Validation of nonlinear simulation metamodels

M. Isabel Reis dos Santos
Departamento de Matemática
Instituto Superior Técnico
Av. Rovisco Pais, 1049-001
Lisboa, PORTUGAL
email: prs@digitais.ist.utl.pt

Acácio M. O. Porta Nova
Secção autónoma de Economia e gestão
Instituto Superior Técnico
Av. Rovisco Pais, 1049-001
Lisboa, PORTUGAL
email: apnova@ist.utl.pt

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, using several linear regression polynomials results in a poor fit. The use of a nonlinear regression modeling 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.

KEY WORDS
Discrete event simulation; Nonlinear metamodels; Nonlinear regression; Metamodel validation.

1. Introduction

Frequently, the main objective in digital simulation studies is the prediction and sensitivity analysis of 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 queuing 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* [1]. 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.

The construction of discrete event simulation metamodels often uses traditional linear regression procedures. In particular, the general *linear* regression model has been intensively studied (e.g. [2], [3], [4] and [5]). However, polynomials are unable to produce a global fit to curves of arbitrary shape. Moreover, in real-life systems, nonlinearity is common and approximation using polynomials becomes unrealistic. Consequently, in these situations, polynomials often fail to provide good fits (e.g., in problems involving queueing systems [6]). An alternative that provides better and more realistic global fits is the use of statistical nonlinear regression techniques [7].

After estimating the unknown metamodel parameters, the metamodel can only be used to analyze simulation output if it is 'good enough'. So, once the metamodel has been estimated it is advisable to check if the hypothetical metamodel is, in fact, an accurate representation of the simulation model. For this purpose, valid statistical validation techniques from nonlinear regression are used.

This paper is organized as follows. In Section 2. we present the general nonlinear metamodel, and techniques for validating it. In Section 3. we describe an actual problem concerning a center for inspecting and repairing automobiles. We consider several candidate metamodels, including linear and nonlinear ones, and we select the metamodel that provides the best fit. This selection is based on regression techniques of fit and validation described in Section 2.. Section 5. is reserved for conclusions.

2. Nonlinear Regression Metamodels

Consider the following nonlinear simulation metamodel:

$$Y = f(\mathbf{X}, \theta) + \epsilon, \tag{1}$$

where Y is a univariate response, $\mathbf{X} = (X_1, \ldots, X_\mathsf{d})$ is a vector of d explanatory variables, $\theta = (\theta_1, \ldots, \theta_\mathsf{m})$ represents a vector of unknown parameters, ϵ is the error and f is an unknown function. Suppose also, an experimental design consisting of n different design points, $\{X_{il} : i = 1, \ldots, \mathsf{n}; \mathsf{l} = 1, \ldots, \mathsf{d}\}$. For each design point, r independent replications of the simulation model are carried out and the experiment yields $\{(Y_{ij}, \hat{\sigma}_i) : i = 1, \ldots, \mathsf{n}; \mathsf{j} = 1, \ldots, \mathsf{r}\}$, where Y_{ij} is the relevant system response and $\hat{\sigma}_i$ is the estimated standard deviation at the design point i based on r observations and is calculated using (8.4) of [8]

$$\hat{\sigma}_i = \left[\frac{1}{\mathsf{r}-1} \sum_{j=1}^{\mathsf{r}} (Y_{ij} - \bar{Y}_{i.})^2 \right]^{1/2},$$

where $\bar{Y}_{i.} = \frac{1}{\mathsf{r}}\sum_{j=1}^{\mathsf{r}} Y_{ij}$. This leads us to express the metamodel (1) as a replicated simulation metamodel

$$Y_{ij} = f(\mathbf{X}_{i.}, \theta) + \epsilon_{\mathbf{ij}}, \tag{2}$$

for $i = 1, \ldots, \mathsf{n}$ and $j = 1, \ldots, \mathsf{r}$, where $\epsilon_{ij} \sim$ NID $(0, \sigma_i^2)$, with $\sigma_i > 0$.

For large n and given appropriate regularity conditions (see Proposition 1 of [7], the vector of unknown parameters θ is approximated through the estimated generalized nonlinear least squares estimator

$$\hat{\theta} \approx \theta^* + [\mathbf{F^T}\hat{\mathbf{\Sigma}}^{-1}\mathbf{F}]^{-1}\mathbf{F^T}\hat{\mathbf{\Sigma}}^{-1}[\bar{\mathbf{Y}} - \mathbf{f}], \tag{3}$$

where θ^* is the exact value of θ, $\mathbf{f} = \mathbf{f}(\theta^*) = (f(\mathbf{X}_{1.}, \theta^*), \ldots, \mathbf{f}(\mathbf{X}_{\mathsf{n}.}, \theta^*))^{\mathbf{T}}$, $\mathbf{F} = \mathbf{F}(\theta^*)$ the Jacobian matrix of $\mathbf{f}$, evaluated at θ^*, $\bar{\mathbf{Y}} = (\bar{Y}_{1.}, \ldots, \bar{Y}_{\mathsf{n}.})^T$ and $\hat{\mathbf{\Sigma}}$ is the diagonal matrix $\hat{\mathbf{\Sigma}} = \mathrm{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_{\mathsf{n}}^2)$. In order to simplify the notation, we omit that $\mathbf{f}$ and $\mathbf{F}$ are evaluated at θ^*.

2.1 Validation Procedure

After estimating the metamodel parameters, we must test the ability of the estimated metamodel to approximate the simulation model response (i.e., to ascertain if the estimated metamodel adequately fits the simulation data). To test the adequacy of the metamodel (2), we assume that the responses are normally distributed (when simulation responses are averages, central limit theorems ensure normality), $\mathsf{n} > \mathsf{m} = \mathsf{rank}(\mathbf{F})$ and $\mathsf{r} > \mathsf{n}$, and we propose the following lack-of-fit test that is an adaptation of Rao's test [9] to nonlinear models:

$$\begin{aligned} F &= c\left[\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \hat{\theta})\right]^T \hat{\mathbf{\Sigma}}^{-1}\left[\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \hat{\theta})\right] \\ &= c\sum_{i=1}^{\mathsf{n}}\left[\frac{\bar{Y}_{i.} - f(\mathbf{X}_{i.}, \hat{\theta})}{\hat{\sigma}_i}\right]^2, \end{aligned} \tag{4}$$

where $c = [\mathsf{r}(\mathsf{r} - \mathsf{n} + \mathsf{m})]/[(\mathsf{n} - \mathsf{m})(\mathsf{r} - 1)]$.

When the metamodel is valid, F 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 [10] and [6]. In our problem, the PRESS statistic has the form $\mathrm{PRESS} = \sum_{i=1}^{\mathsf{n}}\sum_{j=1}^{\mathsf{r}}[Y_{ij} - f(\mathbf{X}_{i.}, \hat{\theta}_{(-j)})]^2/\hat{\sigma}_i^2$, where $\hat{\theta}_{(-j)}$ 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 $\mathrm{SSE} = \sum_{i=1}^{\mathsf{n}}\sum_{j=1}^{\mathsf{r}}[Y_{ij} - f(\mathbf{X}_{i.}, \hat{\theta})]^2/\hat{\sigma}_i^2$ and mean sum of squares $\mathrm{MSE} = \mathrm{SSE}/(\mathsf{rn} - \mathsf{m})$.

In double cross-validation we split the data intuitively into two subsets of approximately equal 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 we calculate two values of the coefficient of determination, R^2: the first one is based on observations from the subset used in building it, R^2_{bui}, and the second one is based on the other half, R^2_{val}. Moreover, we compare the estimated parameters of both metamodels.

3. An Inspection and Repair Center

The problem studied in this paper is a center for inspecting and repairing automobiles. The time between car arrivals to the center is normally distributed with mean μ and variance 15 minutes. One inspector waits for a car and the time required to inspect it 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 inspection and leave the center. The other 15 percent must go to the repair section where two repairmen 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 mean 60 minutes.

Our goal is to express the average time in system, Y (response), as a function of the mean time between arrivals, μ (decision variable). We consider 14 combinations of simulation input, μ, selected intuitively, $\{\mu_i : i = 1, 14\} = \{1, 5, 10, 15, 20, 23, 26, 29, 32, 35, 40, 50, 60, 90\}$. At each design point, we run Welch's procedure [11] 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 metamodel, where each replication contains $l = 2000$ observations, that is,

$$\bar{\mu}(l, W) = \begin{cases} \frac{\sum_{w=-W}^{W} \bar{\mu}_{l+w}}{2W+1} & if \;\; l \geq W+1, \\ \frac{\sum_{w=-(l-1)}^{l-1} \bar{\mu}_{l+w}}{2l-1} & if \;\; 1 \geq l < W+1, \end{cases}$$

with $\bar{\mu}_{l+w} = \sum_{i=1}^{l+w} Y_{i,(l+w)}/(l + w)$ and where W is Welch's window. For example, at the design point $\mu_i = 10$, we delete 100 observations from the beginning of the run and only the remaining 600 observations, approximately 85% of the observations in run, are used to estimate the response Y (see Table 1 and Figure 1).

Table 1. Initial data deletion.

	Observations		Welch's
μ_i	Deleted	In run	window
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 carry out $\mathsf{r} = 30$ replications of each of the $\mathsf{n} = 14$ design points; in order to apply Rao's validation test, r must

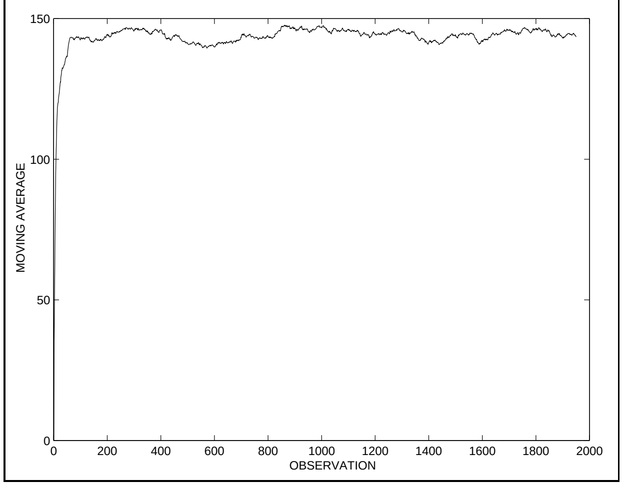

Figure 1. Moving average $\bar{\mu}(2000, 50)$ based on 20 replications.

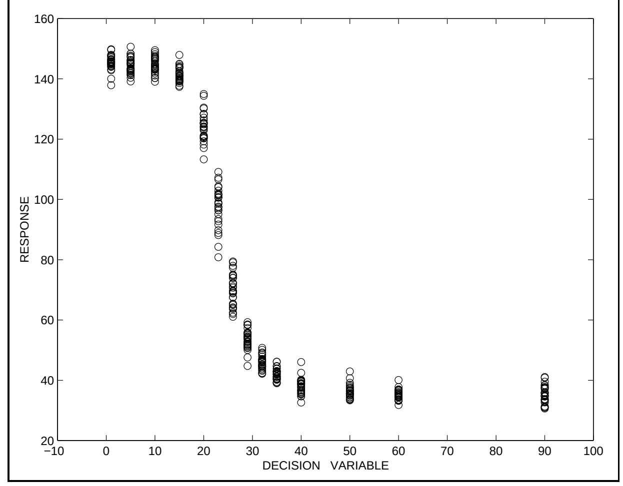

Figure 2. Visualization of simulation results.

be greater than n, and since r is greater than nine, we obtain an appropriate estimate for $\hat{\sigma}_i$, $i = 1, \ldots, \mathsf{n}$ [12].

With the objective of identifying a curve that might fit the input/output of the simulation program, we performed a visual check, plotting the pairs (X_i, Y_{ij}) $(i = 1, \ldots, \mathsf{n}, \mathsf{j} = 1, \ldots, \mathsf{r})$ and comparing them with the graphical representations of some functional relationships. The candidate nonlinear curves, that resemble the simulation data represented in Figure 2, are in Table 2. '$atan$' is the arc tangent function and the others are three sigmoidal growth models (see [13], pages 329, 338 and 340): 'Logistic' is the logistic model, 'Weibull' is the Weibull growth curve and 'MMF' is the Morgan-Mercer-Flodin family. We also considered polynomial functions of degree r, with $r = 2, \ldots, 10$.

Table 2. Some functional relationships.

Model	Expression
atan	$y = \theta_1 + \theta_2 \arctan(\theta_3 x + \theta_4)$
Logistic	$y = \frac{\theta_1}{1+\theta_2 e^{-\theta_3 x}} + \theta_4$
Weibull	$y = \theta_1 - (\theta_1 - \theta_2) e^{-(\theta_3 x)^{\theta_4}}$
MMF	$y = \theta_2 - \frac{\theta_2 - \theta_1}{1+(\theta_3 x)^{\theta_4}}$

We measure the variance heterogeneity through the quantity

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

(see [4]), and we obtained $het = 3.946$ (quite different from 1), so we are in the presence of non constant variances. Thus we will use the method of nonlinear weighted least squares in the case of nonlinear curves and weighted least squares for polynomials. The estimators for the cases included in Table 2, were obtained using Lavenberg-Marquard method implemented in MATLAB, with termination tolerance equal to 10^{-6} and maximum number of function evaluations equal to 600 (default is 100 $\times$ number of parameters). In the cases of the estimation using polynomials with degree r, with $r = 4, \ldots, 10$, we obtained matrices close to singular or badly scaled. Since the results might be inaccurate, we therefore rejected these metamodels.

Table 3. Metamodel diagnostics.

Metamodel	SSE	MSE	PRESS	SSE/PRESS
$atan$	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

4. Validating the Metamodel

With the objective of checking the validity of the remaining hypothesized metamodels, we evaluated the statistics presented in Section 2.1.

The SSE and PRESS statistics for the nonlinear models exhibit similar values, in contrast to the linear polynomial models (see Table 3). Also, the SSE values are large for the linear models, compared to the nonlinear ones. Based on these results, we conclude that polynomial functions have lack of predictive validity, so they are not good approximations for the target simulation model.

To gain more insight into the predictive validity, we analyze the results of double cross-validation (see Table 4). 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.

Rao's test was used in order to select the metamodel that better approximated the simulation results, by ordering

Table 4. Double cross-validation test.

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

Table 5. Rao test.

Metamodel	$atan$	logistic	MMF	Weibull
F_{Rao}	6.108	9.102	2.206	22.673

the F_{Rao} values and comparing them with the F critical value $F^{1-\alpha}_{\mathsf{n-m,r-n+m}} = F^{0.95}_{10,20} = 2.348$. The elected metamodel according to this criterion is the one based on the MMF curve (see Table 5).

5. Conclusions

This article stresses the importance of reliable nonlinear metamodels in simulation studies. In the example discussed here, a poor fit was obtained when various polynomial metamodels were used, leading to a demand for more precise and flexible models. Linear models are comparably simpler to fit than nonlinear ones, but 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 fitting of complex curves. In the example illustrated in this paper, a set of sigmoidal growth models and arc-tangent function were compared against each other in a nonlinear regression context.

It is of prime importance 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 anyone of these tests fails, the model is rejected. Firstly, the SSE and PRESS statistics lead to the rejection of the linear models. Then, the nonlinear metamodels were checked using double-cross validation. Finally, the generalization of Rao's test of lack-of-fit to nonlinear models was used to refine the curve selection process, allowing us to elect the best fitting curve: the MMF metamodel.

The use of nonlinear metamodels requires an extensive catalog of curves and a more complex and time consuming regression software. The selection of a good fitting curve 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 curve is rather straightforward. The regression and validation software can be reused repeatedly, as soon as the user supplies a function to implement the chosen curve and an initial value. Finally, the increased computation time, when compared to linear regression procedures, is becoming less important with the ever growing computing power of recent computers. Nevertheless, the computation time required for nonlinear model regression can be several orders of magnitude smaller than the one needed to run the simulation model.

References

[1] R. R. Barton. Metamodels for simulation input-output relations. In *Proceedings of the Winter Simulation Conference*, pages 289–299, 1992.

[2] R. P. Panis, R. H. Myers, and E. C. Houck. Combining regression diagnostics with simulation metamodels. *European Journal of Operational Research*, 13:85–94, 1994.

[3] J. D. Tew and J. R. Wilson. Validation of simulation analysis methods for the Schruben-Margolin correlation-induction strategy. *Operations Research*, 40(1):87–103, 1992.

[4] J. P. C. Kleijnen. Regression metamodels for simulation with common random numbers: Comparison of validation tests and confidence intervals. *Management Science*, 38(8):1164–1185, August 1992.

[5] J. P. C. Kleijnen, A. J. Burg, and R. Th. Ham. Generalization of simulation results, practicality of statistical methods. *European Journal of Operational Research*, 3:50–64, 1979.

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

[7] M. I. Santos and A. M. Porta Nova. The main issues in nonlinear simulation metamodel estimation. In *Proceedings of the Winter Simulation Conference*, pages 502–509, 1999.

[8] J. P. C. Kleijnen and W. V. Groenendaal. *Simulation, A Statistical Perspective*. John Wiley and Sons, New York, NY, USA, 1992.

[9] C. R. Rao. Some problems involving linear hypotheses in multivariate analysis. *Biometrika*, 46:49–58, 1959.

[10] J. Neter, W. Wasserman, and M. H. Kutner. *Applied Linear Regression Models*. Richard D. Irwin Inc., segunda edition, 1989.

[11] P. D. Welch. The statistical analysis of simulation results. In S. S. Lavenberg, editor, *Computer Performance Modeling Handbook*, pages 268–328. Academic Press, New York, NY, USA, 1983.

[12] M. L. Deaton, M. R. Reynolds, and R.H. Myers. Estimation and hypothesis testing in regression in the presence of nonhomogeneous error variances. *Communications in Statistics B - Simulation and Computation*, 12(1):45–66, 1983.

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