STATISTICAL FITTING AND VALIDATION OF NONLINEAR SIMULATION METAMODELS: A CASE STUDY

M. Isabel Reis dos Santos and Acácio M. O. Porta Nova
Instituto Superior Técnico
Av. Rovisco Pais, 1049-001 Lisboa, PORTUGAL
apnova@ist.utl.pt

Key Words: discrete event simulation; nonlinear regression metamodels; metamodel validation.

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 polynomials results in a poor fit. The use of a nonlinear 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 nonlinear metamodels and double-cross 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 (e.g., in queueing systems) and it becomes increasingly difficult to make decisions about design modifications in the target system. Whenever possible, it is more suitable to construct a simple mathematical relationship that relates the

inputs and outputs of the computer simulation model, that is, a model of the simulation model, or *metamodel* (Barton 1992). Metamodels are very useful in design optimization and "what if?" questions–all this, 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.

Traditional 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, Myers and Houck (1994), Kleijnen (1992), Porta Nova and Wilson (1989) and Kleijnen, Burg and Ham (1979). However, polynomials are unable to produce a global fit to curves of arbitrary shape. Moreover, in real-life systems, nonlinearity is common and the approximation using polynomials becomes unrealistic. Consequently, in these situations, polynomials often fail to provide good fits, namely in problems involving queueing systems (Friedman and Friedman 1985). An alternative that provides better and more realistic global fits is the use of statistical nonlinear regression techniques; see Santos and Porta Nova (1999, 2001).

However, any metamodel can only be used to analyze simulation output if it is "good enough". So, after estimating the metamodel, it is advisable to check if the hypothetical metamodel is, in fact, an accurate representation of the simulation model. For this purpose, robust statistical validation techniques from nonlinear regression are used.

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

2. NONLINEAR 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 can not be observed in the real system and 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 represented by the output variable of the simulation model. As a result, the simulation model can be represented by

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

where Y is a univariate response, $\mathbf{Z} = (Z_1, \ldots, Z_k)^T$ is a vector of input variables and $\mathbf{a}$ represents a set of random streams that drive the simulation at $\mathbf{Z}$. The vector $\mathbf{Z}$, in the simulation of a supermarket, can include the mean interarrival time, the mean service time and the number of physical lanes. The response Y 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 $\mathbf{X} = (X_1, \ldots, X_{\mathsf{d}})$ is a vector of d explanatory variables, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{\mathsf{m}})^T$ represents a vector of unknown parameters, ϵ represents the error and f is an unknown function simpler than η (the error ϵ includes effects of the inadequacy of f as a representation of ϕ and intrinsic effects encountered in any stochastic simulation model); see Figure 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's. For example, in the $M/M/1$ queue, the utilization factor $X = \rho = \lambda/\mu$ (where λ is the arrival rate and μ is the service rate) can be a better explanatory variable than λ and μ. In this paper, f is a nonlinear function of the unknown parameter vector $\boldsymbol{\theta}$, so we are dealing with nonlinear metamodels. The unknown parameter vector must be estimated.

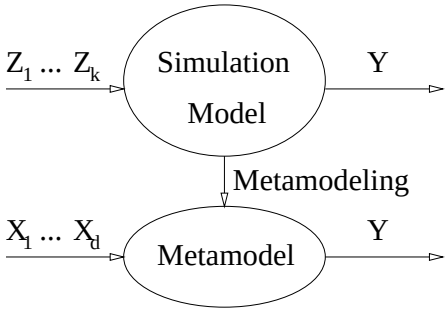

Figure 1: Simulation model *vs.* metamodel

In the $M/M/1$ queue, an example of a nonlinear 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, whereas a linear metamodel could be

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

where Y represents the average waiting time in the queue.

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

$$\hat{\sigma}_i^2 = \sum_{j=1}^{\mathsf{r}} (Y_{ij} - \bar{Y}_{i.})^2 / [\mathsf{r}(\mathsf{r} - 1)] \qquad i = 1, \ldots, \mathsf{n}. \tag{3}$$

The average

$$\bar{Y}_{i.} = \sum_{j=1}^{\mathsf{r}} Y_{ij}/\mathsf{r}, \qquad i = 1, \ldots, \mathsf{n}$$

is the metamodel response of interest.

This allows us to express the metamodel (2) as

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

where ϵ_i are independent random variables with $\epsilon_{ij} \sim \mathrm{N}\,(0, \sigma_i^2)$, and $\sigma_i > 0$.

Before estimation, one or more hypothetical metamodels must be selected. Ideally, the form of the metamodel should be dictated by theoretical considerations. For example, in the M/M/1 queueing system, an hypothetical 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 hypothetical metamodels should be made visually, just like we compare empirical histograms with known density functions for selecting a specific random distribution. Following this perspective, it is convenient to represent the dispersion diagrams (or scatter plots) of the response *versus* each independent variable, plotting, for each fixed $l = 1, \ldots, \mathsf{d}$, the corresponding pairs (X_{il}, Y_{ij}), where $i = 1, \ldots, \mathsf{n}$ and $j = 1, \ldots, \mathsf{r}$. Then, we should compare the graphical representations with different analytical curves from a comprehensive catalog, in order to select the best candidates.

2.1 Nonlinear Metamodel Estimation

The metamodel estimation procedure uses the nonlinear least squares method for estimating the unknown parameters of the hypothetical metamodel. This well known method from regression analysis minimizes the error sum of squares

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

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 $\mathbb{R}^{\mathsf{m}}$. For most nonlinear metamodels, the $\mathrm{SSE}(\boldsymbol{\theta})$ function can not be minimized analytically and, as a consequence, an iterative numerical method is used. We chose the Levenberg-Marquardt method because it is almost an unanimous opinion that, for many nonlinear

least squares problems, this method works very well. However, in problems with large residuals, Levenberg-Marquardt algorithms can 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.

The SSE($\boldsymbol{\theta}$) function in (5), in contrast to the linear case, can have several local minima, in addition to the global minimum. Thus, in many situations, the best that we can expect is that the numerical method will converge to a local minimum. For $\hat{\boldsymbol{\theta}}$ to be a local minimum, it is sufficient that: (i) the partial derivatives of SSE($\boldsymbol{\theta}$) with respect to θ_1, ..., θ_{m} be zero; and (ii) the Hessian matrix of SSE($\boldsymbol{\theta}$), calculated at $\hat{\boldsymbol{\theta}}$, be positive definite.

In Proposition 1, the nonlinear least squares estimator $\hat{\boldsymbol{\theta}}$ is established. Under the hypothesis that the ϵ_i are independent and normally distributed, $\epsilon_i \sim \mathrm{N}\left(0, \sigma_i^2/\mathsf{r}\right)$, and assuming some further regularity conditions, it is shown that $\hat{\boldsymbol{\theta}}$ is asymptotically normally distributed as $\mathsf{N} = \mathsf{nr} \to \infty$ (see the verification of these results in the Appendix).

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

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

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

where $\boldsymbol{\theta}^*$ *is the exact value of* $\boldsymbol{\theta}$*,* $\mathbf{f} = \mathbf{f}(\boldsymbol{\theta}^*) = (f(\mathbf{X}_{1.}, \boldsymbol{\theta}^*), \ldots, f(\mathbf{X}_{\mathsf{n}.}, \boldsymbol{\theta}^*))^T$*,* $\mathbf{F} = \mathbf{F}(\boldsymbol{\theta}^*)$ *is the Jacobian matrix of* $\mathbf{f}$*, evaluated at* $\boldsymbol{\theta}^*$*,* $\bar{\mathbf{Y}} = (\bar{Y}_{1.}, \ldots, \bar{Y}_{\mathsf{n}.})^T$ *and* $\boldsymbol{\Sigma}$ *is the diagonal matrix* $\boldsymbol{\Sigma}$ $= \mathrm{diag}(\sigma_1^2, \ldots, \sigma_{\mathsf{n}}^2)$*. 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 $\hat{\boldsymbol{\Sigma}} = \mathrm{diag}[\hat{\sigma}_1^2, \ldots, \hat{\sigma}_{\mathsf{n}}^2]$ and, as a result, we have the estimated generalized nonlinear least squares estimator.

3. METAMODEL VALIDATION

In order to use the hypothetical metamodel (as a surrogate of the simulation model) for prediction and sensitivity analysis, we have to be sure that it is, in fact, 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.

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 nonlinear models and has the following requirements: (i) $\mathsf{n} > \mathsf{m} = \mathsf{rank}(\mathbf{F})$ and $\mathsf{r} > \mathsf{n} (> \mathsf{n} - \mathsf{m})$ (so that $\hat{\mathbf{\Sigma}}$ is nonsingular); (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 conditions of applicability, the generalization of Rao's test has the form:

$$\begin{aligned} F_{Rao} &= \frac{\mathsf{r}(\mathsf{r}-\mathsf{n}+\mathsf{m})}{(\mathsf{n}-\mathsf{m})(\mathsf{r}-1)} \left[\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \hat{\boldsymbol{\theta}})\right]^T \hat{\mathbf{\Sigma}}^{-1} \left[\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \hat{\boldsymbol{\theta}})\right] \\ &= \frac{\mathsf{r}(\mathsf{r}-\mathsf{n}+\mathsf{m})}{(\mathsf{n}-\mathsf{m})(\mathsf{r}-1)} \sum_{i=1}^{\mathsf{n}} \left[\frac{\bar{Y}_{i.} - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}})}{\hat{\sigma}_i}\right]^2 . \end{aligned} \tag{8}$$

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

The predictive validity is verified using double cross-validation and an adaptation of the

prediction sum of squares, PRESS; see Neter, Wasserman and Kutner (1989) and Friedman and Friedman (1985). In our problem, the PRESS statistic has the form

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

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

3.2 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 dimension. Then, a regression metamodel is developed for each subset and used for prediction on 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_{bld}, 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. Since the errors have generally unequal variances, we propose the following approximated $1 - \alpha$ two-sided confidence rectangle for the estimated generalized nonlinear least squares estimator of the regression parameter θ_k $(k = 1, \ldots, \mathsf{m})$:

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

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

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

(see Proposition 1) and $F_{Rao} = F_{\mathsf{n}-\mathsf{m},\mathsf{r}-\mathsf{n}+\mathsf{m}}$ in (8). This confidence rectangle is obtained adapting (3.13) of Kleijnen (1992) to nonlinear 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 inter-arrival times are normally distributed with mean μ and a variance of 15. Only one inspector services the cars and the time that he needs to inspect one is uniformly distributed between 15 and 25 minutes. In the inspection queue, space is available for only six cars. On the average, 85 percent of the cars pass the inspection and leave the center. The other 15 percent 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 $\mathsf{l} = 2000$ observations, that is,

$$\bar{\mu}(\mathsf{l}, W) = \begin{cases} (2W+1)^{-1} \sum_{w=-W}^{W} 1/\mathsf{r} \sum_{i=1}^{\mathsf{r}} Y_{i,\mathsf{l}+w}, & \text{if} \quad \mathsf{l} \geq W+1, \\ (2\mathsf{l}-1)^{-1} \sum_{w=-(\mathsf{l}-1)}^{\mathsf{l}-1} 1/\mathsf{r} \sum_{i=1}^{\mathsf{r}} Y_{i,(\mathsf{l}+w)} & \text{if} \quad \mathsf{l} < W+1. \end{cases}$$

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 I).

TABLE I: 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

We carried out $\mathsf{r} = 30$ replications of each of the $\mathsf{n} = 14$ 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 $\hat{\sigma}_i$, $i = 1, \ldots, \mathsf{n}$ (Deaton, Reynolds and Myers 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 $\{(X_i, Y_{ij}) : i = 1, \ldots, \mathsf{n}, j = 1, \ldots, \mathsf{r}\}$; see Figure 2. Then, we performed a visual check of the scatter plot, comparing it with graphical representations of some functional relationships from an appropriate catalog. The nonlinear curves that seemed to be good candidates are represented in Figure 3. 'Arctan' is based on the arc tangent function and the others are three sigmoidal growth models (Seber and Wild 1989), pages 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 $r = 2, \ldots, 10$.

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

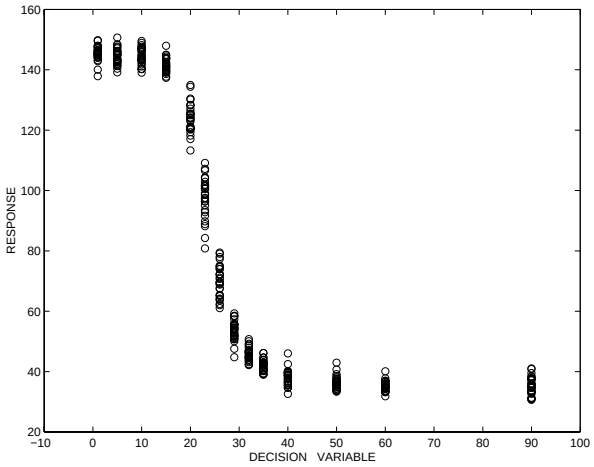

Figure 2: Visualization of simulation results.

quantity

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

(Kleijnen 1992). We obtained $het = 3.946$ (quite different from 1), and so we are in the presence of non homogeneous variances. Thus, we will use nonlinear weighted least squares for nonlinear curves and weighted least squares for polynomials. The nonlinear 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 had to be rejected.

To check the validity of the remaining hypothesized metamodels, we evaluated the statistics presented in Section 3. The SSE and PRESS statistics for the nonlinear models exhibit similar values, in contrast to the linear polynomial models (see Table II). Also, the SSE values are large for the linear models, compared to the nonlinear ones. Based on these results,

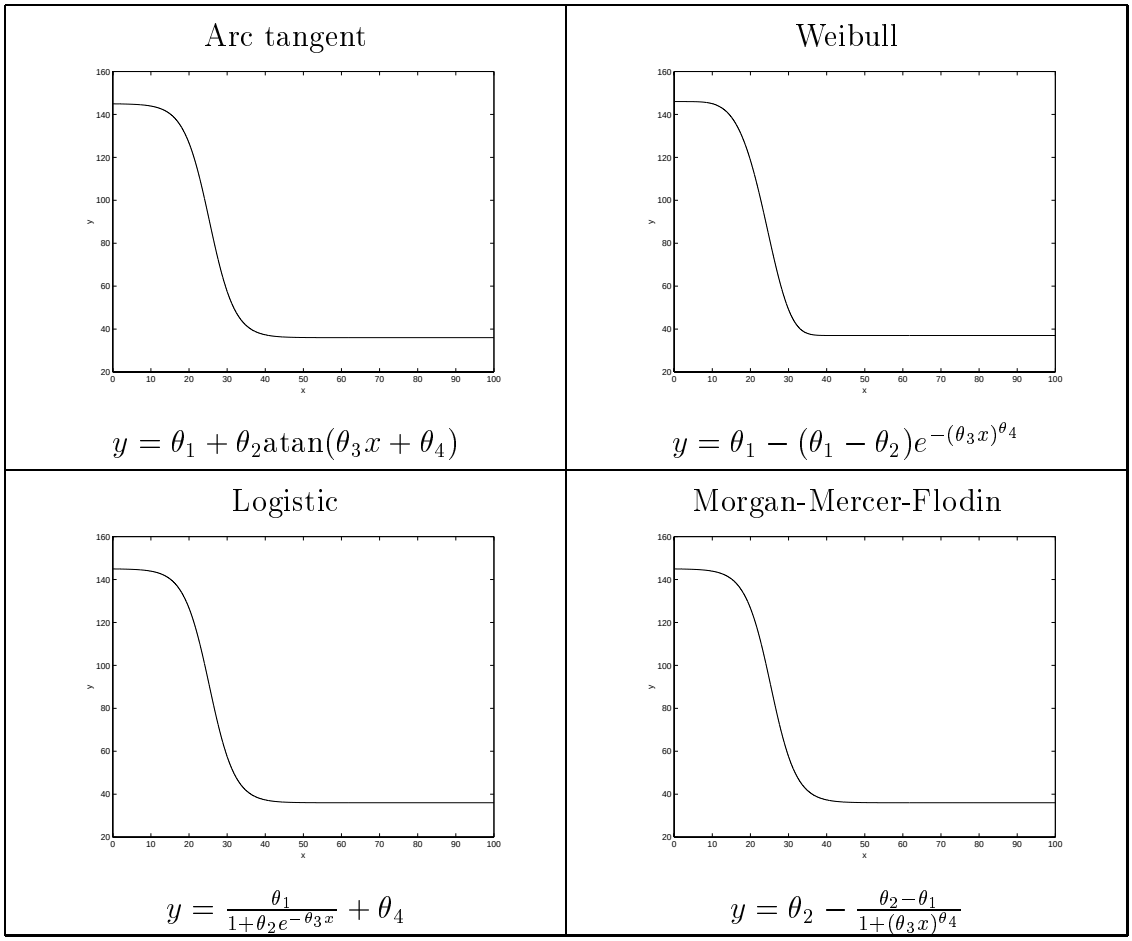

Figure 3: Candidate nonlinear functional relationships.

we conclude that polynomial functions have lack of predictive validity, and so they are not good approximations for the target simulation model.

TABLE II: Metamodel diagnostics.

Metamodel	SSE	MSE	PRESS	SSE/PRESS
Arctan	1554.23	3.73612	1608.21	0.966
Logistic	1647.62	3.96063	1696.78	0.971
MMF	1366.68	3.28529	1418.26	0.964
Weibull	2224.96	5.34847	2272.22	0.979
Pol2	39714.1	95.2377	15162.9	2.619
Pol3	40041.4	96.2533	15105.4	2.651

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, for each $i = 1, \ldots, \mathsf{n}$, the normal probability plots for the set of original simulation responses, $\{Y_{ij} : j = 1, \ldots, \mathsf{r}\}$, were obtained. All of the resulting graphics appear to be nearly linear, but the slope varies with the corresponding design point. This agrees with the fact that the value of *het* is quite different from 1. In Figure 4, results for some experimental points are displayed. 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, $F^{1-\alpha}_{\mathsf{n}-\mathsf{m},\mathsf{r}-\mathsf{n}+\mathsf{m}} = F^{0.95}_{10,20} = 2.348$. The elected metamodel, according to this criterion, is the one based on the MMF curve–all others are rejected (see Table III).

TABLE III: Rao's test.

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

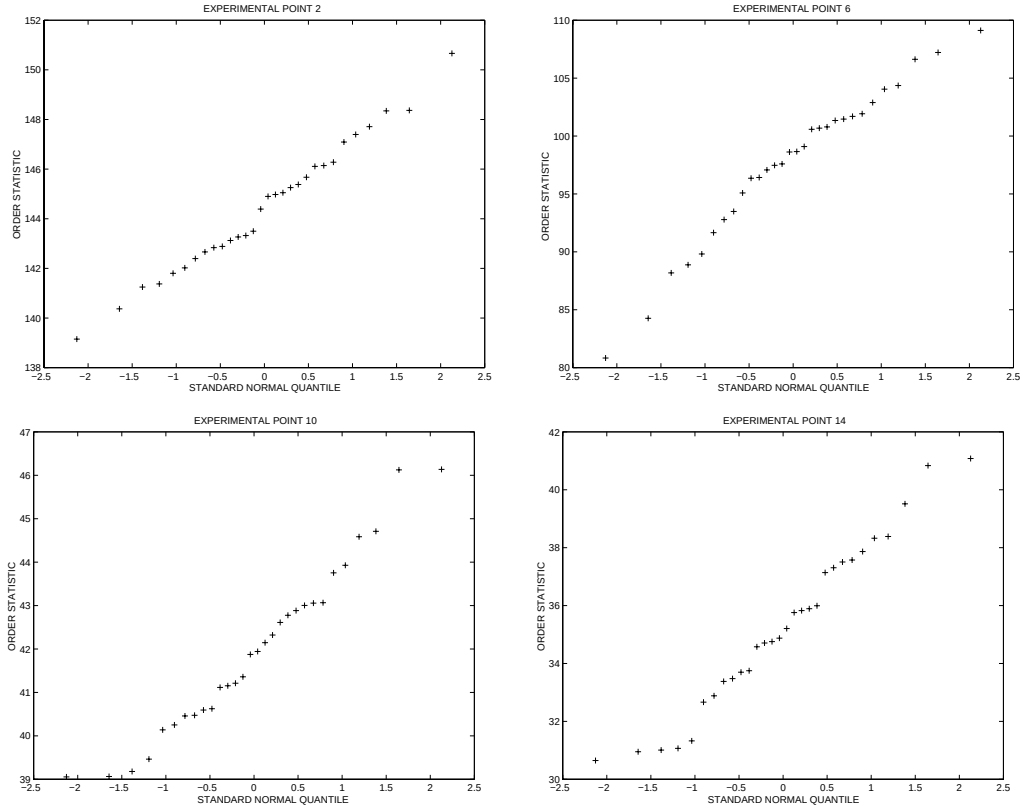

Figure 4: Normal probability plots (experimental points 2, 6, 10 e 14).

To gain more insight into the predictive validity of the metamodels, we analyzed the results of double cross-validation (see Table IV). In each model, we observe a good agreement between the coefficients obtained based in 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 V). The standard deviations of the estimators are also shown. We observe that the standard deviations and the CI half-lengths have rather 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.

TABLE IV: Double cross-validation test.

	Arctan		Logistic	
Coefficient	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	
Coefficient	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 V: 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

5. CONCLUSIONS

This paper stresses the importance of using reliable nonlinear 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 nonlinear ones, but they are unable to ensure a global fit to curves of arbitrary shape. Nonlinear regression metamodels are advantageous, because they do not have this limitation, allowing an adequate adherence 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 model is rejected.

The use of nonlinear metamodels requires an extensive catalog of curves and a 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 rather 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. Nevertheless, the computation time required for obtaining nonlinear regression metamodels can be orders of magnitude smaller than the time needed to run and analyze the actual simulation model.

APPENDIX

(i) Verification of (6).

We pointed out that the errors in the nonlinear metamodel (4) have unequal variances–the setup for generalized or weighted least squares. We also observe that, minimizing problem (4) is equivalent to minimizing

$$[\bar{Y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta})]^T \left(\frac{1}{\mathsf{r}}\boldsymbol{\Sigma}\right)^{-1} [\bar{Y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta})]$$

with respect to $\boldsymbol{\theta}$; see Seber and Wild (1989), Section 2.1.4. This is also equivalent to minimizing $[\bar{Y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta})]^T \boldsymbol{\Sigma}^{-1} [\bar{Y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta})]$, the error sum of squares corresponding to the problem

$$\bar{Y}_{i.} = \mathbf{f}(\mathbf{X}, \boldsymbol{\theta}) + \bar{\boldsymbol{\epsilon}}_{i.}, \quad i = 1, \ldots, \mathsf{n}. \tag{9}$$

$\boldsymbol{\Sigma}$ is a symmetric positive definite matrix, that accepts the Cholesky decomposition: $\boldsymbol{\Sigma} = \mathbf{U}^T\mathbf{U}$, where $\mathbf{U}$ is an upper triangular matrix. Multiplying the nonlinear model (9), on the left, by $\mathbf{R} = (\mathbf{U}^T)^{-1}$, we obtain $\mathbf{W} = \mathbf{g}(\mathbf{X}, \boldsymbol{\theta}) + \boldsymbol{\eta}$, where $\mathbf{W} = \mathbf{R}\bar{\mathbf{Y}}$, $\mathbf{g}(\mathbf{X}, \boldsymbol{\theta}) = \mathbf{R}\mathbf{f}(\mathbf{X}, \boldsymbol{\theta})$ and $\boldsymbol{\eta} = \mathbf{R}\boldsymbol{\epsilon}$, with $\boldsymbol{\epsilon} = (\epsilon_{1.}, \ldots, \epsilon_{\mathsf{n}.})^T$.

Then, we observe that $\mathrm{E}[\boldsymbol{\eta}] = \mathbf{0}$ and $\mathrm{Var}[\boldsymbol{\eta}] = \mathbf{R}\mathrm{Var}[\boldsymbol{\epsilon}]\mathbf{R}^T = 1/\mathsf{r}\mathbf{R}\boldsymbol{\Sigma}\mathbf{R}^{\mathsf{T}}$. But $\boldsymbol{\Sigma} = \mathbf{U}^T\mathbf{U}$ (Cholesky decomposition) and $\mathbf{R} = (\mathbf{U}^T)^{-1}$. Thus,

$$\mathrm{Var}[\boldsymbol{\eta}] = 1/\mathsf{r}(\mathbf{U}^{\mathsf{T}})^{-1}\mathbf{U}^{\mathsf{T}}\mathbf{U}\left[(\mathbf{U}^{\mathsf{T}})^{-1}\right]^{\mathsf{T}} = 1/\mathsf{r}\mathbf{I}_{\mathsf{n}},$$

where $\mathbf{I}_{\mathsf{n}}$ is the identity matrix of order n. We conclude that problem (9) has been transformed into an ordinary least squares problem. Thus, the least squares estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^* + [\mathbf{G}^T\mathbf{G}]^{-1}\mathbf{G}^T[\mathbf{W} - \mathbf{g}], \tag{10}$$

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

But, since $\mathbf{g}(\mathbf{X}, \boldsymbol{\theta}) = \mathbf{R}\mathbf{f}(\mathbf{X}, \boldsymbol{\theta})$, we have $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{R}\partial\mathbf{f}(\mathbf{X}, \boldsymbol{\theta})/\partial\boldsymbol{\theta}^T = \mathbf{R}\mathbf{F}(\boldsymbol{\theta})$. Besides,

$\mathbf{W} = \mathbf{R}\bar{\mathbf{Y}}$ and $\mathbf{R}^T\mathbf{R} = \boldsymbol{\Sigma}^{-1}$, thereof (10) is equivalent to:

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

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

(i) Verification of (7).

Result (7) is obtained by applying Theorem 2.1, item (i), of Seber and Wild (1989), to (10): $\hat{\boldsymbol{\theta}} \sim \mathrm{N}_p\left[\boldsymbol{\theta}, 1/\mathsf{r}(\mathbf{G}^T\mathbf{G})^{-1}\right]$. Since $\mathbf{G} = \mathbf{R}\mathbf{F}$ and $\mathbf{R}^T\mathbf{R} = \boldsymbol{\Sigma}^{-1}$, we obtain $\hat{\boldsymbol{\theta}} \sim \mathrm{N}_p\left[\boldsymbol{\theta}, 1/\mathsf{r}(\mathbf{F}^T\mathbf{R}^T\mathbf{R}\mathbf{F})^{-1}\right]$, and then $\hat{\boldsymbol{\theta}} \sim \mathrm{N}_p\left[\boldsymbol{\theta}, 1/\mathsf{r}(\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{F})^{-1}\right]$.

BIBLIOGRAPHY

Barton R. R. (1992). "Metamodels for Simulation Input-Output Relations", in *Proceedings of Winter Simulation Conference*, 289-299.

Deaton, M. L., Reynolds, M. R. and 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. and Friedman H. H. (1985). "Validating the simulation metamodel: Some practical approaches", *Simulation*, 45 (3), 144-146.

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., van den Burg, A. J. and 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. and Groenendaal, W. V. (1992). *Simulation, A Statistical Perspective*. John Wiley and Sons, New York, NY, USA.

Kleijnen, J. P. C. and Sargent, R. G. (2000). "A Methodology for fitting and Validating Metamodels in Simulation", *European Journal of Operational Research*, 120, 14-29.

Neter, J., Wasserman, W. and Kutner, M. H. (1989). *Applied Linear Regression Models*. Richard D. Irwin Inc., second edition.

Panis, R. P., Myers, R. H. and Houck, E. C. (1994). "Combining Regression Diagnostic with Simulation Metamodels", *European Journal of Operational Research*, 13, 133-141.

Porta Nova, A. M. and 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. and Porta Nova, A. M. (2001). "Validation of nonlinear simulation metamodels", in *Proceedings of Applied Simulation and Modeling*, 421-425.

Santos, M. I. and Porta Nova, A. M. (1999). "The Main Issues in Nonlinear Simulation Metamodel Estimation", in *Proceedings of Winter Simulation Conference*, 502-509.

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

White, H. (1980). "Nonlinear Regression on Cross-Section Data", *Econometrica*, 48 (3), 721-746.

Welch P. D. (1983). "The statistical analysis of simulation results", in S. S. Lavenberg, editor, *Computer Performance Modeling Handbook*, 268-328. Academic Press, New York, NY, USA.