Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Operational Research 171 (2006) 53–63
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Stochastics and Statistics

Statistical fitting and validation of non-linear simulation metamodels: A case study

M. Isabel Reis dos Santos [a,*], Acácio M.O. Porta Nova [b]

[a] *Department of Mathematics, Center for Management Studies, Technical University of Lisbon (IST), Av. Rovisco Pais, 1049-001 Lisboa, Portugal*

[b] *Department of Engineering and Management, Center for Management Studies, Technical University of Lisbon (IST), Av. Rovisco Pais, 1049-001 Lisboa, Portugal*

Received 13 January 2004; accepted 27 July 2004
Available online 27 October 2004

Abstract

Linear regression metamodels have been widely used to explain the behavior of computer simulation models, although they do not always provide a good global fit to smooth response functions of arbitrary shape. In the case study discussed in this paper, the use of several linear regression polynomial results in a poor fit. The use of a non-linear regression metamodeling methodology provides simple functions that adequately approximate the behavior of the target simulation model. The importance of metamodel validation is emphasized by using the generalization of Rao's test to non-linear metamodels and double cross-validation.

Keywords: Simulation; Regression; Non-linear metamodels; Metamodel validation

1. Introduction

Frequently, the main objective in discrete event simulation studies is the prediction and sensitivity analysis of a system response, for different combinations of a particular set of controllable input variables. However, it is not generally an easy task to interpret the large amounts of data yielded by simulation runs and it becomes increasingly difficult to make decisions about design modifications in the target system (e.g., in queuing systems). Detailed models are particularly valuable for representing explicitly the underlying phenomena. However, much of our knowledge of the world comes from low-resolution sources. Whenever possible, it is more suitable to construct a simple mathematical relationship that relates the inputs and

* Corresponding author. Tel.: +351 2142 33297; fax: +351 2142 33268.
E-mail addresses: prs@asterix.ist.utl.pt (M.I. Reis dos Santos), apnova@ist.utl.pt (A.M.O. Porta Nova).

0377-2217/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2004.07.067

outputs of the computer simulation model; that is, a model of the simulation model—called a *metamodel* (Barton, 1998). Blanning (1974) suggested the use of metamodels to perform sensitivity analysis for all kinds of models in management science and Kleijnen (1975) proposed some statistical tools to make metamodels commonly usable. There are different types of metamodels; for example, polynomial regression models (which uses linear regression), artificial neural networks (one type of neural networks, named regression supervised neural nets, is trained using non-linear regression), splines (which is based on piecewise polynomial functions) and Kriging (which uses interpolation); see Kleijnen (in press), van Beers and Kleijnen (2004), Barton (1994), Hurrion (1992), and Másson and Wang (1990). Metamodels are very useful in simulation optimization and 'what if?' questions, without having to perform additional simulation runs. Also, the simple mathematical expression of a metamodel can expose—more clearly than the simulation model—the fundamental nature of the system input–output relationships.

Traditionally linear regression procedures are frequently used for constructing simulation metamodels. In particular, the general *linear* regression model has been intensively studied—e.g., Kleijnen and Sargent (2000), Panis et al. (1994), Kleijnen (1992), Porta Nova and Wilson (1989) and Kleijnen et al. (1979). However, polynomials are unable to produce a global fit to curves of arbitrary shape. Moreover, in real-life systems, non-linearity is common and the approximation using polynomials becomes unrealistic. Consequently, in these situations, polynomials often fail to provide good fits, namely in problems involving queuing systems (Friedman and Friedman, 1985). An alternative that may provide better and more realistic global fits is non-linear regression; see Santos and Porta Nova (1999, 2001).

However, any metamodel can be used to analyze simulation output only if it is 'good enough'. So, after estimating the metamodel, it is advisable to check if the estimated (fitted) metamodel is, in fact, an accurate representation of the simulation model (Kleijnen and Sargent, 2000). For this purpose, robust statistical validation techniques from non-linear regression are used.

This paper is organized as follows. In Section 2, estimation procedures for the general non-linear metamodel are presented. In Section 3, the issue of metamodel validation is discussed. In Section 4, an actual problem concerning a center for inspecting and repairing automobiles is analyzed and several candidate metamodels, including linear and non-linear ones, are considered. Section 5 gives conclusions and suggestions for further work.

2. Non-linear regression metamodels

A simulation model attempts to describe the relationship between a set of input parameters and variables and the output of the real system. As a consequence, the most important variables and parameters should be selected and represented. A parameter is a quantity that cannot be observed in the real system, whereas a variable is directly observable (Kleijnen and Groenendaal, 1992); customer arrival times is an example of a variable, and the arrival rate of a Poisson process is an example of a parameter. The response of the real system is modeled by the output variable of the simulation program. If the problem has several output variables, separate metamodels can be developed for each output (Kleijnen and Sargent, 2000, p. 15). As a result, the simulation model can be represented by

$$W = \eta(\mathbf{Z}, \mathbf{a}), \tag{1}$$

where W is a univariate response, $\mathbf{Z} = (Z_1, \ldots, Z_k)^{\mathrm{T}}$ is a vector of input variables and $\mathbf{a}$ represents a set of random streams that drive the simulation at $\mathbf{Z}$. For example, in the simulation of a supermarket, the vector $\mathbf{Z}$ can include the mean interarrival time, the mean service time and the number of physical lanes. The response W can be, for example, the delay in the queue or the time in the system.

Assume that the simulation model can be represented by the simulation metamodel

$$Y = f(\mathbf{X}, \boldsymbol{\theta}) + \epsilon, \tag{2}$$

where $\mathbf{X} = (X_1, \ldots, X_d)$ is a vector of d explanatory variables, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^{\mathrm{T}}$ represents a vector of

unknown parameters, ϵ represents the error and f is an unknown function simpler than η (the error ϵ includes both effects of the inadequacy of f as a representation of η, and intrinsic effects encountered in any stochastic simulation model); see Fig. 1. The variable X_i may be the same as the simulation variable Z_i, or a transformation of one or more variables Z_j. For example, in the $M/M/1$ queue, the utilization factor $X = \rho = \lambda/\mu$ (where $Z_1 = \lambda$ is the arrival rate and $Z_2 = \mu$ is the service rate) can be a better explanatory variable than λ and μ. In this paper, f is a non-linear function of the unknown parameter vector $\boldsymbol{\theta}$, so we are dealing with non-linear metamodels. The unknown parameter vector must be estimated.

In the $M/M/1$ queue, an example of a non-linear metamodel is

$$Y = \frac{X\theta_1}{1 - X\theta_2} + \epsilon,$$

where the decision variable $X = \rho$ and the response Y represents the expected number of customers in the queue (Santos and Porta Nova, 1999), whereas a linear metamodel could be

$$Y = \theta X^2 + \epsilon,$$

where Y represents the average waiting time in the queue (Kleijnen and Groenendaal, 1992); or $\ln(Y) = \theta \ \ln X + \varepsilon$ (Friedman and Friedman, 1985).

Suppose that a simulation experiment is performed according to some experimental design, consisting of n different design points, $\{X_{il} : i = 1, \ldots, n; l = 1, \ldots, d\}$. For each design point, r_i independent replications of the simulation model are obtained, so the simulation experiment yields $\{(W_{ij}, \hat{\sigma}_i^2) : i = 1, \ldots, n; j = 1, \ldots, r_i\}$, where W_{ij} is the jth observation at experimental point i and $\hat{\sigma}_i$ is the estimated variance at the design point i, based on r_i observations,

$$\hat{\sigma}_i^2 = \sum_{j=1}^{r} (W_{ij} - \overline{W}_{i.})^2 / [r_i(r_i - 1)], \quad i = 1, \ldots, n. \tag{3}$$

The average

$$\overline{W}_{i.} = \sum_{j=1}^{r} W_{ij}/r_i, \quad i = 1, \ldots, n,$$

is the metamodel response of interest. The simulation model (1) defines a statistic population of observations W. The members of the population correspond to all possible pseudorandom number seeds, that is, in theory the population has an infinite dimension. So, in a classical hypothesis, the population has Gaussian distribution.

This allows us to express the metamodel (2) as

$$Y_{ij} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \epsilon_{ij}, \quad i = 1, \ldots, n, \ j = 1, \ldots, r_i, \tag{4}$$

where $Y_{ij} = W_{ij}$, ϵ_{ij} are independent random variables with $\epsilon_{ij} \sim \mathrm{N}(0, \sigma_i^2)$, and $\sigma_i > 0$. We shall assume the same number of replications per design point, $r_i = r$ $(i = 1, \ldots, n)$. Note that if and only if each input combination is replicated an equal number of times, we may replace the individual observations Y_{ij} by their averages $\overline{Y}_{i.}$, and apply least squares to the vector with the n averages $\overline{Y} = (\overline{Y}_{1.}, \ldots, \overline{Y}_{n.})^{\mathrm{T}}$ (Kleijnen and Groenendaal, 1992, p. 152); since this hypothesis is not restrictive (it is not difficult to verify it using an adequate experimental design), it is possible to use the dimension n instead of $N = \sum_{i=1}^{n} r$.

Before estimation, one or more types of metamodels must be selected. Ideally, the form of the metamodel should be dictated by theoretical considerations. For example, in the $M/M/1$ queuing system, a metamodel for the expected queue length Y might be $Y = \theta_1 X^2/(1 + \theta_2 X)$, where X is the utilization factor $X = \lambda/\mu$. However, in many simulation studies we have little or no idea about the relationship between the simulation response and the decision variables. In these cases, we suggest that the choice of metamodels should be made visually, just like we compare empirical histograms with known density functions for selecting

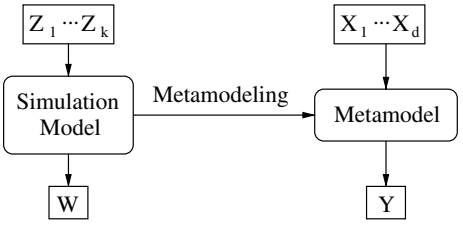

Fig. 1. Simulation model versus metamodel.

a specific random distribution. Following this approach, it is convenient to represent the dispersion diagrams (or scatter plots) of the response versus each independent variable, plotting, for each fixed $l = 1, \ldots, d$, the corresponding pairs (X_{il}, Y_{ij}), where $i = 1, \ldots, n$ and $j = 1, \ldots, r$. Then, we should compare the graphical representations with different analytical curves from a comprehensive catalog, in order to select the best candidates.

2.1. Non-linear metamodel estimation

The metamodel estimation procedure may use the non-linear least squares method for estimating the unknown parameters of the hypothetical metamodel; there are other methods for model estimation like maximum likelihood estimation commonly used in Kriging metamodels (Seber and Wild, 1989, Chapter 2; van Beers and Kleijnen, 2004). The well known non-linear least squares method from regression analysis minimizes the sum of squared errors (SSE)

$$\mathrm{SSE}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{j=1}^{r} [Y_{ij} - f(\mathbf{X}_{i.}, \boldsymbol{\theta})]^2 / \sigma_i^2, \tag{5}$$

that is, it obtains a vector $\hat{\boldsymbol{\theta}}$ such that $\mathrm{SSE}(\hat{\boldsymbol{\theta}}) < \mathrm{SSE}(\boldsymbol{\theta})$, for all $\boldsymbol{\theta}$ in a region of $\mathbb{R}^m$. For most non-linear metamodels, the $\mathrm{SSE}(\boldsymbol{\theta})$ function cannot be minimized analytically and, as a consequence, an iterative numerical method is used. We chose the Levenberg–Marquardt method because the almost unanimous opinion is that, for many non-linear least squares problems, this method works very well. However, in problems with large residuals, Levenberg–Marquardt algorithms may converge unacceptably slowly—or may even not converge at all. In these cases, it is convenient to use numerical methods adapted to each situation—see, for example, Seber and Wild (1989, Section 14.3).

In contrast to the linear case, the $\mathrm{SSE}(\boldsymbol{\theta})$ function in (5) can have several local minima, in addition to the global minimum. For $\hat{\boldsymbol{\theta}}$ to be a local minimum, it is sufficient that: (i) the partial derivatives of $\mathrm{SSE}(\boldsymbol{\theta})$ with respect to $\theta_1, \ldots, \theta_m$ be zero; and (ii) the Hessian matrix of $\mathrm{SSE}(\boldsymbol{\theta})$, calculated at $\hat{\boldsymbol{\theta}}$, be positive definite.

In Proposition 1, the non-linear least squares estimator $\hat{\boldsymbol{\theta}}$ is established. Under the hypothesis that the ϵ_{ij} are independent and normally distributed, $\epsilon_{ij} \sim \mathrm{N}(0, \sigma_i^2)$, and assuming some further regularity conditions, it is shown that $\hat{\boldsymbol{\theta}}$ is asymptotically normally distributed as $N = nr \to \infty$ (see the verification of these results in Appendix A).

Proposition 1. *Given appropriate regularity conditions* (White, 1980) *and for large* $N = nr$, *the least squares estimator of* $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, *in* (4) *satisfies, approximately*:

$$\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^* + [\mathbf{F}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{F}]^{-1} \mathbf{F}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} [\overline{\mathbf{Y}} - \mathbf{f}], \tag{6}$$

$$\hat{\boldsymbol{\theta}} \sim \mathrm{N}_m \left(\mathbf{0}, \frac{1}{r} [\mathbf{F}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{F}]^{-1} \right), \tag{7}$$

where $\boldsymbol{\theta}^*$ *is the exact value of* $\boldsymbol{\theta}$, $\mathbf{f} = \mathbf{f}(\boldsymbol{\theta}^*) = (f(\mathbf{X}_{1.}, \boldsymbol{\theta}^*), \ldots, f(\mathbf{X}_{n.}, \boldsymbol{\theta}^*))^{\mathrm{T}}$, $\mathbf{F} = \mathbf{F}(\boldsymbol{\theta}^*)$ *is the Jacobian matrix of* $\mathbf{f}$, *evaluated at* $\boldsymbol{\theta}^*$, $\overline{\mathbf{Y}} = (\overline{Y}_{1.}, \ldots, \overline{Y}_{n.})^{\mathrm{T}}$ *and* $\boldsymbol{\Sigma}$ *is the diagonal matrix* $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. *In order to simplify the notation, we omit that* $\mathbf{f}$ *and* $\mathbf{F}$ *are evaluated at* $\boldsymbol{\theta}^*$.

Frequently, $\boldsymbol{\Sigma}$ must be replaced by $\widehat{\boldsymbol{\Sigma}} = \mathrm{diag}[\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2]$ and, as a result, we have the estimated generalized non-linear least squares estimator.

3. Metamodel validation

In order to use the estimated metamodel (as a surrogate of the simulation model) for prediction and sensitivity analysis, we have to be sure that it is indeed an accurate representation of the simulation model. For this purpose, we use robust statistical validation techniques from regression analysis. The validation procedure tests the metamodel adequacy and the metamodel validity with respect to the simulation model. To verify the metamodel adequacy, we investigate if the deterministic portion of the metamodel is adequate in the statistical sense and if the predictive capacity of the metamodel is satisfactory. We assert the validity with respect to the simulation model investigating if the metamodel is sufficiently close to the simulation model, taking into account the general

objective of using the metamodel (Kleijnen and Sargent, 2000).

3.1. Metamodel adequacy

To test the adequacy of the metamodel (4), we propose a lack-of-fit test that is an adaptation of Rao's test (Rao, 1959) to non-linear models and has the following requirements: (i) $n > m = \text{rank}(\mathbf{F})$ and $r > n(>n-m)$ (so that $\widehat{\boldsymbol{\Sigma}}$ is non-singular); (ii) the simulation response is normally distributed—in fact, it is enough to require the distribution to be symmetric (Kleijnen and Groenendaal, 1992).

Given the previous requirements (i) and (ii), the generalization of Rao's test has the form:

$$F_{\text{Rao}} = \frac{r(r-n+m)}{(n-m)(r-1)}[\overline{\mathbf{Y}} - \mathbf{f}(\mathbf{X},\hat{\boldsymbol{\theta}})]^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}^{-1}[\overline{\mathbf{Y}} - \mathbf{f}(\mathbf{X},\hat{\boldsymbol{\theta}})] = \frac{r(r-n+m)}{(n-m)(r-1)}\sum_{i=1}^{n}\left[\frac{\overline{Y}_{i.} - f(\mathbf{X}_{i.},\hat{\boldsymbol{\theta}})}{\hat{\sigma}_i}\right]^2. \quad (8)$$

When the metamodel is valid, then F_{Rao} is approximately distributed as an $F_{n-m,r-n+m}$ distribution for large N. Smaller values of F_{Rao} correspond to a better approximation metamodel, consequently an ideal fit corresponds to $F_{\text{Rao}} = 0$.

The predictive validity is verified using double cross-validation and an adaptation of the prediction sum of squares, PRESS; see Neter et al. (1989) and Friedman and Friedman (1985). In our problem, the PRESS statistic has the form

$$\text{PRESS} = \sum_{i=1}^{n}\sum_{j=1}^{r}[Y_{ij} - f(\mathbf{X}_{i.},\hat{\boldsymbol{\theta}}_{(-j)})]^2/\hat{\sigma}_i^2,$$

where $\hat{\boldsymbol{\theta}}_{(-j)}$ is the estimated parameter vector based on the set that we obtain if we delete the jth replication in all experimental points. Other useful statistics are the error sum of squares, $\text{SSE}(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\sum_{j=1}^{r}[Y_{ij} - f(\mathbf{X}_{i.},\hat{\boldsymbol{\theta}})]^2/\hat{\sigma}_{i.}^2$, and the mean sum of squares, $\text{MSE} = \text{SSE}(\hat{\boldsymbol{\theta}})/(N-m)$; in order to simplify the notation, we will write SSE instead of $\text{SSE}(\hat{\boldsymbol{\theta}})$.

3.2. Double cross-validation with respect to the simulation model

We suggest the use of the double cross-validation method to validate the metamodel with respect to the simulation model. In double cross-validation, we split the data intuitively into two subsets of, approximately, the same size ($n/2$). Then, a regression metamodel is developed for each subset and used for prediction of the other subset of the data. In particular, for each metamodel, two values of the coefficient of determination, R^2, are calculated: the first one, R^2_{bui}, is based on the observations from the subset used to build it, and the second one, R^2_{val}, is based on the remaining observations, for validation purposes. Moreover, we compare the parameter estimators of both metamodels.

3.3. Confidence intervals

After validation, and only if the validation tests do not reject the regression metamodel, we can build confidence intervals for the unknown metamodel parameters. We propose the following approximated $1-\alpha$ two-sided confidence rectangle for the estimated parameter $\hat{\theta}_k$ $(k = 1,\ldots,m)$:

$$\hat{\theta}_k \pm t^{\alpha/(2m)}_{r-n+m-1}[\widehat{\text{Var}}[\hat{\theta}_k]]^{1/2} \times \left[\frac{1 + F_{\text{Rao}}(n-m)/(r-n+m)}{1+(n-m)/(r-1)}\right]^{1/2},$$

where $\widehat{\text{Var}}[\hat{\theta}_k]$ is the estimated variance of $\hat{\theta}_k$ calculated by

$$\widehat{\text{Var}}[\hat{\theta}_k] = (r\widehat{\mathbf{F}}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{F}})^{-1}_{kk},$$

(see Proposition 1) and $F_{\text{Rao}} = F_{n-m,r-n+m}$ in (8). This confidence rectangle is obtained adapting (3.13) of Kleijnen (1992) to non-linear simulation metamodels and, then, applying the Bonferroni method.

4. Application: An inspection and repair center

In this paper, we analyze a car inspection and repair center. The car interarrival times are normally distributed with mean μ and a variance of 15; negative numbers correspond to simultaneous events ('censured' normal distribution). Only one inspector services the cars; the time that he needs to inspect one car is uniformly distributed between

15 and 25 minutes. In the inspection queue, space is available for only six cars. On the average, 85% of the cars pass the inspection and leave the center. The other 15% must go to the repair section, where two mechanics work side-by-side. After being repaired, the cars have to go back to the inspection queue. The time required to repair a vehicle is exponentially distributed with a mean of 60 minutes.

Our goal is to express the average time in the system, Y (response), as a function of the mean time between arrivals, μ (decision variable). We considered 14 combinations of simulation input, $\{\mu_i : i = 1, 14\} = \{1, 5, 10, 15, 20, 23, 26, 29, 32, 35, 40, 50, 60, 90\}$, unevenly spaced to take into account the different rates of variation of the output. At each design point, we ran Welch's procedure (Welch, 1983), in order to determine adequate run durations and points for initial-data deletion. Welch's moving averages are based on 20 replications of the simulation model, where each replication contains 2000 observations, that is,

$$\bar{\mu}(j, W) = \begin{cases} r(2W+1)^{-1} \sum_{s=-W}^{W} \sum_{i=1}^{r} Y_{i,j+s}, \\ \quad \text{if } j \geqslant W+1, \\ r(2j-1)^{-1} \sum_{s=-(j-1)}^{j-1} \sum_{i=1}^{r} Y_{i,j+s} \\ \quad \text{if } j < W+1, \end{cases}$$

where W is Welch's window. For example, at the design point $\mu_i = 10$, we deleted 100 observations from the beginning of the run and we used only the remaining 600 observations to estimate the response Y; see Table 1.

We carried out $r = 30$ replications of each of the $n = 14$ design points; in order to apply Rao's validation test, r must be greater than n, and since r is greater than nine, we can obtain an appropriate estimate for $\hat{\sigma}_i$, $i = 1, \ldots, n$ (Deaton et al., 1983).

With the objective of identifying a curve that might fit the input–output relationship of the simulation program, we built the corresponding dispersion diagram, plotting the pairs $\{(X_i, Y_{ij}) : i = 1, \ldots, n, j = 1, \ldots, r\}$; see Fig. 2. Then, we visually checked the scatter plot, comparing it

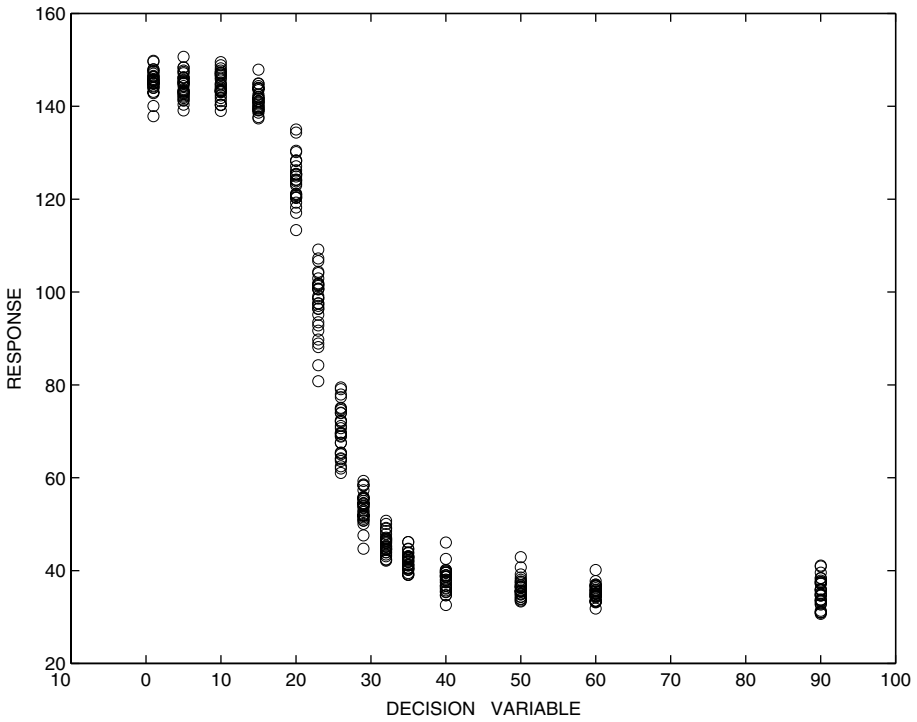

Fig. 2. Visualization of simulation results.

Table 1
Initial data deletion

μ_i	Observations		Welch's window
	Deleted	In run	
1, 5, 10	100	700	50
15	150	1000	100
20	200	1400	150
23, 26, 29, 32, 35	200	1400	300
40, 50, 60	100	700	200
90	50	350	200

with graphical representations of some functional relationships from an appropriate catalog; the non-linear curves that seemed to be good candidates are represented in Fig. 3. 'Arctan' is based on the arc tangent function and the others are three sigmoidal growth models (Seber and Wild, 1989, pp. 329, 338 and 340): 'Logistic' is the logistic model, 'Weibull' is Weibull's growth curve and 'MMF' is the Morgan–Mercer–Flodin family. We also considered polynomial functions of degree r, with $r = 2, \ldots, 10$.

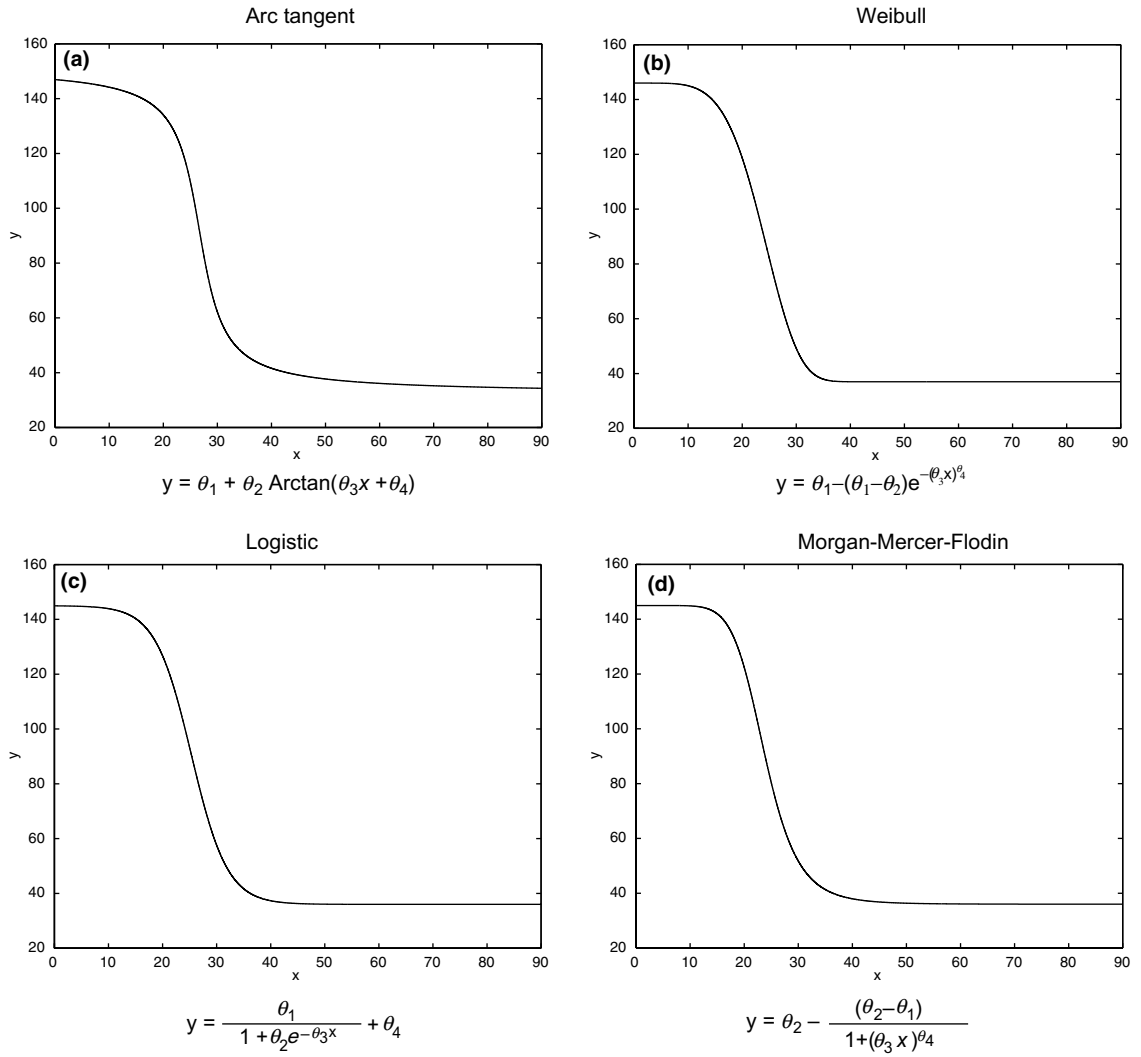

Fig. 3. Candidate non-linear functional relationships.

4.1. Estimating and validating the metamodel

Before estimating the metamodel parameters, we must first check if the response has a constant variance across design points. We measure the variance heterogeneity through the quantity

$$\text{het} = \frac{\max_{i=1,n}\hat{\sigma}_i}{\min_{i=1,n}\hat{\sigma}_i}$$

Table 2
Metamodel diagnostics

Metamodel	SSE	MSE	PRESS	SSE/PRESS
Arctan	494.561	1.23025	510.829	0.968
logistic	537.977	1.33825	554.005	0.971
MMF	437.985	1.08951	454.28	0.964
Weibull	734.765	1.82777	751.049	0.978
pol2	15147.5	37.5867	15162.9	0.998
pol3	15090.1	37.5376	15105.4	0.998

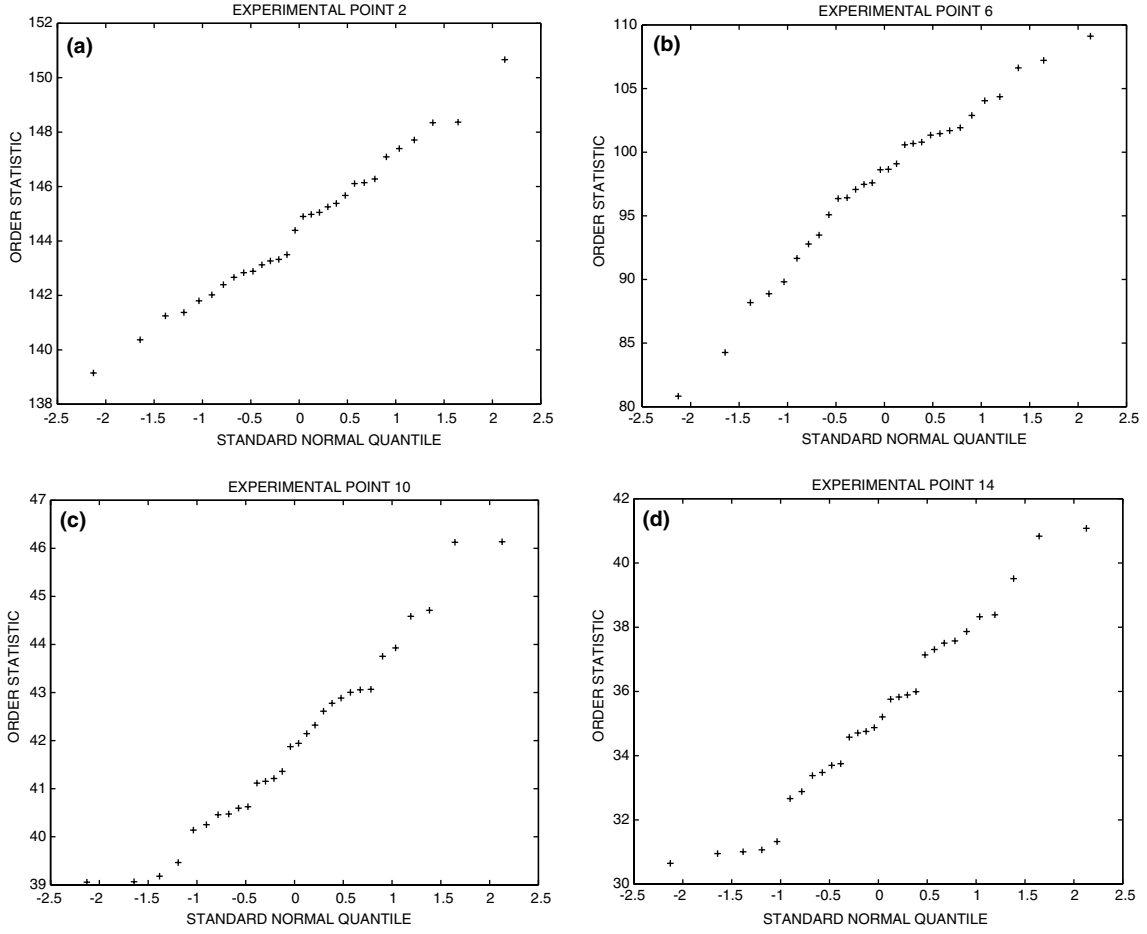

Fig. 4. Normal probability plots (experimental points 2, 6, 10 and 14).

(Kleijnen, 1992). We obtained het = 3.946 (quite different from 1), and so we have heterogeneous variances. Thus, we will use non-linear weighted least squares for non-linear curves, and weighted least squares for polynomials. The non-linear parameter estimators were obtained using the Levenberg–Marquardt method implemented in MATLAB, with the termination tolerance equal to 10^{-6} and the maximum number of function evaluations equal to 600 (the default is $100 \times$ the number of parameters). When we tried to fit polynomials of degree r, with $r = 4, \ldots, 10$, we obtained matrices that were close to singular or badly scaled. Since the results might be highly inaccurate, these metamodels were rejected.

To check the validity of the remaining metamodels, we evaluated the statistics presented in Section 3. The SSE and PRESS statistics are quite close, so MSE is a reasonably valid indicator of the selected model's predictive capability; see Table 2. As a consequence, we conclude that polynomial models have lack of predictive validity, and so they are not good approximations for the target simulation model.

Before using Rao's test, it is convenient to verify if the simulation responses are normally distributed. Since the variance depends on the design point ($i = 1, \ldots, n$) normal probability plots for their original simulation responses, $\{Y_{ij}: j = 1, \ldots, r\}$, were obtained. All of the resulting graphics appear to be nearly linear, but the slope varies with the corresponding design point; see Fig. 4. This agrees with the fact that the value of het is quite different from 1. Thus, there is no evidence to reject the normality of the response, at each design point, with the variance depending on the design point. As a result, Rao's test can be used to select the metamodel that better approximates the simulation results, comparing the F_{Rao} values with the F critical value, $F^{1-\alpha}_{n-m,r-n+m} = F^{0.95}_{10,20} = 2.348$. The elected metamodel, according to this criterion, is the one based on the MMF curve—all others are rejected; see Table 3. Also, we observe that MMF is the model that has the smallest PRESS in the Table 2 (i.e., MMF is the curve that has the smallest prediction errors).

To gain more insight into the predictive validity of the metamodels, we analyzed the results of double cross-validation; see Table 4. In each model,

Table 3
Rao's test

Metamodel	Arctan	Logistic	MMF	Weibull
F_{Rao}	6.108	9.102	2.206	22.673

Table 4
Double cross-validation test

Coefficient	Arctan		Logistic	
	Subset 1	Subset 2	Subset 1	Subset 2
$\hat{\theta}_1$	91.494	92.474	−109.0	−109.7
$\hat{\theta}_2$	−37.9318	−38.4031	1925.8	2193.7
$\hat{\theta}_3$	0.3296	0.3213	0.3	0.3
$\hat{\theta}_4$	−7.8454	−7.6333	145.4	146.4
R^2_{bui}	0.9922	0.9971	975.5	0.9882
R^2_{val}	0.9786	1.0113	989.2	1.0023
	MMF		Weibull	
	Subset 1	Subset 2	Subset 1	Subset 2
$\hat{\theta}_1$	144.58	145.29	36.9817	37.427
$\hat{\theta}_2$	35.681	35.855	145.50	146.45
$\hat{\theta}_3$	0.0422	0.0420	0.0386	0.0385
$\hat{\theta}_4$	7.7327	7.7974	4.9014	5.0609
R^2_{bui}	1.0003	0.9934	0.9841	0.9828
R^2_{val}	0.9866	1.0073	0.9706	0.9969

Table 5
95% Confidence intervals for individual parameters

Metamodel coefficient	Estimator $\hat{\theta}$	Standard deviation	Confidence interval
θ_1	144.882	0.1185	144.882 ± 0.3067
θ_2	35.797	0.0761	35.797 ± 0.1968
θ_3	0.0421	0.0001	0.0421 ± 0.0003
θ_4	7.776	0.0704	7.776 ± 0.1821

we observe a good agreement between the coefficients obtained based on subsets 1 and 2. Also, the coefficients of determination are quite similar.

Finally, we obtained confidence intervals for the individual metamodel parameters, with coverage probability $1-\alpha = 0.95$; see Table 5. The standard deviations of the estimators are also shown. We observe that the standard deviations and the confidence interval half-lengths have quite small values, compared with the absolute values of the estimators. These are good indicators of the precision of the regression estimators obtained in this work.

5. Conclusions

This paper stresses the importance of using valid reliable non-linear metamodels in simulation studies. In the example discussed here, a poor fit was obtained when various polynomial metamodels were tried, leading to a demand for more precise and flexible models. Linear models are considerably simpler to fit than non-linear ones, but they are unable to ensure a global fit to curves of arbitrary shape. Non-linear regression metamodels are attractive, because they do not have this limitation, allowing to complex curves.

It is generally much more convenient to have a ready-to-use and reliable metamodel, rather than a more expensive and hard to calibrate simulation model. In order to ensure that a specific metamodel provides an adequate substitute for the simulation model, a series of adequacy tests must be performed. If any one of these tests fails, the metamodel is rejected.

The use of non-linear metamodels requires an extensive catalog of curves and more complex and time consuming regression software. The selection of good candidate curves for the fitting process influences dramatically the resulting metamodel precision, as shown in the example. However, once a comprehensive catalog of curves is provided, the choice of an adequate metamodel is quite straightforward. The regression and validation software can be repeatedly used, as soon as the user supplies a trial function and an initial solution. Finally, the increased computation time when compared to linear regression procedures, is becoming less important with the ever growing computing power of personal computers. Moreover, the computation time required for obtaining non-linear regression metamodels can be orders of magnitude smaller than the time needed to run and analyze the actual simulation model.

Appendix A

(i) *Verification of (6).* Consider the usual decomposition

$$\begin{aligned}\mathrm{SSE}(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\sum_{j=1}^{r}\frac{1}{\sigma_i^2}\left[Y_{ij}-f(\mathbf{X}_{i.},\boldsymbol{\theta})\right]^2 \\ &= \sum_{i=1}^{n}\sum_{j=1}^{r}\frac{1}{\sigma_i^2}\left(Y_{ij}-\overline{Y}_{i.}\right)^2 \\ &\quad + r\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\left[\overline{Y}_{i.}-f(\mathbf{X}_{i.},\boldsymbol{\theta})\right]^2,\end{aligned}$$

where $\sum_{i=1}^{n}\sum_{j=1}^{r}(Y_{ij}-\overline{Y}_{i.})^2/\sigma_i^2$ is the pure error sum of squares.

To find the least squares estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, $\mathrm{SSE}(\boldsymbol{\theta})$ must be minimized with respect to $\boldsymbol{\theta}$, which is equivalent to minimizing

$$\sum_{i=1}^{n}\frac{r}{\sigma_i^2}\left[\overline{Y}_{i.}-f(\mathbf{X}_{i.},\boldsymbol{\theta})\right]^2. \tag{9}$$

But (9) is the SSE ($\boldsymbol{\theta}$) of the problem

$$\mathbf{W} = \mathbf{g}(\mathbf{X},\boldsymbol{\theta}) + \boldsymbol{\eta}, \tag{10}$$

where $\mathbf{W} = \mathbf{R}\overline{\mathbf{Y}}$, $\mathbf{g}(\mathbf{X},\boldsymbol{\theta}) = \mathbf{R}\mathbf{f}(\mathbf{X},\boldsymbol{\theta})$ and $\boldsymbol{\eta} = \mathbf{R}\boldsymbol{\epsilon}$, with $\mathbf{R} = (\mathbf{U}^{\mathrm{T}})^{-1}$ where $\mathbf{U}$ is obtained from the decomposition $\boldsymbol{\Sigma} = \mathbf{U}^{\mathrm{T}}\mathbf{U}$ (assumption: $\boldsymbol{\Sigma}$ is a symmetric positive definite matrix, that accepts the Cholesky decomposition, $\boldsymbol{\Sigma} = \mathbf{U}^{\mathrm{T}}\mathbf{U}$) and $\boldsymbol{\epsilon} = (\bar{\epsilon}_{1.},\ldots,\bar{\epsilon}_{n.})^{\mathrm{T}}$.

Then, we observe that $E[\boldsymbol{\eta}] = \mathbf{0}$ and $\mathrm{Var}[\boldsymbol{\eta}] = \mathbf{R}\mathrm{Var}[\boldsymbol{\epsilon}]\mathbf{R}^{\mathrm{T}} = 1/r\mathbf{R}\boldsymbol{\Sigma}\mathbf{R}^{\mathrm{T}}$. But $\boldsymbol{\Sigma} = \mathbf{U}^{\mathrm{T}}\mathbf{U}$ (Cholesky decomposition) and $\mathbf{R} = (\mathbf{U}^{\mathrm{T}})^{-1}$. Thus, $\mathrm{Var}[\boldsymbol{\eta}] = 1/r(\mathbf{U}^{\mathrm{T}})^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{U}[(\mathbf{U}^{\mathrm{T}})^{-1}]^{\mathrm{T}} = 1/r\mathbf{I}_n$, where $\mathbf{I}_n$ is the identity matrix of order n. We conclude that the problem (10) is an ordinary least squares problem. Thus, the least squares estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^* + [\mathbf{G}^{\mathrm{T}}\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}[\mathbf{W} - \mathbf{g}], \tag{11}$$

where $\mathbf{G} = \partial\mathbf{g}(\mathbf{X}, \boldsymbol{\theta})/\partial\boldsymbol{\theta}^{\mathrm{T}}$ is the Jacobian matrix of $\mathbf{g}$, and we omit, in our notation, that both $\mathbf{g}$ and $\mathbf{G}$ are evaluated at $\boldsymbol{\theta}^*$; see Theorem 2.1 in Seber and Wild (1989).

But, since $\mathbf{g}(\mathbf{X}, \boldsymbol{\theta}) = \mathbf{R}\mathbf{f}(\mathbf{X}, \boldsymbol{\theta})$, we have $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{R}\partial\mathbf{f}(\mathbf{X}, \boldsymbol{\theta})/\partial\boldsymbol{\theta}^{\mathrm{T}} = \mathbf{R}\mathbf{F}(\boldsymbol{\theta})$. Besides, $\mathbf{W} = \mathbf{R}\overline{\mathbf{Y}}$ and $\mathbf{R}^{\mathrm{T}}\mathbf{R} = \boldsymbol{\Sigma}^{-1}$, therefore (11) is equivalent to

$$\begin{aligned}\hat{\boldsymbol{\theta}} &\approx \boldsymbol{\theta}^* + [\mathbf{F}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}\mathbf{R}\mathbf{F}]^{-1}(\mathbf{R}\mathbf{F})^{\mathrm{T}}[\mathbf{R}\overline{\mathbf{Y}} - \mathbf{R}\mathbf{f}(\mathbf{X}, \boldsymbol{\theta}^*)] \\ &= \boldsymbol{\theta}^* + [\mathbf{F}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{F}]^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}\mathbf{R}[\overline{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta}^*)] \\ &= \boldsymbol{\theta}^* + [\mathbf{F}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{F}]^{-1}\mathbf{F}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}[\overline{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta}^*)].\end{aligned}$$

Thus, the approximate result (6) has been established.

(ii) *Verification of (7).* The result (7) is obtained by applying Theorem 2.1 in Seber and Wild (1989), item (*i*), to the problem (11): $\hat{\boldsymbol{\theta}} \sim \mathrm{N}_p[\boldsymbol{\theta}, 1/r(\mathbf{G}^{\mathrm{T}}\mathbf{G})^{-1}]$. Since $\mathbf{G} = \mathbf{R}\mathbf{F}$ and $\mathbf{R}^{\mathrm{T}}\mathbf{R} = \boldsymbol{\Sigma}^{-1}$, we obtain $\hat{\boldsymbol{\theta}} \sim \mathrm{N}_p[\boldsymbol{\theta}, 1/r(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}\mathbf{R}\mathbf{F})^{-1}]$, and then $\hat{\boldsymbol{\theta}} \sim \mathrm{N}_p[\boldsymbol{\theta}, 1/r(\mathbf{F}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{F})^{-1}]$.

References

Barton, R.R., 1998. Simulation metamodels. In: Medeiros, D.J., Watson, E.F., Carson, J.S., Manivannan, M.S. (Eds.), Proceedings of Winter Simulation Conference, pp. 167–174.

Barton, R.R., 1994. Metamodeling: A state of the art review. In: Tew, J.D., Manivannan, S., Sadowski, D.A., Seila, A.F. (Eds.), Proceedings of Winter Simulation Conference, pp. 237–244.

Blanning, R.W., 1974. The sources and uses of sensitivity information. Interfaces 4 (4), 32–38.

Deaton, M.L., Reynolds, M.R., Myers, R.H., 1983. Estimation and hypothesis testing in regression in the presence of non-homogeneous error variances. Communications in Statistics B 12 (1), 45–66.

Friedman, L.W., Friedman, H.H., 1985. Validating the simulation metamodel: Some practical approaches. Simulation 45 (3), 144–146.

Hurrion, R.D., 1992. Using a neural network to enhance the decision making quality of a visual interactive simulation model. Journal of the Operational Research Society 43 (4), 333–341.

Kleijnen, J.P.C., in press. An overview of the design and analysis of simulation experiments for sensitivity analysis. European Journal of Operational Research. PII:S0377-2217(04)00083-9.

Kleijnen, J.P.C., 1992. Regression metamodels for simulation with common random numbers: Comparison of validation tests and confidence intervals. Management Science 38 (8), 1164–1184.

Kleijnen, J.P.C., 1975. A comment on Blanning's metamodel for sensitivity analysis: The regression metamodel in simulation. Interfaces 5 (3), 21–23.

Kleijnen, J.P.C., van den Burg, A.J., van der Ham, R.Th., 1979. Generalization of simulation results. Practicality of statistical methods. European Journal of Operational Research 3, 50–64.

Kleijnen, J.P.C., Groenendaal, W.V., 1992. Simulation: A Statistical Perspective. John Wiley and Sons, New York, USA.

Kleijnen, J.P.C., Sargent, R.G., 2000. A methodology for fitting and validating metamodels in simulation. European Journal of Operational Research 120, 14–29.

Másson, E., Wang, Y., 1990. Invited review: Introduction to computation and learning in artificial neural networks. European Journal of Operational Research 47, 1–28.

Neter, J., Wasserman, W., Kutner, M.H., 1989. Applied Linear Regression Models, second ed. Richard D. Irwin Inc.

Panis, R.P., Myers, R.H., Houck, E.C., 1994. Combining regression diagnostic with simulation metamodels. European Journal of Operational Research 13, 133–141.

Porta Nova, A.M., Wilson, J.R., 1989. Estimation of multiresponse simulation metamodels using control variates. Management Science 35 (11), 1316–1333.

Rao, C.R., 1959. Some problems involving linear hypotheses in multivariate analysis. Biometrika 46, 49–58.

Santos, M.I., Porta Nova, A.M., 2001. Validation of non-linear simulation metamodels. In: Hamza, M.H. (Ed.), Proceedings of Applied Simulation and Modeling, pp. 421–425.

Santos, M.I., Porta Nova, A.M., 1999. The main issues in non-linear simulation metamodel estimation. In: Farrington, P.A., Nembhard, H.B., Sturrock, D.T., Evans, G.W. (Eds.), Proceedings of Winter Simulation Conference, pp. 502–509.

Seber, G.A.F., Wild, C.J., 1989. Nonlinear Regression. John Wiley and Sons, New York, USA.

van Beers, W.C.M., Kleijnen, J.P.C., 2004. Kriging interpolation in simulation: A survey. In: Ingalls, R.G., Rossetti, M.D., Smith, J.S., Peters, B.A. (Eds.), Proceedings of Winter Simulation Conference.

Welch, P.D., 1983. The statistical analysis of simulation results. In: Lavenberg, S.S. (Ed.), Computer Performance Modeling Handbook. Academic Press, New York, pp. 268–328.

White, H., 1980. Nonlinear regression on cross-section data. Econometrica 48 (3), 721–746.