Statistical Adjustments to Engineering Models

V. ROSHAN JOSEPH*and SHREYES N. MELKOTE†

Georgia Institute of Technology, Atlanta, GA 30332

Abstract

Statistical models are commonly used in quality improvement studies. However, such models tend to perform poorly when predictions are made away from the observed data points. On the other hand, engineering models derived using the underlying physics of the process do not always match satisfactorily with reality. This article proposes engineering-statistical models that overcomes the disadvantages of engineering models and statistical models. The engineering-statistical model is obtained through some adjustments to the engineering model using experimental data. The adjustments are done in a sequential way and are based on empirical Bayes methods. We also develop approximate frequentist procedures for adjustments that are computationally much easier to implement. The usefulness of the methodology is illustrated using a problem of predicting surface roughness in a micro-cutting process and the optimization of a spot welding process.

Key Words: Bayesian methods; Calibration; Computer experiments; Gaussian process; Prediction; Semi-empirical models; Validation.

*Dr. Joseph is an Associate Professor in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Tech. He is a Member of ASQ. His email address is roshan@gatech.edu.

†Dr. Melkote is a Professor in the George W. Woodruff School of Mechanical Engineering at Georgia Tech. His email address is smelkote@gatech.edu.

Introduction

Models describing the performance of a physical process are essential for prediction, control, and optimization. In general, there are two approaches for developing models. One approach is to develop models based on the engineering/physical laws governing the process, which include analytical models and finite element models. We call such models engineering models. For example, in a metal cutting process, an analytical model can be developed using the slip line field method for the analysis of stresses and strains induced in the material during the cutting (Oxley 1989). A finite element model for the same process can also be developed, which accounts for the process geometry/kinematics and the intricate physics of finite material deformation during cutting (Marusich and Ortiz 1995). Another approach to develop models is to postulate statistical models and estimate them based on the data generated from the process. Although engineering models are quite common in product design (Santner, Williams, and Notz 2003), statistical models are more commonly used in quality improvement (Wu and Hamada 2000).

Both modeling approaches have drawbacks. Predictions derived from engineering models are often not accurate. This is because engineering models are developed based on several simplifying assumptions, which may not hold true in practice. On the other hand, statistical models can give good prediction at points close to the observed data. However, when trying to predict in regions away from data, predictions can be poor and may at times make no physical sense. Moreover, the experimental data required for estimating the statistical models can be expensive.

Our objective is to develop a modeling approach that utilizes both engineering models and statistical models. We term such models *engineering-statistical models*. This approach combines the advantages of the two models and thus, mitigates their individual drawbacks. Engineering-statistical models are expected to produce more realistic predictions than engineering models and are less expensive to estimate than statistical models.

To illustrate the idea, consider the problem of predicting surface roughness as a function of feed in a micro-cutting process. Figure 1 shows an illustration of a turning operation

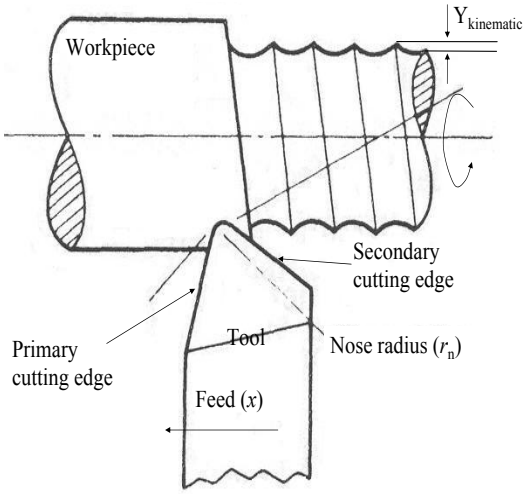

Figure 1: Illustration of turning operation showing surface roughness.

(adapted from Shaw 1997). As can be seen in the figure, the waveform left on the finished surface is mainly due to the tool nose geometry and feed. An analytical model for the peak-to-valley surface roughness (Y) considering only the kinematics of machining is given by (Shaw 1997)

$$Y_{kinematic} = \frac{x^2}{8r_n}, \tag{1}$$

where x is the feed (in microns/revolution) and r_n is the tool nose radius (in microns). Figure 2 shows the surface roughness data obtained from micro-turning of an aluminium alloy using a diamond tool with nose radius of 800 microns (Liu and Melkote 2006). Clearly, the predictions of the engineering model (dashed line) are poor. Now consider the statistical modeling approach. Suppose that we have data only up to $x = 20$. A quadratic model is fitted to the data and is also shown in the figure (dotted line). We can see that although the statistical model is useful for predicting surface roughness in the range 5-20 microns/revolution, it behaves poorly outside this range. An engineering-statistical model is also fitted using the data with $x \leq 20$ and is shown in the figure (solid line). The details of the model and estimation will be given later. Clearly, the engineering-statistical model is much better than both the engineering model and the statistical model. Note that the statistical model gave the best fit to the data, but seems to be the worst among the three models. The engineering-statistical

model came out as the best model because it behaves like the engineering model and is also close to the data. This provides the intuition that statistical adjustments to the engineering model should be made in such a way that they bring the engineering model closer to the data by making minimal changes to it.

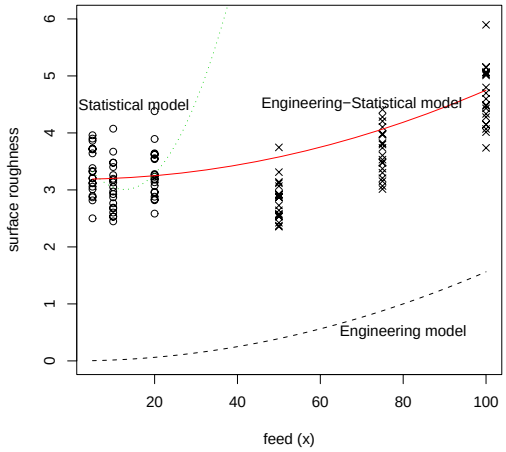

Figure 2: Comparison of different models. The statistical model and the engineering-statistical model are fitted using data for which feed ≤ 20.

The basic idea of improving engineering models using experimental data is not new. The so-called mechanistic models are obtained by estimating the unknown parameters (known as calibration parameters) in the engineering model from real data (see, for examples, Box and Hunter 1962, Kapoor et al. 1998). However, such models are not general. For example, there can be an engineering model with no calibration parameters and thus, the mechanistic modeling approach cannot be used. Moreover, such models do not correct for model inadequacy. Kennedy and O'Hagan (2001) proposed a model using Gaussian process that is capable of accounting for the model inadequacy. Other recent work in this area include Reese et al. (2004), Higdon et al. (2004), Bayarri et al. (2007), and Qian and Wu (2008). Our approach for developing engineering-statistical models is built on this earlier work. However, our approach differs in terms of the modeling and estimation, which will be explained in later sections.

Methodology

Let Y be the output of the physical process and $\boldsymbol{x} = (x_1, \cdots, x_p)'$ be the factors. The output is random due to the presence of noise (uncontrollable) factors and measurement error. Denote the random error by ϵ. Then

$$Y = \mu(\boldsymbol{x}) + \epsilon, \tag{2}$$

where $\mu(\boldsymbol{x})$ is the mean of Y at a given $\boldsymbol{x}$ and $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Our objective is to find the unknown function $\mu(\boldsymbol{x})$. What we have is the engineering model $f(\boldsymbol{x}; \boldsymbol{\eta})$ and the output from a physical experiment $(\boldsymbol{x}_1, y_1)$, ..., $(\boldsymbol{x}_n, y_n)$, where $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_q)'$ denotes the unknown calibration parameters. Note that the argument of $\boldsymbol{\eta}$ is omitted from $\mu(\boldsymbol{x})$, because the y_i's are generated from (2) with $\boldsymbol{\eta}$ fixed at its true value $\boldsymbol{\eta}^*$. Of course, $\boldsymbol{\eta}^*$ is unknown to the investigator and needs to be estimated from the physical observations.

A Bayesian approach is useful in formulating this problem. Because the engineering model is available before obtaining the data, it can be used to postulate a prior distribution for $\mu(\boldsymbol{x})$. Specifically, the output of the engineering model is taken as the mean of the prior distribution. Then, based on the data we can obtain the posterior distribution of $\mu(\boldsymbol{x})$. The posterior distribution incorporates information from the engineering model as well as information from the data. This is exactly what we need. The engineering-statistical model is simply the mean of the posterior distribution.

The main task in the Bayesian approach is in postulating an appropriate prior distribution for $\mu(\boldsymbol{x})$. Kennedy and O'Hagan (2001) made a major advancement in this approach by postulating a Gaussian process prior for $\mu(\boldsymbol{x})$ that incorporates $f(\boldsymbol{x}; \boldsymbol{\eta})$ in its mean part. Different from them, we propose to postulate the prior in two stages. This leads to a sequential model building strategy, which helps in identifying a prediction model that makes minimal changes to the engineering model. The strategy is depicted in Figure 3 and is explained below.

For ease of exposition, first consider an engineering model without calibration parameters. We later explain how calibration parameters can be incorporated into the analysis. The first

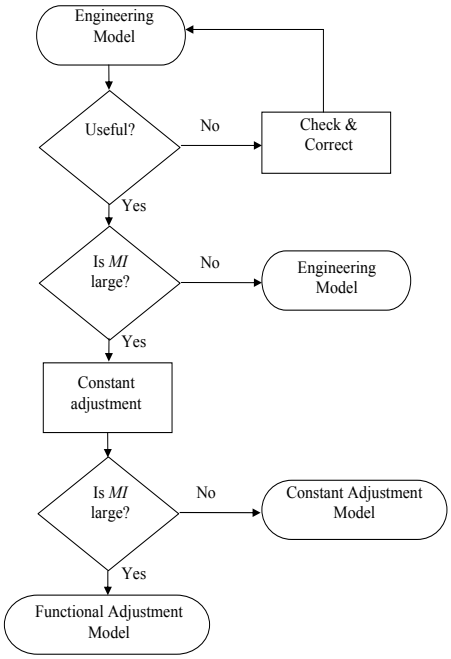

Figure 3: Sequential model building strategy.

step in the sequential model building strategy is to check if the engineering model is useful or not. This can be done by using graphical plots such as a y_i versus $f(\boldsymbol{x}_i)$ plot. If there is a positive correlation, then we conclude that the engineering model is useful for prediction. Otherwise, the engineer should go back and verify the assumptions made in deriving the engineering model and make corrections to it. (If it is not possible to do this, we may resort to fitting a pure statistical model.) If the engineering model is deemed useful, then we can check how good the engineering model is. Denote the predictor at this stage by $\hat{\mu}^E(\boldsymbol{x})$, which is equal to $f(\boldsymbol{x})$. Define a model inadequacy measure:

$$MI = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}_i^E)^2,$$

where $\hat{\mu}_i^E = f(\boldsymbol{x}_i)$. If the engineering model is good, then $MI \approx \sigma^2$ for large n (estimation of σ^2 will be discussed later). Therefore, if MI is small enough (close to or less than σ^2), then we can stop and use the engineering model for prediction. Otherwise, we proceed to

make statistical adjustments.

First we do a simple location-scale adjustment:

$$\mu(\boldsymbol{x}) - f(\boldsymbol{x}) = \beta_0 + \beta_1(f(\boldsymbol{x}) - \bar{f}),$$

where $\bar{f} = \sum_{i=1}^{n} f(\boldsymbol{x}_i)/n$. The prior for $\mu(\boldsymbol{x})$ is postulated through the parameters: $\beta_0 \sim \mathcal{N}(0, \tau_0^2)$ and $\beta_1 \sim \mathcal{N}(0, \tau_1^2)$, where β_0 and β_1 are independent. Note that the mean of $\mu(\boldsymbol{x})$ is $f(\boldsymbol{x})$. We call this a *constant adjustment model*, because two constants are used for the adjustments. The predictor at this stage is denoted by $\mu^C(\boldsymbol{x}) = f(\boldsymbol{x}) + \beta_0 + \beta_1(f(\boldsymbol{x}) - \bar{f})$. We again compute the model inadequacy measure:

$$MI = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}_i^C)^2,$$

where $\hat{\mu}_i^C = f(\boldsymbol{x}_i) + \hat{\beta}_0 + \hat{\beta}_1(f(\boldsymbol{x}_i) - \bar{f})$ and $\hat{\beta}_j$ denotes the estimate of β_j from the data. If MI is small enough (close to or less than σ^2), we stop and use the constant adjustment model for prediction. Otherwise, we proceed to make a more sophisticated adjustment to the engineering model.

Let

$$\mu(\boldsymbol{x}) - \mu^C(\boldsymbol{x}) = \delta(\boldsymbol{x}; \boldsymbol{\alpha}),$$

where $\delta(\boldsymbol{x}; \boldsymbol{\alpha})$ is used for capturing the inadequacy of the constant adjustment model. We call this a *functional adjustment model*, because the functional form of the predictor is different from the engineering model. The form of the function δ is obtained using residual analysis from the constant adjustment model. A simple method is to use a linear model given by $\delta(\boldsymbol{x}; \boldsymbol{\alpha}) = \sum_{i=0}^{m} \alpha_i u_i(\boldsymbol{x})$, where u_i's are known functions in $\boldsymbol{x}$. The prior for $\boldsymbol{\alpha}$ is chosen to make $E(\delta(\boldsymbol{x}; \boldsymbol{\alpha})) = 0$. Thus, we must have $E(\boldsymbol{\alpha}) = 0$ in the linear model. Now, $\mu^C(\boldsymbol{x})$ and $\boldsymbol{\alpha}$ can be estimated from data. The functional adjustment predictor is given by $\hat{\mu}^F(\boldsymbol{x}) = \hat{\mu}^C(\boldsymbol{x}) + \delta(\boldsymbol{x}; \hat{\boldsymbol{\alpha}})$. Although simultaneously estimating $\mu^C(\boldsymbol{x})$ and $\boldsymbol{\alpha}$ from the data may lead to better fitting, we propose to do it in two stages, i.e., use the estimate of $\mu^C(\boldsymbol{x})$ from the previous stage and estimate only $\boldsymbol{\alpha}$ in the current stage. This two-stage estimation procedure reduces the tendency for the statistical model to replace the engineering model

by forcing it to be present in the final predictor. This may not give the best fit in terms of the observed data, but as illustrated in Figure 2, it may give a better prediction at the unobserved points. Moreover, it helps in avoiding certain identifiability problems and makes the computations much simpler.

The functional adjustment model is closely related to the model in Kennedy and O'Hagan (2001). Their model can be written as $\mu(\boldsymbol{x}) = \rho_0 + \rho_1 f(\boldsymbol{x}) + \delta(\boldsymbol{x})$, where $f(\boldsymbol{x})$ and $\delta(\boldsymbol{x})$ are assumed to be independent Gaussian processes. A Gaussian process prior is used for $f(\boldsymbol{x})$ because the engineering model is assumed to be computationally intensive to evaluate and could be observed only at some finite locations. First, following the suggestion of Bayarri et al. (2007), we do not use physical data to estimate $f(\boldsymbol{x})$. Therefore, we only need to consider models for given $f(\boldsymbol{x})$. When $f(\boldsymbol{x})$ is complex, it will be replaced by an easy-to-evaluate metamodel (a model that approximates $f(\boldsymbol{x})$). Apart from the two-stage fitting of the functional adjustment model, there are other differences in our approach compared to that of Kennedy and O'Hagan. The prior specification in our model is quite different. Kennedy and O'Hagan assume independent and improper priors for ρ_0 and ρ_1, whereas because of the reparameterization, we use dependent and proper priors for these two parameters. Different from the Gaussian process model, we use a linear/nonlinear regression model for $\delta(\boldsymbol{x})$. This enables us to identify important factors affecting the model discrepancy through variable selection and use minimal adjustments to the engineering model. This agrees with our overall objective of identifying a simple adjustment model, which is a major difference from the Kennedy and O'Hagan's approach.

The unknown calibration parameters ($\boldsymbol{\eta}$) can be easily incorporated into the sequential model building strategy by treating them as hyperparameters. This is quite legitimate because $\boldsymbol{\eta}$ enters the Bayesian model only through $f(\boldsymbol{x}; \boldsymbol{\eta})$, which is taken as the mean of the prior distribution of $\boldsymbol{\mu}(\boldsymbol{x})$. We start the procedure with a least squares estimate of $\boldsymbol{\eta}$ and in the next stage obtain an empirical Bayes estimate from the constant adjustment model. Because $\boldsymbol{\eta}$ does not appear in $\delta(\boldsymbol{x}; \boldsymbol{\alpha})$, the fitting of functional adjustment model remains the same as before. This approach is markedly different from the previous approaches (see, Kennedy and O'Hagan 2001, Bayarri et al. 2007), wherein $\boldsymbol{\eta}$ and δ are simultaneously

estimated from the data. Researchers have pointed out identifiability problems with such an approach and associated convergence problems with Markov chain Monte Carlo implementation (see, e.g., Loeppky, Bingham, and Welch 2006). This identifiability problem is completely avoided by virtue of our two-stage fitting procedure.

Note that Kennedy and O'Hagan (2001) and others directly fit the functional adjustment model, but we do it only at the last stage of our sequential model building strategy and only if it is necessary. This is because, a functional adjustment model may completely change the basic shape of the engineering model. Such a major change should be considered necessary only when the constant adjustment model is not working well. To draw an analogy, a surgery may be considered for treatment only when the patient is not satisfactorily responding to the prescribed drugs.

We use empirical Bayes methods for estimation of the models and derive explicit expressions for some of the results. This makes our approach much easier to implement than the procedures in Kennedy and O'Hagan (2001), Reese et al. (2004), Higdon et al. (2004), Bayarri et al. (2007), and Qian and Wu (2008). We also develop approximate frequentist procedures to our empirical Bayes approach, which are computationally very efficient and should appeal to many practitioners. Of course, there is no free lunch; we admit that uncertainties about the hyperparameters are not accounted for in our approach and therefore, the prediction intervals are not as accurate as those obtained by using hierarchical Bayes methods.

In the next two sections, we explain the details of fitting constant and functional adjustment models.

Constant Adjustment Model

To make the exposition simple, consider first an engineering model with no calibration parameters. The details of the adjustments when calibration parameters are present will be

discussed later. The constant adjustment model is given by

$$Y - f(\boldsymbol{x}) = \beta_0 + \beta_1(f(\boldsymbol{x}) - \bar{f}) + \epsilon,$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$, $\beta_0 \sim \mathcal{N}(0, \tau_0^2)$, $\beta_1 \sim \mathcal{N}(0, \tau_1^2)$, and they are independent. For the moment, assume that σ^2 is known.

We have the data $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)$. Let $\boldsymbol{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1)'$, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$. Let $\boldsymbol{F}$ be an $n \times 2$ matrix, whose first column is a column of 1's and the second column is $(f_1 - \bar{f}, \ldots, f_n - \bar{f})'$ and $\boldsymbol{f} = (f_1, \ldots, f_n)'$, where $f(\boldsymbol{x}_i) = f_i$. Then, the constant adjustment model can be written as

$$\boldsymbol{y} - \boldsymbol{f} = \boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}), \quad \boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}), \tag{3}$$

where $\boldsymbol{0}$ is a vector of 0's, $\boldsymbol{I}$ is the n-dimensional identity matrix, and $\boldsymbol{\Sigma} = diag\{\tau_0^2, \tau_1^2\}$. The posterior distribution of $\boldsymbol{\beta}$ is given by

$$\boldsymbol{\beta}|\boldsymbol{y} \sim \mathcal{N}\left((\boldsymbol{F}'\boldsymbol{F} + \sigma^2\boldsymbol{\Sigma}^{-1})^{-1}\boldsymbol{F}'(\boldsymbol{y} - \boldsymbol{f}),\ \sigma^2(\boldsymbol{F}'\boldsymbol{F} + \sigma^2\boldsymbol{\Sigma}^{-1})^{-1}\right). \tag{4}$$

Denote the posterior mean of $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}} = (\boldsymbol{F}'\boldsymbol{F} + \sigma^2\boldsymbol{\Sigma}^{-1})^{-1}\boldsymbol{F}'(\boldsymbol{y} - \boldsymbol{f})$. Then, the constant adjustment predictor is given by

$$\hat{\mu}^C(\boldsymbol{x}) = f(\boldsymbol{x}) + \hat{\beta}_0 + \hat{\beta}_1(f(\boldsymbol{x}) - \bar{f}). \tag{5}$$

The formulas can be simplified. Let $\tilde{\boldsymbol{\beta}} = (\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'(\boldsymbol{y} - \boldsymbol{f})$, which is the least squares estimate of $\boldsymbol{\beta}$. We obtain

$$\tilde{\beta}_0 = \bar{y} - \bar{f} \ \text{ and } \ \tilde{\beta}_1 = \sum_{i=1}^{n}(y_i - f_i)(f_i - \bar{f})/S, \tag{6}$$

where $S = \sum_{i=1}^{n}(f_i - \bar{f})^2$. Then,

$$\hat{\beta}_0 = \frac{\tau_0^2}{\tau_0^2 + \sigma^2/n}\tilde{\beta}_0 \ \text{ and } \ \hat{\beta}_1 = \frac{\tau_1^2}{\tau_1^2 + \sigma^2/S}\tilde{\beta}_1. \tag{7}$$

Thus, $\hat{\beta}_0$ and $\hat{\beta}_1$ are shrinkage estimates of their least squares estimates.

If the model inadequacy measure MI for the constant adjustment model is small, then it can be used for prediction. The $(1-\alpha)$-Bayesian prediction interval is given by

$$\hat{\mu}^C(\boldsymbol{x}) \pm z_{\alpha/2}\sigma \left\{1 + \frac{1}{n+\sigma^2/\tau_0^2} + \frac{(f(\boldsymbol{x})-\bar{f})^2}{S+\sigma^2/\tau_1^2}\right\}^{1/2}, \tag{8}$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. Note that we had assumed the engineering model to be known. However, in some cases the engineering model can be quite complex and observable only at some finite locations. As mentioned before, in such cases the predictions can be made by replacing $f(\boldsymbol{x})$ with a metamodel $\hat{f}(\boldsymbol{x})$, i.e., $\hat{\mu}^C(\boldsymbol{x}) = \hat{f}(\boldsymbol{x}) + \hat{\beta}_0 + \hat{\beta}_1(\hat{f}(\boldsymbol{x}) - \bar{f})$. However, the prediction interval needs to be modified to account for the uncertainty in the metamodel. Let $v(\boldsymbol{x})$ denote the prediction variance of the metamodel at the location $\boldsymbol{x}$. Assume that the engineering model evaluations are available at the locations of the physical experiment. Then, as shown in the Appendix, an approximate prediction interval is given by

$$\hat{\mu}^C(\boldsymbol{x}) \pm z_{\alpha/2}\sigma \left\{1 + \frac{1}{n+\sigma^2/\tau_0^2} + \frac{(f(\boldsymbol{x})-\bar{f})^2 + v(\boldsymbol{x})}{S+\sigma^2/\tau_1^2} + (1+\hat{\beta}_1)^2 v(\boldsymbol{x})\right\}^{1/2}. \tag{9}$$

Of course, the predictions can be made only after specifying the values of σ^2, τ_0^2, and τ_1^2. First consider the case of σ^2. It can be estimated from replicates and prior knowledge. Suppose that we have r replicates at d distinct values of $\boldsymbol{x}$. Thus, $n = rd$. Let y_{ij} be the jth replicate in the ith group, $i = 1, \ldots, d$ and $j = 1, \ldots, r$. Let $s_i^2 = \sum_{j=1}^r (y_{ij} - \bar{y}_i)^2/(r-1)$ be the sample variance of the ith group and $s^2 = \sum_{i=1}^d s_i^2/d$ the pooled variance. Assume an inverse gamma prior distribution for σ^2 with parameters $\nu/2$ and $\nu\sigma_0^2/2$, where σ_0^2 is an estimate of σ^2 from the prior knowledge (which could come from gage repeatability and reproducibility studies and experience with the product/process) and ν controls its uncertainty. Then, $\sigma^2|s^2$ also has an inverse gamma distribution with parameters $(d(r-1)+\nu)/2$ and $(d(r-1)s^2 + \nu\sigma_0^2)/2$. Therefore, σ^2 can be estimated by

$$\hat{\sigma}^2 = \frac{d(r-1)s^2 + \nu\sigma_0^2}{d(r-1)+\nu}, \tag{10}$$

which is a weighted average of s^2 and σ_0^2 (This estimate lies between the posterior mean and posterior mode of $\sigma^2|s^2$ and thus is closer to the posterior median). However, after obtaining

this estimate we act as though it was known. In other words, we assume $d(r-1)+\nu$ to be large enough to neglect the uncertainty of $\hat{\sigma}^2$ in the subsequent inference.

The hyperparameters τ_0^2 and τ_1^2 can be estimated from data using empirical Bayes methods. By integrating out $\boldsymbol{\beta}$ from (3), we obtain the marginal distribution of $\boldsymbol{y}$ as $\mathcal{N}(\boldsymbol{f}, \boldsymbol{F\Sigma F}' + \sigma^2\boldsymbol{I})$. Thus, the marginal log-likelihood is given by

$$l = -\frac{1}{2}\log det(\boldsymbol{F\Sigma F}' + \sigma^2\boldsymbol{I}) - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{f})'(\boldsymbol{F\Sigma F}' + \sigma^2\boldsymbol{I})^{-1}(\boldsymbol{y}-\boldsymbol{f}), \tag{11}$$

where the constant term is omitted. Now l can be maximized with respect to τ_0^2 and τ_1^2 to obtain their estimates.

We can derive explicit expressions for the maximum likelihood estimates. It can be shown that maximizing l with respect to τ_0^2 and τ_1^2 is equivalent to minimizing

$$\log(n\tau_0^2+\sigma^2) + \log(\tau_1^2 S + \sigma^2) + \frac{n\tilde{\beta}_0^2}{n\tau_0^2+\sigma^2} + \frac{S\tilde{\beta}_1^2}{\tau_1^2 S+\sigma^2}.$$

Under the nonnegativity constraints $\tau_0^2 \geq 0$ and $\tau_1^2 \geq 0$, we obtain

$$\hat{\tau}_0^2 = \left(\tilde{\beta}_0^2 - \sigma^2/n\right)_+ \text{ and } \hat{\tau}_1^2 = \left(\tilde{\beta}_1^2 - \sigma^2/S\right)_+, \tag{12}$$

where $(x)_+ = x$ if $x > 0$ and 0 otherwise. Substituting in (7), we obtain

$$\hat{\beta}_0 = \left(1 - \frac{1}{z_0^2}\right)_+ \tilde{\beta}_0 \text{ and } \hat{\beta}_1 = \left(1 - \frac{1}{z_1^2}\right)_+ \tilde{\beta}_1, \tag{13}$$

where

$$z_0 = \frac{\tilde{\beta}_0}{\sigma/\sqrt{n}} \text{ and } z_1 = \frac{\tilde{\beta}_1}{\sigma/\sqrt{S}}. \tag{14}$$

Thus, $\hat{\beta}_j$ is shrunk to 0 when $|z_j| \geq 1$. Note that z_j is the test statistic for testing the hypotheses $H_0 : \beta_j = 0$. Thus, although we started with a Bayesian approach, the final results are close to a frequentist approach. As shown by Joseph and Delaney (2008), the best approximated version of the frequentist testing to the Bayesian shrinkage estimation uses a critical value of $\sqrt{2}$. This amounts to making the shrinkage coefficient $(1-1/z_j^2)_+$ equal to 0 when the shrinkage coefficient is less than 1/2 and 1 otherwise. Thus, to obtain a constant adjustment model, we may use simple linear regression to fit $y_i - f_i = \beta_0 + \beta_1(f_i -$

$\bar{f}) + \epsilon_i$ and force β_j to be 0 if $|z_j| < \sqrt{2}$. This clearly shows the simplicity of the proposed approach. Because of its simplicity and ease of implementation, which will be more evident from the next section, we recommend using the frequentist approach over the empirical Bayes approach.

By letting τ_0^2 and τ_1^2 go to ∞ in (8), we obtain the familiar prediction interval for the frequentist approach given by

$$\hat{\mu}^C(\boldsymbol{x}) \pm z_{\alpha/2}\sigma \left\{1 + \frac{1}{n} + \frac{(f(\boldsymbol{x}) - \bar{f})^2}{S}\right\}^{1/2}. \tag{15}$$

Note that if any of the coefficients is truncated to 0, then the corresponding term in the prediction variance should be removed (note that $1/n$ and $(f(\boldsymbol{x}) - \bar{f})^2/S$ correspond to β_0 and β_1, respectively). The approximate prediction interval that incorporates the uncertainty in the metamodel can be obtained from (9) as

$$\hat{\mu}^C(\boldsymbol{x}) \pm z_{\alpha/2}\sigma \left\{1 + \frac{1}{n} + \frac{(f(\boldsymbol{x}) - \bar{f})^2 + v(\boldsymbol{x})}{S} + (1 + \hat{\beta}_1)^2 v(\boldsymbol{x})\right\}^{1/2}, \tag{16}$$

where $\hat{\beta}_1 = \tilde{\beta}_1$ if $|z_1| < \sqrt{2}$ and 0 otherwise.

We illustrate the methodology using the surface roughness example given in the introduction. From (1) with $r_n = 800$, we obtain $f_i = x_i^2/6400$. The plot of the measured surface roughness values (y_i) against the f_i values is shown in Figure 4. We can see that there is a positive correlation between y and f. Thus, the engineering model is useful for prediction. We can also see from Figure 4 that y does not go to 0 as f goes to 0. Moreover, the variance in y seems to be approximately constant over different levels of f. These indicate that a location-scale model with an additive error term is appropriate. Otherwise, some transformations may be applied.

In this example, we have 20 measured values of surface roughness at each of the six levels of feed. Thus, $r = 20$ and $d = 6$. Without using any prior knowledge, we obtain $\hat{\sigma}^2 = s^2 = .183$. The model inadequacy measure for the engineering model can be computed as

$$MI = \frac{1}{120}\sum_{i=1}^{120}(y_i - f_i)^2 = 9.12,$$

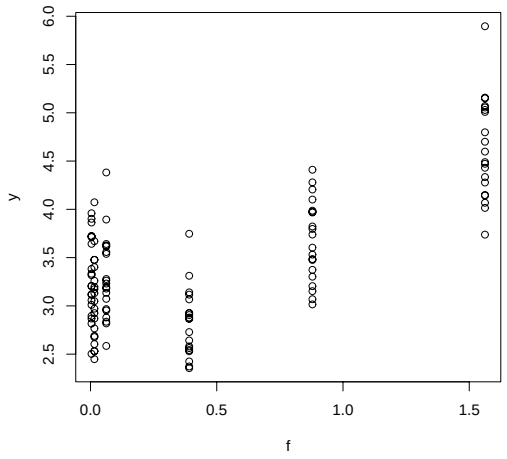

Figure 4: Plot of y against f in the surface roughness example.

which is much larger than $\hat{\sigma}^2$ and therefore, the engineering model is not adequate.

A hypothesis test can be performed to check this. Because the engineering model is deterministic $f_{ij} = f_{i\cdot}$ for all $j = 1, \ldots, r$. Let $\mu_{i\cdot}$ be the true mean value for the ith group, $i = 1, \ldots, d$. Then, testing the adequacy of the engineering model can be formulated as $H_0 : \mu_{i\cdot} = f_{i\cdot}$ against $H_1 : \mu_{i\cdot} \neq f_{i\cdot}$, $i = 1, \ldots, d$. Now, H_0 can be rejected if $\sum_{i=1}^{d} (\bar{y}_i - f_{i\cdot})^2 > \sigma^2 / r \chi^2_{d,\alpha}$, where $\chi^2_{d,\alpha}$ is the upper α-quantile of the χ^2_d distribution. This can be written in terms of MI as

$$MI > \frac{r-1}{r} s^2 + \frac{\sigma^2}{n} \chi^2_{d,\alpha}.$$

Note that when σ^2 is not known, it should be replaced with $\hat{\sigma}^2$. At $\alpha = .05$, $19/20 \times .183 + .183/120 \times 12.59 = .193 < MI$, which confirms that the engineering model is not good for prediction. Hence we proceed to fit the constant adjustment model.

Using a simple linear regression for $y - f(x)$ on $f(x) - \bar{f}$, we obtain $\tilde{\beta}_0 = 2.98$ and $\tilde{\beta}_1 = -.11$. From (14), $|z_0| = 2.98/\sqrt{.183/120} = 76.2$ and $|z_1| = .11/\sqrt{.183/39.1} = 1.60$. Thus, the empirical Bayes method gives $\hat{\mu}^C(x) - f(x) = 2.98 - .07(f(x) - .4857)$, whereas the frequentist approximation gives

$$\hat{\mu}^C(x) - f(x) = 2.98 - .11(f(x) - .4857). \tag{17}$$

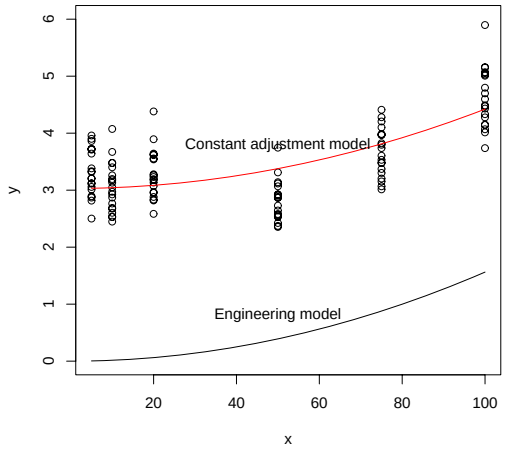

Figure 5: Constant adjustment model in the surface roughness example.

Hereafter, we use the model obtained through frequentist approximation.

A plot of the predictor is shown in Figure 5, which clearly shows great improvement compared to the engineering model. To quantify the performance, we can compute the model inadequacy measure for the constant adjustment model as

$$MI = \frac{1}{120}\sum_{i=1}^{120}(y_i - \hat{\mu}_i^C)^2 = .255.$$

This is much smaller than the engineering model and only slightly larger than $\hat{\sigma}^2 = .183$, showing that the constant adjustment model has greatly improved the prediction. If the constant adjustment model is true, then $\sum_{i=1}^{d}(\bar{y}_i - \hat{\mu}_{i.}^C)^2/(\sigma^2/r)$ follows a χ^2_{d-2} distribution (χ^2_{d-1} if β_0 or β_1 is truncated to 0). Therefore, the constant adjustment model is inadequate if

$$MI > \frac{r-1}{r}s^2 + \frac{\sigma^2}{n}\chi^2_{d-2,\alpha}.$$

Note that if we replace σ^2 with s^2 and use the approximation $\chi^2_{d-2,\alpha} \approx (d-2)F_{d-2,d(r-1),\alpha}$, then the above test reduces to the well known lack-of-fit test in regression analysis (see, e.g., Neter et al. 1996, page 123). The two tests are approximately the same when $d(r-1)$ is large. We obtain at $\alpha = .05$, $19/20 \times .183 + .183/120 \times 9.49 = .188 < MI$, and thus, the

constant adjustment model still has some model inadequacy. Therefore, we proceed to fit a functional adjustment model.

Functional Adjustment Model

The functional adjustment model is given by

$$Y - \mu^C(\boldsymbol{x}) = \delta(\boldsymbol{x}; \boldsymbol{\alpha}) + \epsilon,$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$, $\delta(\boldsymbol{x}; \boldsymbol{\alpha})$ captures the discrepancy between the observations and the constant adjustment model, and $\boldsymbol{\alpha}$ is a set of unknown parameters. Kennedy and O'Hagan (2001) proposed using a Gaussian process prior for the model inadequacy term in their model. Such priors are popular in the analysis of deterministic computer experiments (see Sacks et al. 1989). However, the data from the physical process are corrupted by random errors and therefore, there is no compelling reason to use such priors in this problem. A more common approach is to use linear and nonlinear regression models.

Consider a linear model

$$\delta(\boldsymbol{x}; \boldsymbol{\alpha}) = \alpha_0 + \sum_{i=1}^{m} \alpha_i u_i(\boldsymbol{x}), \tag{18}$$

where $u_i(\boldsymbol{x})$'s are known functions of $\boldsymbol{x}$ and $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_m)'$ are unknown parameters. Center the u_i's such that $\sum_{j=1}^{n} u_i(\boldsymbol{x}_j) = 0$. Now we can put a prior on $\boldsymbol{\alpha}$, say $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{0}, \gamma^2 \boldsymbol{R})$ (see Joseph 2006a, Joseph and Delaney 2007). As discussed before, we use $\hat{\mu}^C(\boldsymbol{x})$ from the previous stage and estimate only $\boldsymbol{\alpha}$ at this stage. Let $\hat{\boldsymbol{\mu}}^C$ be the fitted values from the constant adjustment model and $\boldsymbol{U}$ be the $n \times (m+1)$ model matrix corresponding to (18), i.e., the first column of $\boldsymbol{U}$ is a column of 1's and the other m columns correspond to the $u_i(\boldsymbol{x})$'s which take values depending on the n data points. Then, the functional adjustment predictor is given by

$$\hat{\mu}^F(\boldsymbol{x}) = \hat{\mu}^C(\boldsymbol{x}) + \sum_{i=0}^{m} \hat{\alpha}_i u_i(\boldsymbol{x}), \tag{19}$$

where $\hat{\mu}^C(\boldsymbol{x})$ is given in (5) and

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{U}'\boldsymbol{U} + \frac{\sigma^2}{\gamma^2}\boldsymbol{R}^{-1})^{-1}\boldsymbol{U}'(\boldsymbol{y} - \hat{\boldsymbol{\mu}}^C).$$

As before, the hyperparameters γ^2 and the parameters in $\boldsymbol{R}$ can be estimated by maximizing the log-likelihood (for given $\mu^C(\boldsymbol{x}) = \hat{\mu}^C(\boldsymbol{x})$)

$$l = -\frac{1}{2}\log det(\gamma^2 \boldsymbol{U}\boldsymbol{R}\boldsymbol{U}' + \sigma^2 \boldsymbol{I}) - \frac{1}{2}(\boldsymbol{y} - \hat{\boldsymbol{\mu}}^C)'(\gamma^2 \boldsymbol{U}\boldsymbol{R}\boldsymbol{U}' + \sigma^2 \boldsymbol{I})^{-1}(\boldsymbol{y} - \hat{\boldsymbol{\mu}}^C).$$

When $\boldsymbol{U}$ is an orthogonal matrix, the empirical Bayes approach leads to shrinkage estimates of $\boldsymbol{\alpha}$ and thus can be approximated with a frequentist hypothesis testing procedure that forces a parameter to 0 when the corresponding $|z|$ statistic is less than $\sqrt{2}$ (see Joseph and Delaney 2008). When $\boldsymbol{U}$ is not orthogonal, we may use a variable selection technique such as forward selection with a $|z|$ critical value-to-enter as $\sqrt{2}$. Because empirical Bayes estimation shrinks the coefficients, it is even more appropriate to use methods that do shrinkage and variable selection simultaneously such as the nonnegative garrote (Breiman 1995) and the least absolute shrinkage and selection operator (Tibshirani 1996). The above frequentist procedures are computationally very efficient and extremely simple to implement.

The two stage estimation creates complications in deriving explicit expressions for the prediction intervals. We are able to provide formulas only for the case of the frequentist approach. Let $\boldsymbol{w}(\boldsymbol{x})' = (1, f(\boldsymbol{x}) - \bar{f})$ and $\boldsymbol{u}(\boldsymbol{x})' = (u_1(\boldsymbol{x}), \ldots, u_m(\boldsymbol{x}))$. Note that when using least squares estimation, we can set $\alpha_0 = 0$ in (18) (because the sum of the residuals are 0) and therefore, here $\boldsymbol{U}$ is an $n \times m$ matrix. Then, as shown in the Appendix, the prediction interval is given by

$$\hat{\mu}^F(\boldsymbol{x}) \pm z_{\alpha/2}\sigma\{1 + \frac{1}{n} + \frac{(f(\boldsymbol{x}) - \bar{f})^2}{S} + \boldsymbol{u}(\boldsymbol{x})'[(\boldsymbol{U}'\boldsymbol{U})^{-1} - \boldsymbol{A}]\boldsymbol{u}(\boldsymbol{x})\}^{1/2}, \tag{20}$$

where $\boldsymbol{A} = (\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}$. Of course, if any of the parameters are truncated to 0, then the corresponding terms should be omitted in the calculation of the prediction variance. Also, noting that $f(\boldsymbol{x})$ enters into the functional adjustment model only through the constant adjustment term, the approximate prediction interval that incorporates the uncertainty in the metamodel $\hat{f}(\boldsymbol{x})$ can be obtained as in (16).

Consider the surface roughness example. Based on the plot of residuals from the constant adjustment model against x and after some trial and error, we choose $u_1(x) = x$ and $u_2(x) = \log(1 + x)$. It is not easy to choose a reasonable $\boldsymbol{R}$ matrix in this particular case. Therefore,

we resort to the frequentist approach. By regressing the residuals against x and $\log(1+x)$ (after centering), we obtain $\tilde{\alpha}_1 = .015$, and $\tilde{\alpha}_2 = -.593$. For both coefficients $|z| > \sqrt{2}$ and therefore, we obtain

$$\hat{\mu}^F(x) - \hat{\mu}^C(x) = .015(x - 43.33) - .593(\log(1+x) - 3.35),$$

where $\hat{\mu}^C(x)$ is given in (17). This model gives a good fit (see Figure 6). The inadequacy measure for the functional adjustment model is

$$MI = \frac{1}{120}\sum_{i=1}^{120}(y_i - \hat{\mu}_i^F)^2 = .215,$$

which is smaller than that of the constant adjustment model (.255) and only slightly larger than $\hat{\sigma}^2 = .183$. Because two more parameters have been estimated, $\sum_{i=1}^{d}(\bar{y}_i - \hat{\mu}_{i.}^F)^2/(\sigma^2/r)$ now follows approximately a χ^2_{d-4} distribution. At $\alpha = .05$, $19/20 \times .183 + .183/120 \times 5.99 = .183 < MI = .215$ and therefore, the functional adjustment model is barely inadequate. This could be because the surface roughness is slightly higher at $x = 20$ and slightly lower at $x = 50$. Although more terms can be added into the functional adjustment model to capture this discrepancy, such an up and down behavior for surface roughness is unexpected. Instead, the data collection scheme should be blamed for this observed discrepancy. Therefore, for all practical purposes the foregoing functional adjustment model can be considered adequate.

As suggested by Kennedy and O'Hagan (2001) and others, we may use a Gaussian process prior for δ. Under the two-stage estimation procedure, this amounts to fitting a simple kriging model with a nugget term and zero mean to the residuals (see, e.g., Joseph 2006b). The simple kriging predictor using a Gaussian correlation function is also plotted in Figure 6. We can see that it gives a slightly better fit ($MI = .197$) than the linear model. However, it exhibits the unrealistic up and down pattern and thus the linear model is a better choice for this particular example. Although the two approaches are closely related (Joseph 2006a), the linear model has the advantage that it can be implemented using the well-established techniques of variable selection and thus, can easily identify a functional adjustment model that makes minimal changes to the engineering model.

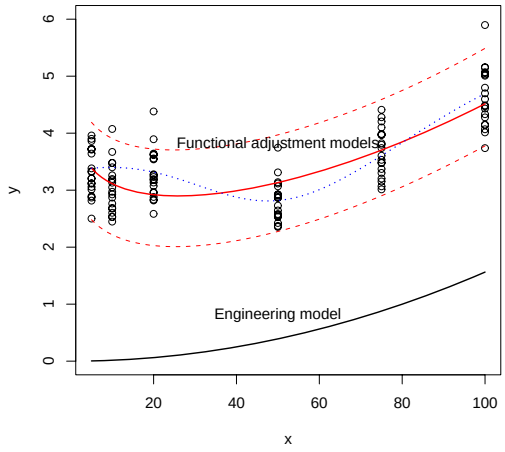

Figure 6: Functional adjustment models in the surface roughness example using a linear model (solid) and a simple kriging model (dotted). The 95% prediction intervals with the linear model are also shown (dashed).

Calibration Parameters

Now we discuss the statistical adjustments when calibration parameters are present in the engineering model. As an example, consider again the problem of predicting surface roughness in micro-cutting. The engineering model given in (1) is purely based on the kinematics of the cutting process. The data clearly show that the actual surface roughness is more than the kinematic surface roughness. Moreover, the kinematic surface roughness decreases as feed decreases, whereas the actual surface roughness seems to increase or remain constant below a certain feed. This phenomenon was also observed by Shaw (1997) who gave a likely explanation. According to Shaw the plastic side flow during cutting is most significant at very small feeds and could be partly responsible for the rise in surface roughness below a certain feed.

Liu and Melkote (2006) did an in-depth study on the surface roughness due to plastic side flow during micro-cutting. As shown in Figure 7, the furrow or ridge formed because of the plastic side flow generated by the secondary cutting edge of the tool will add to the

surface roughness. Thus, the actual surface roughness will be the sum of the kinematic surface roughness and the surface roughness associated with the plastic side flow. Based on the similarity with the analysis of a scratch test performed on metals, Liu and Melkote (2006) proposed the following model to account for the roughness due to plastic side flow: $Y_{plastic} = \eta_0 + \eta_1 \log(\rho(x))$, where $\rho(x)$ is known as the rheological factor and η_0 and η_1 are unknown calibration parameters. Thus, the new engineering model for the surface roughness is given by

$$f(x; \boldsymbol{\eta}) = Y_{kinematic} + Y_{plastic} = \frac{x^2}{8r_n} + \eta_0 + \eta_1 \log(\rho(x)). \tag{21}$$

The rheological factor is a complicated function of x, which can be calculated using a combination of analytical formulas and finite element simulations (see Liu and Melkote 2006 for details).

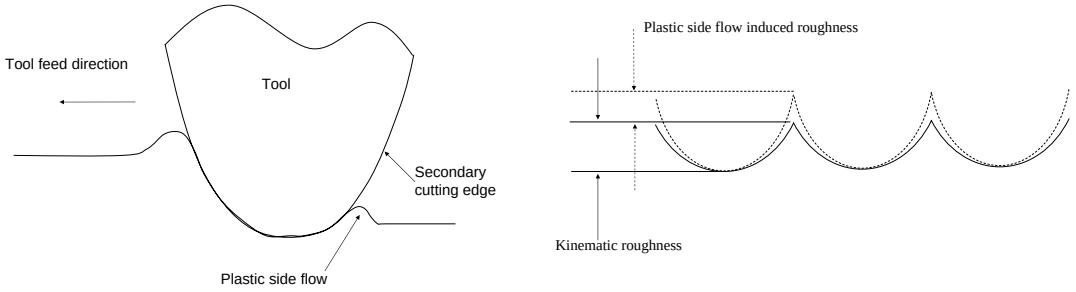

Figure 7: Illustration of surface roughness generated due to plastic side flow.

We start with a least squares estimate of $\boldsymbol{\eta}$, denoted by $\tilde{\boldsymbol{\eta}}$. Usually, engineers will be able to specify an approximate interval for the calibration parameters: $[\boldsymbol{\eta}_L, \boldsymbol{\eta}_U]$, i.e., $\eta_{Li} \leq \eta_i \leq \eta_{Ui}$, for all $i = 1, \ldots, q$. Thus, we use linear/nonlinear regression to get the least squares estimate

$$\tilde{\boldsymbol{\eta}} = \arg \min_{\boldsymbol{\eta} \in [\boldsymbol{\eta}_L, \boldsymbol{\eta}_U]} \sum_{i=1}^{n} [y_i - f_i(\boldsymbol{\eta})]^2, \tag{22}$$

where $f_i(\boldsymbol{\eta}) = f(\boldsymbol{x}_i; \boldsymbol{\eta})$. If y_i's and $f_i(\tilde{\boldsymbol{\eta}})$ are positively correlated and the model inadequacy is large, then we proceed to fit a constant adjustment model. We treat $\boldsymbol{\eta}$ as hyperparameters.

The constant adjustment model is given by

$$Y - f(\boldsymbol{x};\boldsymbol{\eta}) = \beta_0 + \beta_1(f(\boldsymbol{x};\boldsymbol{\eta}) - \bar{f}(\boldsymbol{\eta})) + \epsilon, \tag{23}$$

where $\bar{f}(\boldsymbol{\eta}) = \sum_{i=1}^{n} f_i(\boldsymbol{\eta})/n$. As before $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ and $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Let $\boldsymbol{f}(\boldsymbol{\eta}) = (f_1(\boldsymbol{\eta}), \cdots, f_n(\boldsymbol{\eta}))'$ and define the model matrix $\boldsymbol{F}(\boldsymbol{\eta})$ similarly. Then, the hyperparameters can be estimated by maximizing

$$\begin{aligned} l &= -\frac{1}{2}\log det(\boldsymbol{F}(\boldsymbol{\eta})\boldsymbol{\Sigma}\boldsymbol{F}'(\boldsymbol{\eta}) + \sigma^2\boldsymbol{I}) \\ &\quad -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\eta}))'(\boldsymbol{F}(\boldsymbol{\eta})\boldsymbol{\Sigma}\boldsymbol{F}'(\boldsymbol{\eta}) + \sigma^2\boldsymbol{I})^{-1}(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\eta})), \end{aligned} \tag{24}$$

with respect to τ_0^2, τ_1^2, and $\boldsymbol{\eta}$.

The foregoing optimization can be simplified as follows. First fix $\boldsymbol{\eta}$ and maximize l with respect to τ_0^2 and τ_1^2. We obtain

$$\hat{\tau}_0^2(\boldsymbol{\eta}) = \left(\tilde{\beta}_0^2(\boldsymbol{\eta}) - \sigma^2/n\right)_+ \text{ and } \hat{\tau}_1^2(\boldsymbol{\eta}) = \left(\tilde{\beta}_1^2(\boldsymbol{\eta}) - \sigma^2/S(\boldsymbol{\eta})\right)_+,$$

where $\tilde{\beta}_j(\boldsymbol{\eta})$ is the least squares estimate of β_j and $S(\boldsymbol{\eta}) = \sum_{i=1}^{n}(f_i(\boldsymbol{\eta}) - \bar{f}(\boldsymbol{\eta}))^2$. Substituting in (24), we obtain the profile log-likelihood of $\boldsymbol{\eta}$. It is easy to show that maximizing the profile log-likelihood is equivalent to minimizing

$$\begin{aligned} A(\boldsymbol{\eta}) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}[y_i - f_i(\boldsymbol{\eta})]^2 + \log(n\hat{\tau}_0^2(\boldsymbol{\eta}) + \sigma^2) \\ &+ \log(\hat{\tau}_1{}^2(\boldsymbol{\eta})S(\boldsymbol{\eta}) + \sigma^2) - \frac{n^2\hat{\tau}_0^2(\boldsymbol{\eta})\tilde{\beta}_0^2(\boldsymbol{\eta})}{\sigma^2(n\hat{\tau}_0^2(\boldsymbol{\eta}) + \sigma^2)} - \frac{\hat{\tau}_1^2(\boldsymbol{\eta})S^2(\boldsymbol{\eta})\tilde{\beta}_1^2(\boldsymbol{\eta})}{\sigma^2(\hat{\tau}_1^2(\boldsymbol{\eta})S(\boldsymbol{\eta}) + \sigma^2)}. \end{aligned}$$

After some simplification, we obtain (constant term omitted)

$$\begin{aligned} A(\boldsymbol{\eta}) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}[y_i - f_i(\boldsymbol{\eta})]^2 + \log(1 + (z_0^2(\boldsymbol{\eta}) - 1)_+) \\ &+ \log(1 + (z_1^2(\boldsymbol{\eta}) - 1)_+) - (z_0^2(\boldsymbol{\eta}) - 1)_+ - (z_1^2(\boldsymbol{\eta}) - 1)_+, \end{aligned} \tag{25}$$

where $z_0(\boldsymbol{\eta}) = \sqrt{n}\tilde{\beta}_0(\boldsymbol{\eta})/\sigma$ and $z_1(\boldsymbol{\eta}) = \sqrt{S(\boldsymbol{\eta})}\tilde{\beta}_1(\boldsymbol{\eta})/\sigma$. Now, by minimizing $A(\boldsymbol{\eta})$ under the constraints $\boldsymbol{\eta} \in [\boldsymbol{\eta}_L, \boldsymbol{\eta}_U]$, we can obtain $\hat{\boldsymbol{\eta}}$. Then, the estimates of β_0 and β_1 can be obtained as

$$\hat{\beta}_0 = \left(1 - \frac{1}{z_0^2(\hat{\boldsymbol{\eta}})}\right)_+ \tilde{\beta}_0(\hat{\boldsymbol{\eta}}) \text{ and } \hat{\beta}_1 = \left(1 - \frac{1}{z_1^2(\hat{\boldsymbol{\eta}})}\right)_+ \tilde{\beta}_1(\hat{\boldsymbol{\eta}}). \tag{26}$$

The foregoing empirical Bayes estimation is closely related to the following nonlinear least squares estimation:

$$\min_{\beta_0,\beta_1,\boldsymbol{\eta}} \frac{1}{\sigma^2}\sum_{i=1}^{n}[y_i - f_i(\boldsymbol{\eta}) - \beta_0 - \beta_1(f_i(\boldsymbol{\eta}) - \bar{f}(\boldsymbol{\eta}))]^2. \tag{27}$$

By minimizing with respect to β_0 and β_1, we obtain

$$\begin{aligned} Q(\boldsymbol{\eta}) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}[y_i - f_i(\boldsymbol{\eta}) - \tilde{\beta}_0(\boldsymbol{\eta}) - \tilde{\beta}_1(\boldsymbol{\eta})(f_i(\boldsymbol{\eta}) - \bar{f}(\boldsymbol{\eta}))]^2 \\ &= \frac{1}{\sigma^2}\sum_{i=1}^{n}[y_i - f_i(\boldsymbol{\eta})]^2 - z_0^2(\boldsymbol{\eta}) - z_1^2(\boldsymbol{\eta}). \end{aligned} \tag{28}$$

For large values of $z_0^2(\boldsymbol{\eta})$ and $z_1^2(\boldsymbol{\eta})$, $A(\boldsymbol{\eta})$ behaves like $Q(\boldsymbol{\eta})$ and therefore, minimizing $A(\boldsymbol{\eta})$ is approximately equivalent to minimizing $Q(\boldsymbol{\eta})$. Thus, a frequentist version of the empirical Bayes approach is to fit a nonlinear regression model $y_i = f_i(\boldsymbol{\eta}) + \beta_0 + \beta_1(f_i(\boldsymbol{\eta}) - \bar{f}(\boldsymbol{\eta})) + \epsilon_i$ and truncate β_0 and β_1 to 0 if their $|z|$ statistics are less than $\sqrt{2}$. Note that if $\beta_0 = 0$ and/or $\beta_1 = 0$, then $\boldsymbol{\eta}$ should be reestimated from the reduced model. This is quite intuitive and very easy to implement. The implementation can be easily done using a standard statistical software without requiring the need for additional programming work.

If the model inadequacy for the constant adjustment model is large, then we can proceed to fit a functional adjustment model. As explained in the previous section, this can be done by fitting a linear/nonlinear regression model to the residuals from the constant adjustment model. By virtue of the two-stage estimation, we do not need to estimate $\boldsymbol{\eta}$ again. This is extremely simple compared to the procedures in Kennedy and O'Hagan (2001), Bayarri et al. (2007), and others. One could argue that because $\boldsymbol{\eta}$ is fixed at $\hat{\boldsymbol{\eta}}$ when estimating the functional adjustment model, we are not getting a good estimate of the true value of the calibration parameters. However, since the functional form of the model is altered, the calibration parameters lose their physical meaning and therefore the notion of "true" value becomes questionable. Moreover, as shown by Loeppky et al. (2006), final prediction is not affected much by $\boldsymbol{\eta}$, because the errors in its estimation get compensated through the model inadequacy part $\delta(\boldsymbol{x};\boldsymbol{\alpha})$.

We explain the approach using the new surface roughness model. The rheological values are computed using a combination of analytical formulas and finite element simulations for the six values of feed as $\rho(5) = 521$, $\rho(10) = 508$, $\rho(20) = 490$, $\rho(50) = 463$, $\rho(75) = 460$, and $\rho(100) = 475$. First we use the least squares method to obtain an initial estimate of $\boldsymbol{\eta}$. In this example, this can be easily done by using a simple linear regression for $y - x^2/(8r)$ against $\log(\rho(x))$. We obtain $\tilde{\boldsymbol{\eta}} = (-27.1, 4.86)'$. The model inadequacy measure can be computed as $MI = .209$, which is very small compared to the kinematic surface roughness model ($MI = 9.12$). Thus, the new surface roughness model that includes the effect of plastic side flow gives a much better prediction than the kinematic surface roughness model.

Because the prediction error from the calibrated engineering model is small, in this example, we may not need any statistical adjustments. However, for illustrative purposes, we fit a constant adjustment model. By minimizing $Q(\boldsymbol{\eta})$ in (28), we obtain $\hat{\boldsymbol{\eta}} = (-30.47, 5.41)'$. Then using simple linear regression $y_i = f_i(\hat{\boldsymbol{\eta}}) + \beta_0 + \beta_1(f_i(\hat{\boldsymbol{\eta}}) - \bar{f}(\hat{\boldsymbol{\eta}})) + \epsilon_i$, we obtain $\hat{\beta}_0 = 0$ and $\hat{\beta}_1 = .23$. Thus, the constant adjustment predictor is

$$\hat{\mu}^C(x) \;\; = \;\; f(x; \hat{\boldsymbol{\eta}}) + .23(f(x; \hat{\boldsymbol{\eta}}) - 3.46).$$

The model is plotted in Figure 8, which shows a reasonably good fit to the data ($MI = .199$). We may consider fitting a functional adjustment model to obtain further improvement over the constant adjustment model, however, in this case little would be achieved by doing that.

An Example: Spot Welding Experiment

Consider the spot welding experiment given in Higdon et al. (2004) and Bayarri et al. (2007). There are three factors in this experiment: load (x_1), current (x_2), and gage (x_3), and the objective is to optimize the nugget diameter of the weld (y). The engineering model is quite complex and contains one calibration parameter (η).

The first step is to approximate the engineering model. Because there is no random error in the engineering model simulations, an interpolating model such as kriging is suitable for the model approximation (Sacks et al. 1989). We fit an ordinary kriging model with Gaussian

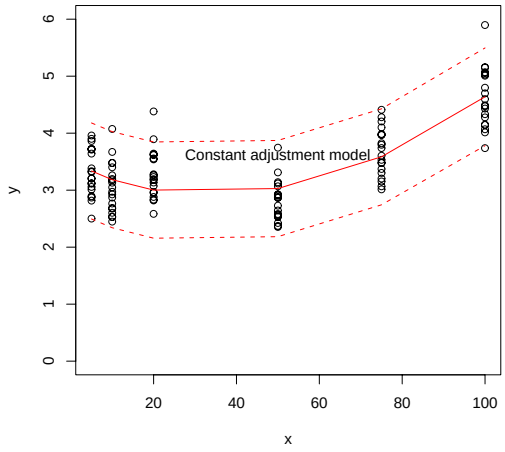

Figure 8: Constant adjustment model fitted using the new model for surface roughness. The 95% prediction intervals computed using (15) are also shown (dashed).

correlation function: $\psi((\boldsymbol{x}_j, \eta_j)', (\boldsymbol{x}_k, \eta_k)') = \exp(-\sum_{i=1}^{3} \theta_i (x_{ij} - x_{ik})^2 - \theta_4(\eta_j - \eta_k)^2)$. Using the 35 simulations in Table 3 of Bayarri et al. (2007), the maximum likelihood estimate of the correlation parameters is obtained as $\hat{\theta}_1 = .19$, $\hat{\theta}_2 = 1.12$, $\hat{\theta}_3 = .13$, and $\hat{\theta}_4 = 2.94$ (all the variables are scaled to $[-1, 1]$.) The ordinary kriging predictor is given by

$$\hat{f}(\boldsymbol{x}; \eta) = 6.17 + \boldsymbol{\psi}(\boldsymbol{x}, \eta)' \boldsymbol{\Psi}^{-1} (\boldsymbol{f} - 6.17\ \mathbf{1}), \tag{29}$$

where $\boldsymbol{\psi}(\boldsymbol{x}, \eta)$ is a 35×1 vector with ith element $\psi((\boldsymbol{x}, \eta)', (\boldsymbol{x}_i, \eta_i)')$, $\boldsymbol{\Psi}$ is the 35×35 correlation matrix, and $\mathbf{1}$ is a vector of 1's.

Physical data are obtained from a 12-run experiment with 10 replicates per run (see Table 4 of Bayarri et al. (2007)). The least squares estimate of the calibration parameter can be obtained from (22) with f replaced with $\hat{f}$. It is known that $\eta_L = .8$ and $\eta_U = 8$. Thus, we obtain $\tilde{\eta} = 2.1$. The model inadequacy $MI = .675$ is much higher than $\hat{\sigma}^2 = s^2 = .20$ (obtained from the replicates) and therefore, we proceed to fit a constant adjustment model. We take the frequentist approach. The constant adjustment predictor is then given by

$$\hat{\mu}^C(\boldsymbol{x}) = \hat{f}(\boldsymbol{x}; 2.08) - .12(\hat{f}(\boldsymbol{x}; 2.08) - 6.12).$$

The $MI = .669$, which shows no significant improvement over the engineering model. Therefore, we fit a functional adjustment model. Consider the main effects and two-factor interactions of the three variables. Thus, after centering, $u_1 = x_1$, $u_2 = x_2 - .03$, $u_3 = x_3$, $u_4 = x_1x_2$, $u_5 = x_1x_3$, and $u_6 = x_2x_3 - .33$. We find that all the effects except u_5 are significant ($|z| > \sqrt{2}$). Thus, we obtain the following functional adjustment model

$$\hat{\mu}^F(\boldsymbol{x}) - \hat{\mu}^C(\boldsymbol{x}) = .12x_1 - .21(x_2 - .03) + .65x_3 + .44x_1x_2 + .40(x_2x_3 - .33). \tag{30}$$

The model inadequacy $MI = .229$ is not significant at 5% level and thus, it is a good model for prediction. The functional adjustment model is plotted in Figure 9 for various combinations of current, load, and gage. Clearly, the predictions are much closer to the data than those from the calibrated engineering model ($\hat{f}(\boldsymbol{x}; 2.1)$). The model from Bayarri et al. (2007) is also plotted in the figure (see Figure 5 in their paper). The prediction performances of the two models near the data points are comparable, but Bayarri et al.'s model completely changes the shape of the engineering model, which is undesirable. Whereas the proposed model approximately preserves the shape of the engineering model and therefore, it is more reliable and clearly better for optimizing the nugget diameter. Moreover, the fitting in Bayarri et al. uses Markov Chain Monte Carlo methods, which are computationally much more intensive compared to the proposed approach.

Conclusions

Engineering models and statistical models have their own pros and cons. It is shown that through careful integration we can develop models that mitigate the drawbacks of both types of models. Such models can help engineers develop improved process control strategies and find better processing conditions.

The statistical adjustments correct the engineering model without trying to understand the mistakes made in modeling the physics of the process. This is good in the sense that we are able to quickly develop models that yield more realistic predictions. The drawback is that we are not able to determine what went wrong in the engineering model. Take for

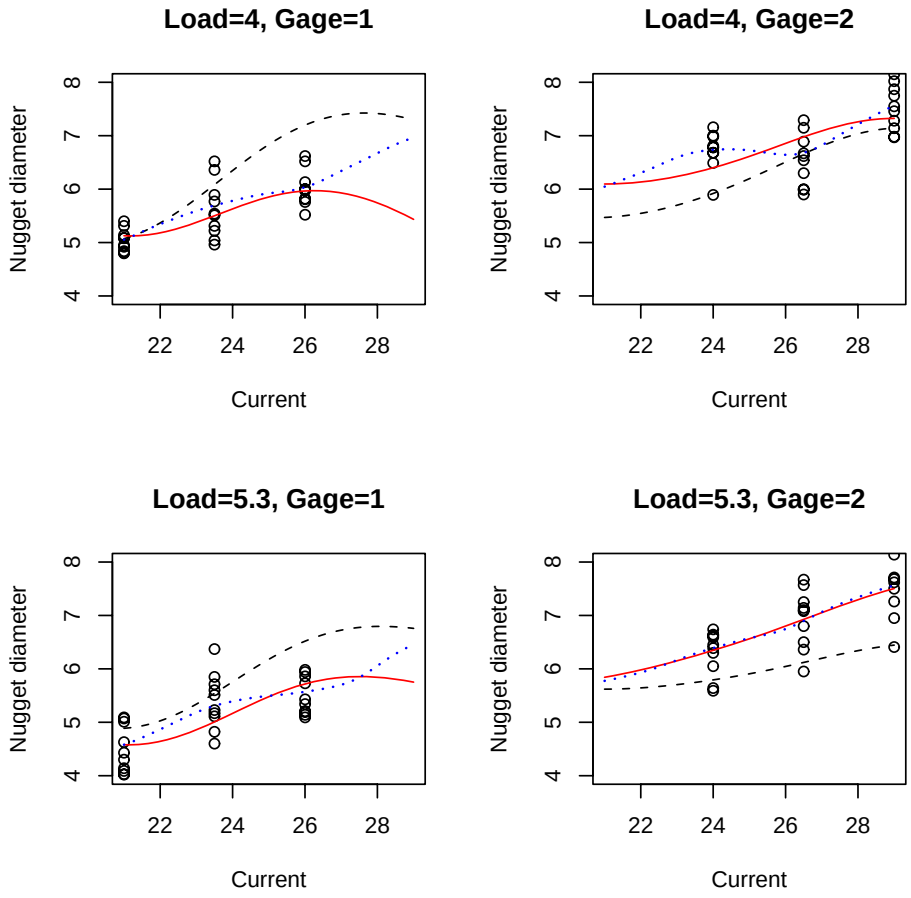

Figure 9: Calibrated engineering model in (29) with $\eta = 2.1$ (dashed), functional adjustment model in (30) (solid), and the model from Bayarri et al. (dotted) in the spot welding example.

example the case of surface roughness prediction in the micro-cutting process. The functional adjustment model based on the kinematic surface roughness model shown in Figure 6 gives excellent prediction, almost as good as the surface roughness model that includes the effect of plastic side flow. However, the surface roughness model that takes into account the effect of plastic side flow should be considered a better model, because we are able to explain the micro-cutting phenomenon in a better way. This of course does not diminish the utility of statistical adjustments to engineering models. In reality, practicing engineers in industries do not have the time to identify the mistakes in the underlying assumptions of engineering

models, develop new theory, and make corrections to the predictions. The engineering-statistical models are a better alternative for them.

Although not directly evident, the engineering-statistical models may help researchers in pin-pointing the mistakes in the assumptions and guide them in developing a better engineering model. We leave this important topic for future research.

Appendix: Proofs

Proof of Equation (9)

Assume that the engineering model is evaluated at $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \ldots, \boldsymbol{x}_N\}$, which contains the locations for the physical experiment. Therefore, $\hat{\beta}_0$, $\hat{\beta}_1$, and S are constants for given $\boldsymbol{f}_c = (f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N))'$. As mentioned before, only the information from $\boldsymbol{f}_c$ is used for constructing the metamodel. Thus, let $\hat{f}(\boldsymbol{x}) = E(f(\boldsymbol{x})|\boldsymbol{f}_c)$ and $v(\boldsymbol{x}) = var(f(\boldsymbol{x})|\boldsymbol{f}_c)$. Then,

$$\begin{aligned} E(\mu(\boldsymbol{x})|\boldsymbol{y}, \boldsymbol{f}_c) &= E\{E[\mu(\boldsymbol{x})|\boldsymbol{y}, \boldsymbol{f}_c, f(\boldsymbol{x})]\} \\ &= E\{f(\boldsymbol{x}) + \hat{\beta}_0 + \hat{\beta}_1(f(\boldsymbol{x}) - \bar{f})|\boldsymbol{y}, \boldsymbol{f}_c\} \\ &= \hat{f}(\boldsymbol{x}) + \hat{\beta}_0 + \hat{\beta}_1(\hat{f}(\boldsymbol{x}) - \bar{f}), \end{aligned}$$

and

$$\begin{aligned} var(\mu(\boldsymbol{x})|\boldsymbol{y}, \boldsymbol{f}_c) &= E\{var[\mu(\boldsymbol{x})|\boldsymbol{y}, \boldsymbol{f}_c, f(\boldsymbol{x})]\} + var\{E[\mu(\boldsymbol{x})|\boldsymbol{y}, \boldsymbol{f}_c, f(\boldsymbol{x})]\} \\ &= E\{\frac{1}{n + \sigma^2/\tau_0^2} + \frac{(f(\boldsymbol{x}) - \bar{f})^2}{S + \sigma^2/\tau_1^2}|\boldsymbol{y}, \boldsymbol{f}_c\} \\ &\quad + var\{f(\boldsymbol{x}) + \hat{\beta}_0 + \hat{\beta}_1(f(\boldsymbol{x}) - \bar{f})|\boldsymbol{y}, \boldsymbol{f}_c\} \\ &= \frac{1}{n + \sigma^2/\tau_0^2} + \frac{(\hat{f}(\boldsymbol{x}) - \bar{f})^2 + v(\boldsymbol{x})}{S + \sigma^2/\tau_1^2} + (1 + \hat{\beta}_1)^2 v(\boldsymbol{x}). \end{aligned}$$

Using a normal approximation to the distribution of $\mu(\boldsymbol{x}) + \epsilon|\{\boldsymbol{y}, \boldsymbol{f}_c\}$, we obtain the prediction interval given in (9).

Proof of Equation (20)

Let $\tilde{\boldsymbol{\beta}} = (\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'(\boldsymbol{y} - \boldsymbol{f})$ and $\tilde{\boldsymbol{\alpha}} = (\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'(\boldsymbol{y} - \boldsymbol{f} - \boldsymbol{F}\tilde{\boldsymbol{\beta}})$ be the least squares estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, respectively. Then, the prediction variance of a new observation is

given by

$$\begin{aligned} V &= var\{f(\boldsymbol{x}) + \boldsymbol{w}(\boldsymbol{x})'\tilde{\boldsymbol{\beta}} + \boldsymbol{u}(\boldsymbol{x})'\tilde{\boldsymbol{\alpha}} + \epsilon\} \\ &= var\{\boldsymbol{w}(\boldsymbol{x})'\tilde{\boldsymbol{\beta}} + \boldsymbol{u}(\boldsymbol{x})'\tilde{\boldsymbol{\alpha}}\} + \sigma^2 \\ &= var\{\boldsymbol{w}(\boldsymbol{x})'(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'(\boldsymbol{y}-\boldsymbol{f}) + \boldsymbol{u}(\boldsymbol{x})'(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'(\boldsymbol{y}-\boldsymbol{f}-\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'(\boldsymbol{y}-\boldsymbol{f}))\} + \sigma^2 \\ &= var\{\boldsymbol{a}(\boldsymbol{x})'(\boldsymbol{y}-\boldsymbol{f})\} + \sigma^2, \end{aligned}$$

where

$$\boldsymbol{a}(\boldsymbol{x})' = \boldsymbol{w}(\boldsymbol{x})'(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}' + \boldsymbol{u}(\boldsymbol{x})'(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'(\boldsymbol{I}-\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}').$$

Therefore,

$$\begin{aligned} V &= \sigma^2\boldsymbol{a}(\boldsymbol{x})'\boldsymbol{a}(\boldsymbol{x}) + \sigma^2 \\ &= \sigma^2\{1 + \boldsymbol{w}(\boldsymbol{x})'(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{u}(\boldsymbol{x})'(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'(\boldsymbol{I}-\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}')\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{u}(\boldsymbol{x})\} \\ &= \sigma^2\{1 + \frac{1}{n} + \frac{(f(\boldsymbol{x})-\bar{f})^2}{S} + \boldsymbol{u}(\boldsymbol{x})'[(\boldsymbol{U}'\boldsymbol{U})^{-1} - \boldsymbol{A}]\boldsymbol{u}(\boldsymbol{x})\}, \end{aligned}$$

where $\boldsymbol{A} = (\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}$. Thus, $f(\boldsymbol{x}) + \boldsymbol{w}(\boldsymbol{x})'\tilde{\boldsymbol{\beta}} + \boldsymbol{u}(\boldsymbol{x})'\tilde{\boldsymbol{\alpha}} + \epsilon$ has a normal distribution with mean $\hat{\mu}^F(\boldsymbol{x})$ and variance V. Therefore, the prediction interval is given by $\hat{\mu}^F(\boldsymbol{x}) \pm z_{\alpha/2}\sqrt{V}$.

Acknowledgments

This research was supported by U. S. National Science Foundation grant CMMI-0654369.

References

Bayarri, M. J.; Berger, J. O.; Paulo, R.; Sacks, J.; Cafeo, J.; Cavendish, J.; Lin, C. H.; Tu, J. (2007). "A Framework for Validation of Computer Models". *Technometrics* 49, pp. 138-154.

Box, G. E. P. and Hunter, W. G. (1962). "A Useful Method for Model-Building". *Technometrics* 4, pp. 301-318.

Breiman, L. (1995). "Better Subset Regression Using the Nonnegative Garrote". *Technometrics* 37, pp. 373-384.

Higdon, D.; Kennedy, M.; Cavendish, J. C.; Cafeo, J. A.; and Ryne, R. D. (2004). "Combining Field Data and Computer Simulations for Calibration and Prediction". *SIAM Journal of Scientific Computing* 26, pp. 448-466.

Joseph, V. R. (2006a). "A Bayesian Approach to the Design and Analysis of Fractionated Experiments". *Technometrics* 48, pp. 219-229.

Joseph, V. R. (2006b). "Limit Kriging". *Technometrics* 48, pp. 458-466.

Joseph, V. R. and Delaney, J. D. (2007). "Functionally Induced Priors for the Analysis of Experiments". *Technometrics* 49, pp. 1-11.

Joseph, V. R. and Delaney, J. D. (2008). "Analysis of Optimization Experiments". *Journal of Quality Technology* 40, pp. 282-298.

Kapoor, S. G.; DeVor, R. E.; Zhu, R.; Gajjela, R.; Parakkal, G.; and Smithey, D. (1998). "Development of Mechanistic Models for the Prediction of Machining Performance: Model Building Methodology". *Machining Science and Technology* 2, pp. 213-238.

Kennedy, M. C. and O'Hagan, A. (2001). "Bayesian Calibration of Computer Models" (with discussion). *Journal of Royal Statistical Society - Series B* 63, pp. 425-464.

Liu, K. and Melkote, S. N. (2006). "Effect of Plastic Side Flow on Surface Roughness in Micro-Turning Process". *International Journal of Machine Tools and Manufacture* 46, pp. 1778-1785.

Loeppky, J. L.; Bingham, D.; and Welch, W. J. (2006). "Computer Model Calibration or Tuning in Practice". *Technical Report* 221, Department of Statistics, University of British Columbia, Canada.

Marusich, T. D. and Ortiz, M. (1995). "Modeling and Simulation of High-Speed Machining". *International Journal for Numerical Methods in Engineering* 38, pp. 3675-3694.

Neter, J.; Kutner, M. H.; Nachtsheim, C. J.; and Wasserman, W. (1996). *Applied Linear Statistical Models*. McGraw-Hill, Boston, MA.

Oxley, P. L. B. (1989). *Mechanics of Machining*. Ellis Harwood Ltd., Chichester.

Qian, P. and Wu, C. F. J. (2008). "Bayesian Hierarchical Modeling for Integrating Low-Accuracy and High-Accuracy Experiments". *Technometrics* 50, pp. 192-204.

Reese, C. S.; Wilson, A. G.; Hamada, M.; Martz, H. F.; and Ryan, K. J. (2004). "Integrated Analysis of Computer and Physical Experiments". *Technometrics* 46, pp. 153-164.

Sacks, J.; Welch, W. J.; Mitchell, T. J.; and Wynn, H. P. (1989). "Design and Analysis of Computer Experiments". *Statistical Science* 4, pp. 409-423.

Santner, T. J.; Williams, B. J.; and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York, NY.

Shaw, M. C. (1997). *Metal Cutting Principles*. Oxford University Press, London.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the LASSO". *Journal of the Royal Statistical Society, Ser. B* 58, pp. 267-288.

Wu, C. F. J. and Hamada, M. (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, Wiley, New York, NY.