faculteit der economische wetenschappen

RESEARCH MEMORANDUM

[redacted]

TILBURG UNIVERSITY
DEPARTMENT OF ECONOMICS
Postbus 90153 - 5000 LE Tilburg
Netherlands

[redacted]

DESIGN OF SIMULATION EXPERIMENTS

J.P.C. Kleijnen
Katholieke Hogeschool Tilburg
5000 LE Tilburg, Netherlands

A contribution to:
Encyclopedia of Systems and Control,
Editor-in-chief: M.G. Singh
Subject Editor: B.P. Zeigler
Publisher: Pergamon Press, Oxford

December 1982

Design of Simulation Experiments

A simulation model is run for different combinations of its parameter values. Besides the effects of parameter values the effects of different model structures are investigated, but for brevity's sake this introduction speaks of parameters where parameters and model structures are meant (see Sect. 1). A sizable number of runs is necessary for validation, sensitivity analysis, what-if questions, optimization, etc. If the simulation model has many parameters - as most realistic models do - then the exploration of the parameters' effects becomes problematic. The following approaches are popular:

(i) Change one parameter at a time.

(ii) Investigate all parameter combinations.

Both approaches concentrate on relatively few parameters, because investigating all parameters of possible relevance is thought to be impossible. This contribution will present designs that are more efficient and more effective, i.e., fewer simulations runs are needed than in the approaches (i) and (ii); contrary to approach (i) interactions among factors can be detected; moreover, it becomes possible - if needed - to explore the possible relevance of a great many parameters, say, a thousand parameters. Applications of the proposed designs are found in Kleijnen (1979) and in the publications referenced below.

1. Quantitative and qualitative factors

The introduction spoke of a simulation model's parameters and structure. The statistical literature speaks of quantitative and qualitative factors. For example, in a queuing model factor 1 may be the traffic load ρ (= λ/μ with arrival rate λ and service rate μ); factor 2 may be the number of servers (a quantitative, discrete variable); factor 3 may reflect the rule determining the order in which customers are served (a qualitative variable whose changes affect the simulation program). By definition a factor is not constant in the experiment but changes over the simulation runs. This contribution concentrates on factors that can assume only two values or levels in the simulation experiment. For example, the priority rule (factor 3) is either first-come-first-served (FCFS) or smallest-jobs-first (SJF). And the traffic load (factor 1) is studied only at its "low" value, say ρ = 0.70, and at its "high" value, say ρ = 0.95. Restricting the factors to two levels means that it is still possible to detect whether the factors do affect the simulation model's response (apart from pathological situations, i.e., the response curve is hill-shaped and happens to reach identical values at the two selected factor levels). Obviously interpolation or extrapolation to other factor levels, makes no sense when the factor is qualitative. For a quantitative factor it does make sense. However, this contribution concentrates on the qualitative question: do the factors affect the response-yes or no? In a later phase of the investigation, more detailed questions can be asked, such as: how much does the response change when factor 1 changes by one unit; which combination of factor levels yields the maximum response? The techniques of this contribution apply to these more detailed questions but this issue will not be further discussed; see the general literature on experimental designs, e.g., Daniel (1976). Text-books tailored to the needs of simulation practitioners are Kleijnen (1975 and 1983b). Also see the contribution by Kleijnen (1983a)

The matnematical representation is as follows. Suppose the simulation experimenter distinguishes k factors. In the above example k was only three but k may be much higher, say k = 1000. Each factor j (j = 1,...,k) can assume only two values or levels so that k binary variables x_{ij} are introduced: $x_{ij} = -1$ if factor j is "switched off" in run i with the simulation model, and $x_{ij} = +1$ if factor j is switched on in run i (i = 1,...,n). For instance, if factor 3 (priority rule) is FCFS then it might be said that factor 3 is switched off; however, the association between a factor's two levels and the values minus and plus one is completely arbitrary and hence that association may be randomly determined (toss a coin). Consequently, $x_{2i} = -1$ may mean that the number of servers is at its high level. (Note that if a qualitative factor would assume more than two levels, then binary variables assuming the values zero and plus one are necessary; see Kleijnen, 1975.)

2. Three approaches to the design of experiments

Using the binary variables x_{ij} three different approaches to the design of simulation experiment are shown in Table 1, where the constant one is not explicitly displayed but only its sign.

(i) One-factor-at-a-time.

The first run is made in the base position, i.e., all factors are off: $x_{1j} = -1$ (j = 1,2,3). Next each factor is changed in turn while keeping all other factors at their base position: $x_{ij} = +1$ and $x_{ij'} = -1$ (i = j+1 and j ≠ j'). This yields n responses y_i (i = 1,...,n) where n = k+1. Factor effects, say γ_j, are estimated by

$$\hat{\gamma}_j = y_i - y_1 \quad (j = 1,\ldots,k) \quad (i = j+1) \tag{1}$$

Note that the levels +1 and -1 were associated arbitrarily so that the sign of the factor effect has no importance, i.e., in Eqn. (1) the absolute value could have been taken. Approach (i) does yield valid estimators of the factor effects γ. To compare this approach to the other approaches the accuracy of the estimators should be quantified. For simplicity's sake - and in accordance with statistical tradition - the responses are assumed to have a constant variance $\sigma_i^2 = \sigma^2$ $(i = 1,\ldots n)$. Then Eqn. (1) implies that the variance of the effect estimators is a constant, say σ_γ^2:

$$\sigma_\gamma^2 = 2\sigma^2 \tag{2}$$

(ii) All combinations of factor levels: full factorial design.
Approach (ii) requires more runs than approach (i) does. The advantages are:

- Efficiency: the estimators of γ have smaller variance "per run" (see below).

- Effectiveness: interactions among factors can be estimated.

The "common sense" estimator of γ_1 is obtained by averaging the responses observed when factor 1 is on respectively off, and taking the difference between these two averages:

$$\hat{\gamma}_1 = (y_2+y_4+y_6+y_8)/4 - (y_1+y_3+y_5+y_7)/4 \tag{3}$$

In general:

$$\hat{\gamma}_j = \sum_{i=1}^{n} x_{ij}.y_i/(n/2) \text{ with } n = 2^k \tag{4}$$

Now consider the following regression model, where E represents noise (capitals denote random variables):

$$Y_i = \beta_0 + \beta_1 . x_{i1} + \ldots + \beta_j . x_{ij} + \ldots + \beta_k . x_{ik} + E_i \tag{5}$$

Eqn. (4) in the contribution by Kleijnen (1983a) shows that the Least Squares estimates of β_j in Eqn. (5) are equal to half the $\hat{\gamma}$ values defined by Eqn. (4): $\hat{\beta}_j = 0.5\hat{\gamma}_j$ $(j = 1,\ldots k)$. Since the experimenter's purpose it to detect whether the response is sensitive to switching a factor on, β and γ give exactly the same information, namely factor j is important if the t statistic corresponding to either $\hat{\beta}_j$ or $\hat{\gamma}_j$ is significant; see Eqn (16) in Kleijnen (1983a). From Eqn. (4) and the assumption of independent responses with variance σ^2, it follows that

$$\sigma_\gamma^2 = \frac{4}{n^2} \sum_{i=1}^{n} (x_{ij})^2 . \sigma_i^2 = \frac{4\sigma^2}{n} = \frac{\sigma^2}{2} \quad \text{with} \quad n = 2^k = 8 \tag{6}$$

Comparison of Eqn. (6) and Eqn. (2) shows that the variance reduced with a factor 4 while the number of observations increased with a factor 2. In general, the variance reduction becomes dramatic as n increases, i.e., as the number of factors increases (since $n = 2^k$).

Approach (ii) is not only more efficient than the one-factor-at-a-time approach, it is also more effective in that it permits the estimation of interactions among factors. Intuitively formulated, factors 1 and 2 are said to show interaction if the effect of factor 1 also depends on the levels of factor 2. Graphically, interaction means that the response curves are not parallel: $\Delta E(Y)/\Delta x_1$ is not constant but depends on x_2. In mathematical symbols, Eqn. (5) is augumented with the interactions between two factors (denoted by $\beta_{jj'}$, with $j \neq j'$), interactions among three factors ($\beta_{jj'j''}$) and so on. For example, three factors yield:

$$Y_i = \beta_0 + \beta_1 . x_{i1} + \beta_2 . x_{i2} + \beta_3 . x_{i3} + \\ + \beta_{12} . x_{i1} . x_{i2} + \beta_{13} . x_{i1} . x_{i3} + \beta_{23} . x_{i2} . x_{i3} + \\ + \beta_{123} . x_{i1} . x_{i2} . x_{i3} + E_i \quad (7)$$

Interactions among three or more factors are difficult to interpret and are usually assumed to be zero. However, if interactions are assumed to be negligible then approach (iii) becomes relevant.

(iii) Incomplete factorial designs.

If no interactions at all are assumed so that Eqn. (5) holds, then approach (ii) uses $2^3 = 8$ runs to estimate only four parameters, namely, the factor effects γ_j ($j = 1,2,3$) and the overall effect γ_0. These four parameters can also be estimated from the four runs in Table 1 which form a 2^{3-1} design (only a fraction 2^{-1} of all 2^3 combinations is actually simulated). The parameters are estimated strictly analogous to Eqn. (3) or Eqn. (4), e.g.,

$$\hat{\gamma}_1 = (y_2+y_4)/2 - (y_1+y_3)/2 \quad (8)$$

And Eqn. (6) still applies but now with $n = k+1 << 2^k$:

$$\sigma^2_{\gamma} = \frac{4\sigma^2}{n} = \sigma^2 \quad \text{with} \quad n = k+1 = 4 \quad (9)$$

Comparison with Eqn. (2) shows that the approaches (i) and (iii) use the same number of runs but approach (iii) reduces the variance by a factor 2 if $k = 3$. This variance reduction further improves as k increases: in Eqn. (2) the variance remains a constant while in Eqn. (9) n equals k + 1 (more precisely, n equals k+1 rounded upwards to the next multiple of 4; notation: n = [k+1]; e.g. if k = 30 then n = 32; see Kleijnen, 1975).

In general, if the interactions are assumed to be negligible then k factors can be investigated in only n = [k+1] runs. These designs are much more efficient than the one-factor-at-a-time approach is, and they require fewer runs than a full factorial design does. The designs of approach (iii) have been tabulated; see Kleijnen (1975). They are further investigated in the next section.

3. Incomplete factorial designs

If the experimenter has confidence in his assumption of negligible interactions then he can use only n = [k+1] runs. After he has estimated the factor effects he may double-check his assumption by simulating one or more extra factor-level combinations to validate the (first-order) model of Eqn. (5); also see Eqn. (27) in the contribution by Kleijnen (1983a). If from the beginning he has doubts about the assumption of negligible interactions then he should make more than n = [k+1] runs. One attractive design type requires n = 2k runs (more exactly, 2k is rounded upwards to the next multiple of eight, e.g. k = 5 requires n = 16). The latter type yields estimators of the main effects (first-order effects) γ_j or β_j which remain unbiased even if two-factor interactions $\gamma_{jj'}$ or $\beta_{jj'}$ are important. Note that this type still requires not 2^k but only 2k runs, e.g., if k = 8 then not 256 but only 16 combinations are simulated.

If actually interactions <u>are</u> important then incomplete factorial designs may give very misleading results. For instance, design (iii) in Table 1 would result in an estimator of β_3 with expected value equal to β_3 plus the two-factor interaction β_{12} (the column corresponding to interaction β_{12} equals $x_{i1}\ x_{i2}$ and is identical to x_{i3}; also see Eqn. 7).

Many more types of incomplete factorials can be derived. For instance, in a harbor simulation Kleijnen et. al. (1979) studied six factors; not all interactions were thought to be important; actually six of the fifteen two-factor interactions were suspected to be important. A design with sixteen runs yielded estimates of all parameters thought to be relevant. Other types yield unbiased estimators of all two-factor interactions, at the price of more runs. Special designs have been constructed for optimizing k variables, applying Response Surface Methodology (then the variables are not binary but are cardinal); also see Sect. 4 in Kleijnen, 1983a).

In general, experimental design theory shows how estimators of specific factor effects are biased by other effects. For instance, design (iii) in Table 1 yields an estimator of γ_3 which is biased if the interaction γ_{12} is important. The choice of a design is based on the postulated regression model. For instance, the simple model of Eqn. (5) leads to design (iii) of Table 1. More on regression (meta)models can be found in the contribution by Kleijnen (1983a).

4. Too many factors: screening

For pedagogical reasons first designs were presented for the case that the experimenter wishes to investigate relatively few factors. Actually in the first stage of an investigation the simulation model usually contains a great many conceivably important factors. The model builder hopes that of these many factors only a few are really important; otherwise he would have to report that "everything depends on everything else" and a parsimonious, scientific explanation breaks down.

Several approaches are possible :

(i) Simply assume that the response is insensitive to changes in most parameters and structural relationships, and concentrate on a few remaining factors. In this approach the experimenter will never become aware of the limitations of his conclusions.

(ii) Vary all factors randomly. Suppose the experimenter has a list of, say, a thousand potentially important parameters: $k = 1000$. The designs listed in Table 1 require at least 1001 runs. Suppose further that it is impractical to make so many runs (limited computer time, etc.), i.e., a practical restriction is: $n \ll k$. A simple solution is provided by a random design: sample the plus one respectively minus one values in an $n \times k$ table (such as Table 1) with probability a half respectively. This sampling process can be refined (make each column have an equal number of plus and minus one values; reject column j if the correlation with column j' is plus or minus one, where $j' < j$). Factor effects can still be estimated by Eqn. (4). However, these estimators are no longer Least Squares estimators! (Mathematically speaking, Least Squares requires inversion, but the matrix $\underset{\sim}{x} = \{x_{ij}\}$ is singular because $n < k$.) Eqn. (4) results in biased estimators.[1)]

1) $E(\underset{\sim}{\hat{B}}) = E(\underset{\sim}{x}'.\underset{\sim}{y}/n) = (1/n).\underset{\sim}{x}'.E(\underset{\sim}{y}) = (1/n).\underset{\sim}{x}'.\underset{\sim}{x}.\underset{\sim}{\beta}$. In orthogonal designs, such as listed in Table 1, $\underset{\sim}{x}'.\underset{\sim}{x} = n.\underset{\sim}{I}$ by definition. Then $E(\underset{\sim}{\hat{B}}) = \underset{\sim}{\beta}$. If $n < k$ than not all columns of $\underset{\sim}{x}$ can be orthogonal: $\underset{\sim}{x}'.\underset{\sim}{x} \neq n.\underset{\sim}{I}$ so that $E(\underset{\sim}{\hat{B}}) \neq \underset{\sim}{\beta}$.

(iii) Group the k individual factors into g groups of factors ($g \ll k$) and investigate these groups as if they were factors. For example, if $k = 1000$ then $g = 10$ groups each of 100 factors can be formed. Suppose that the individual factors 1 through 100 are important. Changing the individual factors 1 through 100 from minus one (off) to plus one (on) when going from run 1 to run 2 will then not affect the response. Conversely, if runs 1 and 2 yield the same response, then the experimenter may conclude that none of the first hundred individual factors are important (this conclusion can be proved to be correct under mild assumptions; see Kleijnen, 1975). Group-screening enables the experimenter to eliminate many factors after very few runs: $n \ll k$ (since $n = g+1$ or $n = 2g$ where $g \ll k$). After the screening phase the important factors can be further investigated, applying the designs of the preceding section. Applications of group screening are rare. One explanation may be that experimental designs have been applied mainly to physical systems, not to abstract systems like simulation models. In physical systems it is extremely difficult to control a thousand factors from run to run. In simulation, however, all factors are controlled by the computer program and its input, so that group screening should become more popular. This type of design was indeed applied in the simulation of a strategic airlift and a computer system; see (Nolan and Mastroberti 1972) and (Schatzoff and Tillman, 1975).

Bibliography

Daniel, C., 1976, Applications of statistics to industrial experimentation, John Wiley & Sons, Inc., New York.

Kleijnen, J.P.C., 1974/1975, Statistical techniques in simulation

(In two volumes). Marcel Dekker, Inc., New York. (Russian translation: Publishing House "Statistics", Moscow, 1978.)

Kleijnen, J.P.C., 1979, The role of statistical methodology in simulation. In: Methodology in systems modelling and simulation, edited by B. Zeigler, M.S. Elzas, G.J. Klir and T.I. Ören, North-Holland Publishing Company, Amsterdam.

Kleijnen, J.P.C., 1983a, Regression metamodel summarization of model behaviour. Encyclopedia of Systems and Control.

Kleijnen, J.P.C., 1983b, Statistical Tools for Simulation Practitioners. Marcel Dekker, Inc., New York.

Kleijnen, J.P.C., A.J. van den Burg and R.T. van der Ham, 1979. Generalization of simualtion results: practicality of statistical methods. European Journal of Operational Research, 3, pp. 50-64.

Nolan, R.L. and R. Mastroberti, 1972, Productivity estimates of the strategic airlift system by the use of simulation. Naval Research Logistics Quarterly, 19, pp. 737-752.

Schatzoff, M. and C.C. Tillman, 1975, Design of experiments in simulation validation. IBM Journal of Research and Development, 19, no. 3, p. 252-262.

J.P.C. Kleijnen

Catholic University Tilburg

Tilburg, Netherlands

Index

validation, sensitivity analysis, what-if, optimization, regression analysis, least squares.

Table 1. Experimental Designs for Three Factors

(i) One-factor-at-a-time

Run i	x_1	x_2	x_3
1	−	−	−
2	+	−	−
3	−	+	−
4	−	−	+

(ii) All combinations of factor levels

Run i	x_1	x_2	x_3
1	−	−	−
2	+	−	−
3	−	+	−
4	+	+	−
5	−	−	+
6	+	−	+
7	−	+	+
8	+	+	+

(iii) An incomplete factorial design: 2^{3-1} design

Run i	x_1	x_2	x_3
1	−	−	+
2	+	−	−
3	−	+	−
4	+	+	+

IN 1981 REEDS VERSCHENEN:

96 B.B. Van der Genugten
Central limit theorems for least squares estimators in linear regression models including lagged variables.

97 A.J.M. Swanenberg
Rationing and price dynamics in a simple market-game.

98 A.J. Hendriks
Theoretische en practische problemen van het distributieplanologisch onderzoek.
Referaten themadag RSA op 14 mei 1981.

99 A.J.J. Talman en L. Van der Heyden
Algorithms for the linear complementarity problem which allow an arbitrary starting point.

100 J.P.C. Kleijnen
Cross-validation using the t statistic.

101 J.P.C. Kleijnen
Statistical aspects of simulation: an updated survey. Version 1.

102 Dr. Wim G.H. van Hulst
On the concept of divergence in the theory of industrial organization.

103 H. Gremmen, T. van Bergen, J. Hotterbeekx
Waar liggen de Nederlandse comparatieve voordelen? Voorlopige versie.

104 A.L. Hempenius en P.G.H. Mulder
Multiple Failure Rates and Observations of Time Dependent Covariables. (Part 1: Theory.)

105 M.H.C. Paardekooper
A Newton-like method for error analysis. Applied to linear continuous systems and eigenproblems.

106 G. van der Laan en A.J.J. Talman
Note of the path following approach of equilibrium programming.

IN 1982 REEDS VERSCHENEN:

107 Aart J. de Zeeuw
Hierarchical decentralized optimal control in econometric policy models.

108 Arie Kapteyn en Tom Wansbeek
Identification in Factor Analysis.

109 G. van der Laan en A.J.J. Talman
Simplical Algorithms for finding Stationary Points, a unifying description

110 Pieter Boot
Ekonomiese betrekkingen tussen Oost en West Europa.

111 B.B. van der Genugten
The asymptotic behaviour of the estimated generalized least squares method in the linear regression model.

112 Jack P.C. Kleijnen / Anton J. van Reeken
Principles of computer charging in a university-like organization.

113 H. Tigelaar.
The informative sample size for dynamic multiple equation systems with moving average errors.

114 Drs. H.G. van Gemert, Dr. R.J. de Groof en Ir. A.J. Markink
Sektorstruktuur en Economische Ontwikkeling

115 P.H.M. Ruys
The tripolar model: a unifying approach to change.

116 Aart J. de Zeeuw
Policy solutions for Mini-Interplay, a Linked Model for two Common Market Countries.

IN 1982 REEDS VERSCHENEN (vervolg):

117 F.J.M. van Doorne en P.H.M. Ruys
Die Struktur einer Sprechhandlung in Habermas' Forschungsprogramm. Formale Analyse mit den Mitteln des tripolaren Modells.

118 A.J.J. Talman en G. van der Laan
Simplicial Approximation of Solutions to the Non-Linear Complementarity Problem

119 Pieter Boot
The East German Planning System Reconsidered

120 J.P.C. Kleijnen
Regression Metamodel Summarization of Model Behaviour

121 P.G.H. Mulder en A.L. Hempenius
Evaluating life time in a competing risks model for a chronic desease.

122 A.J.J. Talman en G. van der Laan
From Fixed Point to Equilibrium