Journal of Statistical Computation and Simulation
Vol. 00, No. 0, Month 2004, pp. 1–15

ESTIMATION OF NONLINEAR SIMULATION METAMODELS USING CONTROL VARIATES

M. ISABEL REIS SANTOS[a,*] and ACÁCIO M. O. PORTA NOVA[b,†]

[a]*Department of Mathematics and Centre for Management Studies, Technical University of Lisbon, Av. Rovisco Pais, 1049-001 Lisboa, Portugal;* [b]*Department of Engineering and Management and Centre for Management Studies, Technical University of Lisbon, Av. Rovisco Pais, 1049-001 Lisboa, Portugal*

(Revised 8 January 2004; In final form 17 August 2004)

The method of control variates has been intensively used for reducing the variance of estimated (linear) regression metamodels in simulation experiments. In contrast to previous studies, this article presents a procedure for applying multiple control variates when the objective is to estimate and validate a nonlinear regression metamodel for a single response, in terms of selected decision variables. This procedure includes robust statistical regression techniques for estimation and validation. Assuming joint normality of the response and controls, confidence intervals and hypothesis tests for the metamodel parameters are obtained. Finally, results for measuring the efficiency of the use of control variates are discussed.

Keywords: Control variates; Nonlinear metamodels; Simulation; Variance reduction techniques; Modeling

1 INTRODUCTION

Computer simulation models are commonly used for estimating and validating metamodels. A simulation metamodel is a mathematical relationship between the input (input parameters or design variables) and the output (response) of the computer simulation model; see Barton (1992). If this auxiliary model is an accurate representation of the simulation model, it can be very useful for prediction and sensitivity analysis, although it uses fewer computer resources, when compared with the more time consuming and expensive simulation program. To improve the efficiency of metamodel estimation, it is common to use the control variates technique. This technique is one of the most widely used variance reduction methods, because it is not very difficult to implement, it is a general method and it does not alter the underlying stochastic process.

Many authors have studied the method of control variables in the context of linear metamodel estimation; see, for example, Nozari *et al.* (1984), Porta Nova and Wilson (1989) and Shih and Song (1995). In particular, the polynomial form of the general linear regression model has been extensively analyzed. However, polynomials are unable to produce a global fit to curves of arbitrary shape. Moreover, in real-life systems nonlinearity is common

* Corresponding author. E-mail: prs@digitais.ist.utl.pt or prs@asterix.ist.utl.pt
† E-mail: apnova@ist.utl.pt

ISSN 0094-9655 print; ISSN 1563-5163 online
DOI: 10.1080/00949650412331299201

and approximation using polynomials becomes unrealistic. Consequently, in these problems, polynomials often do not provide good fits, *e.g.*, in problems involving queuing systems (Friedman and Friedman, 1985). An alternative that provides better and more realistic global fits is the use of statistical nonlinear regression techniques (Santos and Porta Nova, 1999). In this article, we apply the method of control variates to the estimation and validation of a nonlinear metamodel of a single simulation response, expressed in terms of multiple inputs. In fact, the procedure presented here is a generalization of the work of Nozari *et al.* (1984) to nonlinear simulation metamodels.

This article is organized as follows. In Section 2, we formulate the nonlinear metamodeling problems, both without and with control variates. In Section 3, we obtain some distribution-free results. In Section 4, we discuss the metamodel estimation problem under a joint normality assumption. The minimum variance ratio and the loss factor are also obtained. In Section 5, the methodology described in this article is illustrated using a simple $M/M/1$ queuing system. Finally, Section 6 is reserved for conclusions.

2 NONLINEAR METAMODEL ESTIMATION

2.1 Nonlinear Metamodels

Consider an experimental design consisting of n different design points, defined by the d decision variables $\{X_{il}\colon i = 1, \ldots, n; l = l, \ldots, d\}$. For each design point, r independent replications of the simulation model are carried out and the experiment yields $\{(Y_{ij}, \mathbf{C}_{ij})\colon i = 1, \ldots, n, j = 1, \ldots, r\}$, where Y_{ij} is the relevant system response and $\mathbf{C}_{ij}$ is a vector of q concomitant control variables, with a known mean. Without loss of generality, we assume $\mathcal{E}[\mathbf{C}_{ij}] = 0$, with $i = 1, \ldots, n$ and $j = 1, \ldots, r$. Suppose that the simulation model (computer program) can be represented by the metamodel

$$Y_{ij} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \varepsilon_{ij}, \tag{1}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, r$, where $\varepsilon_{ij} \sim \text{NID}(0, \sigma^2)$, with $\sigma > 0$, and $\boldsymbol{\theta}$ is an $m \times 1$ vector of unknown parameters. Under mild regularity conditions, every nonlinear control variable scheme behaves asymptotically like a linear control variable scheme; see Glynn and Whitt (1989) and Loh (1994). As a result, we only consider linear schemes involving control variables. Thus, we assume that the error $\varepsilon_{ij} = Y_{ij} - f(\mathbf{X}_{i.}, \boldsymbol{\theta})$, in problem (1), has a linear regression on the control vector $\mathbf{C}_{ij}$, with an unknown $q \times 1$ vector of control coefficients $\boldsymbol{\delta}$ and an error ε_{ij}. This way, the simulation model can also be represented by the replicated simulation metamodel

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

with $i = 1, \ldots, n$ and $j = 1, \ldots, r$.

Let $\mathbf{Z}$ be the following random matrix:

$$\mathbf{Z} = \begin{bmatrix} Y_{11} & C_{111} & \cdots & C_{11q} \\ \vdots & \vdots & & \vdots \\ Y_{n1} & C_{n11} & \cdots & C_{n1q} \\ \vdots & \vdots & & \vdots \\ Y_{1r} & C_{1r1} & \cdots & C_{1rq} \\ \vdots & \vdots & & \vdots \\ Y_{nr} & C_{nr1} & \cdots & C_{nrq} \end{bmatrix} \tag{3}$$

Assume that the row vector $\mathbf{Z}_{l.}$ is continuous and has the same probability density function for all $l = 1, \ldots, N = nr$, such that the following dispersion matrix exists:

$$\Sigma = \mathcal{D}[\mathbf{Z}_{l.}] = \begin{bmatrix} \sigma^2 & \boldsymbol{\sigma}_{YC} \\ \boldsymbol{\sigma}_{CY} & \Sigma_C \end{bmatrix}, \tag{4}$$

where $\sigma^2 = \text{Var}[Y_{ij}]$. As a consequence, $\Sigma_C = \mathcal{D}[\mathbf{C}_{ij}]$ is nonsingular and also positive definite with probability one, for all replications of all experimental points (Porta Nova and Wilson, 1989). The covariance vector between Y_{ij} and $\mathbf{C}_{ij}$, denoted by $\boldsymbol{\sigma}_{YC} = \mathcal{C}[Y_{ij}, \mathbf{C}_{ij}]$, is assumed to be constant for all $i = 1, \ldots, n$ and $j = 1, \ldots, r$, with $\boldsymbol{\sigma}_{CY} = \boldsymbol{\sigma}_{YC}^{\mathrm{T}}$.

To simplify the estimation procedure, instead of problems (1) and (2), we consider respectively the equivalent least squares problems, in which the individual observations, at each design point, are replaced by their averages across runs:

$$\bar{Y}_{i.} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \bar{\varepsilon}_{i.}, \quad i = 1, 2, \ldots, n, \tag{5}$$

with $\bar{\varepsilon}_{i.} \sim \text{NID}(0, \sigma_{\bar{Y}}^2)$, $\sigma_{\bar{Y}}^2 = \text{Var}[\bar{Y}_{i.}] = \sigma^2/r$ and

$$\bar{Y}_{i.} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \bar{\mathbf{C}}_{i.}\boldsymbol{\delta} + \bar{\varepsilon}_{i.}, \quad i = 1, 2, \ldots, n, \tag{6}$$

where $\bar{\mathbf{C}}_{i.} = (\bar{C}_{i.1}, \ldots, \bar{C}_{i.q})$, with $\bar{C}_{i.k} = 1/r \sum_{j=1}^{r} C_{ijk}$.

2.2 Objectives

In this article, two kinds of results are exposed:

(i) Assuming that the metamodel (2) is valid, we obtain the approximated minimum variance ratio, the nonlinear least squares estimator $\hat{\boldsymbol{\delta}}$ (for the true vector of control coefficients $\boldsymbol{\delta}$), and the corresponding controlled nonlinear least squares estimator $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$ (to estimate the true vector of metamodel coefficients $\boldsymbol{\theta}$).

(ii) Assuming that the response and the control variables have a joint multivariate normal distribution, we derive the approximated loss factor and we construct asymptotic confidence regions for $\boldsymbol{\theta}$. We also propose procedures for testing hypotheses about the metamodel parameters.

3 GENERAL RESULTS ON METAMODEL ESTIMATION

In this section, we present results on metamodel estimation using control variables that do not depend on the assumption of joint normality between the response and the control variables.

3.1 Minimum Variance Ratio

If the Jacobian matrix $\mathbf{F}$ of $\mathbf{f} = (f(\mathbf{X}_{1.}, \boldsymbol{\theta}^*), \ldots, f(\mathbf{X}_{n.}, \boldsymbol{\theta}^*))^{\mathrm{T}}$ has full column rank m, then we apply result (12.21) of Seber and Wild (1989) to problem (5), obtaining the following asymptotic ordinary nonlinear least squares estimator of $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^* + (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}[\bar{\mathbf{Y}} - \mathbf{f}], \tag{7}$$

where $\boldsymbol{\theta}^*$ is the exact value of $\boldsymbol{\theta}$ and $\bar{\mathbf{Y}} = (\bar{Y}_{1.}, \ldots, \bar{Y}_{n.})^{\mathrm{T}}$ (to simplify the notation, we use $\mathbf{F} = \mathbf{F}(\boldsymbol{\theta}^*)$ and $\mathbf{f} = \mathbf{f}(\boldsymbol{\theta}^*)$). The mean and the covariance of this estimator are obtained applying (12.23) of Seber and Wild (1989) to problem (5),

$$\mathcal{E}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}^*, \quad \mathcal{D}[\hat{\boldsymbol{\theta}}] = \frac{\sigma^2}{r}(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}. \tag{8}$$

When control variables are observed, and for a fixed vector of control coefficients $\boldsymbol{\phi}$, the least squares estimator of $\boldsymbol{\theta}$ is given approximately by

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) \approx \boldsymbol{\theta}^* + (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}[\bar{\mathbf{Y}} - \bar{\mathbf{C}}\boldsymbol{\phi} - \mathbf{f}], \tag{9}$$

where

$$\bar{\mathbf{C}} = \begin{bmatrix} \bar{C}_{1.1} & \cdots & \bar{C}_{1.q} \\ \vdots & & \vdots \\ \bar{C}_{n.1} & \cdots & \bar{C}_{n.q} \end{bmatrix}$$

This estimator is obtained representing problem (6) in the form $\bar{Y}_{i.} - \bar{\mathbf{C}}_{i.}\boldsymbol{\phi} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \bar{\varepsilon}_{i.}$ and then determining the ordinary nonlinear least squares estimator as in Eq. (7) (considering a fixed $\boldsymbol{\phi}$).

The approximation (9) is equivalent to

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) \approx \boldsymbol{\theta}^* + (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathrm{F}^{\mathrm{T}}[\bar{\mathbf{Y}} - \mathbf{f}] - (\mathrm{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\bar{\mathbf{C}}\boldsymbol{\phi},$$

and, when control variables are not used, the least squares estimator is given by Eq. (7), therefore

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) \approx \hat{\boldsymbol{\theta}} - (\mathrm{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\bar{\mathbf{C}}\boldsymbol{\phi},$$

that is, $\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) \neq \hat{\boldsymbol{\theta}}$, or in other words, observing control variables with known means results in a different estimator of $\boldsymbol{\theta}$. As a consequence, if the random matrix has a probability density function, then using $\mathcal{E}[\bar{\mathbf{C}}] = 0$ and $\mathcal{E}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}^*$, we obtain

$$\mathcal{E}[\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})] = \boldsymbol{\theta}^*. \tag{10}$$

To obtain the dispersion matrix $\mathcal{D}[\hat{\boldsymbol{\theta}}(\boldsymbol{\delta})]$, it is useful to write the estimator (9) in the form $\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) \approx \mathbf{A}[\bar{\mathbf{Y}} - \bar{\mathbf{C}}\boldsymbol{\phi}] + \mathbf{b}$, where $\mathbf{A} = (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}$ and $\mathbf{b} = \boldsymbol{\theta}^* - (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{f}$. Thus, we can write

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})] = \mathbf{A}\mathcal{D}[\bar{\mathbf{Y}} - \bar{\mathbf{C}}\boldsymbol{\phi}]\mathbf{A}^{\mathrm{T}}, \tag{11}$$

where $\bar{\mathbf{Y}} - \bar{\mathbf{C}}\boldsymbol{\phi} = (\bar{Y}_{1.} - \bar{\mathbf{C}}_{1.}\boldsymbol{\phi}, \ldots, \bar{Y}_{n.} - \bar{\mathbf{C}}_{n.}\boldsymbol{\phi})^{\mathrm{T}}$. As this work is in the context of the method of independent replications, we have

$$\mathrm{Cov}[\bar{Y}_{i.} - \bar{\mathbf{C}}_{i.}\boldsymbol{\phi}, \bar{Y}_{i',} - \bar{\mathbf{C}}_{i',}\boldsymbol{\phi}] = 0, \quad i \neq i'. \tag{12}$$

Moreover,

$$\mathrm{Var}[\bar{Y}_{i.} - \bar{\mathbf{C}}_{i.}\boldsymbol{\phi}] = \sigma_{\bar{Y}}^2 + \boldsymbol{\phi}^{\mathrm{T}}\Sigma_{\bar{C}}\boldsymbol{\phi} - 2\boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\sigma}_{\bar{C}\bar{Y}}. \tag{13}$$

The vector of control coefficients that minimizes this variance is given by $\boldsymbol{\delta} = \Sigma_{\bar{C}}^{-1}\boldsymbol{\sigma}_{\bar{C}\bar{Y}}$; see Eqs. (8) and (9) of Lavenberg and Welch (1981). But $\Sigma_{\bar{C}} = \Sigma_C/r$ and $\boldsymbol{\sigma}_{\bar{C}\bar{Y}} = \boldsymbol{\sigma}_{CY}/r$, and as a result

$$\boldsymbol{\delta} = \Sigma_C^{-1}\boldsymbol{\sigma}_{CY.} \tag{14}$$

Substituting Eq. (14) in Eq. (13), we obtain

$$\text{Var}[\bar{Y}_{i.} - \bar{\mathbf{C}}_{i.}\boldsymbol{\delta}] = \frac{1}{r}\tau^2, \tag{15}$$

where

$$\tau^2 = \sigma^2 - \boldsymbol{\sigma}_{YC}\Sigma_C^{-1}\boldsymbol{\sigma}_{CY}. \tag{16}$$

Equation (12), (15) and (16) imply that $\mathcal{D}[\bar{\mathbf{Y}} - \bar{\mathbf{C}}\boldsymbol{\delta}] = 1/r(\sigma^2 - \boldsymbol{\sigma}_{YC}\boldsymbol{\Sigma}_C^{-1}\boldsymbol{\sigma}_{CY})\mathbf{I}_n$. Substituting this dispersion matrix in Eq. (11) and as $\mathbf{A} = (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}$, we have

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\boldsymbol{\delta})] = \frac{1}{r}(\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}_{YC}\Sigma_C^{-1}\boldsymbol{\sigma}_{CY})(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}. \tag{17}$$

Using Eqs. (8) and (17), we conclude that the maximum reduction in variance that is possible to obtain, with the use of control variables is given approximately by the minimum variance ratio

$$\eta(\boldsymbol{\delta}) = \frac{|\mathcal{D}[\hat{\boldsymbol{\theta}}(\boldsymbol{\delta})]|}{|\mathcal{D}[\hat{\boldsymbol{\theta}}]|} \approx 1 - \rho_{YC}^2, \tag{18}$$

where $\rho_{YC}^2 = \boldsymbol{\sigma}_{YC}\boldsymbol{\Sigma}_C^{-1}\boldsymbol{\sigma}_{CY}/\sigma^2$ is the multiple correlation coefficient between Y_{ij} and C_{ij}, $i = 1, \ldots, n$ and $j = 1, \ldots, r$. In general, as in the linear case, $\boldsymbol{\delta}$ is unknown, so it must be estimated and, as a consequence, the variance will increase. We will use a loss factor to quantify the percentage increase in variance when $\boldsymbol{\delta}$ must be estimated.

3.2 Controlled Nonlinear Least Squares Estimator

To obtain estimators for the unknown true parameters $\boldsymbol{\delta}$ and $\boldsymbol{\theta}$, $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$, we resort to the method of nonlinear least squares. Given appropriate regularity conditions (Seber and Wild, 1989), then for large N, the least squares estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ in Eq. (6) satisfy, approximately:

$$\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) \approx \boldsymbol{\theta}^* + (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}[\bar{\mathbf{Y}} - \mathbf{f} - \mathbf{C}\hat{\boldsymbol{\delta}}], \tag{19}$$

$$\hat{\boldsymbol{\delta}} \approx (\mathbf{C}^{\mathrm{T}}\mathbf{P}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{P}[\bar{\mathbf{Y}} - \mathbf{f}], \tag{20}$$

where

$$\mathbf{P} = \mathbf{I}_n - \mathbf{F}(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}. \tag{21}$$

These results are obtained as follows. Taylor's series expansion of $f_i = f(\mathbf{X}_{i.}, \boldsymbol{\theta})$ about the point $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ yields $\mathbf{f}(\boldsymbol{\theta}) \approx \mathbf{f}(\boldsymbol{\theta}^*) + \mathbf{F}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)$. As a result, Eq. (6) becomes

$$\bar{\mathbf{Y}} - \mathbf{f}(\boldsymbol{\theta}^*) \approx \mathbf{F}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) + \mathbf{C}\boldsymbol{\delta} + \bar{\varepsilon}. \tag{22}$$

Applying Eqs. (6) and (8) of Searle (1971), pp. 341 and 342, to this (linearized) problem (22), we can write

$$\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^* \approx (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}(\bar{Y} - \mathbf{f}) - (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{C}\hat{\boldsymbol{\delta}},$$

As a consequence, we obtain Eq. (19). Finally, Eqs. (20) and (21) are obtained using Eqs. (10) and (11) of Searle (1971), p. 342.

3.3 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 (6), we propose the following F-test for lack of fit (see Seber and Wild (1989), p. 32):

$$F = \frac{(\mathrm{SSE} - \mathrm{SSPE})/(n - m - q)}{\mathrm{SSPE}/(N - n)},$$

where, in this situation, $\mathrm{SSE} = \sum_{i=1}^{n} \sum_{j=1}^{r} [Y_{ij} - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}}) - \mathbf{C}_{ij}\hat{\boldsymbol{\delta}}]^2$ is the usual residual sum of squares and $\mathrm{SSPE} = \sum_{i=1}^{n} \sum_{j=1}^{r} (Y_{ij} - \bar{Y}_{i.})^2$ is the pure error sum of squares. When the metamodel is valid and if there exists a parameterization for which it can be adequately approximated by a linear model, then F is roughly distributed as an $F_{n-m-q,n(r-1)}$ distribution.

4 RESULTS FOR NORMAL NONLINEAR METAMODELS

To simplify the presentation of the results, when the response and the control variables have a joint normal distribution, we will introduce some additional notations and hypotheses. Consider the models (5) and (6). Suppose that, for the lth design point,

$$\mathbf{Z}_{l.} \sim \mathrm{N}_{q+1}((\boldsymbol{\mu}_l, \mathbf{0}^{\mathrm{T}}), \Sigma), \tag{23}$$

where Σ is given by Eq. (4). As a result, the $N \times (q + 1)$ random matrix, defined by Eq. (3), has a multivariate normal distribution

$$\mathbf{Z} \sim \mathrm{N}_{N,q+1}(\boldsymbol{\mu}_Z, \Xi, \Sigma), \tag{24}$$

with unknown $\boldsymbol{\sigma}$ and $\mathcal{E}[\mathbf{Z}] = \boldsymbol{\mu}_{\mathbf{Z}} = (\boldsymbol{\mu}_Y, \mathbf{0})$, where

$$\boldsymbol{\mu}_Y = (f(\mathbf{X}_{1.}, \boldsymbol{\theta}), \ldots, f(\mathbf{X}_{n.}, \boldsymbol{\theta}), \ldots, f(\mathbf{X}_{1.}, \boldsymbol{\theta}), \ldots, f(\mathbf{X}_{n.}, \boldsymbol{\theta}))^{\mathrm{T}}.$$

The dispersion matrix between the ith and the kth rows is $\mathcal{D}[Z_{i.}, Z_{k.}] = \Xi_{ik}\Sigma$ for $1 \le i, k \le N$ and the dispersion matrix between the jth and the lth columns is $\mathcal{D}[Z_{.j}, Z_{.l}] = \Sigma_{jl}\Xi$ for $1 \le j, l \le q + 1$. Suppose also that Ξ and $\boldsymbol{\sigma}$ are positive definite. Moreover, the rows of $\mathbf{Z}$ are mutually independent, because they correspond to independent executions of the simulation program. As a consequence, we consider $\Xi = \mathbf{I}_N$ in the following development. If the $q \times q$ matrix Σ_C is positive definite, then the conditional distribution of $\mathbf{Y}$ given $\mathbf{C}$ is given by

$$\mathbf{Y}|\mathbf{C} \sim \mathrm{N}_N(\boldsymbol{\mu}_{Y.C}, \tau^2\mathbf{I}_N), \tag{25}$$

with τ_2 given by Eq. (16) and

$$\boldsymbol{\mu}_{Y.C} = \boldsymbol{\mu}_Y + \mathbf{C}\Sigma_C^{-1}\boldsymbol{\sigma}_{CY}; \tag{26}$$

see Theorem 17.2-g) of Arnold (1981). As a consequence, conditioning on $\mathbf{C}$, we conclude that the correct metamodel is Eq. (2). In the following development, we will see that the asymptotic nonlinear least squares estimators are unbiased, both conditionally and unconditionally. Moreover, the approximated confidence region, for the true metamodel parameter vector, centered at $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$ will also be obtained.

4.1 Distribution of the Controlled Estimator

If $\mathbf{Z} \sim \mathrm{N}_{N,q+1}(\boldsymbol{\mu}_Z, \mathbf{I}_N, \Sigma)$ with unknown $\boldsymbol{\theta}$ and Σ, then we will use the fact that the conditional distribution of $\mathbf{Y}$ given $\mathbf{C}$ is normal and given by Eq. (25). Conditioning on $\mathbf{C}$, we see that the correct nonlinear metamodel for a normal response is

$$\mathbf{Y} = \tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}) + \mathbf{C}\boldsymbol{\delta} + \mathbf{e}, \tag{27}$$

where $\tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}) = (f(\mathbf{X}_{1.}, \boldsymbol{\theta}), \ldots, f(\mathbf{X}_{1.}, \boldsymbol{\theta}), \ldots, f(\mathbf{X}_{n.}, \boldsymbol{\theta}), \ldots, f(\mathbf{X}_{n.}, \boldsymbol{\theta}))^{\mathrm{T}}$ (a vector with N components) and $\mathbf{e} = (\varepsilon_{11}, \ldots, \varepsilon_{1r}, \ldots, \varepsilon_{n1}, \ldots, \varepsilon_{nr})^{\mathrm{T}}$. Using Taylor's expansion in a neighborhood of $(\mathbf{X}, \boldsymbol{\theta}^*)$, we obtain

$$\tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}) \approx \tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}^*) + \tilde{\mathbf{F}}(\boldsymbol{\theta} - \boldsymbol{\theta}^*), \tag{28}$$

where $\tilde{\mathbf{F}}$ is the Jacobian matrix of $\tilde{\mathbf{f}}$, calculated at $(\mathbf{X}, \theta^*)$. As a result, Eq. (27) can be rewritten as

$$\mathbf{G} \approx \tilde{\mathbf{F}}\lambda + \mathbf{C}\boldsymbol{\delta} + \mathbf{e} \tag{29}$$

where $\mathbf{G} = \mathbf{Y} - \tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}^*)$ and $\lambda = \boldsymbol{\theta} - \boldsymbol{\theta}^*$. Applying Searle (1971), p. 342, to problem (29), we obtain $\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}}) \approx (\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^{\mathrm{T}}\mathbf{G} - (\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^{\mathrm{T}}\mathbf{C}\hat{\boldsymbol{\delta}}$, where $\hat{\boldsymbol{\delta}} = (\mathbf{C}^{\mathrm{T}}\tilde{\mathbf{P}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\tilde{\mathbf{P}}\mathbf{G}$ and $\tilde{\mathbf{P}} = \mathbf{I}_N - \tilde{\mathbf{F}}(\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^{\mathrm{T}}$. That is,

$$\hat{\lambda}(\hat{\boldsymbol{\delta}}) \approx \mathbf{B}\mathbf{G} \quad \text{with} \quad \mathbf{B} = (\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^{\mathrm{T}}[\mathbf{I}_N - \mathbf{C}(\mathbf{C}^{\mathrm{T}}\tilde{\mathbf{P}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\tilde{\mathbf{P}}]. \tag{30}$$

As Eq. (25) is verified, we have approximately $\mathbf{G}|\mathbf{C} \sim \mathrm{N}_N(\boldsymbol{\mu}_{G.C}, \tau^2\mathbf{I}_N)$, where $\boldsymbol{\mu}_{G.C} = \boldsymbol{\mu}_{Y.C} - \tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}^*)$. Using this result and Eq. (30), we can apply Theorem 17.2-d) of Arnold (1981) and therefore $\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}})|\mathbf{C} \sim \mathrm{N}_m(\mathbf{B}\boldsymbol{\mu}_{G.C}, \tau^2\mathbf{B}\mathbf{B}^{\mathrm{T}})$. Suppose that $\tilde{\mathbf{F}}$ has rank m, then $\tilde{\mathbf{P}}$ is an orthogonal projection of $\mathbb{R}^N$ into $\mathcal{R}(\tilde{F})^{\perp}$; see Seber and Wild (1989). As $\tilde{\mathbf{P}}$ is an orthogonal projection, then $(\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{P}} = \tilde{\mathbf{P}}\tilde{\mathbf{F}}(\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1} = 0$ and, as a consequence, $\mathbf{B}\mathbf{B}^{\mathrm{T}} = (\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1}[\mathrm{I}_m + \tilde{\mathbf{F}}^{\mathrm{T}}\mathbf{C}(\mathbf{C}^{\mathrm{T}}\tilde{\mathbf{P}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\tilde{\mathbf{F}}(\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1}]$. Applying Nozari (1982), p. 121, to the linearized problem (29), we obtain $\mathcal{E}[\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] = \lambda$. As a result, $\mathbf{B}\boldsymbol{\mu}_{G.C} = \boldsymbol{\lambda}$ and we have $\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}})|\mathbf{C} \sim \mathrm{N}_m(\lambda, \tau^2\mathbf{B}\mathbf{B}^{\mathrm{T}})$. But $\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}}) = \hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*$ and $\boldsymbol{\lambda} = \boldsymbol{\theta} - \boldsymbol{\theta}^*$, therefore $\mathcal{E}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] = \boldsymbol{\theta}$ (given C, $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$ is an unbiased estimator of θ), $\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] = \tau^2\mathbf{B}\mathbf{B}^{\mathrm{T}}$ and

$$\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C} \sim \mathrm{N}_m(\boldsymbol{\theta}, \tau^2\mathbf{B}\mathbf{B}^{\mathrm{T}}). \tag{31}$$

To construct a confidence region for $\boldsymbol{\theta}$ centered at $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$, we will present an estimator of τ^2. Applying Theorem 3.3 of Porta Nova (1985) to problem (29), we obtain

$$\hat{\mathbf{e}}^{\mathrm{T}}\hat{\mathbf{e}}|\mathbf{C} \sim W_1(N - m - q, \tau^2, 0),$$

that is, $\hat{\mathbf{e}}^{\mathrm{T}}\hat{\mathbf{e}}|\mathbf{C} \sim \tau^2\chi^2_{N-m-q}$, where $\hat{\mathbf{e}} = \mathbf{G} - \tilde{\mathbf{F}}\hat{\lambda} - \mathbf{C}\hat{\boldsymbol{\delta}}$; see Theorem 17.6-b) of Arnold (1981). But $\mathbf{G} = \mathbf{Y} - \tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}^*)$ and $\hat{\lambda} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*$, then $\hat{\mathbf{e}} = \mathbf{Y} - \tilde{\mathbf{f}}(\mathbf{X}, \hat{\boldsymbol{\theta}}) - \mathbf{C}\hat{\boldsymbol{\delta}}$. As a result, given $\mathbf{C}$, an unbiased estimator for τ^2 is given by

$$\hat{\tau}^2|\mathbf{C} = \frac{\hat{\mathbf{e}}^{\mathrm{T}}\hat{\mathbf{e}}}{N - m - q} = \frac{1}{N - m - q}\sum_{i=1}^{n}\sum_{j=1}^{r}[Y_{ij} - f(\mathrm{X}_{i.}, \hat{\boldsymbol{\theta}}) - \mathbf{C}_{ij}\hat{\boldsymbol{\delta}}]^2, \tag{32}$$

where

$$\hat{\tau}^2|\mathbf{C} \sim (N - m - q)^{-1}\tau^2\chi^2_{N-m-q}. \tag{33}$$

The unbiasedness results from a property of the χ^2 distribution: $E[\hat{\tau}^2|\mathbf{C}] = E[\hat{\mathbf{e}}^{\mathrm{T}}\hat{\mathbf{e}}]/(N - m - q) = (N - m - q)\tau^2/(N - m - q) = \tau^2$.

4.2 Asymptotic Variance Ratio and Loss Factor

Just like Venkatraman and Wilson (1986) and Porta Nova and Wilson (1989), and in contrast to Nozari *et al.* (1984) and Rubinstein and Marcus (1985), we consider that the adequate generalization of the performance measures (variance ratio and loss factor), introduced by Lavenberg *et al.* (1982), is based on the unconditioned dispersion matrix of the controlled coefficients in metamodel (2). As a consequence, we will now obtain the unconditioned variance of $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$.

As Eq. (31) is verified, we have

$$\mathcal{E}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] = \boldsymbol{\theta} \quad \text{and} \quad \mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] = \tau^2 \mathbf{B}\mathbf{B}^{\mathrm{T}}.$$

Therefore,

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})] = \mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] + \mathcal{D}[\mathcal{E}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] = \mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] \tag{34}$$

and

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] \approx \tau^2 (\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1}[\mathrm{I}_m + \tilde{\mathbf{F}}^{\mathrm{T}}\mathbf{C}(\mathbf{C}^{\mathrm{T}}\tilde{\mathbf{P}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\tilde{\mathbf{F}}(\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1}].$$

Let $\mathbf{U} = (\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^{\mathrm{T}}\mathbf{C}$ and $\mathbf{V} = \mathbf{C}^{\mathrm{T}}\tilde{\mathbf{P}}\mathbf{C}$. Then, we can write

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] \approx \tau^2 (\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1} + \tau^2 \mathbf{U}\mathbf{V}^{-1}\mathbf{U}^{\mathrm{T}}. \tag{35}$$

Since $\mathbf{C} \sim \mathrm{N}_{N,q}(\mathbf{0}, \mathbf{I}_N, \Sigma_C)$ and $\tilde{\mathbf{P}}$ is an orthogonal projection of $\mathbb{R}^N$ into a space of dimension $N - m(\mathcal{R}(\mathbf{F})^{\perp})$, we apply Theorem 17.7 of Arnold (1981) to obtain $\mathbf{V} = \mathbf{C}^{\mathrm{T}}\tilde{\mathbf{P}}\mathbf{C} \sim W_q(N - m, \Sigma_C)$.

Using the fact that $(\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^{\mathrm{T}}(\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{P}} = 0$ and, as $\tilde{\mathbf{P}}$ is positive definite, we apply Theorem 17.7-b.2) of Arnold (1981) and we can conclude that $\mathbf{U}$ and $\mathbf{V}$ are independent. Taking the expected value in Eq. (35), we obtain

$$\mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] \approx \tau^2 (\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1} + \tau^2 \mathcal{E}[\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^{\mathrm{T}}]$$

Using the results in Appendix of Nozari *et al.* (1984), we have

$$\mathcal{E}[\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^{\mathrm{T}}] = \frac{q}{N - m - q - 1}(\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1}.$$

Combining the two previous results, we obtain

$$\mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] \approx \tau^2 \frac{N - m - 1}{N - m - q - 1}(\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1}.$$

But $\tilde{\mathbf{F}}^{\mathrm{T}} = [\mathbf{F}^{\mathrm{T}} \cdots \mathbf{F}^{\mathrm{T}}]$, therefore $(\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1} = \mathbf{F}^{\mathrm{T}}\mathbf{F}/r$ and, as a result,

$$\mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] \approx \tau^2 \frac{N - m - 1}{r(N - m - q - 1)}(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1},$$

that is, from Eq. (34), we have the following asymptotic result

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})] \approx \tau^2 (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1} \frac{N - m - 1}{r(N - m - q - 1)}.$$

This approximation, in conjunction with Eqs. (8) and (16), allows us to obtain the following approximated generalized variance ratio:

$$\eta(\hat{\boldsymbol{\delta}}) = \frac{|\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})]|}{|\mathcal{D}[\hat{\boldsymbol{\theta}}]|} \approx \frac{N - m - 1}{N - m - q - 1}\frac{\tau^2}{\sigma^2} = \frac{N - m - 1}{N - m - q - 1}(1 - \rho_{YC}^2). \tag{36}$$

Comparing this with the minimum variance ratio (18), we observe a degradation of the maximum variance reduction, namely the loss factor:

$$\mathrm{LF}(\hat{\boldsymbol{\delta}}) = \frac{N-m-1}{N-m-q-1}. \tag{37}$$

4.3 Asymptotic Confidence Regions for the Metamodel Coefficients

The objective of this section is to determine a confidence rectangle consisting of m confidence intervals for $\boldsymbol{\theta}_j$, $j = 1, \ldots, m$. In fact, it is simpler to represent graphically and to explain the meaning of a confidence rectangle of this type, when compared with the more common approximated confidence ellipsoid.

As Eq. (31) is verified, then using Theorem 3.10 of Arnold (1981), we can ensure that conditioning on $\mathbf{C}$:

$$\frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*)^{\mathrm{T}}(\mathbf{B}\mathbf{B}^{\mathrm{T}})^{-1}(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*)}{\tau^2} | \mathbf{C} \sim \chi^2_m \tag{38}$$

On the other hand, Eq. (33) can be rewritten as

$$(N-m-q)\frac{\hat{\tau}^2}{\tau^2}|\mathbf{C} \sim \chi^2_{N-m-q}. \tag{39}$$

Combining Eqs. (38) and (39), we obtain

$$\frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*)^{\mathrm{T}}(\mathbf{B}\mathbf{B}^{\mathrm{T}})^{-1}(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*)}{m\hat{\tau}^2} | \mathbf{C} \sim F_{m,N} - m - q.$$

Given $\mathbf{C}$, an asymptotic confidence region for $\boldsymbol{\theta}$, with conditional coverage probability of atleast $1-\alpha$, is given by

$$\left\{ \boldsymbol{\theta}^* \colon \frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*)^{\mathrm{T}}(\mathbf{B}\mathbf{B}^{\mathrm{T}})^{-1}(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*)}{m\hat{\tau}^2} \leq F_{m,N-m-q;1-\alpha} \right. . \tag{40}$$

Although this confidence region has conditional coverage of atleast $1-\alpha$, it has also unconditional coverage of atleast $1-\alpha$.

Let $\hat{\boldsymbol{\theta}}_i(\hat{\boldsymbol{\delta}})$ be the ith component of the vector $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$ and let $\hat{\tau}^2[\mathbf{B}\mathbf{B}^{\mathrm{T}}]_{ii}$ be the corresponding variance estimator (the ith diagonal element of $\hat{\tau}^2\mathbf{B}\mathbf{B}^{\mathrm{T}}$). As $\hat{\boldsymbol{\theta}}_i(\hat{\boldsymbol{\delta}})$ is conditionally independent of $\hat{\tau}^2[\mathbf{B}\mathbf{B}^{\mathrm{T}}]_{ii}$ given $\mathbf{C}$, the results (31) and (33) imply that

$$\frac{\hat{\boldsymbol{\theta}}_i(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*_i}{\hat{\tau}\sqrt{[\mathbf{B}\mathbf{B}^{\mathrm{T}}]_{ii}}}|\mathbf{C} \sim t_{N-m-q},$$

where t_{N-m-q} represents the Student t-distribution with $N-m-q$ degrees of freedom. As a consequence, using the Bonferroni method, a confidence rectangle for $\boldsymbol{\theta}$ with conditional coverage probability of atleast $1-\alpha$ has the form

$$\hat{\boldsymbol{\theta}}_k(\hat{\boldsymbol{\delta}}) \pm t_{N-m-q;1-\alpha/(2p)}\hat{\tau}[\mathbf{B}\mathbf{B}^{\mathrm{T}}]^{1/2}_{kk}, \quad k = 1, \ldots, p, \tag{41}$$

where $1 \leq p \leq m$. This confidence region has conditional coverage probability of atleast $1-\alpha$, therefore it also has unconditional coverage probability of atleast $1-\alpha$.

4.4 Hypothesis Testing on the Metamodel Coefficients

Suppose that we want to test the hypothesis H_0: $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ versus H_1: $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$. As Eq. (31) is verified and using Theorem 3.10 of Arnold (1981), we observe that conditioning on $\mathbf{C}$:

$$\frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0)^{\mathrm{T}}(\mathbf{BB}^{\mathrm{T}})^{-1}(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0)}{\tau^2} \sim \chi^2_m, \tag{42}$$

if H_0 is true. Combining Eqs. (42) and (39) and following an identical procedure to the one considered in Section 4.3, we reject H_0, with confidence level $100(1-\alpha)\%$, if

$$\frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0)^{\mathrm{T}}(\mathbf{BB}^{\mathrm{T}})^{-1}(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0)}{m\hat{\tau}^2} > F_{m,n-m-q;1-\alpha}. \tag{43}$$

5 NUMERICAL RESULTS FOR QUEUING APPLICATION

We illustrate our methodology using a simple $M/M/1$ queuing system. We assume that customers arrive according to a Poisson process with a constant expected arrival rate, λ, and that service times follow an exponential distribution with a constant expected service time, $1/\mu \equiv 1$. The performance measure of interest is the average waiting time in the queue. The objective is to express this response as a function of the queue utilization factor, $\rho = \lambda/\mu$ (a single decision variable). The available concomitant output variables, that can be used as controls, are the average service time and the average inter-arrival time. In this experiment, after some minor adjustments, 12 ($n = 12$) different values for ρ were considered: $\{\rho_i : i = 1, 12\} = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.55, 0.6, 0.7, 0.75, 0.85, 0.9, 0.95\}$. There were $r = 20$ replications for each of the $n = 12$ design points. Different replications used the same value for the independent variable ρ_i, but different pseudo-random number seeds. Each of these 20 replications started with an empty and idle system (no customers waiting). At each design point, we ran Welch's procedure (Welch, 1983) to determine the length of each simulation run and the initial data deletion. For example, at the design point $\rho = 0.95$, we ignored 3500 observations from the beginning of each run and we used only the remaining 36,500 observations (approximately 85% of the number of observations in the run), while considering a Welch window of 10,000; see Table I.

We then compared the dispersion diagram based on the collected data (Fig. 1) with commonly available theoretical curves.

To relate the average waiting time in the queue with the utilization factor, we chose the hyperbolic metamodel $Y_{ij} = \theta_1 X_i/(1 + \theta_2 X_i) + \varepsilon_{ij}$ (where Y_{ij} is the average queue waiting time during the jth run at experimental point i), with $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma_i^2)(i = 1, \ldots, 12; j =$

TABLE I Initial data deletion.

	Observations		
ρ_i	*Deleted*	*In run*	*Welch's window*
0.10, 0.20, 0.30	500	3,500	1,000
0.40	1,000	7,000	1,000
0.50	1,500	10,000	1,000
0.55, 0.60, 0.70, 0.75	1,500	10,000	4,000
0.85	2,000	20,000	8,000
0.90	2,500	20,000	10,000
0.95	3,500	40,000	10,000

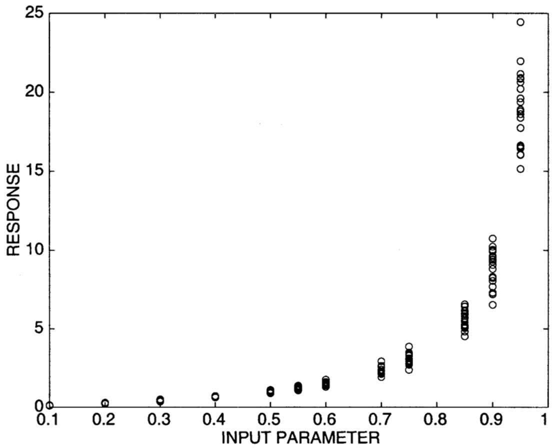

FIGURE 1 Dispersion diagram of the $M/M/1$ queue.

$1, \ldots, 20$), and σ_i^2 varies with i. For stabilizing the variance, we took logarithms on both sides of the above expression, obtaining

$$\log Y_{ij} = \log \frac{\theta_1 X_i}{1 + \theta_2 X_i} + v_{ij} \quad i = 1, \ldots, 12, \quad j = 1, \ldots, 20, \tag{44}$$

with $v_{ij} = \log(l - \varepsilon_{ij}/\mathrm{E}[Y_{ij}])$. $\mathrm{E}[v_{ij}] \approx 0$, because ε_{ij} is small when compared with $E[Y_{ij}]$ and $\mathrm{Var}[v_{ij}]$ is approximately constant for all $i = 1, \ldots, n$ and $j = 1, \ldots, r$. We measure the variance heterogeneity using the quantity

$$\text{het} = \frac{\max\limits_{i=1,\ldots,12} \widehat{\mathrm{Var}}[v_{ij}]}{\min\limits_{i=1,\ldots,12} \widehat{\mathrm{Var}}[v_{ij}]},$$

with

$$\widehat{\mathrm{Var}}[v_{ij}] = \left[\frac{1}{r-1} \sum_{j=1}^{r} \left(\log Y_{ij} - \frac{1}{r} \sum_{j=1}^{r} \log Y_{ij} \right)^2 \right]^{1/2};$$

see Kleijnen (1992). We obtained a het value approximately constant and equal to 1. For improving the efficiency of metamodel estimation, we chose the following control variates

$$C_{kij} = \frac{t_{kij} - \mu_{ki}}{\zeta_{ki}}, \quad k = 1, 2, \quad i = 1, \ldots, n, \quad j = 1, \ldots, r,$$

where t_{1ij} is the average inter-arrival time and t_{2ij} is the average service time. Both were sampled from exponential distributions with known means and variances: $\mathrm{E}[t_{1ij}] = \mu_{1i} = 1/\rho_i$, $\mathrm{Var}[t_{1ij}] = \zeta_{1i}^2 = 1/\rho_i^2$, $\mathrm{E}[t_{2ij}] = \mu_{2i} = 1/\mu^2 = 1$ and $\mathrm{Var}[t_{2ij}] = \zeta_{2i}^2 = 1/\mu^2 = 1$. The hypothesized controlled problem is

$$\log Y_{ij} = \log \frac{\theta_1 X_i}{1 + \theta_2 X_i} + \delta_1 C_{1ij} + \delta_2 C_{2ij} + v_{ij}, \quad i = 1, \ldots, 12 \;\; j = 1, \ldots, 20.$$

5.1 Estimation and Validation

We obtained the least squares estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$ using the Levenberg–Marquart method, implemented in MATLAB, with a termination tolerance of 10^{-6} and maximum number of function evaluations equal to 600; see Table II.

After estimation, the validation of the controlled metamodel was carried out. As $F_{n-m-q,N-n;\alpha} = F_{8,228;0.05} \approx 1.95$, with $N = 240$, based on the F-test, we do not reject the metamodel with control variables; see Table III. As a consequence, on the basis of this validation procedure, we do not reject the controlled metamodel.

5.2 Confidence Regions and Hypothesis Testing

As $F_{m,N-m-q;\alpha} = F_{2,236;0.05} = 3.034$ and $\hat{\tau}_2 = 8.184 \times 10^{-3}$, the 95% approximated confidence ellipsoid for θ, centered at $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$, is given by Eq. (40):

$$\left\{\boldsymbol{\theta}* : (\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*)^{\mathrm{T}}(\mathbf{BB}^{\mathrm{T}})^{-1}(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*) \leq 4.966 \times 10^{-2}\right\},$$

where

$$\mathbf{BB}^{\mathrm{T}} = \begin{bmatrix} 3.2891 \times 10^{-2} & 1.8495 \times 10^{-3} \\ 1.8495 \times 10^{-3} & 2.2099 \times 10^{-4} \end{bmatrix}.$$

The corresponding approximated confidence rectangle for $\boldsymbol{\theta}$, with coverage probability of atleast $1 - \alpha = 0.95$, from Eq. (41), is given in Table IV. In the construction of this confidence rectangle, we used $t_{N-m-q;1-\alpha/(2m)} = t_{236;0.9875} \approx 2.256$.

We tested the hypothesis H_0: $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0 = (1.0, -1.0)^{\mathrm{T}}$ versus H_1: $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0 = (1.0, -1.0)^{\mathrm{T}}$, with a confidence level of 0.95%, using $F_{m,N-m-q;1-\alpha} = F_{2,236;0.95} = 3.034$. We obtained $(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0)^{\mathrm{T}}(m\hat{\tau}^2\mathbf{BB}^{\mathrm{T}})^{-1}(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0) - 0.3848$ (43). Therefore, we do not reject the null hypothesis H_0.

TABLE II Estimated metamodel coefficients.

Metamodel coefficients	*Direct estimator* $\hat{\boldsymbol{\theta}}$	*Controlled estimator* $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$
θ_1	0.9982	1.0001
θ_2	−0.9992	−0.9991

TABLE III Testing for lack-of-fit.

Source	*d.f.*	*Sum of squares*	*Mean of squares*	*F*
Lack-of-fit	8	0.008036	0.001004	0.119
Pure error	228	1.923	0.008435	

TABLE IV Approximated Bonferroni 95% confidence intervals.

Metamodel coefficients	*Controlled estimator*
θ_1	$1.0001188 \pm 3.35 \times 10^{-3}$
θ_2	$-0.9991349 \pm 2.74 \times 10^{-4}$

5.3 Experimental Variance Ratio and Loss Factor

In order to estimate the variance ratios and loss factors, we adapted the procedure described by Porta Nova and Wilson (1989) to our situation. Thus, we performed a meta-experiment with $K = 30$ independent replications of the basic experiment, consisting of 12 design points ($\rho = 0.1, 0.2$, etc.) and 20 independent replications for each design point. For each kth replication of the basic experiment ($k = 1, \ldots, K$), we calculate the direct estimator $\hat{\boldsymbol{\theta}}_k$ and the control-variate estimator $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})^k$ for the metamodel coefficient vector $\boldsymbol{\theta}$. From the random sample $\{\hat{\boldsymbol{\theta}}^k : 1 \le k \le K\}$, we compute a unbiased estimator for $\mathcal{D}[\hat{\boldsymbol{\theta}}]$ as follow:

$$\hat{\mathcal{D}}[\hat{\boldsymbol{\theta}}] = \frac{1}{K}\sum_{k=1}^{K}(\hat{\boldsymbol{\theta}}^k - \bar{\hat{\boldsymbol{\theta}}})(\hat{\boldsymbol{\theta}}^k - \bar{\hat{\boldsymbol{\theta}}})^{\mathrm{T}}, \quad \text{where} \quad \bar{\hat{\boldsymbol{\theta}}} = \frac{1}{K}\sum_{k=1}^{K}\hat{\boldsymbol{\theta}}^k;$$

and similarly from the random sample $\{\hat{\boldsymbol{\theta}}^k : 1 \le k \le K\}$, we compute an unbiased estimator $\hat{\mathcal{D}}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})]$ of $\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})]$. On the basis of this estimators, we compute the following estimator of the variance ratio (36), which we call the observed variance ratio:

$$\hat{\eta}(\hat{\boldsymbol{\delta}}) = \frac{|\hat{\mathcal{D}}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})]|}{|\hat{\mathcal{D}}[\hat{\boldsymbol{\theta}}]|}. \tag{45}$$

To obtain the minimum variance ratio (18), at each design point we compute unbiased estimators for the variance of Y, σ^2, covariance vector between Y and $\mathbf{C}$, $\boldsymbol{\sigma}_{YC}$, and despersion matrix of $\mathbf{C}$, Σ_C. For example, if Y_i^k denotes the mean response observed at the ith design point on the kth independent replication of the basic experiment, then the variance of Y is estimated by

$$(\hat{\sigma}^2)^i = \frac{1}{K}\sum_{k=1}^{K}(Y_i^k - \bar{Y})^2,$$

where $\bar{Y} = \sum_{j=1}^{K} Y_i^j / K$. As a result, a pooled estimator of σ^2 based on all n experimental points is given by: $\bar{\hat{\sigma}}^2 = \sum_{i=1}^{n}(\hat{\sigma}^2)^i / n$. The other estimators are obtained in a similar way: $\bar{\hat{\boldsymbol{\sigma}}}_{YC} = \sum_{i=1}^{n} \hat{\boldsymbol{\sigma}}_{YC}^i / n$, where

$$\hat{\Sigma}_C^i = \frac{1}{K}\sum_{k=1}^{K}(\bar{\mathbf{C}}_{i.}^k - \bar{\bar{\mathbf{C}}}_{i.})^{\mathbf{T}}(\bar{\mathbf{C}}_i^k - \bar{\bar{\mathbf{C}}}_{i.}),$$

with $\bar{\bar{\mathrm{C}}}_{i.} = \sum_{k=1}^{K} \bar{\mathrm{C}}_{i.}^k / K$; and $\bar{\hat{\boldsymbol{\sigma}}}_{YC} = \sum_{i=1}^{n} \hat{\boldsymbol{\sigma}}_{YC}^i / n$, where

$$\hat{\boldsymbol{\sigma}}_{YC}^i = \frac{1}{K}\sum_{k=1}^{K}(\bar{Y}_{i.}^k - \bar{\bar{Y}}_{i.})(\bar{\mathbf{C}}_{i.}^k - \bar{\bar{\mathbf{C}}}_{i.}).$$

Using this numerical values, we calculate the following estimator of Eq. (18) that we call the estimated minimum variance ratio:

$$\hat{\eta}(\boldsymbol{\delta}) = 1 - \frac{\bar{\hat{\boldsymbol{\sigma}}}_{YC}\bar{\hat{\boldsymbol{\Sigma}}}_C^{-1}\bar{\hat{\boldsymbol{\sigma}}}_{YC}^{\mathrm{T}}}{\bar{\hat{\sigma}}^2} \tag{46}$$

Multiplying Eq. (46) by the loss factor (37), we obtain the predicted variance ratio

$$\ddot{\eta} = \hat{\eta}(\boldsymbol{\delta})\mathrm{LF}(\hat{\boldsymbol{\delta}}).$$

TABLE V Estimation of the variance ratios and loss factors.

Estimated variance ratios			*Loss factor*		
	Actual				
Minimum $\eta(\hat{\boldsymbol{\delta}})$	*Predicted* $\ddot{\eta}(\hat{\boldsymbol{\delta}})$	*Observed* $\hat{\eta}(\hat{\boldsymbol{\delta}})$	*True LF* $(\hat{\boldsymbol{\delta}})$	*Estimated* $\widehat{LF}(\hat{\delta})$	*%Error*
0.613	0.618	0.993	1.008	1.620	60.6

The observed loss factor is given by

$$\widehat{\mathrm{LF}}(\hat{\boldsymbol{\delta}}) = \frac{\hat{\eta}(\hat{\boldsymbol{\delta}})}{\ddot{\eta}(\boldsymbol{\delta})}.$$

The predicted variance ratio can be compared with the observed variance ratio and the observed loss factor can be compared with the theoretical loss factor. The numerical results are reported in Table V.

In the example described here, the maximum percentage reduction in generalized variance that can be achieved using control variates is approximately $100[1 - \ddot{\eta}(\boldsymbol{\delta})]\% = 38.7\%$. Although we do not know δ, it must be estimated. Nevertheless the resulting estimator $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$ has a smaller variance, when compared with the estimator $\hat{\boldsymbol{\theta}}$ without control variates. The numerical results presented here are in agreement with the theoretical results developed in this article. The error obtained for the loss factor is similar to the value obtained by Porta Nova and Wilson (1989), 57.7%. These authors analyzed a queuing network simulation in the context of a linear metamodel estimation.

6 CONCLUSIONS

The main objective of this article is to establish some important results on the use of multiple control variates for improving the precision of nonlinear regression metamodel estimation. This technique can be useful in many situations where it is possible to identify effective concomitant control variables. Because nonlinear regression models are better than linear models, in capturing the shape of arbitrary mathematical functions, we emphasize the importance of using valid nonlinear metamodels in simulation studies. In addition, nonlinear metamodels allow us to characterize the precision of the fit by the use of confidence intervals and they are more robust than linear models when extrapolating from the actual experimental domain.

For experimental designs with a sufficiently large number of experimental points and under certain regularity conditions, the efficiency of metamodel estimation can be improved using the method of control variables. However, whether a regression metamodel is used in the simulation context, it must be validated. The validation can be made using, for example, the lack of fit F-test present in the statistical literature on nonlinear regression models. In our experimental study, we observed a marked sensitivity of the variance ratio $\eta(\hat{\boldsymbol{\delta}})$ and the loss factor LF $(\hat{\boldsymbol{\delta}})$ with respect to the validity of the assumed controlled problem (2). As a consequence, it is imperative to resort to statistical validation techniques, like the above mentioned F-test, to verify the capability of the controlled metamodel in representing the simulation model.

References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, New York, USA. **Q1**

Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis*. John Wiley and Sons, New York, USA.

Barton, R. R. (1992). Metamodels for simulation of input–output relations. In: *Proceedings of the Winter Simulation Conference*, pp. 289–299.

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

Glynn, P. W. and Whitt, W. (1989). Indirect estimation via $L = \lambda W$. *Operations Research*, **37**(1), 82–103.

Loh, W. W. (1994). On the method of control variates. Unpublished *PhD dissertation*, Department of Operations Research, Stanford University, Stanford CA (Downloadable from website http://www-or.stanford.edu/simlib/theses.html).

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.

Lavenberg, S. S. and Welch, P. D. (1981). A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations. *Management Science*, **27**(3), 322–335.

Lavenberg, S. S., Moeller, T. L. and Welch, P. D. (1982). Statistical results on control variables with application to queuing network simulation. *Operations Research*, **30**, 182–202.

Nozari, A. (1982). On simulation experiments for multipopulation problems. Unpublished *PhD dissertation*, Pennsylvania State University.

Nozari, A., Arnold, S. F. and Pegden, C. D. (1984). Control variates for multipopulation simulation experiments. *IIE Transactions*, **16**, 159–169.

Porta Nova, A. M. (1985). A generalized approach to variance reduction in discrete-event simulation using control variables. Unpublished *PhD dissertation*, University of Texas.

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

Rubinstein, R. Y. and Marcus, R. (1985). Efficiency of multivariate control variates in Monte Carlo simulation. *Operations Research*, **33**(3), 661–677.

Santos, M. I. and Porta Nova, A. M. (1999). The main issues in nonlinear simulation metamodel estimation. In: *Proceedings of the Winter Simulation Conference*, pp. 502–509.

Searle, S. R. (1971). *Linear Models*. John Wiley and Sons, New York, USA.

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

Shih, N. H. and Song, W. T. (1995). Estimating simulation metamodels using correlated control variates. *Communications in Statistics B –Simulation and Computation*, **24**(1), 79–90.

Venkatraman, S. and Wilson, J. (1986). The efficiency of control variates in multire-sponse simulation. *Operations Research Letters*, **5**(1), 37–42.

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

Journal …**J. Statist. Comput. Simul.,** Article ID …**GSCS 041060**

TO: CORRESPONDING AUTHOR

AUTHOR QUERIES - TO BE ANSWERED BY THE AUTHOR

The following queries have arisen during the typesetting of your manuscript. Please answer the queries.

Q1	Anderson (1984) is not cited in the text but present in the reference list please check.	

Production Editorial Department, Taylor & Francis Ltd.
4 Park Square, Milton Park, Abingdon OX14 4RN

Telephone: +44 (0) 1235 828600
Facsimile: +44 (0) 1235 829000