Proceedings of the 1994 Winter Simulation Conference
ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila

DESIGNING SIMULATION EXPERIMENTS FOR EVALUATING MANUFACTURING SYSTEMS

James J. Swain
Phillip A. Farrington
Department of Industrial and Systems Engineering
University of Alabama in Huntsville
Huntsville, AL 35899

ABSTRACT

Simulation experiments can benefit from proper planning and design, which can often increase the precision of estimates and strengthen confidence in conclusions drawn from the simulations. While simulation experiments are broadly similar to any statistical experiment, there are a number of differences. In particular, it is often possible to exploit the control of random numbers used to drive the simulation model. To illustrate the methodology described, four examples drawn from manufacturing are used.

1 INTRODUCTION

Stochastic simulations produce random outputs, or samples, which can be used to estimate parameters for the system under study. The statistical design of experiments applies to the planning of simulation experiments, just as in the case of physical experiments. In both cases, statistical design of experiments is employed to provide reliable and efficient estimates and to insure that the objectives of a study can be met.

The design, or planning, of statistical experiments comprises a number of details, including choice of sample size, randomization scheme, and treatment plan. In physical experiments, this randomization specifies how experimental units are assigned to treatments, and the order in which they are run. In simulation experiments, the method of choosing random numbers, through the choice of independent or dependent random sequences, plays a similar, critical role.

In experiments with multiple factors, the design also includes a treatment plan, consisting of the combinations of the different levels of each factor that are to be run (e.g., as a factorial experiment), including the specification of blocking factors. The proper treatment plan can yield efficient estimates of important factors, and avoid threats such as the confounding of several effects which would make it impossible to reach an unambiguous conclusion. Because the properties of statistical estimators are largely determined by how the experiment is performed, a poorly planned experiment cannot be salvaged by analysis, no matter how sophisticated. Hence the importance of planning.

While simulation experiments share similarities with physical experiments, there are some important differences. Because observations within a simulation are generally dependent, the statistical unit will often be based upon replication or batch statistics, rather than on observations from within a simulation run. Given increased user control over a simulation model, additional opportunities are available for design of experiments that are not possible in physical experiments. Because the random numbers (or the random number streams) that drive the simulation are among the specifications of the experiment, they can be manipulated as part of the experimental design in ways that are impossible in physical experimentation (see Nelson, 1992a).

1.1 A Failure Time Model

The first example is not a simulation at all, but will be used to illustrate the analysis of simulation models. An electronic device has a lifetime that we will assume can be represented by a Weibull distribution, with shape parameter 0.5 and scale parameter of 10,000 hours. Our objectives are to estimate the mean lifetime of such devices, and the reliability, $R(t) = R(T > t)$, at $t = 1000$ hours. (In this simplified case, these are known to be $\mu = 20,000$ hours and $R(1000) = .729$.) As posed, this example is a static problem, and could be approached most effectively as an exercise in sampling. However, the outputs can also be considered representative of more complicated, dynamic simulations whose outputs exhibit similar variability and skewness.

1.2 A Tool Crib

A manufacturing operation maintains a tool crib, where tooling is available for machinery on the shop floor. Two clerks are currently employed, and each can handle 25 requests per hour, exponentially distributed. Demand for tooling varies over the day. There is a peak demand of 70 tools per hour at 8 am, which decreases to a constant rate of 40 per hour between 10am and noon. There is a second peak at noon, of 60 parts per hour, which again decreases to 40 parts per hour from 1 p.m. to 2 p.m. Between 2 p.m. and 4p.m., the demand again climbs to a peak of 60 parts per hour. The two clerks handle all requests, and remain on duty until the last request is received at 4 p.m.

The manager would like to consider two alternative staffing policies: employing either another clerk, or using a "helper", who can handle only 15 requests per hour. The two alternatives will be compared to the current operation in terms of the daily average delay, the peak delay, and proportion of delays over 10 minutes.

This example serves to illustrate the case of a terminating, or non-steady state simulation problem. In this case, the tool crib closes each day, with no carry over of customers from one day to the next, and the arrival process varies with time.

1.3 An Inspection Station with Rework

As part of an electronics manufacturing operation, several inspection stations are included to visually inspect solder joints and part placement prior to burn-in testing of the completed component boards. Boards will either pass the test or be held for minor rework when no inspection is taking place. It is desired to know how large buffers for incoming and rework parts are required for these stations, if boards arrive at a rate of 30 per hour, inspection has a capacity of 45 per hour, both according to a Poisson process. Rework takes 5 minutes per board and 5 per cent of the boards require rework.

If we take the view that this production system will operate relatively unchanged for a long period of time, then this system can be treated as a nonterminating system for which steady state parameters are to be determined. Conceptually, we seek parameters (such as the mean) which are independent of starting conditions and represent the behavior of the system over the long run.

1.4 An Electronic Assembly Line

A simulation model of an electronic assembly line consists of 55 stations, including board loaders, solder paste, automated and hand insertion of components, ovens, test and rework, and packaging. The line consists of three sub lines in which automated component insertion takes place, merging into a single line where inspection and testing, rework, solder bake, and packaging takes place. The purpose of the simulation is to better understand the dynamics of the process, and in particular to determine which process variables are most influential in controlling the process throughput.

The simulation model has already been run with no downtime, resulting in a throughput of roughly 5200 parts per day (two shifts). After several iterations of data collection and expert assistance with respect to equipment downtime and repair, the model produces approximately 4000 parts per day, which compares favorably with the 3950 part per day observed in practice. The model has a standard deviation of 550 parts per day and the real system has a standard deviation of 450 parts per day, so the difference of 50 units per day appears relatively small.

Experimentation has shown that the system is dependent upon the downtime. Because the downtime information is hard to collect, a sensitivity experiment is needed to determine which machines affect the throughput the most. Further refinement to the production process can then focus on these machines.

2 NOTATION

This section will identify some fundamental concepts and establish our notation. It is presumed that the reader has some familiarity with the construction of simulation models and basic issues involved in output analysis. Many textbooks on simulation cover these topics, among which Fishman (1972), Law and Kelton (1991), and Banks and Carson (1984) are very well known. Charnes (1993) provides a useful introduction to output analysis. The chapter by Nelson (1992a) offers a practical overview of output analysis, with a valuable emphasis on the central role of random number streams in the design of experiments.

For simplicity, we will denote observations within simulation runs (or replications) W_j, $j = 1, \ldots, m$ for random variables such as delays in queue, and Z_j for (0,1) indicator variables. Because the W_j are autocorrelated within a run, the usual variance estimator, $S^2_W(m)/m$ is likely to be severely biased as an estimator for the variance of the sample, within-run averages $\bar{W}(m)$. This bias is particularly severe when

autocorrelations are large and positive, a situation often observed in queueing and manufacturing cases, in which case variances will be underestimated.

Two approaches are used to obtain independent observations, from which a reliable variance estimate can be constructed. One method is to make independent replications, or runs, of the simulation program, through the use of independent (non overlapping) random numbers. The other method is to group observations; the batch means so obtained will be independent if the batch size is large enough, and they will be normally distributed by the Central Limit Theorem. Independent replications can be used for terminating or nonterminating simulations, whereas the method of batch means only finds application in steady state simulation such as example 3.

In the method of independent replications, the simulation seeds are chosen so that each replication, $i = 1, \ldots, r$ is independent. For run i, let Y_i be the sample average, $\bar{W}$ or $\bar{Z}$, for that run. Let the average over r runs be given by $\bar{Y}(r)$. Then if

$$S_Y^2(r) = \sum_{i=1}^{r} (Y_i - \bar{Y}(r))^2/(r-1),$$

then $S_Y^2(r)/r$ is an unbiased estimator of $\mathrm{Var}[\bar{Y}(r)]$. In the case of indicator variables, $\mathrm{Var}(\bar{Y}(r))$ is estimated by $\bar{Y}(r)(1-\bar{Y}(r))/r$.

For instance, in the failure time model, twenty independent observations of failure times yielded an average MTBF of $\bar{X}(20)$ of 16,070 with $S_X(20)$ of 20,700. An approximate probable error in MTBF is

$$14,680 = t_{19;.975} S_X(20)/\sqrt{20}.$$

Care must be used with this data, as the sample is highly skewed and in this case only about 82% of confidence intervals constructed with a sample size of 20 contained the true MTBF of 20,000 when a confidence level of 95% was used. This coverage error decreased as the sample size increased, but a sample size in excess of 200 would be required to obtain the nominal 95% coverage.

The same 20 observations provided the estimate $\bar{Z}(20)$ of 0.75 for the reliability of the system at 1,000 hours. A probable error for this estimate is

$$0.19 = z_{.975}\{\bar{Z}(20)[1-\bar{Z}(20)]\}^{\frac{1}{2}}/\sqrt{20}.$$

The coverage of confidence intervals based on normal theory was again low due to the skewness of the sample, but nominal coverage could be obtained for sample sizes as low as 100 observations.

3 SAMPLE PLANNING

In many experimental designs, the choice of the proper experimental unit is an important consideration. In simulation studies, the autocorrelation of observations within replications means that the basic experimental unit is either a run average or a batch mean, and in both cases the ideal is a basis for independent, identically distributed observations for each treatment condition.

Once experimental units are decided, a remaining problem is to determine how many are required. One of the basic problems in the design of experiments is to determine the number of experimental runs needed to obtain a specified level of precision or to achieve a specified level of power in a test of hypothesis. We will consider sample size determination for the computation of confidence intervals for specified absolute and relative precision, using a two-stage sampling procedure.

It has often been observed, that design of experiments requires knowledge of the system being studied, and the greater that knowledge, the better the design. In many simulation studies, production runs are preceded by debugging and calibration runs from which much can be learned about a simulation. This information can be valuable in properly designing the study, and making most effective use of the production runs.

3.1 Random Numbers

We return to the question of random number seeds and their importance in simulation design. Most simulations obtain random inputs from pseudo-random numbers, which are transformed into the random variates specified by the model (e.g., normal, exponential). Pseudo-random numbers are deterministic sequences of numbers, whose behavior approximates independent (uniform) variates. The sequence used during the simulation depends on the specification of an initial seed, or equivalently, by the specification of the point within the overall sequence that the seed is to be taken. This latter method, when the seeds are arranged to be (say) 100,000 variates apart, makes the seeding of different generators easier.

The specification of a random number seed is somewhat akin to the problem of sample randomization and blocking encountered in physical experimentation. For instance, although many generators have acceptable higher order (multidimensional) behavior, it is common to allocate different generators to the random processes being simulated as an additional method of insuring independence between runs, much

as complete randomization is used in physical experiments. This allocation also improves synchronization between runs, which is important whenever common random numbers or antithetic variates are used.

However, there are times when complete randomization is not the most efficient sampling plan. For instance, when two treatments are being contrasted, observing paired observations is often done to reduce the variance of the difference. Pairing, by providing a common basis for making the pairwise comparison, causes the paired observations to be correlated, and this correlation can be used to advantage. Suppose observations $(Y_i^{(1)}, Y_i^{(2)})$ represent the paired observations; then $\mathrm{Var}(\bar{Y}^{(1)} - \bar{Y}^{(2)})$ has the following form:

$$\mathrm{Var}(\bar{Y}^{(1)}) + \mathrm{Var}(\bar{Y}^{(2)}) - 2\mathrm{Cov}(\bar{Y}^{(1)}, \bar{Y}^{(2)})$$

which is less than the case of independence whenever the covariance between $\bar{Y}^{(1)}$ and $\bar{Y}^{(2)}$, $\mathrm{Cov}(\bar{Y}^{(1)}, \bar{Y}^{(2)})$, is positive.

In simulation experiments, this pairing is accomplished by using common random numbers; the paired runs use the same random numbers seeds to drive the respective simulations. When the random numbers are synchronized and this strategy is successful (i.e., a positive covariance results), a more precise contrast is obtained.

For instance, in the tool crib model, the two alternative designs for the tool crib are substantially better than the existing system with two clerks. To quantify the difference between the two alternatives, common random numbers can be used so that the same stream of demands for tools are experienced by each of the two systems, and approximately comparable service times are used in each (the services are not exact as the "helper" is slower than the third clerk).

To implement common random numbers, the arrival and service processes can be assigned a different random number generator and seeded separately. A correlation of 0.98 is observed in a set of 15 pairs of average delays, when the model is run in this way. With independently seeded observations, a 95% confidence interval of differences in mean delay is (-1.59,0.08), whereas a similar interval based on paired differences is (-2.21,-1.22). The paired comparison is more conclusive in this case.

When statistics are based on sums, antithetic variates can be used to reduce the variance. As with common random numbers, let $(Y_i^{(1)}, Y_i^{(2)})$ represent paired observations. Then $\mathrm{Var}(\bar{Y}^{(1)} + \bar{Y}^{(2)})$ has the following form:

$$\mathrm{Var}(\bar{Y}^{(1)}) + \mathrm{Var}(\bar{Y}^{(2)}) + 2\mathrm{Cov}(\bar{Y}^{(1)}, \bar{Y}^{(2)})$$

which is less than the case of independence whenever the $\mathrm{Cov}(\bar{Y}^{(1)}, \bar{Y}^{(2)})$ term is negative.

In multifactor and regression experiments, an experimental plan using combinations of independent, common, and antithetic random number streams can be used (Schruben and Margolin, 1978) to improve the precision of estimators.

3.2 Terminating Simulations

As noted, the experimental unit in a terminating simulation is a replication (i.e., a single run of the model), where the length of the simulation run is generally determined by the context of the problem. For instance, if information is desired on system performance per day, or portion of a day (e.g., rush hour), then the model run length should match the time frame for which information is desired.

Where the replication outputs are approximately normally distributed, confidence intervals can be constructed for the parameters of interest using the Student's t-distribution, as mentioned previously. Sample size is typically determined by the desired precision of the confidence interval (e.g., in terms of the half-width of the interval). Since the variance of the simulation output is rarely known, recourse is made to two stage procedures, in which an initial estimate of the variance is made, from which the overall sample requirements can be computed.

When an initial estimate of the variance is not available, n_0 replications are made to estimate the variance of the run statistic. The initial sample (n_0) needs to be large enough to insure that a reliable estimate of the variance can be made; 20 and 40 are values typically cited for this purpose. If an absolute error of size h is desired, where

$$h \geq \mid \bar{Y}(r) - \mu \mid,$$

then the required sample can be determined using

$$n(h) = \min \{i \geq n_0 \mid t_{i-1;1-\alpha/2} S_Y(n_0)/\sqrt{i}\}$$

A large sample approximation is given by

$$n(h) \approx (z_{1-\alpha/2} S_Y(n_0)/h)^2.$$

To obtain a relative error of γ,

$$\gamma \geq \mid \bar{Y}(r) - \mu \mid / \mid \mu \mid,$$

then use

$$n(\gamma) = \min \{i \geq n_0 \mid \frac{t_{i-1;1-\alpha/2} S_Y(n_0)/\sqrt{i})}{\mid \bar{Y}(r) \mid} \leq \gamma'\}$$

where $\gamma' = \gamma/(1-\gamma)$ is used in place of γ because μ is estimated is by $\bar{Y}(r)$.

In cases where simulation output is not normally distributed, nonparametric statistics may be substituted for the usual normal-based statistics, or the sample size may be increased. Non-normal statistics occur particularly when failure times and other extreme value statistics are being observed, as in the failure time data. These outputs exhibit considerable skew, with the result that the actual confidence level of a normal-based confidence interval can be significantly degraded, unless large sample sizes are used.

3.3 Non-Terminating Simulations

The analysis of non-terminating systems presents the modeler with two challenges: eliminating initial conditions bias and obtaining a valid estimate of the variance parameter. The latter is usually managed by production of independent observations through batching or replications. Initial conditions bias is reduced by increasing the run length, in effect, swamping out the initial observations which produce the bias, or by selective deletion of initial observations; the two strategies can also be used in combination.

Two strategies are commonly employed for non-terminating models — either independent replications with initial deletions, or a single run (with or without initial deletions) with approximate independence obtained via batching of the observations. A different set of sampling units is employed in each.

An ideal strategy is the use of very long, independent replications. Where resources are limited, the two strategies represent compromises over two conflicting methods: long runs for initial condition bias versus independent replications for variance estimation. Whitt (1991) provides a thorough discussion of this problem, slightly favoring long runs over shorter, independent replications, while cautioning against either strategy always being best.

The proper choice of a truncation point for the replication deletion strategy, or the proper batch size in batch means procedures, are difficult questions. Charnes (1993) provides a brief list of prominent methods for initialization detection, to which the graphical approach of Welch (1983) can be added. In the Welch procedure observations from several runs are averaged and smoothed, and the resultant plot visually inspected to determine the transient portion of the simulation.

The Welch procedure can be used with the inspection and rework model to help determine a truncation point for the warm up period. Twenty replications of the first 1400 waiting times were averaged, and then a smoothed curved using a 201 adjacent points (each point is the smoothed average of itself and the 100 points before and after it). A truncation point of between 200 and 250 initial parts is suggested here.

Choice of the proper batch size is also more art than science. As a general rule, independence is more critical than the sample size, particularly beyond 20 batches, so that 20 long batches are preferred over 100 shorter ones. Even when using that rule, the run length of the simulation still must be determined. Several sequential procedures have been suggested and are reviewed in Law and Kelton (1982), and these have the advantage that if the initial choice of a run length does not appear to be sufficient, more samples are collected.

4 SINGLE SYSTEM CONSIDERATIONS

When experimentation involves a single system, whether terminating or nonterminating, attention turns to parameter estimation, sensitivity analysis, or perhaps response surface or metamodel estimation for summarization of the output as a function of input parameters of the model.

Parameter estimation is usually done by a confidence interval which summarizes the best estimate of the parameter (e.g., mean waiting time in system) and an interval representing values consistent with the natural variability of the model.

When k parameters are being estimated simultaneously using confidence intervals, each of confidence level $100(1-\alpha)\%$, the overall rate of errors will approach $k\alpha$. A common method to control the experimentwise error rate α_E is based on the Bonferroni inequality, and leads to the choice $\alpha' = \alpha_E/k$ for each individual confidence interval. An alternative approach is to construct a simultaneous interval based upon a joint multivariate interval, Charnes and Kelton (1988).

In cases involving the sample mean, which is based on the sum of sample observations, pairing of replicates that are antithetic can be used, in much the same way that common random numbers are employed to produce positive correlations between differences in random variables. Antithetic sampling involves negative correlations between pairs of observations. A look at example two illustrates the principle: consider pairing runs in which low arrival rates in one run are balanced by a corresponding high arrival rates in the antithetic run. That is, if the first run has a lower than average arrival rate, the second will have above average arrivals (and vice versa). The average of the paired observation will then have less variance than two independent observations, reducing the overall variance. All correlation is within pairs, so the paired means can be used as a basis for

variance computation.

Where sampling is done by monotonic variate generation methods such as the inverse CDF method, antithetic outputs can be constructed by replacing the input random numbers $\{U_i\}$ with the set $\{\bar{U}_i\}$, where $\bar{U} = 1 - U$. Antithetic outputs are harder to arrange where random variate generation is not by the inverse CDF method, however, Fishman and Huang (1983) provide some guidance for other cases with their rotation and reflection sampling methods. Another interesting approach is based on conditional sampling: the antithetic sample is made with respect to a statistic, such as the mean. For a given mean, its antithetic value is computed (based on the sampling distribution), and the antithetic sample is selected to be random, but conditional on the value of that antithetic mean. Cheng (1983) summarizes the procedure and identifies distributions for which conditional sampling can be performed.

Other strategies are available for improving the precision of parameters estimated from simulation. We mention one other: control variate sampling. Here the sample output is observed together with a control variate C which is correlated with the output variable, and whose mean $\mathrm{E}[C]$ is known. We may correct the sample statistic $\bar{Y}$ by an amount proportional to the error between the control variate sample mean $\bar{C}$ and its actual value, $C - \mathrm{E}[C]$,

$$\bar{Y}(\beta) = \bar{Y}(r) - \beta(\bar{C}(r) - \mathrm{E}[C]).$$

This is equivalent to regression of Y_i on the control C_i and is similar to the analysis of covariance procedure sometimes used to correct a sample statistic for the value of a factor which cannot be controlled. Where there is any correlation between the control C_i and the output Y_i the control variate estimator has less variance, $\mathrm{Var}(\bar{Y}(\beta)) \leq \mathrm{Var}(\bar{Y})$ if $\beta = \mathrm{Cov}(\bar{Y}, \bar{C})/\mathrm{Var}(\bar{Y})$. An estimate of β can be obtained from the regression of Y_i on the control C_i.

Regression models are also useful for examining the sensitivity of the simulation output to parameters of the simulation, or in the construction of metamodels used to summarize model behavior over a range of values of the parameters. Such a summary model can be used for prediction, or be incorporated into an optimization procedure. In the sensitivity case, the independent variables X are the settings of the parameters for particular runs around a nominal value. In the metamodel case, the independent variables are the settings of parameters or external factors at which simulations are to be performed. Similar models are constructed in both cases, but the range of the independent variables is often larger for metamodels.

In both cases, response surface methods provide useful experimental designs for the efficient estimation of the regression coefficients. There are well known designs for first and second order models, and even more complicated models.

The regression model is of the form

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where the matrix X represents the settings of the independent variables (the design matrix), $\boldsymbol{\beta}$ the coefficients of the regression model, and the $\boldsymbol{\epsilon}$ error associated with the model, usually assumed random, but consisting of effects from terms not included in the model. A least squares estimator of the regression coefficients is provided by $\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{Y}$. When the replication variance varies between treatment combinations, a generalized least squares regression model can be used.

For first order models, the values of the regression coefficients β_i can be interpreted as the partial derivative of the output with respect to a unit change in the parameter under consideration, $\partial Y/\partial X_i$. Kleijnen (1990) discusses the use of regression models in sensitivity analysis of simulation models, while Biles and Swain (1979) illustrate its use in optimization.

In the electronic assembly example, five factors are investigated to determine their effect on the throughput. To reduce the total number of experiments run, the actual design was a 2^{5-2} fractional factorial design (ABC=CDE). Each point was replicated four times, and there were ten observations taken at the center of the design. This design allows us to estimate the main effects, while losing the ability to determine any of the interactions involved. However, since our purpose is to identify the most important factors, loss of the interactions should not be too important. The results are summarized in Table 1. The standard error of these regression coefficients is 106, from which it is clear that the bottomside placement is the key factor.

The assignment rule of Schruben and Margolin (1978) for common random numbers across all design

Table 1: Factors and Estimated Regression Coefficients for Electronic Assembly Model

Label	Factor	Coefficient
A	Topside Placement	105
B	Bottomside Placement	278
C	Visual Inspection	61
D	Incircuit Test	-64
E	Radial Component Test	158

points, or a combination of common random numbers and antithetic variates, can be readily applied to response surface designs. The utility of these designs and the efficiency of these seeding strategies, has led to a number of developments. Two recent examples include Donohue, Houck, and Myers (1992) for first and second order models, and Tew (1992) for the use of central composite designs.

5 COMPARING MULTIPLE SYSTEMS

A common design problem is the comparison of several systems to determine the best, or at least to screen out the worst contenders so that attention can be focused on a smaller subset of system configurations. For ease of exposition we consider the case of two systems, then the case of more than two systems.

For two systems, common random numbers are employed to increase the precision in the estimation of the contrast between the systems. The analysis is the paired t-test. That is, the observations $D_i = Y_i^{(1)} - Y_i^{(2)}$, are used to construct a confidence interval on the mean difference, μ_D. If $Y^{(1)}$ and $Y^{(2)}$ are correlated, $\mathrm{Var}(\bar{D})$, is less than what would be obtained by independent sampling, but the D_i are independent whatever the correlation between the paired outputs, $(Y_i^{(1)}, Y_i^{(2)})$.

In the second example, common random numbers are implemented in part by blocking on arrival seeds, so that each comparison is exposed to the same pattern of arrivals. Synchronization is obtained whether an inverse CDF method is used to generate the random variates, or not. The output data can be analyzed as a randomized complete block design. As in physical experiments, an improvement results from the elimination of the inter-block variation from the overall error. Individual comparisons can be made with a multiple-range test, such as the studentized multiple range test.

As the number of design alternatives increases, ranking and selection procedures are used in place of the usual procedures for estimation and hypothesis testing. Instead, these methods either select the best system, or identify a subset that contains the best or consists of the best of k systems. Based on procedures developed by Dudewicz and Dalal (1975), these are sequential procedures for determining the necessary sample size to pick subsets so that the probability of making a correct selection is bounded below. Goldsman, Nelson, and Schmeiser (1991) provide an illustration of several methods for selecting the best of several systems.

6 REMARKS

This article as previously published in the 1993 Winter Simulation Conference Proceedings, with the title "Design of Simulation Experiments with Manufacturing Applications". It has been edited and the illustrations expanded for 1994.

REFERENCES

Banks, J., and J. S. Carson, II. 1984. *Discrete-event system simulation.* Englewood Cliffs, New Jersey: Prentice-Hall.

Biles, W. E., and J. J. Swain. 1979. Mathematical programming and the optimization of computer simulations, *Mathematical programming study 11*, 189–207. North-Holland Publishing.

Cheng, R. C. H. 1983. Random Samples with known sample statistics: with application to variance reduction. In *Proceedings of the 1983 Winter Simulation Conference*, ed. S. Roberts, J. Banks, and B. Schmeiser, 395–400. Institute of Electrical and Electronics Engineers, Washington, D. C.

Donohue, J. M., E. C. Houck, and R. H. Myers. 1992. Sequential experimental designs for simulation metamodels. In *Proceedings of the 1992 Winter Simulation Conference*, ed. J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, 539–547. Institute of Electrical and Electronics Engineers, Washington, D. C.

Dudewicz, E. J., and S. R. Dalal. 1975. Allocation of observations in ranking and selection with unequal variances. *Sankya B37*, 28–78.

Fishman, G. S. 1978. *Principles of discrete event simulation.* New York: John Wiley.

Fishman, G. S., and B. D. Huang. 1983. Antithetic variates revisited. *Communications of the ACM 26*: 964–971.

Goldsman, D., B. L. Nelson, and B. Schmeiser. 1991. Methods for selecting the best system. In *Proceedings of the 1991 Winter Simulation Conference*, ed. B. L. Nelson, W. D. Kelton, and G. M. Clark, 177–186. Institute of Electrical and Electronics Engineers, Phoenix, Arizona.

Kleijnen, J. P. C. 1990. Sensitivity analysis of simulation experiments: regression analysis and statistical design. Technical Report, Tilburg University, The Netherlands.

Law, A. M. and W.D. Kelton. 1991. *Simulation modeling and analysis.* 2d ed. New York: McGraw-Hill.

Law, A. M. and W.D. Kelton. 1982. Confidence intervals for steady-state simulations, II: A survey of sequential procedures. *Management Science 28*:

550–562.

Nelson, B. L. 1992a. Statistical Analysis of Simulation Results. In *Handbook of industrial engineering*, ed. G. Salvendy. New York: John Wiley.

Nelson, B. L. 1992b. Designing Efficient Experiments. In *Proceedings of the 1992 Winter Simulation Conference*, ed. J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, 126–132. Institute of Electrical and Electronics Engineers, Washington, D. C.

Schruben, L. W., and B. H. Margolin. 1978. Pseudorandom number assignment in statistically designed simulation and distribution sampling experiments. *Journal of the American Statistical Association 73*: 505–525.

Tew, J. D. 1992. Using Central Composite Designs in simulation experiments. In *Proceedings of the 1992 Winter Simulation Conference*, ed. J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, 529–538 Institute of Electrical and Electronics Engineers, Washington, D. C.

Welch, P. D. 1983. The statistical analysis of simulation results. In *The computer performance modeling handbook*, ed. S. S. Lavenberg, 268–328. Academic Press: New York.

Whitt, W. 1991. The efficiency of one long run versus independent replications in steady-state simulation, *Management Science 37*: 645–666.

ACKNOWLEDGEMENTS

This research was partially supported by NSF grant DDM-9012020. Support by the Chrysler's Huntsville Electronics Division is also gratefully acknowledged. The authors are grateful for the assistance and comments of Dr. Sherri L. Messimer, Dr. John Evans, Monica Cottingham, Hank Czarnecki, and Kim Pickel.

AUTHOR BIOGRAPHIES

JAMES J. SWAIN is an Assistant Professor in the Department of Industrial and Systems Engineering at the University of Alabama in Huntsville. He received his B.A., B.S., and M.S. from the University of Notre Dame in 1974, 1975, and 1977, respectively. From 1977 to 1979 he was employed in the Management Information Department of Air Products and Chemicals. He earned his Ph.D. in Industrial Engineering from Purdue University in 1982, and was an Assistant Professor of Industrial and Systems Engineering at Georgia Tech from then until 1990. He has also taught at Purdue, the University of Miami, and Georgia State University, and held a variety of editorial positions with Lionheart Publishing. His research interests include Monte Carlo methods and variance reduction, statistical modeling, and output analysis. He has been active in the WSC and TIMS College on Simulation, and served as the *Proceedings* Editor for WSC 1992.

PHILLIP A. FARRINGTON is an Assistant Professor in the Department of Industrial and Systems Engineering at the University of Alabama in Huntsville. He received his B.S. and M.S. from the University of Missouri-Columbia in 1983 and 1984. From 1985 to 1987 he was employed as a Manufacturing Engineer by the A. B. Chance Company of Centralia, Missouri. He earned his Ph.D. in Industrial Engineering and Management from Oklahoma State University in 1991. His research interests include the design and analysis of manufacturing systems, the efficient modeling of complex systems, statistical modeling, and output analysis.