Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

THE MAIN ISSUES IN NONLINEAR SIMULATION METAMODEL ESTIMATION

M. Isabel Reis dos Santos

Departamento de Matemática
Instituto Superior Técnico
Av. Rovisco Pais, 1049 Lisboa, PORTUGAL

Acácio M. O. Porta Nova

Secção Autónoma de Economia e Gestão
Instituto Superior Técnico
Av. Rovisco Pais, 1049 Lisboa, PORTUGAL

ABSTRACT

In this paper, we investigate and discuss some of the main issues concerning the estimation of nonlinear simulation metamodels. We propose a methodology for identifying a tentative functional relationship, estimating the metamodel coefficients and validating the simulation metamodel. This approach is illustrated with a simple queueing system. Finally, we draw some conclusions and identify topics for further work in this area.

1 INTRODUCTION

The use of discrete event simulation models produces significant amounts of output data making it hard to interpret that data, or to try to predict the system behavior for a slightly different experimental environment. A simulation *metamodel* simplifies the simulation model itself, exposing more clearly the fundamental nature of the system input-output relationships.

To build a discrete event simulation metamodel, we use classical statistical procedures, borrowed from regression analysis. The objective is to determine a (relatively simple) functional relationship between the system response and selected decision variables. Thus, it becomes much easier (and cheaper), not only to analyze the simulation output, but to predict how the real system will react to specific combinations of the set of controllable input variables. It is also straightforward to perform sensivity analyses of the simulation model parameters and "what if?" questions—all this, without having to perform additional simulation runs. However, extra care must be taken when collecting the simulation data, fitting the metamodel and, especially, validating it. Since we will be using mainly well known and robust statistical procedures, this approach is more likely to gain confidence and acceptance from simulation practitioners as well.

Linear models are relatively simple to fit and manipulate, and so their use becomes rather attractive. In particular, Kleijnen has been especially active in this area; see, for instance, Kleijnen (1992), Kleijnen, Burg and Ham (1979) and Kleijnen and Groenendaal (1992). Porta Nova and Wilson (1989) discuss the estimation of a general linear multivariate simulation metamodel, as well as its use in the context of variance reduction, with the technique of control variables.

Since reality is hardly linear, linear models are acceptable approximations only in smaller or larger neighborhoods of the design points under consideration. However, Friedman and Friedman (1985) reported a significant lack-of-fit, when queue length in the M/M/s queue was expressed, in a linear fashion, in terms of the arrival and service rates and the number of parallel servers, s. They point out that this is a common problem in metamodels of queueing systems, since the above decision variables are known to be "... intricately related in a nonlinear fashion". As in this case, if a nonlinear simulation metamodel is firmly based in theory and we extrapolate from the region where it was developed, it is rather unlikely that it will produce fundamentally wrong predictions. Unfortunately, this does not happen with most polynomial models. Another advantage of nonlinear models is that they usually have a much smaller number of parameters, when compared with linear models.

Simulation practitioners might raise some questions... Is it feasible, in practical terms, to fit a meaningful nonlinear metamodel to a realistic simulation model of an actual system? Is the eventual improvement worth the additional time and complexity of nonlinear *vs.* linear statistical approaches?

Consequently, in this paper, we address and discuss some of the main issues involved in the use of nonlinear metamodels to analyze simulation output. In this context, it becomes even more important to test the metamodel validity. We will use valid statistical procedures to determine the lack-of-fit of the model, as well as its predictive capability. We feel that our approach will only be useful if it can be understood and applied by any simulation practitioner. Thus, we will propose a methodology for graphically sum-

marizing the simulation data, selecting from a catalog of target functional relationships and estimating and validating a specific metamodel.

This paper is organized as follows. In Section 2, we discuss the estimation of a general nonlinear simulation metamodel. In Section 3, we investigate the validation of the simulation metamodel. In Section 4, we present a methodology for iterative identification, estimation and validation of simulation metamodels, and illustrate its application using the M/M/s queue. Finally, in Section 5, we draw some conclusions and recommendations for future work in this area.

2 METAMODEL ESTIMATION

In general, we can say that simulation models try to approximate reality, while simulation metamodels are approximations of the simulation models themselves. Thinking of simulation as an input-output transformation, we are lead to the notion that simulation is basically a function, although rather complicated, that cannot usually be expressed by a simple expression. But it may be possible to approximate, with a single formula (a metamodel), what the simulation actually does.

When building a simulation model, we should represent the most important variables and parameters. Zeigler (1976) defines a parameter as a quantity that can not be observed in the real system, while a variable is directly observable. Client arrival times or the number of servers in a queue are examples of variables. The arrival rate, λ, and the service rate, μ, of a Poisson process—see Section 1.2 in Kleijnen and Groenendaal (1992)—are examples of parameters. When a simulation program is executed, parameters are well known input values. The response of the real system is represented by the output variable Y of the simulation model.

In this paper, following Kleijnen and Groenendaal (1992), we represent the simulation model (or program) by a mathematical function, ϕ:

$$Y = \phi(\mathbf{Z}, \mathbf{r}), \tag{1}$$

where Y is the system response, $\mathbf{Z} = Z_1, \ldots, Z_k$ are the input variables and parameters and $\mathbf{r}$ represents the set of random number streams that drive the simulation at $(Z_1, \ldots, Z_k)$. In a queueing system, the dependent variable or response might be, for instance, the average queue length or the mean system sojourn time.

The approximating function, of the above simulation program, is the following nonlinear metamodel:

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

where the independent or explanatory variables $\mathbf{X} = (X_1, ..., X_d)$ belong to a subset of R^d, $\boldsymbol{\theta} = (\theta_1, ..., \theta_p) \in \Theta \subseteq R^p$ is a vector of parameters to be estimated, ϵ represents the error and f is an unknown function. The error, ϵ, includes both effects due to the inadequacy of f as a representation of ϕ, as well as intrinsic effects, always present in any stochastic simulation model—they depend on $\mathbf{r}$ in (1). Sometimes, X_i, in (2), is identical to the simulation variable or parameter Z_j, in (1); for instance, the arrival or service rates in a queue. In other cases, X_i may be a transformation of one or more Z_j's; again, in the specific case of queueing systems, $X_1 = Z_1/Z_2$ may constitute a better explanatory variable (if we consider the "traffic intensity", $\rho = \lambda/\mu$). Consequently, the parameters and input variables $\mathbf{Z}$ of the simulation model (1) determine the independent variables $\mathbf{X}$ of the simulation metamodel (2). The coefficients $\boldsymbol{\theta}$ in (2) are designated metamodel parameters and must be estimated.

2.1 Data for Analysis

In practice, the mathematical conditions associated with the metamodel (2) may or may not be satisfied. Thus, we start by postulating a specific form for model (2) and, then, we test its validity. The approach is: (i) we first choose, for the model, a function that may closely follow the output variable Y, throughout the region to which the data belong; then, (ii) we estimate the parameters of the "elected" model; and, finally, (iii) we investigate if the model is, in fact, adequate or not. That is, if it can be used to forecast the system behavior or not.

In order to build a meaningful simulation metamodel, we have to determine a sufficient number, m, of *design points* (that is, combinations of the d explanatory or decision variables) that will cover the relevant part of the decision region under study, $\{X_{il} : l = 1, d\}$, for $i = 1, m$. These design points must be *unique*—that is, any two combinations of the decision variables must have, at least, one different element.

Although the estimation of a nonlinear simulation metamodel might be discussed in the context of other methods for output analysis, we felt that *independent replications* were particularly well suited for this purpose. Thus, we first choose adequate values for parameters and suitable random distributions for the stochastic components in the simulation model. Then, we perform an appropriate number of model runs, n, for each of the m design points, using independent random streams, and collect the data on the relevant system response, $\{Y_{ij} : i = 1, m;\ j = 1, n\}$. Finally, we use classical statistical procedures to compute point estimators or confidence intervals for the response, from the above random sample of size n. The number of replications, n, at each design point may now be much smaller than the number that is generally used in a common simulation study.

2.2 Least Squares Estimation

We assume that the simulation model (1) can be modeled through the *replicated* simulation metamodel

$$Y_{ij} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \epsilon_{ij}, \tag{3}$$

for $i = 1, m$ and $j = 1, n$, where $\epsilon_{ij} \sim \text{NID}\left(0, \sigma_i^2\right)$, with $\sigma_i > 0$. Then, the population's conditional expectation and variance are $\text{E}[Y_{ij}|\mathbf{X}_{i.}] = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) = \mu_i$ and $\text{Var}[Y_{ij}|\mathbf{X}_{i.}] = \sigma_i^2$. As such, the simulation output at each design point, $\{Y_{ij} : j = 1, \ldots, n\}$, for $i = 1, \ldots, m$, can be interpreted as n independent observations from the normal distribution $\text{N}(\mu_i, \sigma_i^2)$. Thus, for estimation purposes, we can consider, instead, an equivalent LS problem, in which the individual observations, at each design point, are replaced by their averages:

$$\bar{Y}_{i.} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \bar{\epsilon}_{i.}, \qquad i = 1, 2, \ldots, m, \tag{4}$$

with $\text{Var}[\bar{Y}_{i.}] = \sigma_i^2/n$ and $\bar{\epsilon}_{i.} \sim \text{N}(0, \sigma_i^2/n)$.

To estimate the parameters, $\boldsymbol{\theta}$, in the metamodel (3), we apply the nonlinear least squares (LS) method. In contrast to the linear case, for most nonlinear models, the system of normal equations cannot be solved analytically; so, we must resort to an iterative method. We will discuss how the Gauss-Newton method can be used to obtain approximately the asymptotic nonlinear LS estimator, $\hat{\boldsymbol{\theta}}$; see Section 2.1.3 in Seber and Wild (1989).

Proposition 1 *Given appropriate regularity conditions—see White (1980)—and for large m, the LS estimator of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, in (3) satisfies, approximately:*

$$\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^* + [\mathbf{F}^T \boldsymbol{\Sigma}^{-1} \mathbf{F}]^{-1} \mathbf{F}^T \boldsymbol{\Sigma}^{-1} [\bar{\mathbf{Y}} - \mathbf{f}], \tag{5}$$

$$\hat{\boldsymbol{\theta}} \sim \text{N}_p \left(\mathbf{0}, \frac{1}{n} [\mathbf{F}^T \boldsymbol{\Sigma}^{-1} \mathbf{F}]^{-1} \right). \tag{6}$$

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

Verification We point out that in the nonlinear metamodels (3) and (4), the errors have *unequal* variances—the setup for *generalized* or *weighted* LS. Consequently, to determine the WLS estimator, $\hat{\boldsymbol{\theta}}$, we minimize

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

with respect to $\boldsymbol{\theta}$; see Section 2.1.4 in Seber and Wild (1989). But, this is equivalent to minimizing $[\bar{Y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta})]^T \boldsymbol{\Sigma}^{-1} [\bar{Y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta})]$.

$\boldsymbol{\Sigma}$ is a symmetric positive definite matrix, that accepts the Cholesky decomposition:

$$\boldsymbol{\Sigma} = \mathbf{U}^T \mathbf{U}, \tag{7}$$

where $\mathbf{U}$ is an upper triangular matrix. Multiplying the nonlinear model (4) through $\mathbf{R} = (\mathbf{U}^T)^{-1}$, we obtain

$$\mathbf{W} = \mathbf{g}(\mathbf{X}, \boldsymbol{\theta}) + \boldsymbol{\eta}, \tag{8}$$

where $\mathbf{W} = \mathbf{R}\bar{\mathbf{Y}}$, $\mathbf{g}(\mathbf{X}, \boldsymbol{\theta}) = \mathbf{R}\mathbf{f}(\mathbf{X}, \boldsymbol{\theta})$ and $\boldsymbol{\eta} = \mathbf{R}\bar{\boldsymbol{\epsilon}}$, with $\bar{\boldsymbol{\epsilon}} = (\bar{\epsilon}_{1.}, \ldots, \bar{\epsilon}_{m.})^T$.

Then, we observe that $\text{E}[\boldsymbol{\eta}] = \mathbf{0}$ and $\text{Var}[\boldsymbol{\eta}] = \mathbf{R}\text{Var}[\bar{\boldsymbol{\epsilon}}]\mathbf{R}^T = (1/n)\mathbf{R}\boldsymbol{\Sigma}\mathbf{R}^T$. But $\boldsymbol{\Sigma}$ allows the decomposition (7) and $\mathbf{R} = (\mathbf{U}^T)^{-1}$; thus, $\text{Var}[\boldsymbol{\eta}] = (1/n)(\mathbf{U}^T)^{-1}\mathbf{U}^T\mathbf{U}\left[(\mathbf{U}^T)^{-1}\right]^T = (1/n)\mathbf{I}_m$, where $\mathbf{I}_m$ is the identity matrix of order m. We conclude that the problem (4) has been transformed into an *ordinary* LS (OLS) problem. Thus, the OLS estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^* + [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T [\mathbf{W} - \mathbf{g}], \tag{9}$$

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 Theorem 2.1 in Seber and Wild (1989).

But, since $\mathbf{g}(\mathbf{X}, \boldsymbol{\theta}) = \mathbf{R}\mathbf{f}(\mathbf{X}, \boldsymbol{\theta})$, we have $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{R}\partial \mathbf{f}(\mathbf{X}, \boldsymbol{\theta}) / \partial \boldsymbol{\theta}^T = \mathbf{R}\mathbf{F}(\boldsymbol{\theta})$. Besides, $\mathbf{W} = \mathbf{R}\bar{\mathbf{Y}}$ and $\mathbf{R}^T\mathbf{R} = \boldsymbol{\Sigma}^{-1}$, thereof (9) 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 (5) has been established.

Result (6) is obtained by applying Theorem 2.1 in Seber and Wild (1989), item (i), to the problem (9): $\hat{\boldsymbol{\theta}} \sim \text{N}_p \left[\boldsymbol{\theta}, (1/n)(\mathbf{G}^T\mathbf{G})^{-1}\right]$. Since $\mathbf{G} = \mathbf{R}\mathbf{F}$ and $\mathbf{R}^T\mathbf{R} = \boldsymbol{\Sigma}^{-1}$, we obtain $\hat{\boldsymbol{\theta}} \sim \text{N}_p \left[\boldsymbol{\theta}, (1/n)(\mathbf{F}^T\mathbf{R}^T\mathbf{R}\mathbf{F})^{-1}\right]$, and then $\hat{\boldsymbol{\theta}} \sim \text{N}_p \left[\boldsymbol{\theta}, (1/n)(\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{F})^{-1}\right]$.

As as estimator of $\boldsymbol{\Sigma}$, we can use

$$\hat{\boldsymbol{\Sigma}} = \text{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_m^2), \tag{10}$$

where $\hat{\sigma}_i^2$ is given by

$$\hat{\sigma}_i^2 = \frac{1}{n-1} \sum_{j=1}^{n} (Y_{ij} - \bar{Y}_{i.})^2; \tag{11}$$

see page 151 in Kleijnen and Groenendaal (1992).

3 METAMODEL VALIDATION

The purpose of the simulation metamodel validation is to investigate whether the metamodel adequately approximates the behavior of the input/output generated by a simulation program. The assessment of this adequacy is necessarily subjective. However, the simulation responses of interest are generally averages, and so, central limit effects ensure normality. In the next section, we present statistical tests that help detecting the lack of fit associated with the deterministic portion of the proposed nonlinear regression metamodel. In Section 3.2, the metamodel predictive capability is tested, by using the holdout sample method.

3.1 Model Adequacy

In the replicated metamodel (3), if the total number of observations, $N = \sum_{i=1}^{m} n = mn$, is large, we can use the following rough F-test for lack of fit, proposed by Seber and Wild (1989), page 82:

$$F = \frac{(SSE - SSPE)/(m-p)}{SSPE/(N-m)}, \tag{12}$$

where $SSE = \sum_{i=1}^{m}\sum_{j=1}^{n} w_i[Y_{ij} - f(\mathbf{X}_{i\cdot}\hat{\boldsymbol{\theta}})]^2$ is the usual error sum of squares (or, residual sum of squares), $SSPE = \sum_{i=1}^{m}\sum_{j=1}^{n} w_i[Y_{ij} - \bar{Y}_{i\cdot}]^2$ is the pure error sum of squares and $w_i = 1/\sigma_i^2$ are the weights. If there exists a parametrization for which the model can be adequately approximated by a linear model, then F will be roughly distributed as an $F_{m-p,N-m}$ distribution, when the model is valid.

As an additional statistic for testing the metamodel validity, we also propose the coefficient of determination R^2,

$$R^2 = \frac{SSR}{SST},$$

where $SSR = \sum_{i=1}^{m}\sum_{j=1}^{n}\left[f(\mathbf{X}_{i\cdot},\hat{\boldsymbol{\theta}})/\sigma_i - \bar{Y}_{\cdot\cdot}\right]^2$ is the regression sum of squares, $SST = \sum_{i=1}^{m}\sum_{j=1}^{n}\left[Y_{ij}/\sigma_i - \bar{Y}_{\cdot\cdot}\right]^2$ is the total sum of squares and $\bar{Y}_{\cdot\cdot} = 1/(mn)\sum_{i=1}^{m}\sum_{j=1}^{n} Y_{ij}/\sigma_i$ is the grand mean of the observations. Since R^2 always increases as we add more explanatory variables, we could also use an R^2 adjusted for the number of parameters, p:

$$R^2_{adj} = 1 - (1-R^2)\frac{N-1}{N-p}.$$

3.2 Predictive Validity

To test the predictive validity of the metamodel, we use a *data splitting* (or *cross-validation*) procedure; see Sections 12.6 and 15.4 in Neter, Wasserman and Kutner (1989). We build a new regression model using only about two-thirds of the N observations (*model-building* sample). Splits of the data are made intuitively. The holdout observations (*validation* or *prediction* set) are used to test the regression model. For example, if we have $N = 120$ observations, then the model-building and the holdout groups will have 72 and 48 observations, respectively. We evaluate the coefficient of determination, R^2, for both the model-building and the holdout cases. If the two values of R^2 are very close, then we can conclude that the model does have predictive validity.

In order to obtain more information about the predictive capability of the metamodel, we can compute the mean squared prediction error, denoted by $MSPR$:

$$MSPR = \frac{1}{m^*}\sum_{i=1}^{m^*}[Y_i - f(\mathbf{X}_{i\cdot},\hat{\boldsymbol{\theta}})]^2, \tag{13}$$

where $f(\mathbf{X}_{i\cdot},\hat{\boldsymbol{\theta}})$ is the predicted value for the ith validation case, based on the model-building data set, Y_i is the value of the response variable in the ith validation case and m^* is the number of cases in the validation data set (holdout sample); see Neter, Wasserman and Kutner (1989), page 466.

In our case, problem (3), we compute the $MSPR$ through

$$MSPR = \frac{1}{\hat{m}n}\sum_{i=1}^{\hat{m}}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}[Y_{ij} - f(\mathbf{X}_{i\cdot},\hat{\boldsymbol{\theta}})]^2,$$

where $\hat{m}$ is the number of levels (design points) of $\mathbf{X}$ and n is the number of replications in each level. Values of $MSPR$ close to the MSE computed for the regression fitted to the model-building sample, are an indication that the MSE gives an appropriate measure of the predictive capability of the model. If MSE is much smaller than $MSPR$, then we should use $MSPR$ as an indicator of the predictive capability of the metamodel.

The regression coefficients for the holdout group are then estimated and we compare for consistency with the estimated regression coefficients based in the model-building group.

Another useful statistic for testing the metamodel predictive validity is the *prediction sum of squares*, $PRESS$,

procedure; see Neter, Wasserman and Kutner (1989), page 450. In our case, this procedure has to be adapted and so we have the following quantity:

$$PRESS = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{1}{\sigma_i^2} [Y_{ij} - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}}_{(-i)})]^2, \qquad (14)$$

where $\hat{\boldsymbol{\theta}}_{(-i)}$ is the estimated parameter vector based on the set that we obtain if we delete the replications that correspond to case i.

If $PRESS$ and SSE are quite close, then MSE may be a valid indicator of the predictive capability of the selected model. A disadvantage is the necessity of doing m distinct regressions. To perform each of the m estimations, we have to use an iterative procedure. This is usually time-consuming and we may have an additional problem choosing adequate starting values for the iterations.

4 APPLICATION

In this section we illustrate the application of our methodology by means of an example. For this purpose, we modeled the $M/M/s$ queue, with a single service facility and a single waiting line. Demands were assumed to arrive according to a Poisson process with a constant average arrival rate, λ, and service times were assumed to follow an exponential distribution with a constant average service time $1/\mu \equiv 1$. Our goal was to express the average waiting time in the queue (the response) as a function of the queue utilization factor, $\rho = \lambda/\mu$ (a single decision variable). We considered the following twelve ($m = 12$) different values for ρ (and λ), $\{\rho_i : i = 1, 9\} =$ {.1, .2, .3, .4, .5, .55, .6, .7, .75, .85, .9, .95}. We decided to perform $n = 10$ replications of each of the $m = 12$ design points; we chose, for n, a number greater than nine, in order to obtain an appropriate estimate for σ_i^2, $i = 1, ..., m$; see Deaton, Reynolds and Myers (1983). Different replications use the same value for the independent variable ρ_i, but different pseudorandom number seeds. Each of these 10 replications starts in the empty state (no customers waiting). In order to account for the presence of initializing bias at each design point, we ran Welch's procedure (Welch 1983), for increasing number of observations and window widths. Consequently, the deleted observations were made to correspond to about 15% of the total number of observations in each run. For instance, for $\rho = .1$, the number of observations in each run was 3,500, the first 500 were deleted and a window of 1,000 was enough; for $\rho = .95$, each run included 40,000 and 3,500 were deleted, for a window of 20,000.

Taking into account the discussion in Section 2.1, the collected simulation data for the metamodel estimation is summarized in Table 1. X_i represents the utilization factor corresponding to the ith design point. $\bar{Y}_{i.}$ is the average across runs of the waiting time in queue for the ith design point, with utilization factor X_i.

Table 1: Simulation Data for Metamodel Estimation

i	X_i	$\bar{Y}_{i.}$	$\hat{\sigma}_i/\sqrt{n}$
1	0.10	0.110601	0.000131466
2	0.20	0.248065	0.000533648
3	0.30	0.429343	0.00145294
4	0.40	0.670248	0.00146284
5	0.50	0.987577	0.00335524
6	0.55	1.21614	0.00637383
7	0.60	1.50915	0.0154060
8	0.70	2.38149	0.0742331
9	0.75	3.09417	0.152555
10	0.85	5.72853	0.201175
11	0.90	8.95594	0.932344
12	0.95	18.8805	4.09515

Our **suggested procedure, for fitting the simulation metamodel**, consists of the steps that follow.

1. **Identifying** a tentative nonlinear relation between the response and the decision variable.

   Ideally, we should select a curve based on physical justifications. Pragmatically, we usually do that visually, just like we compare empirical histograms with known density functions for selecting a random distribution. A convenient first step is to represent the *dispersion diagram* (or scatterplot) of the response and the decision variable, plotting the corresponding pairs (X_i, Y_{ij}), for $i = 1, m$ and $j = 1, n$. In Figure 1, we graphically display the results of our experiment. We observe that the average waiting time in queue is actually related in a nonlinear fashion with the utilization factor. If we had two decision variables, we might draw contour curves or use a three dimensional

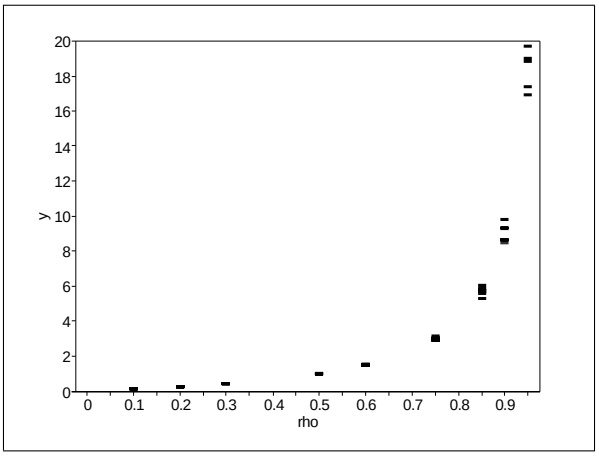

Figure 1: Dispersion Diagram for $M/M/1$ Queue

visualization software, instead. In order to identify the type of nonlinear relation, we advance to the next step.

2. **Selecting** a curve from a catalog of typical nonlinear functional relationships.

To facilitate the identification of tentative nonlinear relations, we might build a catalog of different functional relationships, with their graphical representations. Due to space restrictions, we reproduce in Figures 2 and 3, only a small part of one such catalog. Comparing the actual dispersion diagram of Figure 1 with this part of the catalog of curves, it is likely that an hyperbole might fit the data. So, we chose this functional relationship for the tentative simulation metamodel relating the average waiting time in the $M/M/1$ queue with the utilization factor:

$$\bar{Y}_{i.} = \theta_1 X_i/(1+\theta_2 X_i) + \bar{\epsilon}_{i.}, i = 1, \ldots, 12,$$

with $\bar{\epsilon}_{i.} \sim \mathrm{N}\left(0, \sigma_i^2/10\right)$ and $Y_{ij} \sim \mathrm{N}(\mu_i, \sigma_i^2)$. As we mentioned in Section 2.2, this hypothesis of normality is generally satisfied when the simulation responses are averages, which is the case (we are analyzing the average waiting time in queue). Note that the selected function (a hyperbole) is not linearizable.

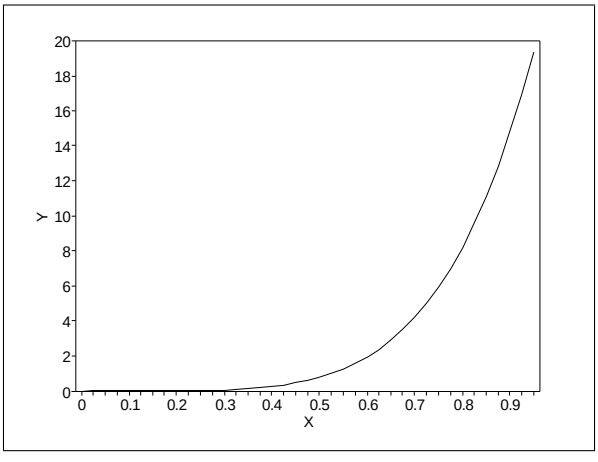

Figure 2: Monomial, $y = \theta_1 x^{\theta_2}, \theta_1 = 25, \theta_2 = 5$

For this particular system, we know the theoretical expected steady-state response. To illustrate what might have happened, if we had chosen to fit another functional relationship to the M/M/1 data (for instance, the monomial in Figure 2, $Y_{ij} = \theta_1 X_i^{\theta_2} \epsilon_{ij}$), we perform the next two steps of the procedure for both models above. Since the metamodel to fit is now linearizable, we take the decimal logarithm of both sides and obtain $\log Y_{ij} = \log\theta_1 + \theta_2 \log X_i + \log\epsilon_{ij}$, the familiar simple linear regression equation.

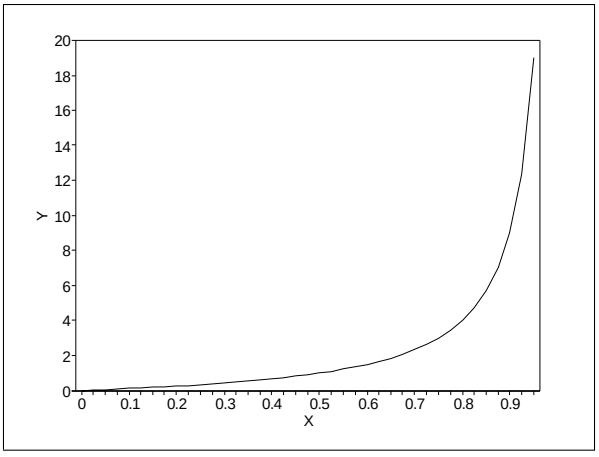

Figure 3: Hyperbole, $y = \frac{\theta_1 x}{1+\theta_2 x}, \theta_1 = 1, \theta_2 = -1$

3. **Estimating** the nonlinear simulation metamodel.

From Figure 1, we observe that the variance of the response increases with the utilization factor. Thus, we must use the least squares estimator given by (5) and satisfying (6), with $n = 10$ and $p = 2$:

$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \frac{1}{10}[\mathbf{F}^T \boldsymbol{\Sigma}^{-1} \mathbf{F}]^{-1}\right),$$

where $\mathbf{F} = \mathbf{F}(\theta_1^*, \theta_2^*)$.

In order to obtain the approximate least squares estimator of θ_1 and θ_2, we used the iterative Gauss-Newton method. The approximate solution was found when $|(SSE^{k+1} - SSE^k)/SSE^{k+1}| < 10^{-6}$ for five consecutive values of k, where SSE^k is the residual sum of squares in iteration k.

In Table 2, we present the main results of the estimation of the selected simulation metamodel: the approximate LSE of θ_1 e θ_2, as well as their corresponding asymptotic standard errors. We also reproduce the corresponding values for the monomial case (Model 2).

Table 2: Estimated Metamodel Coefficients

Model	Coeff.	Estimator	Standard error
1	θ_1	1.0004	0.00882
	θ_2	-1.0000	0.00174
2	$\log\theta_1$	0.7451	0.00548
	θ_2	2.0254	0.01351

However, the metamodel has to be validated, to determine if it is indeed the "elected" model. This topic is dealt with in the next step.

4. **Validating** the nonlinear simulation metamodel.

Table 3 reproduces the significance tests performed on the estimated metamodel. We can observe that the proposed model explains rather well the simulation model response, through the factor $X = \rho$: clearly, the

F test for lack of fit is not significant ($F_{10,108} \approx 1.93$). Thus, there is no evidence to reject this model and try another one. We should also emphasize that the graphical analysis of the residuals also suggests that the assumptions of regression analysis are met. However, some authors argue that, in the linear case, this F test is not very sensitive to departures from the normality and homogeneous variance assumptions; see Kleijnen, Burg and Ham (1979). That is, it may happen that the test may have a low probability of detecting if the fitted model does provide a good fit. The nonlinear case is treated performing a linearization. Consequently, the same problem may occur in the nonlinear case. Thus, the acceptance of the fitted metamodel can not depend solely on the value of this statistic.

In the same table, we show the corresponding values we would have obtained, had we chosen the monomial function. It is quite obvious that there is evidence to reject the monomial fit.

Table 3: Testing for Lack-of-Fit

Mod	Source	D f	SS	MS	F
1	Lack	10	1.891	0.189	0.189
	Error	108	108.0	1.000	
2	Lack	10	3580.	358.0	358.0
	Error	108	108.0	1.000	

Besides testing the metamodel adequacy, we must also test its predictive validity. We notice, in Table 4, that the values of $PRESS = 111.005$ and $SSE = 109.891$ are rather close. This supports the validity of the fitted regression metamodel. It also stresses the importance of MSE as an indicador of the predictive capability of this model.

Again, we also present the corresponding diagnostics for Model 2 (the monomial function). The values are consistent with the rejection of this model.

Table 4: Metamodel Diagnostics

Statistic	Hyperbole	Monomial
SSE	109.891	3688.38
$PRESS$	111.005	6603.47
MSE	0.9313	31.2575
R^2	0.9444	0.883554
R^2_{adj}	0.9491	0.893258

The main results of the metamodel validation procedure, based on the regressions on the model-building and validation data sets, are reproduced in Table 5. We present the estimated regression model coefficients, their standard deviations and some other related statistics. Notice that there is good agreement between the two sets of estimated regression coefficients, and between the values of statistics MSE and R^2 for both cases. The fact that $MSPR$ is not significantly different from MSE implies that the mean squared error, MSE, based on the model-building data set, is a reasonably valid indication of the predictive capability of the fitted regression model. These validation results support the appropriateness of the selected simulation metamodel.

Table 5: Metamodel Validity Test

	Hyperbole		Monomial	
Statistic	M-bld	Valid	M-bld	Valid
$\hat{\theta_1}$	.9965	.9940	–	–
$\log\hat{\theta_1}$	–	–	.7870	.6407
$\hat{\sigma}(\hat{\theta_1})$	.0117	.0099	–	–
$\hat{\sigma}(\log\hat{\theta_1})$	–	–	.0958	.1135
$\hat{\theta_2}$	-1.001	$-.9982$	1.996	1.940
$\hat{\sigma}(\hat{\theta_2})$	.0021	.0033	.0028	.0057
SSE	66.99	46.48	1912.	642.9
$PRESS$	73.23	–	4885.	–
MSE	.9569	1.010	27.31	13.98
$MSPR$	–	1.333	–	50.93
R^2	.9011	.9855	.5896	.9033
R^2_{adj}	.9093	.9867	.6238	.9113

On the other hand, for the monomial (model 2), the same statistics convey the opposite information. So, as before, this model would be rejected based on this validation procedure.

In conclusion, we can say that, although the linear(ized) model is much simpler to fit, we strongly feel that there will be many situations in which the advantages of a nonlinear model will overcome the extra time and computation needed.

5 CONCLUSIONS

In this work, we have addressed some of the most important issues involved in the estimation of nonlinear simulation metamodels. Although they are more complex and time-consuming than their linear counterparts, nonlinear metamodels account for a larger part of the variability of the simulation output and have fewer parameters. We have shown that it is feasible, for an informed practitioner, to apply our proposed procedure for metamodel estimation. The statistical procedures that we propose for validation also seem to be reasonably discriminating between a good and a bad choice for the metamodel structure.

We strongly feel that more work needs to be done in this area. The catalog of functional relationships has to be enlarged with more functions of one or two independent variables. It would be useful to come up with specific visually oriented approaches to help in the curve selection, when the metamodel includes more than two independent variables. Finally, the construction of confidence intervals

and hypotheses testing are two additional topics deserving further attention.

REFERENCES

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

Friedman, L. W. and H. H. Friedman. 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):1164–1185.

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

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

Neter, J., W. Wasserman, and M. H. Kutner. 1989. *Applied Linear Regression Models*. R. R. Donnelley & Sons Company, second edition.

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

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

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

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

Zeigler, B., 1976. *Theory of Modeling and Simulation*. John Wiley and Sons, New York, NY, USA.

AUTHOR BIOGRAPHIES

M. ISABEL REIS DOS SANTOS is a lecturer in the Department of Mathematics at the Superior Technical Institute (IST) of the Technical University of Lisbon. She is a Ph.D. candidate in industrial engineering and management at IST. She received a B.S. degree in applied mathematics and computation and a M. S. degree in applied mathematics from IST.

ACÁCIO M. O. PORTA NOVA is an associate professor in the Autonomous Section of Economics and Management at the Superior Technical Institute (IST) of the Technical University of Lisbon. He received a B.S. degree in electrical engineering from IST. He received a Ph.D. in operations research from the University of Texas at Austin.