Technical Report CS81019-R*

A METHODOLOGY FOR VALIDATING MULTIVARIATE RESPONSE SIMULATION MODELS BY USING SIMULTANEOUS CONFIDENCE INTERVALS

by

Osman Balci

and

Robert G. Sargent**

Department of Computer Science
Virginia Polytechnic Institute & State University
Blacksburg, Virginia 24061

November 1981

---

*Cross listed as Working Paper #81-009 in the Department of Industrial Engineering and Operations Research, Syracuse University, Syracuse, New York 13210

**Currently with the Department of Industrial Engineering and Operations Research, Syracuse University, Syracuse, New York 13210

ABSTRACT

This paper deals with the substantiation that a multivariate response self- or trace-driven simulation model, within its domain of applicability, possesses a satisfactory range of accuracy consistent with the intended application of the model.

A methodology is developed by using simultaneous confidence intervals to do this substantiation with respect to the mean behavior of a simulation model that represents an observable system. A trade-off analysis can be performed and judgement decisions can be made as to what data collection budget to allocate, what data collection method to use, how many observations to collect on each of the model and system response variables, and what confidence level to choose for producing the range of accuracy with satisfactory lengths. The methodology is illustrated for self-driven steady-state and trace-driven terminating simulations.

1. INTRODUCTION

One of the most important steps in the development of a simulation model is determining whether the representation of the computerized model has sufficient accuracy for the purpose for which the model is to be used. "Substantiation that a computerized (simulation) model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model" is usually referred to as (simulation) model validation [21] and is the definition used in this paper.

The validity or the range of accuracy of a simulation model should be evaluated with regard to experimental frame(s) and this evaluation should be made only in terms of the purpose for which the model is developed. As defined by Zeigler [25], an experimental frame "... characterizes a limited set of circumstances under which the real system (or the model) is to be observed or experimented with." A model that is valid in one experimental frame may be completely invalid in another. Therefore, the validity of a simulation model should only be evaluated with respect to a set of experimental frames determined by the purpose for which the model is intended, and not with respect to all possible experimental frames (or all sets of conditions) [19, 20]. Sometimes, the model sponsor, model user, or a third party will specify an acceptable range of accuracy which is the amount of accuracy that is required for the simulation model to be valid under a given experimental frame. In this case, the validity or the range of accuracy of the simulation model must be evaluated with respect to the acceptable range of accuracy specified.

It is generally preferable to use some form of objective analysis to substantiate that a computerized simulation model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model [3, 4, 20]. In this paper, simultaneous confidence intervals (s.c.i.), which is a form of objective analysis, will be used to determine the validity of a multivariate response simulation model.

In using a statistical procedure for validation, one should consider the type of the simulation model with regard to the way it is driven and with regard to the way its output is analyzed. There are basically two types of simulation models with regard to the way they are driven: self- and trace-driven simulation models. <u>Self-driven</u> (<u>distribution-driven</u> or <u>probabilistic</u>) <u>simulation</u> [11] is a technique which uses random numbers in sampling from distributions or stochastic processes. <u>Trace-driven</u> (or <u>retrospective</u> [17]) <u>simulation</u> is a technique which combines measurement and simulation by using the actual data collected on the system as the model input [11, 23].

There are basically two types of simulation models with regard to analysis of the output: steady-state and terminating simulation models [7, 14]. A <u>steady-state simulation</u> "is one for which the quantity of interest is defined as a limit as the length of the simulation goes to infinity" [14]. A <u>terminating simulation</u> "is one for which any quantities of interest are defined relative to the interval of simulated time $[0,T_E]$, where T_E, a possibly degenerate random variable, is the time that a specified event E occurs" [14].

The validity or the range of accuracy of a multivariate response simulation model can be expressed in terms of s.c.i. for the differences between the corresponding model and system response variables. A confidence interval (c.i.) constructed for the difference between the jth model and system response variables can represent the range of accuracy of the jth model response variable. The length of the c.i. then becomes the length of the <u>range of accuracy</u>.

It is usually desirable to have the length of the range of accuracy as short as possible. The length or the expected length of the c.i. can be decreased by either increasing the sample sizes of observations or by decreasing the confidence level. However, increasing the sample sizes will increase the cost of data collection. Thus, a trade-off analysis may be necessary to determine appropriate values for the sample sizes, confidence level(s), and data collection budget.

Schedules and graphs can be constructed to show the relationships among the sample sizes, confidence levels, estimates of half lengths of the range of accuracy of the simulation model, and cost of data collection. These schedules and graphs can be used to make judgement decisions regarding the trade-offs among the parameters of interest.

The purpose of this paper is to give a methodology for validating a multivariate response simulation model with respect to its mean behavior by using s.c.i. The methodology is given in section 2 and the Bonferroni method for constructing the range of accuracy is discussed in section 3. Section 4 contains some examples that illustrate the methodology by using the Bonferroni s.c.i. and section 5 contains the conclusions.

2. METHODOLOGY

The validity, with respect to the mean behavior, of a multivariate response stochastic simulation model representing a stochastic observable system can be evaluated by examining the differences between the population means of the corresponding model and system performance measures when the model is run with the "same" input data that drive the real system. The model and system performance measures are represented by the model and system response variables.

Assuming that there are k response variables from the model and from the system, let $(\underline{\mu}^m)' = [\mu_1^m,\mu_2^m,\ldots,\mu_k^m]$ and $(\underline{\mu}^s)' = [\mu_1^s,\mu_2^s,\ldots,\mu_k^s]$ be the k dimensional vectors containing the population means of the model and system response variables, respectively.

The validity of a simulation model can be expressed as a range of accuracy. The range of accuracy for the mean behavior of a multivariate response simulation model can be expressed in terms of s.c.i.

Basically, there are three approaches for constructing the s.c.i. to express the range of accuracy for the mean behavior.

In the first approach, the range of accuracy is determined by using a statistical procedure that gives the 100(1-γ)% s.c.i. for $\underline{\mu}^m-\underline{\mu}^s$ as

$$[\underline{\ell},\underline{u}] \tag{1}$$

where $\underline{\ell}' = [\ell_1,\ell_2,\ldots,\ell_k]$ and $\underline{u}' = [u_1,u_2,\ldots,u_k]$. We can be 100(1-$\gamma$)% confident that the true differences between the population means of the model and system response variables are simultaneously contained within the range of accuracy of the simulation model given by (1).

In the second approach, the 100(1-γ^m)% s.c.i. are first constructed for $\underline{\mu}^m$ as

$$[\underline{\ell}^m,\underline{u}^m] \tag{2}$$

where $(\underline{\ell}^m)' = [\ell_1^m,\ell_2^m,\ldots,\ell_k^m]$ and $(\underline{u}^m)' = (u_1^m,u_2^m,\ldots,u_k^m]$. Then, the 100(1-$\gamma^s$)% s.c.i. are constructed for $\underline{\mu}^s$ as

$$[\underline{\ell}^s,\underline{u}^s] \tag{3}$$

where $(\underline{\ell}^s)' = [\ell_1^s,\ell_2^s,\ldots,\ell_k^s]$ and $(\underline{u}^s)' = [u_1^s,u_2^s,\ldots,u_k^s]$. Finally, the range of accuracy of the simulation model is determined by the following s.c.i. for $\underline{\mu}^m-\underline{\mu}^s$ with a confidence level of at least

$(1-\gamma^m-\gamma^s)$ [10, p. 466]:

$$[\underline{\ell}^m-\underline{u}^s,\ \underline{u}^m-\underline{\ell}^s]. \tag{4}$$

Letting $\gamma = \gamma^m + \gamma^s$, $\ell_i = \ell_i^m - u_i^s$ and $u_i = u_i^m - \ell_i^s$ for $i = 1,2,\ldots,k$, we can be at least $100(1-\gamma)\%$ confident that the true differences between the population means of the model and system response variables are simultaneously contained within the range of accuracy given by $[\underline{\ell},\underline{u}]$.

In the third approach, the model and system response variables are observed in pairs and the range of accuracy is determined by the $100(1-\gamma)\%$ s.c.i. for $\underline{\mu}^d$, the population means of the differences of paired observations, as

$$[\underline{\ell}^d,\underline{u}^d] \tag{5}$$

where $(\underline{\ell}^d)' = [\ell_1^d,\ell_2^d,\ldots,\ell_k^d]$ and $(\underline{u}^d)' = [u_1^d,u_2^d,\ldots,u_k^d]$. Letting $\ell_i = \ell_i^d$ and $u_i = u_i^d$, for $i = 1,2,\ldots,k$, we can be $100(1-\gamma)\%$ confident that the true population means of the differences of paired observations collected from the model and system response variables are simultaneously contained within the range of accuracy of the simulation model given by $[\underline{\ell},\underline{u}]$.

The approach for constructing the s.c.i. to express the range of accuracy of a simulation model should be chosen with respect to the way the simulation model is driven and the statistical procedure used in constructing the s.c.i. The range of accuracy is determined by the s.c.i. constructed by using the observations that are collected from the model and system response variables by running the simulation model with the "same" input data that drive the real system. If the simulation model is self-driven, then the "same" indicates that the model input data are coming independently from the same populations or stochastic process of the system input data. Since the model and system input data are independent from each other, but coming from the same populations or stochastic process, then the model and system output data are expected to be identical and independent from each other. Hence, the range of accuracy of a multivariate response self-driven simulation model can be determined by the s.c.i. that are constructed by using one of the aforementioned three approaches. If the simulation model is trace-driven, then the "same" indicates that the model input data are exactly the same as the system input data. In this case, the model and system output data are expected to be dependent and identical to each other. Therefore, the third aforementioned approach should be used for constructing the s.c.i. to determine the range of accuracy of a multivariate response trace-driven simulation model.

Sometimes, the model sponsor, model user, or a third party will specify an acceptable range of accuracy for a specific simulation model. This specification can be made for the mean behavior of a stochastic simulation model as

$$L_j \leq \mu_j^m - \mu_j^s \leq U_j, \quad j = 1,\ldots,k \tag{6}$$

where L_j and U_j are the lower and upper bounds of the acceptable difference between the population means of the jth model and system response variables. In this case, the range of accuracy of the simulation model should be checked if it satisfies the acceptable range of accuracy given.

It is desirable to construct the range of accuracy with half lengths as short as possible. The shorter the half lengths, the more meaningful the specification of the range of accuracy will be. The half lengths are affected by the values of confidence level(s), variances of the model and system response variables, and sample sizes of observations. The half lengths can be shortened by decreasing the confidence level(s). Variance reduction techniques [6] can be used, in some cases when collecting observations from a simulation model, to decrease the variability of the observations and thus obtain a shorter range of accuracy. The half lengths can also be shortened by increasing the sample sizes of observations. However, increasing the sample sizes of observations will increase the cost of data collection. In those cases where the cost of data collection is high and is an important factor to consider, a tradeoff analysis should be performed to choose appropriate values for the sample sizes, confidence level(s), and data collection budget.

For a real system represented by a simulation model with k response variables, let $\underline{n}' = [n_1, n_2, \ldots, n_k]$ and $\underline{N}' = [N_1, N_2, \ldots, N_k]$ be the k dimensional vectors containing the sample sizes of observations on the model and system response variables, respectively. Let $\underline{c}' = [c_1, c_2, \ldots, c_k]$ and $\underline{C}' = [C_1, C_2, \ldots, C_k]$ be the k dimensional vectors containing the unit costs of collecting one observation from the model and system response variables, respectively. Let $\underline{c}_0' = [c_{01}, c_{02}, \ldots, c_{0k}]$ and $\underline{C}_0' = [C_{01}, C_{02}, \ldots, C_{0k}]$ be the k dimensional vectors containing the overhead data collection costs for the model and system response variables, respectively. Total cost of data collection on the model (Tc) and on the system (TC) can usually be evaluated as a linear function of the sample sizes and overhead costs and can be stated as follows by assuming that the collection of observations on one response variable is done independently from the collection of observations on another response variable:

$$Tc = \underline{c}_0'\underline{1} + \underline{c}'\underline{n}, \tag{7}$$

$$TC = \underline{C}_0'\underline{1} + \underline{C}'\underline{N}, \text{ and} \tag{8}$$

$$CDC = Tc + TC \tag{9}$$

where $\underline{1}$ is a k dimensional vector of ones and CDC is the total cost of data collection on the model and system.

The equations (7) and (8) are used to determine the cost of data collected for validation only and are stated in general assuming unequal sample sizes. Usually, however, one would simultaneously collect observations from a simulation model with equal sample sizes. In this case, Tc would be evaluated as

$$Tc = c_t + c_m n \tag{10}$$

where $c_t = \underline{c}_0'\underline{1}$ and $c_m = \underline{c}'\underline{1}$. In those cases where the real system is observed simultaneously with equal sample sizes, TC would be evaluated as

$$TC = C_t + C_s N \tag{11}$$

where $C_t = \underline{C}_0'\underline{1}$ and $C_s = \underline{C}'\underline{1}$.

The unit and overhead data collection costs should be estimated with respect to the data collection method employed. There may be alternative values of the model and system unit and overhead costs corresponding to alternative methods of data collection. The analyst may wish to make a trade-off analysis with respect to alternative methods of data collection and upon this analysis conclude which data collection method to employ in validation.

For a given data collection budget B, it is possible to select different values for the sample sizes of observations n and N. (Unequal sample sizes are considered for the purpose of generality.) It is desirable to select the sample sizes in such a way that the length of the range of accuracy of the simulation model will be minimized for the given data collection budget. A function, $f(n_j, N_j; j = 1,\ldots,k)$, can be determined in terms of the sample sizes from the half lengths (H_j, $j = 1,\ldots,k$) of the s.c.i. that determine the lengths of the range of accuracy. This function should be determined in such a way that the half lengths or the expected half lengths of the s.c.i. decrease as the value of this function decreases. Thus, the sample sizes n and N can be selected, to produce the range of accuracy of the simulation model with minimum lengths for a given data collection budget, by solving the following optimization problem.

$$
\begin{aligned}
\text{Minimize:}\quad & f(n_j, N_j;\ j = 1,\ldots,k) \\
\text{Subject to:}\quad & \sum_{j=1}^{k} c_j n_j + \sum_{j=1}^{k} C_j N_j + \sum_{j=1}^{k} (c_{0j} + C_{0j}) \le B \\
& \sum_{j=1}^{k} n_j + \sum_{j=1}^{k} N_j \ge q \\
& n_j \ge r, \qquad j = 1,\ldots,k \\
& N_j \ge R, \qquad j = 1,\ldots,k \\
& n_j, N_j \text{ integer} \qquad j = 1,\ldots,k
\end{aligned} \tag{12}
$$

where q is the minimum total sample size requirement and r and R are the minimum sample size requirements for the statistical procedure used in constructing the s.c.i.; and the other parameters are used as defined before.

The optimal sample sizes $\underline{n}^*$ and $\underline{N}^*$ that are determined by solving (12) and the estimates of variances of the response variables that are obtained from a pilot run or data are both used in estimating the half lengths of the $100(1-\gamma)\%$ s.c.i. that determine the range of accuracy of the simulation model. The expected half lengths or the half lengths of the s.c.i. which are estimated by using $\underline{n}^*$ and $\underline{N}^*$ will be the shortest among all possible values of the half lengths that are obtained from all possible values of $\underline{n}$ and $\underline{N}$ for a given data collection budget B. Since the half lengths of the s.c.i. determine the half lengths of the range of accuracy of the simulation model, we will be concerned with the half length estimates rather than the expected half lengths. H_j^*, $j = 1,\ldots,k$, will represent the half lengths of the range of accuracy that are minimum in length because of the use of $\underline{n}^*$ and $\underline{N}^*$ for a given data collection budget.

Schedules and graphs can be constructed to show the relationships among the unit and overhead data collection costs ($\underline{c}$, $\underline{C}$, $\underline{c}_0$, $\underline{C}_0$), data collection budget (B), optimal sample sizes of observations ($\underline{n}^*$, $\underline{N}^*$), total cost of data collection (CDC), confidence level $(1-\gamma)$, and estimates of minimum half lengths of the range of accuracy of the simulation model (H_j^*, $j = 1,\ldots,k$). The objective function and the constraints of the optimization problem (12) should be determined specifically for the statistical procedure used in constructing the range of accuracy and with respect to the assumption underlying the equality of the sample sizes. Construction of the schedules should be based upon the approach chosen for determining the range of accuracy and upon the assumption underlying the equality of the sample sizes.

If approach I for developing the range of accuracy is chosen and it is assumed or required that $n_j = n$, $N_j = N$, $j = 1,\ldots,k$, then the schedules in Table 1 are constructed for "a" values of c_m, c_t, C_s, and C_t; for "b" values of B; and for "c" values of γ. (In Table 1, "ε" means "contained in".)

If approach II is chosen and it is assumed or required that $n_j = n$, $N_j = N$, $j = 1,\ldots,k$, then the schedules in Table 2 are constructed. In Table 2, H_j^{m*} and H_j^{s*}, $j = 1,\ldots,k$, represent the estimates of the half lengths of the s.c.i. that are minimum in length because of the use of the optimal sample sizes n^* and N^* for the given data collection budget on the model (B^m) and on the system (B^s), respectively. B^m and B^s are specified to cover the cost of data collection on the model ($Tc = c_t + c_m n^*$) and on the system ($TC = C_t + C_s N^*$), respectively. The specifications of B^m and B^s and the collection of observations on the model and on the system are done independently and the s.c.i. for the model and for the system response variables are constructed separately with joint confidence levels of $1-\gamma^m$ and $1-\gamma^s$, respectively.

TABLE 1. Schedules. $\{(\underline{\mu}^m-\underline{\mu}^s) \in [\underline{\ell},\underline{u}];\ n_j = n,\ N_j = N,\ j = 1,\ldots,k\}$

c_m,c_t	C_s,C_t	B	n^*	N^*	CDC	γ	H_1^*	...	H_k^*
						γ_1	H^*_{1111}	...	H^*_{k111}
		B_{11}	n^*_{11}	N^*_{11}	CDC_{11}	⋮	⋮	...	⋮
						γ_c	H^*_{111c}	...	H^*_{k11c}
c_{m1},c_{t1}	C_{s1},C_{t1}	⋮	⋮	⋮	⋮	⋮	⋮	...	⋮
						γ_1	H^*_{11b1}	...	H^*_{k1b1}
		B_{1b}	n^*_{1b}	N^*_{1b}	CDC_{1b}	⋮	⋮	...	⋮
						γ_c	H^*_{11bc}	...	H^*_{k1bc}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...	⋮
						γ_1	H^*_{1a11}	...	H^*_{ka11}
		B_{a1}	n^*_{a1}	N^*_{a1}	CDC_{a1}	⋮	⋮	...	⋮
						γ_c	H^*_{1a1c}	...	H^*_{ka1c}
c_{ma},c_{ta}	C_{sa},C_{ta}	⋮	⋮	⋮	⋮	⋮	⋮	...	⋮
						γ_1	H^*_{1ab1}	...	H^*_{kab1}
		B_{ab}	n^*_{ab}	N^*_{ab}	CDC_{ab}	⋮	⋮	...	⋮
						γ_c	H^*_{1abc}	...	H^*_{kabc}

TABLE 2. Schedules. $\{\underline{\mu}^m \epsilon [\underline{\ell}^m, \underline{u}^m];\ \underline{\mu}^s \epsilon [\underline{\ell}^s, \underline{u}^s];\ H_j^* = H_j^{m*} + H_j^{s*},\ n_j = n,\ N_j = N,\ j=1,\ldots,k\}$

c_m, c_t	B^m	n^*	Tc	γ^m	H_1^{m*}	...	H_k^{m*}	C_s, C_t	B^s	N^*	TC	γ^s	H_1^{s*}	...	H_k^{s*}
				γ_1^m	H_{1111}^{m*}	...	H_{k111}^{m*}					γ_1^s	H_{1111}^{s*}	...	H_{k111}^{s*}
	B_{11}^m	n_{11}^*	Tc_{11}	⋮	⋮	...	⋮		B_{11}^s	N_{11}^*	TC_{11}	⋮	⋮	...	⋮
				γ_c^m	H_{111c}^{m*}	...	H_{k11c}^{m*}					γ_c^s	H_{111c}^{s*}	...	H_{k11c}^{s*}
c_{m1}, c_{t1}	⋮	⋮	⋮	⋮	⋮	...	⋮	C_{s1}, C_{t1}	⋮	⋮	⋮	⋮	⋮	...	⋮
				γ_1^m	H_{11b1}^{m*}	...	H_{k1b1}^{m*}					γ_1^s	H_{11b1}^{s*}	...	H_{k1b1}^{s*}
	B_{1b}^m	n_{1b}^*	Tc_{1b}	⋮	⋮	...	⋮		B_{1b}^s	N_{1b}^*	TC_{1b}	⋮	⋮	...	⋮
				γ_c^m	H_{11bc}^{m*}	...	H_{k1bc}^{m*}					γ_c^s	H_{11bc}^{s*}	...	H_{k1bc}^{s*}
⋮	⋮	⋮	⋮	⋮	⋮	...	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...	⋮
				γ_1^m	H_{1a11}^{m*}	...	H_{ka11}^{m*}					γ_1^s	H_{1a11}^{s*}	...	H_{ka11}^{s*}
	B_{a1}^m	n_{a1}^*	Tc_{a1}	⋮	⋮	...	⋮		B_{a1}^s	N_{a1}^*	TC_{a1}	⋮	⋮	...	⋮
				γ_c^m	H_{1a1c}^{m*}	...	H_{ka1c}^{m*}					γ_c^s	H_{1a1c}^{s*}	...	H_{ka1c}^{s*}
c_{ma}, c_{ta}	⋮	⋮	⋮	⋮	⋮	...	⋮	C_{sa}, C_{ta}	⋮	⋮	⋮	⋮	⋮	...	⋮
				γ_1^m	H_{1ab1}^{m*}	...	H_{kab1}^{m*}					γ_1^s	H_{1ab1}^{s*}	...	H_{kab1}^{s*}
	B_{ab}^m	n_{ab}^*	Tc_{ab}	⋮	⋮	...	⋮		B_{ab}^s	N_{ab}^*	TC_{ab}	⋮	⋮	...	⋮
				γ_c^m	H_{1abc}^{m*}	...	H_{kabc}^{m*}					γ_c^s	H_{1abc}^{s*}	...	H_{kabc}^{s*}

Notice that the minimum estimates of half lengths of the range of accuracy of the simulation model (H_j^*, j = 1,...,k) are equal to $H_j^{m*} + H_j^{s*}$, j = 1,...,k, respectively. The schedules in Table 2 are constructed for "a" values of c_m, c_t, C_s, and C_t; for "b" values of B^m and B^s; and for "c" values of γ^m and γ^s. Although the values of "a", "b", and "c" that are chosen for the model can be different from the values chosen for the system, they are shown to be the same in Table 2 for the purpose of simplicity. In a similar way, schedules can be constructed for approach III with equal and unequal sample sizes and for approach I and II with unequal sample sizes [2].

The model sponsor, model user, and model builder, individually or together, can perform a trade-off analysis by using the schedules and/or the graphs of the data contained in the schedules as will be illustrated later in section 4. The trade-offs among the cost of data collection, sample sizes of observations, confidence levels, and estimates of minimum half lengths of the range of accuracy of the simulation model can be examined and judgement decisions can be made to determine appropriate values for the data collection budget, sample sizes of observations, and confidence levels to produce the range of accuracy with satisfactory half lengths.

Next, the methodology for validating multivariate response simulation models with respect to the mean behavior by using s.c.i. will be presented.

1. Determine the experimental frame under which the validity of the simulation model is going to be tested. Go to 2.

2. Specify the acceptable range of accuracy, if there exists one, for the population means with respect to the intended application of the model as

$$L_j \leq \mu_j^m - \mu_j^s \leq U_j, \quad j = 1,\ldots,k$$

where L_j and U_j are the lower and upper bounds of the acceptable difference between the population means of the jth model and system response variables. Go to 3.

3. If the simulation model is self-driven, go to 4; otherwise, if it is trace-driven, go to 5.

4. Choose one of the three approaches for constructing the s.c.i. that determine the range of accuracy of the simulation model and determine an appropriate statistical procedure for constructing the s.c.i. with respect to the approach chosen. Go to 6.

5. Determine an appropriate statistical procedure, with respect to the third approach, for constructing the s.c.i. that determine the range of accuracy of the simulation model. Go to 6.

6. If a trade-off analysis among cost of data collection, sample sizes, confidence levels, acceptable range of accuracy, and estimates of minimum half lengths of the range of accuracy of the simulation model is desired, go to 7; otherwise go to 9.

7. Construct the schedules and graphs with respect to the statistical approach and procedure chosen by using the optimal sample sizes. Go to 8.

8. Examine the trade-offs among the parameters by studying the schedules and/or the graphs of the data contained in the schedules. Make judgement decisions to determine appropriate values for the data collection budget, sample sizes of observations, and confidence level(s) to produce satisfactory half lengths of the range of accuracy. The estimates of minimum half lengths of the s.c.i. that determine the range of accuracy of the simulation model should be chosen to be less than or equal to the half lengths of the acceptable range of accuracy if there is one specified. Go to 10.

9. Determine the sample sizes of observations and confidence level(s). Go to 10.

10. Collect data from the real system and from the simulation model by running the simulation model with the "same" input data that drive the real system. Go to 11.

11. Determine the range of accuracy of the simulation model by constructing the $100(1-\gamma)\%$ s.c.i. $[\underline{\ell},\underline{u}]$ using the s.c.i. approach and statistical procedure selected. Go to 12.

12. If an acceptable range of accuracy is specified, go to 13; otherwise go to 16.

13. Determine if $[\underline{\ell},\underline{u}] \varepsilon [\underline{L},\underline{U}]$. If yes, go to 14; otherwise go to 15.

14. We are at least $100(1-\gamma)\%$ confident that the true differences between the population means of the model and system response variables are contained within the acceptable range of accuracy $[\underline{L},\underline{U}]$. Thus, the model is valid with respect to the acceptable range of accuracy under the experimental frame and the validity of the model is given by the range of accuracy $[\underline{\ell},\underline{u}]$ at a confidence level of (at least) $100(1-\gamma)\%$. Terminate.

15. At least one of the response variables' range of accuracy is not contained within its acceptable range of accuracy. If this incompatibility is believed to be created because of the values chosen for the sample sizes, confidence levels, and/or the estimated parameters, then go back to 6 and choose new values; otherwise revise the model and go to 6.

16. If the (at least) $100(1-\gamma)\%$ range of accuracy for the mean behavior of the simulation model is satisfactory, conclude that the model is valid under the experimental frame and the validity of the model is given by the range of accuracy $[\underline{\ell},\underline{u}]$ at a confidence level of (at least) $100(1-\gamma)\%$ and terminate; otherwise, if the range of accuracy is believed to be unsatisfactory because of the values chosen for the sample sizes, confidence level, and/or the estimated para-

meters, then go back to 6 to choose new values; otherwise revise the model and go to 6.

3. BONFERRONI METHOD FOR CONSTRUCTING THE RANGE OF ACCURACY

Suppose that $[\ell_j^m, u_j^m]$ is a $100(1-\gamma_j^m)\%$ c.i. for μ_j^m and $[\ell_j^m, u_j^m]$, $j = 1,\ldots,k$, may be dependent. Then, using the Bonferroni inequality [10, 16], the probability that all k c.i.'s simultaneously contain their true population means is given by

$$\Pr\{\mu_j^m \in [\ell_j^m, u_j^m] \text{ for all } j = 1,\ldots,k\} \geq 1 - \sum_{j=1}^{k} \gamma_j^m. \tag{13}$$

Thus, the s.c.i. for $\underline{\mu}^m$ are stated as $[\underline{\ell}^m, \underline{u}^m]$ with a joint confidence level that is greater than or equal to $1 - \sum_{j=1}^{k} \gamma_j^m$.

Any parametric or nonparametric [8, 15] statistical procedure can be used to construct the c.i.'s with respect to one of the three approaches given in section 2 and then the Bonferroni inequality can be used to make a simultaneous inference. The range of accuracy of the simulation model can thus be stated with a minimum overall level of confidence. For example, one of the five methods that have been suggested in the simulation literature can be used for constructing a c.i. for the population mean of a model response variable and, in a similar manner, for the population mean of a system response variable. These methods are: (1) the method of replications, (2) batch means methods, (3) regenerative methods, (4) time series methods, and (5) spectral analysis methods. (See [5, 6, 9, 10, 12, 13, 14, 18, 22] for descriptions and details of these methods.)

The Bonferroni method of constructing the range of accuracy will allow either equal or unequal confidence levels of the various response variables to be used in constructing the joint or overall model confidence level as can readily be seen from (13). For the sake of simplicity, equal confidence levels will be used in this paper. (The use of unequal confidence levels, if their use is so desired, is straightforward.)

In this section, classical statistical procedures with the Bonferroni method will be considered for constructing the range of accuracy of a multivariate response self- or trace-driven simulation model with respect to each one of the three approaches of section 2.

Approach I

In this approach, the range of accuracy for the mean behavior of the simulation model is determined by the s.c.i. (1).

Let n be the sample size of independent observations from the model response variables $x_1,\dots,x_k$ that are normally distributed with unknown population means $\mu_1^m,\dots,\mu_k^m$ and variances $\sigma_1^2,\dots,\sigma_k^2$, respectively. Let N be the sample size of independent observations from the system response variables $y_1,\dots,y_k$ that are normally distributed with unknown population means $\mu_1^s,\dots,\mu_k^s$ and variances $\sigma_1^2,\dots,\sigma_k^2$, respectively. (The variances of the model response variables are assumed to be equal to the variances of the corresponding system response variables.) The x_j's may or may not be independent. Similarly, the y_j's may or may not be independent. Then, the c.i.'s for $\mu_j^m - \mu_j^s$, $j = 1,\ldots,k$, each with a confidence level of $1-\gamma_j$, are given [24] as

$$(\bar{x}_j - \bar{y}_j) \pm t_{\gamma_j/2,n+N-2} S_j \sqrt{\frac{1}{n} + \frac{1}{N}}, \quad j = 1,\ldots,k \tag{14}$$

where

$$\bar{x}_j = (1/n) \sum_{i=1}^{n} x_{ji}, \quad j = 1,\ldots,k \tag{15}$$

$$\bar{y}_j = (1/N) \sum_{i=1}^{N} y_{ji}, \quad j = 1,\ldots,k \tag{16}$$

$$S_{mj}^2 = [1/(n-1)] \sum_{i=1}^{n} (x_{ji} - \bar{x}_j)^2, \quad j = 1,\ldots,k \tag{17}$$

$$S_{sj}^2 = [1/(N-1)] \sum_{i=1}^{N} (y_{ji} - \bar{y}_j)^2, \quad j = 1,\ldots,k \tag{18}$$

$$S_j^2 = \frac{(n-1)S_{mj}^2 + (N-1)S_{sj}^2}{n+N-2}, \quad j = 1,\ldots,k \tag{19}$$

and $t_{\gamma_j/2,n+N-2}$ is the upper $\gamma_j/2$ percentage point of the t distribution with degrees of freedom $n+N-2$.

Thus, according to the Bonferroni inequality [10, 16], the probability that all k c.i.'s given by (14) simultaneously contain the true differences between the population means of the model and system response variables is greater than or equal to the joint confidence level of $1-\gamma = 1-\sum_{j=1}^{k}\gamma_j$.

Noting that the half lengths of (14) can be shortened in terms of the sample sizes by minimizing $(1/n + 1/N)$, the objective function and the constraints of the optimization problem (12) are determined for the method of constructing (14) as follows:

$$
\begin{aligned}
\text{Minimize:}\quad & (n+N)/nN \\
\text{Subject to:}\quad & c_m n + C_s N + c_t + C_t \leq B \\
& n + N \geq 3 \\
& n \geq 1 \\
& N \geq 1 \\
& n,N \text{ integer}
\end{aligned}
\tag{20}
$$

The optimal sample sizes n^* and N^* that are used in constructing the schedules in Table 1 are obtained by solving (20). A solution algorithm for (20) is given in [4].

Approach II

In this approach, the range of accuracy for the mean behavior of the simulation model is determined by the s.c.i. (4) that are produced from the s.c.i. (2) and (3) by using the Bonferroni method.

Let n be the sample size of independent observations from the model response variables $x_1,\ldots,x_k$ that are normally distributed with unknown population means $\mu_1^m,\ldots,\mu_k^m$ and variances $\sigma_{m1}^2,\ldots,\sigma_{mk}^2$, respectively. Let N be the sample size of independent observations from the system response variables $y_1,\ldots,y_k$ that are normally distributed with unknown population means $\mu_1^s,\ldots,\mu_k^s$ and variances $\sigma_{s1}^2,\ldots,\sigma_{sk}^2$, respectively. The x_j's may or may not be independent. Similarly, the y_j's may or may not be independent. Then, the c.i.'s for μ_j^m, $j = 1,\ldots,k$, each with a confidence level of $1-\gamma_j^m$, are given [24] as

$$\bar{x}_j \pm t_{\gamma_j^m/2,n-1}\sqrt{S_{mj}^2/n}\quad,\qquad j = 1,\ldots,k, \tag{21}$$

where $\bar{x}_j$ and S_{mj}^2 are given by (15) and (17). Thus, the probability that all k c.i.'s given by (21) simultaneously contain the true population means of the model response variables is greater than or equal to the joint confidence level of $1-\gamma^m = 1-\sum_{j=1}^{k}\gamma_j^m$. Similarly, the c.i.'s for μ_j^s, $j = 1,\ldots,k$, each with a confidence level of $1-\gamma_j^s$, are given as

$$\bar{y}_j \pm t_{\gamma_j^s/2,N-1}\sqrt{S_{sj}^2/N}\quad,\qquad j = 1,\ldots,k \tag{22}$$

where $\bar{y}_j$ and S^2_{sj} are given by (16) and (18). The probability that all k c.i.'s given by (22) simultaneously contain the true population means of the system response variables is greater than or equal to the joint confidence level of $1-\gamma^s = 1- \sum_{j=1}^{k} \gamma_j^s$.

Finally, the s.c.i. (4) with a confidence level of at least $(1-\gamma^m-\gamma^s)$ are obtained by using the interval limits of (21) and (22) as

$$(\bar{x}_j-\bar{y}_j) \pm (t_{\gamma_j^m/2,n-1}\sqrt{S^2_{mj}/n} + t_{\gamma_j^s/2,N-1}\sqrt{S^2_{sj}/N}), \; j=1,\ldots,k. \quad (23)$$

Noting that the half lengths of (21) can be shortened in terms of the sample size by minimizing 1/n, the objective function and the constraints of the optimization problem (12) are determined for the method of constructing (21) as follows:

$$\begin{aligned} &\text{Minimize:} && 1/n \\ &\text{Subject to:} && c_m n + c_t \le B^m \\ &&& n \ge 2 \\ &&& n \text{ integer} \end{aligned} \quad (24)$$

Denoting the largest integer less than or equal to x by $\lfloor x \rfloor$, the optimal solution to (24) is $n^* = \lfloor (B^m-c_t)/c_m \rfloor$ if $\lfloor (B^m-c_t)/c_m \rfloor \ge 2$; otherwise (24) is infeasible. Similarly, N^* can be found and used with n^* in constructing the schedules in Table 2.

Approach III

In this approach, the range of accuracy for the mean behavior of the simulation model is determined by the s.c.i. (5).

Let N be the sample size of independent observations from the model response variables $x_1,\ldots,x_k$ with unknown population means $\mu_1^m,\ldots,\mu_k^m$, and variances $\sigma_{m1}^2,\ldots,\sigma_{mk}^2$. Let N also be the sample size of independent observations from the system response variables $y_1,\ldots,y_k$ with unknown population means $\mu_1^s,\ldots,\mu_k^s$, and variances $\sigma_{s1}^2,\ldots,\sigma_{sk}^2$. Let x_{ji} and y_{ji} be the ith paired observations on the jth model and system response variables. In this case, the observations x_{ji} and y_{ji} occur in pairs so that the two observations are related.

Let $d_{ji} = x_{ji} - y_{ji}$, $i = 1,\ldots,N$ and $j = 1,\ldots,k$. Assuming that d_{ji}, $i = 1,\ldots,N$ are normally distributed for each $j = 1,\ldots,k$, the c.i.'s for $\mu_j^d (= \mu_j^m - \mu_j^s)$, $j = 1,\ldots,k$, each with a confidence level of $1-\gamma_j$, are given [24] as

$$\bar{d}_j \pm t_{\gamma_j/2,N-1}\sqrt{S_{dj}^2/N}\ , \qquad j = 1,\ldots,k \tag{25}$$

where

$$\bar{d}_j = (1/N)\sum_{i=1}^{N} d_{ji}, \qquad j = 1,\ldots,k \tag{26}$$

and

$$S_{dj}^2 = [1/(N-1)\sum_{i=1}^{N}(d_{ji}-\bar{d}_j)^2, \qquad j = 1,\ldots,k\ . \tag{27}$$

Thus, according to the Bonferroni inequality, the probability that all k c.i.'s given by (25) simultaneously contain the true population means of the differences between the paired observations is greater than or equal to the joint confidence level of $1-\gamma = 1-\sum_{j=1}^{k}\gamma_j$.

Noting that the half lengths of (25) can be shortened in terms of the sample size by minizing 1/N, the objective function and the constraints of the optimization problem (12) are determined for the method of constructing (25) as follows:

$$\begin{aligned} &\text{Minimize:} \quad 1/N \\ &\text{Subject to:} \quad (c_m + C_s)N + c_t + C_t \leq B \\ &\qquad\qquad\quad N \geq 2 \qquad\qquad (28) \\ &\qquad\qquad\quad N \text{ integer} \end{aligned}$$

The optimal sample size N^* is obtained by solving (28) and is used in constructing the schedules similar to the ones in Table 1. Denoting the largest integer less than or equal to x by $\lfloor x \rfloor$, the optimal solution to (28) is $N^* = \lfloor (B-c_t-C_t)/(c_m+C_s) \rfloor$ if $\lfloor (B-c_t-C_t)/(c_m+C_s) \rfloor \geq 2$; otherwise (28) is infeasible.

The Roy-Bose method of constructing s.c.i. [16] can also be used with respect to each one of the three approaches to determine the range of accuracy of the simulation model with an overall level of confidence [2]. However, when the precisions of the Bonferroni and Roy-

Bose s.c.i. are compared in terms of their expected lengths, it is found that when constructing an interval for each of the k response variables, the Bonferroni intervals are shorter for the same number of observations, especially when the number of response variables is large [2, 16].

4. EXAMPLES

In this section, the methodology of section 2 and the statistical procedures of section 3 are illustrated for two cases: (1) self-driven steady-state simulation, and (2) trace-driven terminating simulation.

In the first case, a multivariate response self-driven simulation model representing an M/M/1 queueing system is considered. The simulation model is represented by a computerized self-driven model of M/M/1 with an arrival rate (a_r) of 1 and a service rate (s_r) of 1/0.76. Similarly, the real system is represented by a computerized self-driven model of M/M/1 with $a_r = 1$ and $s_r = 1/0.75$.

In the second case, a multivariate response trace-driven simulation model representing an $E_2/E_2/1$ queueing system is considered. The simulation model is represented by a computerized trace-driven model of $E_2/E_2/1$ with $a_r = 0.7$ and $s_r = 0.99$. The real system is represented by a computerized model of $E_2/E_2/1$ with $a_r = 0.7$ and $s_r = 1$. The model and system response variables are observed in pairs by running the simulation model with the same trace-data that drive the real system. The trace-driven simulation is obtained by using the same sequence of random numbers to generate the same arrival pattern to the model and to the system and by using another sequence of random numbers to generate the same pattern of service times in the model and in the system.

The random variate generation is done on an IBM 370 by using the Inverse Transform Method [6] and the multiplicative congruential random number generator $W_n = 7^5 W_{n-1} (\bmod\ 2^{31}-1)$. In each of the simulations in this section, the initial (starting) conditions are assumed to be an empty system and the first arrival takes place at time zero.

4.1 SELF-DRIVEN STEADY-STATE SIMULATION

A multivariate response self-driven simulation model (M/M/1, $a_r = 1$, $s_r = 1/0.76$) representing an M/M/1 queueing system ($a_r = 1$, $s_r = 1/0.75$) has two response variables (performance measures) of interest, namely, the utilization of the server (response variable 1), and the average waiting time of customers in the system (response variable 2). The steps of the methodology of section 2 will be followed for determining the range of accuracy of the simulation model with respect to its mean behavior.

The experimental frame under which the validity of the simulation model is going to be tested is determined by the Poisson arrival process with rate a_r, exponential service times with rate s_r, and the first-come first-served queue discipline. Assuming that the intended application of the model is to analyze the mean behavior of the system with respect to the performance measures chosen, the acceptable range of accuracy for the population means is specified as

$$
\begin{aligned}
-0.035 \le \mu_1^m - \mu_1^s \le 0.035 \\
-0.450 \le \mu_2^m - \mu_2^s \le 0.450 \; .
\end{aligned}
\tag{29}
$$

In Step 4 of the methodology, approach II is chosen to determine the range of accuracy and the statistical procedure of section 3 for approach II is chosen to be used assuming that the underlying assumptions will be satisfied. Supposing that a trade-off analysis is desired, we go to Step 7 to construct the schedules in Table 2.

Estimates of variances of the model and system response variables are needed to construct the schedules. In a pilot run of the simulation model, five independent observations (batches) are obtained in steady-state (after deleting the transient period of the first 10,000 customers) from each of the model response variables by using the method of batch means with a batch size of 5,000 customers. Similarly, five independent observations are obtained from the system and the estimates of the variances are found to be $S_{m1}^2 = 0.00058$, $S_{m2}^2 = 0.093834$, $S_{s1}^2 = 0.000311$, and $S_{s2}^2 = 0.071064$.

Assuming that $c_t = \$150$, $C_t = \$250$, $c_m = \$6$, and $C_s = \$8$, and obtaining n^* and N^* from (24) for several different values of B^m and B^s, the schedules in Table 2 are constructed. Using the data contained in the schedules, Figures 1, 2, 3, and 4 are developed. In Figure 1, the relationships among the minimum half length estimates of the c.i. for the population mean of the first model response variable, significance levels, and the data collection cost are shown. Similarly, Figures 2, 3, and 4 show the relationships for the second model response variable, the first and the second system response variable, respectively.

The trade-offs among the parameters can now be examined by studying the graphs and/or the schedules. It is desirable to choose the confidence level and the sample sizes (number of batches), at a reasonable cost, in such a way that the range of accuracy of the simulation model that will be determined will have half lengths that are no longer than the half lengths of the acceptable range of accuracy specified. As a result of a judgemental analysis of the trade-offs among the parameters, let us assume that the following have been found satisfactory for the intended application of the model: $n^* = 25$, $Tc = \$300$, $\gamma^m = 0.05$, $H_1^{m*} = 0.0117$, $H_2^{m*} = 0.1483$, $N^* = 20$, $TC = \$410$, $\gamma^s = 0.05$, $H_1^{s*} = 0.0097$, and $H_2^{s*} = 0.1469$.

In Step 10 of the methodology, the simulation model and the system are run for 25 and 20 batches in steady-state, respectively, for 5,000 customers in each batch after deleting the first 10,000 customers in the transient period. The following results are obtained: $\overline{x}_1 = 0.7626$, $\overline{x}_2 = 3.1767$, $S_{m1}^2 = 0.00028$, $S_{m2}^2 = 0.05546$, $\overline{y}_1 = 0.7504$, $\overline{y}_2 = 3.0401$, $S_{s1}^2 = 0.00021$, and $S_{s2}^2 = 0.05922$. The univariate normality of the observations collected on the model

FIGURE 1

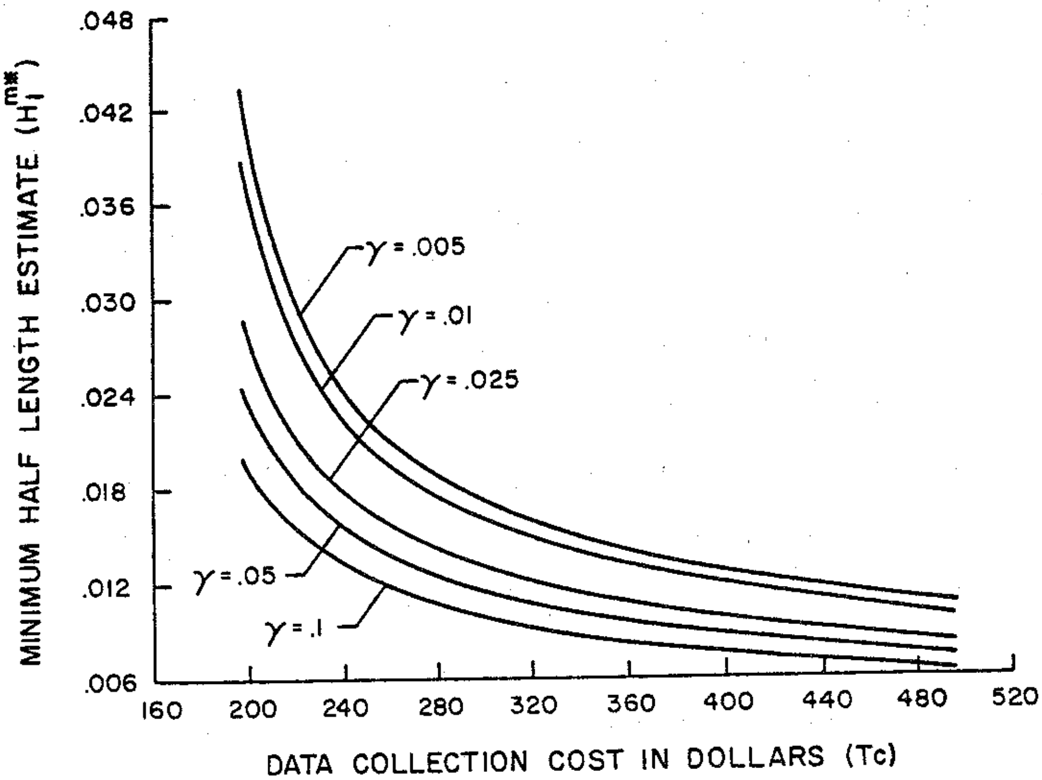


FIGURE 2

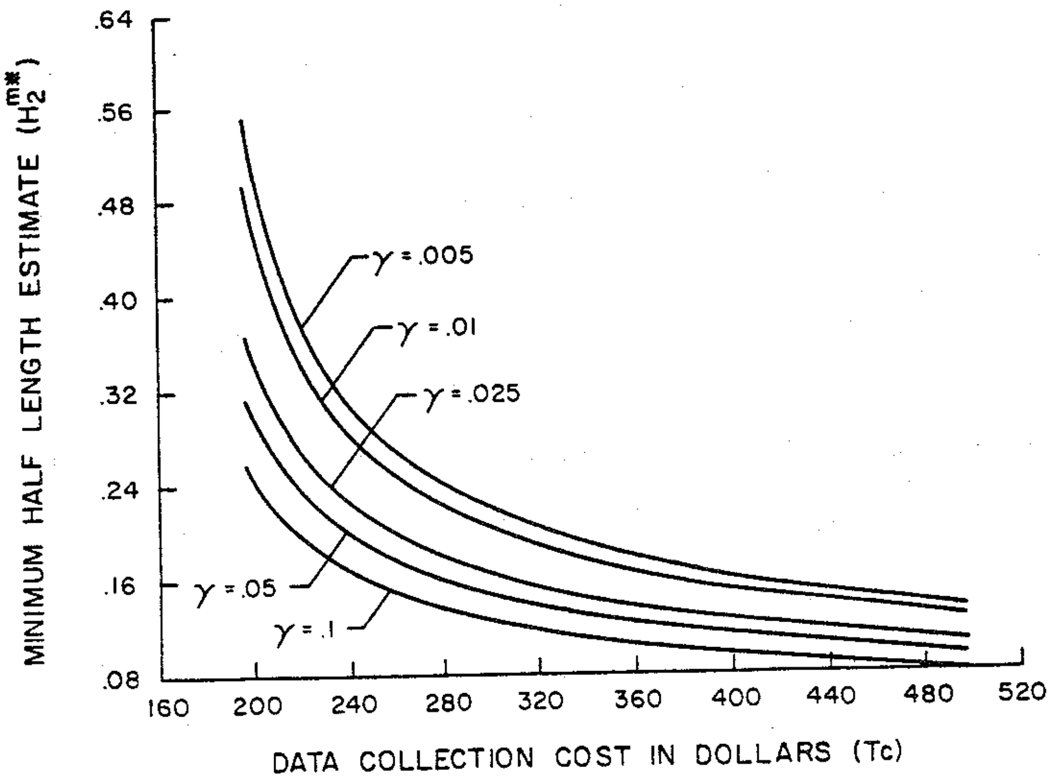


FIGURE 3

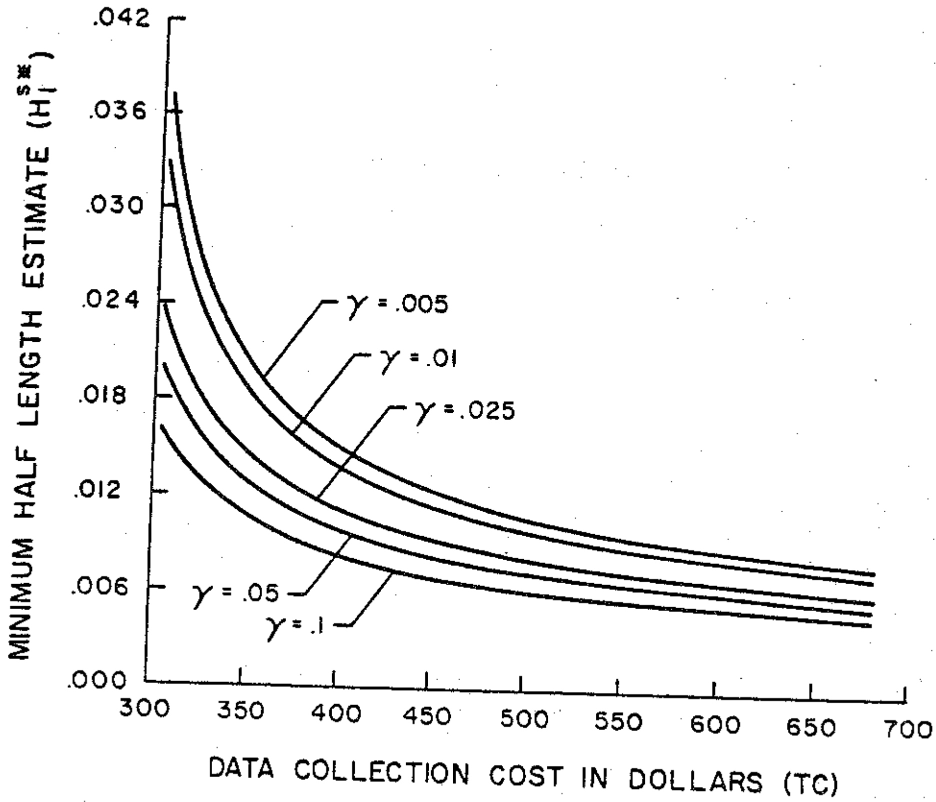
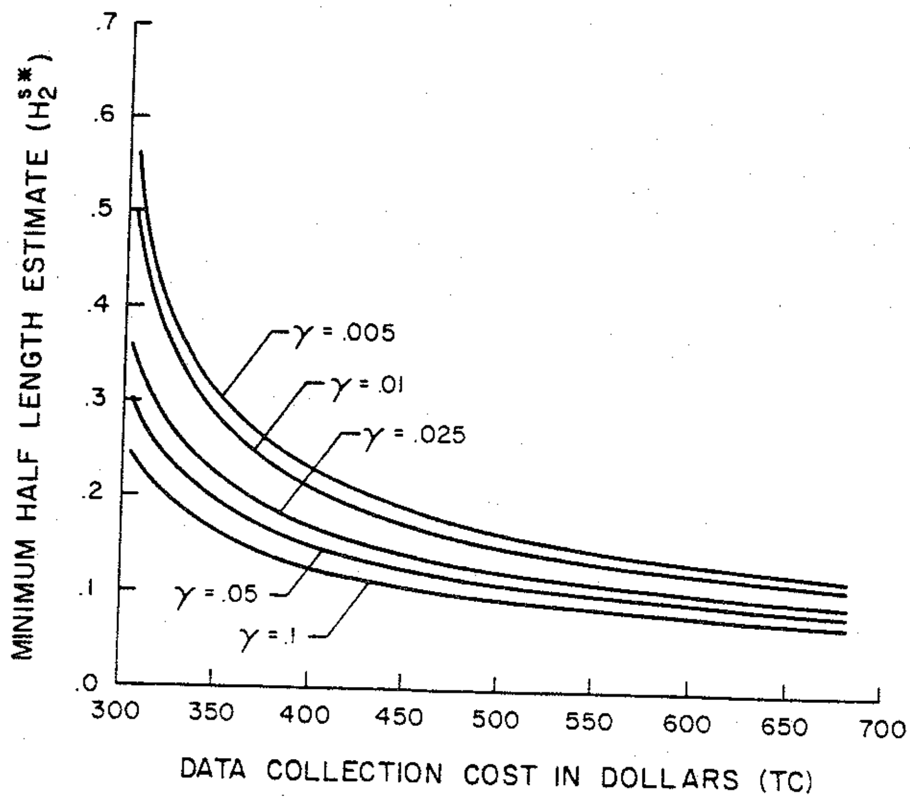


FIGURE 4



and system response variables is tested by using the Box-Cox transformation test [1]. The univariate normality is accepted for $\underline{x}_1$, $\underline{x}_2$, $\underline{y}_1$, and $\underline{y}_2$, with approximate significance levels of 0.66, 0.65, 0.78, and 0.86, respectively.

In Step 11, the range of accuracy of the simulation model is constructed for several values of joint confidence level by using (23) and is presented in Table 3. Since the 80%, 90%, and 95% range of accuracy is completely contained within the acceptable one (29),

TABLE 3. Range of Accuracy of the Self-Driven Steady-State Simulation Model.

M/M/1 Model ($a_r = 1$, $s_r = 1/0.76$, $n^* = 25$) M/M/1 System ($a_r = 1$, $s_r = 1/0.75$, $N^* = 20$)	
$\geq$ 80%	$(\mu_1^m - \mu_1^s) \in [-0.0015, 0.0259]$ $(\mu_2^m - \mu_2^s) \in [-0.0745, 0.3477]$
$\geq$ 90%	$(\mu_1^m - \mu_1^s) \in [-0.0039, 0.0283]$ $(\mu_2^m - \mu_2^s) \in [-0.1115, 0.3847]$
$\geq$ 95%	$(\mu_1^m - \mu_1^s) \in [-0.0059, 0.0303]$ $(\mu_2^m - \mu_2^s) \in [-0.1428, 0.4160]$
$\geq$ 99%	$(\mu_1^m - \mu_1^s) \in [-0.0118, 0.0362]$ $(\mu_2^m - \mu_2^s) \in [-0.2344, 0.5076]$

it is concluded, in Step 14, that we are at least 95% confident that the true differences between the population means of the model and system response variables are contained within the acceptable range of accuracy (29). Thus, the model is valid with respect to the acceptable range of accuracy under the experimental frame and the validity of the model is given by the range of accuracy in Table 3 at a joint confidence level of at least 95%.

4.2 TRACE-DRIVEN TERMINATING SIMULATION

A multivariate response trace-driven simulation model ($E_2/E_2/1$, $a_r = 0.7$, $s_r = 0.99$) representing an $E_2/E_2/1$ queueing system ($a_r = 0.7$, $s_r = 1$) has two response variables of interest, namely, the average queue length for the first 500 customers (response variable 1) and the average waiting time in the system for the first 500 customers (response variable 2). The steps of the methodology will be followed for validating the model with respect to its mean behavior.

The experimental frame is determined by the Erlang-2 arrival process with rate a_r, Erlang-2 service times with rate s_r, and the first-come first-served queue discipline. It is assumed that there is no acceptable range of accuracy specification made. The statistical procedure of section 3 for approach III is chosen in Step 5, assuming that the underlying assumptions will be satisfied. Supposing that a trade-off analysis is desired, we go to Step 7 to construct the schedules similar to the ones in Table 1.

Pilot runs are made to estimate the variances and ten independent observations are obtained in pairs by using the method of replications and the same sequences of random numbers for the model and the system. Then, the variances of differences of the paired observations are estimated to be $S_{d1}^2 = 0.090315$ and $S_{d2}^2 = 0.192665$.

Assuming that $c_t = \$200$, $C_t = \$1,400$, $c_m = \$10$, and $C_s = \$35$, and obtaining N^* from (28) for several different values of B, the schedules are constructed and Figures 5 and 6 are developed. Since there is no acceptable range of accuracy specified, we desire to choose the confidence level and the sample size, at a reasonable cost, in such a way that the half lengths of the range of accuracy will be as short as possible. Let us assume that, as a result of a judgemental analysis, the following parameter values have been found satisfactory: $N^* = 30$, CDC = \$2950, $\gamma = 0.1$, $H_1^* = 0.1122$, and $H_2^* = 0.16388$.

In Step 10, the simulation model and the system are replicated 30 times, for 500 customers in each replication, by using the same sequences of random numbers for the model and the system. The following results are obtained: $\overline{d}_1 = 0.0820$, $\overline{d}_2 = 0.1255$, $S_{d1}^2 = 0.09699$, and $S_{d2}^2 = 0.20161$. The univariate normality of the differences between the paired observations is tested by using the Box-Cox transformation test. The univariate normality is accepted for $\underline{d}_1$ and $\underline{d}_2$ with approximate significance levels of 0.35 and 0.45, respectively.

FIGURE 5

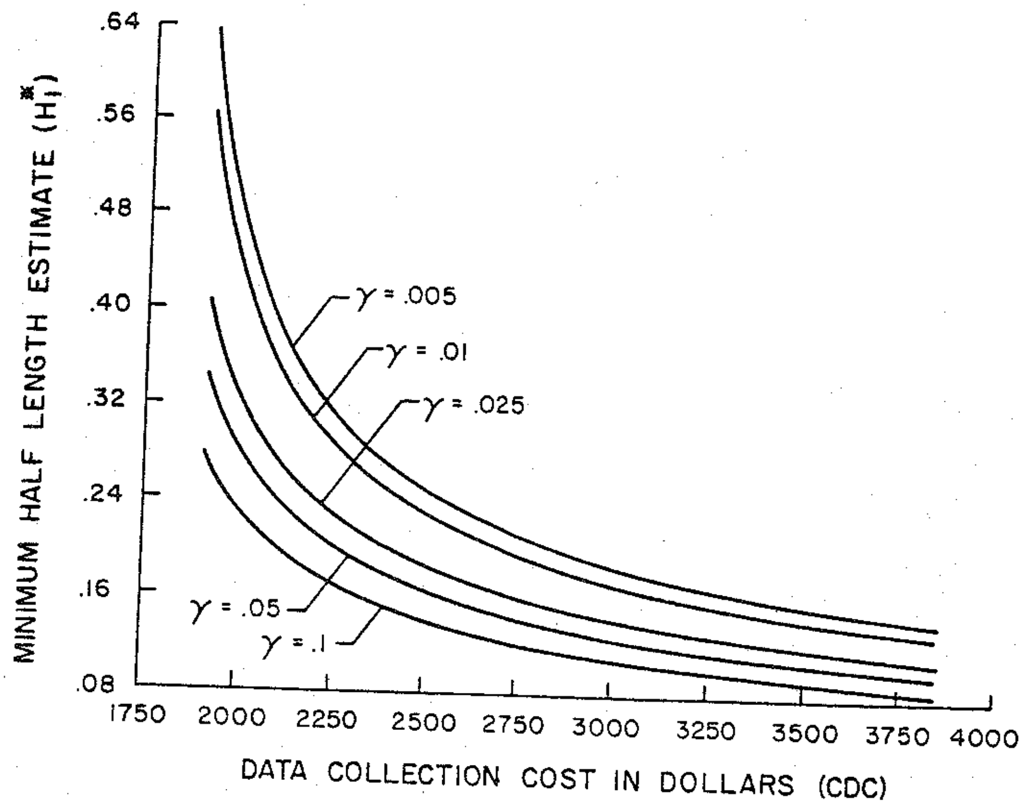
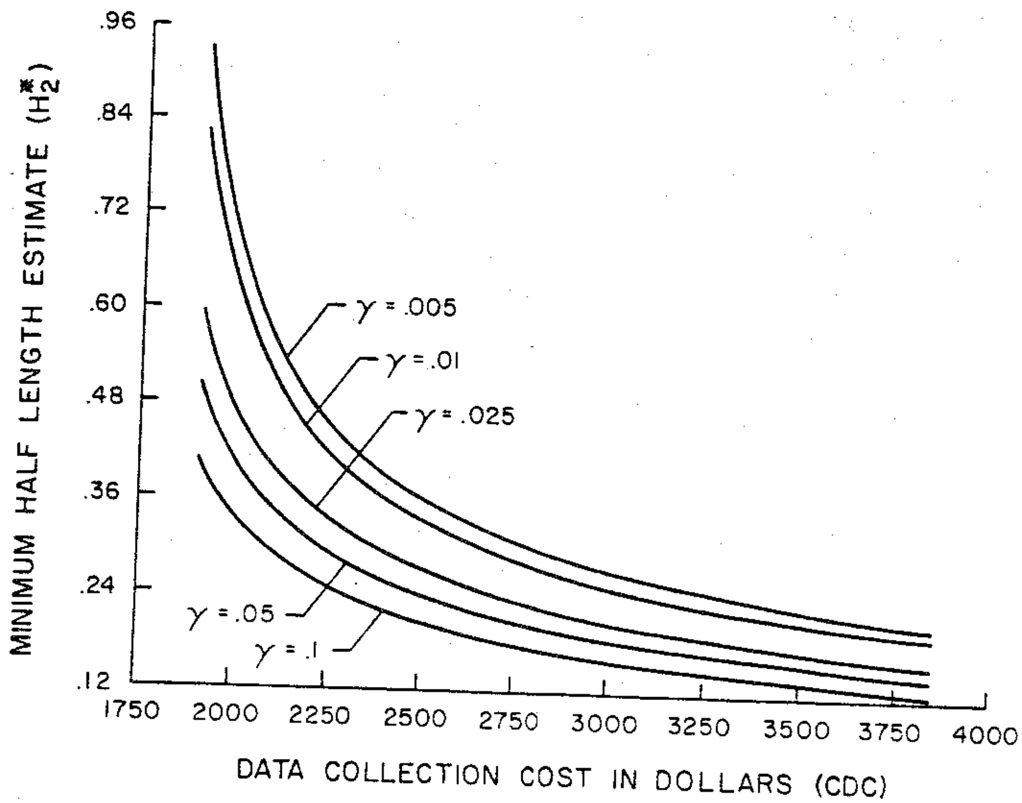


FIGURE 6



In Step 11, the range of accuracy is constructed for several values of joint confidence level by using (25) and is presented in Table 4. Assuming that the range of accuracy for the mean behavior of the simulation model with 97.5% confidence is satisfactory, it is concluded, in Step 16, that the model is valid under the experimental frame and the validity of the model is given by the range of accuracy in Table 4 at a joint confidence level of at least 97.5%.

TABLE 4. Range of Accuracy of the Trace-Driven Terminating Simulation Model.

$E_2/E_2/1$ Model $(a_r = 1,\ s_r = 0.99,\ N^* = 30)$ $E_2/E_2/1$ System $(a_r = 1,\ s_r = 1.00,\ N^* = 30)$	
$\geq$ 90%	$\mu_1^d \in [-0.0343, 0.1983]$ $\mu_2^d \in [-0.0422, 0.2932]$
$\geq$ 95%	$\mu_1^d \in [-0.0540, 0.2180]$ $\mu_2^d \in [-0.0706, 0.3216]$
$\geq$ 97.5%	$\mu_1^d \in [-0.0705, 0.2345]$ $\mu_2^d \in [-0.0944, 0.3454]$
$\geq$ 99.5%	$\mu_1^d \in [-0.1175, 0.2815]$ $\mu_2^d \in [-0.1621, 0.4131]$

5. CONCLUSIONS

A methodology using s.c.i. is presented for validating a multivariate response self- or trace-driven simulation model of an observable system. The methodology provides three approaches for constructing the s.c.i. to express the range of accuracy of a simulation model with respect to its mean behavior. Any statistical procedure can be used to construct the range of accuracy with respect to one of the three approaches. The methodology is illustrated by using the Bonferroni inequality to make a simultaneous inference and state the range of accuracy of the simulation model with a minimum overall level of confidence.

The methodology includes the use of schedules and graphs to show the relationships among the sample sizes of observations, cost of data collection, confidence levels, and estimates of minimum half lengths of the s.c.i. that determine the range of accuracy of the simulation model. The model sponsor, model user, and model builder, individually or together, can perform a trade-off analysis by using the schedules and/or graphs and can make judgement decisions as to what data collection budget to allocate, what data collection method to use, how many observations to collect on each of the model and system response variables, and what confidence level to choose to produce the range of accuracy with satisfactory half lengths.

The methodology using the Bonferroni s.c.i. is illustrated for validating self-driven steady-state and trace-driven terminating simulation models.

REFERENCES

[1] Andrews, D.F., Gnanadesikan, R., and Warner, J.L., "Methods for Assessing Multivariate Normality," in Multivariate Analysis III, P.R. Krishnaiah, Ed., Academic Press, New York, 1973, pp. 95-116.

[2] Balci, O., "Statistical Validation of Multivariate Response Simulation Models," Ph.D. Dissertation, Syracuse University, July 1981.

[3] Balci, O. and Sargent, R.G., "Bibliography on Validation of Simulation Models," Newsletter - TIMS College on Simulation and Gaming, Vol. 4, No. 2, (Spring 1980), pp. 11-15.

[4] Balci, O. and Sargent, R.G., "A Methodology for Cost-Risk Analysis in the Statistical Validation of Simulation Models," Communications of ACM, Vol. 24, No. 4, (April 1981), pp. 190-197.

[5] Crane, M.A. and Lemoine, A.J., An Introduction to the Regenerative Method of Simulation, Lecture notes in Control and Information Sciences, Vol. 4, Springer-Verlag, New York, 1977.

[6] Fishman, G.S., Principles of Discrete Event Simulation, Wiley-Interscience, New York, 1978.

[7] Gafarian, A.V. and Ancker, C.J., "Mean Value Estimation from Digital Computer Simulation," Operations Research, Vol. 14, (1966), pp. 25-44.

[8] Gibbons, J.D., Nonparametric Methods for Quantitative Analysis, American Science Press, Columbus, 1976.

[9] Heidelberber, P. and Lewis, P.A.W., "Regression-Adjusted Estimates for Regenerative Simulations, with Graphics," Communications of ACM, Vol. 24, No. 4, (April 1981), pp. 260-273.

[10] Kleijnen, J.P.C., Statistical Techniques in Simulation, Marcel Dekker, Inc., New York, Vol. 1, 1974, Vol. 2, 1975.

[11] Kobayashi, H., Modeling and Analysis: An Introduction to System Performance Evaluation Methodology, Addison-Wesley, Reading, MA, 1978.

[12] Law, A.M., "Confidence Intervals for Steady-State Simulations, II: A Survey of Sequential Procedures," Technical Report 78-6, University of Wisconsin-Madison, July 1978.

[13] Law, A.M., "Confidence Intervals for Steady-State Simulations, I: A Survey of Fixed Sample Size Procedures," Technical Report 78-5, University of Wisconsin-Madison, March 1979.

[14] Law, A.M., "Statistical Analysis of the Output Data from Terminating Simulations," Naval Research Logistics Quarterly, Vol. 27, (March 1980), pp. 131-143.

[15] Lehmann, E.L., (Dabrera, H.J.M.), Nonparametrics: Statistical Methods Based on Ranks, Holden-Day, San Francisco, CA, 1975.

[16] Morrison, D.F., Multivariate Statistical Methods, McGraw-Hill, New York, 1976.

[17] Rubinstein, R.Y., Simulation and the Monte Carlo Method, John Wiley & Sons, New York, 1981.

[18] Sargent, R.G., "An Introduction to Statistical Analysis of Simulation Output Data," Proc. NATO AGARD Symposium on Modeling and Simulation of Avionics Systems and Command, Control, and Communication Systems, (No. 268), Paris, France, October 1979, pp. 3-1/3-13.

[19] Sargent, R.G., "Validation of Simulation Models," Proc. Winter Simulation Conference, San Diego, California, December 1979, pp. 497-503.

[20] Sargent, R.G., "Verification and Validation of Simulation Models," in Progress in Modeling and Simulation, Edited by F.E. Cellier, Academic, London, 1982.

[21] Schlesinger, S., et al., "Terminology for Model Credibility," Simulation, (March 1979), pp. 103-104.

[22] Schriber, T.J. and Andrews, R.W., "A Conceptual Framework for Research in the Analysis of Simulation Output," Communications of ACM, Vol. 24, No. 4, (April 1981), pp. 218-232.

[23] Sherman, S.W., "Trace Driven Modeling: An Update," Proc. Symposium on the Simulation of Computer Systems IV, (August 1976), pp. 87-91.

[24] Walpole, R.E. and Myers, R.H., Probability and Statistics for Engineers and Scientists, Macmillan, New York, 1978.

[25] Zeigler, B.P., Theory of Modelling and Simulation, John Wiley and Sons, New York, 1976.