Communications in Statistics—Simulation and Computation®, 36: 123–137, 2007

ISSN: 0361-0918 print/1532-4141 online
DOI: 10.1080/03610910601096288

Taylor & Francis
Taylor & Francis Group

Regression Analysis

Estimating and Validating Nonlinear Regression Metamodels in Simulation

M. ISABEL REIS DOS SANTOS AND
ACÁCIO M. O. PORTA NOVA

Department of Mathematics and Centre for Management Studies,
Technical University of Lisbon (IST), Lisboa, Portugal

Frequently, the main objective of statistically designed simulation experiments is to estimate and validate regression metamodels, where the regressors are functions of the design variables and the dependent variable is the system response. In this article, a weighted least squares procedure for estimating the unknown parameters of a nonlinear regression metamodel is formulated and evaluated. Since the validity of a fitted regression model must be tested, a method for validating nonlinear regression simulation metamodels is presented. This method is a generalization of the cross-validation test proposed by Kleijnen (1983) in the context of linear regression metamodels. One drawback of the cross-validation strategy is the need to perform a large number of nonlinear regressions, if the number of experimental points is large. In this article, cross-validation is implemented using only one nonlinear regression. The proposed statistical analysis allows us to obtain Scheffé-type simultaneous confidence intervals for linear combinations of the metamodel's unknown parameters. Using the well-known $M/M/1$ example, a metamodel is built and validated with the aid of the proposed procedure.

Keywords Cross-validation; Nonlinear least squares; Parameter estimation; Simulation.

Mathematics Subject Classification Primary 93E03; Secondary 93E24.

1. Introduction

Digital simulation is a commonly used tool in the analysis of complex systems. However, running a simulation program for a large number of input combinations can be very time-consuming and it may become very hard to interpret the statistical results. Whenever possible, it is preferable to use a simple mathematical model to relate the system response with the decision variables, that is, a model of the

Received December 21, 2005; Accepted July 6, 2006
Address correspondence to M. Isabel Reis Dos Santos, Department of Mathematics and Centre for Management Studies, Technical University of Lisbon (IST), Av. Rovisco Pais, 1049-001 Lisboa, Portugal; E-mail: maria.santos@tagus.ist.utl.pt

simulation model, or a *metamodel* (Barton, 1998). Metamodels are very useful for prediction and sensitivity analysis of the system response for different factor settings. Besides, a metamodel can expose more clearly, than the simulation model, the essential nature of the system input–output relationship.

Simulation metamodels have received a lot of attention in recent years. The most widely used ones have been *linear* regression metamodels (e.g., Friedman and Friedman, 1985; Kleijnen, 1983, 1992; Kleijnen et al., 1979; Porta Nova and Wilson, 1989; Panis et al., 1994). A strong reason for that choice is the availability, in many statistical software packages, of analysis tools to obtain estimators for the unknown parameters of the metamodel, characterize the metamodel fit, and use the metamodel for prediction. In real-life problems (e.g., problems involving queueing systems), however, the relationship between the response and the input variables is usually nonlinear, and polynomials do not have the ability to provide a global fit to curves of arbitrary shape. As a consequence, polynomials often have poor fits in that kind of problems. An alternative that usually produces more realistic and better fits, is the use of *nonlinear* regression metamodels. The nonlinear metamodels can be estimated using nonlinear least squares, assuming non homogeneous error variances, and can be validated using statistical tools (Santos and Porta Nova, 1999). Additional validation tests are applied to a case study (Santos and Porta Nova, 2001). Instead least squares criterion, maximum likelihood estimation may be used. However, with normal errors, the maximum likelihood estimate is the same as the least squares estimate (Seber and Wild, 1989, p. 576). In Santos and Porta Nova (2006), the previous work is combined and augmented with statistical confidence intervals. A nonlinear metamodel regression methodology is introduced in Santos et al. (2006), including a statistical test for variance heterogeneity, estimation assuming constant error variances, validation with respect to the real system, and graphical analysis of the residuals. It also introduces normal probability slopes as measures, and extends the curve selection from a catalog, as suggested in Santos and Porta Nova (2006). The methodology is fully implemented in Matlab as an automated process.

Simulation can be used for constructing nonlinear regression metamodels that appropriately describe the behavior of real systems. The unknown parameters of the regression metamodels are estimated using the weighted least squares method. Before a metamodel can be used to make decisions, it must be validated. Kleijnen (1983) proposed a cross-validation test for validating linear regression metamodels in simulation. One drawback of this method, observed by Kleijnen, is the need to obtain a large number of regressions. Panis et al. (1994), however, showed that the test can be implement using only a single regression. They used a relationship between the usual regression residuals and the deleted residuals, that involves the diagonal elements of the *hat* matrix, to speed up the computation in the cross-validation method. For nonlinear metamodels, the referred drawback becomes even more important since each regression requires an iterative numerical method for solving nonlinear systems of equations (e.g., the Levenberg–Marquardt method). In this article, the Kleijnen cross-validation test is adapted to nonlinear metamodels, and it is shown that, also in this case, we need only one regression to implement the method.

This article is organized as follows. In Sec. 2 the general nonlinear regression metamodel is presented. The cross-validation test for this metamodel and confidence intervals for the metamodel's unknown parameters are developed in Sec. 3. In Sec. 4

it is shown how the cross-validation test can be implement with only a single fitted nonlinear regression model. In Sec. 5 numerical results for an $M/M/1$ queueing application are discussed. The validation techniques of Sec. 3 are used to validate a metamodel constructed with results from the simulation program that represents this system. Also, the validation with respect to the real system is performed. Section 6 is reserved for conclusions.

2. Nonlinear Regression Metamodels in Stochastic Simulation

Frequently, the purpose of a simulation experiment is to estimate a metamodel for a selected response, that is, a mathematical model of the expected response, expressed in terms of relevant decision variables. The metamodel can be a linear or nonlinear function of the unknown coefficients. An example of a metamodel preferred by some simulation practitioners is the following linear regression metamodel

$$Y = \theta_0 + \sum_{k=1}^{d} \theta_k X_k + \epsilon.$$

with d decision variables (Kleijnen and van Groenendaal, 1992). An example of a nonlinear metamodel is the logistic model

$$Y = \theta_1/(1 + \theta_2\ e^{-\theta_3 X}) + \epsilon,$$

with one decision variable, X (Seber and Wild, 1989).

Suppose that the underlying simulation model has k factors or decision variables. Suppose further that each simulation run yields a single response, denoted by W; W can represent, for example, the average time in the system. Although simulation designs are usually used to generate multivariate response outputs, in practice, each output is analyzed individually (Kleijnen et al., 2005).

A simulation experiment is performed according to some experimental design, consisting of n different factor combinations or design points, $\{Z_{il} : i = 1, \ldots, n; l = 1, \ldots, d\}$. Each design point is replicated r_i times using non overlapping, pseudo-random number streams. Thus, the simulation experiment yields $N = \sum_{i=1}^{n} r_i$ independent observations, $\{W_{ij}, i = 1, \ldots, n; j = 1, \ldots, r_i\}$; W_{ij} denotes the jth observation at the ith experimental point. We assume that no common random numbers are used. Common random numbers may increase the efficiency but they also complicate the regression analysis (Kleijnen, 1987, p. 170). For simplicity's purpose it is assumed that all experimental points are replicated an equal number of times ($r_i = r$); that is, factor combinations with higher variability are not replicated more often. Let $\bar{\mathbf{W}} = (\overline{W}_{1.}, \ldots, \overline{W}_{n.})^T$ be the average simulation output vector, containing the sample averages of the response,

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

This allows us to formulate a nonlinear metamodel of the form

$$Y_i = f(\mathbf{X}_{i.}, \boldsymbol{\theta}^*) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where ϵ_i are independent random variables with $\epsilon_i \sim \mathrm{N}(0, \sigma_i^2/r)$, $\sigma_i = \sqrt{\mathrm{Var}[W_{ij}]} > 0$, and Y_i is the metamodel prediction for the expected simulation response $\mathrm{E}[\overline{W}_{i.}]$.

Each variable X_i can either be identical to a variable in the set $\{Z_1, \ldots, Z_k\}$ or a transformation of one or more elements of this set; for example, in the $M/M/1$ queue system it may be useful to consider $X = Z_1/Z_2$, where Z_1 and Z_2 are the arrival and the service rates, respectively. We assume that the variance σ_i^2, $i = 1, \ldots, n$, is estimated by

$$\hat{\sigma}_i^2 = \sum_{j=1}^{r} (W_{ij} - \overline{W}_{i.})^2/(r-1). \tag{2}$$

The metamodel's unknown parameters are estimated using nonlinear least squares, that is, $\hat{\boldsymbol{\theta}}$ is the vector such that $S(\hat{\boldsymbol{\theta}}) < S(\boldsymbol{\theta})$ over $\boldsymbol{\theta} \in \boldsymbol{\Theta} \in \Re^m$, where $S(\boldsymbol{\theta})$ is the error sum of squares $S(\boldsymbol{\theta}) = \sum_{i=1}^{n} r[\overline{W}_{i.} - f(\mathbf{X}_{i.}, \boldsymbol{\theta})]^2/\sigma_i^2$. For large N and given appropriate regularity conditions, the nonlinear weighted least squares estimator of $\boldsymbol{\theta}$ in (1) is approximately (Santos and Porta Nova, 2006)

$$\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^* + \left[\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\right]^{-1}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\left[\overline{\mathbf{W}} - \mathbf{f}\right], \tag{3}$$

where $\boldsymbol{\theta}^*$ is the exact value of $\boldsymbol{\theta}$, $\mathbf{f} = \mathbf{f}(\boldsymbol{\theta}^*) = (f(\mathbf{X}_{1.}, \boldsymbol{\theta}^*), \ldots, f(\mathbf{X}_{d.}, \boldsymbol{\theta}^*))^T$ defined in (1), $\mathbf{F} = \mathbf{F}(\boldsymbol{\theta}^*)$ is the Jacobian matrix of $\mathbf{f}$ evaluated at $\boldsymbol{\theta}^*$, $\boldsymbol{\Sigma}$ is the diagonal matrix $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, and $\hat{\boldsymbol{\theta}}$ is asymptotically normally distributed

$$\hat{\boldsymbol{\theta}} \sim \mathrm{N}_m\left(\boldsymbol{\theta}^*, \frac{1}{r}\left[\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\right]^{-1}\right). \tag{4}$$

In practice, $\boldsymbol{\Sigma}$ must be replaced by

$$\widehat{\boldsymbol{\Sigma}} = \text{diag}\left[\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2\right] \tag{5}$$

and, as a result, the estimated generalized nonlinear least squares estimator is obtained. The consequences of replacing the deterministic matrix $\boldsymbol{\Sigma}$ by its estimator will be discussed in the next section.

3. Validation and Confidence Interval Procedures

After the metamodel's unknown parameters have been estimated, the metamodel can be used for future work only if it acceptably represents the behavior of the input/output transformation that produced the simulation data. In order to assess the adequacy of the fitted metamodel, a lack-of-fit test is proposed. This test is an adaptation, to nonlinear regression metamodels, of the cross-validation test presented by Kleijnen (1983).

3.1. *Cross-Validation Procedure*

The cross-validation procedure presented here is based on the comparison of predicted values, obtained using the fitted metamodel, with observations that were not used for metamodel estimation. For nonlinear regression metamodels this validation test consists of the following steps:

1. Consider the data set that is obtained by removing, from the original data set, information corresponding to the ith experimental point, that is,

$\{(\mathbf{X}_{i.}, W_{ij}) : j = 1, \ldots, r\}$. Based on this data set, obtain the corresponding nonlinear weighted least squares estimator, denoted by $\hat{\boldsymbol{\theta}}_{-i}$.

2. Predict the response for the experimental point i, using $\hat{\boldsymbol{\theta}}_{-i}$, that is, (1) gives

$$\widehat{Y}_{i,-i} = f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}}_{-i}).$$

Since W_{ij} $(j = 1, \ldots, r)$ is not used in the estimation in Step (1), $\widehat{Y}_{i,-i}$ and $\overline{W}_{i.}$ are statistically independent and, as a consequence,

$$\mathrm{Var}[\overline{W}_{i.} - \widehat{Y}_{i,-i}] = \mathrm{Var}[\overline{W}_{i.}] + \mathrm{Var}[\widehat{Y}_{i,-i}],$$

where $\mathrm{Var}[\overline{W}_{i.}] = \sigma_i^2/r$. In order to obtain the variance of the predicted value $\widehat{Y}_{i,-i}$, we use Taylor's series expansion of $f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}}_{-i})$ about the point $(\mathbf{X}_{i.}, \boldsymbol{\theta}^*)$ and write

$$\widehat{Y}_{i,-i} \approx f(\mathbf{X}_{i.}, \boldsymbol{\theta}^*) + \mathbf{f}_i^T(\hat{\boldsymbol{\theta}}_{-i} - \boldsymbol{\theta}^*),$$

where $\mathbf{f}_i^T = \mathbf{f}_i^T(\boldsymbol{\theta}^*)$. As a result, we have $\mathrm{Var}[\widehat{Y}_{i,-i}] \approx \mathbf{f}_i^T \mathrm{Var}[\hat{\boldsymbol{\theta}}_{-i}]\mathbf{f}_i$. Since (4) applies, we obtain

$$\mathrm{Var}\big[\widehat{Y}_{i,-i}\big] \approx \frac{1}{r}\mathbf{f}_i^T\big[\mathbf{F}_{-i}^T\boldsymbol{\Sigma}_{-i}^{-1}\mathbf{F}_{-i}\big]^{-1}\mathbf{f}_i, \tag{6}$$

where $\mathbf{F}_{-i} = \mathbf{F}_{-i}(\boldsymbol{\theta}^*)$.

3. Compare $\widehat{Y}_{i,-i}$ with the sample average obtained from simulation, $\overline{W}_{i.}$, using the following t-statistic

$$t_i = \frac{\overline{W}_{i.} - \widehat{Y}_{i,-i}}{\big\{\widehat{\mathrm{Var}}[\overline{W}_{i.}] + \widehat{\mathrm{Var}}[\widehat{Y}_{i,-i}]\big\}^{1/2}} \tag{7}$$

where $\widehat{\mathrm{Var}}[\overline{W}_{i.}] = \hat{\sigma}_i^2/r$ with $\hat{\sigma}_i^2$ given by (2), and

$$\widehat{\mathrm{Var}}[\widehat{Y}_{i,-i}] \approx \frac{1}{r}\hat{\mathbf{f}}_i^T\big[\widehat{\mathbf{F}}_{-i}^T\widehat{\boldsymbol{\Sigma}}_{-i}^{-1}\widehat{\mathbf{F}}_{-i}\big]^{-1}\hat{\mathbf{f}}_i,$$

with $\widehat{\mathbf{F}}_{-i} = \mathbf{F}_{-i}(\hat{\boldsymbol{\theta}}_{-i})$ and $\hat{\mathbf{f}}_i^T = \hat{\mathbf{f}}_i^T(\hat{\boldsymbol{\theta}}_{-i})$; see (6).

These three steps are repeated for each $i = 1, \ldots, n$ and, consequently, a set of n dependent t-statistics are obtained. The metamodel is rejected if

$$\max_{i=1,\ldots,n} |t_i| > t_{r-1;\alpha_E/(2n)},$$

using the Bonferroni's inequality, where α_E is the experiment-wise error rate (valid for the whole experiment), that is, each individual t-statistic is tested at the α_E/n level of significance.

3.2. *Confidence Intervals*

We now develop $100(1-\alpha)\%$ simultaneous confidence intervals for $\mathbf{a}^T\hat{\boldsymbol{\theta}}$ for all $\mathbf{a} \in \Re^m$, enabling the testing of the null-hypothesis that a given factor effect

(for example, its main effect) is unimportant or a given factor (that is, its main effect and higher-order effects) is unimportant. The nonlinear weighted least squares estimator $\hat{\boldsymbol{\theta}}$ satisfies approximately

$$\hat{\boldsymbol{\theta}} \sim \mathrm{N}_m\left(\boldsymbol{\theta}^*, \frac{1}{r}\left[\widehat{\mathbf{F}}^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{F}}\right]^{-1}\right), \tag{8}$$

where $\widehat{\mathbf{F}} = \mathbf{F}(\hat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{\Sigma}}$ is given by (5) (Santos and Porta Nova, 2006). As a result, applying Theorem 3.3.3 of Anderson (1984), we obtain approximately

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^T r\widehat{\mathbf{F}}^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{F}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \sim \chi^2_m. \tag{9}$$

As a consequence, an approximate confidence region, for $\boldsymbol{\theta}^*$, with a probability of coverage equal to $1 - \alpha$, is the following:

$$\left\{\boldsymbol{\theta}^* : (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^T r\widehat{\mathbf{F}}^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{F}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \le \chi^2_{m;\alpha}\right\}. \tag{10}$$

In order to present the simultaneous confidence intervals for $\mathbf{a}^T\boldsymbol{\theta}^*$ for all linear combinations $\mathbf{a}^T\hat{\boldsymbol{\theta}}$, where $\mathbf{0} \neq \mathbf{a} \in \Re^m$, the usual linear theory for Scheffé's *S*-method is applied (Seber, 1977, Ch. 5). First, the Cauchy–Schwartz inequality allows us to write (Seber and Wild, 1989, Appendix A7.2)

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^T r\widehat{\mathbf{F}}^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{F}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = \sup_{\mathbf{a}:\mathbf{a}\neq\mathbf{0}} \frac{\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^T\mathbf{a}\right]^2}{\mathbf{a}^T\left(r\widehat{\mathbf{F}}^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{F}}\right)^{-1}\mathbf{a}}.$$

Using this equality and the confidence region (10), we have approximately

$$\left\{\mathbf{a}^T\boldsymbol{\theta}^* : |\mathbf{a}^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)| \le \left(\chi^2_{m;\alpha}\mathbf{a}^T\left[r\widehat{\mathbf{F}}^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{F}}\right]^{-1}\mathbf{a}\right)^{1/2}\right\}.$$

As a result, for all $\mathbf{a}$ real such that $\mathbf{a} \neq \mathbf{0}$, an approximate $100(1 - \alpha)\%$ confidence interval for $\mathbf{a}^T\boldsymbol{\theta}^*$ is

$$\mathbf{a}^T\hat{\boldsymbol{\theta}} \pm \left(\frac{1}{r}\chi^2_{m;\alpha}\right)^{1/2}\left(\mathbf{a}^T\left[\widehat{\mathbf{F}}^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{F}}\right]^{-1}\mathbf{a}\right)^{1/2}. \tag{11}$$

In order to test the hypothesis $\mathrm{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $\mathrm{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, it is possible to use (10) and to calculate the statistic

$$\chi = r(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\widehat{\mathbf{F}}^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{F}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \tag{12}$$

which has an approximate distribution χ^2_m when H_0 is true. In this case, H_0 is rejected with a confidence level of $1 - \alpha$, if $\chi > \chi^2_{m;\alpha}$.

Suppose that we now partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ and $\widehat{\mathbf{F}} = (\widehat{\mathbf{F}}_1, \widehat{\mathbf{F}}_2)$, where $\boldsymbol{\theta}_2$ is $m_2 \times 1$, $\widehat{\mathbf{F}}_1$ is $n \times m_1$, $\widehat{\mathbf{F}}_2$ is $n \times m_2$, and we wish to test the hypothesis $\mathrm{H}_{02} : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{02}$ versus $\mathrm{H}_{12} : \boldsymbol{\theta}_2 \neq \boldsymbol{\theta}_{02}$. Then analogously to (12) we have the test statistic

$$\tilde{\chi} = r(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{02})^T(\widehat{\mathbf{C}}^{22})^{-1}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{02}); \tag{13}$$

see the verification in Appendix. This statistic is approximately distributed as $\chi^2_{m_2}$ when H_{02} is true, and

$$(\widehat{\mathbf{C}}^{22})^{-1} = (\widehat{\mathbf{R}}\widehat{\mathbf{F}}_2)^T(\mathbf{I}_n - \hat{\mathbf{P}}_{\mathrm{RF}_1})\widehat{\mathbf{R}}\widehat{\mathbf{F}}_2$$

where $\widehat{\mathbf{R}}^T\widehat{\mathbf{R}} = \widehat{\boldsymbol{\Sigma}}$ and $\hat{\mathbf{P}}_{\mathrm{RF}_1} = \widehat{\mathbf{R}}\widehat{\mathbf{F}}_1[(\widehat{\mathbf{R}}\widehat{\mathbf{F}}_1)^T(\widehat{\mathbf{R}}\widehat{\mathbf{F}}_1)]^{-1}(\widehat{\mathbf{R}}\widehat{\mathbf{F}}_1)^T$ is a projection matrix (see Seber and Wild, 1989, Appendix A11).

4. Practical Implementation

The cross-validation method consists of obtaining n individual t-statistics, t_i, where for each evaluation of t_i, a nonlinear regression is required. Since the number of experimental points may be large, and the nonlinear regression software is complex and time consuming, the proposed method may become rather inefficient. An alternative proposed by Kleijnen (1983), in the linear case, is to validate the metamodel at a randomly selected subset of the set of all experimental points; but in this case, potential information is lost. Also in the linear case, Panis et al. (1994) showed that only one regression is needed to implement the test. In the nonlinear situation it remains possible to implement the method by making a single regression, with the help of the following results (proved in the Appendix).

Proposition 4.1. *Suppose* $\epsilon_i \sim \mathrm{N}(0, \sigma_i^2/r)$ *(with* $\sigma_i > 0$*) in* (1) *and let* $\overline{W}_{i.} - \widehat{Y}_{i,-i}$ *and* $\overline{W}_{i.} - \widehat{Y}_i$ *be the residuals that correspond to the nonlinear regression where the experimental point* i $(i = 1, \ldots, n)$ *is eliminated and the nonlinear regression with all experimental points, respectively. Then*

$$\overline{W}_{i.} - \widehat{Y}_{i,-i} = \frac{\overline{W}_{i.} - \widehat{Y}_i}{1 - h_{ii}}, \tag{14}$$

where the "leverage" h_{ii} *is the ith diagonal element of the hat matrix*

$$\mathbf{H} = \mathbf{R}\mathbf{F}(\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{R}^T,$$

corresponding to the regression with all experimental points, and

$$\mathbf{R} = \mathrm{diag}[1/\sigma_1, \ldots, 1/\sigma_n].$$

Moreover,

$$\mathrm{Var}[\widehat{Y}_{i,-i}] = \frac{\sigma_i^2 h_{ii}}{r(1 - h_{ii})}. \tag{15}$$

In practice, when necessary, $\mathbf{F}$, $\boldsymbol{\Sigma}$, and $\mathbf{R}$ can be replaced by $\widehat{\mathbf{F}} = \mathbf{F}(\hat{\boldsymbol{\theta}})$, $\widehat{\boldsymbol{\Sigma}} = \mathrm{diag}[\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2]$ and $\widehat{\mathbf{R}} = \mathrm{diag}[1/\hat{\sigma}_1, \ldots, 1/\hat{\sigma}_n]$, respectively.

In fact, substituting (14) and (15) into the t-statistic (7), we obtain a simplified version of this statistic that can be evaluated using only one nonlinear regression:

$$t_i = \frac{\overline{W}_{i.} - \widehat{Y}_i}{\hat{\sigma}_i\sqrt{(1 - h_{ii})/r}}.$$

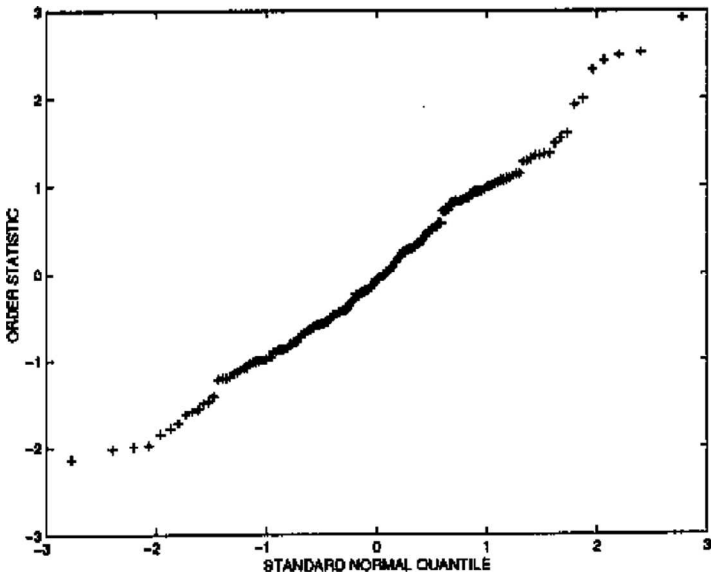

Figure 1. Normal probability plot.

The residuals $\overline{W}_{i.} - \widehat{Y}_{i,-i}$, in (14), are called the deleted residuals and play an important role in regression *diagnostics*. As in the linear case, when building a nonlinear regression metamodel it is of the utmost importance to closely inspect the aptness of the fitted metamodel, by making use of diagnostic techniques based on analysis of the residuals (see Sec. 3.1 and Figs. 1 and 2).

5. Numerical Example: $M/M/1$ Simulation

In this section we use the simple $M/M/1$ queueing system to illustrate the application of our validation and confidence interval procedures. We assume that customers arrive according to a Poisson process with a constant expected arrival rate, λ, and that service times follow an exponential distribution with a constant expected

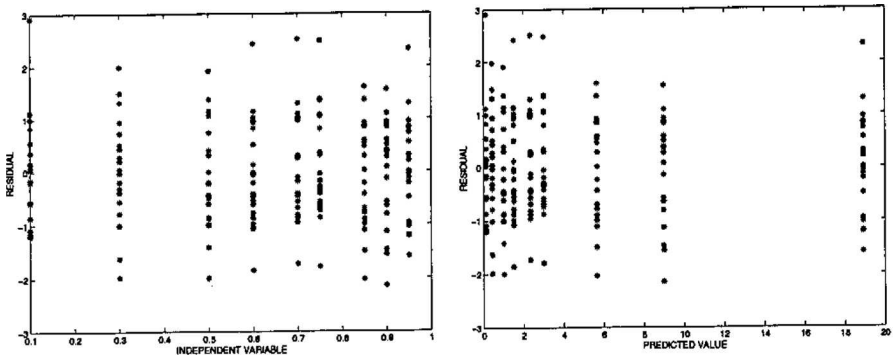

Figure 2. Residuals *versus* the independent variable and the predicted values.

Table 1
Simulation input combinations for $M/M/1$

Combination	1	2	3	4	5	6	7	8	9
λ	0.10	0.30	0.5	0.6	0.70	0.75	0.85	0.90	0.95

service time, $1/\mu$. The performance measure of interest is the steady-state average waiting time in the queue. Our goal is to express this response as a function of the queue utilization factor $\rho = \lambda/\mu$. We fix $1/\mu = 1$, and we select intuitively nine combinations of the λ; see Table 1.

This simulation is an example of a non terminating simulation. For the purposes of output analysis, we distinguish between two types of simulation experiments, *terminating* and *non terminating*, depending on whether there is a specified natural event or set of events that stop the simulation; see, for example, Law and Kelton (2000). The warm-up problem arises in non terminating simulation, where the initial conditions defined by the analyst introduces bias in the simulation output means. A simple and general method for mitigating the initialization bias is the Welch's (1983) procedure. At each design point, we ran this procedure in order to determine the length of each simulation and the initial-data deletion. Welch's moving average is based on 20 replications of the simulation model:

$$\bar{\mu}(j, L) = \begin{cases} (r(2L+1))^{-1} \sum_{s=-L}^{L} \sum_{i=1}^{20} W_{i,j+s}, & \text{if } j \geq L+1, \\ (r(2j-1))^{-1} \sum_{s=-(j-1)}^{j-1} \sum_{i=1}^{20} W_{i,j+s} & \text{if } j < L+1. \end{cases}$$

For example, at the design point $(1/\mu, 1/\lambda) = (1, 2.0)$ (that corresponds to a traffic rate $\rho = \lambda/\mu = 0.5$), we used Welch's window equal to 1,000 and we deleted 1,500 observations from the beginning of the run. Thus, only the remaining 8,500 observations (approximately 85% of the observations per run) were used to estimate the response W.

We carry out $r = 20$ replications of each of the $n = 9$ design points in Table 1; r must be at least nine in order to obtain an appropriate estimate for $\hat{\sigma}_i$, $i = 1, \ldots, n$, (Deaton et al., 1983).

Based on knowledge from queueing theory, we propose the metamodel

$$Y_i = \theta_1 \left(\frac{X_i}{1 + \theta_2 X_i} \right)^{\theta_3} + \epsilon_i, \quad i = 1, \ldots, n. \tag{16}$$

where $X = \rho = \lambda/\mu$.

The variance heterogeneity is measured by the quantity (Kleijnen, 1992)

$$het = \max_{i=1,n} \hat{\sigma}_i / \min_{i=1,n} \hat{\sigma}_i.$$

A value quite different from 1 is obtained: $het = 256.13$. As a consequence, we infer that error variance is not constant. Thus the method of (nonlinear estimated)

Table 2
Estimates for the unknown parameters $\boldsymbol{\theta}$ of the nonlinear metamodel

Regression coefficient	Estimate	Standard deviation
θ_1	0.999622	1.90×10^{-3}
θ_2	−0.999659	5.28×10^{-4}
θ_3	1.000819	1.75×10^{-3}

weighted least squares is used. The estimators for the unknown parameters, in (16), were obtained using the Levenberg–Marquard method implemented in MATLAB, with a termination tolerance equal to 10^{-6} and a maximum number of function evaluations equal to 600 (the default is $100 \times$ number of parameters). This gives Table 2.

In general, the Levenberg–Marquard algorithms are robust and work well in the solution of nonlinear least squares problems. Nevertheless, their linear convergence can be very slow, and they may even not converge if we are solving problems with large residuals. In the latter case, methods for large residual problems can be used (Seber and Wild, 1989, Sec. 14.3).

Our analysis of the weighted residuals indicated that the normality assumption for the underlying metamodel is met, i.e., the normal probability plot in Fig. 1 is approximately linear. We observe that in Fig. 2 the weighted residuals are within a horizontal band centered around 0 and that there is no systematic trend toward being positive or negative (for example, negative residuals for small X values and positive residuals for large X values). These graphs suggest that our nonlinear fitted regression metamodel is appropriate.

In order to ensure that the metamodel is a valid substitute for the simulation model, the cross-validation test proposed in Sec. 3.1 is used, with an experiment-wise error rate of $\alpha_E = 0.10$ so each individual t_i, $i = 1, \ldots, 9$ is tested at the significance level of $\alpha = \alpha_E/9 = 0.01$; see Table 3. Based on the critical value $t_{r-1;\alpha/2} = t_{19;0.01/2} =$

Table 3
Numerical results of the cross-validation test

Combination	t_i
1	−0.247506
2	1.15903
3	−2.67520
4	1.44943
5	2.14220
6	1.48058
7	−0.0641741
8	−0.332590
9	0.228656
$\max \lvert t_i \rvert = 2.67520$	

2.861, we conclude that there is no evidence for rejecting the model, since all t_i are much smaller than the critical value. Observe that no leverage value h_{ii} is large (all h_{ii} are smaller than twice the mean leverage value 0.8889, $i = 1, \ldots, 9$).

The validation test allows us to choose the metamodel $Y = X/(1 - X)$ as representative of the simulation model. Next, since $\chi^2_{m;\alpha}/r = \chi^2_{3;0.05}/20 = 7.81$, we obtained the 95% approximate confidence region for $\boldsymbol{\theta}$, centered at $\hat{\boldsymbol{\theta}}$ (see Eq. (10)):

$$\left\{\boldsymbol{\theta}^* : (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^T \widehat{\mathbf{F}}^T \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\mathbf{F}} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \leq 0.3905\right\},$$

where

$$\widehat{\mathbf{F}}^T \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\mathbf{F}} = \begin{bmatrix} 23.706 & -68.646 & 5.9140 \\ -68.646 & 651.31 & -123.01 \\ 5.9140 & -123.01 & 42.510 \end{bmatrix} \times 10^3.$$

The corresponding approximate simultaneous confidence intervals for $\mathbf{a}^T\boldsymbol{\theta}^*$ for all linear combination $\mathbf{a}^T\hat{\boldsymbol{\theta}}$ ($\mathbf{0} \neq \mathbf{a} = [a_1\ a_2\ a_3]^T \in \Re^3$) are (see Eq. (11)):

$$0.999622a_1 - 0.999659a_2 + 1.000819a_3$$
$$\pm\left(4.5a_1^2 + 0.35a_2^2 + 3.8a_3^2 + 1.6a_1a_2 + 3.2a_1a_3 + 1.8a_2a_3\right) \times 10^{-5}. \quad (17)$$

For this numerical example, the 95% confidence interval for each θ_i was computed from Eq. (17) as

$$\theta_1 \in 0.999622 \pm 4.5 \times 10^{-5}, \quad \theta_2 \in -0.999659 \pm 0.35 \times 10^{-5},$$
$$\theta_3 \in 1.000819 \pm 3.8 \times 10^{-5}.$$

In linear regression—especially polynomial regression—we can use classical design of experiments. In nonlinear problems no general experimental designs exist. For each particular nonlinear problem, a good enough experimental design needs to

Table 4
Analytical solution *versus* metamodel prediction

Exp. point i	Analytical solution Y_i	Metamodel $\widehat{Y}_{i,-i}$	Absolute relative error $\lvert\delta_i\rvert$
1	0.11111	0.11082	2.63×10^{-3}
2	0.42857	0.42786	1.67×10^{-3}
3	1.0000	0.99878	1.22×10^{-3}
4	1.4999	1.4983	1.11×10^{-3}
5	2.3333	2.3307	1.11×10^{-3}
6	3.0000	2.9964	1.20×10^{-3}
7	5.6666	5.6562	1.84×10^{-3}
8	9.0000	8.9737	2.92×10^{-3}
9	18.9999	18.874	6.65×10^{-3}
average % absolute relative error $= 2.26 \times 10^{-3}$			

be found tentatively, with the aid of a computer (see Kleijnen and van Beers, 2004; Santner et al., 2003); that is, an experimental design that allows the construction of a valid simulation metamodel.

Since the analytical formula for the expected steady-state waiting time in an $M/M/1$ system is given by $\lambda/[\mu(\mu-\lambda)]$, we can compare the metamodel with the true I/O function. Based on the results of Table 4, we compute an average absolute relative prediction error of 0.2%, so the metamodel appears to be a reliable and valid approximation; the absolute relative prediction errors are calculated by $|\delta_i| = |(Y_i - \widehat{Y}_{i,-i})/Y_i|$, $i = 1, \ldots, 9$ (Kleijnen and Standridge, 1988).

6. Conclusions

The output of a simulation experiment can be analyzed using a nonlinear regression metamodel that approximates the input/output behavior of the simulation model. The metamodel's parameters can be estimated with the least squares method. Simulation is an important application of the estimation and validation of regression statistical models. Whichever may be the application context (e.g., simulation) of a regression metamodel, its validity must be tested before using it. In this article, a validation test is introduced and illustrated with an example; also, the Scheffé-type confidence intervals are derived. The proposed validation test is a generalization to nonlinear metamodels of Kleijnen's (1983) cross-validation test. This test should always be used together with regression diagnostic tools like, for example, standard diagnostic residual plots for the estimated metamodel. The validation tools give assurance in the use of a regression model as a substitute of a simulation model.

The use of nonlinear regression metamodels in simulation is very important because such metamodels can expose, more clearly than the simulation model itself, the fundamental nature of the simulation model and the real system. Also, they allow us to answer inverse questions. On the other hand, the software used for constructing nonlinear regression metamodels is very time-consuming. However, the advances in modern computer implementation of statistical procedures have made the difference of computing time between nonlinear and linear regression methods less significant. Also, the implementation of the cross-validation test using only a single regression saves much computation time. Finally, when using a metamodel, the computation time used in 'what if' analysis is significantly smaller than required for executing several simulation runs.

Appendix

Verification of (13)

Let $\widehat{\mathbf{C}} = \widehat{\mathbf{F}}^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{F}} = \widehat{\mathbf{F}}^T\widehat{\mathbf{R}}^T\widehat{\mathbf{R}}\widehat{\mathbf{F}}$, $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1^T, \hat{\boldsymbol{\theta}}_2^T)^T$, where $\hat{\boldsymbol{\theta}}_1$ is $m_1 \times 1$ and $\hat{\boldsymbol{\theta}}_2$ is $m_2 \times 1$, and

$$\widehat{\mathbf{C}}^{-1} = \begin{bmatrix} \widehat{\mathbf{C}}_{11} & \widehat{\mathbf{C}}_{12} \\ \widehat{\mathbf{C}}_{21} & \widehat{\mathbf{C}}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \widehat{\mathbf{C}}^{11} & \widehat{\mathbf{C}}^{12} \\ \widehat{\mathbf{C}}^{21} & \widehat{\mathbf{C}}^{22} \end{bmatrix}.$$

We note from (8) that

$$\hat{\boldsymbol{\theta}}_2 \sim \mathrm{N}_{m_2}\left(\boldsymbol{\theta}_2^*, \frac{1}{r}\widehat{\mathbf{C}}^{22}\right),$$

approximately. Partitioning $\widehat{\mathbf{F}} = (\widehat{\mathbf{F}}_1, \widehat{\mathbf{F}}_2)$, where $\widehat{\mathbf{F}}_1$ is $n \times m_1$ and $\widehat{\mathbf{F}}_2$ is $n \times m_2$, we have

$$\begin{aligned}
\left(\widehat{\mathbf{C}}^{22}\right)^{-1} &= \widehat{\mathbf{C}}_{22} - \widehat{\mathbf{C}}_{21}\widehat{\mathbf{C}}_{11}^{-1}\widehat{\mathbf{C}}_{12} \\
&= \widehat{\mathbf{F}}_2^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{F}}_2 - \widehat{\mathbf{F}}_2^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{F}}_1\left[\widehat{\mathbf{F}}_1^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{F}}_1\right]^{-1}\widehat{\mathbf{F}}_1^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{F}}_2 \\
&= \widehat{\mathbf{F}}_2^T\widehat{\mathbf{R}}^T\widehat{\mathbf{R}}\widehat{\mathbf{F}}_2 - \widehat{\mathbf{F}}_2^T\widehat{\mathbf{R}}^T\widehat{\mathbf{R}}\widehat{\mathbf{F}}_1\left[\widehat{\mathbf{F}}_1^T\widehat{\mathbf{R}}^T\widehat{\mathbf{R}}\widehat{\mathbf{F}}_1\right]^{-1}\widehat{\mathbf{F}}_1^T\widehat{\mathbf{R}}^T\widehat{\mathbf{R}}\widehat{\mathbf{F}}_2 \\
&= \widehat{\mathbf{F}}_2^T\widehat{\mathbf{R}}^T\{\mathbf{I}_n - \widehat{\mathbf{R}}\widehat{\mathbf{F}}_1\left[\widehat{\mathbf{F}}_1^T\widehat{\mathbf{R}}^T\widehat{\mathbf{R}}\widehat{\mathbf{F}}_1\right]^{-1}\widehat{\mathbf{F}}_1^T\widehat{\mathbf{R}}^T\}\widehat{\mathbf{R}}\widehat{\mathbf{F}}_2 \\
&= \left(\widehat{\mathbf{R}}\widehat{\mathbf{F}}_2\right)^T\{\mathbf{I}_n - \left(\widehat{\mathbf{R}}\widehat{\mathbf{F}}_1\right)\left[\left(\widehat{\mathbf{R}}\widehat{\mathbf{F}}_1\right)^T\left(\widehat{\mathbf{R}}\widehat{\mathbf{F}}_1\right)\right]^{-1}\left(\widehat{\mathbf{R}}\widehat{\mathbf{F}}_1\right)^T\}\widehat{\mathbf{R}}\widehat{\mathbf{F}}_2 \\
&= \left(\widehat{\mathbf{R}}\widehat{\mathbf{F}}_2\right)^T\left(\mathbf{I}_n - \hat{\mathbf{P}}_{\mathrm{RF}_1}\right)\widehat{\mathbf{R}}\widehat{\mathbf{F}}_2
\end{aligned}$$

where $\widehat{\mathbf{P}}_{\mathrm{RF}_1}$ is an orthogonal projection. Then, analogously to (9), applying Theorem 3.3.3 of Anderson (1984) we have approximately

$$\left(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{02}^*\right)^T r\left(\widehat{\mathbf{C}}^{22}\right)^{-1}\left(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{02}^*\right) \sim \chi_{m_2}^2,$$

and, as a result, the test statistic (13) that is approximately distributed as $\chi_{m_2}^2$ when $\mathrm{H}_{02} : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{02}$ is true.

***Verification of* (14)**

Let $\boldsymbol{\theta}$ be in a small neighborhood of $\boldsymbol{\theta}^*$, the true value of $\boldsymbol{\theta}$. Then the following linear Taylor's expansion is valid (Seber and Wild, 1989, p. 23):

$$f(\mathbf{X}_{i.}, \boldsymbol{\theta}) \approx f(\mathbf{X}_{i.}, \boldsymbol{\theta}^*) + \sum_{j=1}^{m} \frac{\partial f(\mathbf{X}_{i.}, \boldsymbol{\theta}^*)}{\partial \theta_j}(\theta_j - \theta_j^*). \tag{18}$$

Substituting (18) into (1), we have approximately

$$\overline{Y}_{i.} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}^*) + \sum_{j=1}^{m} \frac{\partial f(\mathbf{X}_{i.}, \boldsymbol{\theta}^*)}{\partial \theta_j}(\theta_j - \theta_j^*) + \boldsymbol{\epsilon}_i;$$

that is,

$$\overline{Y}_{i.} - f(\mathbf{X}_{i.}, \boldsymbol{\theta}^*) = \sum_{j=1}^{m} \frac{\partial f(\mathbf{X}_{i.}, \boldsymbol{\theta}^*)}{\partial \theta_j}(\theta_j - \theta_j^*) + \boldsymbol{\epsilon}_i,$$

or

$$\mathbf{Z} = \mathbf{F}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) + \boldsymbol{\epsilon}. \tag{19}$$

Applying the linear regression theory to (19) (see, for example, (9) in Panis et al., 1994), we have

$$Z_i - \widehat{Z}_{i,-i} = \frac{Z_i - \widehat{Z}_i}{1 - h_{ii}}, \tag{20}$$

where h_{ii} is the ith diagonal element of $\mathbf{H} = \mathbf{RF}(\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{R}^T$. Since $Z_i = \overline{Y}_{i.} - f(\mathbf{X}_{i.}, \boldsymbol{\theta}^*)$, $\widehat{Z}_i = \widehat{Y}_i - f(\mathbf{X}_{i.}, \boldsymbol{\theta}^*)$ and $\widehat{Z}_{i,-i} = \widehat{Y}_{i,-i} - f(\mathbf{X}_{i.}, \boldsymbol{\theta}^*)$, (20) is equivalent to (14).

Verification of (15)

Consider $g(\mathbf{X}_{i.}, \boldsymbol{\theta}) = \sqrt{r} f(\mathbf{X}_{i.}, \boldsymbol{\theta})/\sigma_i$, $\varepsilon_i = \sqrt{r}\epsilon_i/\sigma_i$ and $W_i = \sqrt{r}\overline{Y}_{i.}/\sigma_i$. Then, we have approximately $\varepsilon_i \sim \mathrm{N}(0, 1)$ in the problem $W_i = g(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \varepsilon_i$ and (6) can be written, approximately, as

$$\mathrm{Var}[\widehat{Y}_{i,-i}] = \frac{\sigma_i^2}{r}\mathbf{g}_i^T\left[\mathbf{G}_{-i}^T\mathbf{G}_{-i}\right]^{-1}\mathbf{g}_i, \tag{21}$$

where $\mathbf{G}$ is the Jacobian matrix $\mathbf{G} = \mathbf{RF}$. The application of the Sherman–Morrison–Woodbury Theorem (Myers, 1990) to (21) allows us to obtain

$$\mathrm{Var}\left[\widehat{Y}_{i,-i}\right] = \frac{\sigma_i^2}{r}\mathbf{g}_i^T\left[(\mathbf{G}^T\mathbf{G})^{-1} + \frac{(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{g}_i\mathbf{g}_i^T(\mathbf{G}^T\mathbf{G})^{-1}}{1 - h_{ii}}\right]\mathbf{g}_i.$$

But $\mathbf{g}_i^T(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{g}_i = h_{ii}$, so

$$\mathrm{Var}\left[\widehat{Y}_{i,-i}\right] = \frac{\sigma_i^2 h_{ii}}{r(1 - h_{ii})}.$$

References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley and Sons.

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

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, Simulation and Computation* 12(1):45–66.

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

Kleijnen, J. P. C. (1983). Cross-validation using the t statistic. *European Journal of Operational Research* 13:133–141.

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

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., Standridge, C. R. (1988). Experimental design and regression analysis in simulation: an FMS case study. *European Journal of Operational Research* 33:257–261.

Kleijnen, J. P. C., van Groenendaal, W. (1992). *Simulation, A Statistical Perspective*. New York: John Wiley and Sons.

Kleijnen, J. P. C., van Beers, W. C. M. (2004). Application-driven sequential designs for simulation experiments: Kriging metamodeling. *Journal of the Operational Research Society* 55(8):876–883.

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., Sanchez S. M., Lucas T. W., Cioppa, T. M. (2005). A user's guide to the brave new world of designing simulation experiment. *INFORMS Journal on Computing* 17(3):263–289.

Law, A. M., Kelton, W. D. (2000). *Simulation Modeling and Analysis*. 3rd ed. New York: McGraw-Hill.

Myers, R. H. (1990). *Classical and Modern Regression with Applications*. Boston: Duxbury Press.

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.

Santner, T. J., Williams, B. J., Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. New York: Springer-Verlag.

Santos, M. I., Porta Nova, A. M. (1999). The main issues in nonlinear simulation metamodel estimation. In: Farrington, P., Nembhard, H., Sturrok, D., Evans, G., eds. *Proceedings of Winter Simulation Conference*, pp. 502–509.

Santos, M. I., Porta Nova, A. M. (2001). Validation of nonlinear simulation metamodels. In: Hamza, M., ed. *Proceedings of Applied Simulation and Modeling*, pp. 421–425.

Santos, M. I., Porta Nova, A. M. (2006). Statistical fitting and validation of nonlinear simulation metamodels: a case study. *European Journal of Operational Research* 171(1): 53–63.

Santos, M. I., Porta Nova, A. M., Santos, P. M. (2006). A methodology for constructing and implementating nonlinear simulation metamodels. Working Paper. Technical University of Lisbon.

Seber, G. A. F. (1977). *Linear Regression Analysis*. New York: John Wiley and Sons.

Seber, G. A. F., Wild, C. J. (1989). *Nonlinear Regression*. New York: John Wiley and Sons.

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