IDENTIFYING THE IMPORTANT FACTORS IN SIMULATION MODELS WITH MANY FACTORS

Bert Bettonvil and Jack P.C. Kleijnen

Department of Information Systems and Center for Economic Research (CentER)

Tilburg University

Postbox 90153

5000 LE Tilburg, Netherlands

Fax:+3113-663377

E-mail: BERT@kub.nl

kleijnen@kub.nl

Version 3: November 1994

Version 2: September 1993

Version 1: January 1993

IDENTIFYING THE IMPORTANT FACTORS IN SIMULATION MODELS WITH MANY FACTORS

BERT BETTONVIL and JACK P.C. KLEIJNEN

Department of Information Systems and Center for Economic Research (CentER),

Tilburg University, Postbox 90153, 5000 LE Tilburg, Netherlands

Fax:+3113-663377; E-mail: BERT@kub.nl

Simulation models may have many parameters and input variables (together called factors), while only a few factors are really important (parsimony principle). For such models this paper presents an effective and efficient screening technique to identify and estimate those important factors. The technique extends the classical binary search technique to situations with more than a single important factor. The technique uses a low-order polynomial approximation to the input/output behavior of the simulation model. This approximation may account for interactions among factors. The technique is demonstrated by applying it to a complicated ecological simulation that models the increase of temperatures worldwide.

KEYS WORDS: Sensitivity Analysis; Selection and Ranking; Experimental Design

1. INTRODUCTION

This paper presents a novel technique for *screening*, which is the process of searching for a few important factors among a great many potentially important factors. Whereas in real-life experiments (such as industrial experiments) it is impossible to control hundreds of factors, computer simulation programs may have many inputs. For example, stochastic (random) simulation models of computer and telecommunication networks

have many 'factors': parameters (such as service and arrival rates), input variables (say, number of servers), and behavioral relationships (queuing discipline or priority rule). Another example is the case study in this paper: we shall illustrate our technique by applying it to a complicated ecological deterministic simulation that models the effects of increasing quantities of carbon dioxide (CO_2) and other gases on temperatures worldwide; we shall screen as many as 281 factors. Morris (1991) also states: 'Often, computational models have as many as several hundred input variables'.

Screening may be applied in the initial phase of many simulation studies. Screening yields a list of important factors, which can be further explored in later experiments. Our technique also estimates the magnitudes of the individual important factor effects. In a follow-up experiment the input-output behavior of the simulation model can be studied more precisely (including possible curvature of the response surface); this is not part of our paper. Screening and more precise fitting is discussed in, among others, Welch, Buck, Sacks, Wynn, Mitchell, and Morris (1992).

Let us have a closer look at the options that statistical analysts have in practice. In this paper we concentrate on simulation models with N factors, N being so big that it is practically impossible to execute a number of simulation runs (observations) roughly equal to N. This means that one simulation run requires so much computer time that the users of the simulation results do not wish to wait until all N runs will have been executed. For example, our ecological simulation model with 281 factors takes 15 minutes of computer time per run, so 281 runs would have taken roughly 70 hours of computer time. Of course, computer speed increases steadily with technological progress, at decreasing costs. But information technology (IT) increases the users' aspiration levels: they want to explore bigger, more realistic simulation models. So in practice, many simulation models will continue to have so many factors that substantially fewer than N runs can be executed. This assumption eliminates the type of designs developed by Morris (1991).

Note that even if the number of runs were about N, then the statistical analysts would still have to make *assumptions*. They may assume that the input/output behavior can be approximated by a first order polynomial. That assumption permits designs with roughly N runs, such as one at a time designs and Plackett-Burman designs, which are *resolution III* designs; see Kleijnen (1987, pp.301-303) and Morris (1991, p.162).

If the analysts do not know for sure that all interactions between *pairs* of factors are negligible, then

they may double the number of runs to 2N: *resolution IV* designs based on the foldover principle (see (9) and section 3).

If they are certain that besides main effects there are only interactions between *pairs* of factors, then they can use a design of *resolution V*. These designs require many runs; for example, even if N is as small as 9 then 128 runs are needed; see Kleijnen (1987, p.309).

If they do not wish to assume that *quadratic* effects (curved response surfaces) are negligible, then each factor must be simulated for at least three values, and the number of runs further increases. A well-known class of designs are *central composite* designs; see Kleijnen (1987, pp.312-316) and Morris (1991, p.162).

As we mentioned before, simulation models may have too many factors to perform N runs. Then practitioners often assume that they know which factors are unimportant, and *only a few intuitively selected factors* are investigated in a resolution III design, usually a one at a time design; see Van Groenendaal (1994).

A general scientific principle when studying such large systems is *aggregation*; for example, in economics the production volumes of individual companies are aggregated into the General National Product (GNP). In the case of large simulation models, *group screening* may be applied: individual factors are combined into groups. Group screening goes back a long time: see Jacoby and Harrison (1962), Li (1962), Patel (1962), and Watson (1961). Our technique also uses group screening, but we use a different design and analysis; our technique turns out to be more efficient (see §2 and §5).

To know *for sure* that individual effects do not cancel out within a group, certain *assumptions* must be made. All classic group screening techniques assume a first order polynomial (no interactions, quadratic effects, etc.) with known signs for the main effects. Kleijnen (1975) proved that two-factor interactions do not bias the main effect estimators if a resolution IV design is used for the group factors.

If practitioners feel that they know the directions of *all but a few* factor effects, then the latter factors may be treated separately, while the former factors are investigated by group screening. In our ecological case study we examined the very small group of factors with unknown directions, using the traditional one-at-a-time method; however, for most factors the system analysts felt confident when they had to specify the signs of the factor effects (see §3).

Note that *Latin Hypercube Sampling* (LHS) is an experimental design that is often used to investigate deterministic simulation models (computational models), but LHS tackles a fundamentally different problem: the inputs are assumed to have a certain distribution; see Kleijnen (1994) and Morris (1991).

Our technique treats the simulation model as a *black box*. It approximates the input-output behavior of the simulation model through a *low-order polynomial*; this approximation is also known as a metamodel. More specifically, we start with a first order polynomial, which in Analysis of Variance (ANOVA) is known as a model with main effects only. If deemed necessary, we augment that approximation with two-factor interactions. The advantage of the black box approach is that it is simple and applies to *all* types of random and deterministic simulation. The disadvantage is that the approach cannot exploit the special structure of the simulation model at hand (our technique does use prior knowledge about the *signs* of the factor effects). Techniques that do exploit this structure are Perturbation Analysis and Score Function; see Ho and Cao (1991) and Rubinstein and Shapiro (1993). These techniques, however, are meant for random discrete event simulations, and they do not consider screening.

The main objective of this paper is to inform practitioners about a novel technique for screening large simulation models. Space limitations do not permit the presentation of all technicalities, so for mathematical details we refer to the doctoral dissertation, Bettonvil (1990).

We organize the paper as follows. In § 2 we explain the basics of our technique, first assuming the simplest approximation, namely a first order polynomial model; next we augment this approximation with interactions among factors. In § 3 we present a case study: we apply our technique to the ecological simulation mentioned above. In § 4 we revisit the assumptions of our technique. In §5 we examine the efficiency of our technique. In §6 we give conclusions.

2. BASICS OF THE SEQUENTIAL BIFURCATION PROCEDURE

Let s( ) denote the mathematical function that is specified by the computer simulation program; v_j the j*th* factor among the N factors of that program (j=1,2,...,N); and y the simulation output:

$$y = s(v_1, \ldots, v_j, \ldots, v_N, r). \quad (1)$$

2.1 First Order Approximation and Concomitant Experimental Design

The simplest approximation of the simulation model (1) is a *first order polynomial*, which is also known as a metamodel with main effects β_j and the overall mean β_0. We call the resulting approximation errors 'negligible', if they are 'small' relative to the factor effects. In this section we assume that the approximation errors are zero.

In section 1 we mentioned that all group screening techniques assume *known signs* for the main effects. Therefore we can define L_j^* as the level of factor j that generates a low value for the output y, and H_j^* as the level of factor j that generates a higher value for y, provided this factor has any effect at all. Note that if an increase of factor j reduces the output, then we have $L_j^* > H_j^*$. Observe further that the domain of the factor is limited by the area of experimentation; also see section 4 (Assumptions revisited).

It is convenient to transform the original quantitative factors v_j in (1) linearly into *standardized* variables x_j that assume values 0 and 1, where 0 corresponds with the level that generates a lower output.

Note that qualitative factors (such as priority rule) with two levels only (say, FIFO versus LIFO) may be represented in a similar way. If a qualitative factor has more than two levels, then we may consider only those two levels that generate extreme outputs (the minimum and the maximum outputs), assuming we know which two levels give extreme values. However, we have no experience with qualitative factors so far.

Because we shall revisit the assumptions of our technique (§4), we number all our assumptions:

Assumption 1: A first-order polynomial gives 'negligible' approximation errors over the experimental domain of the simulation model.

Assumption 2: The *direction* of the influence that a factor has on the output y, *if* that factor has any effect at all, is *known*.

Because of assumption 1 the simulation model (1) is approximated by (identical to)

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_j x_j + \ldots + \beta_N x_N. \quad (2)$$

In sensitivity analysis the overall mean (β_0) is not interesting (when using the approximation for prediction purposes, β_0 is important). Indeed, β_0 is not estimated by our technique, as we shall see.

Assumption 2 implies that all main effects are non-negative: $\beta_j \geq 0$. In this section we call all positive effects important (they are bigger than the approximation errors, which are zero in this section).

Next we shall present out *experimental design*. We introduce the symbol $y_{(j)}$ to denote the simulation output observed when the factors #1 through #j are switched on while the remaining factors are off; that is, the first j original factors have the values H_j^* (the corresponding standardized variables are 1) and the (N - j) remaining original factors have the values L_j^* (their standardized variables are 0). Then (2) yields

$$y_{(j)} = \beta_0 + \beta_1 + ... + \beta_j \quad (j=0,1, ... ,N). \tag{3}$$

Consequently the sequence $\{y_{(j)}\}$ is nondecreasing in j.

We explain our technique informally, using the example in Jacoby and Harrison (1962): there are 2^7 = 128 factors but only the factors #68, #113, and #120 are important (obviously our technique does not know that there are only three important factors, let alone the magnitudes of these factor effects). We further introduce the symbol $\beta_{j\text{-}j'}$ to denote the sum of β_j through $\beta_{j'}$ (j' > j); for example, $\beta_{1\text{-}128}$ denotes the sum of β_1 through β_{128}. Our technique is *sequential*; that is, the factor combinations to be simulated (observed) depend on the results of previous combinations simulated. In stage #0 we always observe $y_{(0)}$ (all factors low) and $y_{(N)}$ (all factors high). An important sum of effects $\beta_{j\text{-}j'}$ makes us obtain a new observation; see the symbol ↓ in Figure 1. In the example we notice $y_{(0)} < y_{(128)}$, so we infer $\beta_{1\text{-}128} > 0$.

Our technique also uses *bifurcation*: we split the group of factors into two (sub)groups of equal size. This procedure assumes that the total number of factors is a power of two. If this is not the case, then we proceed as if we added *dummy factors* after the last factor so that the new set of factors consists of 2^m factors with 2^m the smallest power of two not smaller than the original number of factors, see §5. In the example, however, we have 2^7 factors.

So in the next stage (#1) we observe $y_{(64)}$. We see $y_{(64)} = y_{(0)}$. Hence we *eliminate* all factors in the first half. So after only three observations we have already inferred that the factors 1 through 64 have no effects and that there is at least one important factor in the second half of the group of 128 factors.

In stage #2 we concentrate on the remaining factors (labelled 65 through 128). Again we bifurcate. And we go on, until after only 16 observations we have identified the three important factors and have computed their individual main effects, indicated by ↑ in Figure 1.

INSERT FIGURE 1

Note that our technique has the following characteristics. If we find that a (sub)group of factors is 'important', then this group contains at least one important factor. In the next stage we split this group into two subgroups. We obtain one new observation, switching on all factors from #1 up to halfway the important original group; for example, in stage #2 we switch on factors #1 through #96. We compare this new observation with two old observations, to infer the sums of factor effects. That comparison leads to the elimination of at most one subgroup; obviously it is impossible that both subgroups are unimportant and eliminated. A formal description of sequential bifurcation is given in Bettonvil (1990, pp.13-22).

Note further that if there is only one important factor, then our procedure is simply classical binary search; see, for example, Brassard and Bratley (1988).

2.2 Quantifying Importance

So far we assumed that a factor j is declared important if its effect β_j is positive, not zero. Our technique implies that a group of factors that seems important at the end of a stage, is split into two smaller subgroups; in the next stage those two subgroups are investigated. The new observation splits the sum of effects of the factors in the original group into two sums: sum #1 equals the sum of the effects of the factors in the first subgroup; similarly, sum #2 equals the sum of effects of the second subgroup. Obviously no individual effect can exceed the sum of which it is a part.

Now consider the example in Figure 2: there are 24 factors; the sizes of their main effects are shown by the vertical solid bars; assumptions 1 and 2 hold (no negative main effects, no interactions, no approximation errors).

Suppose that stage #0 (not displayed in Figure 2) yields $y_{(0)}$ = 0.0 and $y_{(24)}$ = 2388.2. Hence after two observations we know an upper limit for each of individual effect, denoted by U(2), which is 2388.2 -- 0.0 = 2388.2. If we were interested only in effects exceeding 2388.2, we would stop! Suppose, however, that we are also interested in effects smaller than 2388.2.

In the next stage we take observation #3, namely $y_{(16)}$, and compare this observation with the two preceding ones: we compute $y_{(24)} - y_{(16)}$ and $y_{(16)} - y_{(0)}$. We concentrate on the largest difference; say, $y_{(24)} - y_{(16)}$. Then at this moment this value is the most stringent upper limit we have for *all* individual effects (β_1 through β_{24}). Next we make this limit sharper, taking an observation within the group that gives this limit; in the example, we take $y_{(20)}$. Again we compare this observation with its two predecessors: we compute $y_{(24)} - y_{(20)}$ and $y_{(20)} - y_{(16)}$. And we compute the largest difference over all groups that have not been split so far: $U(4) = \max (y_{(24)} - y_{(20)}, y_{(20)} - y_{(16)}, y_{(16)} - y_{(0)})$. Obviously this maximum does not exceed the preceding maximum, $y_{(24)} - y_{(16)}$.

So in general, as we proceed through various stages, groups get smaller and upper limits decrease. The dotted horizontal lines in Figure 2 are the upper limits after 11 through 17 observations. These limits decrease from U(11) = 383.6 to U(17) = 139.7. For example, after 13 observations we know that the factors #17 and #20 are 'important'; that is, they have effects larger than 217.9, contrary to the remaining factors. After 17 observations we have identified the eight most important factors (#14, #17, #18, and #20 through #24), all exceeding 139.7, and we know that the remaining factors have effects smaller than 139.7. (Clustering is bad in this example, so this is not a favorable configuration for our technique; see §5.)

INSERT FIGURE 2.

This example demonstrates a characteristic that we consider a strength of our technique: it does not require the practitioners to specify *a priori* a critical value (say) δ that must be exceeded by an effect in order to be declared important. As soon as the users consider a sum of effects to be small, the investigation of this subgroup can be stopped. Their critical value is not completely subjective: δ depends on the problem at hand, i.e. is it worthwhile to perform extra observations to find factors with effects smaller than δ, while already knowing all effects exceeding δ. We shall see an application in section 3.

2.3 Two-factor Interactions and Concomitant Experimental Design

By definition, interaction means that the effect of a specific factor depends on the levels of other factors. It is simple to adapt our technique to approximations expanded with either two-factor interactions only or with all interactions. The latter approximation seems less realistic; for completeness' sake it is presented in Bettonvil (1990, pp.220-227).

In this section we replace Assumption 1 by the following assumption.

*Assumption 1**: A first-order polynomial augmented with cross-products between two factors gives 'negligible' approximation errors over the experimental domain of the simulation model.

Hence the simulation model (1) is approximated, not by (2), but by

$$y = \beta_0 + \sum_{j=1}^{N} \beta_j x_j + \beta_{12} x_1 x_2 + ... + \beta_{N-1,N} x_{N-1} x_N. \quad (4)$$

It is convenient to use a different *parametrization*: $z_j = -1$ iff $x_j = 0$, and $z_j = 1$ iff $x_j = 1$. Then (4) is equivalent to

$$y = \gamma_0 + \sum_{j=1}^{N} \gamma_j z_j + \gamma_{12} z_1 z_2 + ... + \gamma_{N-1,N} z_{N-1} z_N, \quad (5)$$

where

$$\gamma_{jj'} = \beta_{jj'}/4 \text{ with } j,j' = 1,...N \text{ and } j \neq j', \quad (6)$$

$$\gamma_j = \beta_j/2 + \sum_{j' \neq j} \beta_{jj'}/4 \text{ with } j = 1,...N, \quad (7)$$

$$\gamma_0 = \beta_0 + \sum_{j=1}^{N} \beta_j/2 + \sum_{j=1}^{N-1} \sum_{j'=j+1}^{N} \beta_{jj'}/4 \ . \quad (8)$$

In the first-order approximation, all main effects were non-negative ($\beta_j \geq 0$), because of Assumption 2 (known directions) and the appropriate choice of factor levels (L_j^*, H_j^*). What does *known directions* of factor

effects mean when there are *interactions*? The standard literature on experimental design and ANOVA defines the *main effect* of factor j as the difference between

(a) the average output when factor j is switching *on*, when averaging over all 2^{N-1} combinations of the remaining (N-1) factors, and

(b) the average output when factor j is switched *off*, again averaging over all 2^{N-1} combinations.

It is easily verified that in our case the main effect of factor j is $2\gamma_j$. Therefore we replace assumption 2 by

*Assumption 2**: All N main effects in (5) are non-negative: $\gamma_j \geq 0$.

We call a factor *important* iff its main effect is 'important'. This means that a factor with an unimportant main effect but an important interaction goes undetected; we shall return to this issue in section 4 (Assumptions revisited).

Next we shall present out *experimental design*. We can simply adapt our technique, as follows. We *double* the number of runs: if in the first-order approximation we observed $y_{(j)}$ (defined in (3)), we now also observe the output with the first j factors switched *off* and the remaining factors switched on: $y_{-(j)}$ (obviously $y_{-(0)} = y_{(N)}$ and $y_{-(N)} = y_{(0)}$). We call $y_{-(j)}$ the *mirror* observation of $y_{(j)}$. The idea of doubling the number of runs in order to identify main effects in the presence of two-factor interactions resembles the *foldover* principle, presented in Box and Wilson (1951, p.35).

It is simple to prove that $y_{(j)} - y_{-(j)}$ is a non-decreasing function of j, and that

$$\{y_{(j)} - y_{-(j)}\} - \{y_{(j-1)} - y_{-(j-1)}\} = 4\,\gamma_j. \tag{9}$$

So now our technique proceeds as follows. In stage #0 all factors are off and all factors are on respectively, which yields $y_{(0)}$ and $y_{(N)}$. Because of (9) we compute

$$\{y_{(N)} - y_{-(N)}\} - \{y_{(0)} - y_{-(0)}\} = 2y_{(N)} - 2y_{(0)}. \tag{10}$$

Suppose this contrast is 'important': $\gamma_j \geq 0$ for some j. Then we proceed to stage #1 and obtain $y_{(N/2)}$ and $y_{-(N/2)}$. Because of (9) we compute

$$\{y_{(N/2)} - y_{-(N/2)}\} - \{y_{(0)} - y_{-(0)}\} \tag{11}$$

and

$$\{y_{(N)} - y_{-(N)}\} - \{y_{(N/2)} - y_{-(N/2)}\}. \tag{12}$$

Iff (11) yields zero, we eliminate the first half of the N factors; otherwise we proceed to the next stage. The second half of factors is treated analogously. Ultimately we arrive at the individual important factors and we compute (9).

In section 2.2 we explained that it is not necessary to a priori quantify a threshold δ for the importance of the effects β_j in (2). Now β_j is replaced by γ_j in (5), which is the approximation augmented with two-factor interactions and reparametrized. The results of section 2.2 apply to interactions, mutatis mutandis, as we shall see in the case study of section 3.

3. CASE STUDY: A LARGE ECOLOGICAL SIMULATION MODEL

To demonstrate the application of our technique in practice, we apply this technique to a *large* simulation model that has *not* been constructed for the purpose of testing our technique (so our test is 'unbiased'). The users of this large model are interested in screening their model. This model is a deterministic computational model of the 'greenhouse effect': how do temperatures increase worldwide, as the fraction of gases such as carbon dioxide (CO_2) in the atmosphere increases? The model has been developed by the Dutch National Institute of Public Health and Environmental Protection (RIVM). The model is called IMAGE, Integrated Model to Assess the Greenhouse Effect. It consists of many modules (submodels): for details we refer to Rotmans (1990).

An essential part of IMAGE represents the worldwide circulation of CO_2 in the atmosphere, the oceans, and the terrestrial biosphere. We apply our technique to this part. The output y denotes the simulated CO_2 concentration in the year 2100. We investigate as many as 281 factors, and specify a metamodel with two-factor interactions; see (5). The ecological experts (not we) must specify the ranges of these 281 factors.

Further we use the experts' prior knowledge to label the factors such that their effects are sorted in increasing order, we hope (also see §4 where we revisit assumptions).

Stage #0 gives $y_{(0)}$ = 988 and $y_{(281)}$ = 1496. Since it seems obvious that the difference between these two outputs is 'important', we proceed to stage #1, where we observe $y_{(256)}$ and its mirror observation $y_{-(256)}$; see §4. And so on.

Table 1 gives the results after 77 pairs of observations. The upper limits for the main effects γ_j have decreased to 2.5 (remember Figure 2). We have found 15 factors with main effects γ_j that exceed 2.5. The other 281 - 15 = 266 factors have effects smaller than 2.5 and, in view of the factor effects already found, are declared unimportant. (In this table some ranges start from the high value in order to make the signs positive; see § 2.1).

INSERT TABLE 1

How are we assured that the results of our technique applied to this simulation, are correct? This simulation is so complex that nobody really knows the true factor effects! Prior to our findings, the ecological experts expected eight factors to be important, namely the factors labeled #250, #246, #19, #281, #13, #86, #20 (see Table 1) and a factor called DIFF (not listed in Table 1). So it turns out that seven of their factors are also declared important by our technique, and one of their factors is not declared important. We verify the results of our technique as follows.

We execute an additional experiment with only these eight factors. Such a small number of factors can be observed in a *resolution IV design* (by definition, a resolution IV design gives estimators of all main effects not biased by possible two-factor interactions; the estimators for the two-factor interactions are biased and estimate specific sums of these interactions). For eight factors a resolution IV design requires 16 observations. Table 2 gives the results. In that table, factor #250 is relabeled as factor (1), factor #246 becomes (2), and so on (left hand side of table). The symbol (1,2) (right hand side of table) denotes the interaction between the factors 1 and 2, but actually resolution IV means that the interaction estimator is aliased with other two-factor interactions. The symbol (0) denotes the estimate of the grand mean γ_0 in the approximation (5), computed from the resolution IV design. The other symbols have obvious meanings.

INSERT TABLE 2.

Comparing Tables 1 and 2 shows that our technique indeed detected the seven factors expected to be important by the ecologists. Our technique also gives roughly the same point estimates as the resolution IV design gives. The latter design shows that the factor DIFF is indeed 'unimportant', namely smaller than 2.5 (the upper limit reached by our technique after 77 pairs of observations). Table 2 suggests that two-factor interactions are unimportant for these eight factors.

We point out that some of the important factors detected by our technique, were neglected by the ecologists in their original experiments (namely factors listed in Table 1 but not in Table 2). So our results are surprising and they provide new insight for the users.

4. ASSUMPTIONS REVISITED

Assumption 1 stated that the first-order polynomial (2) gives zero approximation errors; assumption 1* claimed that adding two-factor interactions (resulting in (5)) gives zero approximation errors. First we called factors important when they have positive main effects β_j in (2) and γ_j in (5) respectively. Next we redefined important effects, introducing the upper limits illustrated by Figure 2.

To examine these assumptions, we first emphasize that in screening we are interested, not in local (marginal) effects, but in *overall* effects: what happens if a factor changes from its lower level (L_j^*) to its higher level (H_j^*)? The experimental area should be selected such that the simulation model is valid in that area. A simulation model is *valid* only when the variables remain within a certain area, which Zeigler (1976) calls the 'experimental frame'. Indeed, in the ecological case study we found that we had to reduce the experimental area because the two input combinations corresponding with all inputs 'off' and 'on' respectively, gave an unrealistically low and high value respectively for the response variable (that is, the ecologists rejected these two response values).

It is natural to express the experimental frame in terms of the original variables v_j; also see (1). Related to the experimental frame is the *experimental area*. For example, when investigating an anti-fever

medicine, one experiments with (simulated) patients who have a temperature between (say) 37 and 42 degrees centigrade (not between 0 and 100 degrees). Our technique uses the experimental area defined by the N-dimensional rectangle that corresponds with (L_j^*, H_j^*).

We further distinguished v_j (the original variables of the simulation model), x_j (standardized variables that are either 0 or 1), and z_j (-1 or +1); see (1) through (5). Factor effects should be measured at the *center* of the experimental area; in other words, the model in z is relevant, not the model is x (the model in x measures factor effects at one of the corners of the experimental area, namely the point where $x_j = 0$ for all j). For details we refer to Bettonvil and Kleijnen (1990).

Note that in section 2.1 we used the variables x, because they simplify the explanation of our technique when there are only main effects. If there are no interactions, then both x and z give essentially the same factor effects: $\gamma_j = \beta_j/2$; see (7) (the factor 2 is well-known in experimental design theory).

Let us next consider three cases:

a) Monotonic response surfaces, that is, the simulation output increases as an input increases.

b) Non-monotonic response surfaces. Then the output might happen to be approximately the same at the two extreme values investigated in our technique (L^* and H^*). In that case we falsely infer that this factor is unimportant. Non-monotonicity may be quantified by quadratic effects ($\gamma_{jj}z_j^2$ is added to the metamodel). Such effects can be estimated by observing the output at a value halfway the two extremes. We have not investigated this complication, but it certainly deserves future research; also see Andres and Hajas (1993).

c) Interactions only. Consider the following academic example: the simulation response (y) is determined solely by the interaction between two factors (say, $y = \gamma_{12}z_1z_2$). Then our technique fails (unless the analysts switch to a logarithmic scale: $\log y = \log(\gamma_{12}) + \log(z_1) + \log(z_2)$).

In other words, our technique is not a panacea. But then again, no technique is: a scientific analysis always requires assumptions. We propose that in order to make progress, the analysts explicitly specify an approximation to the complicated, nonlinear input/output behavior of simulation models. Also see section 1 (Introduction).

We can also drop the assumption of perfect fit, and model the *approximation errors* (say) ε statistically, using the standard assumptions of *white noise*: errors are additive and distributed normally; iff the approximation is adequate, the errors have zero expectation; the errors have constant variance (say) σ^2;

and they are independently distributed. For white noise Bettonvil (1990, pp.49-142) developed three statistical techniques. These techniques work only if the signal/noise ratio is high, say, $\gamma_j/\sigma > 6$. For such high signal/noise ratios, however, we may as well ignore noise. Ignoring noise is not so bad: in screening there are so many factors that only those factors with a high signal/noise ratio should be detected.

Note that we might use a covariance stationary process (instead of white noise) to model the approximation errors. Sacks, Welch, Mitchell, and Wynn (1989) use such a process to model the systematic effects of the inputs (not the noise part). If the simulation model itself is stochastic, then the intrinsic noise may exceed the fitting errors. We have not investigated these complications.

Next we shall investigate the remaining assumptions. Assumption 2 stated that all N main effects in the first order approximation (4) are non-negative ($\beta_j \geq 0$); assumption 2* stated that all main effects in the approximation with two-factor interactions (5) are non-negative ($\gamma_j \geq 0$). So these assumptions postulated *known signs* of the main effects.

When a factor does not meet the assumption of known signs, the simulation response may show so: the response decreases when switching this factor on, assuming its negative effect exceeds the sum of the positive effects of the other factors within its group. In our case study this happened twice; however, the response decreased so little that we ignored it.

Nevertheless, aggregation of factors may give wrong conclusions. For example, if there are only two important factors and they have effects with the same sizes but opposite signs, then in stage #0 they will cancel and we shall terminate the search, concluding that there are no important factors at all.

We further point out that as the search continues, the groups get smaller and hence the *probability* of having important factors within the same group decreases (unless we have tried to place important factors within the same group, in order to increase the efficiency of our technique). Of course, this low probability is no guarantee.

If practitioners feel that they know the directions of all but a few factor effects, then the 'unknown' factors may be treated separately, while the 'known' factors are investigated by our method. In the ecological case study we examined the very small group of factors with unknown directions through the traditional one-at-a-time method; however, for most factors the system analysts felt confident when they had to specify the signs of the factor effects.

We give one more example. Consider a logistics system that is modeled as a network with nodes representing individual machines; through that network travel individual jobs (orders, customers, transactions). Suppose the response of the system is the throughput per year. Then we may assume that higher speeds of nodes have non-negative effects on the response; and so does an increasing arrival rate of jobs at the source of the network. Our screening technique can identify the 'bottleneck' nodes, that is, the nodes that have speeds with important effects on the throughput of the total system.

We conclude that the assumption of known signs may be more realistic than it may seem at first sight.

5. EFFICIENCY OF SEQUENTIAL BIFURCATION

Above we investigated the effectiveness of our technique: can sequential bifurcation detect important factors? Now we examine its efficiency. We measure efficiency by the number of simulation runs (observations), not by the precision of the main effect estimators: assumptions 1 and 1* state that the approximation errors are negligible; hence (2) and (4) have no error terms. Consequently the estimated effects (see (11) and (12)) are 100% precise. We emphasize that the alternative group screening designs also assume negligible error terms!

If N is not a power of two, then at first sight it seems obvious to split the factors (nearly) equally. For example, a group of six factors would be split into two groups of three, the latter into groups of two and one. This can be described as adding *dummy factors* at positions four and eight. However, as Bettonvil (1990, pp.40-43) demonstrates, it is more efficient to split six factors into two groups of sizes four and two. This can be described as adding dummies at the end: clustering improves efficiency. In general, if the size of a group is not equal to a power of two, and it must split, then the first subgroup has a size equal to the largest possible power of two. In the ecological case study we split the 281 factors into two groups, one group of 256 factors, and one group of 25 factors. 25 Factors are split into groups of sizes 16 and 9.

The clustering argument implies that our technique is most efficient if the original factors are labelled from 1 through N in *increasing* order of importance. After stage #0 the important factors are then clustered in the second group, not in the first group of factors. So our technique can take advantage of a priori knowledge about factor effects.

Next we shall compare our technique with some other group screening techniques, already mentioned in section 1; all these techniques use the assumptions 1 and 2, explicitly or implicitly.

First we examine the *maximum* number of observations needed to find the k important factors among the $N = 2^m$ factors. This number, say n, represents the worst case, caused by lack of clustering.

In *sequential bifurcation* n is given by the following equation if the number of important factors is a power of two; otherwise this equation gives an approximation:

$$n = 1 + k \{\log_2(2N/k)\}; \quad (13)$$

see Bettonvil (1990). For example, in the Jacoby and Harrison case (Figure 1), N is 128 and k is 3, so the approximation gives n = 20.2. Because the important factors show some clustering, the actual number of observations is smaller, namely 16.

Two-stage group screening, introduced by Watson (1961), has already been evaluated by Mauro (1984).

Note that Mauro and Burns (1984) show that random designs are less efficient than two-stage group screening designs are.

Both Patel (1962) and Li (1962) generalize two-stage group screening to *multi-stage group screening*. They assume that each factor has an a priori probability (say) p of being important, which yields an optimal group size and an optimal number of stages.

Morris (1987) devises *multiple grouping*, which is a variant of two-stage group screening. Bettonvil (1990) derives the formula for n (because it is a complicated formula that requires many details on Morris's technique, we do not display that formula).

Jacoby and Harrison (1962)'s sequential bifurcation uses two observations where we use a single one: they use a less efficient design at each stage.

Note that *search linear* models were introduced by Srivastava (1975), assuming that the number of important factors is small but *known.* Srivastava gives specific designs only for k equal to one or two. Ghosh (1979) generalizes this approach to multi-stage designs in such a way that his technique resembles ours. His number of observations equals ours. Because his technique uses more restrictive assumptions than we do, it

will not be further considered in this paper.

The maximum number of observations required by the various group screening techniques is shown in Table 3 for the case N = 1024. We add some comments:

1. The number of observations in Two-stage group screening and in Morris' s group screening depend on *a priori* guesses of the number of important factors; we report the number of observations for optimal guesses.

2. For multi-stage group screening we use the formula for the expected number of observations, which is smaller than the worst case number.

Table 3 shows that, even though the data in that table favor our competitors, our technique has better worst case performance. The comparisons in Table 3 (and Table 4) also hold for other values of N, as Bettonvil (1990, pp.25-32) proves for general N and k.

Note that N = 1024 may be a realistic number; that is, not too high. For example, in the ecological case study we have N = 281, but we restricted our study to a subset of modules of the whole simulation model. Morris (1991) also states: 'Often, computational models have as many as several hundred input variables'.

INSERT TABLE 3

Besides worst case results we also derive the *expected* number of observations. Using the a priori probability p and assuming $N = 2^m$, Bettonvil (1990) derives that the expected number of observations in Sequential Bifurcation is

$$1 + N - \sum_{j=1}^{m} 2^{m-j}(1-p)^{2^j} \quad . \tag{15}$$

Morris (1987) gives the expected number of observations for his technique and for two-stage group screening. Table 4 shows that our technique is more efficient under this criterion too.

INSERT TABLE 4

Note that statistical efficiency should be distinguished from 'handling' efficiency: sequential procedures require that the analysts switch back and forth between the simulation program and the statistical analysis and design program. Two-stage procedures are more efficient in that respect. The development of parallel computer architectures may further complicate the implementation of sequential procedures; also see Kleijnen and Annink (1992).

6. CONCLUSIONS

Simulation models may have many factors, whereas only a few factors are really important. To detect these important factors we derive the sequential bifurcation technique. This group screening uses a design and analysis that requires fewer simulation observations than other group screening techniques do.

Our experimental design depends on the simulation model's approximation or metamodel. If that approximation assumes first-order effects only, then the number of observation is half the number used in approximations that assume two-factor interactions. In general, however, we recommend a more cautious approach, that is, we recommend the metamodel with interactions.

Our technique can be applied without a priori quantifying the importance of main effects: as simulation outputs are observed, upper limits for the main effects become available, and the users can stop as soon as these limits are judged to be sharp enough, depending on the system being simulated.

Sequential bifurcation has already been applied to a complicated ecological simulation model, giving surprising results that are confirmed by a classical resolution IV experiment.

ACKNOWLEDGEMENTS

We thank three anonymous referees, the Associate Editor, and the Editor for their constructive comments on the first two versions of this paper. We further thank Dr. Peter Sander (Professor in the School of Industrial Engineering and Management, Eindhoven University of Technology), who -together with Jack Kleijnen-

supervised the doctoral work of Bert Bettonvil. Bettonvil was financially supported jointly by the Eindhoven and Tilburg universities (Samenwerkingsorgaan Brabantse Universiteiten, SOBU).

BIBLIOGRAPHY

ANDRES, T.H., and W.C. HAJAS (1993), "Using iterated fractional factorial design to screen parameters in sensitivity analysis of a probabilistic risk assessment model", presented at the *Joint International Conference on Mathematical Methods and Supercomputing in Nuclear Applications*.

BETTONVIL, B. (1990), *Detection of Important Factors by Sequential Bifurcation*, Tilburg University Press, Tilburg.

BETTONVIL, B. and J.P.C. Kleijnen (1990), Measurement scales and resolution IV designs. *American Journal of Mathematical and Management Sciences* 10, 309-322.

BOX, G.E.P. and K.B. WILSON (1951), "On the experimental attainment of optimum conditions", *Journal of the Royal Statistical Society, Series B*, 13, 1-38.

BRASSARD, G. and P. BRADLEY (1988), *Algorithmics: Theory & Practice*, Prentice Hall, Englewood Cliffs, NJ.

GHOSH, S. (1979), "On single and multistage factor screening procedures", *Journal of Combinatorics, Information & System Sciences* 4, 275-284.

Ho, Y. and X. Cao (1991), *Perturbation Analysis of Discrete Event Systems*. Kluwer, Dordrecht.

JACOBY, J.E. and S. HARRISON (1962), "Multi-variable experimentation and simulation models", *Naval Research Logistic Quarterly* 9, 121-136.

KLEIJNEN, J.P.C. (1975), "Screening designs for poly-factor experimentation", *Technometrics* 17(4), 487-493.

KLEIJNEN, J.P.C. (1987), *Statistical tools for simulation practitioners*, Marcel Dekker, Inc., New York.

KLEIJNEN, J.P.C. (1994), "Sensitivity analysis versus uncertainty analysis: when to use what?' *Predictability and nonlinear modeling in natural sciences and economics*, Kluwer, Dordrecht, The Netherlands, 1994.

KLEIJNEN, J.P.C. and B. ANNINK (1992), "Vector computers, Monte Carlo simulation, and regression analysis: an introduction". *Management Science*, 38, 170-181.

LI, C.H. (1962), "A sequential method for screening experimental variables", *American Statistical Association Journal* 57, 455-477.

MAURO, C.A. (1984), "On the performance of two-stage group screening experiments", *Technometrics* 26(3), 255-264.

MAURO, C.A. and K.C. BURNS (1984), "A comparison of random balance and two-stage group screening designs: a case study", *Communications in Statistics: Theory and Methods* 13, 2625-2647.

MORRIS, M.D. (1987), "Two-stage factor screening procedures using multiple grouping assignments", *Communications in Statistics: Theory and Methods* 16, 3051-3067.

MORRIS, M.D. (1991), "Factorial plans for preliminary computational experiments", *Technometrics* 33(2), 161-174.

PATEL, M.S. (1962), "Group screening with more than two stages", *Technometrics* 4(2), 209-217.

ROTMANS, J. (1990), *IMAGE: An Integrated model to Assess the Greenhouse Effect*, Kluwer Academic Publishers, Dordrecht/Boston/London.

RUBINSTEIN, R.Y. and A. SHAPIRO (1993), *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization via the Score Function Method*, John Wiley & Sons. New York.

SACKS, J., W.J. WELCH, T.J. MITCHELL and H.P. WYNN (1989), "Design and analysis of computer experiments" (including Comments), *Statistical Science* 4(4), 409-435.

SRIVASTAVA, J.N. (1975), "Designs for searching nonnegligible effects", in J.N. Srivastava (ed.) *A Survey of Statistical Design and Linear Models*, North-Holland Publishing Company, Amsterdam, 507-519.

VAN GROENENDAAL, W. (1994), *Investment analysis and decision support for gas transmission on Java*, Tilburg University Press, Tilburg.

WELCH, J.W., R.J. BUCK, J. SACKS, H.P. WYNN, T.J. MITCHELL, and M.D. MORRIS (1992), "Screening, Predicting, and Computer Experiments", *Technometrics* 34(1), 15-25.

WATSON, G.S. (1961), "A study of the group screening method", *Technometrics* 3(3), 371-388.

ZIEGLER, B. (1976), *Theory of modelling and simulation*, Wiley Interscience, New York.

TABLE 1

Important Factors in IMAGE according to Sequential Bifurcation

label j	effect γ_j	range L^*_j	range H^*_j	meaning
250	12.7475	0.0	0.6	shift from temperate forest to agricultural land
246	8.3725	0.9	1.4	shift from temperate forest to grassland
19	7.35	19.0	21.0	residence time in the (thick) cold mixed layer
237	7.0925	0.0	0.3	shift from temperate forest to open tropical forest
243	6.8	0.5	0.0	shift from human area to temperate forest
242	5.6	0.4	0.0	shift from agricultural land to temperate forest
241	5.26	0.5	0.0	shift from grassland to temperate forest
240	5.2075	0.3	0.0	shift from open tropical forest to temperate forest
281	4.8305	0.41	0.38	biotic stimulation factor
13	4.46	0.81	0.79	rate of precipitation of carbon in the oceans
86	3.5525	0.11	0.095	fraction of charcoal formed upon burning of branches
239	3.4175	0.2	0.0	shift from closed tropical forest to temperate forest
22	3.4125	1282.23	1482.23	initial area of ecosystem 1 (tropical closed forest)
20	3.0475	2.27	2.37	circulating massflow (Gordon flow)
244	2.8625	0.2	0.0	shift from semi-desert to temperate forest

TABLE 2

Important Factors in IMAGE according to Resolution IV Design

(1) # 250	13.07	(0)	1224.33
(2) # 246	8.54	(1,2)	-0.05
(3) # 19	7.31	(1,3)	0.10
(4) # 281	5.16	(1,4)	0.09
(5) # 13	4.42	(1,5)	0.08
(6) # 86	3.54	(1,6)	0.07
(7) # 20	3.10	(1,7)	0.07
(8) DIFF	1.98	(1,8)	0.06

TABLE 3

Maximum number of observations if N = 1024

Technique	Number of important factors (k)								
	0	1	2	3	4	5	6	7	8
Two-stage group screening	2	68	96	116	136	152	168	180	192
Multi-stage group screening	2	20	35	49	62	74	85	96	107
Morris group screening	2	12	33	53	73	80	89	100	113
Jacoby & Harrison	3	21	39	55	71	85	99	113	127
Sequential Bifurcation	2	12	21	29	37	44	51	58	65

TABLE 4

Expected number of observations for N = 1024

Technique	A priori probability p			
	.0001	.001	.01	.1
Two-stage group screening	21.9	64.3	198.2	521.1
Morris " "	10.3	26.4	130.3	521.1
Sequential Bifurcation	3.0	11.4	70.5	374.2

FIGURE I

Finding 3 important variables amongst 128 candidates.

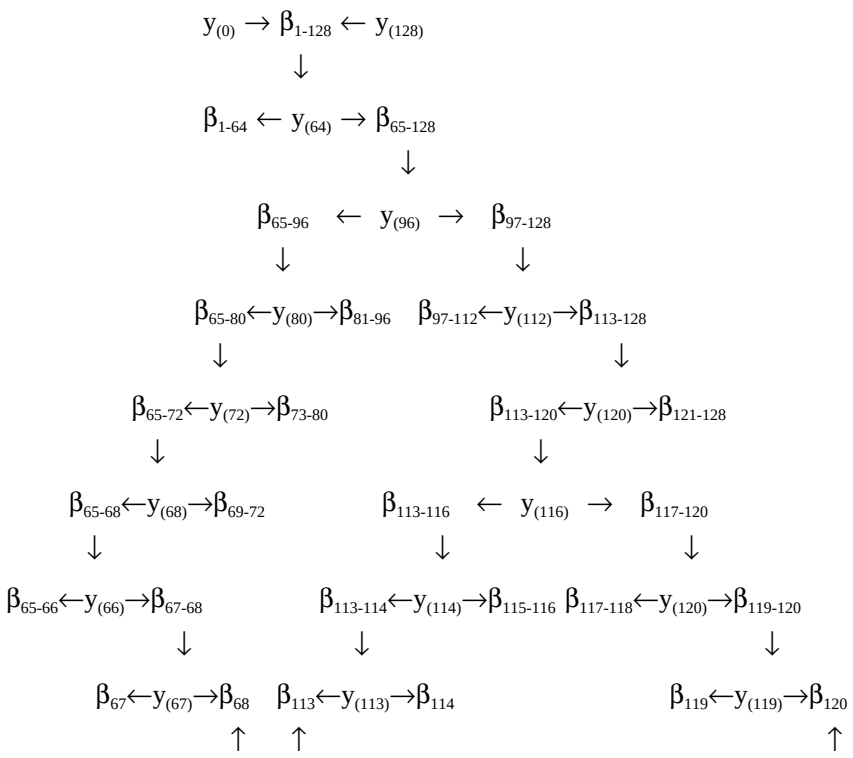

FIGURE 2. Main effects and upper limits in various stages.