Tilburg University

Sensitivity Analysis of Simulation Models

Kleijnen, J.P.C.

Publication date:
2009

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Kleijnen, J. P. C. (2009). *Sensitivity Analysis of Simulation Models*. (CentER Discussion Paper; Vol. 2009-11). Operations research.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

No. 2009–11

SENSITIVITY ANALYSIS OF SIMULATION MODELS

By Jack P.C. Kleijnen

February 2009

ISSN 0924-7815

Sensitivity Analysis of Simulation Models

Jack P.C. Kleijnen

Department of Information Management / CentER
Tilburg University, Postbox 90153, 5000 LE Tilburg,
Netherlands
<http://center.uvt.nl/staff/kleijnen/>

Abstract

This contribution presents an overview of sensitivity analysis of simulation models, including the estimation of gradients. It covers classic designs and their corresponding (meta)models; namely, resolution-III designs including fractional-factorial two-level designs for first-order polynomial metamodels, resolution-IV and resolution-V designs for metamodels augmented with two-factor interactions, and designs for second-degree polynomial metamodels including central composite designs. It also reviews factor screening for simulation models with very many factors, focusing on the so-called ‘sequential bifurcation’ method. Furthermore, it reviews Kriging metamodels and their designs. It mentions that sensitivity analysis may also aim at the optimization of the simulated system, allowing multiple random simulation outputs.

Keywords: simulation, sensitivity analysis, gradients, screening, Kriging, optimization, Response Surface Methodology, Taguchi

JEL: C0, C1, C9

Sensitivity analysis of simulation models estimates the change in the simulation output as the simulation input changes. This input consists of (say) k input variables x_j with $j = 1, \dots, k$ (e.g., machines in manufacturing facility design) or parameters (e.g., production rates)—these variables and parameters are called ‘factors’ in the statistical theory on Design Of Experiments, which is usually abbreviated to DOE. The simulation output may be a single response (a scalar or univariate) or multiple responses (a vector or multivariate). This contribution focuses on a single response, assuming that in case of multiple outputs the sensitivity analysis is applied per output. Realistic simulation models have multiple inputs: $k > 1$. The inputs may vary either locally or globally. Local small changes enable the estimation of the *gradient*, the vector with

the first-order partial derivatives (say) $\nabla(w) = (\partial w/\partial x_1, \dots, \partial w/\partial x_k)$ where $w(x_1, \dots, x_k)$ denotes the Input/Output (I/O) function implicitly defined by the given simulation model (computer program or code). Gradients play an important role in simulation optimization. The inputs, however, may be integers (e.g., the number of machines) so the gradient is not defined. Or the users may be interested in output changes caused by changing an input from its minimum to its maximum in the *experimental area* (called the ‘experimental frame’ in [1] or the ‘domain of admissible scenarios’ given the goals of the simulation study; various goals are discussed in [2] and [3]). Such global sensitivity analysis is also called ‘What If’ analysis. Notice that the terminology differs among different scientific disciplines and application areas.

Sensitivity analysis is needed for any type of simulation model—be it discrete-event, continuous, or hybrid (see Sections 2.4.1.2 and 2.4.1.3 of the encyclopedia), terminating or nonterminating (Sections 2.4.1.4 and 2.4.1.5). The sensitivity analysis in this contribution uses statistical methods (techniques) that have been developed for the design and analysis of three types of experiments:

- real-life (physical) systems;
- deterministic simulation models;
- random (stochastic) simulation models.

For *real-life* systems, DOE started with agricultural experiments in the 1920s (Sir Ronald Fisher), followed by chemical experiments in the 1950s (George Box), and social systems (e.g., educational and service systems) nowadays. DOE for real systems is covered extensively by many textbooks; e.g., [4] and [5].

In *deterministic simulation*, DOE gained popularity with the increased use of ‘computer codes’ for the design (in an engineering, not a statistical sense) of airplanes, automobiles, TV sets, chemical processes, computer chips, etc.—in Computer Aided Engineering (CAE) and Computer Aided Design (CAD). This domain often uses the term *DACE*, Design and Analysis of Computer Experiments, instead of DOE. The classic article on DACE is [6]; a classic textbook is [7].

Random simulation includes discrete-event models such as queuing and inventory models, but also stochastic difference equation models. This type of simulation is very popular in Operations Research and Management Science. DOE for random simulation is the focus of several textbooks by Kleijnen; the most recent one is [8].

Note that deterministic simulation models become random if inputs are unknown so their values are sampled from statistical distribution functions; again see [8].

Sensitivity analysis using DOE views the simulation model as a *black box*; i.e., only the simulation I/O data are used (the values of internal variables and specific functions implied by the simulation’s computer modules are not observed). *White-box* approaches are Perturbation Analysis and the Score Function or Likelihood Ratio method, which estimate the gradient for local—not global—sensitivity analysis and for optimization.

Whereas in real-life experimentation it is not practical to investigate many factors, realistic simulation experiments may have hundreds of factors. Moreover, whereas in real-life experiments it is hard to vary a factor over many values, in simulation experiments this restriction does not apply (Latin Hypercube Sampling or LHS is a design that has as many values per factor as it has combinations; see Section 3 below). Consequently, a multitude of *scenarios*—combinations of factor values—may be observed. Furthermore, simulation experiments are well-suited to the application of sequential designs instead of ‘one shot’ designs; see Sections 2 and 3 below. So a change of mindset of the experimenter is necessary; also see [9].

Design and *analysis* of experiments are like ‘chicken and egg’. The analysis uses—implicitly or explicitly—a *metamodel* (also called response surface, surrogate, emulator, etc.), which is an approximation of the I/O function implied by the underlying simulation model; i.e., the experiment yields I/O data that are used to estimate this function. There are different *types* of metamodels. The most popular type are polynomials of first or second order (degree); see Section 1 below. Obviously, a first-order polynomial implies a gradient that remains constant over the whole experimental area. Another metamodel type are Kriging or Gaussian Process models, which enable the approximation of more general I/O functions; such a Kriging metamodel implies a gradient that changes over the whole experimental area, as Section 3 will show. References to other metamodel types are given by [8], p. 8. Different metamodels require different designs; e.g., first-order polynomials require two values per factor, whereas Kriging requires many more values per factor—as the next sections will show. The adequacy of the design and metamodel combination depends on the *goal* of the experiment; again see [2]. This goal may be sensitivity analysis, including *factor screening*—or briefly screening—which denotes the search for the really important factors among the many factors that are varied in a simulation experiment. Sensitivity analysis may also serve optimization,

because sensitivity analysis may provide the gradient for the important factors.

So, sensitivity analysis may also serve optimization. There are optimization methods closely related to the design and analysis methods discussed in this contribution. The latter methods allow multiple random simulation outputs, selecting one output as the goal (objective) output while the other random simulation outputs must satisfy prespecified target values (thresholds). Generalized Response Surface methodology (GRSM) fits a series of local low-order polynomials to the simulation I/O data; see [10]. Kriging is combined with Mathematical Programming in [11]. ‘Robust’ simulation-optimization in the sense of Taguchi [12] is discussed in [13].

The following overview of the remainder of this contribution enables readers to decide which sections they might wish to skip.

Section 1 covers classic designs and their corresponding metamodels, namely Resolution-III (R-III) designs including 2^{k-p} designs for first-order polynomials, Resolution-IV (R-IV) and Resolution-V (R-V) designs for two-factor interactions, and designs for second-degree polynomials including Central Composite Designs (CCDs). Compared with these designs, the traditional approach of changing one factor at a time is inferior; again see [8], p. 33.

Section 2 reviews screening, focussing on sequential bifurcation. Traditionally, simulation analysts use prior knowledge to select a few factors, assuming that these factors are the most important ones. In a recent case-study with 92 factors, sequential bifurcation identified a shortlist with the ten most important factors—after investigating only 19 combinations.

Section 3 reviews Kriging and its designs, especially LHS. Kriging models are fitted to data that are obtained for larger experimental areas than the areas often used when fitting low-order polynomials; i.e., Kriging models are *global* rather than local. Kriging is used for prediction; its final goals are sensitivity analysis and optimization.

Section 4 summarizes this contribution.

1 Classic Designs and Metamodels

Classic designs and their corresponding metamodels are detailed in many DOE textbooks; again see [4] and [5]; DOE for simulation experiments is detailed in [8]. To illustrate these designs, consider the following example with $k = 6$ factors.

To estimate the *first-order polynomial* metamodel, at least $k + 1 = 7$ combinations need to be simulated. Table 1 presents a 2_{III}^{7-4} design, which should be read as follows. The columns with the symbols **1** through **7**

Combi.	1	2	3	4 = 1.2	5 = 1.3	6 = 2.3	7 = 1.2.3
1	-	-	-	+	+	+	-
2	+	-	-	-	-	+	+
3	-	+	-	-	+	-	+
4	+	+	-	+	-	-	-
5	-	-	+	+	-	-	+
6	+	-	+	-	+	-	-
7	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+

Table 1: A one-sixteenth fractional factorial design for seven factors

give the values of the corresponding factor in the experiment; **4 = 1.2** means that the value of factor 4 equals the product of the values of the factors 1 and 2 in the corresponding combination (abbreviated to ‘Combi’); the symbol - means that the factor has the standardized value -1—which corresponds with the lowest value in the original scale—and + means that the factor has the standardized value +1—highest value in the original scale. This table gives a R-III design, which means that it enables the unbiased estimation of the coefficients of the first-order polynomial—assuming this polynomial is an adequate approximation of the I/O function. This design investigates only 8 combinations; i.e., it investigates only a fraction $2^{-4} = 1/16$ of all $2^7 = 128$ combinations of a full-factorial two-level design.

This R-3 design enables the *Ordinary Least Squares* (OLS) estimation of the coefficients of the first-order polynomial; these coefficients represent the $k = 6$ first-order effects (say) β_j ($j = 1, \dots, 6$) and the intercept (overall mean) β_0 , which corresponds with a column that had only the value +. OLS is the classic estimation method in linear regression analysis, assuming *white noise*; i.e., the metamodel’s residuals are Normally (Gaussian), Independently, and Identically Distributed (NIID) with zero mean. A first-order polynomial for k factors implies the gradient $\nabla(w) = (\beta_1, \dots, \beta_k)$, so the most important factor is the one with the maximum absolute value of the first-order effect.

Next consider a R-IV design. By definition, this design ensures that the estimated first-order effects (say) $\widehat{\beta_j}$ are *not biased* by the two-factor interactions $\beta_{j;j'}$ ($j < j'$; $j' = 2, \dots, 6$). Such a design is easily constructed, as follows. Let **D** denote the R-III design; i.e., replace - in Table 1 by -1, and + by 1 so the 8×7 design matrix **D** results. Then add $-\mathbf{D}$ to **D** to get a R-IV design (so the R-IV design doubles the number of combinations).

To estimate the *individual* interactions, a R-V design is needed. An

example is a 2_V^{6-1} design; however, its 32 combinations take too much computer time if a simulation run is computationally expensive. In that case, it is better to use a *saturated* design, which by definition has a number of combinations (say) n that equals the number of metamodel parameters q ; e.g. $k = 6$ implies $q = 1 + 6 + 15 = 22$. A type of saturated R-V design is Rechtschaffner’s design; see [8], p. 49.

If the analysts use a *second-degree polynomial* approximation for the I/O function, then a CCD enables the estimation of the k ‘purely quadratic effects’ (say) $\beta_{j;j}$. A CCD augments the R-V design with the ‘central point’ of the experimental area (which has standardized values 0) and $2k$ ‘axial points’, which change each factor one-at-a-time by $-c$ and c units where $c > 0$. The CCD is rather wasteful in case of expensive simulation models, because it uses five values per factor (instead of the minimum, three) and it is not saturated. Alternatives for CCDs are discussed in [5] and [8].

Many simulation applications of these classic designs are provided by [8]. The assumptions of these designs and their metamodels stipulate univariate output and white noise. Reference [8] discusses multivariate (multiple) outputs, nonnormality of the output (solved through either jackknifing or bootstrapping), variance heterogeneity (e.g., as the traffic rate increases, both the mean and the variance of the waiting time increase), and Common Random Numbers or CRN; see Section 2.4.4.1 (the correlations created by CRN may be incorporated through either Generalized Least Squares or OLS computed per replicate), and testing the validity of low-order polynomial metamodels (through either the F lack-of-fit statistic or cross-validation).

2 Screening: Sequential Bifurcation

R-III designs (discussed in Section 1) are called screening designs by some authors; e.g., [14]. Screening is related to ‘sparse’ effects, ‘parsimony’, ‘Pareto principle’, ‘Occam’s razor’, ‘20-80 rule’, and ‘curse of dimensionality’. Unfortunately, screening is not yet much applied in practice; instead, analysts experiment with a few intuitively selected factors only. Nevertheless, the following case study illustrates the need for screening. In [15], a greenhouse deterministic simulation model has 281 factors. The politicians, however, want to reduce the release of CO_2 gasses, and start with legislation for a few factors only. Another case study is presented in [16], concerning a discrete-event simulation model of a supply chain of the Ericsson company in Sweden. This simulation has 92 factors; sequential bifurcation identifies a shortlist with 10 factors after simulating only 19 combinations.

There are several types of screening designs. Sequential bifurcation

is very efficient—it is supersaturated; i.e., $n < k$ —and effective if its assumptions are satisfied; also see [17]. It uses the following metamodel *assumptions*.

1. A first-order polynomial augmented with two-factor interactions is an adequate approximation.
2. All first-order effects have known signs and are nonnegative.
3. There is ‘strong heredity’; i.e., if a factor has no important main effect, then this factor does not interact with any other factor; also see [18].

The role of these assumptions may be demonstrated, as follows. The first step of sequential bifurcation aggregates all factors into a single group, and tests whether or not that group of factors has an important effect; no *cancellation* of effects occurs, given the assumptions 1 and 2. If that group indeed has an important effect—which is very likely in the first step—then the second step splits the group into two subgroups—it bifurcates—and tests each of these subgroups for importance. In the next steps, sequential bifurcation splits important subgroups into smaller subgroups, and discards unimportant subgroups. In the final step, all individual factors that are not in subgroups identified as unimportant, are estimated and tested.

This procedure may be interpreted through the following *metaphor*. Imagine a lake that is controlled by a dam. The goal of the experiment is to identify the highest (most important) rocks; actually, sequential bifurcation not only identifies but also measures the height of these ‘rocks’. The dam is controlled in such a way that the level of the murky water slowly drops. Obviously, the highest rock first emerges from the water! The most-important-but-one rock turns up next, etc. Sequential bifurcation stops when the analysts feel that all the ‘important’ factors are identified; once it stops, the analysts know that all remaining (unidentified) factors have smaller effects than the effects of the factors that have been identified. Obviously, this property is important for practice.

There is a need for more research on sequential bifurcation:

- To control the overall probability of correctly classifying the individual factors as important or unimportant, the number of replicates need to be derived; currently, a statistical test is applied to each subgroup individually (by definition, replicates use the same input combination, but different random number streams).
- After sequential bifurcation (with its assumptions) stops, the resulting shortlist of important factors should be validated.

- Practical simulation models have multiple simulation outputs (instead of a single one); sequential bifurcation should be investigated for this situation.
- There is a need for software that implements sequential bifurcation.
- A contest may be organized to challenge different screening methods to find the important factors of different simulation models. Such ‘testbeds’ are popular in Mathematical Programming, but not yet in simulation.

3 Kriging

Originally, Kriging was developed for interpolation in geostatistical or spatial sampling by the South African mining engineer Danie Krige; see the classic textbook [19]. Later on, Kriging was applied to the I/O data of deterministic simulation models; again see [6] and [7]. Only recently, Kriging has also been applied to random simulation models; see [20], and also [21], [22], and [23]. The track record of Kriging in deterministic simulation holds great promise for Kriging in random simulation.

Like most Kriging publications, this contribution focuses on the simplest type of Kriging called *Ordinary* Kriging, which assumes

$$w(\mathbf{d}) = \mu + \delta(\mathbf{d}) \quad (1)$$

where $w(\mathbf{d})$ denotes the simulation output for input combination $\mathbf{d}$, μ is the output averaged over the whole experimental area, and $\delta(\mathbf{d})$ is the additive noise that forms a stationary covariance process with zero mean. This Kriging uses the *linear* predictor

$$y(\mathbf{d}) = \boldsymbol{\lambda}'\mathbf{w} \quad (2)$$

where the Kriging weights $\boldsymbol{\lambda}$ are not constants—whereas the regression parameters $\boldsymbol{\beta}$ are—but decrease with the *distance* between the ‘new’ input $\mathbf{d}$ to be predicted and the ‘old’ points collected in the $n \times k$ design matrix $\mathbf{D}$. The *optimal* weights are proven to be

$$\boldsymbol{\lambda}_o = \boldsymbol{\Gamma}^{-1}[\boldsymbol{\gamma} + \mathbf{1} \frac{1 - \mathbf{1}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}}{\mathbf{1}'\boldsymbol{\Gamma}^{-1}\mathbf{1}}] \quad (3)$$

where $\boldsymbol{\Gamma} = (\text{cov}(w_i, w_{i'}))$ with $i, i' = 1, \dots, n$ is the $n \times n$ matrix with the covariances between the old outputs; $\boldsymbol{\gamma} = (\text{cov}(w_i, w_0))$ is the n -dimensional vector with the covariances between the n old outputs w_i and w_0 , the output of the combination to be predicted which may be

either new or old. Actually, these last three equations imply that the predictor may be written as

$$y(\mathbf{d}) = \hat{\mu} + \boldsymbol{\gamma}(\mathbf{d})' \boldsymbol{\Gamma}^{-1}(\mathbf{w} - \hat{\mu} \mathbf{1}) \quad (4)$$

where

$$\hat{\mu} = (\mathbf{1}' \boldsymbol{\Gamma}^{-1} \mathbf{1})^{-1} \mathbf{1}' \boldsymbol{\Gamma}^{-1} \mathbf{w}.$$

The covariances collected in $\boldsymbol{\Gamma}$ and $\boldsymbol{\gamma}$ are often based on the *correlation function*

$$\rho = \exp\left[-\sum_{j=1}^k \theta_j h_j^{p_j}\right] = \prod_{j=1}^k \exp[-\theta_j h_j^{p_j}] \quad (5)$$

where h_j denotes the distance between the input d_j of the new and the old combinations, θ_j denotes the importance of input j (the higher θ_j is, the less effect input j has), and p_j denotes the smoothness of the correlation function (e.g., $p_j = 2$ implies an infinitely differentiable function). So-called exponential and Gaussian correlation functions have $p = 1$ and $p = 2$ respectively. The correlation function (5) implies that the weights are relatively high for inputs close to the input to be predicted. Finally, the weights imply that for an old input the predictor equals the observed output at that input:

$$y(\mathbf{d}_i) = w(\mathbf{d}_i) \text{ if } \mathbf{d}_i \in \mathbf{D}, \quad (6)$$

so all weights are zero except the weight of the observed output; i.e., the Kriging predictor is an *exact interpolator* (the OLS regression predictor minimizes the Sum of Squared Residuals, so it is not an exact interpolator—unless the design is saturated); also see the third paragraph that follows.

The *gradient* follows from (4) and (5); e.g., assuming a single input and a Gaussian correlation function, [8], p. 156 derives that the gradient $\nabla(y)$ equals

$$\left(-2\theta(d_0 - d_1)e^{-\theta(d_0 - d_1)^2}, \dots, -2\theta(d_0 - d_n)e^{-\theta(d_0 - d_n)^2}\right) \cdot \boldsymbol{\Gamma}^{-1}(\mathbf{w} - \hat{\mu} \mathbf{1}).$$

A major problem is that *the correlation function is unknown*, so both the type and the parameter values must be estimated. To estimate the parameters, the standard Kriging literature and software uses Maximum Likelihood Estimators (MLEs). The estimation of the correlation functions, the corresponding optimal Kriging weights, the Kriging predictor, and the corresponding gradient can all be done through DACE, which is software that is well documented and free of charge; see [24].

The interpolation property (6) is attractive in *deterministic* simulation, because the observed simulation output is unambiguous. In *random*

simulation, however, the observed output is only one of the many possible values. For random simulations, [20] replaces $w(\mathbf{d}_i)$ by the average observed output $\overline{w}_i$, and gives examples in which the Kriging predictions are much better than the regression predictions.

The Kriging literature virtually ignores problems caused by replacing the optimal weights $\boldsymbol{\lambda}$ in (2) by the estimated optimal weights (say) $\widehat{\boldsymbol{\lambda}}_0$, which makes the Kriging predictor a *nonlinear* estimator. The literature uses the predictor variance—*given* the Kriging weights $\boldsymbol{\lambda}$. This equation implies a zero variance in case the new point w_0 equals an old point w_i ; also see (6). Furthermore this equation tends to underestimate the true variance. Finally, this conditional variance and the true variance do not reach their maxima for the same input combination, which is important in sequential designs; see [25].

In random simulation, each input combination is replicated a number of times so a simple method for estimating the true gradient is *distribution-free bootstrapping*. The basics of bootstrapping are explained in [26]. To estimate the true gradient in Kriging, this bootstrapping resamples with replacement the (say) m_i replicates for combination i ($i = 1, \dots, n$). This sampling gives the bootstrapped average $\overline{w}_i^*$ where the superscript $*$ is the usual symbol to denote a bootstrapped observation. These n bootstrapped averages $\overline{w}_i^*$ give the bootstrapped estimated optimal weights $\widehat{\boldsymbol{\lambda}}_0^*$, the corresponding bootstrapped Kriging predictor y^* , and the bootstrapped gradient $\nabla(w)^*$. To decrease sampling effects, this whole procedure is repeated B times (e.g., $B = 100$), which gives $\nabla(w)_b^*$ with $b = 1, \dots, B$. These B values enable estimation of the mean and variance of the estimated gradient. These two estimates enable testing whether any gradient component is zero. Also see [27].

To get the Kriging I/O data, experimenters often use *LHS*. This design assumes that a valid metamodel is more complicated than a low-order polynomial (assumed by classic designs; again see Section 1). LHS does not assume a specific metamodel. Instead, LHS focuses on the design space formed by the k -dimensional unit cube defined by the k standardized inputs. LHS is a space filling design; other types of space-filling designs are discussed in [8].

Instead of such a (one-shot) space-filling design, a *sequentialized* design may be used. In general, sequential statistical procedures are known to require fewer observations than fixed-sample procedures. Sequential designs imply that simulation I/O data are analyzed—so the data generating process is better understood—*before* the next input combination is selected. This property implies that the design depends on the specific underlying simulation model; i.e., the design is customized (tailored or application-driven, not generic). Moreover, simulation experiments (un-

like real-life experiments) proceed sequentially. The following sequential design for Kriging in sensitivity analysis is developed in [27]:

1. Start with a *pilot* experiment, using some small generic space-filling design (e.g., a LHS design).
2. Fit a *Kriging* metamodel to the simulation I/O data that are available at this step (in the first pass of this procedure, these data are the data resulting from Step 1).
3. Consider (but do not yet simulate) a set of *candidate* input combinations that have not yet been simulated and that are selected through some space-filling design; select as the next combination to be actually simulated, the candidate combination that has the *highest predictor variance* (computed through bootstrapping).
4. Use the combination selected in Step 3 as the input combination to be simulated (expensive!), and obtain the corresponding simulation output.
5. Re-fit a Kriging model to the I/O data augmented with the data from Step 4.
6. Return to Step 3, until the Kriging metamodel is acceptable for its goal, sensitivity analysis.

The resulting design is indeed *customized*; i.e., which combination has the highest predictor variance (computed in Step 3) is determined by the underlying simulation model; e.g., if the true (unknown) I/O function is a simple hyperplane within a subspace of the total experimental area, then this design selects relatively few points in that part of the input space. A sequential design for constrained optimization (instead of sensitivity analysis) is presented in [11].

There is a need for more research on Kriging:

- Kriging *software* needs further improvement; e.g., Kriging should allow random outputs with variance heterogeneity and correlations created by CRN.
- Sequential designs may benefit from *asymptotic proofs* of their performance; e.g., does the design approximate the ‘optimal’ design (optimal designs are discussed in several publications including the recent [28])?

- More experimentation and analyses may be done to derive *rules of thumb* for the sequential design's parameters, such as the size of the pilot design and the initial number of replicates.
- *Stopping rules* for sequential designs based on a measure of accuracy may be investigated.
- Nearly all Kriging publications assume univariate output, whereas in practice simulation models have *multivariate output*; see [7] pp. 101-116, [29], and [30].
- Often the analysts know that the simulation's I/O function has certain properties, e.g., monotonicity (see [8], p. 162, [31], and [32]). Most metamodels (such as Kriging and regression) do not preserve these properties.

4 Summary

This contribution gave an overview of sensitivity analysis through the statistical design and analysis of simulation experiments. It covered classic designs and their analysis through corresponding metamodels; namely, R-III designs (such as 2^{k-p} designs) for first-order polynomials, R-IV and R-V designs for two-factor interactions, and designs such as CCD for second-degree polynomials. It also reviewed factor screening for simulation experiments with very many factors, focusing on the supersaturated sequential bifurcation method. Furthermore, it reviewed Kriging and its designs (especially LHS). Topics for future research were also presented. The challenge now is to apply these methods to more simulation models in practice.

References

- [1] Zeigler B.P., H. Praehofer, and T.G. Kim (2000), *Theory of modeling and simulation; second edition*. Academic Press, San Diego
- [2] Kleijnen, J.P.C. and R.G. Sargent (2000), A methodology for the fitting and validation of metamodels in simulation.) *European Journal of Operational Research*, 120, no.1, pp. 14-29
- [3] Law, A.M. (2007), *Simulation modeling and analysis; fourth edition*. McGraw-Hill, Boston
- [4] Montgomery, D. C. (2009), *Design and analysis of experiments; 7th edition*, Wiley, Hoboken, New Jersey
- [5] Myers, R.H. and D.C. Montgomery (2002), *Response surface methodology: process and product optimization using designed experiments; second edition*. Wiley, New York

- [6] Sacks, J., W.J. Welch, T.J. Mitchell, and H.P. Wynn (1989), Design and analysis of computer experiments (includes Comments and Rejoinder). *Statistical Science*, 4, no. 4, pp. 409-435
- [7] Santner, T.J., B.J. Williams, and W.I. Notz (2003), *The design and analysis of computer experiments*. Springer-Verlag, New York
- [8] Kleijnen, J.P.C. (2008), *Design and analysis of simulation experiments*, Springer
- [9] Kleijnen, J.P.C., S.M. Sanchez, T.W. Lucas, and T.M. Cioppa (2005), State-of-the-art review: a user's guide to the brave new world of designing simulation experiments. *INFORMS Journal on Computing*, 17, no. 3, pp. 263-289
- [10] Angün, E., J. Kleijnen, D. den Hertog, and G. Gürkan (2008), Response surface methodology with stochastic constraints for expensive simulation. *Journal of the Operational Research Society* (accepted)
- [11] Kleijnen J.P.C., W. van Beers, and I. van Nieuwenhuyse (2009), Constrained optimization in simulation: a novel approach. Working Paper
- [12] Taguchi, G. (1987), *System of experimental designs, volumes 1 and 2*. UNIPUB/Krauss International, White Plains, New York
- [13] Dellino, G., J.P.C. Kleijnen, and C. Meloni (2009), Robust optimization in simulation: Taguchi and Response Surface Methodology. Working Paper
- [14] Yu, H-F. (2007), Designing a screening experiment with a reciprocal Weibull degradation rate. *Computers & Industrial Engineering*, 52, no. 2, pp. 175-191
- [15] Bettonvil, B. and J.P.C. Kleijnen (1997), Searching for important factors in simulation models with many factors: sequential bifurcation. *European Journal of Operational Research*, 96, pp. 180-194
- [16] Kleijnen, J.P.C., B. Bettonvil, and F. Persson (2006), Screening for the important factors in large discrete-event simulation: sequential bifurcation and its applications. In: *Screening: Methods for experimentation in industry, drug discovery, and genetics*, edited by A. Dean and S. Lewis, Springer-Verlag, New York, pp. 287-307
- [17] Xu, J., F. Yang, and H. Wan (2007), Controlled sequential bifurcation for software reliability study. *Proceedings of the 2007 Winter Simulation Conference*, edited by S.G. Henderson, B. Biller, M-H. Hsieh, J. Shortle, J.D. Tew, and R.R. Barton, pp. 281-288
- [18] Wu, C.F.J. and M. Hamada (2000), *Experiments; planning, analysis, and parameter design optimization*. Wiley, New York
- [19] Cressie, N.A.C. (1993), *Statistics for spatial data: revised edition*. Wiley, New York
- [20] Van Beers, W. and J.P.C. Kleijnen (2003), Kriging for interpolation

- in random simulation. *Journal of the Operational Research Society*, no. 54, pp. 255-262
- [21] Ankenman, B., B.L. Nelson, and J. Staum (2008), Stochastic Kriging for simulation metamodeling. *Proceedings of the 2008 Winter Simulation Conference*, edited by S.J. Mason, R.R. Hill, L. Moench, O. Rose, pp. 362–370
 - [22] Gramacy, R.B. and H.K.H. Lee (2008), Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, 103, no. 483, pp. 1119-1130
 - [23] Yin, J., S.H. Ng, and K.M. Ng (2008), Kriging model with modified nugget effect. *Proceedings of the 2008 IEEE International Conference on Industrial Engineering and Engineering Management*, to appear
 - [24] Lophaven, S.N., H.B. Nielsen, and J. Sondergaard (2002), DACE: a Matlab Kriging toolbox, version 2.0. IMM Technical University of Denmark, Lyngby
 - [25] Den Hertog, D., J.P.C. Kleijnen, and A.Y.D. Siem (2006), The correct Kriging variance estimated by bootstrapping. *Journal of the Operational Research Society*, 57, no. 4, pp. 400-409
 - [26] Efron, B. and R.J. Tibshirani (1993), *An introduction to the bootstrap*. Chapman & Hall, New York
 - [27] Van Beers, W. and J.P.C. Kleijnen (2008), Customized sequential designs for random simulation experiments: Kriging metamodeling and bootstrapping. *European Journal of Operational Research*, 186, no. 3, pp. 1099-1113
 - [28] Melas, V.B. (2006), *Functional approach to optimal experimental design* Springer, Berlin
 - [29] Chilès J-P., P. Delfiner. 1999. *Geostatistics: modeling spatial uncertainty*. Wiley, New York
 - [30] Wackernagel, H. (2003), *Multivariate geostatistics: an introduction with applications; third edition*. Springer-Verlag, Berlin
 - [31] Siem, A.Y.D. (2007), *Property preservation and quality measures in meta-models*. Ph.D. dissertation, Tilburg University, Tilburg, Netherlands
 - [32] Velikova, M.V. (2006), *Monotone models for prediction in data mining*. Ph.D. dissertation, Tilburg University, Tilburg, Netherlands