Sensitivity analysis and related analysis

Kleijnen, J.P.C.

Publication date:
1995

\samo95.v2

Sensitivity Analysis and Related Analyses:
a Survey of Statistical Techniques

Jack P.C. Kleijnen

Department of Information Systems and Auditing/Center for Economic Research (CentER)
School of Management and Economics
Tilburg University (Katholieke Universiteit Brabant)
5000 LE Tilburg
Netherlands
Internet: kleijnen@kub.nl
Fax: +3113-663377

International Symposium
SAMO95
Theory and applications of ***S***ensitivity ***A***nalysis of ***M***odel ***O***utput incomputer simulation
25-27 September 1995
Belgirate, Italy

International Symposium
organized by
European Commission DGXII/F
RTD Action Energy
and
Joint Research Centre
Environment Institute

Version 2: 16 October 1995
Version 1: 21 September 1995
(Version 1 was titled ’Sensitivity Analysis, Uncertainty Analysis, and Validation:
a Survey of Statistical Techniques and Case Studies’)

Sensitivity Analysis and Related Analyses:
a Survey of Statistical Techniques

Jack P.C. Kleijnen

Department of Information Systems and Auditing/Center for Economic Research (CentER)
School of Management and Economics
Tilburg University (Katholieke Universiteit Brabant)
5000 LE Tilburg
Netherlands
Internet: kleijnen@kub.nl; fax: +3113-663377

Abstract

This paper reviews the state of the art in five related types of analysis, namely (i) sensitivity or what-if analysis, (ii) uncertainty or risk analysis, (iii) screening, (iv) validation, and (v) optimization. The main question is: when should which type of analysis be applied; which statistical techniques may then be used? This paper distinguishes the following five stages in the analysis of a simulation model. 1) Validation: the availability of data on the real system determines which type of statistical technique to use for validation. 2) Screening: in the simulation's pilot phase the really important inputs can be identified through a novel technique, called sequential bifurcation, which uses aggregation and sequential experimentation. 3) Sensitivity analysis: the really important inputs should be subjected to a more detailed analysis, which includes interactions between these inputs; relevant statistical techniques are design of experiments (DOE) and regression analysis. 4) Uncertainty analysis: the important environmental inputs may have values that are not precisely known, so the uncertainties of the model outputs that result from the uncertainties in these model inputs should be quantified; relevant techniques are the Monte Carlo method and Latin hypercube sampling. 5) Optimization: the controllable inputs should be steered; a relevant technique is Response Surface Methodology (RSM), which combines DOE, regression analysis, and steepest-ascent hill-climbing. This approach with its five stages implies that sensitivity analysis should precede uncertainty analysis. This paper briefly discusses several case studies for each phase.

Keywords: Sensitivity analysis, what-if, uncertainty analysis, risk analysis, validation, optimization, regression analysis, least squares, design of experiments, screening, Latin hypercube sampling, perturbation

1. Introduction

Sensitivity analysis or SA (also called what-if analysis) is applied in simulation studies of very different real-life systems, in all kinds of disciplines (that use mathematical models): chemistry, physics, engineering, economics, management science, and so on. Moreover, the theoretical aspects of SA are studied in mathematics and statistics. Unfortunately, the definition of SA (and of related analyses such as Response Surface Methodology or RSM and validation) varies over and within these many disciplines! Because terminologies differs so much, communication among different disciplines is difficult. Yet, cross-fertilization among these disciplines is certainly possible and fruitful (for example, Jan Tinbergen received the Nobel Prize in economics, but got his academic training in physics).

Especially the roles of SA and uncertainty analysis (or UA, also called risk analysis) seem unclear (see Kleijnen 1994). It seems wise to first consider questions that might be asked by the *users* of simulation models (these users are the clients of the simulation modelers, and 'the customer is king'):

(i) What is the probability of a nuclear accident happening at our site?

(ii) How big is the chance that the financial investment in our project will turn out to be unprofitable?

(iii) What is the probability of customers having to wait longer than three minutes at our supermarket's checkout lanes?

Next consider questions asked by simulation modelers (or analysts):

(i) What happens if parameter h (with $h = 1, ..., k$) of our simulation model is changed?

(ii) Which are the really important parameters among the hundreds of parameters of our total simulation model (that consists of several modules or submodels)?

Uncertainty about a simulated system's response (outcome) may have two different causes:

(i) The system's process is well-known, so it is represented by a *deterministic* model; however, its *parameters* are not known exactly. Many examples can be found in classical physics. Karplus (1983) gives an interesting survey of different types of models in different disciplines, ranging from astronomy through ecology to social sciences, using white box, grey box, and black box models respectively.

(ii) Some models are *intrinsically stochastic*: without the randomness the problem disappears. In the supermarket example the customer arrival and service times are random, which creates stochastic customer waiting times. In an ecological model an example may be the wind's speed and direction. Helton (1993, 1995) speaks of 'stochastic or aleatory uncertainty'. Also see Van Asselt and Rotmans (1995, pp. 11-13) and Zeigler (1976).

Uncertainty is the central problem in *mathematical probability theory*. This discipline, however, has two schools.

(i) The *objectivists or frequentists* school: for example, when throwing a dice many times, the probability of throwing a six can be defined as the limit of the frequency of throwing a six (obviously this limit is 1/6 for a fair dice).

(ii) The *subjectivists* school: an event may be unique (for example, to-morrow's weather at a particular place); yet a 'probability' (between zero and one) may be assigned to a particular outcome (for example, sunny all day). Many UA studies concern such unique events; for example, the following questions by the users were mentioned above: (i) What is the probability of a nuclear accident happening at our site? (ii) How big is the chance that our financial investment in this project will turn out to be unprofitable? Note that *Bayesians* try to combine prior subjective data with new factual data.

Zeigler wrote a seminal book on the theory of modeling and simulation, distinguishing between *input variables* and *parameters*; see Zeigler (1976). A variable can be directly observed; an example is the number of check-out lanes in a supermarket. A parameter, however, can *not* be observed directly, so its value must be inferred from observations; an example is the arrival rate of customers. Hence mathematical statistics may be used to quantify the probability of certain values for a parameter (see §5.1). When applying design of experiments (such as fractional factorial designs) to simulation models, this paper will use the term *factor* for parameters, input variables, and modules that are changed from (simulation) run to run. (For modules treated as factors also see Helton et al. 1995, McKay 1995, pp. 51-54, and Van Asselt and Rotmans 1995.)

The *objective* of this paper is to *review the state of the art* in five related types of analysis, namely (i) SA, (ii) UA, (iii) screening, (iv) validation, and (v) optimization. The main question is: when should which type of analysis be applied; which statistical techniques may then be used? What, however, is meant by these terms in this paper?

Sub (i): The vast literature on simulation does not provide a standard definition of SA. This paper defines *sensitivity analysis* as the systematic investigation of the reaction of the simulation response to either *extreme* values of the model's quantitative factors (parameters and input variables) or to drastic changes in the model's qualitative factors (modules). For example, what happens to the mean waiting time in a supermarket, when the arrival rate doubles; what happens if a 'fast lane' is opened for customers with less than (say) seven articles? So the focus is not on marginal changes in inputs (such small perturbations or local sensitivities are discussed elsewhere at this conference; also see the survey in Helton 1993).

Note that this paper concentrates on a single response per run; multiple responses are only briefly mentioned in §6 (on RSM); also see McKay (1995).

Sub (ii): In *uncertainty analysis*, values of the model inputs are sampled from prespecified distributions, to quantify the consequences of the uncertainties in the model inputs, for the model outputs. So in UA the input values range between the extreme values investigated in SA. The goal of UA is to quantify the *probability* of specific output values, whereas SA does not tell how likely a specific result is. The differences between SA and UA will be further explored later on.

Sub (iii): *Screening* is needed whenever a simulation study is still in its early, *pilot* phase and *many factors* may be conceivably important (dominant). Straightforward numerical experiments with such a simulation model may require too much computer time.

Sub (iv): *Validation* concerns the question: is the simulation model an adequate representation of the corresponding system in the real world? Model validity should be of major interest to both users and analysts.

Sub (v): From the users' viewpoint the important model inputs should be split into two groups, namely inputs that are under the users' *control* (policy variables) versus *environmental* inputs, which are not controllable. Specifically, users want to ask 'what if' questions: what happens if controllable inputs are changed (scenario analysis), what if other inputs change (UA)? Users may wish to *optimize* controllable inputs.

This paper covers both practical and theoretical aspects (which is in line with the objective of this conference). It covers these aspects at an advanced tutorial level; that is, the readers are assumed to be experienced professionals in simulation. For technical details, a bibliography with 77 references is included. The review synthesizes the latest research results in a unified treatment of the various analysis types. It also briefly discusses selected applications of these analyses. This paper combines, updates, and revises Kleijnen (1994, 1995b).

Unfortunately, nobody can be an expert in the many disciplines covered by this conference! This paper is biased by more than 25 years of experience with simulation, especially its statistical aspects and its applications to problems in business, economics, environmental, agricultural, military, and computer systems.

The *main conclusion* will be that the simulation analysts should distinguish *five phases* in a simulation study; the *techniques* to be used in these phases are as follows.

(i) *Validation* may use regression and DOE, especially if there are no data on the input/output (I/O) of the simulation model or its modules. If there are ample data, then the simulation model may be validated by a special type of regression analysis.

(ii) *Screening* may use sequential bifurcation, which is a simple, efficient, and effective technique.

(iii) SA may use *regression analysis* to generalize the results of the simulation experiment, since regression analysis characterizes the I/O behavior of the simulation model. The estimators of the effects in the regression model should be based on *DOE*, which gives better estimators of main (first-order) effects and interactions among factors; these designs require fewer simulation runs than intuitive designs (such as changing one factor at a time or 'ceteris paribus' approach).

(iv) UA uses *Monte Carlo* sampling, possibly including variance reduction techniques such as *Latin hypercube sampling* or LHS, possibly combined with regression analysis.

(v) *Optimization* of the controllable inputs may use *Response Surface Methodology* or RSM, which builds on regression analysis and DOE; see (iii). Note that in nuclear engineering the term RSM refers to SA without optimization; see Olivi (1980) and Rao and Sarkar (1995).

An important conclusion is that SA should precede UA. McKay (1995, p.33), however, proposes a different sequence of steps in UA.

These statistical techniques have already been *applied* many times in practical simulation studies, in many domains (also see the many contributions at this conference). These techniques make simulation studies give more general results, in less time.

For didactic reasons, this paper is organized as follows. In §2 screening and especially sequential bifurcation are discussed. In §3 SA through regression analysis and DOE is explained. In §4 validation is briefly addressed. In §5 UA is discussed, distinguishing between deterministic and stochastic simulations. In §6 optimization through RSM is explained. In §7 conclusions are summarized. A bibliography ends this paper.

2. Screening: Sequential Bifurcation

Screening is the search for the few (say) k really important factors among the great many (say) N potentially important factors ($k << N$). In practice, experiments with simulated systems often do involve many factors. Andres (1995) gives an example with 3300 factors that affect deep geological waste disposal. Rahni et al. (1995) screen 390 factors in an experimental building (a physical simulation, not a mathematical simulation), developed to study energy flows in buildings. Another example is the case study in this paper, which has 281 factors; this study concerns a complicated deterministic simulation model of the 'greenhouse' phenomenon (the increase in temperatures worldwide, caused by increasing quantities of carbon dioxide or CO_2 and of other gases in the atmosphere). McKay (1995) discusses case studies with 84 and 36 factors respectively. Other examples would be stochastic simulations of queueing systems, such as logistic, computer, and telecommunication networks with nodes that represent individual servers (machines). These queueing models have many parameters (service and arrival rates), input variables (number of servers), and behavioral relationships (queuing discipline or priority rule). In summary, simulation models with *several hundred factors* are common; also see Morris (1991).

The problem is that a simulation run may require so much *computer time* that the number of runs (say) n must be much smaller than the number of factors ($n << N$). For example, the greenhouse model with 281 factors takes 15 minutes of computer time per run, so 282 runs would have taken roughly 70 hours of computer time (on the computer available at that time: PC 386DX, 16 Mhz). In general, the analysts and their clients may not wish to wait until (roughly) N runs will have been generated. (The greenhouse model is part of an even bigger model, called 'IMAGE', Integrated Model to Assess the Greenhouse Effect, which was developed at the Dutch institute RIVM, National Institute of Public Health and Environmental Protection; see Rotmans 1990).

Whenever there are many factors the analysts should assume that only a few factors are really important ($k << N$): *parsimony, Occam's razor*. Moreover, to solve the screening problem ($k << N$, $n << N$), one general scientific principle can be applied: *aggregation*. Indeed, aggregation has been used in the study of many large systems; for example, in economic theory and practice the production volumes of the individual companies are aggregated into the Gross National Product or GNP.

More specifically, in the experimentation with large systems (real or simulated), analysts have applied *group screening*: they combine individual factors into groups and experiment with these groups as if they were individual factors. The theory on group screening goes back a long time: Jacoby and Harrison (1962), Li (1962), Patel (1962), and Watson (1961). A more recent publication is Morris (1987).

Practical applications of group screening, however, are rare. The reason may be that in experiments with real systems it is impossible to control hundreds of factors. And in simulation, most analysts seem to be unaware of group screening. Four simulation applications of group screening are summarized in Kleijnen (1987, p. 327); other applications are given in Cochran and Chang (1990) and Rahni et al. (1995).

This paper focusses on a special group screening technique, namely *sequential bifurcation* or SB. SB uses a special design; it also applies a different analysis. SB proceeds sequentially (or stagewise) and splits up (or bifurcates) the aggregated inputs as the experiment proceeds, until finally the important individual inputs are identified and their effects are estimated. SB turns out to be more efficient than competing group screening techniques. Moreover, SB has been used in practice, namely in the greenhouse simulation mentioned above, and in the building thermal deterministic simulation in De Wit (1995). For the greenhouse simulation SB found the 15 most important inputs among the 281 factors after only 144 runs. SB lead to surprising results: the technique identified some factors as being important that the ecological experts assumed to be unimportant, and vice versa. In De Wit's building simulation, SB found the 16 most important inputs among the 82 factors after only 50 runs. De Wit checked these results by applying a different screening technique, namely Morris's (1991) 'randomized one-factor-at-a time designs', which took 328 runs.

How general are the results of these SB applications; can SB be applied with confidence to other simulation models? Scientists must always make *assumptions* to make progress. More specifically, any screening technique must use assumptions. In practice, the simulation analysts often leave their assumptions implicit. Frequently the analysts assume that they know which factors are unimportant, and they investigate only a few *intuitively* selected factors; also see Bankes and Lucas (1995). (Often they apply an inefficient or ineffective design: they change one factor at a time; see Van Groenendaal (1994).) All group screening techniques (including SB) make the following two *assumptions*:

(i) a *low-order polynomial approximation* or metamodel for the I/O behavior of the simulation model;

(ii) *known* signs or directions for the first-order or main effects.

Sub (i): A metamodel implies that the underlying simulation model is treated as a *black box*. The advantage of a low-order polynomial is that it is simple and that it may apply 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. Note that other techniques do use the specific structure of the specific simulation model; examples are importance sampling (see §5.1) and differential or perturbation analysis (see Helton 1993 for continuous systems and Ho and Cao 1991 for discrete-event systems). However, these sophisticated techniques have not yet been applied to simulations with hundreds of factors.

Low-order polynomials are often used in DOE with its concomitant Analysis of Variance (ANOVA), applied to real or simulated systems; for a survey see Kleijnen and Van Groenendaal(1992).

The classic group screening designs assume a first-order polynomial. In SB a first-order polynomial requires only half the number of runs that an approximation with two-factor interactions does (foldover principle; see §3.3.2 and also Andres 1995 and Kleijnen 1987). In general, however, a more cautious approach is recommended, namely a metamodel with interactions.

Sub (ii): Group screening assumes known signs or directions for the effects, in order to know with certainty that individual effects do *not compensate* each other within a group. In practice the sign of a factor may be known indeed (the magnitude, however, is unknown, so simulation is used). For example, in the greenhouse study the ecological experts felt confident when they had to specify the signs of the factor effects, for most factors. In queueing networks the response may be throughput, so the analysts may assume that higher speeds of servers have non-negative effects on the response; and so does an increasing arrival rate of jobs at the source of the network.

If the analysts feel that they do *not* know the signs of *a few* factors, then they may treat these factors separately. Indeed, in the greenhouse case study there is a very small group of factors with unknown signs. These factors can be investigated in a traditional design.

A special characteristic of SB is that the analysts do not need to quantify *a priori* how big a factor effect should be in order to be called *important*. As simulation outputs are generated sequentially, SB computes upper limits for the factor effects, and the analysts can stop the simulation experiment as soon as they find these limits sharp enough. Obviously, the analysts may make these limits depend on the system being simulated.

The main objective of this section is to inform simulation analysts about a novel technique for the screening of large simulation models. Therefore concepts are emphasized; for technical details the readers is referred to the more than 200 pages of the doctoral dissertation, Bettonvil (1990) or to the summary paper by Bettonvil and Kleijnen (1995). Different screening techniques are Andres's (1995) Iterated Fractional Factorial Design (IFFD) and McKay's (1995) replicated LHS design.

3. Sensitivity Analysis: Regression Analysis and DOE

Sensitivity analysis was defined as the systematic investigation of the reaction of the simulation response to either *extreme* values of the model's quantitative factors (parameters and input variables) or to drastic changes in the model's qualitative factors (modules). This section assumes that the number of factors is relatively small, that is, the screening phase is over. An example with fourteen factors will be mentioned in §6 (section on optimization). In the sensitivity phase, regression analysis may be used to approximate the I/O behavior of the simulation model. This regression analysis gives better results when the simulation experiment is well designed, using classical statistical designs such as fractional factorials.

3.1 Introduction: Graphical Methods

Practitioners often make a *scatter plot* with on the y-axis the simulation response (say, average waiting time) and on the x-axis the values of one factor (for example, service rate). This graph indicates the I/O behavior of the simulation model, treated as a black box. It shows whether this factor has a positive or negative effect on the response, whether that effect remains constant over the domain or experimental area of the factor, and so on. Also see Helton (1993, pp.347-349).

The practitioners may further analyze this scatter plot: they may fit a curve to these (x, y) data, for example, a straight line ($y = \beta_0 + \beta_1 x$). Of course, they may fit other curves, such as a quadratic (second degree polynomial), an exponential, or a logarithmic curve (using paper with a log scale).

To study *interactions* between factors, they may combine scatter plots per factor. For example, they drew one scatter plot for different values of the service rate, given a certain number of servers. They can now superimpose plots for different numbers of servers. Intuitively, the waiting time curve for a low number of servers should lie above the curve for a high number of servers (if not, the simulation model is probably wrong; see the discussion on validation in §4). If the response curves are not parallel, there are interactions, by definition.

However, superimposing many plots is cumbersome. Moreover, their interpretation is subjective: are the response curves really parallel and straight lines? These shortcomings are removed by regression analysis.

3.2 Regression analysis

A regression metamodel was defined as an approximation of the I/O behavior of the underlying simulation model. Consider the *second degree polynomial*

$$Y_i = \beta_0 + \sum_{h=1}^{k} \beta_h x_{ih} + \sum_{h=1}^{k} \sum_{h'=h}^{k} \beta_{hh'} x_{ih} x_{ih'} + E_i \qquad (1)$$
$$(i = 1, \ldots, n)$$

with

- Y_i: simulation response of factor combination i (stochastic variables are shown in capitals);
- β_0: overall mean response or regression intercept;
- β_h: main or first-order effect of factor h;
- x_{ih}: value of the standardized factor h in combination i (see Equation (2) below);
- $\beta_{hh'}$: interaction effect of the factors h and h' with $h \neq h'$;
- β_{hh}: quadratic effect of factor h;
- E_i: fitting error of the regression model for factor combination i;
- n: number of simulated factor combinations.

First ignore interactions and quadratic effects, for didactic reasons. Then the *relative importance* of a factor is obtained by sorting the absolute values of the main effects β_h, provided the factors are *standardized*. So let the original (non-standardized) factor h be denoted by z_h. In the simulation experiment z_h ranges between a lowest value (say) l_h and an upper value u_h; that is, either the simulation model is not valid outside that range (see the discussion on validation in §4) or in practice that factor can range over that domain only (for example, the number of servers can vary only between one and five). The variation (or spread) of factor h is measured by $a_h = (u_h - l_h)/2$; its location (or mean) by $b_h = (u_h + l_h)/2$. Then the following standardization is appropriate:

$$x_{ih} = (z_{ih} - b_h)/a_h. \qquad (2)$$

Note that other measures besides $|\hat{\beta}_h|$ have been proposed; see McKay (1995); also see Helton (1993), Saltelli (1995), and Sobol (1995). More research and applications are needed.

The classic *fitting algorithm*, which determines the regression parameters $\beta = (\beta_0, \beta_1, \ldots, \beta_{kk})'$ in Equation (1), uses the *ordinary least squares* (OLS) criterion. Software for this algorithm is abundant.

If *statistical assumptions* about the fitting error are added, then there are better algorithms. Consider the following assumptions.

It is realistic to assume that the variance of the stochastic fitting error E varies with the input combination of the random simulation model: $\mathrm{var}(E_i) = \sigma_i^2$. (So both the mean and the variance of Y, the response of the stochastic simulation, depend on the input.) Then *weighted least squares* or WLS (with the standard deviations σ_i as weights) yields unbiased estimators of the factor effects, but with smaller variances than OLS gives.

To improve the SA, *common* pseudorandom number seeds should be used when simulating different factor combinations. Then, however, *generalized least squares* or GLS should be applied to get minimum variance, unbiased estimators. Unfortunately, the variances and covariances of the simulation responses Y_i are unknown, so they must be estimated. The following equation gives the classic covariance estimator, assuming d independent replications or runs per factor combination (so Y_{ig} and $Y_{i'g}$ are correlated, but Y_{ig} and $Y_{ig'}$ are not):

$$c\hat{o}v(Y_i, Y_{i'}) = \sum_{g=1}^{d} (Y_{ig} - \bar{Y}_i)(Y_{i'g} - \bar{Y}_{i'})/(d - 1). \tag{3}$$

The resulting *estimated GLS* gives good results; see Kleijnen and Van Groenendaal (1992).

To make statistical inferences (for example, about the importance of the factor effects), a *Gaussian* distribution is normally assumed. To satisfy this assumption, the analysts may apply transformations such as the logarithmic and the rank transformations. Alternatively, the analysts might hope that the results are not sensitive to 'mild' non-normality. Moreover, they may analyze both the original and the transformed data, and see whether both analyses give the same qualitative results. See Andres (1995), Kleijnen (1987), Kleijnen, Bettonvil, and Van Groenendaal (1995), and Saltelli (1995).

Of course, it is necessary to check the fitted regression metamodel: is it an adequate approximation of the underlying simulation model? Therefore the metamodel may be used to predict the outcome for a *new* factor combination of the simulation model. So replace β in the specified metamodel by the estimate $\hat{\beta}$, and substitute a new input combination (there are n old combinations). Compare the regression prediction $\hat{y}_{n+1}$ with the simulation response y_{n+1}.

A refinement is *cross-validation*: do not add a new combination (which requires computer time), but eliminate one old combination (say combination i) and re-estimate the regression model from the remaining n - 1 combinations. Repeat this elimination for all values of i (with i = 1, ..., n; see Equation (1)). Statistical details are discussed in Kleijnen and Van Groenendaal (1992); also see Helton (1993, pp. 347-356).

Applications of regression metamodeling will be discusses below.

Note that the analysts may use a covariance stationary process (instead of white noise) to model the systematic effects of the inputs; see Sacks, Welch, Mitchell, and Wynn (1989).

3.3 Design of Experiments

The preceding subsection (§3.2) used regression metamodels to approximate the I/O behavior of simulation models. Such a metamodel has (say) q regression parameters in the vector β, which measure the effects of the k factors; for example, q equals k + 1 if there are no high-order effects (see §3.3.1), but if there are interactions between factors, then q increases with $k(k - 1)/2$.

It is obvious that to get unique, unbiased estimators of these q effects, it is necessary to simulate at least q factor combinations ($n \geq q$). Moreover, *which* n combinations to simulate (provided that $n \geq q$), can be determined such that the accuracy of the estimated factor effects is maximized (variance minimized). This is the goal of the statistical theory on DOE (which Fisher started in the 1930s and Taguchi continues today).

3.3.1 Main Effects Only

Consider a first-order polynomial, which is a model with only k main effects, besides the overall mean (see the first two terms in the right-hand side of Equation (1)).

In practice, analysts usually first simulate the 'base' situation, and next they change *one factor at a time*; hence $n = 1 + k$. See Helton et al. (1995, p.290) and Van Groenendaal (1994).

However, DOE concentrates on *orthogonal* designs, that is, designs that satisfy

$$\boldsymbol{x}'\boldsymbol{x} = n\boldsymbol{I} \tag{4}$$

where **bold** letters denote matrices, and

$\boldsymbol{x} = (x_{ij})$: design matrix with $i = 1, ..., n$; $j = 0, 1, ..., k$; $n > k$;
$x_{i0} = 1$: dummy factor corresponding with β_0;
x_{ih}: see text below Equation (1);
$\boldsymbol{I}$: identity matrix (this capital letter does not denote a stochastic variable).

Orthogonal designs give estimators of β that are unbiased and have smaller variances than the estimators resulting from designs that change one factor at a time.

Orthogonal designs are tabulated in many publications. The analysts may also learn how to construct those designs; see Kleijnen (1987). Recently, software has been developed to help the analysts specify these designs; see Őren (1993).

A well-known class of orthogonal designs are 2^{k-p} fractional factorials; for example, a simulation with $k = 7$ factors requires $n = 2^{7-4} = 8$ factor combinations (runs) to estimate the main effects plus overall mean. Actually, these 2^{k-p} designs also require 8 runs when $4 \leq k \leq 7$. See Kleijnen (1987).

References to many *simulation applications* of these designs can be found in Kleijnen (1987) and Kleijnen and Van Groenendaal (1992).

In practice, however, it is unknown whether only main effects are important. Therefore orthogonal designs with $n \approx k + 1$ should be used only in optimization (see §6). Moreover these designs are useful as building blocks if interactions are accounted for; see §3.3.2.

3.3.2 Main Effects Biased by Interactions?

It seems prudent to assume that *interactions* between pairs of factors may be important. Then the k main effects can still be estimated without bias caused by these interactions. However, the number of simulated factor combinations must be *doubled* (foldover principle; also see §2); for example, $k = 7$ now requires $n = 2 \times 8 = 16$. These designs also give an indication of the importance of interactions (also see §3.3.3).

Note that designs that change one factor at a time (see §3.3.1) do not enable estimation of interactions!

Details, including *simulation applications* are presented in Kleijnen (1987) and Kleijnen and Van Groenendaal (1992). Recent applications include the simulation of a decision support system (DSS) for the investment analysis of gas pipes in Indonesia, and a simulation model for the Amsterdam police; see Van Groenendaal (1994) and Van Meel (1994) respectively.

3.3.3 Factor Interactions

Suppose the analysts wish to estimate the *individual two-factor interactions* $\beta_{hh'}$ with $h \neq h'$; see Equation (1). There are $k(k - 1)/2$ such interactions. Then many more simulation runs are necessary; for example, $k = 7$ now requires $n = 2^{7-1} = 64$ factor combinations. Therefore practitioners apply these designs only for small values for k. Kleijnen (1987) gives details, including applications.

Of course, if k is really small, then all 2^k combinations should be simulated, so that all interactions (not only two-factor interactions) can be estimated. In practice, these *full factorial* designs are sometimes used indeed (but high-order interactions are hard to interpret). An example with $k = 6$ is provided in Rao and Sarkar (1995). Also see Kleijnen (1987).

3.3.4 Quadratic Effects: Curvature

If the quadratic effects β_{hh} in Equation (1) are to be estimated, then at least k extra runs are needed (since $h = 1, ..., k$). Moreover, each factor must be simulated for more than two values.

Popular in statistics and in simulation are *central composite designs*. They have five values per factor, and require many runs ($n >> q$); for example, $k = 2$ factors have $q = 6$ effects, but $n = 9$ factor combinations are simulated. An application in nuclear engineering is given by Rao and Sarkar (1995). For more details see Kleijnen (1987) and Kleijnen and Van Groenendaal (1992).

Applications are found in the *optimization* of simulation models (see §6).

Note that simulating as many as five values per factor resembles UA, in the sense that the range of factor values is well covered.

4. Validation

This paper is confined to the role of SA (§3) in validation; other statistical techniques for validation and verification are discussed in Kleijnen (1995a). Obviously, validation is one of the first problems that must be solved in a simulation study; for didactic reasons, however, validation is discussed in this part of the paper.

True validation requires that *data* on the real system be available. In practice, the amount of data varies greatly: data on failures of nuclear installations are rare, whereas data on electronically monitored systems (such as computers and supermarkets) are abundant.

If data are available, then many statistical techniques can be applied. For example, simulated and real data on the response, can be compared through the Student statistic for paired observations, assuming the simulation is fed with real-life input data: *trace driven simulation*. A better test uses well-known regression analysis, but does *not* test whether real and simulated data lie on a straight line through the origin! Instead, the difference between simulated and real data is regressed on their sum; this novel test is discussed in Kleijnen, Bettonvil, and Van Groenendaal (1995).

However, if no data are available, then the following type of SA can be used. The analysts and their clients do have *qualitative* knowledge of certain parts of the real system; that is, they do know in which direction certain factors affect the response of the corresponding module in the simulation model (also see the discussion on sequential bifurcation in §2). If the regression metamodel (see §3) gives an estimated factor effect with the *wrong sign*, this is a strong indication of a wrong simulation model or a wrong computer program.

Applications are given in Kleijnen, Van Ham, and Rotmans (1992), who discuss the greenhouse model IMAGE, and Kleijnen (1995c), who discusses a military model, namely the hunt for mines on the bottom of the sea. These applications further show that the validity of a simulation model is restricted to a certain domain of factor combinations. This domain corresponds with the *experimental frame* in Zeigler (1976), defined as the limited set of circumstances under which the real system is to be observed or experimented with.

Moreover, the regression metamodel shows which factors are most important. For the important *environmental inputs* the analysts should try to collect data on the values that occur in practice. If they do not succeed in getting accurate information, then they may use the UA of the next section.

5. Uncertainty Analysis: Monte Carlo and Latin Hypercube Sampling

As the preceding section mentioned, the analysts may be unable to collect reliable data on important environmental inputs, that is, the values that may occur in practice are uncertain. Then the analysts may apply UA. The goal of UA is to quantify the *probability* of specific output values, whereas SA (as defined in §3) does not tell how likely a specific result is. The differences between SA and UA are further explored below.

5.1 The basics of uncertainty analysis

First the analysts derive a *probability function* for the input values. This distribution may be estimated from sample data, if those data are available; otherwise this distribution must be based on subjective expert opinions (also see Helton 1993, pp. 337-341, Helton et al. 1992, chapter 2, p. 4, Helton et al. 1995, p. 288, and Kraan and Cooke 1995). Popular distribution types are uniform, loguniform, triangular, beta, normal, and lognormal distributions. Usually the inputs are assumed to be statistically independent. Nevertheless, correlated inputs are discussed in Bedford and Meeuwissen (1995), Helton (1993, pp.343-345), Helton et al. (1992, chapter 3, p. 7), Meeuwissen and Cooke (1994), Reilly, Edmonds, Gardner, and Brenkert (1987), and Reilly (1995).

Next the analysts use pseudorandom numbers to sample input values from those distributions: *Monte Carlo or distribution sampling*.

UA often uses *Latin hypercube sampling* (LHS), which forces the sample of size (say) n to cover the whole experimental area; for example, in case of a single input, this input’s domain is partitioned into (say) s equally likely subintervals and each subinterval is sampled s/n times. See Helton (1993, pp. 341-343).

This paper’s message is that LHS is recommended as a *variance reduction technique* or VRT, not as a *screening* technique. For screening purposes the inputs should be changed to their extreme values, whereupon their effects should be computed; see the discussion on screening in §2. Of course, the larger sample in LHS gives more insight than the small sample in screening does; however, for a large number of factors such a large sample is assumed to be impossible. Also see Banks (1989) versus Downing et al. (1986) and McKay (1992).

The sampled input values are fed into the simulation model. This subsection is focused on *deterministic* simulation models (the next subsection covers stochastic models). Hence, *during* a simulation run all its inputs are deterministic; for example, the input is constant or shows exponential growth. From run to run, however, the (sampled) inputs vary; for example, constants or growth percentages change. These sampled inputs yield an estimated distribution of output or response values. That distribution may be characterized by its location (measured by the mean, modus, and median) and its dispersion (quantified by the standard deviation and various percentiles or quantiles, such as the 90% quantile). For a basic introduction to UA see Helton (1993) and Kleijnen and Van Groenendaal (1992, pp. 75-78).

Which quantities sufficiently summarize a distribution function, depends on the users’ *risk attitude*: risk neutral (in that case the mean is a statistic that characterizes the whole distribution sufficiently), risk aversion, or risk seeking; see Balson, Welsh, and Wilson (1992) and Bankes (1993, p. 444). The former authors further distinguish between *risk assessment* (defined as risk analysis in this paper) and *risk management* (risk attitude, possible countermeasures); also see Brehmer, Eriksson, and Wulff (1994), Hora (1995), and Van Asselt and Rotmans (1995).

Combining UA with *regression analysis* gives estimates of the effects of the various inputs; that is, regression analysis shows which inputs contribute most to the uncertainty in the output. (Mathematically, this means that in Equation (1) the deterministic independent variables x_{ik} are replaced by random variables.) Because more values are sampled per factor, more complicated metamodels might now be used. Indeed, for prediction purposes these metamodels may be made really complicated; for example, splines may be used. For explanatory purposes and SA, however, simple metamodels may be preferred; also see Kleijnen (1979). Note that Helton et al. (1991, 1992) call this combination of uncertainty and regression analysis ’sensitivity analysis’.

UA is applied in *business and economics*. Hertz (1964) introduced this analysis into investment analysis: what is the probability of a negative Net Present Value? Krumm and Rolle (1992) give recent applications in the Du Pont company. Birge and Rosa (1995) and Van Groenendaal and Kleijnen (1995) also discuss investment analysis issues. UA in business applications may be implemented through add-ons (such as @RISK and Crystal Ball) that extend spreadsheet software (such as Lotus 1-2-3 and Excel). Moreover, these add-ons are augmented with distribution-fitting software (such as BestFit) and optimization software (such as What'sBest).

In the *natural sciences*, UA is also popular. For example, in the USA the Sandia National Laboratories has performed many uncertainty analyses for nuclear waste disposal (Helton et al. 1991, 1992). Oak Ridge National Laboratory has investigated radioactive doses absorbed by humans (Downing et al. 1985). Nuclear reactor safety has been investigated for the Commission of the European Communities (Olivi 1980 and Saltelli and Homma 1992). UA has also been performed at the Dutch RIVM (Harbers 1993, Janssen et al. 1992). Three environmental studies for the electric utility industry were presented in Balson et al. (1992). UA in the natural sciences has been implemented through software such as LISA (see Saltelli and Homma 1992, p. 79), PRISM (Reilly et al. 1987), and UNCSAM (Janssen et al. 1992).

Note that UA is also used in the analysis of computer security; see Engemann and Miller (1992) and FIPS (1979).

The beginning of this section (§5) mentioned that a basic characteristic of UA is that information about the inputs of the simulation model is not reliable; therefore the analysts do not consider a single 'base value' per input variable, but a distribution of possible values. Unfortunately, the form of that distribution must be specified (by the analysts together with their clients). There is the danger of *software driven* specification; that is, the analysts concentrate on the development of software that implements a variety of statistical distributions, but their clients are not familiar at all with the implications of these distributions; also see Easterling (1986). Bridging this gap requires intensive collaboration between model users, model builders, and software developers.

Consequently, it may be necessary to study the effects of the *specification* of the input distributions (and of other types of inputs such as scenarios). This type of SA may be called *robustness analysis*. Examples are given by Helton et al. (1992, section 4.6); also see Helton et al. (1995), Janssen et al. (1992), Kleijnen (1987, pp. 144-145), and McKay (1995, p. 31).

Robustness analysis may also use more sophisticated, faster sampling techniques that are based on *importance sampling or likelihood ratios*, which changes the original input distribution. Technical details can be found in Beckman and McKay (1987), Kleijnen and Rubinstein (1996), and Rubinstein and Shapiro (1993).

Note that importance sampling is also very useful (if not indispensable) whenever *rare events* must be simulated, such as nuclear accidents and buffer overflows in reliable telecommunication networks. See Helton et al. (1995, p.290), Kleijnen and Rubinstein (1995), Rubinstein and Shapiro (1993), and Sarkar and Rief (1995).

5.2 Uncertainty analysis of stochastic models

The type of question answered by UA is 'what is the chance of ...?' So the model must contain some random element. In §5.1 that randomness was limited to the inputs of the model, whereas the model itself was deterministic. However, as the Introduction (§1) mentioned, some models are intrinsically *stochastic*: without the randomness the problem disappears. Examples are queueing models, where the customer interarrival times may be independent drawings from an exponential distribution with parameter λ (so its mean is $1/\lambda$). This parameter is an input of the simulation queueing model. That model generates a stochastic time series of customer waiting times. The question may be: what is the probability of customers having to wait longer than 15 minutes? For simple queueing models this question can be answered analytically or numerically, but for more realistic models the analysts use simulation. Mathematical statistics is needed to determine how many customers must be simulated in order to estimate the response with prespecified accuracy; see Kleijnen and Van Groenendaal (1992, pp. 187-197).

Note that 'stochastic uncertainty is a property of the system being studied, while subjective uncertainty is a property of the analysts performing the study'; see Helton et al. (1995, p. 287).

How to apply UA to such a queueing simulation? Suppose the interarrival parameter λ is estimated from a sample of r independent interarrival times. Then the *central limit theorem* implies that the distribution of the estimated interarrival parameter $\hat{\Lambda}$ approximates a normal distribution. Hence the parameter value $\hat{\lambda}$ can be sampled from this distribution, and be used as input to the queueing simulation. That simulation is run for 'enough' customers. Next the procedure is repeated: sample $\hat{\lambda}$, and so on. For details see Kleijnen (1983).

Instead of relying on the central limit theorem, Cheng and Holland (1995) apply *bootstrapping*. However, the question remains *which* response to report to the users: the unconditional, ex post variance as do Cheng and Holland (1995) and also Brady and Hillestad (1995, p.30); the ex post variance, mean, and various quantiles, as Haverkort and Meeuwissen (1995) do; or the conditional moments (conditioned on the values of the estimated parameters)? Also see Kleijnen (1983).

In summary, UA has hardly been applied to stochastic models such as queueing models (SA has been employed in many simulation studies; see §3). Helton (1993, pp. 356-358) and Helton et al. (1991, 1992, 1995) discuss UA of stochastic models in the natural sciences (nuclear power plants, the spreading of nuclides). So UA of stochastic simulation models is an interesting area for further research.

6. Optimization: Response Surface Methodology

The *controllable inputs* should be steered (by the decision makers) into the right direction. For example, in the greenhouse case the government should restrict emissions of the gases concerned; in queueing problems, management may add more servers (such as check-out lanes at a supermarket). Strictly speaking, *optimization* means maximization under restrictions; the best-known example is linear programming. In this paper the term optimization is also used when restrictions are absent or ignored. There are many mathematical techniques for finding optimal values for the decision variables of nonlinear implicit functions (such functions may indeed be formulated by simulation models), possibly with stochastic noise; examples of such techniques are genetic algorithms, simulated annealing, and tabu search. However, this paper is limited to Response Surface Methodology (RSM). RSM combines regression analysis and DOE (see §3) with a hill-climbing technique called steepest ascent.

First consider four *general characteristics of RSM*; then some details:

(i) RSM relies on first-order and second-order polynomial regression metamodels, now

called *response surfaces* (see Equation 1 in §3.2).
(ii) It uses the *statistical designs* of DOE (see §3.3).
(iii) It is augmented with the mathematical (not statistical) technique of *steepest ascent*, to determine in which direction the decision variables should be changed.
(iv) It uses the mathematical technique of *canonical analysis* to analyze the shape of the optimal region: does that region have a unique maximum, a saddle point or a ridge?

Next consider some details. RSM begins by selecting a *starting point*. Because RSM is a heuristic (no success guaranteed!), several starting points may be tried later on, if time permits.

RSM explores the *neighborhood* of that point. Locally the response surface is approximated by a first-order polynomial in the decision variables (Taylor series expansion).

The main effects β_h (see Equation 1) are estimated, using a design with $n \approx k + 1$ (see §3.3.1). Suppose $\hat{\beta}_1 >> \hat{\beta}_2 > 0$. Then obviously the increase of decision variable 1 (say) z_1 should be larger than that of z_2. The *steepest ascent path* means $\Delta z_1/\Delta z_2 = \hat{\beta}_1/\hat{\beta}_2$ (no standardization; also see next paragraph).

Unfortunately, the steepest ascent technique does not quantify the *step size* along this path. Therefore the analysts may try a specific value for the step size. If that value yields a lower response, then this value should be reduced. Otherwise, one more step is taken. Ultimately, the response must decrease, since the first-order polynomial is only an approximation. Then the procedure is *repeated*: around the best point so far, a new first-order polynomial is estimated, after simulating $n \approx k + 1$ combinations of z_1 through z_k. And so on.

In the neighborhood of the top, a hyperplane can *not* be an adequate representation. To detect this lack of fit, the analysts may use cross-validation (see §3.2). Other diagnostic measures are R^2 (where R^2 denotes the multiple correlation coefficient), and modern statistics such as PRESS, discussed in Kleijnen (1987).

So when a hyperplane no longer approximates the local I/O behavior well enough, then a second-order polynomial is fitted (see §3.3.4).

Finally, the *optimal* values of z_h are found by straightforward differentiation of the fitted quadratic polynomial. A more sophisticated evaluation is *canonical analysis*.

Consider the following *case study*. A decision support system (DSS) for production planning in a steel tube factory is simulated and is to be optimized. There are fourteen decision variables, and two response variables (namely, a production and a commercial criterion); one response variable is maximized, whereas the other one forms a side-restriction. Simulation of one combination takes six hours of computer time, so searching for the optimal combination can not be performed using only common sense. Details can be found in Kleijnen (1993).

More applications can be found in Hood and Welch (1993), Kleijnen (1987, 1995d), and Kleijnen and Van Groenendaal (1992).

7. Conclusions

The problem addressed in this paper (see especially §1) is that there are *five related analysis types*: (i) sensitivity analysis (SA) or what-if analysis, (ii) uncertainty analysis (UA) or risk analysis, (iii) screening, (iv) validation, and (v) optimization. And the question is: *when* should *which* type of analysis be applied; which *statistical techniques* should be used?

This paper gave a *survey* of these issues, emphasizing *statistical* procedures. Such procedures yield reproducible, objective, quantitative results.

Sensitivity analysis determines which model inputs are really important. From the users' perspective, the important inputs are either controllable or not. The controllable inputs may be optimized. The values of the uncontrollable inputs may be well-known, in which case these values can be used for validation of the model. If the values of the uncontrollable inputs are not well known, then the likelihood of their values can be quantified objectively or subjectively, and the probability of specific output values can be quantified by *uncertainty analysis*.

More specifically, SA means that the model is subjected to *extreme value* testing. A model is valid only within its experimental frame (which was defined as the limited set of circumstances under which the real system is to be observed or experimented with). Mathematically that frame might be defined as the hypercube formed by the k standardized inputs x_{ih} of the model (more complicated definitions allow for restrictions such as linear conditions on the input combinations (say) $\sum_{h=1}^{k} x_{ih} = 1$). Experimental designs such as 2^{k-p} fractional factorials specify *which* combinations are actually observed or simulated; for example, a 2^{-p} fraction of the 2^k corner points of that hypercube. The n observed input combinations and their corresponding responses are analyzed through a regression (meta)model, which is an approximation of the simulation model's I/O behavior. That regression model quantifies the importance of the simulation inputs.

This paper proposed the following *five stages* in the analysis of a simulation model.

Stage 1. Obviously, *validation* is one of the first problems that must be solved in a simulation study. The availability of data on the real system determines the type of statistical technique to be used for validation. Regression analysis, however, may be applied, whether data are available or not, albeit through different regression models (see §4).

Stage 2. When the simulation study is still in its *pilot* phase, then very many inputs may be conceivably important. The really important inputs can be identified through Bettonvil and Kleijnen (1995)'s *sequential bifurcation*, which is a *screening* technique that is based on aggregation and sequential experimentation.

Stage 3. The important inputs found in stage 2 are investigated in a *more detailed sensitivity analysis*, including *interactions* between these inputs. This investigation may use *design of experiments* (DOE), which includes classical designs such as 2^{k-p} fractional factorial designs. Such designs give estimators of the effects in the *regression metamodel* that are better: minimum variance, unbiased linear estimators.

Stage 4. The important inputs should be split into two groups: inputs that are under the decision makers' control versus environmental inputs. Though the important *environmental inputs* cannot be controlled, information on the values they are likely to assume might be gathered. If the value of such an input is not precisely known, then the chances of various values can be quantified through a *probability function*. If a sample of data is available, then this function can be estimated objectively, applying mathematical statistics; otherwise subjective expert opinions are used. UA quantifies the uncertainties of the model outputs that result from the uncertainties in the model inputs. Output uncertainty is quantified through a statistical distribution. This analysis uses the Monte Carlo technique. This Monte Carlo experiment has smaller variance when applying *Latin hypercube sampling* (LHS). LHS is a variance reduction technique (VRT), not a screening technique.

SA does not tell *how likely* a particular combination of inputs (specified by a statistical design) is, whereas UA does account for the probabilities of input values.

Combining UA with regression analysis shows which non-controllable inputs contribute most to the uncertainty in the output.

UA of stochastic simulation models is an interesting area for further research.

Stage 5. The *controllable inputs* should be steered into the right direction. *Response Surface Methodology* (RSM) is a heuristic technique that combines DOE, regression analysis, and steepest ascent, in order to find the model inputs that give better model responses, possibly the best response.

Applications of the recommended techniques for these five stages are quite plentiful (see references).

An important conclusion is that SA should precede UA. Each type of analysis may apply its own set of statistical techniques, for example, 2^{k-p} fractional designs in SA, and LHS in UA. Some techniques may be applied in both analyses, for example, regression modelling. Hopefully, this paper succeeded in explaining *when* to use *which* technique! Yet, sensitivity and risk analyses remain *controversial* topics; communication within and among scientific disciplines is certainly needed.

References

Andres T.H. (1995), Sampling methods and sensitivity analysis for large parameter sets. SAMO95.

Balson W.E., Welsh J.L., and Wilson D.S. (1992), Using decision analysis and risk analysis to manage utility environmental risk. *Interfaces*, 22, no. 6, 126-139.

Bankes S. (1993), Exploratory modeling for policy analysis. *Operations Research,* 41, no.3, 435-449.

Bankes S. and Lucas T. (1995), Statistical approaches for the exploratory modeling of large complex models. SAMO95.

Banks J. (1989), Testing, understanding and validating complex simulation models, *Proceedings of the 1989 Winter Simulation Conference*.

Beckman R.J. and McKay M.D. (1987), Monte Carlo estimation under different distributions using the same simulation. *Technometrics*, 29, no. 2, 153-160.

Bedford T. and Meeuwissen A.M.H. (1995), The distribution of maximum entropy distributions given fixed rank correlation, and numerical approximations for use in sensitivity analyses. SAMO95

Bettonvil, B., *Detection of Important Factors by Sequential Bifurcation,* Tilburg University Press, Tilburg, 1990.

Bettonvil, B. and Kleijnen, J.P.C. (1995), Searching for important factors in simulation models with many factors: sequential bifurcation. Tilburg University.

Birge J.R. and Rosa C.H. (1995), Modeling investment uncertainty in the costs of global CO_2 emission policy. *European Journal of Operational Research* 83: 466-488.

Brady S.D. and Hillestad R.J. (1995), Modeling the external risks of airports for policy analysis. (MR-605-EAC/VW) RAND European-American Center for Policy Analysis, Delft (Netherlands).

Brehmer B., Eriksson E.A., and Wulff P. (1994), Risk management (feature issue), *European Journal of Operational Research* 75: 477-566.

Cheng R.C.H. and Holland W. (1995), The sensitivity of computer simulation experiments to errors in input data. SAMO95.

Cochran, J.K and Chang J. (1990), Optimization of multivariate simulation output models using a group screening method. *Computers Industrial Engineering*, 18, no. 1 , 95-103.

De Wit M.S. (1995), Uncertainty analysis in building thermal modelling. SAMO95.

Downing D.J., Gardner R.H., and Hoffman F.O. (1985), An examination of response-surface methodologies for uncertainty analysis in assessment models. *Technometrics*, 27, 151-163.

Downing D.J., Gardner R.H., and Hoffman F.O. (1986), Response to Robert G. Easterling. *Technometrics*, 28, no.1, 92-93.

Easterling, R.G. (1986), Discussion of Downing, Gardner, and Hoffman (1985). *Technometrics*, 28, no.1, 91-92.

Engemann K.J and Miller H.E. (1992), Operations risk management at a major bank. *Interfaces*, 22, no. 6, 140-149.

FIPS (1979), *Guidelines for automatic data processing risk analysis*. FIPS PUB 65 (Federal Information Processing Standards Publication), Washington.

Harbers A. (1993), *Onzekerheidsanalyse op een simulatiemodel voor het broeikaseffect.* (Uncertainty analysis of a simulation model for the greenhouse effect.) Tilburg University.

Haverkort B.R. and Meeuwissen A.M.H. (1995), Sensitivity and uncertainty analysis of Markov-reward models.

Helton J.C. (1993), Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering and System Safety*, 42, 327-367.

Helton J.C. (1995), Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. SAMO.

Helton J.C. et al. (1995), Effect of alternative conceptual models in a preliminary performance assessment for the waste isolation pilot plant. *Nuclear Engineering and Design*, 251-344.

Helton J.C., Garner J.W., McCurley R.D. and Rudeen D.K. (1991), *Sensitivity analysis techniques and results for performance assessment at the waste isolation pilot plant*. Sandia Report, SAND90-7103.

Helton J.C., Garner J.W., Rechard R.P., Rudeen D.K., and Swift P.N. (1992), *Preliminary comparison with 40 CFR part 191, subpart B for the waste isolation pilot plant, vol. 4: uncertainty and sensitivity analysis*. Sandia Report, SAND91-0893/4.

Hertz D.B. (1964), Risk analysis in capital investments. *Harvard Business Review*, 95-106.

Ho,Y. and X. Cao. 1991. *Perturbation analysis of discrete event systems*. Dordrecht: Kluwer.

Hora S.C. (1995), Sensitivity, uncertainty, and decision analyses in the prioritization of research. SAMO95

Iman, R.L. and Conover W.J. (1980), Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. (Including comments and rejoinder). *Communications in Statistics*, A9, no. 17, 1749-1874.

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

Janssen, P.H.M., P.S.C. Heuberger, and R. Sanders (1992), *UNCSAM 1.0: a software package for sensitivity and uncertainty analysis manual*. National Institute of Public Health and Environmental Protection (RIVM), Bilthoven, The Netherlands.

Karplus, W.J. (1983), The spectrum of mathematical models. *Perspectives in Computing*, no. 3, 4-13.

Kleijnen J.P.C. (1974/1975), *Statistical techniques in simulation, volumes I and II*. Marcel Dekker Inc., New York. (Russian translation, Publishing House "Statistics", Moscow, 1978.)

Kleijnen J.P.C. (1979) Regression metamodels for generalizing simulation results. *IEEE Transactions on Systems, Man, and Cybernetics*. SMC-9, no. 2, 93-96.

Kleijnen (1983) Risk analysis and sensitivity analysis: antithesis or synthesis? *Simuletter*, 14, no. 1-4, 64-72.

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

Kleijnen, J.P.C. (1993), Simulation and optimization in production planning: a case study, *Decision Support Systems* 9: 269-280.

Kleijnen J.P.C. (1994), Sensitivity analysis versus uncertainty analysis: when to use what? Predictability and nonlinear modelling in natural sciences and economics, edited by J Grasman and G. van Straten, Kluwer, Dordrecht, The Netherlands, 1994, pp.322-333. (Preprinted in Kwantitatieve Methoden, 15, no. 45, February 1994, pp. 3-15.)

Kleijnen J.P.C. (1995a), Verification and validation of simulation models. *European Journal of Operational Research* 82: 145-162.

Kleijnen J.P.C. (1995b), Sensitivity Analysis and Optimization in Simulation: Design of Experiments and Case Studies. *Proceedings of the 1995 Winter Simulation Conference* (edited by C. Alexopoulos, K. Kang, W.R. Lilegdon, D. Goldsman).

Kleijnen, J.P.C. (1995c), Case-study: statistical validation of simulation models. *European Journal of Operational Research* (in press).

Kleijnen, J.P.C. (1995d), Sensitivity analysis and optimization of system dynamics models: regression analysis and statistical design of experiments. *System Dynamics Review* (accepted).

Kleijnen and Rubinstein (1996) *European Journal of Operational Research* (forthcoming).

Kleijnen J.P.C. and W. Van Groenendaal (1992), *Simulation: a Statistical Perspective*. Wiley, Chichester.

Kleijnen, J.P.C., G. van Ham and J. Rotmans (1990), Techniques for sensitivity analysis of simulation models: a case study of the CO2 greenhouse effect. *Simulation*, 58, no. 6, June 1992, pp. 410-417.

Kraan, B.C.P. and Cooke R.M. (1995) Joint CEC/USNRC accident consequence code uncertainty analysis using expert judgement. *SAMO95*

Krumm, F.V. and Rolle, C.F. (1992), Management and application of decision and risk analysis in Du Pont. *Interfaces*, 22, no. 6, 84-93.

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

McKay M.D. (1992), Latin hypercube sampling as a tool in uncertainty analysis of computer models. *1992 Winter Simulation Conference Proceedings*, Association for Computing Machinery, New York.

McKay M.D. (1995), Evaluating prediction uncertainty. Los Alamos National Laboratory, NUREG/CR-6311 (LA-12915-MS).

Meeuwissen A.M.H. and Cooke R.M. (1994), Tree dependent random variables. Technical University Delft, 1994.

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.

Olivi L. (1980), *Response surface methodology in risk analysis; synthesis and analysis methods for safety and reliability studies.* Edited by G. Apostolakis, S. Garibra and G. Volta, Plenum Publishing Corporation, New York.

Ő̋ren, T.I. 1993. Three simulation experimentation environments: SIMAD, SIMGEST, and E/SLAM. In *Proceedings of the 1993 European Simulation Symposium*. La Jolla: Society for Computer Simulation.

Pacheco, N.S. (1988), Session III: simulation certification, verification and validation, *SDI Testing: the Road to Success; 1988 Symposium Proceedings International Test & Evaluation Association*, ITEA, Fairfax (Virginia 22033).

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

Rahni N., Ramdani N., Candau Y., and Dalicieux P. (1995), Sensitivity analysis of dynamic buildings energy simulation models using group screening and sampling methods. SAMO95.

Rao G.P. and Sarkar P.K. (1995), Sensitivity studies of air scattered neutron dose from particle accelerators. SAMO95.

Reilly J.M., Edmonds J.A., Gardner R.H., and Brenkert A.L. (1987), Uncertainty analysis of the IEA/ORAU CO_2 emissions model. *The Energy Journal*, 8, no. 3, 1-29.

Reilly T. (1995), The effect of correlated variables on a one-way sensitivity analysis. SAMO95.

Rotmans, J. (1990), *IMAGE: An Integrated Model to Assess the Greenhouse Effect*. Kluwer, Dordrecht, Netherlands.

Rubinstein, R.Y. and A. Shapiro (1993). *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization via the Score Function Method*, Wiley, New York.

Sacks J., Welch W.J., Mitchell T.J., and Wynn H.R. (1989), Design and analysis of computer experiments (includes comments and rejoinder). *Statistical Science*, vol. 4, no. 4, 409-431.

Saltelli A. (1995), On the use of rank transformation in sensitivity analysis. SAMO95.

Saltelli A., Andres T.H., and Homma T. (1993), Sensitivity analysis of model output: an investigation of new techniques. *Computational Statistics and Data Analysis*, 15, 211-238.

Saltelli A. and Homma T. (1992), Sensitivity analysis for model output: performance of black box techniques on three international benchmark exercises. *Computational Statistics and Data Analysis*, 13, 73-94.

Sarkar P.K. and Rief H. (1995), Differential operator sampling for self-optimizing Monte Carlo simulations. SAMO95.

Sobol I.M. (1995), Sensitivity analysis of nonlinear models using sensitivity indices. SAMO95.

Van Asselt M. and Rotmans J. (1995), Uncertainty in integrated assessment modelling: a cultural perspective-based approach. RIVM, Bilthoven, Netherlands.

Van Groenendaal W. (1994), *Investment analysis and DSS for gas transmission on Java*. Tilburg (Netherlands): Tilburg University.

Van Groenendaal W. and Kleijnen (1995), ... SAMO

Van Meel, J. 1994. *The dynamics of business engineering*. Delft (Netherlands): Delft University.

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

Zeigler, B. (1976), *Theory of Modelling and Simulation.* Wiley Interscience, New York.

AUTHOR BIOGRAPHY

JACK P.C. KLEIJNEN is Professor of Simulation and Information Systems in the Department of Information Systems and Auditing; he is also associated with the Center for Economic Research (CentER). Both the Department and the Center are within the School of Management and Economics of Tilburg University (Katholieke Universiteit Brabant) in Tilburg, Netherlands. He received his Ph.D. in Management Science at Tilburg University. His research interests are in simulation, mathematical statistics, information systems, and logistics. He published six books and more than 130 articles; lectured at numerous conferences throughout Europe, the USA, Turkey, and Israel; he was a consultant for various organizations; and is a member of several editorial boards. He spent some years in the USA, at different universities and companies. He was awarded a number of fellowships, both nationally and internationally.