0)^{-1}\mathbf{S}(1)'|/|\mathbf{S}(0)|$, the $D \times D$ matrix $\mathbf{S}^*(0) = \sum_{b=1}^{B-1}(\mathbf{Y}_b - \overline{\mathbf{X}})(\mathbf{Y}_b - \overline{\mathbf{X}})'$, the $D \times D$ matrix $\mathbf{S}(0) = \sum_{b=2}^{B}(\mathbf{Y}_b - \overline{\mathbf{X}})(\mathbf{Y}_b - \overline{\mathbf{X}})'$, the $D \times D$ matrix $\mathbf{S}(1) = \sum_{b=2}^{B}(\mathbf{Y}_b - \overline{\mathbf{X}})(\mathbf{Y}_{b-1} - \overline{\mathbf{X}})'$, and the scalars $s = \sqrt{(D^4-4)/(2D^2-5)}$, $r = D^2/2 - 1$, and $k = B - 1/2$. R has approximately the F distribution with D^2 and $ks-r$ degrees of freedom (Anderson 1984). A procedure for selecting the number and size of the batches is to begin with some maximum number of batches, $B = B_{\max}(D)$, compute the sequence of batch means, $\{\mathbf{Y}_b\}_{b=1}^{B}$, and test H_0 with R. If H_0 is rejected, decrease B, compute the new sequence $\{\mathbf{Y}_b\}_{b=1}^{B}$, and test H_0 again with R. Continue until either H_0 is not rejected, or until a prespecified minimum number of batches, $B_{\min}(D)$, is reached. If H_0 is not rejected, compute the region given above. If the minimum number of batches is reached, run the simulation for a longer amount of time, and repeat this procedure.

Schmeiser (1982) provides guidelines for choosing the minimum and maximum number of batches for the univariate batch means method ($D = 1$), but an open topic for further research is the best choice of the minimum and maximum number of batches for multivariate analysis. Yang and Nelson (1992) give some guidelines for choosing $B_{\min}(D)$ for $D \leq 5$; however, little else has been published to date on this topic.

4.3 Simultaneous Confidence Intervals

As in the terminating simulation case, an analyst will probably want to construct individual confidence intervals on the true mean of each component process. One way to accomplish this is to use the univariate batch means method with each component process taken individually (see Schmeiser 1982), while being mindful of the Bonferroni Inequality when choosing the percentile of the t distribution used to construct each interval.

An alternative is to use the elements of $\mathbf{S}$ from the MBM method to compute $\overline{x}_d \pm t_{\alpha/(2D),B-1}\sqrt{s_{dd}/B}$ for $d = 1, \ldots, D$. Note that by using this alternate method, the analyst is forcing the batch sizes to be the same for each component process. This is not necessarily true when the univariate batch means technique is applied to each process individually. However, by calculating the matrix $\mathbf{S}$ from the MBM method, the analyst can also get an estimate of the correlation among the estimators of the means. This will not be true, in general, for any univariate method applied individually to the component processes.

4.4 Advanced Techniques

More advanced multivariate techniques have been proposed for analyzing data generated by steady-state simulation models.

4.4.1 Autoregressive Models

Charnes and Kelton (1993) use a vector autoregressive (VAR) model to obtain confidence regions on the mean vector, $\vec{\mu}$. This approach differs from the MBM method in that the VAR method uses the information contained in the autocorrelation of the output, while the MBM method attempts to eliminate the autocorrelation by batching. The empirical evidence shows that VAR compares favorably with MBM and other methods.

4.4.2 Spectral Analysis

Kabaila and Nelson (1985) give a frequency domain time-series technique that was used to make inferences on the earth's mean atmospheric response to external forcing. The method uses an estimate of the spectral density function at frequency zero for constructing a confidence region on $\vec{\mu}$. Charnes and Kel-

ton (1993) compare this method empirically to MBM and VAR.

4.4.3 Standardized Time Series

Munoz (1991) extends the method of standardized time series, initially proposed to estimate a single steady-state mean, to the case where simultaneous inferences on the components of $\vec{\mu}$ are desired.

4.4.4 Regenerative Method

The regenerative method is a way to analyze steady-state data in a manner similar to that used in analyzing terminating data. The idea is to identify "naturally occurring" cycles in the output processes from which point estimates of the parameters of interest can be calculated. Seila (1990) and Chen and Cheng (1989) discuss estimation in regenerative simulations.

4.4.5 Variance Reduction Techniques

Yang and Nelson (1988, 1992) discuss the extension of a univariate variance-reduction technique, the use of control variates, to the multivariate case. Yang and Nelson's (1992) work can be used to gain insight as to the minimum number of batches to specify in the MBM method.

4.4.6 Sequential Methods

Raatikainen (1993) gives a sequential procedure for controlling the length of a simulation run so as to obtain confidence intervals on the components of $\overline{\mathbf{X}}$ that are a pre-specified width or smaller. This method is based on the Bonferroni Inequality and does not use the information contained in the off-diagonal elements of $\mathbf{S}$.

5 CONCLUSION

Multivariate methods must be used if the analyst is interested in learning about the correlation structure of the output processes of simulation models. Even if a joint confidence region won't be constructed, it can be informative to calculate the correlation matrix, $\mathbf{C}$, to gain some insight into the behavior of the model.

Constructing an ellipsoidal confidence region is a multivariate method that takes into account the cross covariance among output processes, while simultaneous confidence intervals based on the Bonferroni Inequality do not. However, ellipsoidal confidence regions are harder to interpret, especially for $D \geq 4$, when they can't be plotted. On the other hand, for higher D, the Bonferroni confidence intervals can be very large, and thus not very precise.

Research is continuing in developing and refining techniques for analyzing multivariate simulation output. Perhaps these methods will one day be included as part of the standard output routines in the commonly used simulation software packages.

REFERENCES

Anderson, T. W. 1978. Repeated measurements on autoregressive processes, *Journal of The American Statistical Association*, 73: 371–378.

Anderson, T.W. 1984. *An Introduction to Multivariate Statistical Analysis*, Second Edition. New York: John Wiley.

Bratley, P., B. L. Fox, and L. E. Schrage. 1987. *A Guide to Simulation*, Second Edition. New York: Springer-Verlag.

Charnes, J.M. 1990. Power comparisons for the multivariate batch-means method, In *Proceedings of the 1990 Winter Simulation Conference*, eds. O. Balci, R.P. Sadowski, R.E. Nance, 281–287. Institute of Electrical and Electronics Engineers, New Orleans, Lousiana.

Charnes, J.M. 1991. Multivariate simulation output analysis, In *Proceedings of the 1991 Winter Simulation Conference*, eds. B. L. Nelson, W. D. Kelton, G. M. Clark, 187–193. Institute of Electrical and Electronics Engineers, Phoenix, Arizona.

Charnes, J.M. and W.D. Kelton. 1988. A comparison of confidence region estimators for multivariate simulation output, In *Proceedings of the 1988 Winter Simulation Conference*, eds. M. Abrams, P. Haigh, J.C. Comfort, 458–465. Institute of Electrical and Electronics Engineers, San Diego, California.

Charnes, J. M. and Kelton, W. D. 1993. Multivariate autoregressive techniques for constructing confidence regions on the mean vector, *Management Science*, 39: 1112–1129.

Chen, D.R. and Chen, W.P. 1988. Multivariate inference in stationary simulation using batch means, In *Proceedings of the 1989 Winter Simulation Conference*, E.A. MacNair, K.J. Musselman, and P. Heidelberger, Eds. Institute of Electrical and Electronics Engineers, Piscataway, NJ, 524–533.

Chen , D.R. and Cheng, W.J. 1989. Multivariate statistical analysis for regenerative simulation, *Journal of Management Science*, 6: 43–56.

Chen, D.R. 1991. The relationship between multivariate correlation and coverage rate, presentation at TIMS XXX—SOBRAPO XXIII Joint International Meeting.

Chen, D.R. and A.F. Seila. 1987. Multivariate inference in stationary simulation using batch means, In *Proceedings of the 1987 Winter Simulation Conference*, eds. A. Thesen, H. Grant, W.D. Kelton, 302–304. Institute of Electrical and Electronics Engineers, Atlanta, Georgia.

Gallagher, M. A., Bauer, K. W., and Maybeck, P. S. 1994. Initial data truncation for multivariate output of discrete-event simulations using the Kalman filter. In *Annals of Operations Research*, ed. O. Balci, 53: 419.

Johnson, R.A. and D.W. Wichern 1988. *Applied Multivariate Statistical Analysis*, Second Edition. Englewood Cliffs, New Jersey: Prentice Hall.

Kabaila, P. and G. Nelson. 1985. On confidence regions for the mean of a multivariate time series, *Communications in Statistics—Theory and Methods*, 14: 735–753.

Law, A.M. and W.D. Kelton. 1991. *Simulation Modelling and Analysis*, Second Edition. New York: McGraw-Hill.

Miller, R. G. Jr. 1981. *Simultaneous Statistical Inference*, Second Edition. New York: Springer-Verlag.

Morrison, D.F. 1976. *Multivariate Statistical Methods*, New York: McGraw-Hill.

Munoz, D. F. 1991. Cancellation methods in the analysis of simulation output (confidence intervals). Ph.D. thesis, Department of Operations Research, Stanford University.

Raatikainen, K. E. E. 1993. A sequential procedure for simultaneous estimation of several means, *ACM Transactions on Modeling and Computer Simulation*, 3: 108–133.

Rao, C.R. 1951. An asymptotic expansion of the distribution of Wilk's criterion. *Bulletin of the International Statistical Institute*, 33: 177–180.

Schmeiser, B. 1982. Batch size effects in the analysis of simulation output. *Operations Research*, 30: 556–568.

Schruben, L.W. 1981. Control of initialization bias in multivariate simulation response. *Communications of the ACM*, 24: 246–252.

Seila, A. F. 1982. Multivariate estimation in regenerative simulation, *Operations Research Letters*, 1: 153–156.

Seila, A. F. 1983. Multivariate estimation in simulation, *Proceedings of the 1983 Winter Simulation Conference*, 693–696. Institute of Electrical and Electronics Engineers.

Seila, A. F. 1984. Multivariate simulation output analysis, *American Journal of Mathematical and Management Sciences*, 4: 313–334.

Seila, A.F. 1990. Multivariate Estimation of Conditional Performance Measures in Regenerative Simulation. *American Journal of Mathematical and Management Sciences* 10: 17–50.

Yang, W.-N. and B.L. Nelson. 1988. Multivariate estimation and variance reduction in terminating and steady-state simulation, In *Proceedings of the 1988 Winter Simulation Conference*, eds. M. Abrams, P. Haigh, J.C. Comfort, 466–472. Institute of Electrical and Electronics Engineers, San Diego, California.

Yang, W.-N. and B.L. Nelson. 1992. Multivariate batch means and control variates, *Management Science*, 38: 1415–1431.

AUTHOR BIOGRAPHY

JOHN M. CHARNES is Associate Professor in the management science group of the School of Business at the University of Kansas in Lawrence. His research interests are in multivariate simulation output analysis and using simulation for process improvement. He is the current newsletter editor for the INFORMS College on Simulation.