Proceedings of the 1991 Winter Simulation Conference
Barry L. Nelson, W. David Kelton, Gordon M. Clark (eds.)

MULTIVARIATE SIMULATION OUTPUT ANALYSIS

John M. Charnes

Department of Management Science
University of Miami
Coral Gables, Florida 33124-8237

ABSTRACT

This paper gives an overview of some multivariate statistical techniques that can be used in analyzing discrete-event simulation output. A general discussion of multivariate output is given, as well as methods for analyzing the output from two fundamentally different types (terminating and steady-state) of simulation models. References are provided for more advanced techniques of multivariate output analysis.

1 INTRODUCTION

There has been considerable activity recently by researchers on the problem of making statistical inferences simultaneously on more than one output measure of interest in simulation modeling (Chen and Seila 1987; Yang and Nelson 1988; Charnes and Kelton 1988; Charnes 1990). The intent of this paper is to give an overview of some multivariate output-analytic techniques for simulation practitioners. 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 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.

The next section discusses multivariate output from simulation models and contrasts their analysis to the univariate case. Section 3 discusses terminating simulation models. Section 4 presents some techniques to be used with 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 simple examples serve to illustrate the usefulness of considering multivariate, rather than univariate, output from simulation models.

Example 1. 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 may be 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 customers may be more tolerant of slightly longer delays if they observe corporate customers experiencing long delays when they do.

Example 2. Two states of a simple tandem queueing system are shown in Figure 2. The customers, depicted as open circles ($\circ$), 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 in-

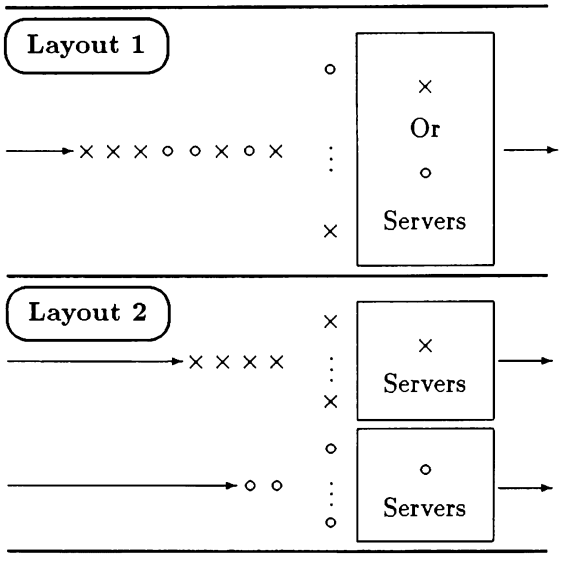

Figure 1: Two Bank Lobby Layouts

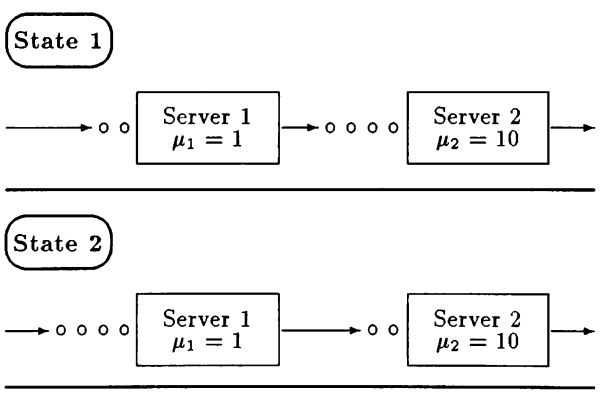

Figure 2: Tandem Queueing System

dividually 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 may miss important information about the system that might 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-analytic 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 comprising 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 SIMULATIONS

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 may 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* (rather than an *interval*). The procedure is based on Hotelling's T^2 distribution, and is the generalization of the univariate t-distribution confidence-interval procedure to higher dimensions. The validity of the procedure rests upon the assumption of multivariate normality of the data. See 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 may be the average cycle time, time-average number in queue d, 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

$$\vec{\mu} = E\left[\mathbf{X}^{(r)}\right] = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_D \end{pmatrix}$$

To form the confidence region, find first

$$\overline{\mathbf{X}} = \frac{1}{R}\sum_{r=1}^{R}\mathbf{X}^{(r)} = \begin{pmatrix} \sum_{r=1}^{R} X_1^{(r)}/R \\ \sum_{r=1}^{R} X_2^{(r)}/R \\ \vdots \\ \sum_{r=1}^{R} X_D^{(r)}/R \end{pmatrix}$$

Then an unbiased estimate of the variance-covariance matrix of the vector of point estimators is

$$\mathbf{S} = \frac{1}{(R-1)}\sum_{r=1}^{R}(\mathbf{X}^{(r)} - \overline{\mathbf{X}})(\mathbf{X}^{(r)} - \overline{\mathbf{X}})',$$

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 $\boldsymbol{\Theta}$ such that

$$(\overline{\mathbf{X}} - \boldsymbol{\Theta})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\Theta}) \leq \frac{D(R-1)}{R(R-D)}F_{D,R-D}(\alpha)$$

where $F_{D,R-D}(\alpha)$ is the upper (100α)th percentile of the F distribution with D and $R-D$ degrees of freedom.

This procedure gives a formula for an ellipsoidal region, the exact shape of which depends upon the magnitudes and algebraic sign 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 may be informative to compute the correlation matrix, $\mathbf{C}$, for the mean vector. This is calculated as

$$c_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}},$$

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 $-1 \leq c_{ij} \leq 1$.

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 can not 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).

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 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 these two assumptions are important for the successful use of this method in practice.

3.2 Simultaneous Confidence Intervals

Usually, the analyst will want to construct individual confidence intervals on the mean of each component process in the output vector. However, one must be careful in constructing more than one confidence interval from simulation output, as 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 difficult to determine; however, the Bonferroni Inequality gives a simple means of setting the individual confidence levels in order to obtain a *lower bound* on the overall level of confidence.

In particular, let C_d denote a confidence statement about the mean value of the dth component of the output vector. If

$$\Pr(C_d \text{ true}) = 1 - \alpha_d \quad \text{for } d = 1, \ldots, D$$

then

$$\Pr(\text{all } C_d \text{ true}) \geq 1 - (\alpha_1 + \alpha_2 + \cdots + \alpha_D).$$

Thus, for example, if each one of D confidence intervals is constructed at the $1 - \alpha/D$ level, one may have an overall level of confidence of at least $1 - \alpha$ that all D parameters lie in the D-dimensional "box" defined by the D confidence intervals.

If the vector $\overline{\mathbf{X}}$ and the matrix $\mathbf{S}$ has been calculated as descibed in the previous section, individual "Bonferroni Intervals" can be constructed as follows on the mean of each parameter of interest:

$$\begin{aligned}
\overline{x}_1 &\pm t_{(R-1)}\left(\frac{\alpha}{2D}\right)\sqrt{\frac{s_{11}}{R}} \\
\overline{x}_2 &\pm t_{(R-1)}\left(\frac{\alpha}{2D}\right)\sqrt{\frac{s_{22}}{R}} \\
&\vdots \\
\overline{x}_D &\pm t_{(R-1)}\left(\frac{\alpha}{2D}\right)\sqrt{\frac{s_{DD}}{R}}
\end{aligned}$$

where $\overline{x}_d$ is the dth element of the vector $\overline{\mathbf{X}}$; $t_{(R-1)}(\frac{\alpha}{2D})$ is the upper $(100\alpha/2D)$th percentile of the t distribution with $(R-1)$ degrees of freedom; and s_{dd} is the dth diagonal element of $\mathbf{S}$.

The advantage of using Bonferroni Intervals is their ease of construction (they are merely the combination of several univariate intervals) and interpretation. 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, individually, the univariate confidence intervals do not obtain their nominal coverage, use of the Bonferroni Inequality will not ensure that the bound on the nominal coverage given by the inequality will obtain.

4 STEADY-STATE SIMULATIONS

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, which means intuitively that 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 running for a suitable length of time to mitigate any bias induced by the non-representative initial conditions. Schruben (1981) gives a multivariate method for deciding when the simulation appears to have reached steady state.

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

4.1 Autocorrelation Function

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*,

$$\mathbf{\Gamma}(h) = E\left[(\mathbf{X}_t - \vec{\mu})(\mathbf{X}_{t+h} - \vec{\mu})'\right],$$

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 $\mathbf{\Gamma}(h)$, $\gamma_{ii}(h)$. 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

$$\mathbf{G}(h) = \frac{1}{T-h}\sum_{t=1}^{T-h}(\mathbf{X}_t - \overline{\mathbf{X}})(\mathbf{X}_{t+h} - \overline{\mathbf{X}})'$$

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 **G**s as

$$r_{ij}(h) = \frac{g_{ij}(h)}{\sqrt{g_{ii}(0)g_{jj}(0)}}$$

Because these are correlations, it will be true that

$$-1 \le r_{ij}(h) \le 1 \quad \forall i, j, h.$$

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 Method

The multivariate batch-means (MBM) method can be used to construct confidence regions on the mean vector in a manner similar to that described for terminating simulations, but to do so with autocorrelated data.

The MBM method attempts to circumvent the autocorrelation problem without losing the information on cross-correlation by grouping the data into (approximately) uncorrelated batches. This is the same idea behind the nonoverlapping batch-means method used with univariate data. To use the MBM method properly, the analyst must use simultaneous observations in order to make the cross correlation meaningful. (The observations need not be equally spaced.)

The procedure is begun by grouping the elements of the output matrix

$$\begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,T} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ X_{D,1} & X_{D,2} & \cdots & X_{D,T} \end{bmatrix}$$

into m *serially* uncorrelated (approximately) batches of length k ($T = km$) and finding the vectors of "batch means" to get

$$\begin{bmatrix} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,m} \\ Y_{2,1} & Y_{2,2} & \cdots & Y_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{D,1} & Y_{D,2} & \cdots & Y_{D,m} \end{bmatrix}$$

where the batch means are computed as

$$Y_{i,j} = \frac{1}{k}\sum_{t=(j-1)k+1}^{jk} X_{i,t}.$$

The method then proceeds as if the vectors

$$\mathbf{Y}_j = \begin{pmatrix} Y_{1,j} \\ Y_{2,j} \\ \vdots \\ Y_{D,j} \end{pmatrix}$$

are iid multivariate normal random vectors. That is, first a point estimate of the mean is found as

$$\overline{\mathbf{X}} = \frac{1}{m}\sum_{j=1}^{m}\mathbf{Y}_j$$

and the sample variance-covariance matrix is computed:

$$\mathbf{S} = \frac{1}{m-1}\sum_{j=1}^{m}(\mathbf{Y}_j - \widehat{\mu})(\mathbf{Y}_j - \widehat{\mu})'.$$

Then a $100(1-\alpha)\%$ confidence region for μ is computed as the set of all vectors $\boldsymbol{\Theta}$ such that

$$(\overline{\mathbf{X}} - \boldsymbol{\Theta})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\Theta}) \le \frac{D(m-1)}{m(m-D)}F_{\alpha;D,m-D}.$$

A critical step in using the MBM method is the determination of the number of vector observations per batch, k (or, equivalently, the number of batches, m). The usual method of making this determination is to assume that the batch-means process can be sufficiently approximated by the first-order, vector-autoregressive (VAR(1)) model

$$\mathbf{Y}_i = \boldsymbol{\Phi}\mathbf{Y}_{i-1} + \epsilon_i \qquad \text{for} \quad i = 2, \ldots, m,$$

where Φ is a $(D \times D)$ matrix of autoregression coefficients and the ϵ_i are $(D \times 1)$ iid vectors of random errors drawn from the multivariate normal distribution. Then k is chosen such that H_0: $\Phi = \mathbf{0}$ is not rejected. Implicit in the use of this model is the assumption that if the first-order serial correlation is zero, then the higher-order serial correlations will also be zero. Yang and Nelson (1991) give some guidelines for choosing k. Charnes (1990) found that a good test statistic for H_0 is the F-approximation to the Wilks likelihood-ratio procedure suggested by Rao (1951).

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 overlapping batch means method with each component process taken individually (see Schmeiser 1982), while being mindful of the Bonferroni Inequality when choosing the values of the t statistic used to construct each interval.

An alternative is to use the elements of $\mathbf{S}$ from the MBM method as follows

$$\begin{aligned}
\overline{x}_1 &\pm t_{(m-1)}\left(\frac{\alpha}{2D}\right)\sqrt{\frac{s_{11}}{m}} \\
\overline{x}_2 &\pm t_{(m-1)}\left(\frac{\alpha}{2D}\right)\sqrt{\frac{s_{22}}{m}} \\
&\vdots \\
\overline{x}_D &\pm t_{(m-1)}\left(\frac{\alpha}{2D}\right)\sqrt{\frac{s_{DD}}{m}}
\end{aligned}$$

Note that by using this alternative method, the analyst is forcing the batch sizes to be the same, which is not necessarily true when the univariate overlapping 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 the overlapping batch means method applied individually to the component processes.

4.4 Other Multivariate Techniques

More advanced multivariate techniques have been proposed for analyzing data generated by stationary processes.

Kabaila and Nelson (1985) give a frequency-domain time-series technique, and Charnes (1989) gives a time-domain time-series technique for constructing confidence regions in the steady-state case. Research is currently underway to compare the performance of these techniques *vis-a-vis* the performance of the methods described above.

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. A recent paper by Seila (1990) discusses estimation in regenerative simulations and gives references.

Yang and Nelson (1988, 1991) discuss the extension of univariate variance reduction techniques to the multivariate case.

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 soon be included as part of the standard output routines in the commonly used simulation software packages.

ACKNOWLEDGEMENT

The author is grateful to the Minnesota Supercomputer Institute for providing the equipment with which to prepare this paper.

REFERENCES

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. 1989. Statistical analysis of multivariate discrete-event simulation output. Ph.D. Thesis,

Department of Operations and Management Science, University of Minnesota, Minneapolis, Minnesota.

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. 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.

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.

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**(3): 735–753.

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

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

Rao, C.R. 1951. An asymptotic expansion of the distribution of Wilk's criterion. *Bulletin of the International Statistical Institute*, **33**, Part 2: 177–180.

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

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**(4): 246–252.

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. 1991. Multivariate batch means and control variates, Technical Report, Department of Industrial and Systems Engineering, The Ohio State University, Columbus, Ohio.

AUTHOR BIOGRAPHY

JOHN M. CHARNES is Assistant Professor in the Management Science Department at the University of Miami in Coral Gables, Florida. His main research interests are in multivariate time series and simulation output analysis. He holds the Bachelor of Civil Engineering, MBA, and Ph.D. degrees from the University of Minnesota. Before he endured graduate school, Professor Charnes was a surveyor, and construction engineer for a subsidiary of S.J. Groves and Sons, a general contractor based in Minneapolis, Minnesota.