Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

ANALYZING TRANSFORMATION-BASED SIMULATION METAMODELS

Maria de los A. Irizarry

Department of Industrial Engineering
University of Puerto Rico
Mayaguez, PR 00681, U.S.A.

Michael E. Kuhl

Department of Industrial and Manufacturing
Systems Engineering
Louisiana State University
Baton Rouge, LA 70803-6409, U.S.A.

Emily K. Lada
Sriram Subramanian
James R. Wilson

Graduate Program in Operations Research
Integrated Manufacturing Systems Engineering Institute
Department of Industrial Engineering
North Carolina State University
Raleigh, NC 27695, U.S.A.

ABSTRACT

We present a technique for analyzing a simulation metamodel that has been constructed using a variance-stabilizing transformation. To compute a valid confidence interval for the expected value of the original simulation response at a selected factor-level combination (design point), we first compute the corresponding confidence interval for the transformed response at that factor-level combination and then untransform the endpoints of the resulting confidence interval. Taking the midpoint of the untransformed confidence interval as our point estimator of the expected simulation response at the selected factor-level combination and approximating the variance of this point estimator via the delta method, we formulate an approximate two-sample Student *t*-test for validating our metamodel-based estimator versus the results of making independent runs of the simulation at the selected factor-level combination. We illustrate this technique in a case study involving the design of a manufacturing cell, and we compare our results with those of a more conventional approach to analyzing transformed-based simulation metamodels. A Monte Carlo performance evaluation shows that significantly better confidence-interval coverage is maintained with the proposed procedure over a wide range of values for the residual variance of the transformed metamodel.

1 A CASE STUDY IN MANUFACTURING CELL DESIGN

In Irizarry et al. (2000a, b) we present an approach to manufacturing-cell design and analysis based on the use of a generic manufacturing-cell simulation model together with effective techniques for response surface estimation and optimization. This methodology consists of four major steps: (i) selection of cell design and operation issues; (ii) development of a comprehensive cell performance measure; (iii) identification of critical design and operation factors for the cell; and (iv) optimization of cell performance as a function of the cell's critical design and operation factors.

The methodology was applied to a cell for the assembly of printed circuit boards. A detailed description of the cell is presented in Irizarry et al. (2000a). In this study we concentrated attention on following the cell operational issues: setup policy (SU), unit load size (UL), lot size (LT), machine minor stoppages (ST), quality policy (QL), and maintenance policy (MA). These became the input factors for the factor-screening experimentation. The performance measure chosen for the evaluation of alternative cell designs was a comprehensive annualized cost function that incorporates ten major cost components as detailed in Irizarry (1996).

1.1 Identification of Critical Design and Operation Factors

Since it is likely that only a few of the selected cell design and operation factors will have a significant impact on cell performance, we performed a factor-screening experiment as the first step of our procedure. In the printed circuit board case study, the factor-screening experiment was a 2^{6-1} fractional factorial design in two blocks of 16 design points each. This is a resolution V design—that is, no main effect or two-factor interaction is confounded with any other main effect or two-factor interaction (Montgomery 1991). Table 1 summarizes the input factors and factor levels selected for the factor-screening experiment.

Table 1: Description of Factor Levels for the Factor-Screening Experiment

Factor	Level 1 (Coded Value –1)	Level 2 (Coded Value +1)
SU	Long (unreduced) setup times	Quick changeovers (75% reduction)
UL	Large (≈ 50% of lot size)	Small (≈ 10% of lot size)
LT	Large lots (groups of customer orders)	Small lots (individual customer orders)
ST	Small rolls of Components	Bigger rolls of Components
QL	Traditional Inspections	Quality at the source
MA	Breakdown Maintenance	Autonomous and Preventive maintenance

The objective of the screening experiment was to identify the important main effects and two-factor interactions. Cell performance at each design point was evaluated using the generic cell simulator that we introduced in Irizarry et al. (2000a). We performed a statistical analysis of the results using the general linear model procedure GLM of SAS (1990).

In the regression analysis for the factor-screening experiment of the case study, the significance probabilities (*P*-values) for the corresponding estimated regression coefficients were used to identify significant main effects and two-factor interactions. For a significance level of 5%, the significant main effects were setup time (SU), unit load size (UL), lot size (LT), minor stoppages (ST), and maintenance (MA). The significant two-factor interactions were: lot size with unit load size (LT×UL); lot size with setup time (LT×SU); lot size with minor stoppage (LT×ST); and setup time with minor stoppages (SU×ST).

In the factor-screening analysis for the case study, the estimated regression coefficient for the maintenance policy (MA) was highly significant; however, this factor was not found to interact with any other factor. A detailed analysis of the cell's annual cost showed that investments and interruptions caused by the high level of maintenance had a negative impact on cell performance. We concluded that maintenance should be set at its low level (MA = –1) in all subsequent experimentation.

1.2 Optimization of Cell Performance

After the factor-screening analysis was complete, a response surface experimental design was used to construct an adequate approximation to the target response surface and to estimate the optimal settings for the significant input factors as well as the expected optimal response. In the case study, the corresponding simulation experiment included two qualitative factors (lot size, LT, and machine stoppages, ST) each at two levels; and for every combination of these qualitative input factors, we performed a 3^2 full factorial simulation experiment to estimate the expected annual cell operating cost as a function of the selected quantitative factors (setup time, SU, and unit load size, UL). The overall simulation *macroexperiment* consisted of four separate 3^2 factorial experiments. Three independent simulation runs were performed at each design point of each 3^2 factorial experiment, yielding a total of 27 runs per experiment and a total of 108 independent simulation runs in the entire macroexperiment. Table 2 summarizes the input-factor levels used in all these scenarios.

The statistical analysis of the results of the simulation macroexperiment was performed using the SAS response surface regression procedure RSREG with a 5% significance level (SAS 1990). We estimated a response surface for each of the four experiments contained in the macroexperiment using a metamodel that is a quadratic function of the two regressors (setup time, SU, and unit load size, UL). Canonical and ridge analyses of the estimated metamodel yielded the optimal design point $\hat{X}^* = (\hat{\mathrm{SU}}^*, \hat{\mathrm{UL}}^*)$ (that is, the stationary point of the fitted response surface); see pages 332–381 of Box and Draper (1987). Residuals were analyzed using the SAS procedure UNIVARIATE, which includes moments, quantiles, stem-and-leaf plots, box plots, and normal probability plots (SAS 1990). Table 3 summarizes the results from this analysis.

We found experiments 1 and 3 to exhibit similar behavior in the shape of their estimated response surfaces, with both metamodels having the same significant regressors as well as similar values for the predicted mean response at the corresponding stationary point. The *P*-values in the lack-of-fit tests for these metamodels are 0.948 and 0.964, respectively. Recall that each *P*-value represents the smallest level of significance at which the observed value of the goodness-of-fit test statistic would cause us to reject the null hypothesis that the corresponding metamodel has the correct functional form. Since the observed *P*-values are much larger than the significance

Table 2: Factor Levels for the Response Surface Macroexperiment

Factor	Level 1 (–1)	Level 2 (0)	Level 3 (+1)
SU	Long (unreduced) setup times	Quick changeovers (37.5% reduction)	Quick changeovers (75% reduction)
UL	Large (≈ 50% of lot size)	Medium (≈30% of lot size)	Small (≈ 10% of lot size)
LT	Large lots (groups of customer orders)		Small lots (individual customer orders)
ST	Small rolls of components		Bigger rolls of components

Table 3: Results from the Statistical Analysis of the Simulation Macroexperiment

Experiment	Significant Regressors	Estimated Optimum Response $\hat{Y}(\hat{X}^*)$	Residual Skewness and Excess Kurtosis	Std. Error of $\hat{Y}(\hat{X}^*)$
1	UL, SU, UL^2, SU^2	\$1,954,905	–0.0071, –1.0607	\$5,136
2	SU, SU^2	–\$1,425,522	0.6259, 1.7765	\$82,588
3	UL, SU, UL^2, SU^2	\$1,937,638	0.0677, –0.5288	\$4,322
4	SU, SU^2	–\$161,378	–0.6946, 5.3033	\$2,290,593

levels commonly used for hypothesis testing, we found the metamodels for experiments 1 and 3 to exhibit no significant lack of fit.

The original metamodels developed for experiments 2 and 4 resulted in negative predictions for the expected annual cost at the stationary point and in other regions of the input-factor space. Moreover, in experiments 2 and 4 we found that the estimated residuals exhibited significant departures from normality—especially in terms of the sample skewness and excess kurtosis of the residuals. Finally the standard error of $\hat{Y}(\hat{X}^*)$ is much larger for experiments 2 and 4 than it is for experiments 1 and 3; and we found that introducing additional design points and then fitting a higher-order metamodel did *not* improve the behavior of the estimated residuals as measured by their sample variance, skewness, and excess kurtosis. We concluded that in order to obtain usable response-surface models in experiments 2 and 4, we must do the following: (a) identify and apply an appropriate variance stabilizing transformation of the original simulation-generated responses (Box and Draper 1987); and (b) augment the set of regressors (independent variables) in the corresponding metamodels to include relevant higher-order terms. These conclusions are consistent with our previous experience in building simulation metamodels for certain types of textile production systems as discussed, for example, in Powell (1992).

2 TRANSFORMATION-BASED METAMODELS

We examined several transformations of the response in our search for metamodels with a better fit to the results observed in experiments 2 and 4; see pages 280–293 of Box and Draper (1987). Moreover, we added design points to these experiments to allow estimation of higher-order metamodels. Based on examination of the fitted response surfaces for experiments 2 and 4 as described in Table 3, we augmented these experiments with selected design points that would allow us to estimate third-order metamodels. Statistical analyses of these more complex metamodels were performed using SAS's general linear model procedure GLM (1990). Table 4 summarizes the results of this follow-up analysis for experiments 2 and 4.

Let D denote the overall design matrix for a response surface experiment so that each row of D has the form $W(X)$, where the first element of $W(X)$ is 1 and the other elements are the corresponding regressors defined in Table 4; thus, for example, at each design point X in experiment 2, we have

$$W(X) = W(\text{SU, UL}) = [1,\ \text{SU},\ \text{UL},\ \text{SU}^2,\ \text{UL}^2,\ \text{SU}^3,\ \text{UL}^3].$$

If the transformation $Z(X) = f[Y(X)]$ achieves normal responses with mean $W(X)\beta$ and constant variance σ_Z^2, then the usual least squares estimator $\hat{\beta}$ of the metamodel coefficient vector β yields the prediction $\hat{Z}(X) = W(X)\hat{\beta}$ that is unbiased and has variance

$$\mathrm{Var}\left[\hat{Z}(X)\right] = \sigma_Z^2 W(X)\left(D^{\mathrm{T}} D\right)^{-1} W^{\mathrm{T}}(X).$$

Table 4: Results for Selected Transformations of the Response in Experiments 2 and 4

	Experiment 2	Experiment 4
Significant Regressors	SU, UL, SU^2, UL^2, SU^3, UL^3	SU, UL^2, SU^2, SU^3
Normalizing Transformation Z(X)=f[Y(X)]	Z(X) = ln[Y(X)]	Z(X) = 10000/Y(X)
Predicted Annual Cost $\hat{Y}(\hat{X}^*) = f^{-1}[\hat{Z}(\hat{X}^*)]$ at the stationary point $\hat{X}^*$	$1,647,899	$1,730,652
Standard Error of $\hat{Y}(\hat{X}^*)$	$26,649	$12,454
Residual Skewness and Excess Kurtosis	0.2690, 1.4656	–0.0404, 0.5661

Thus an estimator of the standard error of $\hat{Z}(X)$ is given by

$$\mathrm{SE}[\hat{Z}(X)] = \hat{\sigma}_Z \sqrt{W(X)(D^{\mathrm{T}}D)^{-1}W^{\mathrm{T}}(X)}\,,$$

where $\hat{\sigma}_Z$ is the standard deviation of the estimated residuals in the response surface (regression) analysis; moreover, a valid $100(1-\alpha)\%$ confidence interval for E[Z(X)] = $\mathrm{E}\{f[Y(X)]\}$ is

$$\hat{Z}(X) \pm H(X), \tag{1}$$

where the half-length of the confidence interval is

$$H(X) = t_{1-\alpha/2,\nu}\,\hat{\sigma}_Z \sqrt{W(X)\left(D^{\mathrm{T}}D\right)^{-1}W^{\mathrm{T}}(X)},$$

a standard result from normal regression theory. In terms of the metamodel for the original untransformed responses, we propose the following $100(1-\alpha)\%$ confidence interval for E[Y(X)] = $\mathrm{E}\{f^{-1}[Z(X)]\}$:

$$\left[\min\left\{\hat{Y}_1(X), \hat{Y}_2(X)\right\}, \max\left\{\hat{Y}_1(X), \hat{Y}_2(X)\right\}\right], \tag{2}$$

where the endpoints of (2) are:

$$\hat{Y}_1(X) = f^{-1}\left[\hat{Z}(X) - H(X)\right] \quad ,$$

$$\hat{Y}_2(X) = f^{-1}\left[\hat{Z}(X) + H(X)\right] \quad .$$

If (1) is a valid $100(1-\alpha)\%$ confidence interval for E[Z(X)], then (2) is an approximate $100(1-\alpha)\%$ confidence interval for E[Y(X)] since in general the operators E[·] and $f^{-1}(\cdot)$ are not commutative; but we believe that (2) will perform well in practice.

As a point estimator of E[Y(X)], we propose the midpoint of (2),

$$\hat{Y}(X) = \frac{1}{2}\left[\hat{Y}_1(X) + \hat{Y}_2(X)\right], \tag{3}$$

as an alternative to the point estimator $f^{-1}[\hat{Z}(X)]$ used in Irizarry et al. (2000b). We believe that (3) is a more robust point estimator that may be used more reliably, for example, in statistical tests for validating the fitted metamodel versus simulation-based estimates of E[Y(X)]. To use (3) in such tests, we estimate $\mathrm{Var}[\hat{Y}(X)]$ using the delta method (Stuart and Ord 1994). In terms of the auxiliary functions

$$g_1\left[\hat{Z}(X), H(X)\right] = \frac{1}{2}\left\{f'\left(f^{-1}\left[\hat{Z}(X) - H(X)\right]\right)\right\}^{-1} + \frac{1}{2}\left\{f'\left(f^{-1}\left[\hat{Z}(X) + H(X)\right]\right)\right\}^{-1}\,, \tag{4}$$

$$g_2\left[\hat{Z}(X), H(X)\right] = \frac{1}{2}\left\{-f'\left(f^{-1}\left[\hat{Z}(X) - H(X)\right]\right)\right\}^{-1} + \frac{1}{2}\left\{f'\left(f^{-1}\left[\hat{Z}(X) + H(X)\right]\right)\right\}^{-1}\,, \tag{5}$$

the estimated standard error of $\hat{Y}(X)$ is given by

$$\mathrm{SE}\left[\hat{Y}(X)\right] \cong \hat{\sigma}_Z \sqrt{W(X)\left(D^{\mathrm{T}}D\right)^{-1}W^{\mathrm{T}}(X)} \times \sqrt{g_1^2\left[\hat{Z}(X), H(X)\right] + g_2^2\left[\hat{Z}(X), H(X)\right]\frac{t_{1-\alpha/2,\nu}^2}{2\nu}}. \tag{6}$$

Next we proceed to apply equations (4)–(6) to the transformations used on experiments 2 and 4 in Irizarry et al. (2000b).

To validate the final metamodel for a simulation experiment, we assess the statistical and practical significance of the difference between the metamodel-predicted mean response and the simulated-generated mean response at the same stationary point. The results from n independent replications of the simulation model at the stationary point are combined with the results of evaluating (3) and (6) to build the following confidence interval for the difference between the metamodel- and simulation-based estimates of optimal cell performance (see equations (15.4.15)–(15.4.17) of Hald 1952):

$$\hat{Y}(X)-\overline{Y}(X) \pm t_{1-\alpha/2,\eta}\left\{\mathrm{SE}^2\left[\hat{Y}(X)\right]+S^2_{\overline{Y}(X)}\right\}^{1/2}, \qquad (7)$$

where: $S^2_{\overline{Y}(X)}$ is the estimated variance of the simulation mean response $\overline{Y}(X)$ based on n independent replications of the simulation at the design point X; and η, our approximation to the "effective" degrees of freedom for the complex variance estimator $\mathrm{SE}^2[\hat{Y}(X)]+S^2_{\overline{Y}(X)}$ in (7), is taken to be

$$\eta=\left\lfloor \frac{\left\{\mathrm{SE}^2\left[\hat{Y}(X)\right]+S^2_{\overline{Y}(X)}\right\}^2}{\dfrac{\mathrm{SE}^4\left[\hat{Y}(X)\right]}{\nu}+\dfrac{S^4_{\overline{Y}(X)}}{n-1}} \right\rfloor, \qquad (8)$$

with ν denoting the degrees of freedom for the residual mean square in the regression analysis used to estimate the associated response surface, and n denoting the number of independent simulation runs performed at X.

2.1 Follow-Up Analysis for Transformation-Based Metamodel in Experiment 2

In the two-dimensional region of interest for experiment 2, we obtained the best results using the normalizing transformation

$$f(y)=\ln(y) \quad \text{so that} \quad f'(y)=y^{-1}. \qquad (9)$$

Inserting (9) into (2)–(5), we obtain

$$\hat{Y}_1(X)=\exp\left[\hat{Z}(X)-H(X)\right],$$

$$\hat{Y}_2(X)=\exp\left[\hat{Z}(X)+H(X)\right],$$

$$g_1\left[\hat{Z}(X),H(X)\right]=\frac{1}{2}\left[\hat{Y}_1(X)+\hat{Y}_2(X)\right]=\hat{Y}(X), \qquad (10)$$

and

$$g_2\left[\hat{Z}(X),H(X)\right]=\frac{1}{2}\left[\hat{Y}_2(X)-\hat{Y}_1(X)\right]. \qquad (11)$$

The standard error of $\hat{Y}(X)$ is estimated by inserting results from equations (10)–(11) in equation (6), yielding

$$\mathrm{SE}\left[\hat{Y}(X)\right]=\mathrm{SE}\left[\hat{Z}(X)\right] \times\sqrt{\hat{Y}^2(X)+[\hat{Y}_2(X)-\hat{Y}_1(X)]^2\frac{t^2_{1-\alpha/2,\nu}}{8\nu}}\ . \qquad (12)$$

Table 5 contains a 95% confidence interval for $\mathrm{E}[\hat{Y}(\hat{X}^*)]$, the expected value of the original simulation response at the stationary point $\hat{X}^*$, together with $\mathrm{SE}[\hat{Y}(\hat{X}^*)]$, the estimated standard error for the predicted response at $\hat{X}^*$. These results are very similar to those reported in Table 6 of Irizarry et al. (2000a). However, we believe that in other situations in which the original responses display more pronounced departures from normality, more reliable statistical tests will result from using the estimators (3) and (6) proposed in this paper. As summarized in Section 3 below, the results of the Monte Carlo study provide strong support for this conclusion.

Table 5: Confidence Interval and Standard Error for the Predicted Response at the Stationary Point in Experiment 2

Statistics for Transformed Responses	Computed Value
$\hat{Z}(\hat{X}^*)$	14.315
$\mathrm{SE}[\hat{Z}(\hat{X}^*)]$	0.017992
Effective degrees of freedom ν	31
$t_{0.975,31}$	2.042
95% Confidence Interval	[14.28, 14.35]
Statistics for Untransformed Responses	**Computed Value**
$\hat{Y}_1(\hat{X}^*)$	$1,588,000
$\hat{Y}_2(\hat{X}^*)$	$1,710,000
$\hat{Y}(\hat{X}^*)$	$1,649,000
$\mathrm{SE}[\hat{Y}(\hat{X}^*)]$	$29,700

2.2 Follow-Up Analysis for Transformation-Based Metamodel of Experiment 4

In this experiment the normalizing transformation with the best results was

$$f(y)=10^4y^{-1} \quad \text{so that} \quad f'(y)=-10^4y^{-2}. \qquad (13)$$

Table 6: Confidence Interval and Standard Error for the Predicted Response at the Stationary Point in Experiment 4

Statistics for Transformed Responses	Computed Value
$\hat{Z}(\hat{X}^*)$	0.005778
$\mathrm{SE}[\hat{Z}(\hat{X}^*)]$	0.00004158
Effective degrees of freedom ν	33
$t_{0.975,32}$	2.0372
95% Confidence Interval	[0.005693 , 0.005863]
Statistics for Untransformed Responses	**Computed Value**
$\hat{Y}_1(\hat{X}^*)$	\$1,756,000
$\hat{Y}_2(\hat{X}^*)$	\$1,706,000
$\hat{Y}(\hat{X}^*)$	\$1,731,000
$\mathrm{SE}[\hat{Y}(\hat{X}^*)]$	\$12,500

By inserting (13) into (2)–(5), we obtain

$$\hat{Y}_1(X)=10^4\left[\hat{Z}(X)-H(X)\right]^{-1},$$

$$\hat{Y}_2(X)=10^4\left[\hat{Z}(X)+H(X)\right]^{-1},$$

$$g_1\left[\hat{Z}(X),H(X)\right]=-\frac{1}{2}10^{-4}\left[\hat{Y}_1^2(X)+\hat{Y}_2^2(X)\right], \qquad (14)$$

and

$$g_2\left[\hat{Z}(X),H(X)\right]=\frac{1}{2}10^{-4}\left[\hat{Y}_1^2(X)-\hat{Y}_2^2(X)\right]. \qquad (15)$$

The statistic for estimating the standard error of $\hat{Y}(X)$ is obtained by evaluating equation (6) with results from (14)–(15). The resulting equation is

$$\mathrm{SE}\left[\hat{Y}(X)\right]=\frac{1}{2}10^{-4}\mathrm{SE}\left[\hat{Z}(X)\right] \times\sqrt{[\hat{Y}_1^2(X)+\hat{Y}_2^2(X)]^2+[\hat{Y}_1^2(X)-\hat{Y}_2^2(X)]^2\frac{t^2_{1-\alpha/2,\nu}}{2\nu}}.$$

A 95% confidence interval and the estimated standard error for the predicted response at the stationary point for results from experiment 4 are presented in Table 6. Again these results are close to those reported in Table 5 of Irizarry et al. (2000b).

2.3 Validation of Final Fitted Metamodel for Experiment 2

In Irizarry et al. (2000b), we concluded that experiment 2 yielded the optimum setting for the input factors defining the operation of the manufacturing cell. To validate the final fitted metamodel for experiment 2, we performed n = 20 independent replications of the simulation model at the estimated stationary point $\hat{X}^*$ for experiment 2, yielding an average annual cost of \$1,736,000 with a sample standard deviation of \$14,000. Table 7 summarizes the result of evaluating (7) and (8) with $X=\hat{X}^*$.

Table 7 shows that the 95% confidence interval (7) for the difference between the metamodel-predicted mean response and the mean response from the simulation runs does not contain zero. Therefore, we concluded that there is a *statistically* significant difference between the metamodel- and simulation-based estimates of the optimal expected cell cost. However, of greater importance is the practical significance of the difference between these two estimates as measured by the percentage deviation of the metamodel-based estimate from the simulation-based estimate,

$$\%\text{ deviation}=100\left\{\hat{Y}(\hat{X}^*)-\overline{Y}(\hat{X}^*)\right\}/\,\overline{Y}(\hat{X}^*)=-5\%.$$

Table 7: 95% Confidence Interval for Difference between Metamodel- and Simulation-Based Predictions of Optimal Cell Performance

Simulation-Based Estimates	Metamodel-Based Estimates
$\overline{Y}(\hat{X}^*)$ = \$1,736,060	$\hat{Y}(\hat{X}^*)$ = \$1,649,000
$S_{\overline{Y}(\hat{X}^*)}$ =\$14,000/$\sqrt{20}$ =\$3,151	$\mathrm{SE}[\hat{Y}(\hat{X}^*)]$ = \$29,700
n = 20	ν = 31
α = 0.05, η = 31, $t_{1-\alpha/2,\eta}$ = 2.040	
A 95% Confidence Interval for $\mathrm{E}[\hat{Y}(\hat{X}^*)-\overline{Y}(\hat{X}^*)]$ is [–\$148,000, –\$26,000].	

In our experience, a –5% deviation of a metamodel-based performance estimate from the corresponding simulation-based estimate is not practically significant; and thus we concluded that the metamodel could be used effectively to evaluate cell performance under other scenarios within the region of interest in the input-factor space.

3 MONTE CARLO STUDY

To compare the performance of the proposed confidence interval (2) for a transformation-based metamodel versus the conventional (or "naïve") method as specified in equations (9) and (10) of Irizarry et al. (2000b), we conducted a Monte Carlo performance evaluation of both techniques. The comparison is primarily based on the empirical coverage probability of nominal 95% confidence intervals for E[$Y(X)$] when the design point X is fixed.

For the Monte Carlo study, we examined the situation in which each simulation-generated response $Y(X)$ has a lognormal distribution, which is approximately representative of many metamodel estimation problems that we have encountered in practice. In this situation, the logarithm transformation of the original response yields a normally distributed variate—that is, we have $Z(X) = f[Y(X)] = \ln[Y(X)] \sim \mathrm{N}(\mathrm{E}[Z(X)], \sigma_Z^2)$. We assume that the transformed metamodel is linear in the unknown metamodel coefficients and has the correct functional form so that the residuals of the transformed metamodel have mean 0. For simplicity we take $\mathrm{E}[Z(X)] = 0$ and $W(X)(D^{\mathrm{T}}D)^{-1}W^{\mathrm{T}}(X) = 1$ so that $\mathrm{Var}[Z(X)] = \sigma_Z^2$.

The experimental procedure consisted of (a) generating 20 independent and identically distributed observations of the transformed response $Z(X) \sim \mathrm{N}(0, \sigma_Z^2)$; and (b) constructing nominal 95% confidence intervals for the mean E[$Y(X)$] of the untransformed response based on the proposed technique (2) as well as the conventional technique based on equations (9) and (10) of Irizarry et al. (2000b). To estimate the actual coverage probabilities delivered by these two confidence-interval procedures, we performed 10,000 replications of each procedure for different values of σ_Z ranging from 0.10 to 2.0. Table 8 summarizes the results of the experimental performance evaluation. Notice that for each entry in Table 8 with corresponding actual coverage probability 95%, the standard error of that entry is $\sqrt{0.95 \cdot 0.05/10000} = 0.002$ or 0.2%; and the standard error of every entry in Table 8 is at most 0.5%

Table 8: Percent Coverage of 95% Confidence Intervals (CIs) for E[$Y(X)$] Based on 10,000 Replications of Each CI

σ_Z	Proposed CI (2)	Conventional CI
0.10	95.5	95.1
0.25	95.2	93.1
0.50	94.8	87.5
0.75	93.8	80.3
1.00	92.8	73.0
1.25	91.4	65.2
1.50	89.1	58.0
1.75	86.9	50.5
2.00	84.5	43.3

Table 8 shows that the proposed confidence interval (2) substantially outperforms the conventional confidence interval based on equations (9) and (10) of Irizarry et al. (2000b) when the residual variance of the transformed metamodel is relatively large. For example, when σ_Z = 1.75, the actual coverage probability of nominal 95% confidence intervals is 86% for the proposed method (2) and 51% for the conventional method. We believe that the results in Table 8 provide good evidence of the robustness of our proposed confidence intervals against large residual variance in the transformed metamodel.

Table 9 summarizes the results of a similar experimental performance comparison of the proposed validation procedure based on displays (7) and (8) versus the conventional procedure based on displays (11) and (12) of Irizarry et al. (2000b).

Table 9: Percent Coverage of 95% CIs for Metamodel Validation Based on 10,000 Replications of Each CI

σ_Z	Proposed CI (7)	Conventional CI
0.10	95.2	94.8
0.25	96.2	94.8
0.50	97.1	94.8
0.75	97.0	94.9
1.00	96.7	94.8
1.25	96.2	94.4
1.50	95.5	93.8
1.75	94.7	93.2
2.00	93.7	92.3

The results in Table 9 suggest that the proposed metamodel validation procedure based on (7) and (8) may slightly outperform the conventional procedure, but the differences in performance do not appear to be practically significant. We believe that a more extensive performance evaluation of the two validation procedures is required before any definitive conclusions can be reached about the advantages and disadvantages of either procedure. This is the subject of ongoing research.

4 CONCLUSIONS AND RECOMMENDATIONS

In our experience applying response surface methodology to simulation experiments, we have found that it is often necessary to work with variance-stabilizing transformations of the original simulation-generated responses in order to obtain approximately normal responses with a nearly constant variance. Unfortunately once a metamodel has been fitted to the transformed responses, there are no readily available, clear-cut guidelines on how the transformed metamodel should be used to make valid inferences about the performance of the underlying simulation model in terms of the original performance measures of interest. In Irizarry et al. (2000b), we followed what might be considered the conventional approach to the analysis of such transformation-based metamodels.

In this paper we have proposed an alternative method for the analysis of transformation-based simulation metamodels which should be more robust than the conventional approach in applications that exhibit a large residual variance for the transformed metamodel. The results of a preliminary Monte Carlo performance evaluation provide substantial support for the claim that the proposed confidence interval (2) generally delivers acceptable coverage probability and significantly outperforms its conventional (naïve) counterpart—provided that the transformed simulation response is approximately normal and the metamodel fitted to the transformed responses is an adequate approximation to the true underlying transformed response surface.

It is also clear from our experimentation that as a point estimator of the mean E[$Y(X)$] of the original (untransformed) simulation response $Y(X)$, the midpoint (3) of the proposed confidence interval (2) can still have significant bias. We are currently investigating methods for effectively reducing the bias of (3) while still allowing the user to validate the fitted metamodel against the results of additional runs of the simulation model.

ACKNOWLEDGMENTS

This work was partly supported by NSF Grant DDM-9215432. We thank Barbara J. Hurley for many enlightening discussions on the development of the cost function. We also thank David Goldsman for several suggestions that improved the exposition.

REFERENCES

Box, G. E. P., and Draper, N. R. 1987. *Empirical model-building and response surfaces*. New York: John Wiley & Sons.

Hald, A. 1952. *Statistical theory with engineering applications*. New York: John Wiley & Sons.

Irizarry, M. 1996. Simulation-based methodology for the design of world-class manufacturing cells. Ph.D. dissertation, Department of Industrial Engineering, North Carolina State University, Raleigh, North Carolina.

Irizarry, M., Wilson, J. R., and Trevino, J. 2000a. A flexible simulation tool for manufacturing cell design, I: Model structure, operation, and case study. *IIE Transactions* to appear. Available on-line as <ftp://ftp.ncsu.edu/pub/eos/pub/jwilson/cs1r2.pdf> [accessed March 25, 2000].

Irizarry, M., Wilson, J.R., And Trevino, J. 2000b. A flexible simulation tool for manufacturing cell design, II: Response surface analysis and case study. *IIE Transactions* to appear. Available on-line via <ftp://ftp.ncsu.edu/pub/eos/pub/jwilson/cs2r3.pdf> [accessed March 25, 2000].

Montgomery, D. C. 1991. *Design and analysis of experiments.* 3d ed. New York: John Wiley & Sons.

Powell, K. A. 1992. Interactive decision support modeling for the textile spinning industry. Master's thesis, Department of Industrial Engineering, North Carolina State University, Raleigh, North Carolina.

SAS Institute. 1990. *SAS/STAT user's guide.* 4th ed. Cary, North Carolina: SAS Institute.

Stuart, A., and Ord, J. K. 1994. *Kendall's advanced theory of statistics, volume 1: Distribution theory.* 6th ed. London: Edward Arnold.

AUTHOR BIOGRAPHIES

MARIA DE LOS A. IRIZARRY is an Assistant Professor in the Department of Industrial Engineering at the University of Puerto Rico, Mayaguez Campus. She teaches and does research in the areas of simulation and ergonomic work design. She is a member of IIE, APICS, and HFES. She is a registered professional engineer. She received a Ph.D. degree from North Carolina State University and an M.S. degree from Texas A&M University, both in industrial engineering. Her e-mail address is <mariai@ece.uprm.edu>.

MICHAEL E. KUHL is an Assistant Professor in the Industrial and Manufacturing Systems Engineering Department at Louisiana State University. He received his B.S.I.E. from Bradley University and his M.S.I.E./O.R. and Ph.D. from North Carolina State University. He is a member of INFORMS, IIE, and ASEE. His interests include simulation input modeling, design and analysis of simulation experiments, and information systems. His email and web addresses are <mkuhl@lsu.edu> and <www.imse.lsu.edu/kuhl>.

EMILY K. LADA is a Ph.D. student in the Operations Research Graduate Program at North Carolina State University. She received a B.A. degree in mathematics from the University of North Carolina at Chapel Hill; and she received an M.S. degree in Operations Research from North Carolina State University. She is a member of Omega Rho and Phi Kappa Phi. Her email address is <eklada@eos.ncsu.edu>.

SRIRAM SUBRAMANIAN is a master's student in the Integrated Manufacturing Systems Engineering Program at North Carolina State University. His e-mail address is <ssubram3@unity.ncsu.edu>.

JAMES R. WILSON is Professor and Head of the Department of Industrial Engineering at North Carolina State University. From 1988 to 1992, he served as a departmental editor of *Management Science;* and since 1997, he has served as an area editor of *ACM Transactions on Modeling and Computer Simulation.* He has also held various offices in INFORMS–College on Simulation; and he currently serves as the corepresentative of that organization to the Board of Directors of the Winter Simulation Conference. He is a member of ASA, ACM, IIE, and INFORMS. His e-mail address is <jwilson@eos.ncsu.edu>, and his web page is <www.ie.ncsu.edu/jwilson>.