Proceedings of the 1996 Winter Simulation Conference
ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain

FIVE-STAGE PROCEDURE FOR THE EVALUATION OF SIMULATION MODELS THROUGH STATISTICAL TECHNIQUES

Jack P.C. Kleijnen

Department of Information Systems and Auditing(BIKA)/Center for Economic Research (CentER)
School of Management and Economics (FEW)
Tilburg University (KUB)
5000 LE Tilburg, NETHERLANDS

ABSTRACT

This paper recommends the following sequence for the evaluation of simulation models. 1) Validation: the availability of data on the real system determines the proper type of statistical technique. 2) Screening: in the simulation's pilot phase the important inputs are identified through a novel technique, namely sequential bifurcation, which uses aggregation and sequential experimentation. 3) Sensitivity or what-if analysis: the important inputs are analyzed in more detail, including interactions between inputs; relevant techniques are design of experiments (DOE) and regression analysis. 4) Uncertainty or risk analysis: important environmental inputs may have values not precisely known, so the resulting uncertainties in the model outputs are quantified; techniques are Monte Carlo and Latin hypercube sampling. 5) Optimization: policy variables may be controlled, applying Response Surface Methodology (RSM), which combines DOE, regression analysis, and steepest-ascent hill-climbing. This paper summarizes case studies for each stage.

1 INTRODUCTION

The main message of this paper is that most simulation studies should go through the following *five stages* and apply the following statistical *techniques* (which have already been applied in many practical simulations).

Stage 1: Validation and verification (V & V)
The analysts may use design of experiments or DOE (for example, a fractional factorial design) if there are no data on the input/output (I/O) of the simulation model. However, if there are enough data, then the simulation model may be validated by a special type of regression analysis (see §4).

Stage 2: Screening: sequential bifurcation (SB)
If the *pilot* phase of a simulation study has *many factors*, then straightforward experimentation may require too much computer time. In those situations the analysts may use SB, which is simple, efficient, and effective (see §2).

Stage 3: Sensitivity analysis (SA)
Sensitivity analysis is defined in this paper as the investigation of the reaction of the simulation response to either *extreme values* of the model's quantitative factors (parameters and input variables) or *drastic changes* in the model's qualitative factors (modules). (This is also called global, instead of local SA.) However, if the model's I/O behavior is non-monotonic, it may be dangerous to consider extreme input values only (see Saltelli, Andres, and Homma 1995). Non-monotonicity may be quantified by quadratic effects. Notice that this paper concentrates on a single response per run; also see Helton (1996) and McKay (1995).

The analysts may use *regression analysis* to generalize the results of the simulation experiment. To obtain better regression estimators, the analysts should apply *DOE* (see §3).

Stage 4: Uncertainty analysis (UA)
In UA (sometimes called risk analysis) the model inputs are sampled from prespecified distributions. So 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.

UA uses *Monte Carlo* sampling, possibly including variance reduction techniques such as *Latin hypercube sampling* or LHS; the results may be analyzed through regression analysis (see §5).

Stage 5: Optimization
To optimize the controllable (policy) inputs, the analysts may use *Response Surface Methodology* (RSM), which builds on regression analysis and DOE (see §6).

This five-stage procedure implies that SA should precede UA, which is controversial (see McKay 1995, pp. 7, 33, and also Fűrbringer and Roulet 1995).

Uncertainty about a simulated system's response may have two different causes: (i) The system's process is represented by a *deterministic* model, but its *parameters* are not known exactly. (ii) Some models are *intrinsically stochastic*: without the randomness the problem disappears; examples are queuing problems. Helton (1996) points out that stochastic uncertainty is a property of the system, whereas subjective uncertainty is a property of the analysts.

Uncertainty is the central problem in *mathematical probability theory*. This discipline, however, has two schools: *frequentists* versus *subjectivists*. Many UA studies concern unique events; for example, a specific investment, a nuclear accident. See Cooke (1995, pp. 4-6) and Winkler (1996).

Note that Bayesians try to combine prior subjective data with new factual data; see Draper (1995). Bayesians average the outcomes, using the probabilities of the various input scenarios. It might be argued, however, that in general it is the job of 'managers' to make decisions, pondering facts and opinions; it is the job of scientists to prepare a basis for such decisions.

Zeigler (1976)'s seminal book on the theory of modeling and simulation distinguishes between *input variables* and *parameters*. By definition, 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). This paper will use the term *factor* for parameters, input variables, and modules that are changed from (simulation) run to run, applying DOE. (For modules treated as factors also see Helton et al. 1995, McKay 1995, pp. 51-54.)

This paper combines, updates, and revises Kleijnen (1994, 1995b, 1996). For didactic reasons, the paper is not organized in the order of the five stages; for example, DOE is simpler explained for SA (stage 3, §3) than for V & V (stage 1, §4). Hence §2 discusses screening, especially SB. §3 explains SA using regression analysis and DOE. §4 briefly addresses V & V. §5 discusses UA, distinguishing between deterministic and stochastic simulations. §6 explains optimization based on RSM. §7 summarizes the conclusions.

2 SCREENING: SEQUENTIAL BIFURCATION

Screening is the search for the few (say) k important factors among the many (say) K potentially important factors ($k << K$). In practice, simulated systems often do involve many factors. Andres (1996) gives an example with 3,300 factors, Bettonvil and Kleijnen (1996) 281, De Wit (1995) 82, and McKay (1995) 84 factors. These examples concern deterministic simulations. Other examples would be stochastic simulations of queueing networks; in this field, however, applications of screening are yet unknown.

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 ($n << K$). Therefore the analysts assume that only a few factors are important ($k << K$): *parsimony, Occam's razor.*

The theory on group screening goes back to the 1960s; see Bettonvil and Kleijnen (1996) and Kleijnen (1995b). This paper recommends a *novel* group screening technique, namely SB, which is more efficient than competing group screening techniques. Moreover, SB has been used in practice, namely in the greenhouse simulation in Bettonvil and Kleijnen (1996), and in the building thermal simulation in De Wit (1995). In De Wit's case study, SB gives the 16 most important inputs among the 82 factors after only 50 runs. De Wit checks these results through a different technique, namely Morris (1991)'s randomized one-factor-at-a time design, which takes 328 runs.

All group screening techniques including SB use the following two assumptions.

Assumption 1: Low-order polynomial metamodel

A metamodel implies that the underlying simulation model is treated as a *black box*. The advantages of a low-order polynomial is that it is simple and it may apply to *all* types of random and deterministic simulation. The disadvantage is that it cannot exploit the special structure of the simulation model at hand (for alternative techniques see Helton 1996, Ho and Cao 1991, and Kleijnen and Rubinstein 1996).

Low-order polynomials are often used in DOE with its concomitant regression analysis or Analysis of Variance (ANOVA). Applications, however, are then limited to systems with a small number of factors (also see §3).

It is prudent not to assume a first-order polynomial, but a metamodel with interactions. In SB such a metamodel requires twice as many runs as a first-order metamodel does (foldover principle; see Andres 1996 and Kleijnen 1987).

Assumption 2: Known signs

Known signs are needed to know with certainty that individual effects do *not compensate* each other within a group. In practice, signs may be known indeed

(magnitudes, however, are unknown, so simulation is used). For example, higher speeds of servers are supposed to have non-negative effects on the throughput.

Details on SB are given in Bettonvil and Kleijnen (1996). Different screening techniques are Andres (1996)'s Iterated Fractional Factorial Design (IFFD) and McKay (1995)'s replicated LHS design (also see Saltelli et al. 1995).

3 SENSITIVITY ANALYSIS: REGRESSION/DOE

Once the important factors are identified (screening, §2), further analysis is recommended, using fewer assumptions: account for interactions and curvature, and assume no known signs. (The metamodeling methodology of this section is discussed at length in Kleijnen and Sargent 1996.)

3.1 Regression Analysis

It is good practice to make *scatter plots* (see Eschenbach 1992, Helton 1996). 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. See Kleijnen (1995b).

Consider the well-known *second degree polynomial* metamodel with Y_i denoting the simulation response of factor combination i (capitals denote stochastic variables); β_0 the overall mean response or regression intercept; β_h the main or first-order effect of factor h; $x_{i,h}$ the value of the standardized factor h in combination i; $\beta_{h,h'}$ the interaction between the factors h and h' with $h < h'$; $\beta_{h,h}$ the quadratic effect of factor h; E_i the additive fitting error of the regression model for factor combination i; and n the number of simulated factor combinations.

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* such that they range between -1 and +1; see Kleijnen (1995b). The original factor h is denoted by z_h, which ranges between a lowest and a highest values ($l_h \leq z_h \leq u_h$). So either the simulation model is not valid outside that range (see §4) or in practice that factor can range over that domain only. DOE addresses the question whether z_h is to be set at the extreme values only or also at intermediate values; see below. (Other measures of factor importance, such as partial correlation coefficients, are discussed in Helton 1996, McKay 1995, Saltelli and Sobol 1995, and Sobol 1996. When the logarithmic scale for Y and z_h is used, the regression parameters denote elasticity coefficients; these coefficients are popular in econometrics.)

To make statistical inferences (for example, about the importance of a factor), a *Gaussian* distribution is normally assumed. To satisfy this assumption, the analysts may apply transformations such as $\log(Y)$. An alternative is to replace both Y and x_h by their ranks: rank regression. Moreover, the analysts may investigate whether the various transformations give the same qualitative results. See Andres (1996), Helton (1996), Kleijnen (1987), Kleijnen, Bettonvil, and Van Groenendaal (1996), and Saltelli and Sobol (1995).

Of course, it is necessary to check whether the fitted regression metamodel is an adequate approximation of the underlying simulation model: (cross)validation (see Kleijnen 1995b and Kleijnen and Sargent 1996).

3.2 DOE

Let q denote the number of regression parameters; for example, q equals $k + 1 + k(k - 1)/2$ if there are main effects and two-factor interactions. Which n combinations to simulate can be determined such that the accuracy of the estimated parameters is maximized (variance minimized); this is the goal of DOE.

Consider a *first-order polynomial* metamodel with k main effects, and an overall mean. By definition, a *resolution III* or R-3 design permits the unbiased estimation of such a first-order polynomial. In *practice*, analysts usually first simulate the 'base-case', and next they change *one factor at a time*; hence $n = 1 + k$. DOE, however, gives *orthogonal* designs, which yield unbiased estimators with *smaller variances*. A well-known class of such designs are 2^{k-p} fractional factorials; for example, a simulation with $4 \leq k \leq 7$ factors requires $n = 2^{7-4} = 8$ factor combinations. Many *simulation applications* of these designs can be found in Kleijnen (1987) and Kleijnen and Van Groenendaal (1992).

In practice, however, it is prudent to assume that *two-factor interactions* may be important. A *resolution IV* or R-4 design permits the unbiased estimation of all k main effects, even if two-factor interactions are present. These designs do not give unbiased estimators of individual two-factor interactions; they may give an indication of the importance of (confounded, biased) interactions. Compared with R-3 designs, R-4 designs require that the number of simulated factor combinations be *doubled*; for example, $k = 7$ now requires $n = 2 \times 8 = 16$ runs. Obviously, changing one factor at a time (a less accurate R-3 design) does not enable estimation of interactions! Applications of R-4 designs

are mentioned in Kleijnen (1995b).

Suppose the analysts wish to estimate the *individual* two-factor interactions. A *resolution V* or R-5 design gives unbiased estimators for all main effects and all two-factor interactions. Obviously many more simulation runs are now necessary; for example, $k = 8$ factors implies $q = 37$ effects. Therefore practitioners study only small values of k; for example, $k < 5$ requires *full or complete factorial two-level designs*, denoted as 2^k. For higher values of k, however, *fractional* factorials are recommended, for example, 2^{8-2} ($n = 64$) for $k = 8$ (so $q = 37$); see Kleijnen (1987).

If all 2^k combinations are simulated, then *high-order interactions* can be estimated. However, these interactions are hard to interpret. Hence, the analysts may apply either transformations such as the logarithmic transformation or they may restrict the experimental domain. Also see Helton (1996), and Saltelli and Sobol (1995).

If factors are quantitative and their *quadratic* effects are to be estimated, then factors must be simulated for more than two values (taking extreme values minimizes the variance of the estimated main effects; see Cheng and Kleijnen 1996). *Central composite designs* have five values per factor. These designs require relatively many runs (the number of parameters is also high): $n >> q$. See Kleijnen (1987) and Kleijnen and Van Groenendaal (1992). Applications are found in the optimization of simulation models (see §6).

Simulating as many as five values per factor resembles UA, in the sense that the range of factor values is well covered (see §5).

4 VALIDATION

This paper confines the discussion of V & V to the role of SA (§3); other statistical techniques for V & V are discussed in Kleijnen (1995a).

If the simulation is fed with real-life input data (trace driven simulation), then Kleijnen, Bettonvil, and Van Groenendaal (1996) propose a novel test. This test uses regression analysis, but does *not* hypothesize that 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.

If no data are available, then the analysts and their clients still have *qualitative* knowledge; that is, they do know in which direction certain factors affect the response (see §2). If the regression metamodel (see §3) gives an estimated effect with the *wrong sign*, this is a strong indication of a wrong simulation model or a wrong computer program. An application is Kleijnen (1995c), concerning the hunt for mines on the bottom of the sea. This case-study further shows that the validity of a simulation model is restricted to a certain domain of factor combinations, called *experimental frame* in Zeigler (1976).

5 UNCERTAINTY ANALYSIS

The regression metamodel (see §3) shows which factors are 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, then they may use UA: its goal is to quantify the *probability* of specific output values (whereas SA does not tell how likely a specific result is).

First the analysts derive a *probability function* for the input values, based on sample data if available; otherwise this distribution must use subjective expert opinions (see Draper 1995, p. 92, Helton et al. 1995, p. 288, Kleijnen 1996). Correlated inputs are discussed in Cooke (1995), Helton (1996), and Kleijnen (1996).

Next the analysts use pseudorandom numbers to sample input values: *Monte Carlo*. They often use *Latin hypercube sampling* (LHS), which forces the sample (of size n) to cover the whole experimental area; see Avramidis and Wilson (1996) and Helton (1996).

This paper claims that LHS is a *variance reduction technique*, not a *screening* technique. For screening purposes the inputs should be changed to their extreme values (see §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 may be impossible. Also see Kleijnen (1996) and McKay (1992).

5.1 Deterministic Simulation

The distribution of the simulation response may be characterized by its location (measured by the mean, modus, or median) and its dispersion (quantified by the standard deviation or various percentiles and quantiles). Which quantities sufficiently summarize the distribution, depends on the users' *risk attitude*: risk neutral (the mean is then sufficient), risk averse, or risk seeking. See Helton (1996) and Kleijnen (1996).

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. (Technically, in the regression metamodel x_{ik} is replaced by X_{ik}.) Because more values are sampled per factor, more complicated metamodels might now be used; for example, for

prediction purposes splines may be used. For explanatory purposes and SA, however, simple metamodels may be preferred; see Kleijnen and Sargent (1996).

Applications in business and economics are, for example, investment studies on the probability of negative Net Present Values; see Kleijnen (1996) and Van Groenendaal and Kleijnen (1996). Applications in the natural sciences are, for example, Sandia investigations on nuclear waste disposal and reactor safety (Helton 1996); Oak Ridge studies of radioactive doses in humans; European Communities investigations on nuclear reactor safety; and Dutch studies on ecology; see Kleijnen (1996).

It is prudent to study the effects of different *specifications* of the input distributions. Sometimes, this type of SA is called *robustness analysis*. Examples are given by Helton et al. (1995), Kleijnen (1987, 1996), and McKay (1995, p. 31). The analysis may also use sophisticated, fast sampling techniques based on *importance sampling* or *likelihood ratios*; see Kleijnen and Rubinstein (1996). (Importance sampling is indispensable whenever rare events are simulated; see Helton 1996, and Kleijnen and Rubinstein 1995.)

5.2 Stochastic Simulation

A well-known example of stochastic simulation is a queueing model with interarrival times sampled from an exponential distribution with parameter λ (Helton 1996 discusses UA of three stochastic models in nuclear science). For this example Kleijnen (1983) discusses UA, assuming that the *central limit theorem* (CLT) applies. For example, suppose λ is estimated from a sample of independent interarrival times, and assume the distribution of the estimated interarrival parameter approximates a normal distribution. Then a value for λ can be sampled from this distribution, and used as input to the simulation. That simulation is run for 'enough' customers. Next the procedure is repeated: sample λ, and so on. Cheng and Holland (1996), however, do not rely on the CLT, but apply *bootstrapping*.

Still, the question remains *which* response to report to the users: the unconditional, ex post variance (see Cheng and Holland 1996); the ex post variance, mean, and various quantiles (see Haverkort and Meeuwissen 1995); or the conditional moments (conditioned on the values of the estimated parameters)? The discussion in Draper (1995, pp. 78, 83) clearly demonstrates how *controversial* this issue is.

6 OPTIMIZATION: RSM

The decision makers should control the *policy variables*. This paper is limited to RSM, which combines regression analysis and DOE (see §3) with a hill-climbing technique called steepest ascent; see Myers and Montgomery (1995).

Four *general characteristics of RSM* are: (i) RSM uses first and second order polynomial regression metamodels, now called *response surfaces* (see §3). (ii) RSM employs the *statistical designs* of DOE (see §3). (iii) RSM augments DOE with the *steepest ascent* technique, to determine in which direction the decision variables should be changed. (iv) RSM uses *canonical analysis* to analyze the shape of the optimal region (unique maximum, saddle point, ridge).

Kleijnen (1995b) mentions several simulation studies that use RSM. Kleijnen (1995d) discusses a system dynamics model for coal transport.

7 CONCLUSIONS

There are several *related types of analysis* in simulation studies. Questions are: *when* should *which* type of analysis be applied; which *statistical techniques* should then be used?

This paper proposes a *five-stage procedure* for the analysis of simulation models: (i) validation and verification (V & V), (ii) screening, (iii) sensitivity analysis (SA) (iv) uncertainty analysis (UA), and (v) optimization.

Each type of analysis may apply its own set of *statistical techniques*, for example, SA may use 2^{k-p} fractional designs, and UA may apply LHS. Some techniques may be applied in both analyses, for example, regression modelling.

Applications of the statistical techniques recommended above, are quite plentiful. Further *research* is needed on UA of stochastic simulation models.

SA and UA remain *controversial* topics; for example, this paper claims that SA should precede UA. *SA* shows which model inputs are really important. From the users' perspective, these 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 simulation model. If, however, these values 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 *UA*.

More specifically, SA means that the model is subjected to *extreme value* testing. A model is valid

only within its experimental frame. Mathematically that frame might be defined as the hypercube formed by the k standardized inputs of the model. 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 (also see Draper 1995, p. 55). 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. DOE gives better estimates of those input effects.

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

Further development of a methodology for the evaluation of complicated simulation models certainly requires continued communication within and among scientific disciplines!

REFERENCES

Andres, T.H. (1996), Sampling methods and sensitivity analysis for large parameter sets. *Journal Statistical Computation and Simulation* (accepted).

Avramidis, A.N. and J.R. Wilson (1996), Integrated Variance Reduction Strategies for Simulation. *Operations Research* 44: 327-346.

Bettonvil, B. and J.P.C. Kleijnen, (1996), Searching for important factors in simulation models with many factors: sequential bifurcation. *European Journal of Operational Research* (accepted).

Cheng, R.C.H. and W. Holland (1996), The sensitivity of computer simulation experiments to errors in input data. *Journal Statistical Computation and Simulation* (accepted).

Cheng, R.C.H. and J.P.C. Kleijnen (1996), Improved design of queueing simulation experiments with highly heteroscedastic responses. *Operations research* (accepted).

Cooke, R.M. (1995), UNICORN: methods and code for uncertainty analysis. The SRD Association, AEA Technology, Thomson House, Risley, Warrington WA3 6AT, UK.

De Wit, M.S. (1995), Uncertainty analysis in building thermal modelling. *SAMO95*, Office for Official Publications of the European Communities, Luxembourg.

Draper, D. (1995), Assessment and propagation of model uncertainty, *Journal Royal Statistical Society, Series B*, 57: 45-97.

Eschenbach, T.G. (1992), Spider plots versus tornado diagrams for sensitivity analysis. *Interfaces*, 22: 40-46.

Fűrbringer, J.-M. and C.A. Roulet (1996), Comparison and combination of factorial and Monte-Carlo design in sensitivity analysis. *Building and Environment*, 30: 505-519.

Haverkort, B.R. and A.M.H. Meeuwissen (1995), Sensitivity and uncertainty analysis of Markov-reward models. *IEEE Transactions on Reliability*, 44: 147-154.

Helton, J.C. (1996), Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. *Journal Statistical Computation and Simulation* (accepted).

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.

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

Kleijnen, J.P.C. (1983) Risk analysis and sensitivity analysis: antithesis or synthesis? *Simuletter*, 14: 64-72.

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

Kleijnen, J.P.C. (1994), Sensitivity analysis versus uncertainty analysis: when to use what? *Predictability and Nonlinear 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, 1994: 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*, 87: 21-34.

Kleijnen, J.P.C. (1995d), Sensitivity analysis and optimization of system dynamics models: regression analysis and statistical design of experiments. *System Dynamics Review*, 11: 1-14.

Kleijnen, J.P.C. (1996), Sensitivity analysis and related analyses: a survey of statistical techniques. *Journal Statistical Computation and Simulation* (accepted)

Kleijnen, J.P.C. and R.Y. Rubinstein (1996) Optimization and sensitivity analysis of computer simula-

tion models by the score function method. *European Journal of Operational Research*, 88: 1-15.

Kleijnen, J.P.C., B. Bettonvil, and W. Van Groenendaal (1996), Validation of simulation models: a novel regression test. *Management Science* (accepted).

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

Kleijnen, J.P.C. and R. Sargent (1996), Metamodeling methodology. Tilburg University, Tilburg, Netherlands.

McKay, M.D. (1992), Latin hypercube sampling as a tool in uncertainty analysis of computer models. *1992 Winter Simulation Conference Proceedings* (edited by J.J. Swain, D. Goldsman, R.C. Crain, and J.R. Wilson).

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

Myers, R.H. and D.C. Montgomery (1995), *Response Surface Methodology*.

Morris, M.D. (1991), Factorial plans for preliminary computational experiments, *Technometrics*, 33: 161-174.

Saltelli, A., T.H. Andres, and T. Homma (1995), Sensitivity analysis of model output; performance of the iterated fractional factorial design method. *Computational Statistics & Data Analysis*, 20: 387-407.

Saltelli, A. and I.M. Sobol (1995), About the use of rank transformation in sensitivity analysis of model output. *Reliability Engineering and System Safety*, 50: 225-239.

Sobol, I.M. (1996), Sensitivity analysis of nonlinear models using sensitivity indices. *Journal Statistical Computation and Simulation* (accepted).

Winkler, R.L. (1996), Uncertainty in probabilistic risk assessment. *Reliability Engineering and System Safety*, Special Issue, edited by J.Helton and D. Burmaster.

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.