CentER

Discussion Paper

No. 2003–104

FINDING THE IMPORTANT FACTORS IN LARGE DISCRETE-EVENT SIMULATION: SEQUENTIAL BIFURCATION AND ITS APPLICATIONS

By J.P.C. Kleijnen, B.W.M. Bettonvil, F. Persson

November 2003

ISSN 0924-7815

TILBURG UNIVERSITY

File name: Dean\SB_Springer_revision_1.doc; printed: 07-11-03; written: 6 November 2003

Finding the important factors in large discrete-event simulation: sequential bifurcation and its applications

Jack P.C. Kleijnen [1*], Bert Bettonvil [1], and Fredrik Persson [2]

[1] Department of Information Systems and Management/Center for Economic Research (CentER), Tilburg University (UvT), Postbox 90153, 5000 LE Tilburg, The Netherlands

[2] Department of Production Economics, Linköpings Universitet,S-581 83 Linköping, Sweden

- Corresponding author: phone: +31-13-4662029; fax: +31-13-4663377; e-mail: kleijnen@UvT.nl
http://center.kub.nl/staff/kleijnen/

Screening, edited by Angela M. Dean and Susan M. Lewis, Springer-Verlag, New York, 2003

Acknowledgement: this research is sponsored by 'Chain networks, Clusters & ICT' ('KLICT/ICES'). We gratefully acknowledge the many useful comments by the editors and an anonymous referee

Table of Contents

Abstract: This contribution discusses experiments with many factors: the case study includes a simulation model with 92 factors. The experiments are guided by sequential bifurcation. This method is most efficient and effective if the true input/output behavior of the simulation model can be approximated through a first-order polynomial possibly augmented with two-factor interactions. The method is explained and illustrated through three related discrete-event simulation models. These models represent three supply chain configurations, studied for an Ericsson factory in Sweden. After simulating 21 scenarios (factor combinations) – each replicated five times to account for noise – a shortlist with the 11 most important factors is identified for the biggest of the three simulation models.

Finding the important factors in large discrete-event simulation models: sequential bifurcation and its applications

Jack P.C. Kleijnen and Bert Bettonvil

Department of Information Systems and Management/Center for Economic Research (CentER), Tilburg University (UvT),
Postbox 90153, 5000 LE Tilburg, The Netherlands

Fredrik Persson

Department of Production Economics, Linköping Institute of Technology,
S-581 83 Linköping, Sweden

1 Introduction

In this introductory section we discuss our definition of screening; our view on simulation versus real-world experiments, screening procedures, including our procedure called sequential bifurcation, and our case study

1.1 A definition of screening

We define screening as the search for the 'most important' factors among a 'large' set of factors in an experiment. For example, in our case study (see Section 1.4) the number of factors is 92 - but we find only 11 factors to be important: so-called 'effect sparsity'. The simplest definition of importance occurs when the factors have additive effects only and when the experiment has a single response (output); i.e., the input-output relation is a first-order polynomial in regression terminology or a 'main effects only' model in analysis of variance (ANOVA) terminology (also see Schonlau and Welch 2003). In such a case, the most important factor is the one with the largest absolute value of its first-order effect or main effect; the least important factor is the one with the effect closest to zero.

The goal of screening is to come up with a *shortlist of important factors* from a long list of potential factors. Depending on the application, this shortlist might lead to a more thorough investigation via additional experiments; see Kleijnen et

al. (2002). In the case study discussed in Section 1.4, we first performed screening; next we did an optimization and robustness analysis for the most important factors; see Kleijnen et al. (2003). In an ecological case study, Bettonvil and Kleijnen (1997) identified a shortlist of factors, including some factors that the ecological experts had not expected to have important effects!

It is also important to know which factors are 'certainly' unimportant so the clients of the simulation analysts are not bothered by details about these factors. Of course, the importance of factors depends on the *experimental domain* (experimental area to be explored; also called 'experimental frame' by Zeigler et al. 2000). The clients must supply information on this domain, including realistic ranges of the individual factors and limits on the *admissible scenarios* or combinations of factor levels; for example, some factor values must add up to 100%. Note that we distinguish between on one hand the simulation analysts, who develop the simulation model and run experiments with this model, and on the other hand their clients, who may be the managers and other users of the real system being simulated.

We view the real or the simulated system as a *black box*: it transforms inputs into outputs. Experiments with such a system are often analyzed through a regression or an ANOVA model. We call such a model a *metamodel*; other names are: auxiliary model, emulator, response surface, etc. Other metamodel types are: Kriging, neural nets (NN), radial functions, splines, etc. Of course, the simulation is itself a model of some real-world system. Hopefully, a parsimonious metamodel can be built that describes the input-output relationship in simple terms (much simpler than the full simulation model). We emphasize the following chicken-and-egg problem: once the design is specified and simulated, metamodel parameters can be estimated; however, the types of metamodels that the analyst desires to investigate should guide the selection of an appropriate design.

1.2 Simulation versus real-world experiments

Classic design of experiments focuses on real-world experiments; see the classic textbook by Box, Hunter, and Hunter (1978) or the recent textbook by Myers and Montgomery (2002). We, however, focus on experiments with computer or simulation models, which may be either deterministic or stochastic; also see Kleijnen et al. (2002). For an introduction to simulation modeling we refer to the most popular textbook (80,000 copies sold), namely Law and Kelton (2000).

In simulation—with its advances in computing power—the analysts are no longer bound by some of the constraints that characterize real-world experiments. This is a challenge, as it requires a new mindset: we argue that the way simulation experiments should be approached is fundamentally different from the way in which real-world experiments should be approached. We now discuss three aspects (labeled a, b, c).

a. In real-world experiments, the analysts must often select a design that is executed in *one shot* (say, one growing season in an agricultural experiment). In simulation, however, the data are collected *sequentially* because a standard computer operates sequentially (a parallel computer is still an exception). Consequently, the analysts may examine the most recent observation before selecting the next design point. For example, the analysts may start with a relatively small design for a very simple metamodel; then test (validate) the adequacy of that model; and only if they reject that model, they augment the original design to enable the estimation of a more complicated model—this is a two-stage design. In this paper, however, we present a design that analyzes each new observation before selecting the next design point—except for the first two observations (which correspond with the two extreme scenarios). Also see Kleijnen et al. (2002) and Kleijnen and Sargent (2000).

When analysts must take samples sequentially in real-world experiments, then the experiment is viewed as prone to validity problems. Hence, the analysts randomize the order of sampling to guard against time-related changes in the experimental environment (such as temperature, humidity, consumer confidence, and learning effects), and perform appropriate statistical tests to determine whether or not the results have been contaminated.

However, most simulation experiments are implemented *sequentially*.—without requiring explicit randomization. An input file can be generated, once a particular design type has been chosen—as we did indeed in our case study. Such a file can be executed sequentially (and efficiently) in *batch mode*; that is, no human intervention is required while the computer executes the sequential design (including rules for selecting the next design point, based on all preceding observations).

b. In real-world experiments, only a small number of factors are typically varied: it is a challenge to control more than, say, ten factors. Actually, many published experiments deal with fewer than five factors (however, several chapters of this book do deal with many more than ten factors in real-world experiments). We emphasize that - in simulation - good computer programming avoids fixing the factors at specific numerical values within the code; instead the computer reads factor values so that the program can be run for many combinations of values. (Of course, the computer should check whether these values are admissible; that is, do these combinations fall within the experimental domain?) *Such a practice can automatically provide a long list of potential factors*. Next, the users should confirm whether they indeed wish to experiment with all these factors or whether they wish to *a priori* fix some factors at nominal (or base) levels. This type of coding helps *unfreeze* the mindset of users who would otherwise be inclined to focus on only a few factors. (For example, Persson and Olhager 2002 simulate only nine combinations of factor levels.)

However, a large set of factors implies that computer time may again become a bottleneck: if a single simulation observation takes (say) one hour, then traditional design of experiments fails and we must use special screening procedures. (In our case study one run originally took three hours; after modification of our simulation code, this became 40 minutes.)

c. *Randomness* occurs is real-world experiments because the experimenters cannot control all factors that affect the experiment's response; for example, human participants have ambition levels that cannot be fully controlled by the experimenters. In computer simulation, these effects can be modeled through random input variables; for example, the arrivals of customers may be modeled as a Poisson process—so the time between two successive arrivals is exponentially distributed. Values for random variables are generated through pseudorandom numbers. (There are also deterministic simulation models, but we focus on random simulation models.) A single simulation run gives an observation on each of the responses of interest to the analysts; for example, in our supply chain simulation (detailed in Section 3), one simulation run represents the operations of that chain during seventeen weeks; the major response is the average costs per week. To obtain independently and identically distributed (i.i.d.) observations, the analysts generate several simulation runs or *replicates*, which all start in the same initial state of the simulated system and use different pseudorandom numbers.

1.3 Screening procedures for simulation

The contributors of this book present various screening methods; two chapters closely related to this chapter are Morris (2003) and Schonlau and Welch (2003). Further, Campolongo, Kleijnen, and Andres (2000) discuss screening methods in a simulation (not real-world) context: one-factor-at-a-time (OAT), Morris's OAT, Cotter's design, Andres's Iterated Fractional Factorial design (IFFD), and sequential bifurcation. They refer to available software, and emphasize key assumptions. Some of these methods – including sequential bifurcation - require fewer observations than there are factors: supersaturated designs. However, the term 'supersaturated' is usually reserved for single-shot designs; see above (Section 1.2) and also Westfall, Young, and Lin (1998).

In this chapter, we explain the procedure of sequential bifurcation. We add some new results for sequential bifurcation in random (not deterministic) simulations, and apply the resulting method to a simulation model developed for Ericsson in Sweden.

1.4 The Ericsson case study: supply chain simulation

By definition, a *supply chain* consists of several 'links', which are separate companies or independent business units of a single large company. Example *links* are retailers, wholesalers, distributors, and factories. Common sense strategies may amplify fluctuations in the demand by final customers—up the supply chain; Lee et al. (1997) identify this amplification as one of the *bullwhip* effects. Banks et al.

(2002) and Kleijnen (2004) discuss simulation in supply chain management. Supply chain management is a hot topic in Operations Research (OR), including simulation; see for example the *Proceedings of the 2002 Winter Simulation Conference* and Simchi-Levi, Kaminsky, and Simchi-Levi (2000).

In this section, we discuss a recent example of simulation that models three alternative configurations for a supply chain in the mobile communications industry in Sweden—centered on the Ericsson company. A central issue in supply chain management is the *lean and mean* strategy; that is, the elimination of links or steps within links hopefully improves the total chain's performance. In the Ericsson case, three supply chain configurations are studied. Each configuration is actually the same chain—'simplified' over time. The Old configuration—existing in 1998—consists of many operations and test stations. The Current (1999) configuration has fewer operational steps. The Next Generation chain is a future scenario with a minimum of operations and tests. Figure 1 shows simplified pictures of the three configurations. We now discuss this figure in some detail.

In Figure 1, a square denotes a test of the products produced by the preceding operation, which is denoted by a circle. There are several types of tests and operations, as the names in the figure show. Products flow through the chain, from left to right in the figure. The chain starts by buying 'raw' products; next, it processes these products; the chain finishes by assembling components into 'final' products (which are sold). The abbreviation SMD stands for 'surface mounted devices' (these electronic devices are mounted on the surface on a circuit board—which is a modern technology compared with devices that are mounted in holes on the board). In Figure 1, the dotted boxes box indicate that the mounting and testing of the circuit board is integrated in the same machine (in this case a very fast mounting machine with integrated vision control of device placements; into this machine is also integrated the heating to make the soldering paste melt and connect the board to the electronic device).

Figure 2 shows the Current simulation model. This figure illustrates that buffers (inventories) are located before and after each test station and operation; products are transported between all operations and test-stations.

The goal of the simulation study is to quantify the relationships among the simulation output—namely, the steady-state mean costs of the whole supply chain—and the simulation inputs (or factors)—such as lead-time and quality. Our ultimate goal (reported in Kleijnen et al. 2003) is to find *robust* solutions for the supply chain problem, so we distinguish two types of factors:

1. Factors that are controllable by the company; for example, Ericsson can manipulate the manufacturing processes and select logistic partners for transportation.
2. Factors that are determined by the environment; examples are demand for products, process yield or percentage of faulty products, and scrap percentage at each test station.

The simulation model of the Old supply chain has 92 factors (for example, operation time of an individual process, number of resources). The model of the Current and the Next Generation supply chain have 78 and 49 factors respectively.

More details follow in Section 3.

Abstract: This contribution discusses experiments with many factors: the case study includes a simulation model with 92 factors. The experiments are guided by sequential bifurcation. This method is most efficient and effective if the true input/output beha 9

2. Sequential bifurcation (sequential bifurcation)

Originally, sequential bifurcation was developed in the doctoral dissertation by Bettonvil (1990). This dissertation is summarized by Bettonvil and Kleijnen (1997), and updated by Campolongo et al. (2000), including a few applications. Other authors have also studied sequential bifurcation: Cheng (1997), Cheng and Holland (1999). sequential bifurcation is related to binary search, which aims to find the single most important factor. First we present the assumptions and notation of sequential bifurcation (Section 2.1); next we illustrate the sequential bifurcation steps (Section 2.2).

2.1 Assumptions and notation of sequential bifurcation

The two basic assumptions of sequential bifurcation are:

Assumption 1: a first-order polynomial possibly augmented with two-factor interactions is an adequate metamodel:

$$Y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \beta_{1:2} x_1 x_2 + ... + \beta_{(k-1);k} x_{k-1} x_k + E \qquad (2.1)$$

where we use capitals for random variables, Greek letters for parameters to be estimated, and the following symbols:

Y: the response of the metamodel (the sequential bifurcation approximation)

k: the total number of factors in the experiment

β_j : the first-order or main effect of factor j with $j = 1, \ldots, k$

$\beta_{j';j}$: the interaction effect of the factors j' and j with $1 \le j' < j \le k$

x_j : value of factor j

E: the noise (randomness caused by both pseudorandom numbers and approximation error)

We observe that this metamodel is linear in its parameters β_j and $\beta_{j';j}$ but nonlinear in its variables x_j . At the end of the case study, we shall try to validate the assumption; see Section 3.

To estimate the parameters β in the simple metamodel specified in (2.1), it is most efficient to experiment with *only two levels* (values) per factor (see any textbook on design of experiments). In practice, it is important that these levels are realistic.

Assumption 2: the signs of all main effects are known so

$$\beta_j \ge 0 \ (j = 1, \ldots, k) \qquad (2.2)$$

We comment that we need this assumption; otherwise, main effects might cancel each other (also see Morris 2003). Our experience is that in practice this as-

sumption is easy to satisfy; i.e., it is easy to define the upper and lower level of each factor such that the upper level does not decrease the expected response. For example, in our case study some factors refer to transportation speed: the higher this speed, the lower the work in process (WIP); hence the lower the costs. (Other examples are provided by Lewis and Dean 2001.) If in a particular case study this assumption seems hard to meet for specific factors, then these factors should be treated 'individually'; i.e., none of these factors should be grouped with other factors.

Consequently, we call the two factor levels (mentioned above) the *low level* and the *high level* respectively. This low level is the factor level that results in the lower expected (simulation) response. The simplest experimental domain is a k – dimensional hypercube - after coding the original factor z_j such that its low value l_j corresponds with the value -1 of the standardized factor x_j; likewise, its high value h_j corresponds with the value +1 of the standardized factor x_j :

$$x_j = \frac{z_j - (h_j + l_j)/2}{(h_j - l_j)/2} \tag{2.3}$$

The scaling in (2.3) implies that we may rank (sort) the factors by their main effects; that is, the most important factor is the one with the highest main effect, etc. We notice that the larger the range of an untransformed factor is, the larger the response difference and hence the main effect of the transformed factor is. (Also see Cheng and Holland 1999 's 'unit cost' effects.)

In our case study, we could not get information on the factor ranges from Ericsson. Therefore we decided to change most factors by 5 % of the base values reported for the existing system by Persson and Olhager (2002). However, transportation speeds between operations we change by 25 %. (Also see again Cheng and Holland, 1999.)

The *efficiency* of sequential bifurcation—measured by the number of observations (simulation runs and hence simulation time)—increases if the individual factors are listed (numbered) such that factors are placed in increasing order of importance (see Bettonvil 1990, pp. 44)

$$\beta_{j'} \leq \beta_j \quad (j' < j) \tag{2.4}$$

We try to realize this efficiency gain by applying prior knowledge about the simulated real system. Because we think that environmental factors are most important, we place these factors last in the list of factors.

We try to increase the efficiency further, and use our knowledge about the simulated real system to keep *similar* factors together; for example, we group all test yield factors together. We conjecture that if one yield factor is unimportant, then all yield factors are likely to be unimportant too. Bettonvil (1990, pp. 40-43) divides factor groups such that the number of factors per resulting subgroup is a power of two. We use his approach as a secondary guideline; for example, in our sequential bifurcation for the Old supply chain we split the first 48 factors into a group of 32 (= 2^5) factors and a group of the remaining factors; see Figure 3,

which shows the results of applying sequential bifurcation to our Ericsson case study; its factors are labeled from 1 through 92, and the continuous horizontal band at the top of the figure indicates that all factors have been grouped together in the first step. (We do not follow Bettonvil's approach if it implies splitting up a group of related factors. Cheng (1997) splits groups into two groups of equal size.)

To explain the sequential bifurcation procedure, we initially assume that a first-order polynomial is an adequate metamodel; i.e., the interactions in (2.1) are zero and the expected value of the noise E is zero:

$$\beta_{j':j} = 0 \quad (j' \neq j); \mu_E = 0. \tag{2.5}$$

Based on Bettonvil (1990) we introduce the following additional sequential bifurcation notation adapted for replicated random responses.

$y_{(j);r}$: observed (simulation) output with the factors 1 through j set to their high levels and the remaining factors set to their low levels, in replication r of factor combination i with $r = 1, \ldots, m_i$; to simplify the notation we assume a constant number of replications for all factor combinations so $m_i = m$

We now define $\beta_{j'\_j}$ as the sum of main effects for factor j' through j:

$$\beta_{j'\_j} = \sum_{h=j'}^{j} \beta_h \tag{2.6}$$

An estimate of this aggregated main effect $\beta_{j'\_j}$ - based on replication r - is

$$\hat{\beta}_{j'\_j;r} = \frac{y_{(j');r} - y_{(j-1);r}}{2} \tag{2.7}$$

For example, we start with observing (simulating) two extreme scenarios (see below: Section 2.2): in scenario 1 all factors are at their low level; in scenario 2 all factors are high. Using the metamodel in (2.1), we obtain

$$E(Y_{(0)}) = \beta_0 - \beta_1 - \ldots - \beta_k \tag{2.8}$$

$$E(Y_{(k)}) = \beta_0 + \beta_1 + \ldots + \beta_k \tag{2.9}$$

so

$$E(Y_{(k)}) - E(Y_{(0)}) = 2(\beta_1 + \ldots + \beta_k) \tag{2.10}$$

which explains that the estimator following from (2.7) is unbiased.

For the individual main effect the estimator is

$$\hat{\beta}_{j;r} = \frac{y_{(j);r} - y_{(j-1);r}}{2} \tag{2.11}$$

From the m replicates we compute the average and the sample variance for each (aggregated or individual) estimated effect; for example, for the individual main effect j we obtain

$$\bar{\hat{\beta}}_j = \frac{\sum_{r=1}^{m} \hat{\beta}_{j;r}}{m}; \qquad s^2(\bar{\hat{\beta}}_j) = \frac{\sum_{r=1}^{m} (\hat{\beta}_{j;r} - \bar{\hat{\beta}})^2}{m(m-1)} \tag{2.12}$$

where the factor m in $s^2(\bar{\hat{\beta}}_j)$ is needed because it estimates the variance of $\bar{\hat{\beta}}_j$, which is an *average* effect computed from m replicates.

The variance estimators of the estimated effects in (2.12) allow unequal response variances and the use of *common* pseudorandom numbers, which is a well-known variance reduction technique in simulation. This technique uses the same pseudorandom numbers to simulate various factor combinations. When this technique creates positive correlations between the responses of these various factor combinations, then the variances of the estimated main effects are obviously reduced. (This technique is similar to blocking in real-world experiments; also see Morris 2003.)

2.2 The steps of sequential bifurcation

A formal description of sequential bifurcation can be found in Bettonvil (1990). Here we illustrate sequential bifurcation through the Old simulation model, which has 92 factors. As Table 1 shows, we *start* sequential bifurcation finding $\bar{y}_{(0)}$ = 3,981,627 and $\bar{y}_{(92)}$ = 34,013,832 where the overline denotes the average computed from the m replicates; this table also displays the individual five replicates. So, the estimated effect of all 92 factors together is $\bar{\hat{\beta}}_{1\_92}$ = (34,013,832 – 3,983,627)/2 = 15,016,102; also see Figure 3 immediately below the first shaded line listing all factor labels from 1 though 92. The standard error of this estimated aggregated effects is $s(\bar{\hat{\beta}}_{1\_92})$ = 128,508.

To *test* the importance of estimated (either aggregated or individual) main effects statistically, we assume that the (simulated) outputs per scenario are normally and independently distributed. Different scenarios may give different variances. Sequential bifurcation allows the analysts to use common pseudorandom numbers for different scenarios. We apply Student's statistic, which is known not to be very sensitive to violation of the normality assumption; we ignore variance heterogeneity when determining the degrees of freedom (see Kleijnen 1987, pp. 14-23). We apply a one-sided test because we assume that all individual effects

are non-negative. We conclude that the sum of 92 main effects is significantly different from zero. Our heuristic uses a fixed *t*–value throughout the whole sequential bifurcation procedure; we do not adjust for multiple testing (see Kleijnen 1987, pp. 41-45). In practice, significance is not essential; importance is: we search for a shortlist of important factors.

In hindsight, we might have used fewer replications in the early stages, as these stages have higher signal/noise. (The noise remains constant - if for simplicity's sake we assume that all outputs have the same variance - whereas the signal decreases during the sequential bifurcation procedure as this signal is the sum of fewer main effects; i.e., as sequential bifurcation proceeds, less aggregation occurs.)

Some reflection proves that this group effect is also an *upper limit* for any individual main effect. The goal of sequential bifurcation is to find the most 'important' factors; that is, the factors that have significant main effects. If, however, we terminate our screening prematurely (for example, because the computer breaks down or our clients get impatient), then sequential bifurcation still estimates the factors with the largest main effects – as we shall see next.

The *next step* is to divide the current group (of 92 factors) into *two* subgroups; this explains the term *sequential bifurcation*. Into one subgroup we place all the 79 controllable factors; so into the other subgroup we put all 13 environmental factors; see the next shaded line in Figure 3. Simulation of this scenario gives the third observation $\bar{y}_{(79)}$ = 9,250,034 (with standard error 14,127). This $\bar{y}_{(79)}$ lies between $\bar{y}_{(0)}$ and $\bar{y}_{(92)}$, as the sequential bifurcation assumptions imply. Comparison of $\bar{y}_{(79)}$ and $\bar{y}_{(0)}$ gives $\hat{\bar{\beta}}_{1-79}$ = 2,634,203 (s = 16,534). Comparison of $\bar{y}_{(92)}$ and $\bar{y}_{(79)}$ gives $\hat{\bar{\beta}}_{80-92}$ = 12,381,899 (s = 128,220). So, this step splits the total effect of the first step ($\hat{\bar{\beta}}_{1-92}$ = 15,016,102) into its two additive components. This step decreases the upper limit for any individual effect in the first subgroup to 2,634,203; for any individual effect in the second subgroup this limit is 12,381,899.

To decide *where* to split a group into its two subgroups, we use several principles—meant to increase the efficiency of our procedure, as explained between (2.4) and (2.5).

Figure 3 also shows our remaining sequential bifurcation steps. We do not split a group any further, when its estimated (aggregated) effect is either non-significant or negative. For example, the estimated effect of factors 50 through 79 is –3,230 with standard error 31,418

Sequential bifurcation stops after 21 steps. The upper limit for any remaining individual factor – denoted by $U(21)$ (also see Figure 5, discussed below) - is then reduced to 12,792. Our list of the 11 most important factors is shown in the bottom-left part of the figure. This list shows that factor 92 is the most important fac-

tor (its estimated main effect is 8,087,149); the smallest factor in this shortlist is factor 88 (its effect is 241,809). (The meaning of these factors in the case study will be presented in Section 3). Remembering that we tried to label the factors such that equation (2.4) is satisfied, we now conclude that indeed factor 92 turns out to be the most important factor; no factor labeled with a number smaller than 42 is significant.

The figure also shows another list of the 11 most important factors – called gamma, not beta. We obtain this list through the more realistic metamodel including interactions, discussed next.

2.3. Two-factor interactions and foldover design in sequential bifurcation

Groups of main effects are estimated without bias by two-factor interactions if sequential bifurcation uses a *foldover* design (also see Morris 2003). By definition, such a design also observes (simulates) the *mirror* scenario of each scenario observed when (as in Section 2.2) assuming no interactions. By definition, a mirror scenario switches all individual factors that were at the high value in the original scenario to the low value. We let $y_{-(j);\,r}$ denote the observed output with the factors 1 through j set to their low levels in replication r and the remaining factors set to their high levels. (This $y_{-(j);\,r}$ is the 'mirror' observation of $y_{(j);\,r}$.) As an example we consider the second shaded line in Figure 4:

$$E(Y_{-(49)}) = \beta_0 - \beta_1 - \ldots - \beta_{49} + \beta_{50} + \ldots + \beta_{92} + \\ + \beta_{1;2} \ldots + \beta_{48;49} - \beta_{1;50} \ldots - \beta_{49;92} + \beta_{50;51} \ldots + \beta_{91;92} \tag{2.13}$$

whereas the original observation gives

$$E(Y_{(49)}) = \beta_0 + \beta_1 + \ldots + \beta_{49} - \beta_{50} - \ldots - \beta_{92} + \\ + \beta_{1;2} \ldots + \beta_{48;49} - \beta_{1;50} \ldots - \beta_{49;92} + \beta_{50;51} \ldots + \beta_{91;92} \tag{2.14}$$

Subtracting these two equations demonstrates that all interactions cancel out.

Bettonvil (1990) shows that the following estimated group effects are unbiased by two-factor interactions:

$$\hat{\beta}_{j'-j;r} = \frac{\left(y_{(j');r} - y_{-(j');r}\right) - \left(y_{(j-1);r} - y_{-(j-1);r}\right)}{4} \tag{2.15}$$

Likewise, the individual main effect estimate is

$$\hat{\beta}_{j;r} = \frac{\left(y_{(j);r} - y_{-(j);r}\right) - \left(y_{(j-1);r} - y_{-(j-1);r}\right)}{4} \tag{2.16}$$

Note that the foldover design doubles the number of scenarios. This principle is also well known in classic design of experiments, and is due to Box and Wilson (1951); see Kleijnen (1987, p. 303).

Sequential bifurcation may give misleading results if (say) two factors have unimportant main effects but an important interaction (also see Lewis and Dean 2001). Therefore we add the following assumption (which Wu and Hamada 2000 call the 'strong heredity' assumption).

Assumption 3: if a factor has no main effect, then this factor does not interact with any other factor:

$$\beta_j = 0 \Rightarrow \beta_{j;\,j'} = 0 \; (\, j \neq j' \,) \tag{2.17}$$

If *a priori* we suspect that this assumption is violated, then we should investigate such a factor after the screening phase.

The foldover design does *not* enable us to estimate *individual* interactions, but it does enable us to estimate whether interactions are important – as follows. We estimate the main effects from the original scenarios, ignoring the mirror scenarios. If the foldover design and the 'original' design give the same conclusions, then interactions are unimportant. Interactions were indeed unimportant in the ecological simulation reported by Bettonvil (1990) and Bettonvil and Kleijnen (1997). In the present case study, however, we shall see that interactions are important. (In a follow-up experiment that includes only the factors declared to be important, we estimate the individual interactions from a so-called resolution-V design; see Kleijnen et al. 2003.)

Analogous to Figure 3, Figure 4 shows the sequential bifurcation steps when we do allow for interactions. Comparing these two figures (or comparing the two lists in the bottom-left part of Figure 3) shows that in this case we find the *same shortlist*. The individual values, however, do differ: *interactions are important*.

Next we apply sequential bifurcation to the other two supply chain configurations (Current and Next generations), and we interpret these sequential bifurcation results through our knowledge of the simulated real system

3 Case study: Ericsson's supply chain simulation models

First we discuss some programming issues. Then we define the inputs and outputs of the three simulation models. Finally, we present results.

3.1 Programming the three simulation models

We give only a short description of the three supply chain configurations and their simulation models; for details we refer to Persson and Olhager (2002). At the start of our sequential bifurcation, we had three simulation models programmed in the Taylor II simulation software for discrete event simulations; see Incontrol (2003). We conduct our sequential bifurcation via Microsoft Excel, using the batch run mode in Taylor II. We store input-output data in Excel worksheets. This set-up fa-

cilitates the analysis of the simulation input-output data, but it constrains the experiment's set-up. For instance, we cannot control the pseudorandom numbers in the batch mode of Taylor II. Hence, we cannot guarantee absence of overlap in the pseudorandom numbers; we conjecture that the probability of overlap is negligible in practice. Neither can we apply common pseudorandom numbers.

To validate the resulting simulation models, Person and Olhager (2002) examined the simulation of the Current supply chain configuration. They could validate this model, since the real-world supply chain existed at that time, and data were available. Next, they developed the other two models, using the Current system's model. (More precisely, they validated the model of the Current configuration through a structured walkthrough with the engineers and managers who are familiar with the system being modeled; also see Balci (1998). Next they developed the other two models from this model. They validated the model of the Old supply chain configuration through a less thorough walkthrough. The model of the Next Generation supply chain was not validated at all: Ericsson deemed this acceptable since this model was built from a validated Current model and this supply chain did not even exist at that time.)

Each simulation model gives several outputs, but we focus on a single output, namely the total cost. (This cost is calculated from inventory-related costs and quality-related costs. The inventory related costs are calculated from the inventory levels throughout the supply chain and the current value of each product at each step in the chain. The quality-related costs concern yield, scrap, and modification time multiplied by the modification cost. The rework affects the inventory related costs with higher levels of inventory as the reworked products once again flow through the supply chain.)

3.2 Inputs and outputs of the three simulation models

In this case study, we apply sequential bifurcation assuming a metamodel consisting of a first-order polynomial augmented with two-factor interactions. This metamodel requires a foldover design, as described in Section 2.3. For example, after the start of our procedure, we set the first 79 factors at their individual high levels, so in the corresponding foldover design point we set these factors at their low levels; the remaining factors are at the levels opposite of the first 79 factors in the same factor combination. At the start of the procedure, we take only two combinations—which happen to form a foldover design. (We could have reduced the number of simulation observations by 50%, had we assumed a first-order polynomial. However, we felt almost certain that interactions are important in this case study. Had we assumed a first-order polynomial, we would have followed different paths towards the important factors. Had we obtained additional observations at the end of the sequential bifurcation to check for interactions, then the path might be rather different so the analysis would have to be done all over again.)

The environmental factors (labeled 80 through 92 in Section 2) are the demand of products, the process yield, and the percentage of scrap at each test-station.(It can be proven that creating one cluster of environmental factors and one cluster of

controllable factors enables estimation of sums of interactions between controllable and environmental factors.)

In this case study, we assume that the outputs of interest are *steady-state* outputs; see Persson and Olhager (2002). Therefore we use a warm-up period. To determine this period, we apply Welch's procedure described by Law and Kelton (2000). This procedure gives a warm-up period of four weeks. (The warm-up period used by Persson and Olhager was only one week; they determined this period through a rule-of-thumb, namely the warm-up period should be three times the longest lead-time in the simulation. Gunn and Nahavandi 2002 show that initial conditions are important in manufacturing simulations.) After this warm-up period, we run each scenario for sixteen weeks of production (we found substantially different outputs for four and eight weeks). We assume that this runlength indeed gives steady-state output.

We label the factors such that all factors have the same meaning in the three simulation models. Therefore, we introduce *dummy factors* for the Current and the Next Generation models that represent those factors that are removed as the supply chain is changed. These dummy factors simplify the calculations and interpretations of the sequential bifurcation results; they have zero effects.

In conclusion, we focus on a single output, namely the expected weekly cost of the total supply chain in the steady state. (Different outputs may have different important factors so the sequential bifurcation paths differ.)

3.3. Results for three simulation models

The *aggregated effects* of the Old supply chain exceed those of the New supply chain, because the former aggregates more (positive) individual effects. For example, the Current simulation model has 14 dummy factors (which have zero effects), so the first sequential bifurcation step gives a smaller main (group) effect: for the Current model this effect is 7,101,983, whereas it is 15,016,102 for the Old model.

Furthermore, the *shortlists* are slightly shorter for the Current and New Generation models. The individual factors on the three shortlists are the same - except for the factors 91 (product demand), 44 and 46 (the latter two factors represent transportation between operations) of the Next generation model. The most important factor (92) is the demand for one of Ericsson's fast selling products. The other factors represent transportation and yield.

Figure 5 illustrates how the estimated *upper limits U* for main effects decrease as new observations are obtained. Furthermore, this figure shows that the most important individual main effect – that of factor 92 - has already been identified and estimated after only ten steps. The next important factor - factor 49 – shows up after sixteen observations, etc.

To *verify* sequential bifurcation's shortlist, we test the effects of the remaining 'unimportant' factors in the Current model, for the following two scenarios. First,

we set all *unimportant factors* at their low values, while keeping the important factors fixed. Secondly, we switch all unimportant factors to their high values, while keeping the important factors fixed. (These two scenarios are not used in sequential bifurcation to reach the shortlist.) We fix the important factors at their base values. These base levels are coded as zero because all the important factors are quantitative; see equation (2.3). We again replicate these scenarios five times. These replicates give averages, standard deviations, and the t statistic. This statistic is 1.53, which is non-significant for any reasonable type-I error rate and any reasonable approximation for the correct number of degrees of freedom (see Section 2.2). So we conclude that the sequential bifurcation shortlist seems valid.

Persson (2003) gives many more details on both the case study and the application of sequential bifurcation to create the shortlists for this case study.

4 Conclusions

The technique of sequential bifurcation is an important and useful method for experiments with simulation models that involve a large number of factors. We have demonstrated the steps of this technique through a case study on three supply chain configurations in the Swedish mobile communications industry. We formalized the assumptions of the technique, and found that in practice these assumptions may not be too restrictive– as our case study illustrates.

In a companion paper—namely, Kleijnen et al. (2003)—we change the metamodel in (2.1) after the screening phase, as follows. For those controllable factors found by sequential bifurcation to be important, we augment (2.1) with *quadratic* effects (for optimization). For those environmental or noise factors identified by sequential bifurcation as important, we create environmental scenarios through Latin Hypercube Sampling (for robustness analysis). Note that Lewis and Dean (2001) also distinguish between controllable and noise factors.

Further research is needed to derive the overall probability of correctly classifying the individual factors as important or unimportant: in our sequential procedure, which tests each factor group individually; also see Lewis and Dean (2001, pp. 663-664), Nelson (2003), and Westfall et al. (1998).

Finally, the extension of sequential bifurcation from a single to multiple responses is an important practical and theoretical problem.

References

Balci, O. (2001), A methodology for accreditation of modeling and simulation applications. *TOMACS*, 11, no. 4, pp. 352-377

Banks, J., S. Buckley, S. Jain, and P. Lendermann (2002), Panel session: opportunities for simulation in supply chain management. *Proceedings of the 2002 Winter Simulation Conference* (edited by E. Yücesan, C.H. Chen, J.L. Snowdon and J.M. Charnes), pp. 1652-11658

Abstract: This contribution discusses experiments with many factors: the case study includes a simulation model with 92 factors. The experiments are guided by sequential bifurcation. This method is most efficient and effective if the true input/output beha 19

Bettonvil, B. (1990), *Detection of important factors by sequential bifurcation*. Tilburg University Press, Tilburg

Bettonvil, B. and J.P.C. Kleijnen (1997), Searching for important factors in simulation models with many factors: sequential bifurcation. *European Journal of Operational Research*, 96, no. 1, pp. 180-194

Box, G.E.P., W.G. Hunter and J.S. Hunter (1978), *Statistics for experimenters: an introduction to design, data analysis and model building*. Wiley, New York

Box, G.E.P. and K.B. Wilson (1951), On the experimental attainment of optimum conditions. *Journal Royal Statistical Society, Series B*, 13, no. 1, pp. 1-38

Campolongo, F., J.P.C. Kleijnen, and T. Andres (2000), Screening methods. In: *Sensitivity analysis*, edited by A. Saltelli, K. Chan, and E.M. Scott, Wiley, Chichester (England), pp. 65-89

Cheng, R.C.H. (1997), Searching for important factors: sequential bifurcation under uncertainty. Proceedings of the 1997 Winter Simulation Conference, edited by S. Andradottir, K.J. Healy, D.H. Withers and B.L. Nelson, pp. 275-280

Cheng, R.C.H. and W. Holland (1999). Stochastic Sequential Bifurcation: Practical Issues. Proceedings of UK Simulation Society Conference, eds D. Al-Dabass and R.C.H. Cheng, UKSIM**,** Nottingham, pp. 74-80

Gunn, B. and S. Nahavandi (2002), Why initial conditions are important. *Proceedings of the 2002 Winter Simulation Conference* (edited by E. Yűcesan, C.H. Chen, J.L. Snowdon and J.M. Charnes), pp. 557-562

Incontrol (2003), Incontrol Enterprise Dynamics (www.EnterpriseDynamics.com)

Kleijnen, J.P.C. (2004), Supply chain simulation: a survey. *International Journal of Simulation and Process Modelling* (forthcoming)

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

Kleijnen, J.P.C., B. Bettonvil, and F. Persson (2003), Robust solutions for supply chain management: simulation, optimization, and risk analysis [for a preprint, see http://center.kub.nl/staff/kleijnen/papers.html]

Kleijnen, J.P.C. and R.G. Sargent (2000), A methodology for the fitting and validation of metamodels in simulation. *European Journal of Operational Research*, 120, no. 1, pp. 14-29.

Kleijnen, J.P.C., S. M. Sanchez, T. W. Lucas, and T. M. Cioppa (2002), A User's Guide to the Brave New World of Designing Simulation Experiments. Working Paper (preprint: http://center.kub.nl/staff/kleijnen/papers.html)

Law, A.M. and W.D. Kelton (2000), Simulation modeling and analysis; Third edition, *McGraw-Hill*, Boston

Lee, H.L., V. Padmanabhan, and S. Whang (1997). The bullwhip effect in supply chains. *Sloan Management Review,* 38, Spring, pp. 93-102

Lewis, S.M. and A.M. Dean (2001), Detection of interactions in experiments on large numbers of factors (Including discussion). *Journal Royal Statistical Society, Series B*, 63, Part 4, pp. 633-672

Morris, M. (2003) An overview of group screening for active factors. In: this book

Myers, R.H. and D.C. Montgomery (20020), *Response surface methodology: process and product optimization using designed experiments; second edition*. Wiley, New York

Nelson, B. L. (2003) Private communication (e-mail: nelsonb@primal.iems.nwu.edu)

Persson, F. (2003) Discrete event simulation of supply chains –modelling, validation, and analysis, Doctoral dissertation Profil 20, Linköping Institute of Technology

Persson, F. and J. Olhager (2002), Performance simulation of supply chain designs. *International Journal of Production Economics*, 77, pp. 231-245

Schonlau, M. and W.J. Welch (2003), Screening the input variables to a computer code via analysis of variance and Visualization, In this book

Simchi-Levi, D., P. Kaminsky, and E. Simchi-Levi (2000). Designing and managing the supply chain: concepts, strategies, and case studies. Irwin/McGraw-Hill, Boston

Westfall, P.H., S.S Young, and D.K.J. Lin (1998), Forward selection error control in the analysis of supersaturated designs. *Statistica Sinica*, 8, pp. 101-117

Wu, C.F.J. and M. Hamada (2000), Experiments; planning, analysis, and parameter design optimization. Wiley, New York

Zeigler, B.P., H. Praehofer, and T.G. Kim (2000), *Theory of modelling and simulation*, 2nd Ed., Academic Press, San Diego

Abstract: This contribution discusses experiments with many factors: the case study includes a simulation model with 92 factors. The experiments are guided by sequential bifurcation. This method is most efficient and effective if the true input/output beha

Figure 1: The three supply chain configurations: (a) the old, (b) the current, (c) the Next Generation

Figure 2: The simulation model of the Current supply chain configuration

Figure 3: Sequential bifurcation assuming a first-order polynomial metamodel, applied to the Old supply chain simulation; displays the parameter estimates at each step

Figure 4: Sequential bifurcation assuming a first-order polynomial plus two-factor interactions metamodel, applied to the Old supply chain simulation; includes upper limits for parameter values

Figure 5. Upper limit $U(i)$ after step i (i = 9, …, 21) and individual main effect estimates (shaded bars) versus the factor label j (j = 1, …, 92) in the Old supply-chain simulation

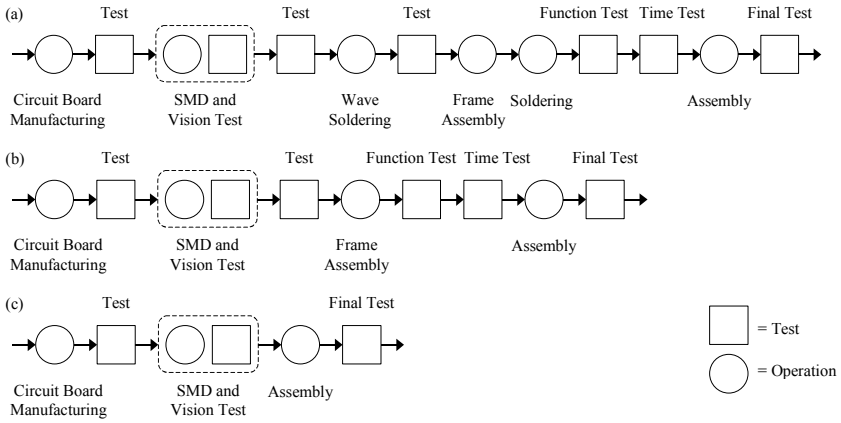

Figure 1: The three supply chain configurations: (a) the old, (b) the current, (c) the next generation

Abstract: This contribution discusses experiments with many factors: the case study includes a simulation model with 92 factors. The experiments are guided by sequential bifurcation. This method is most efficient and effective if the true input/output beha

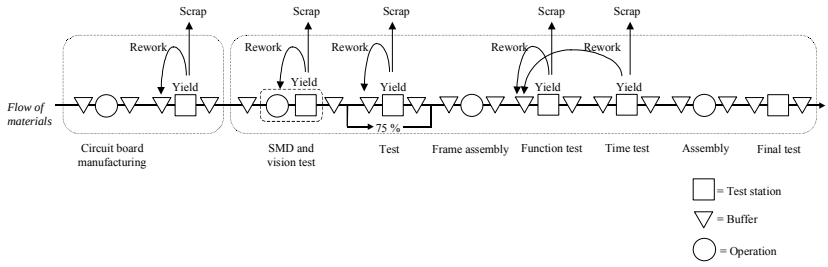

Figure 2: The simulation model of the Current supply chain configuration

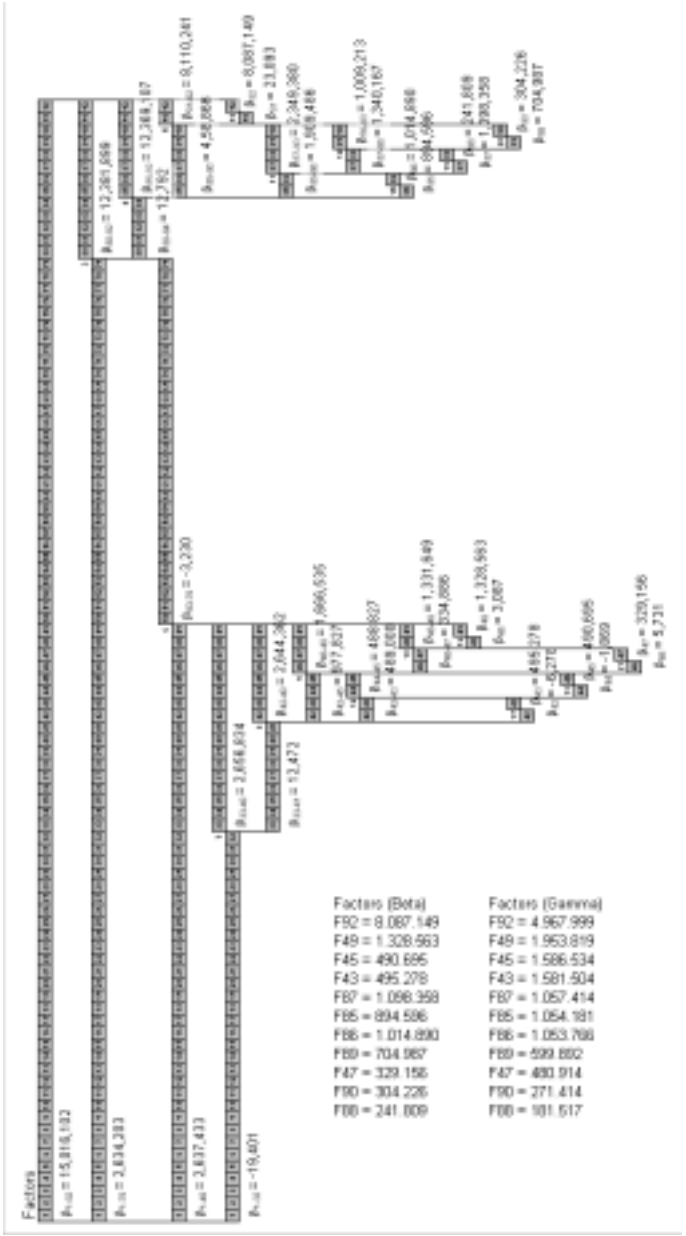

Figure 3: Sequential bifurcation assuming a first-order polynomial metamodel, applied to the Old supply chain simulation; displays the parameter estimates at each step

Abstract: This contribution discusses experiments with many factors: the case study includes a simulation model with 92 factors. The experiments are guided by sequential bifurcation. This method is most efficient and effective if the true input/output beha

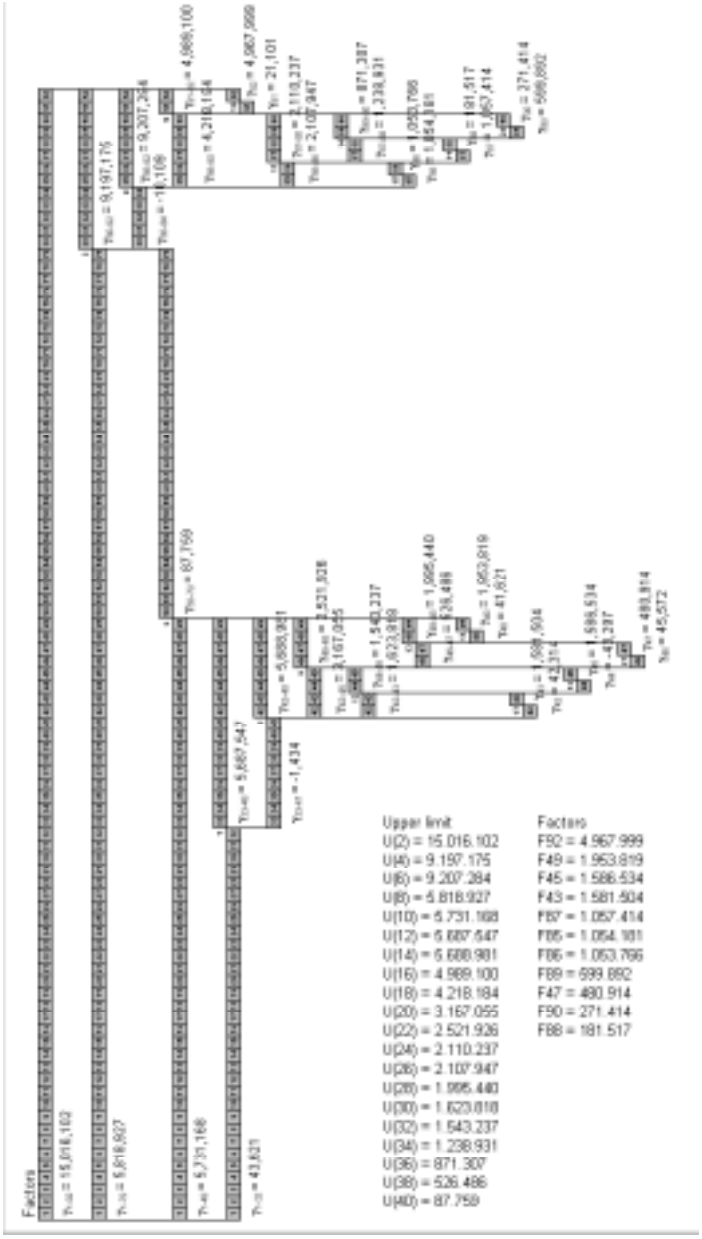

Figure 4: Sequential bifurcation assuming a first-order polynomial plus two-factor interactions metamodel, applied to the Old supply chain simulation; includes upper limits for parameter values

Abstract: This contribution discusses experiments with many factors: the case study includes a simulation model with 92 factors. The experiments are guided by sequential bifurcation. This method is most efficient and effective if the true input/output beha

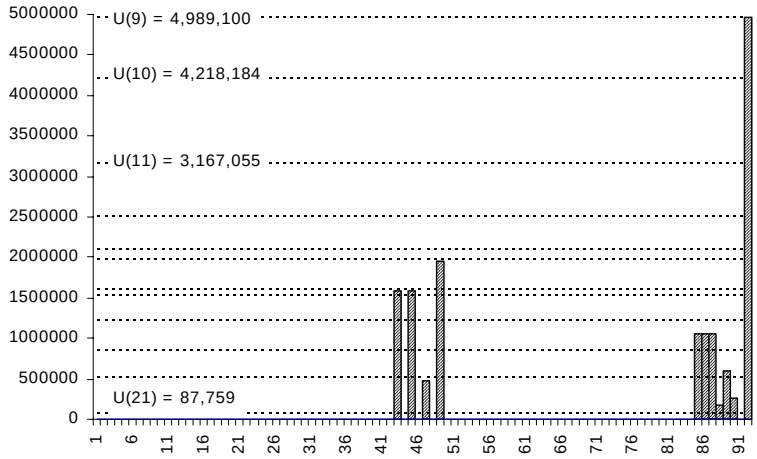

Figure 5. Upper limit $U(i)$ after step i (i = 9, …, 21) and individual main effect estimates (shaded bars) versus the factor label j (j = 1, …, 92) in the Old supply-chain simulation

Table 1: Observations for the first two scenarios simulated in sequential bifurcation for the Old supply chain

Replication	$y_{(0)}$	$y_{(92)}$
1	3,954,024	34,206,800
2	3,975,052	33,874,390
3	3,991,679	33,775,326
4	4,003,475	34,101,251
5	3,983,905	34,111,392
Average	3,981,627	34,013,832
Standard Error	18,633	180,780

Table 2. Important factors identified in sequential bifurcation for the Old supply chain; N/A denotes a dummy factor; ✓ denotes an important factor

		Model		
		Old	Current	Next Generation
Factor				
92	Demand	✓	✓	✓
90	Yield (Circuit board)	✓	✓	✓
89	Yield (Vision test)	✓	✓	✓
88	Yield (SMD test)	✓	✓	N/A
86	Yield (Function test)	✓	✓	✓
85	Yield (Time test)	✓	✓	N/A
87	Yield (Test, after wave soldering)	✓	N/A	N/A
47	Transportation (internal, circuit board factory)	✓	✓	✓
49	Transportation (between factories)	✓	✓	✓
45	Transportation (between SMD and test)	✓	✓	✓
42	Transportation (between wave soldering and test)	✓		