*Proceedings of the 1990 Winter Simulation Conference*
Osman Balci, Randall P. Sadowski, Richard E. Nance (eds.)

# STATISTICS AND DETERMINISTIC SIMULATION MODELS: WHY NOT?

Jack P.C. Kleijnen

Katholieke Universiteit Brabant
(Tilburg University)
5000 LE Tilburg, The Netherlands

## ABSTRACT

First, deterministic simulation models are compared with random simulation models and real-life experiments. In deterministic simulation no mathematical statistics is needed in the experimental design and in the Least Squares curve fitting. Further analysis, however, becomes possible if certain statistical models are specified for the fitting errors. In 1987 we proposed normally identically and indepently distributed errors. Recently Sacks et al. proposed dependent errors with a specific correlation structure. Needs for further research are indicated.

## 1. INTRODUCTION

The unique characteristic of a deterministic simulation model is that its response y is fixed, given the values of its input variables in run i $\mathbf{x}_i = \{x_{i1},...,x_{ij},...,x_{ik}\}$ (i=1,...,n). This characteristic means:

$$\text{var}(y|\mathbf{x}_i) = 0. \quad (1)$$

This property distinguishes these models from random simulation models and real-life experiments. *Random simulation models* use pseudorandom numbers; so different seeds produce different y values, in general. In the analysis of these models, the pseudorandom numbers are treated as if they were truly random numbers, which are distributed uniformly and independently. Hence we view the response y as a random variable with variance

$$\text{var}(y|\mathbf{x}_i) = g(\mathbf{x}_i). \quad (2)$$

Some authors assume that $\text{var}(y|\mathbf{x}_i)$ is independent of $\mathbf{x}_i$, so $\text{var}(y|\mathbf{x}_i)$ reduces to a constant (say) $\sigma^2$. We, however, prefer to assume that the variance depends on the combination of input values . So if an n x k experimental design matrix **X** is selected, then each design "point" i has its own variance $\sigma_i^2$ . This variance can be estimated through replication; that is, the simulation model is fed with different random numbers. Other techniques for variance estimation do exist, but we need not discuss them here; see [Kleijnen 1987].

*Risk Analysis* is an interesting combination of deterministic simulation models and Monte Carlo sampling. Examples of deterministic simulation are financial models (which have become popular since spreadsheet software has become widespread) and ecological models. These models depend on a number of inputs that are unknown. Therefore the user may specify a (prior) distribution of possible input values; for example, a beta distribution. Monte Carlo sampling means that the computer generates values from that distribution, feeds those values into the simulation model, and generates output values, which are summarized in an output distribution. Also see [Iman and Helton 1988].

In *real life experiments* repeated observations of the same system generate different responses y, since the system operates in a noisy environment; that is, the environment is not fully controlled and it perturbes the system. Experimental design textbooks consider only this type of experiments, not simulation experiments; see [Box and Draper 1987].

We emphasize that a *random* simulation model is not completely different from a deterministic simulation model: every time the random model is fed with the same input combination $\mathbf{x}_i$ *and* the same pseudorandom number seed, the same response follows. This problem is "solved" by assuming that the pseudorandom number generator produces truly random numbers, *and* that the seed selection does not affect the distribution of these numbers.

So if a given deterministic simulation model is fed with the input combination $\mathbf{x}_i$, it always generates the same response (say) $y_i$. In other words, a deterministic simulation model provides a perfectly controlled world. Is mathematical statistics still relevant in such an "ideal" world? This paper tries to answer that question as follows. In Section 2 we shall first show that basic ideas of the statistical theory of experimental design (originated by Fisher) still apply. The simulation observations resulting from the experimental design can be analyzed through Least Squares curve fitting and "eye balling" of the results. Further analysis, however, requires statistical assumptions. In Section 3 we shall present the statistical model proposed in Kleijnen [1987]: the fitting errors are normally, independently, and identically distributed. Sacks et al. [1989a] replaced the independence assumption by a stochastic process assumption. In Section 4 we shall sketch future research needs.

## 2. DESIGN AND ANALYSIS WITHOUT STATISTICS

In deterministic simulation models the statistical theory of experimental design should still be applied to select the input combinations. For example, it is obviously not smart to change two inputs (say) $x_1$ and $x_2$ simultaneously in the experiment. And changing one factor at a time does not allow the detection of any interactions among inputs. We also point out that many Response Surface designs have been derived under the assumption that noise can be neglected so that only bias is to be minimized; see [Kleijnen 1987, p. 314; Sacks et al. 1989b, p. 420].

In the *analysis*, however mathematical curve fitting and "eye balling" may be used instead of statistical analysis (that analysis includes regression analysis and Analysis of Variance). So we may apply the mathematical criterion of Least Squares to fit a curve to the simulation data; those data consist of $\{x_i, y_i\}$ with i=1,...,n. That curve is a meta-

model of the underlying simulation model; see [Kleijnen 1987]. To validate this metamodel, mathematical criteria can be used such as the multiple correlation coefficient $R^2$. We prefer to predict the response for a new input combination $\mathbf{x}_{n+1}$, and to "eyeball' the relative prediction error $\hat{y}_{n+1}/y_{n+1}$ where $\hat{y}_{n+1}$ is the value predicted by the metamodel and $y_{n+1}$ is the response of the simulation model for the new input combination. A refinement is "cross validation": delete combination i, fit the metamodel to the remaining simulation data $\{\mathbf{X}_{-i}, y_{-i}\}$; predict the deleted response through $\hat{y}_{-i}$; and "eyeball" the relative prediction error $\hat{y}_{-i}/y_i$; repeat this procedure for i = 1,...,n. Examples are provided by Kleijnen and Standridge [1988] who discuss a deterministic simulation of Flexible Manufacturing Systems (FMS), and Kleijnen et al. [1990] who examine a deterministic ecological simulation.

Once the model as a whole is validated, the individual Least Squares coefficients $b_j$ (j=1,...,k) can be studied. In order to determine which factors are most important, the coefficients $b_j$ may be sorted, provided the factors are standardized; see [Bettonvil and Kleijnen 1989]. If, however, the user wishes to identify coefficients that are so small that they actually reflect "noise", then mathematical statistics becomes necessary, as we shall show in the next section.

The regression model can also be used to *predict* the simulation responses at input values *not* contained in the simulation data {X, y}. (In the validation stage we predicted responses for the "old" input X.) Usually interpolation, not extrapolation, is needed, since the simulation data are based on an experimental design that includes extremal values: the "experimental area" covers the "area of interest". The predictors are

$$\hat{y}_2 = \mathbf{X}_2 \mathbf{b} \tag{3}$$

where $\mathbf{X}_2$ denotes the m x k matrix of new input values ,m $\geq$ 1, and **b** is the vector of Least Squares estimators. These predictors are computed faster than the simulation responses $y_2 = h(\mathbf{X}_2)$ are, where h(.) denotes the simulation model. For example, in a study of the Rotterdam container harbor we answered *ad hoc* management questions through the metamodel; in the beginning we were not completely sure that this approach was adequate, so we checked the predictions by running the expensive (random) simulation model overnight; see [Kleijnen et al. 1978].

To quantity the *uncertainty* of these predictions a statistical model must be specified for the fitting errors **e**: (3) yields

$$\text{var}(\hat{y}_2) = \mathbf{X}_2 \, \text{cov}(\mathbf{b}) \, \mathbf{X}_2', \tag{4}$$

where

$$\text{cov}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \, \text{cov}(\mathbf{e})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{5}$$

In the next section we shall discuss statistical models for **e**, which also specify **cov(e)** in (5).

Note that Sacks et al. [1989a] introduced a more complicated predictor that has the nice property that at the observed input combinations the predictor equals the observed response: $\hat{y}(\mathbf{X}) = h(\mathbf{X})$. In random simulation such an equality is not expected, since the observed responses have zero probability of being observed again when the responses are continuous variables and new seeds are employed. Therefore we test whether $E[\hat{y}(\mathbf{X})] = E[h(\mathbf{X})]$; see [Kleijnen 1990].

## 3. STATISTICAL ERROR MODELS

Kleijnen [1987, pp. 163-164] discussed why a statistical error model may be appropriate in deterministic simulation: "Since infinitely many combinations of simulation parameters [x].... are possible, there are infinitely many errors e. The population of these errors has a specific variance, denoted by $\sigma^2$.... Now we *sample* the simulation parameters.... We may perform this sampling randomly or more or less systematically.... in the metamodeling of deterministic simulation we may model the independent variables as random variables. Consequently, the regression parameter estimator [b] being a function of **x** ... becomes random, and so does $\hat{y}$ ... so that **e** = **y** - $\hat{y}$ is random too ..."

Note that "random designs" were elaborately discussed in *Technometrics*, back in 1959. In these designs the input combinations are sampled by flipping a coin so $P(x_{ij}=1)$ = 0.5 and $P(x_{ij}=-1)$ = 0.5; see [Kleijnen 1987, pp. 321-323].

Recently Sacks et al. [1989a, pp. 41-47] modeled the fitting errors "as a realization of a stochastic process in which the covariance structure of [e] relates to the smoothness of the response". They further assumed that **e** is Gaussian with zero expectation; if two design points are further apart along one of the k axes, then the covariance of the two fitting errors decreases exponentially. Their procedure is computationally complex; it uses a supercomputer.

So Kleijnen [1987] proposed the same *marginal* distribution as Sacks et al. [1989a] assumed. Kleijnen implicitly assumed independent errors, whereas Sacks et al. postulated a stationary stochastic process with a particular covariance function. For simplicity's sake we may stick to the model with independent errors that underlies Ordinary Least Squares (OLS). Sacks et al. 's model, however, seems more realistic, if the response surface is smooth. For, suppose the error is positive for some $\mathbf{x}_i$ in the k-dimensional space. Suppose further that we wish to interpolate for $\mathbf{x}_i + \epsilon$, which is a point "close" to the point $\mathbf{x}_i$. Then $\hat{y}(\mathbf{x}_i + \epsilon)$, the response predicted by OLS, tends to underestimate the true response $h(\mathbf{x}_i + \epsilon)$. Sacks et al. 's procedure does not have that unattractive characteristic. Unfortunately they must assume a *specific* covariance function; moreover, their computations for that function are formidable.

A less fundamental discussion of statistical errors in deterministic simulation can be found in Olivi [1984] and Olivi and Pike [1981]. They distinguished two groups of independent variables, namely controlled variables that are supposed to be of major importance, and uncontrolled variables of minor importance. They sampled the minor variables, which resulted in experimental error. Similarly Owen et al. [1989] considered "the response as a function of the most important inputs, possibly with some noise due to the other inputs".

## 4. FUTURE RESEARCH

In the tradition of Popper a scientific model should be refutable. Therefore tests should be developed that allow the user to *reject* the hypothesis that the fitting errors have a specific distribution. In this case the hypothesis stipulates independent normal errors or normal errors with a specific covariance structure. Also see [Kleijnen 1987, pp. 178-179;

Sacks et al. 1989b, p. 417].

The stochastic process specification introduced by Sacks et al. [1989a] looks promising. *Practitioners* will probably apply this approach, once the conceptual and computational details have been worked out. The "classical" regression analysis and experimental design (based on independent errors) have already been applied to many simulation experiments, as the many references in [Kleijnen1987, p. 241] illustrate.

In practice, simulation models often have *many* inputs. For example, Bettonvil [1990] investigated a deterministic ecological model with 281 inputs. He assumed "white" noise: normally, identically, and independently distributed errors. Sacks et al. [1989a,b] limited their approach to small problems with, for example, six inputs; the computational burden of problems with many inputs seems formidable.

We emphasize that metamodels have two goals: prediction and explanation. For *prediction* purposes the metamodel is a black box; the only question is: does the black box predict "well"? *Explanation* means that the user gets insight into the behavior of the underlying simulation model. For example, Kleijnen and Standridge [1988, p. 261] reported that the final metamodel (after the original metamodel was rejected) explained the behavior of the underlying Flexible Manufacturing System: "Statistical techniques... reduce the drawbacks of an empirical technique like simulation, i.e.,... the regression metamodel... helped the authors to better understand how an FMS works!". Sacks et al. [1989a,b], however, concentrate on prediction.

The statistical analysis of deterministic simulation data is *controversial*. Sacks et al. [1989b, p. 435] stated: "In earlier drafts we did attempt to discuss these philosophical matters [Bayesian and frequentist views] more fully, but we gave up due to differences among the authors!" We did not discuss the Bayesian viewpoint at all; Sacks et al. [1989b] did.

## REFERENCES


Bettonvil, B. (1990), *Sequential Bifurcation: Screening For Important Factors*, Tilburg University Press, Tilburg.

Bettonvil, B. and J.P.C. Kleijnen (1990), "Measurement Scales and Resolution IV Design," *American Journal of Mathematical and Management Sciences* (accepted).

Box, G.E.P. and N.R. Draper (1987), *Empirical Model-Building with Response Surfaces*, John Wiley & Sons, New York, NY.

Iman, R.L. and J.C. Helton (1988), "An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Codes," *Risk Analysis 8*, 1, 71-90.

Kleijnen, J.P.C. (1987), *Statistical Tool for Simulation Practitioners*, Marcel Dekker, Inc., New York, NY.

Kleijnen, J.P.C., (1990), "Regression metamodels for simulation with common random numbers: comparison of techniques," Research memorandum, Katholieke Universiteit Brabant, Tilburg, Netherlands.

Kleijnen, J.P.C., J. Rotmans, and G. van Ham (1990), "Techniques for Sensitivity Analysis of Simulation: A Greenhouse Case Study," Research Memorandum, Katholieke Universiteit Brabant, Tilburg, Netherlands.

Kleijnen, J.P.C. and C.R. Standridge (1988), "Experimental Design and Regression Analysis in Simulation: An FMS Case Study," *European Journal of Operational Research 33*, 3, 257-261.

Kleijnen, J.P.C., A.J. Van den Burg, and R.T. Van der Ham (1979), "Generalization of Simulation Results: Practically of Statistical Methods," *European Journal of Operational Research 3*, 1, 50-64.

Olivi, L. (1984), editor. "Response Surface Methodology Handbook for Nuclear Reactor Safety," EUR 9600, Commission of the European Communities, Joint Research Centre, Ispra, Italy.

Olivi, L. and J.D. Pike (1981), "Response surface methodology for the analysis of reactor safety: methodology, development and implementation," ISPRA, Varese, Italy.

Owen, A., J. Koehler, and S. Sharifzadeh (1989), "Comment," *Statistical Science 4*, 4, 429-430.

Sacks, J., S.B. Miller, and W.J. Welch (1989), "Design for Computer Experiments," *Technometrics 31*, 1, 41-47.

Sacks, J., W.J. Welch, T.J. Mitchell, and H.P. Wynn (1989), Design and Analysis of Computer Experiments," *Statistical Science 4*, 4, 409-435.