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

# DESIGNING PSEUDO-RANDOM NUMBER ASSIGNMENT STRATEGIES FOR SIMULATION EXPERIMENTS

Wheyming Tina Song
Chien-Chou Su

Department of Industrial Engineering
Tsing Hua University
Hsinchu, Taiwan, R.O.C.

## ABSTRACT

This paper investigates the use of an effective pseudo-random number assignment strategy, which we call maximum-blocks strategy, for simulation experiments involving the estimation of quadratic response surface metamodel in $3^k$ factorial designs. The maximum-blocks strategy is shown to perform better than well-known existing quadratic-metamodel assignment strategies, such as the IR, CR, and modified-AR strategies.

## 1 INTRODUCTION

Simulation models are often used to make decisions on alternative system designs. A second-level model of a simulation model, referred to as a *metamodel*, can be used to explore the functional relationship between the mean response of the simulation model and a set of simulation inputs (design variables). The simulation input-output functional relationship is often represented as $\underline{Y} = f(\mathbf{X}) + \underline{\epsilon}$, where $\underline{Y}$ is a vector value of simulation outputs, $f$ is the chosen functional form, $\mathbf{X}$ is the design matrix for simulation experiments, and $\underline{\epsilon}$ is the simulation random error affected by the assignment of pseudo-random number streams. Our work focuses primarily on the issue of modeling $\underline{\epsilon}$. That is, how do we best assign the pseudo-random number streams for simulation experiments to increase the precision of the estimation of the metamodel? In this paper we focus our discussion on a quadratic metamodel for $f$ and on a $3^k$ factorial ($3^k$-FAC) design for $\mathbf{X}$. A thorough discussion of choosing a functional form for $f$ is given by Barton (1992).

Because of the simplicity and the utility of metamodels, estimation of metamodels has become an important research topic. Several strategies have been proposed to increase the precision of estimation by inducing a desired correlation structure among the responses. These strategies are called *correlation-induction strategies* or *pseudo-random number assignment strategies*. No matter which correlation-induction strategy is used, Schruben (1979) showed that the sum of variances of the independent estimable parameters is constant for designs in which $\mathbf{X}$ is *saturated* (nonsingular and square) and *orthogonal*. Therefore, such strategies are also called *variance swapping techniques* because the variances are not reduced but *swapped* (or *shifted*) from more important estimators to less important estimators.

To our knowledge, an optimal assignment strategy for linear metamodels in $2^k$ fractional design has previously been proposed (see Schruben and Margolin 1978, Hussey, Myers, and Houck 1987a, Song and Su 1995a, and Song and Su 1995b), but an optimal assignment strategy for quadratic metamodels has not. In this paper, we continue to pursue an *optimal* assignment strategy for the estimation of a quadratic metamodel in $3^k$-FAC designs. The strategy we investigate here is called the *maximum-blocks* strategy. This strategy is based on the same underlying principle as the multiple-blocks strategy proposed by Hussey, Myers, and Houck (1987a), but has been combined with other variance reduction techniques so as to treat the *quadratic* metamodels. We show empirically that the maximum-blocks strategy is superior to the IR, CR, and modified-AR strategies for various input variables. More empirical results illustrating the efficiency of maximum-blocks strategy for quadratic metamodels, but not being restricted to $3^k$-FAC design, can be seen in Song and Su (1994).

## 2 PROBLEM FORMULATION

### 2.1 Quadratic Metamodels in Three-Level Designs

Consider a $3^k$ factorial design (see Montgomery 1991 for details) for the simulation experiment. That is, the simulation experiment consists of $n$ experimental points, each of which specifies the combinations of $k$ input design variables, each with three levels ($n = 3^k$). Without loss of generality, the three levels of design variables can be specified as $+1$, $0$, and $-1$, respectively. During each run of simulation experiments, we record the discrete time series, say $\{Y_{it}; i = 1, 2, \ldots, n, t = 1, 2, \ldots, T\}$, where $Y_{it}$ denotes observed output at time $t$ for experimental point $i$, $n$ is the number of runs, and $T$ is the length of the time series recorded. Let $Y^{(i)}$ denote the sample response statistic of experimental point $i$ (often the average, $T^{-1}\sum_{t=1}^{T} Y_{it}$ ). Our metamodel assumes a relation between the response statistic $Y^{(i)}$ and the corresponding design variables at experimental point $i$, denoted by $\underline{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$, of the form

$$y^{(i)} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \sum_{j=1}^{k} \beta_{jj} x_{ij}^2 + \sum_{u=1}^{k-1} \sum_{j=u+1}^{k} \beta_{uj} x_{iu} x_{ij} + \epsilon_i, \qquad (1)$$

where $\beta_{jj}$, $j = 1, 2, \ldots, k$ is the $j^{\text{th}}$ *quadratic* coefficient. For convenience, we express Equation (1) in matrix form,

$$\underline{Y} = \mathbf{X}\underline{\beta} + \underline{\epsilon}\ , \qquad (2)$$

where the response vector is $\underline{Y} = (Y^{(1)}, \ldots, Y^{(n)})^t$, the vector of the metamodel coefficients of interest is $\underline{\beta} = (\beta_0, \beta_1, \ldots, \beta_k, \beta_{11}, \ldots, \beta_{kk}, \beta_{12}, \ldots, \beta_{k-1,k})^t$, the vector of error terms is $\underline{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^t$, and the $n \times (1 + 2k + k(k-1)/2)$ design matrix is

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \vdots & \vdots & \cdots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \\ x_{11}^2 & x_{21}^2 & \cdots & x_{n1}^2 \\ \vdots & \vdots & \cdots & \vdots \\ x_{1k}^2 & x_{2k}^2 & \cdots & x_{nk}^2 \\ x_{11}x_{12} & x_{21}x_{22} & \cdots & x_{n1}x_{n2} \\ \vdots & \vdots & \cdots & \vdots \\ x_{1,k-1}x_{1k} & x_{2,k-1}x_{2k} & \cdots & x_{n,k-1}x_{nk} \end{bmatrix}^t .$$

A design suitable for separately estimating the $1+2k+k(k-1)/2$ coefficients in Equation (2) is called a *second-order (response surface) design.* The $3^k$ factorial ($3^k$-FAC) design, which is investigated in this paper, is one commonly used second-order design. Other commonly used second-order designs include the central composite (CC) design (Box and Wilson 1951), Box-Behnken (BB) design (Box and Behnken 1960), and small composite design (Draper 1985). For more references, see Montgomery and Evans (1975), Lucas (1976), Donohue, Houck, and Myers (1992), and Hussey, Myers, and Houck (1987b).

### 2.2 The "Min-V|Min-B" Strategy

A natural estimator of $\beta$ in Equation (2) is the ordinary least squares (OLS) estimator,

$$\hat{\underline{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\underline{\mathbf{Y}},$$

each element of which is a linear combination of the responses $Y^{(1)}, Y^{(2)}, \ldots, Y^{(n)}$. Here the quality of the estimator is usually measured by the associated dispersion matrix,

$$\mathbf{\Sigma}_{\hat{\underline{\beta}}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{\Sigma}_{\underline{Y}}\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1},$$

where $\mathbf{\Sigma}_{\underline{Y}}$ denotes the dispersion matrix of $\underline{Y}$. For example, the D-optimal (the determinant of $\mathbf{\Sigma}_{\hat{\underline{\beta}}}$) and the A-optimal (the trace of $\mathbf{\Sigma}_{\hat{\underline{\beta}}}$) are usually selected as the design criteria for comparing the precision of metamodel coefficients. The assumption behind the use of these design criteria (which are based on $\mathbf{\Sigma}_{\hat{\underline{\beta}}}$) is that the hypothesized metamodel in (2) models the input-output relationship exactly. That is, bias is not an issue in estimating $\hat{\underline{\beta}}$. Rather than construct designs based on the premise that there is no bias effect, Box and Draper (1959, 1963) assumed the existence of unfitted third-order terms and incorporated the bias effect into a performance measure

$$\mathbf{J} = \frac{n \int_{\Omega} E\{\hat{y}(\underline{x}) - E[y(\underline{x})]\}^2 d\underline{x}}{\int_{\Omega} d\underline{x}},$$

which is the *average* mse of the *response* over a region of interest. For example, a spherical region $\Omega$ is defined as $\{\underline{x} | x_1^2 + x_2^2 + \cdots + x_k^2 \leq 1\}$.

The performance measure $\mathbf{J}$ can be written as the sum of two components:

$$\mathbf{J} = \mathbf{B} + \mathbf{V},$$

where $\mathbf{B}$ denotes the squared bias component of the average and $\mathbf{V}$ ($= n\ \mathrm{tr}[\mathbf{\Sigma}_{\hat{\underline{\beta}}}\mathbf{M}]$) denotes the variance component of the average, where $\mathbf{M}$ is called the region moment matrix, which contains known constant entries (see Box and Draper 1963 for a discussion of $\mathbf{B}$, $\mathbf{V}$, and $\mathbf{M}$).

Instead of pursuing a $\mathbf{J}$-optimal strategy, Donohue, Houck, and Myers (1992) used $\mathbf{V}$|Min-$\mathbf{B}$ as a criterion to pursue a Min-$\mathbf{V}$|Min-$\mathbf{B}$ strategy. The Min-$\mathbf{V}$|Min-$\mathbf{B}$ strategy is carried out by determining (1) the optimal values of *scaling factors*, which are the level values of designs, (2) the optimal number of *additional center points*, which indicates the number of replications of the center points, and (3) the optimal pseudo-random number assignment strategy. (We do not need to concern the second issue here since there is no center point for $3^k$-FAC design; however center points are required for other second-order designs, such as BB and CC designs.) Donohue, Houck, and Myers (1992) derived the optimal scaling factors to meet the Min-**B** criterion, which is a necessary condition for a Min-**V**|Min-**B** strategy. In this paper, we also use **V**|Min-**B** as our performance criterion, and we use the optimal scaling factor for a $3^k$-FAC design. Our attention is therefore focused on developing the *optimal* assignment strategy in the sense that **V** is minimized.

### 2.3 Assumptions

The assignment strategy is based on controlling the pseudo-random number stream sets that drive the simulation model. For a simulation system containing $g$ stochastic components, a set of $g$ independent pseudo-random number streams, $\mathbf{R} = (R_1, R_2, \ldots, R_g)$ is required to drive the simulation model for a specified design point. The $i^{\text{th}}$ stream of pseudo-random numbers $R_i$ can be composed as $(r_1, r_2, \ldots)$, where $r_i$'s are independent uniform variates over the interval (0,1). Each of the $g$ streams drives one of the $g$ stochastic components in the simulation system. Let us review four major assumptions concerning stream sets (see Schruben and Margolin 1978 for details):

Assumption 1: All sample response statistics have equal variances. Here, for mathematical convenience, without loss of generality, we assume the common variance of the error terms is unity, i.e., $\text{var}(Y^{(1)}) = \text{var}(Y^{(2)}) = \cdots = \text{var}(Y^{(n)}) = 1$.

Assumption 2: Utilization of the same stream set for any two distinct experimental points results in the same positive correlation, i.e., $\text{corr}(Y^{(i)}(\mathbf{R}), Y^{(j)}(\mathbf{R})) = \rho_+,\ 0 < \rho_+ < 1$, for all $i \neq j$.

Assumption 3: Let $R_i = (r_1, r_2, r_3, \ldots)$ be stream $i$ of the stream set $\mathbf{R}$. Its antithesis is then defined as $\overline{R}_i = (1-r_1, 1-r_2, 1-r_3, \ldots)$ and the antithetic stream set is defined as $\overline{\mathbf{R}} = (\overline{R}_1, \overline{R}_2, \ldots, \overline{R}_g)$. Utilization of a pair of antithetic stream sets, $\mathbf{R}$ and $\overline{\mathbf{R}}$, for any two distinct experimental points results in the same negative correlation, i.e., $\text{corr}(Y^{(i)}(\mathbf{R}), Y^{(j)}(\overline{\mathbf{R}})) = -\rho_-,\ 0 < \rho_- < 1$, for all $i \neq j$.

Assumption 4: Utilization of distinct randomly selected stream sets, $\mathbf{R}_1$ and $\mathbf{R}_2$, results in a zero correlation, i.e., $\text{corr}(Y^{(i)}(\mathbf{R}_1), Y^{(j)}(\mathbf{R}_2)) = 0$, for all $i$, $j$.

Furthermore, we must assume that $\rho_- < \rho_+$ to ensure that $\mathbf{\Sigma}_{\underline{Y}}$ is positive definite.

Under the above assumptions, Schruben and Margolin (1978) showed that there are $p_n$ feasible sign patterns for $\mathbf{\Sigma}_{\underline{Y}}$ and therefore $p_n$ possible assignment strategies, where

$$p_n = \sum_{j=1}^{n} 2^{n-j} \sum_{i=1}^{j} \frac{(-1)^{j-i}\, i^n}{i!\,(j-i)!}.$$

Theoretically, the optimal assignment strategy can be identified by means of an exhaustive search over the $p_n$ possible associates $\mathbf{\Sigma}_{\underline{Y}}$. In practice, however, we are not able to check all possible strategies because of the vast size of $p_n$. For an experiment design with only 10 design points, for example, there are over 5 million possible assignment strategies. Moreover, in each assignment strategy, we can generate an infinite number of further cases, each of which corresponds to a different magnitude of the induced correlation, $\rho_+$ and $\rho_-$.

The maximum-blocks strategy investigated in this paper is a systematic procedure for assigning the stream sets for each design point. The examination of other "well-behaved" strategies is given in Song and Su (1994).

## 3 THE MAXIMUM-BLOCKS STRATEGY

### 3.1 Basic Principle

The maximum-blocks strategies for $3^k$-FAC designs are based on the following two principles:

Blocking Principle: All design points are partitioned into as many orthogonal blocks as possible, with the stipulation that the estimation of the metamodel coefficients must not be confounded with these blocks.

Assignment Principle:
- Inside a block: A common stream set is used for design points in the same block.
- Between the blocks: Arbitrarily group all blocks into pairs. A pair of antithetic stream sets is used for the two blocks in each pair and independent stream sets are used for different pairs.

The blocking principle is consistent with the traditional blocking theory: "Block what you can and randomize what you cannot" (Box, Hunter, and Hunter 1978, p. 101). The principle for assignment inside a block is also consistent with *local control* in the blocking theory: "The environment inside a block should be as homogeneous as possible." To illustrate the principle for assignment between blocks, Song and Su (1994) has performed an exhaustive search of all possible strategies that use the blocking principle and the principle for assignment inside a block.

### 3.2 Procedure for the MB Design for $3^k$-FAC Design

The $3^k$-FAC designs are a class of traditional second-order designs. A discussion of the $3^k$-FAC can be found in many textbooks on experimental design, such as Montgomery (1991). The maximum-blocks strategy for the $3^k$-FAC design is as follows.

Step 1. The maximum possible number of independent defining contrasts, say $s$, is chosen to construct blocks for the $3^k$ factorial points such that the metamodel coefficients $\{\beta_j\}_{j=1}^k$, $\{\beta_{jj}\}_{j=1}^k$, and $\{\beta_{ij}\}_{i\neq j}^k$ are not confounded with the block effect. (A systematic procedure proposed by Franklin 1985 can be used to find the desired defining contrasts.) The defining contrasts are then used to divide the design points into $3^p$ blocks, each containing $3^{k-p}$ points.

Step 2. The $3^p$ blocks constructed in Step 1 are grouped into pairs arbitrarily. Thus $(3^p - 1)/2$ groups are specified.

Step 3. Here $(3^p + 1)/2$ independent stream sets, say $(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{(3^p-1)/2}, \mathbf{R}_{(3^p+1)/2})$, are required. The stream set $\mathbf{R}_i$ is assigned to the design points in one block of the $i^{\text{th}}$ pair and $\overline{\mathbf{R}}_i$ is assigned to the design points in the other block of the same pair. Note that there is one unpaired block; the stream set $\mathbf{R}_{(3^p+1)/2}$ is used for the design points in the unpaired block.

## 4 PERFORMANCE

In this section, we compare the performance of four strategies: the IR, CR, modified-AR, and multiple-blocks strategies. The second-order designs investigated are: $3^k$-FAC designs with $k = 3, 4, 5, 6$, where a 1/3 fractional design is used for $k = 6$.

Table 1 shows the V component for the four strategies for $3^k$-FAC designs. The maximum-blocks strategy has the lowest V component and the CR strategy has the highest. The modified-AR strategy has a lower V component than IR for $k = 3$; however, the modified-AR has a higher V component than IR if $\rho_+/\rho_- > 18/5$ for $k = 4$, if $\rho_+/\rho_- > 54/53.07$ for $k = 5$, or if $\rho_+/\rho_- > 54/43.58$ for $k = 6$. We conclude that for all cases studied here, the maximum-blocks strategy is the most effective of the four strategies. The comparison of the maximum-blocks strategy to other second-order designs, such as Box-Behnken design and central composite designs can be seen in Song and Su (1994).

## 5 SUMMARY

This paper has investigated the use of the maximum-blocks strategy for estimating the second-order metamodels in $3^k$-FAC designs. We have shown that the maximum-blocks strategy outperforms well-known existing strategies, including the IR, CR, and modified-AR strategies.

We can summarize the idea behind the maximum-blocks strategy as follows. First, all design points are partitioned into as many orthogonal blocks as possible such that no metamodel coefficients are confounded. These blocks are then arbitrarily grouped into pairs, each of which contains two blocks. Finally, we assign a common stream set to the design points in the same block, antithetic stream sets to the two blocks in the same pair, and independent stream sets to blocks in different pairs or to an unpaired block if there is one.

The application of the maximum-blocks strategy is not restricted to the $3^k$ factorial design. It also performs well for the Box-Behnken design and central composite designs (see Song and Su 1994). It deserves future investigation on how well the maximum-block strategy can be applied in a real world example or a system with a known analytic solution, such as a queueing system.

## ACKNOWLEDGMENTS

This research was supported by the National Science Council of the Republic of China under grant no. NSC 83-0415-E-007-060. The authors wish to thank David Goldsman and Bruce Schmeiser for their helpful comments.

Table 1: V Component for $3^k$-FAC Designs

| Strategy | $k=3$ | $k=4$ | $k=5$ | $k=6$ (1/3 fraction) |
|---|---|---|---|---|
| IR | 12.58 | 19.5 | 27.93 | 37.15 |
| CR | $12.58+14.42\rho_+$ | $19.5+61.5\rho_+$ | $27.93+215.07\rho_+$ | $37.15+205.85\rho_+$ |
| Modified-AR | $12.58-2.39\rho_+-6\rho_-$ | $19.5+5\rho_+-18\rho_-$ | $27.93+53.07\rho_+-54.00\rho_-$ | $37.15+43.85\rho_+-54.00\rho_-$ |
| Maximum-Blocks | $12.58-3.59\rho_+-6\rho_-$ | $19.5-10.5\rho_+-8\rho_-$ | $27.93-15.36\rho_+-6.82\rho_-$ | $37.15-20.72\rho_+-4.82\rho_-$ |

## REFERENCES


Barton, R. R. 1992. Metamodels for simulation input-output relations. In *Proceedings of the 1992 Winter Simulation Conference*, eds. J. Swain, D. Goldsman, R. Crain, and J. Wilson, 289–299.

Box, G. E. P., and D. W. Behnken. 1960. Some new three level designs for the study of quantitative variables. *Technometrics* **2**, 455–475.

Box, G. E. P., and N. R. Draper. 1959. The choice of a second-order design. *J. Am. Statist. Assoc.* **54**, 622–654.

Box, G. E. P., and N. R. Draper. 1963. A Basis for the selection of a response surface design. *Biometrika* **50**, 335–352.

Box, G. E. P., W. G. Hunter, and J. S. Hunter. 1978. *Statistics for experimenters.* John Wiley and Sons, New York.

Box, G. E. P., and K. B. Wilson. 1951. On the experimental attainment of optimum conditions. *J. Roy. Statist. Soc. Ser.* **B 13**, 1–45.

Draper, N. R. 1985. Small composite designs. *Technometrics* **27**, 173–180.

Donohue, J. M., E. C. Houck, and R. H. Myers. 1992. Some designs for quadratic response surface models in the presence of model misspecification. *Management Science* **38**, 1765–1791.

Franklin, M. F. 1985. Selection defining contrasts and confounded effects in $p^{n-m}$ factorial experiments. *Technometrics* **27**, 165–172.

Hussey, J. R., R. H. Myers, and E. C. Houck. 1987a. Correlated simulation experiments in first-order response surface design. *Operations Research* **35**, 744–758.

Hussey, J. R., R. H. Myers, and E. C. Houck. 1987b. Pseudo-random number assignment in quadratic response surface designs. *IIE Transactions* **19**, 395–403.

Lucas, J. M. 1976. Which response surface design is best—A performance comparison of several types of quadratic response surface designs in symmetric regions. *Technometrics* **18**, 411–417.

Montgomery, D. C. 1991. *Design and analysis of experiments.* John Wiley and Sons, New York.

Montgomery, D. C., and D. M. Evans. 1975. Second-order response surface designs in computer simulation. *Simulation* **23**, 335–352.

Schruben, L. W. 1979. Designing correlation induction strategies for simulation experiments. In *Current issues in computer simulation,* eds. N. R. Adam and A. Dogramaci. Academic Press, New York.

Schruben, L. W., and B. H. Margolin 1978. Pseudo-random number assignment in statistically designed simulation and distribution sampling experiments. *J. Am. Statist. Assoc.* **73**, 504–520.

Song, W.-M., and C.-C. Su. 1994. A pseudo-random number assignment strategy for the estimation of quadratic response surface metamodels. Technical Report 94-20, Department of Industrial Engineering, National Tsing Hua University, Taiwan.

Song, W.-M., and C.-C. Su. 1995a. An extension of the multiple-blocks strategy on estimating simulation metamodels. *IIE Transactions on IE Research*, to appear.

Song, W.-M., and C.-C. Su. 1995b. A three-class variance swapping technique for simulation experiments. *Operations Research Letters*, to appear.


## AUTHOR BIOGRAPHIES

**WHEYMING TINA SONG** is a professor in the Department of Industrial Engineering at the National Tsing Hua University in Taiwan. She received a B.S. degree in statistics and an M.S. degree in industrial engineering from Cheng-Kung University in Taiwan. She then received an M.S. degree in applied mathematics and industrial engineering from the University of Pittsburgh, and her Ph.D. in industrial engineering from Purdue University. Her research interests are the statistical aspects of simulation.

**CHIEN-CHOU SU** received his Ph.D. in industrial engineering from the National Tsing Hua University in Taiwan. He received his B.S. degree in mathematics from Chung-Hsing University in Taiwan and his M.S. degree in statistics from Chung-Yang University in Taiwan.