SENSITIVITY ANALYSIS VERSUS UNCERTAINTY ANALYSIS: WHEN TO USE WHAT?

JACK P.C. KLEIJNEN

CentER and Department of Information Systems and Auditing, Tilburg University (Katholieke Universiteit Brabant), 5000 LE Tilburg, The Netherlands[1]

Abstract
Model validity is of major interest to decision makers and other users of models. From their viewpoint the important model inputs should be split into two groups, namely inputs that are under the decision makers' control and (environmental) inputs that 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? Among the techniques to answer these questions are statistical design of experiments (such as fractional factorial designs) and regression analysis. Controllable inputs can be optimized through Response Surface Methodology (RSM). Sensitivity analysis may further show that some non-controllable inputs of the model are important; yet the precise values of these inputs may not be known. Then risk or uncertainty analysis becomes relevant. Its techniques are Monte Carlo sampling, including variance reduction techniques such as Latin hypercube sampling, possibly combined with regression analysis. A bibliography with 35 references is included.

Keywords: sensitivity analysis, uncertainty analysis, risk analysis, validation, experimental design, regression, screening, Latin hypercube sampling, optimization.

1. Introduction

The *analysis methods* discussed in this paper are known in the literature under such names as sensitivity, what-if, perturbation, risk, and uncertainty analyses. Definitions of these terms vary; for example, Helton, Garner, McCurley, and Rudeen (1991) use a definition of 'sensitivity analysis' that differs substantially from the one used in this report. We define the keyterms in the title of this paper as follows.

Sensitivity analysis or *what-if analysis* is the systematic investigation of the reaction of model outputs to *extreme* values of the model inputs and to drastic changes of the model structure. For example, how does the average waiting time in a queueing model change when the arrival rate doubles; what if the priority rule changes from first-in-first-out (FIFO) to last-in-first-out (LIFO)? So this analysis examines *global*, not local (marginal) sensitivities.

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

[1] Prof. J.P.C. Kleijnen, Vakgroep Bestuurlijke Informatiekunde en Accountancy (BIKA), Faculteit Economische Wetenschappen, Postbus 90153, 5000 LE Tilburg; telefoon 013-662029; e-mail: Kleijnen@kub.nl

model outputs. So the input values range between the extreme values investigated in sensitivity analysis. The goal of uncertainty analysis is to quantify the *probability* of specific output values, whereas sensitivity analysis does not tell how likely a specific result is. The differences between sensitivity analysis and uncertainty analysis will be further explored later on.

Each type of analysis may apply its own set of *statistical techniques*. For example, sensitivity analysis may use 2^{K-P} designs, whereas uncertainty analysis applies either crude Monte Carlo sampling or variance reduction techniques such as Latin hypercube sampling. Some techniques are applied in both analyses; for example, regression modelling. We assume that the reader is familiar with the basics of these techniques, so we do not discuss their technicalities but only their role in both types of analysis.

The *issues* to be solved by sensitivity and uncertainty analyses are discussed under such headings as *validation* and *optimization*. These issues are studied in all scientific disciplines that use mathematical models. Unfortunately, nobody can be an expert in all these disciplines. This paper is coloured by more than 25 years of experience with the technique of simulation, especially its statistical aspects and its application to problems in business, environmental, agricultural, military, and computer systems; see Kleijnen and Van Groenendaal (1992).

Not only is there a variety of related methods and issues, there is also much *software*. This software greatly simplifies the implementation of these methods. We shall refer to software throughout this paper.

All scientific methods are based on *assumptions*, which limit the applicability of these methods. These assumptions may be documented explicitly or they may be left implicit. Many practitioners do not know *when* to use *what* method. The goal of this paper is to explain which questions may be asked in practice, and which methods can answer these questions.

These questions are also discussed in the literature: see Bankes (1993), Downing, Gardner, and Hoffman (1985, 1986), Easterling (1986), Iman and Conover (1980), and McKay (1992). This paper, however, is not a recapitulation of those publications: sensitivity and risk analyses remain *controversial* topics. For example, in §3.1 we shall claim that Latin hypercube sampling should not be applied as a screening technique. Controversies were also observed at the workshop on 'Uncertainty analysis' organized in 1989 by the Dutch 'National Institute of Public Health and Environmental Protection'(abbreviated in Dutch to RIVM), and at the conference on 'Predictability and nonlinear modelling in natural sciences and economics', organized by Wageningen Agricultural University in 1993; see Grasman and Van Straten (1994).

The outline of this paper is as follows. In §2 we discuss model validation and what-if analysis of controllable and non-controllable inputs. This includes sensitivity analysis using statistical design of experiments with its concomitant regression analysis. This sensitivity analysis estimates which inputs are important. If these inputs are controllable, then they may be optimized; otherwise uncertainty analysis may be applied. In §3 we address this uncertainty analysis, which encompasses 1) the basics of that analysis, including applications in economics and the natural sciences, and 2) uncertainty analysis of stochastic simulation models. In §4 we give conclusions. A list of 35 references concludes this paper.

2. Model Validation and What-if Questions

Validation addresses the question: is the conceptual model an accurate representation of the real-life system under study (either an existing or a planned system)? Our discussion is based on Kleijnen (1993), which surveys the validation of models, especially simulation models.

There is no standard theory on validation, neither is there a standard 'box of tools'. We emphasize *statistical techniques*, which may yield reproducible, objective, quantitative data about the quality of models (other techniques -such as computer animation- are discussed in Kleijnen 1993). Data on inputs and outputs of the real system may be available in different quantities. We use this data availability to classify validation techniques, as we shall see.

2.1 Obtaining real data

To obtain a valid model, the analysts should try to measure the real system's input, output, and intermediate variables.

1) Sometimes it is *difficult or impossible* to obtain relevant data. For example, in simulation studies of nuclear war, it is (fortunately) impossible to get the necessary data. Further, by definition, there are no data on future situations; however, there may be data on past situations that may be extrapolated (also see below).

2) Usually, however, it is possible to get *some* data. Typically the analysts have data on the existing system variant. For instance, for the existing manufacturing system the current scheduling rule is well known. And in an ecological model there may be data on the existing situation ('status quo' scenario).

3) In the military it is common to conduct *field tests* in order to obtain data on *future* variants. For example, real mine fields are created, not by the enemy but by the friendly navy; next a mine hunt is executed in this field to collect data.

4) In some applications there is an *overload* of data, namely if these data are collected electronically. For instance, all transactions are recorded at the supermarket's point-of-sale (POS) systems.

The further the analysts go back into the past, the more data they get and the more powerful the validation test will be, *unless* they go so far back that different laws governed the system. For example, many econometric models do not use data prior to 1945, because the economic infrastructure changed drastically during World War II. (Also see 1) above.)

Moreover the data may show *observation error*, which complicates the comparison of real and simulated time series. Observation errors are discussed for a theoretical and a practical situation respectively in Barlas (1989, p. 72) and Kleijnen and Alink (1992).

2.2 Simple techniques for comparing model and real data

Suppose the analysts have succeeded in obtaining at least some data on the real, existing system (if not, the sensitivity analysis of §2.3 can be used). They should then feed these input data into the model, in *historical* order. After running the simulation program with these input data, the analysts obtain a time series of simulation outputs. Those data should be compared with the historical outputs of the existing system. That comparison may use simple techniques; for example, the familiar t statistic may be used to test whether the expected values of the simulated and the real time series are equal; see Kleijnen (1993).

Often simulation is meant to predict *relative responses*, not absolute responses; for example, what is the effect of adding one server to a queueing system? The analysts may then test whether simulated and real responses are positively correlated (without having the same means). This correlation can be estimated and tested through elementary regression analysis; see Kleijnen (1993).

Sometimes, however, simulation is meant to predict *absolute responses*; for example, in mine hunting the question may be whether the probability of detecting mines is so high that it makes sense to do a mine sweep (see Kleijnen and Alink 1992). Another example is: do the costs of a certain scenario outweigh the benefits of that same scenario? The analysts may then combine the test on means with the test on correlation; see Kleijnen (1993).

As we saw (§2.1), there may be very many observations. Then not only the means of the simulated and the real time series and their (cross)correlation can be compared, but also their autocorrelations with lag 1, 2, etc. A sophisticated technique to estimate the autocorrelation structure of the simulated and the historical time series respectively, and to compare these two structures is *spectral analysis*. Unfortunately, that analysis is rather difficult (and -as stated- requires long time series).

2.3 Sensitivity or what-if analysis

Models and submodels (modules) with *unobservable* inputs and outputs can *not* be subjected to the tests of the preceding subsection. The analysts should then apply sensitivity analysis, in order to determine whether the model's behaviour agrees with the judgments of the experts (users and analysts). In case of observable inputs and outputs, it is also useful to apply sensitivity analysis. We shall elaborate these statements in this section.

Sensitivity analysis can be based on the *statistical design of experiments*. Most practitioners apply an inferior design: they change *one input at a time*. This design gives estimated effects of input changes that have relatively high variances. Moreover, such a design cannot estimate interactions among inputs.

Factorial designs change *several* inputs (or factors) simultaneously. For example, a 2^K design consists of all 2^K combinations of K inputs with each input at two values. So if there are three inputs (K = 3), then eight combinations are simulated. Unfortunately, high values of K require too many combinations and hence too much computer simulation time.

Therefore *fractional factorial* designs consider only a fraction of all combinations. For instance, 2^{3-1} designs take only half (2^{-1}) of all 2^3 combinations. In general, 2^{K-P} designs consider only a 2^{-P} fraction of all 2^K combinations. So a 2^{8-2} design includes 64 ($=2^6$) of the 256 ($=2^8$) combinations of the eight inputs; these 64 combinations permit the estimation of interactions between inputs.

There are also designs that consider more than two values per input. For example, *central composite* designs use five values (in order to estimate quadratic effects, to which we shall return).

Tables and software help to decide *how many* input combinations to investigate and *which* combinations to select. Software is advertised in, for example, *OR/MS Today*. Technicalities of experimental designs and their application to a variety of simulation problems are discussed in Kleijnen (1987, pp. 257-407); also see Kleijnen and Van Groenendaal (1992, pp. 168-179).

How can the results of such experiments with simulation models be analyzed; how can these results be used for interpolation and extrapolation? Practitioners often *plot* the model output (say) y versus the input x_k, one plot for each input k with k = 1,..., K. More refined plots are conceivable, for instance, superimposed plots. This practice can be formalized through *regression analysis* (which in experimental design is also known as

Analysis of Variance -or ANOVA- with 'fixed effects'; if inputs are sampled, then ANOVA with random effects may apply; see Kleijnen 1987, pp. 285-293). To explain regression analysis of simulation experiments we first consider some simple situations.

Let y denote *the response* of the simulation model; for example, the average waiting time per day. Then y_i denotes the simulated response in combination i of the K model inputs, with i = 1,..., n where n denotes the total number of simulated combinations; for example, $n = 2^{K-P}$.

If the model is *deterministic*, then each combination is run only once. Deterministic models are abundant in the natural sciences, because the classic laws of nature are deterministic. Many financial spreadsheet models are also deterministic. However, if the model is *stochastic*, then combination i should be run (say) m_i times with $m_i \geq 2$, in order to account for stochastic noise. An example is a queueing simulation with m_i days simulated, each day resulting in one average waiting time; then y_i denotes the average of these m_i average waiting times per day. For details see Kleijnen and Van Groenendaal (1992).

Next we consider the model inputs. Let L_k denote the lowest value of the input (say) z_k in the experiment; analogously H_k denotes its highest value. Then $a_k = (H_k - L_k)/2$ measures the dispersion of input k, whereas $b_k = (H_k + L_k)/2$ quantifies the location of input k. Hence $x_{ik} = (z_{ik} - b_k)/a_k$ denotes the *standardized input* k in combination i. See Kleijnen and Van Groenendaal (1992, pp. 177-179, 183-185), and also Downing et al. (1985. p.156) and Helton et al. (1992, chapter 6, p. 4).

Inputs may be *qualitative*. Examples are priority rules in a queueing simulation and scenarios in environmental simulation. Technically, each qualitative input requires one or more binary variables (see Kleijnen 1987).

Let β_k denote the main or *first order effect* of the standardized input k; it measures how much the response changes as the original input changes from its lowest to its highest value, ignoring high order effects of inputs. Let $\beta_{kk'}$ designate the (two-factor) *interaction* between the inputs k and k'. We ignore interactions among three or more inputs, because they are hard to interpret. At this stage of our exposition we also ignore quadratic effects β_{kk}, which measure curvature: in sensitivity analysis we are interested in main effects only. We do not want the estimators of main effects to be biased by interactions. Quadratic effects -if present- bias the overall effect or grand mean β_0, which is of no interest in sensitivity analysis (β_0 is important when predicting the simulation response through a regression model). There are designs that require only 2K combinations to obtain unbiased estimators of the K main effects, in the presence of (K(K - 1)/2) two-factor interactions; these designs are called 'resolution 4' designs; see Kleijnen (1987, p. 301).

Finally, let e_i represent the *fitting error* in combination i when approximating the input/output behaviour of the simulation model by the simple regression (meta)model or response surface

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \sum_{k=1}^{K-1} \sum_{k'=k+1}^{K} \beta_{kk'} x_{ik} x_{ik'} + e_i. \quad (1)$$

Note that y and x_k may also denote the *ranks* of the original variables. This rank regression is popular in risk analysis. Saltelli and Homma (1992, pp. 79-82) give an elementary survey.

The *ordinary least squares* (OLS) criterion applied to the simulation input/output data (**y**, **X**) gives estimated effects $\hat{\beta}$. Weighted least squares (WLS) can be applied in case of stochastic simulation: the variance of input combination i can be estimated from the m_i replications; we assume that the variances of different input combinations are not constant. Generalized least squares (GLS) can be applied if common pseudorandom number seeds are used, which generate positive correlations among simulation responses of different input combinations (these correlations reduce the variances of the sensitivity estimates). We shall focus on OLS.

Of course, the *validity* of the resulting approximation ($\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$) must be tested. The simplest measure is the well-known R^2 coefficient: the closer R^2 is to one, the better; see Kleijnen (1987, p.193). *Cross-validation* is more in the spirit of validation of models in general. First it deletes simulation input $\mathbf{x}_i$ and the concomitant output data y_i. Next it estimates the regression parameters from the remaining data, which yields $\hat{\boldsymbol{\beta}}_{(-i)}$. Then it employs the resulting estimated regression model to compute the forecast $\hat{y}_i$ for the (deleted) input combination $\mathbf{x}_i$. The forecasted output $\hat{y}_i$ is compared with the simulated output y_i to validate the approximation. This procedure can be repeated for all n input combinations. The implementation of cross-validation is simple, since modern regression software provides statistics known as PRESS, DEFITS, DFBETAS, and Cook's D. See Kleijnen and Van Groenendaal (1992, pp. 156-157).

A *case study* illustrating the application of experimental design and regression analysis is provided by Kleijnen, Rotmans, and Van Ham (1990). They apply these techniques to several modules of IMAGE, which is a deterministic simulation model developed at RIVM for the greenhouse effect of carbon dioxide (CO_2) and other gases. This approach gives estimates of the effects of the various inputs. These estimated effects should have the right *signs*: the policy analysts (not the statisticians) know that certain inputs increase the global temperature. Wrong signs indicate computer errors or conceptual errors. Indeed Kleijnen et al. (1990) give examples of estimated sensitivity estimates with the wrong signs, which lead to correction of the simulation model. The remaining estimated effects show which inputs are important. We shall return to this case study.

One more example is the case study in Kleijnen and Alink (1992), concerning mine hunting at sea by means of sonar. The role of experimental design in the validation of simulation models is also discussed in Pacheco (1988). Regression analysis and their application to simulation in radioactive waste disposal is examined in Helton, Garner, Rechard, Rudeen, and Swift (1992). Alternative techniques are reviewed in these two publications and in Downing et al. (1985), Helton et al. (1991, chapter II), and McKay (1992). Also see Kleijnen (1987, pp. 241-242) and Kleijnen and Van Groenendaal (1992, pp. 147-186).

Classic experimental designs, however, may require too much computer time, when the simulation study is still in its early (pilot) phase and *very many inputs* may be conceivably important. Bettonvil and Kleijnen (1991) present a *screening* technique, called sequential bifurcation. They proceed sequentially (or stagewise) and split up (or bifurcate) the aggregated inputs as the experiment proceeds, until finally the important individual inputs are identified and their effects are estimated. They applied this technique to the RIVM model with 281 inputs, and found the 15 most important inputs after only 144 runs. It is remarkable that the statistical technique identified some inputs that were originally thought to be unimportant by the policy analysts.

Before executing the experimental design, the analysts must determine the experimental domain or experimental frame (the design tells *how* to explore this domain, using the expertise of the statistician). Zeigler (1976, p. 30) defines the *experimental frame* as 'a limited set of circumstances under which the real system is to be observed or experimented with'. He emphasizes that 'a model may be valid in one experimental frame but invalid in another'. We have already mentioned (§2.1) that going far back into the past may yield historical data that are not representative of the current system; that is, the old system was ruled by different laws. Similarly, a simulation model is valid only if its input data remain within a certain area. For example, Bettonvil and Kleijnen's (1991) screening study shows that the greenhouse simulation is valid, only if its input values range over a relatively small area; outside that area the resulting simulation responses (CO_2) had magnitudes that were immediately declared unrealistic by the experts. In general, it is difficult to develop valid models for completely new situations!

Mathematically the experimental frame may be defined as the hypercube formed by the K standardized inputs x_{ik} of the model. In practice, some corner points (combinations of extreme values) may be unrealistic, that is, they fall outside the experimental frame; see Janssen, Heuberger, and Sanders (1991, chapter 5) and Kleijnen (1987, pp. 318-319).

So any simulation model is valid only for a certain area of its inputs. Within that area the (valid) simulation model's input/output behaviour may vary. For example, a *first* order regression model (see equation 1 with the double summation term eliminated) is a good approximation of the input/output behaviour of a queueing simulation model, only if the traffic load is 'low'. When traffic is heavy, a second order regression model (which includes curvature) or a logarithmic transformation may apply. Some researchers fit a (meta)model to the simulation input/output data that holds over the whole experimental area. For example, Sacks, Welch, Mitchell, and Wynn (1989) apply covariance-stationary processes or Kriging to approximate deterministic simulation models. Barton (1992) surveys many new alternatives to the polynomial metamodel presented in (1). These approaches are so new that definitive evaluations cannot be given yet. Moreover, they may aim at fast and accurate prediction (interpolation), not at sensitivity analysis.

The *magnitudes* of the estimates of the effects β show which inputs are *important*. Since the regression model is only an approximation to the simulation model, false conclusion are possible. For example, an input might have an unimportant first order effect but an important quadratic effect in (1), given a certain experimental area: non-monotonic reaction of simulation response to simulation input. Cross-validation might fail to reject the regression metamodel. We consider this example to be 'pathological'. Our approach -like any other approach with finite sample sizes- can not guarantee correct conclusions. Also see Saltelli, Andres, and Homma (1993) 's comparison of the performances of different sensitivity analysis techniques.

Mathematically all inputs x_k are treated the same. However, general systems theory (GST) distinguishes (i) inputs that are under the decision makers' control, and (ii) 'environmental' inputs, which (by definition) are not controllable.

The *controllable inputs* should be steered -by the decision makers- into the right direction. For example, in the greenhouse case the governments should restrict emissions of the gases concerned. There are several *optimization* techniques for simulation models. These models may have multiple responses that are nonlinear, possibly stochastic, complicated functions of their inputs. *Response Surface Methodology* (RSM) is a heuristic sequential technique that combines experimental design (especially 2^{K-P} and central composite designs), regression analysis, and steepest ascent, in order to find the model inputs that give a better (possibly the maximum) model response, in terms of a specific performance criterion; see Kleijnen and Van Groenendaal (1992, pp. 181-185). This reference also summarizes an RSM case study in steel tube manufacturing, namely a production planning system with 14 controllable inputs and several response types.

For the important *environmental inputs* the analysts should try to collect data on the (input) values that occur in practice (and apply the validation techniques of the preceding subsection). If they do not succeed in getting accurate information, then they may use the uncertainty analysis of the next section.

3. Uncertainty Analysis

The analysts may be unable to collect reliable data on the important environmental inputs, that is, the values that may occur in practice are uncertain. Then *uncertainty analysis* or *risk analysis* may be applied.

3.1 The basics of uncertainty analysis

First the analysts derive a *probability function* for the input values. That distribution may be estimated from sample data, if those data are available; otherwise that distribution must be based on subjective expert opinions (also see Helton et al. 1992, chapter 2, p. 4). Popular distributions are uniform, triangular, beta, normal, and lognormal distributions. Usually the inputs are assumed to be statistically independent. Correlated inputs are discussed in Helton et al. (1992, chapter 3, p. 7) and Reilly, Edmonds, Gardner, and Brenkert (1987).

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

Variance reduction techniques (VRTs) are possible. Uncertainty analysis often uses *Latin hypercube sampling* (LHS), which forces the sample (of size n) to cover the whole experimental area (for example, in case of a single input, that input's domain is partitioned into, say, s equally likely subintervals and each subinterval is sampled s/n times). We recommend LHS as a VRT, not as a screening technique. For screening we recommend changing the inputs to their extreme values (specified by a fractional factorial design) and testing if at those values the outputs also change. Also see Banks (1989) and Bettonvil and Kleijnen (1991) versus Downing et al. (1985) and McKay (1992).

The sampled input values are fed into the simulation model. In this subsection we consider *deterministic* simulation models; in §3.2 we shall examine stochastic models. *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, different constants or different growth percentages. This yields 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 quantiles or percentiles, such as the 90% quantile). For a basic introduction to risk analysis see Kleijnen and Van Groenendaal (1992, pp. 75-78).

Which quantities sufficiently summarize a distribution function, depends on the users' *risk attitude*: risk neutral (then 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).

Combining uncertainty analysis 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 eq. (1) the deterministic independent variables x_{ik} are replaced by random variables. Helton et al. (1991, 1992) call this combination of uncertainty and regression analysis 'sensitivity analysis'.

Risk analysis is used 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. Risk analysis 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).

In the *natural sciences*, uncertainty analysis is also popular. For example, in the USA the Sandia National Laboratories performed many uncertainty analyses for nuclear waste disposal (Helton et al. 1991, 1992). Oak Ridge National Laboratory investigated radioactive doses absorbed by humans (Downing et al. 1985). Nuclear reactor safety was investigated for the Commission of the European Communities (Olivi 1980 and Saltelli and Homma 1992). Uncertainty analysis is also performed at RIVM (Harbers 1993, Janssen et al. 1992). Three environmental studies for the electric utility industry are presented in Balson et al. (1992). Uncertainty analysis 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 risk analysis is also used in the analysis of computer security; see Engemann and Miller (1992) and FIPS (1979).

As we saw, a basic characteristic of uncertainty analysis is that information about the inputs of the simulation model is not reliable; so 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). This specification may be *software driven*; 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 sensitivity analysis may be called *robustness analysis*. Examples can be found in Helton et al. (1992, section 4.6); also see Janssen et al. (1992) and Kleijnen (1987, pp. 144-145). Faster sampling techniques for robustness analysis are discussed by Beckman and McKay (1987) and Rubinstein and Shapiro (1993).

3.2 Uncertainty analysis of stochastic models

The type of question answered by uncertainty analysis is 'what is the chance of...?' So the model must contain some random element. In §3.1 that randomness was limited to the inputs of the model, whereas the model itself was deterministic.

However, some models are intrinsically *stochastic*: without the randomness the problem disappears. Examples are queueing models, where the interarrival times may be independent drawings from an exponential distribution with mean (say) λ. This mean is an input of the 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 models this question can be answered analytically or numerically. For more realistic models, simulation is used. Mathematical statistics is used 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).

How to apply risk analysis to such a queueing simulation? Suppose the interarrival mean λ is estimated from a sample of r independent interarrival times. Then the central limit theorem implies that the estimated interarrival mean follows a normal distribution. Hence the mean 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 λ, and so on. For details see Kleijnen (1983).

However, to the best of our knowledge, risk analysis has never been applied to stochastic models such as queueing models (sensitivity analysis has been employed indeed; see §2.3). Helton et al. (1991, 1992) briefly discuss uncertainty analysis of stochastic models in the natural sciences (nuclear power plants, the spreading of nuclides).

4. Conclusions

We gave a *survey* of sensitivity analysis and uncertainty analysis of mathematical models, emphasizing *statistical* procedures, which yield reproducible, objective, quantitative results.

Sensitivity analysis determines which model inputs are really important. 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, sensitivity analysis means that the model is subjected to 'extreme value' testing. A model is valid only within its experimental frame (defined in §2.3 as the limited set of circumstances under which the real system is to be observed or experimented with). Mathematically that frame may be defined as the hypercube formed by the K standardized inputs x_{ik} of the model. Experimental designs (such as 2^{K-P} fractional factorials) specify *which* combinations are actually observed or simulated (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. That regression model is an approximation of the simulation model's input/output behaviour. That regression model gives quantitative measures of the importance of the simulation inputs.

We proposed several steps in sensitivity analysis:

1) When the simulation study is still in its early (pilot) phase, then very many inputs may be conceivably important. The really important inputs can be identified through Bettonvil and Kleijnen (1991)'s *screening* technique based on aggregation and sequential experimentation.

2) These important inputs are further investigated, including *interactions* between these inputs. This estimation can use classical factorial designs such as 2^{K-P} designs.

3) The inputs should be split into two groups: inputs that are under the decision makers' control versus environmental inputs. The *controllable inputs* should be steered into the right direction. Response Surface Methodology (RSM) is a heuristic technique that combines experimental designs (including central composite designs), regression analysis, and steepest ascent, in order to find the model inputs that give better model responses, possibly the best response.

The important *environmental inputs* cannot be controlled (by definition), but information on the values they are likely to assume is wanted. 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. *Uncertainty analysis* 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.

Note that sensitivity analysis does not tell *how likely* a particular combination of the inputs (specified by the statistical designs) is, whereas uncertainty analysis does account for the probabilities of input values.

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

Our conclusion is that sensitivity analysis should precede uncertainty analysis. Each type of analysis may apply its own set of statistical techniques, for example, 2^{K-P} designs in sensitivity analysis; Latin hypercube sampling in uncertainty analysis. Some techniques are applied in both analyses, for example, regression modelling. We hope that we succeeded in explaining *when* to use *which* technique!

References

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.

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

Barlas Y. (1989), Multiple tests for validation of system dynamics type of simulation models, *European Journal of Operational Research*, 42, no. 1, pp. 59-87.

Barton R. (1992), Metamodels for simulation input-output relations. *1992 Winter Simulation Conference Proceedings*, Association for Computing Machinery, New York.

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

Bettonvil, B. and Kleijnen, J.P.C. (1991), Identifying the important factors in simulation models with many factors. Tilburg University.

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.

Grasman J. and Van Straten G. (1994), editors, *Predictability and Nonlinear Modelling in Natural Sciences and Economics*. Wiley, Chichester.

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

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), *Prelimi-*

nary 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.

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.

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.

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), Verification and validation of simulation models. *European Journal of Operational Research* (submitted).

Kleijnen, J.P.C. and Alink, G.A. (1992), Validation of simulation models: Mine-hunting case study. Tilburg University.

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

Kleijnen, J.P.C., Rotmans, J. and Van Ham, G. (1990), Gevoeligheidsanalyse via metamodellen en proefopzetten, toegepast op het broeikassimulatiemodel IMAGE. (Sensitivity analysis through metamodels and experimental designs, applied to the greenhouse simulation model IMAGE). *Kwantitatieve Methoden*, 11, no. 35, pp. 51-72, 83-86.

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.

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.

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.

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).

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.

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., 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.

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

Acknowledgement
This paper is a prepublication of a contribution to the proceedings of the conference titled 'Predictability and nonlinear modeling in natural sciences and economics, 75th Anniversary Conference of the Wageningen Agricultural University', organized in Wageningen, The Netherlands, on April 5-7, 1993, and to be published by Wiley, Chichester, United Kingdom; see Grasman and Van Straten (1994). I also gratefully acknowledge discussions I had with professor Y. Merkuryev, Department of Modelling and Simulation, Riga Technical University, Latvia, while he was visiting Tilburg University, financed by CentER.

Ontvangen: 16-4-1993

Geaccepteerd: 17-11-1993