Bayesian Modeling and Optimization of Functional Responses Affected by Noise Factors

ENRIQUE DEL CASTILLO

The Pennsylvania State University, University Park, PA 16802, USA

BIANCA M. COLOSIMO

Politecnico di Milano, 20133 Milano, Italy

HUSSAM ALSHRAIDEH

Jordan University of Science and Technology, Irbid 22110, Jordan

Experiments in systems where each run generates a curve, that is, where the response of interest is a set of observed values of a function, are common in engineering. In this paper, we present a Bayesian predictive modeling approach for functional response systems. The goal is to optimize the shape, or profile, of the functional response. A robust parameter design scenario is assumed where there are controllable factors and noise factors that vary randomly according to some distribution. The approach incorporates the uncertainty in the model parameters in the optimization phase, extending earlier approaches by J. Peterson (in the multivariate regression case) to the functional response case based on a hierarchical two-stage mixed-effects model. The method is illustrated with real examples taken from the literature and from simulated data, and practical aspects related to model building and diagnostics of the assumed mixed-effects model are discussed.

Key Words: Profile Responses; Robust Design; Signal-Response Systems.

Introduction

Consider an experiment where the response of interest Y is a curvilinear function of a scalar variable s, a variable that may be controllable or uncontrollable. The shape or "profile" of this curve $Y(s)$, observed at different values of s, determines the performance of the system under study. By assigning some real number describing the system performance to each observed curve, we have the classical definition of a *functional* (see, e.g., Gelfand and Fomin (1963)). In the type of experiments we study in this paper, the shape of the curve is assumed to be modified by manipulating n_c controllable factors $\mathbf{x}_c$ and is also affected by the values of n_n noise factors $\mathbf{x}_n$. Following common convention in robust parameter design (RPD), the noise factors are assumed controllable during an experiment, but during regular use of the process or product, they vary randomly. As in traditional RPD for scalar responses, the goal is to find the optimal values of $\mathbf{x}_c$ that make the process response $Y(s \mid \mathbf{x}_c, \mathbf{x}_n)$ robust, or insensitive, to variations in $\mathbf{x}_n$. Optimality in the case with which

Dr. Castillo is a Distinguished Professor of Engineering and Professor of Statistics in the Department of Industrial and Manufacturing Engineering. His email address is exd13@psu.edu.

Dr. Colosimo is an Associate Professor in the Production Technology Group, Dipartimento di Meccanica. Her email address is biancamaria.colosimo@polimi.it.

Mr. Alshraideh is an Assistant Professor in the Department of Industrial Engineering. His email is: haalshraideh@just.edu.jo

we are concerned means achieving a specified target shape for the response $Y(s)$.

The problem we study relates to what Taguchi called *signal response systems* or *dynamic robust design problems* (see, e.g., Taguchi (1987), Miller (2002), McCaskey and Tsui (1997), Miller and Wu (1996), and Wu and Hamada (2000)), also called *multiple target systems* by Joseph and Wu (2002). In this literature, the variable s is called a *signal factor*, selected by the user, usually varied over very few levels in order to reduce the number of experiments. After a signal-response experiment is conducted, the signal factor is adjusted (optimized) to achieve a given response value $Y(s^*)$ and the RPD problem then consists in minimizing the variability of $Y(s^*)$ after this adjustment is made. Instead, the problem with which we are concerned (RPD for a functional response), while similar to the signal-response system problem (the experimental data has an identical structure), refers to obtaining an ideal relationship between Y and s. Each experimental run generates a complete profile consisting of several points of a sampled function over s. Hereafter, we refer to the values of s as the *locations*, terminology common in spatial statistics. A related area in biostatistics is longitudinal analysis, where a function of time (usually, the response to some therapy on different individuals) is observed at a few points in time after treatment started (Fitzmaurice et al. (2004)). Many authors have also considered functional responses in process monitoring (statistical-process control) under the name "profile monitoring" (e.g., see Kim et al. (2003)). Nair et al. (2002) propose various frequentist approaches for achieving robustness in a functional response system and clarified the relations between signal-response systems and the type of functional response systems studied in the present article.

A Typical Example

Nair et al. (2002) studied the design of an electrical alternator. The functional response of interest is the electric current generated as a function of the revolutions per minute (RPMs) at which the alternator operates (this is the location factor s). For each alternator design, the electric current was measured at each of the following RPM values: $s_1 = 1375, s_2 = 1500, s_3 = 1750, s_4 = 2000, s_5 = 2500, s_6 = 3500$, and $s_7 = 5000$. A designed experiment was run that consisted of $8(= n_c)$ controllable factors in the design of the alternator, all varied at three levels, except x_1, which was only varied at two levels. The experiment

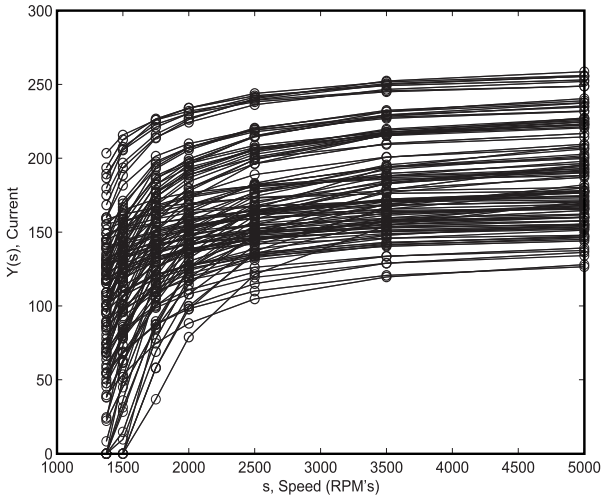


FIGURE 1. Alternator Data, from Nair et al. (2002). The electric-current response for all 108 experiments is shown.

also considered 2 $(= n_n)$ noise factors, x_9 and x_{10}, each at three levels. A total of $N = 108$ electric current vs. RPM profiles were collected, one per run (see Figure 1). Based on these data, the goal is to find settings of the controllable factors that will achieve a specified shape of the electric-current profile $Y(s)$ with high probability in the presence of variability in the noise factors. We return to this example in the fifth section below.

In this paper, we take a Bayesian predictive-modeling approach, initiated by Peterson (2004) in the process optimization field. The proposed method allows one to make predictions about the probability of achieving a desired shape of the functional response (i.e., the method provides a measure of the process "capability" for a given $\mathbf{x}_c$-point) and considers, as part of the probability computations, not only the variability induced by the noise factors but also the uncertainty in the model parameters. The method and auxiliary model-assessment techniques are illustrated with both real examples taken from the literature and with simulated examples. We next provide an overview of the approach and of the remaining contents of this paper.

Overview of the Proposed Modeling and Optimization Approach

Assume the response $Y(s)$ can be observed at several fixed locations $s_1, s_2, \ldots, s_J$. For each experimental run i $(i = 1, \ldots, N)$, where controllable and

noise factors $\mathbf{x}_i = (\mathbf{x}_c, \mathbf{x}_n)_i$ have been tried in a designed experiment, a complete profile or function is observed consisting of J points. Thus, rather than observing a continuous function $Y_i(s \mid \mathrm{x}_i)$, we observe

$$Y_{ij} = g(\mathbf{x}_i; s_j) + \varepsilon_i(s_j), \quad i = 1, \ldots, N, \ j = 1, \ldots, J, \tag{1}$$

where g is some function to be specified/estimated and $\varepsilon_i(s_j)$ is a random error, which can depend on the location s_j. If the N profiles are observed at the same J locations, the errors can be denoted just as ε_{ij}.

An overall summary of the approach presented in this paper is as follows. Suppose a parametric model describes adequately each of the observed profile functions $\mathbf{y}_i = (Y_{i1}, Y_{i,2}, \ldots, Y_{iJ})'$, $i = 1, \ldots, N$. Denote the posterior predictive density for a new profile as $p(\mathbf{y} \mid \mathbf{x}_c, \text{data})$, where "data" is short for all past observed profiles and experimental conditions. If after running the experiment $p(\mathbf{y} \mid \mathbf{x}_c, \text{data})$ is available (either numerically or in closed form), it is possible to perform an optimization where the posterior probability of achieving a function falling within a given specification region $A(s)$ is maximized, that is, we would solve

$$\begin{aligned} \max \quad & p(\mathbf{x}_c) = P(\mathbf{y} \in A(s) \mid \mathbf{x}_c, \text{data}) \\ & \qquad = \int_A p(\mathbf{y} \mid \mathbf{x}_c, \text{data}) d\mathbf{y} \\ \text{subject to:} \quad & \mathbf{x}_c \in R_c, \end{aligned}$$

where R_c is a feasible region for the controllable factors. This is an approach first suggested by Peterson (2004), who proposed it for the multiple-response optimization of linear regression models. The main advantage of this method is that it accounts for the uncertainty in the parameters of the model (via the predictive density, which integrates over the posterior of the model parameters) and considers any correlation present between the responses. It also provides an easy-to-explain probability metric about the "capability" of achieving the specifications $A(s)$.

If noise factors are present in the system and affect the responses (RPD case), a Bayesian approach that provides solutions that are robust with respect to both noise factor variability and regression coefficient uncertainty was proposed by Miro et al. (2004) as an extension to Peterson's method. It consists in solving

$$\begin{aligned} \max \quad & p(\mathbf{x}_c)_{\text{RPD}} \\ & = \int_{\text{all } \mathbf{x}_n} P(\mathbf{y} \in A(s) \mid \mathbf{x}, \text{data}) p(\mathbf{x}_n) d\mathbf{x}_n \qquad (2) \\ \text{subject to:} \quad & \mathbf{x}_c \in R_c. \end{aligned}$$

That is, after obtaining $p(\mathbf{x})$ for fixed $\mathbf{x} = (\mathbf{x}_c, \mathbf{x}_n)$, a second integration is performed over the (assumed-known) density of the noise factors $p(\mathbf{x}_n)$ to obtain the probability of conformance for the RPD problem, $p(\mathbf{x}_c)_{\text{RPD}}$.

If the optimal probability after solving Equation (2) is high, the desired shape of the function $Y(s)$ as defined by region $A(s)$ will be achieved with high probability, despite the noise-factor variability. We would then have a robust solution to the functional RPD problem. For the same amount of computational work, it is possible to instead compute the posterior probability of achieving a given desirability $D(\cdot)$ value of the response $p(D(\mathbf{y}) > D_0 \mid \text{data})$ or compute the posterior probability that one will incur in a quadratic cost $Q(\cdot)$ less than some given value $p(Q(\mathbf{y}) < Q_0 \mid \text{data})$. See Peterson (2004) for more details about these alternative optimization objectives, which may be useful when a specification region is not well defined. In this paper, we focus primarily on computing $p(\mathbf{x}_c)_{\text{RPD}}$ as above.

The sequel of this paper describes the type of mixed-effects model we use for $Y(s)$ and its properties, discusses how one can assess the model fit, explains how to compute the predictive density of a profile $p(\mathbf{y} \mid \mathbf{x}_c, \text{data})$ for the assumed model, and shows how to solve the optimization problems above.

A Hierarchical Model for Functional Data

A useful method for RPD of functional responses consists in modeling the profiles over the locations in a parametric form (stage 1), and then, at a second stage, fitting additional models to the parameter estimates as a function of the controllable and noise factors (stage 2). By changing the controllable factors, the stage 1 parameters can be modified and, therefore, one has control over the profile or shape of the function g in Equation (1). This is the method followed by Wu and Hamada (2000) and Tsui (1999), who refer to it as the *response model* approach. It is also the approach followed by Nair et al. (2002), who consider more general models than typically used for stage 1 in the literature, where usually only simple lines over s are fitted. While their approaches are clearly hierarchical, previous authors have not analyzed the model in a hierarchical way nor used a Bayesian approach as we do here, which allows considering the uncertainty in the model parameters and computing the probability of achieving a desired profile.

The formulation of a two-stage hierarchical model for functional data is as follows. Suppose observations can be modeled as

$$\mathbf{y}_i = \mathbf{S}\boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \sim N_J(\mathbf{0}, \boldsymbol{\Sigma}) \tag{3}$$

and

$$\boldsymbol{\theta}_i = \mathbf{B}\boldsymbol{f}(\mathbf{x}_i) + \mathbf{w}_i, \quad \mathbf{w}_i \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_w) \tag{4}$$

for $i = 1, \ldots, N$, where $\mathbf{y}_i$ is a $J \times 1$ vector containing the observations along profile i, $\mathbf{S}$ is a $J \times p$ matrix of regressors for fitting the stage 1 model (here the regressors are functions of the locations s), $\boldsymbol{\theta}_i$ is a $p \times 1$ vector of stage 1 parameters, $\mathbf{B}$ is a $p \times q$ matrix of parameters—the stage 2 parameters—and $\boldsymbol{f}(\mathbf{x}_i)$ is a $q \times 1$ vector with elements equal to the q terms in the second-stage model evaluated at the experimental settings of run i, namely, $\mathbf{x}_i = (\mathbf{x}_\mathrm{c}, \mathbf{x}_\mathrm{n})_i'$. Thus, each element of the stage 1 parameter vector $\boldsymbol{\theta}_i$ is described by a model containing q parameters. In RPD, a popular model for stage 2 is quadratic in $\mathbf{x}_\mathrm{c}$ and has all control × noise interactions in addition to the noise-factor main effects. It makes sense to also include the noise × noise interactions, which, although have no bearing in the RPD solution, may reduce the probability of conformance if they are significant. Not including them may result in an overestimation of the 'capability' of the process.

Model (3–4) is an instance of the classic Lindley–Smith (1972) hierarchical regression model. By substituting Equation (4) into Equation (3), we get

$$\mathbf{y}_i = \mathbf{S}\mathbf{B}\boldsymbol{f}(\mathbf{x}_i) + \mathbf{S}\mathbf{w}_i + \boldsymbol{\varepsilon}_i \tag{5}$$

or $\mathbf{y}_i \sim N_J(\mathbf{S}\mathbf{B}\boldsymbol{f}(\mathbf{x}_i), \mathbf{S}\boldsymbol{\Sigma}_w\mathbf{S}' + \boldsymbol{\Sigma})$, which is a mixed-effects model. Later on, we will assume that $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_J$, so the within-profile correlation is exclusively modeled through the random effects term $\mathbf{S}\mathbf{w}_i$.

Ware (1985) shows how the mean of $\mathbf{y}_i$ can be written as a linear model. Recognizing that $\mathbf{S}\mathbf{B}\boldsymbol{f}(\mathbf{x}_i)$ is a vector and using the property $\mathrm{vec}(\mathbf{Z_1}\mathbf{Z_2}\mathbf{Z_3}) = (\mathbf{Z_3}' \otimes \mathbf{Z_1})\mathrm{vec}(\mathbf{Z_2})$ for conformable matrices $\mathbf{Z_1}, \mathbf{Z_2}$ and $\mathbf{Z_3}$ (where "vec" is the operator that stacks the columns of a matrix and $\otimes$ is the Kronecker product, see Henderson and Searle (1979)), we arrive at $\mathrm{vec}(\mathbf{S}\mathbf{B}\boldsymbol{f}(\mathbf{x}_i)) = \mathbf{S}\mathbf{B}\boldsymbol{f}(\mathbf{x}_i) = (\boldsymbol{f}(\mathbf{x}_i)' \otimes \mathbf{S})\mathrm{vec}(\mathbf{B})$. We can therefore rewrite Equation (5) as the linear model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{S}\mathbf{w}_i + \boldsymbol{\varepsilon}_i, \tag{6}$$

where we define the $J \times qp$ matrix $\mathbf{X}_i = \boldsymbol{f}(\mathbf{x}_i)' \otimes \mathbf{S}$ and the $qp \times 1$ vector $\boldsymbol{\beta} = \mathrm{vec}(\mathbf{B})$. This is a model widely used in longitudinal analysis (see Laird and Ware (1982) and Fitzmaurice et al. (2004)). The predictive density of model (6) is not available in closed form, and a Gibbs sampling scheme is utilized as described in Appendix A.

Example 1. An Instance of the Mixed Effects Model for Functional Response

Suppose the stage 1 model for $Y(s)$ is linear in s (a "linear profile") and assume the intercept and slope parameters can be, in turn, represented by a main-effects-with-interaction model in two controllable factors. Then we have that $p = 2$, $q = 4$ and the two-stage model (3–4) is, for $i = 1, 2, \ldots, N$,

$$\mathbf{y}_i = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_J \end{pmatrix}_i = \begin{pmatrix} 1 & s_1 \\ 1 & s_2 \\ \vdots & \vdots \\ 1 & s_J \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}_i + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_J \end{pmatrix}_i$$

or $Y_{ij} = \theta_{0i} + \theta_{1i}s_j + \varepsilon_{ij}$ for $j = 1, 2, \ldots, J$, and

$$\begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}_i = \begin{pmatrix} \beta_{00} & \beta_{01} & \beta_{02} & \beta_{012} \\ \beta_{10} & \beta_{11} & \beta_{12} & \beta_{112} \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1x_2 \end{pmatrix}_i + \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}_i$$

or $\theta_{ik} = \beta_{k0} + \beta_{k1}x_{i1} + \beta_{k2}x_{i2} + \beta_{k12}x_{i1}x_{i2} + w_{ik}$, for $k = 0, 1$. To transform this into the linear mixed-effects model of Equation (6), consider first the expected profile for given settings $\mathbf{x}_i$, namely $E[\mathbf{y}_i \mid \mathbf{x}_i] = \boldsymbol{X}_i\boldsymbol{\beta}$, where

$$\boldsymbol{X}_i = \boldsymbol{f}(\mathbf{x}_i)' \otimes \mathbf{S} = (1, x_1, x_2, x_1x_2)_i \otimes \begin{pmatrix} 1 & s_1 \\ 1 & s_2 \\ \vdots & \vdots \\ 1 & s_J \end{pmatrix}$$

$$= (\mathbf{S}, x_{i1}\mathbf{S}, x_{i2}\mathbf{S}, x_{i1}x_{i2}\mathbf{S})$$

(a $J \times pq$ matrix) and, therefore, because $\boldsymbol{\beta} = \mathrm{vec}(\mathbf{B})$ and $\mathbf{B}_{\bullet,j}$ is the jth column of $\mathbf{B}$,

$$\begin{aligned} &E[\mathbf{y}_i|\mathbf{x}_i] \\ &\quad = (\boldsymbol{f}(\mathbf{x}_i)' \otimes \mathbf{S})\mathrm{vec}(\mathbf{B}) \\ &\quad = (\mathbf{S}, x_{i1}\mathbf{S}, x_{i2}\mathbf{S}, x_{i1}x_{i2}\mathbf{S}) \begin{pmatrix} \beta_{00} \\ \beta_{10} \\ \beta_{01} \\ \beta_{11} \\ \beta_{02} \\ \beta_{12} \\ \beta_{012} \\ \beta_{112} \end{pmatrix} \\ &\quad = \mathbf{S}\mathbf{B}_{\bullet,1} + x_{i,1}\mathbf{S}\mathbf{B}_{\bullet,2} + x_{i,2}\mathbf{S}\mathbf{B}_{\bullet,3} + x_{i,1}x_{i,2}\mathbf{S}\mathbf{B}_{\bullet,4} \end{aligned}$$

$$= \begin{pmatrix} \beta_{00} + \beta_{10}s_1 + (\beta_{01} + \beta_{11}s_1)x_{i1} + (\beta_{02} + \beta_{12}s_1)x_{i2} + (\beta_{012} + \beta_{112}s_1)x_{i1}x_{i2} \\ \beta_{00} + \beta_{10}s_2 + (\beta_{01} + \beta_{11}s_2)x_{i1} + (\beta_{02} + \beta_{12}s_2)x_{i2} + (\beta_{012} + \beta_{112}s_2)x_{i1}x_{i2} \\ \vdots \\ \beta_{00} + \beta_{10}s_J + (\beta_{01} + \beta_{11}s_J)x_{i1} + (\beta_{02} + \beta_{12}s_J)x_{i2} + (\beta_{012} + \beta_{112}s_J)x_{i1}x_{i2} \end{pmatrix} \quad (7)$$

or $E[Y_{ij} \mid \mathbf{x}_i] = (\beta_{00}+\beta_{01}x_{i1}+\beta_{02}x_{i2}+\beta_{012}x_{i1}x_{i2})+(\beta_{10}+\beta_{11}x_{i1}+\beta_{12}x_{i2}+\beta_{112}x_{i1}x_{i2})s_j$, for $j = 1, \ldots, J$. Therefore, the intercept and the slope are both controlled by manipulating the factors x_1 and x_2. The complete mixed-effects model is then

$$\begin{aligned} &Y_{ij} \mid \mathbf{x}_i \\ &\quad = (\beta_{00} + \beta_{01}x_{i1} + \beta_{02}x_{i2} + \beta_{012}x_{i1}x_{i2} + w_{i0}) \\ &\qquad + (\beta_{10} + \beta_{11}x_{i1} + \beta_{12}x_{i2} + \beta_{112}x_{i1}x_{i2} + w_{i1})s_j \\ &\qquad + \varepsilon_{ij} \end{aligned}$$

for $j = 1, 2, \ldots, J$. Thus, we see that, in this example, both the intercept and the slope of profile i have random components. □

As discussed by Ware (1985) (see also Fitzmaurice et al. (2004)), model (5) is unnecessarily restricted because the design matrices for the fixed-effects term and for the random-effects term are linked, as they both depend on $\mathbf{S}$. This implies that a complex model for the mean will result necessarily in a complex model for the covariance (induced by the term $\mathbf{S}\mathbf{w}_i$), and such a within-profile-correlation structure may not be what the data show.

If the within-profile correlation is strong and is neglected, the least-squares estimates (on which the Bayesian posteriors we will use are based) will be inefficient and will yield less precise predictions than what one would obtain otherwise. Hence, in this case, $\mathrm{Cov}(\mathbf{y}_i)$ needs to be modeled. To better model the within-profile correlation, we follow a recommendation by Ware (1985) and use instead *different* matrices $\mathbf{S}$ and $\mathbf{S}^*$ in Equation (5), keeping the original $\mathbf{S}$ matrix for the fixed-effects term but selecting $\mathbf{S}^*$ to be equal to the first $p_2 (\leq qp)$ columns of $\boldsymbol{X}_i$ and utilizing, therefore, only p_2 random effects w_k. The model is therefore

$$\mathbf{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \mathbf{S}^*\mathbf{w}_i^* + \boldsymbol{\varepsilon}_i, \quad (8)$$

where we still have that $\boldsymbol{X}_i = \boldsymbol{f}(\mathbf{x}_i)' \otimes \mathbf{S}$ and $\mathbf{w}_i^*$ has p_2 elements only. Note how we can write Equation (8) as $\mathbf{y}_i = [\boldsymbol{X}_i^{(R)}|\boldsymbol{X}_i^{(F)}][\boldsymbol{\beta}^{(R)} \mid \boldsymbol{\beta}^{(F)}]' + \boldsymbol{X}_i^{(R)}\mathbf{w}_i^* + \boldsymbol{\varepsilon}_i = \boldsymbol{X}_i^{(F)}\boldsymbol{\beta}^{(F)} + \boldsymbol{X}_i^{(R)}(\boldsymbol{\beta}^{(R)} + \mathbf{w}_i^*) + \boldsymbol{\varepsilon}_i$, making clear the separation between the fixed (F) and random (R) effects. The following example illustrates a simple instance of covariance matrix that results from using $p_2 = 1$ when $\mathbf{S}$ has an intercept.

Example 1 (Continued)

Suppose in the previous example the mixed-effects model is modified in such a way that $\mathbf{S}^*$ equals to the first column of $\boldsymbol{X}_i = (\mathbf{S}, \mathbf{x}_{i1}\mathbf{S}, \mathbf{x}_{i2}\mathbf{S}, \mathbf{x}_{i1}\mathbf{x}_{i2}\mathbf{S})$ only, i.e., only one ($p_2 = 1$) random effect is used so that $\mathbf{S}^* = \mathbf{1}$ (a $J \times 1$ column vector of ones). Then assuming the errors ε_{ij} are uncorrelated, the induced covariance matrix is

$$\begin{aligned} &\mathrm{Cov}(\mathbf{S}^*\mathbf{w}_i^*) + \sigma^2\mathbf{I}_J \\ &= \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \sigma_w^2(1, 1, \ldots, 1) + \sigma^2\mathbf{I}_J \\ &= \begin{pmatrix} \sigma_w^2 + \sigma^2 & \sigma_w^2 & \cdots & \cdots & \sigma_w^2 \\ & \sigma_w^2 + \sigma^2 & \sigma_w^2 & \cdots & \sigma_w^2 \\ & & \ddots & & \\ & \text{(symmetric)} & & \cdots & \sigma_w^2 + \sigma^2 \end{pmatrix}. \end{aligned}$$

This covariance structure, the simplest besides an an independently and identically distributed (i.i.d.) model, is the so-called "compound symmetric" covariance pattern (Fitzmaurice et al. (2004), p. 77), a pattern common in split-plot experiments. It implies the correlation between the response at any two locations s_k and s_l is constant and equal to $\rho = \mathrm{corr}(Y_{ik}, Y_{il}) = \sigma_w^2/(\sigma_w^2+\sigma^2)$. The random effect makes the intercept random. Choosing more random effects will provide more complex covariance structures than this. A subsequent section illustrates how to determine the number of random effects p_2 and hence, the covariance matrix, that best fits the data using information criteria. □

Alternative Models for Functional Response Experimental Data

An alternative model, not pursued here, is to use a spatial–temporal process to model $Y(\mathbf{x}, s)$. This was first discussed by Fang et al. (2006) and recently studied, with a view toward process optimization, by Hung et al. (2011). These authors, however, fit the model in a frequentist way and therefore the un-

certainty in the model parameters is not considered during the optimization, an issue emphasized by Peterson (2004). Fitting all the parameters of such a model in a full Bayesian way is a rather complex task, and most authors working with these models use a variety of approximations (see, e.g., Qian and Wu (2008)).

A second possibility is to model the stage 1 parameters as functions of the locations s. Santner et al. ((2003), pp. 45–47) describe this approach for the analysis of computer experiments (not necessarily for functional responses) but indicate, again, the difficulties in obtaining a full Bayesian estimation, mainly discussing instead likelihood methods of estimation. A full Bayesian estimation (i.e., estimation of all unknown model parameters via Bayes' theorem, either with closed-form posteriors or with numerically computed ones) of any of these alternative models would, in principle, allow us to solve problem (2). The linear mixed-effects model we adopt can be analyzed in a fully Bayesian manner and this allows one to solve such an optimization problem. The examples in this paper demonstrate that this is a flexible model useful in the analysis and optimization of concrete engineering systems.

A simpler alternative model, feasible when the number of locations J is small and equal for all profiles (a situation common in the "signal response" type of problems alluded in the introduction), is to use a multivariate regression model. A full Bayesian analysis for this model is easy and available in closed form, as briefly explained in Appendix B.

Functional RPD Optimization

The parameters of model (8) are $\boldsymbol{\Theta} = \{\boldsymbol{\beta}, \{\mathbf{w}_i^*\}, \boldsymbol{\Sigma}_w, \sigma^2\}$. Bayesian inference for this model has been studied by some authors (Lange et al. (1992), Chib and Carlin (1999)) and requires Markov chain Monte Carlo (MCMC) sampling because the joint posterior of the parameters is not a known distribution in closed form. Appendix A gives the full conditional distributions of each parameter needed in the Gibbs-sampling algorithm that yields the joint posterior of the model parameters. It also describes the criteria we followed to setup the priors for all parameters.

Finding a Robust Optimal Solution Using the Bayesian Mixed Model

The probability of conformance to a set of specifications, $p(\mathbf{x}_c)_{\mathrm{RPD}}$, is optimized once the predictive density of the response along the profile, $\mathbf{y} \mid \mathbf{x}, \text{data}$, is obtained via the MCMC scheme in Appendix A. As described in that appendix, the Markov chain of the model parameters $\boldsymbol{\Theta}$ is run once until convergence and the second half of the chain is sampled within the optimization whenever a value of $\mathbf{y} \mid \mathbf{x}, \text{data}$ is needed. When noise factors are to be considered, then $\mathbf{x}$ is split into $(\mathbf{x}_c, \mathbf{x}_n)$, where the optimization is done over the $\mathbf{x}_c$ and, each time $\mathbf{y} \mid \mathbf{x}, \text{data}$ is needed, we generate random noise factors according to their distribution and use $\mathbf{y}|(\mathbf{x}_c, \mathbf{x}_n), \text{data}$ instead. This provides numerically the extra level of integration (with respect to $p(\mathbf{x}_n)$) referred to above in the problem of Equation (2). The experimenter should first check if all of the predicted means fall within the specifications. If this is not the case, the optimal probability of conformance will be low. The following example illustrates the RPD optimization for a case where simple models for the mean and covariance structure fit well.

Example 2. Breaking Torque Example

An often analyzed RPD experiment in the literature is the breaking torque RPD experiment by the American Supplier Institute (ASI (1998)). It consists in a crossed array for 8 controllable factors and 2 compounded noise factors in a $N = 72$ run experiment. In each run, four ($J = 4$) values of the breaking torque were collected at locations (line pressures) $s_1 = 0.008$, $s_2 = 0.016$, $s_3 = 0.032$, and $s_4 = 0.064$. The goal is to obtain a breaking torque profile with maximum sensitivity to the line pressure applied. The breaking torque–line pressure function can be fit well with a simple line, $y = \theta_0 + \theta_1 s' + \varepsilon$, where s' is the centered line pressure ($s' = s - \overline{s}$). The ideal shape of this relation is one that has the highest slope θ_1. Thus, the first-stage design matrix is $p \times J = 2 \times 4$,

$$\mathbf{S} = \begin{pmatrix} 1 & -0.022 \\ 1 & -0.014 \\ 1 & 0.002 \\ 1 & 0.034 \end{pmatrix}.$$

Figure 2 shows the 72 observed profiles for the torque–pressure function, together with the lower specifications $L = (5, 10, 20, 40)'$ defined for this problem (the upper specifications were set to numerical "infinity" in this example). We fit the mixed-effects model of Equation (6) with main effects and all noise×control interactions because the design does not allow fitting any control×control interaction. Here we have that $p = 2$ parameters are used in stage 1, namely, the intercept θ_0 and slope θ_1, and $q = 28$ parameters are fitted in each of the two models in stage 2.

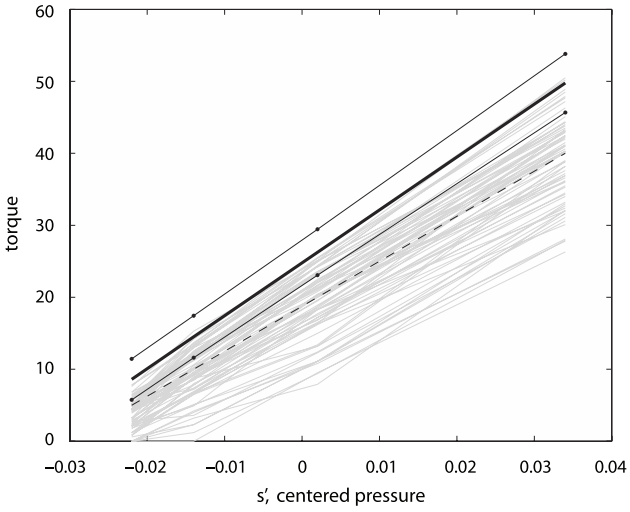


FIGURE 2. Breaking Torque Example (Example 2). Observed 72 torque vs. pressure profiles, lower specification limit (dashed bold line) predicted posterior mean profile (bold line) and 10th and 90th percentiles of the predictive density (thin dark lines) at the optimal solution found.

The best number of random effects p_2 was found to be two, using the information criteria described below. Plots of the residuals discussed later in this paper show no major concern with lack of normality and hence it was concluded this model fits the data well (see Figure 3). The Gibbs-sampling routine described in Appendix A was run for 100,000 iterations, which showed convergence had been achieved (a burn-in period of 50,000 iterations was used; the other examples below required fewer iterations to achieve convergence). This routine was initiated from the priors described in the same Appendix. A flat prior was used on $\boldsymbol{\beta}$, and the other priors were set as in Appendix A, with parameters $\lambda_1 = 2.001$ and $\lambda_2 = 2.5$ (giving $E[\sigma^2] = 0.4$ and $\sqrt{\text{Var}(\sigma^2)} = 12.6$, in the order of magnitude of the standard deviation of the observed profiles; see Figure 2) and $r_1 = 5$ and $r_2 = 2$ (thus anticipating the intercept θ_0 may be more variable numerically than the slope θ_1).

During the optimization, the noise factors were simulated according to an $N_2(\mathbf{0}, \mathbf{I})$ distribution. Maximizing $p(\mathbf{y} \in A \mid \text{data}, \mathbf{x}_c)$ using 1000 random draws for each evaluation of the objective function and the *fmincon* optimization routine from MATLAB (started from 30 different initial points) results in the solution $\mathbf{x}_c^* = (-1, -0.0853, 1, 1, -1, 1, -1, 1)'$, which agrees closely with the solution given by the ASI and also with that given by Lesperance and Park (2003) ($\mathbf{x}_c^* = (-1, \text{NA}, 1, 1, -1, 1, -1, 1)'$), who analyzed these data using generalized linear models (there is some indication that the variance increases with s). These authors did not report an optimal setting for x_2 because it is not clear what setting to use for this factor by looking at its main effect alone. Figure 4 shows how the (posterior) main effect of x_2 over the locations s' has a mixed direction. The computation of these effect plots is discussed in example 3. The solution found yields an estimate $p(\mathbf{y} \in A \mid \text{data}, \mathbf{x}_c) = 0.9348$, which indicates it is likely to meet the lower specification limit for the profile. The optimal predicted mean profile, together

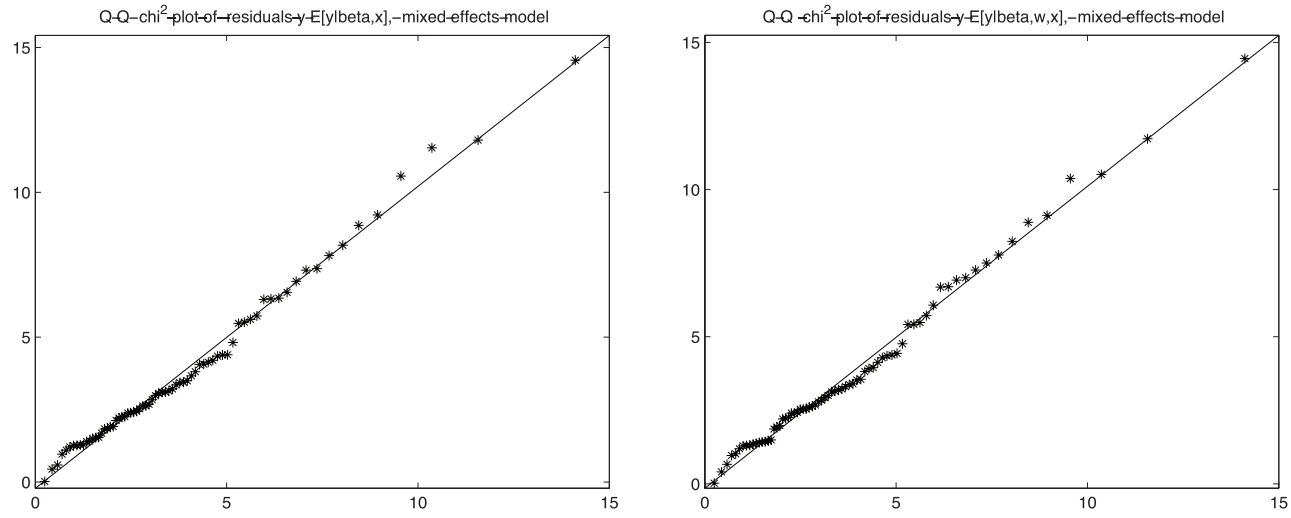
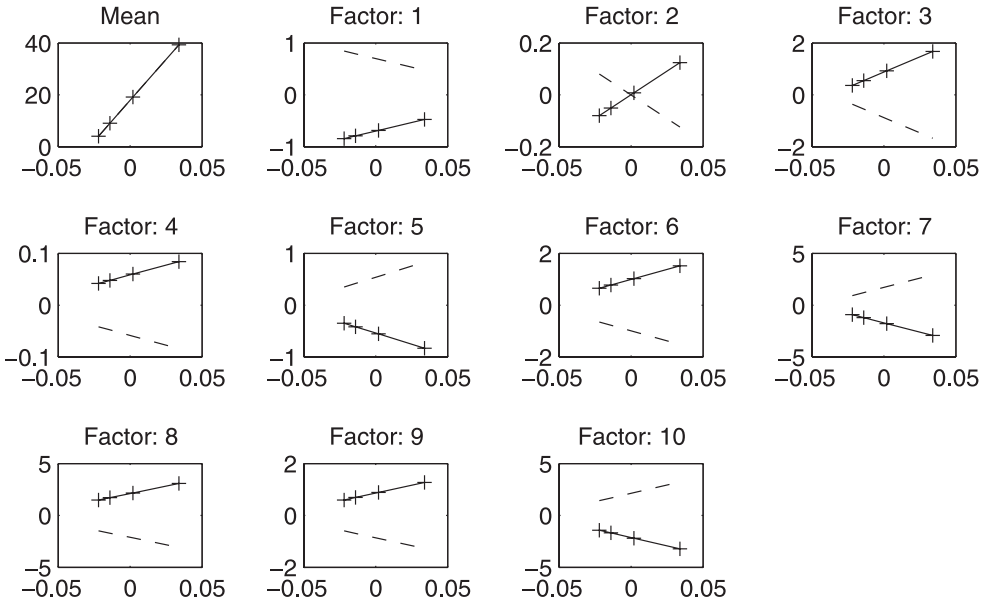


FIGURE 3. Q-Q Plots of the Residuals in the Breaking Torque Example. Left: $\mathbf{r}_1$, right: $\mathbf{r}_2$.



FIGUR 4. Effect Plots Obtained from the Posterior Distributions of $\boldsymbol{\beta}$, Breaking Torque Example. The upper left plot shows the mean profile, and the subsequent plots the additional main effect, over the mean, of changing each factor from its low setting (- - - line) to its high setting (+ - + line). The solution found (-1, ≈ 0, 1, 1, -1, 1, 1, -1, 1) seems reasonable for specifications that call for a 'high' profile (last two factors are the two noise factors, over which no optimization takes place).

with the 10th and 90th percentiles of the predictive density at each value of s', is well above the lower specification, and is shown in Figure 2. □

It is possible that, after conducting the optimization, the probability $p(\mathbf{y} \in A(s)|data, \mathbf{x}_c)$ is too low. This can be due to three possible causes: a) the specifications defining the band $A(s)$ are technologically infeasible for the process, b) the linear mixed-effects model used in the optimization does not fit the data well, or c) there is not enough data to fit the model, and the large parameter uncertainties result in a low probability because the predictive density has large variance. We discuss model diagnostics in the next section to deal with cause b. If the diagnostics indicate a good model fit, then different specifications $A(s)$ should be tried and the optimization redone. In addition, one could try to simulate what would happen if more data were available, following the "preposterior" approach in Peterson (2004). The idea is to replicate the data matrices $[\boldsymbol{f}(\mathbf{x}_1), \boldsymbol{f}(\mathbf{x}_2), \ldots, \boldsymbol{f}(\mathbf{x}_N)]'$, and $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]'$ a certain number of times m, and change N to $m \cdot N$, where m is the number of pseudo-replicates. The MCMC estimation and optimization is then conducted on these combined matrices of real and simulated observations. If the probabilities of conformance are still low, then the cause is probably due to a, i.e., too demanding of specifications because getting more data (running more experiments) is unlikely to reduce the posterior variance and hence to increase the process capability. Another thing to try when probabilities are low is to perform statistical-process monitoring on the profiles at the optimized conditions (Phase I profile control; see Kim et al. (2003)). If all the functional responses fall within the specifications but the maximized $p(\mathbf{x}_c)_{RPD}$ is low, this indicates that the variance model is wrong and needs to be modified.

Model Diagnostics and Covariance-Model Selection

In order for the optimization method proposed here to be relevant in practice, it is important to verify the different modeling assumptions. Many of these assumptions can be verified by computing the residuals. The mixed effects model of Equation (6) allows two types of residuals, namely, the 'population' residuals, $\mathbf{r_1}_i = \mathbf{y}_i - E[\mathbf{y}_i \mid \boldsymbol{X}_i\boldsymbol{\beta}]$, and the 'subject' (or profile) residuals, $\mathbf{r_2}_i = \mathbf{y}_i - E[\mathbf{y}_i \mid \boldsymbol{X}_i\boldsymbol{\beta}, \mathbf{w}_i^*]$. Good modeling practice calls for using 'delete one' cross-validation residuals, where the expectations on the right in the above expressions are estimated conditionally also on all observations except observation i, that is, $E[\mathbf{y}_i \mid \mathbf{Y}_{(i)}]$. Unfortunately, it is very expen-

sive computationally to do this using MCMC techniques because it would require running the MCMC algorithm for each i. However, we can follow an argument in Carlin and Louis ((2000), pp. 204–206), who suggest that the delete-one expectations required in the residual computations can be approximated with the unconditional ones obtained from a single Markov chain, namely, $E[\mathbf{y}_i|\mathbf{Y}_{(i)}] \approx E[\mathbf{y}_i \mid \mathbf{Y}]$. This approximation is good unless the data set is small and profile i is a severe outlier. Hence, residuals diagnostics should be computed after discarding outlier profiles.

One way to identify outlier profiles from the experimental data is to compute $E[\mathbf{w}_i^*|\mathbf{Y}]$ for all i, the posterior mean of the random effects. A large random effect $\mathbf{w}_i^*$ indicates the mean model $\boldsymbol{X}_i\boldsymbol{\beta}$ is not explaining the data well. If $\mathbf{w}_i^* \sim N(\mathbf{0}, \boldsymbol{\Sigma}_w)$, the Mahalanobis distance of the random effects is such that $\mathbf{w}_i^{*'}\Sigma_w\mathbf{w}_i^* \sim \chi^2_{\alpha,p_2}$ (Fitzmaurice et al. (2004)), and this can be used to detect outlier profiles.

The distribution assumptions in model (6) indicate the errors and the random effects should follow vector normal distributions. This can be verified by doing a multivariate test of normality of the residuals and of the random effects. Royston's test (1995) has been shown to be one of the best tests for multivariate normality and was used in what follows.

Example 3. Metal-Injection Data

Govaerts and Noel (2005) report an experiment where there are 25 profiles of the logarithm of the elastic modulus (Y) of a binder used in a metal-injection moulding process as a function of the temperature, which corresponds to the locations s and ranges from 10 to 80 degrees C. The two controllable factors are the ingredients in the binder, namely, xanthan gum concentration (varied from 1 to 5% and coded into x_1,) and chromium nitrate/xanthan concentration ratio (varied from 1:1 to 4:1 and coded into x_2). Each profile had originally 701 locations; this was reduced to 78 equidistant points per profile (by keeping every 9th observation) as the extra locations did not provide additional features for the observed curves and this speeded up all further computations. The objective of the experiment was to obtain a large elastic modulus at lower temperatures while using the lowest chromium concentration, given that it is a pollutant. The specification band was therefore set considering the range of the observed profiles and the conditions above. In their paper, Govaerts and Noel used different methods to fit a model to these profiles, including a nonlinear sigmoidal function that depends on four parameters. We use instead the following linear model based on splines:

$$Y(s) = \theta_0 + \theta_1 \log(s) + \theta_2(\log(s) - 1.4955)_+^2 + \theta_3(\log(s) - 1.7374)_+ + \theta_4(\log(s) - 1.7374)_+^2, \quad (9)$$

where $(x)_+ = \max(0, x)$ denotes the positive part (the locations of the knots were obtained by fitting this model to the trimmed average of $Y(s_j)$ at each location s_j using nonlinear least squares). On a second stage, each θ_i was modeled as a quadratic polynomial in the two factors.

Model (8) was fit using Gibbs sampling from the priors that are described in Appendix A, with parameters $r_1 = 8/5$, $r_k = 1$, $k = 2, \ldots, 5$ (implying more variability is expected a priori on the intercept θ_0 in model 9), and $\lambda_1 = 2.001$, $\lambda_2 = 3.33$ (so that $E[\sigma^2] = 0.3$ and $\sqrt{\mathrm{Var}(\sigma^2)} = 9.48$, in the same order of magnitude of the specifications shown in Figure 8). The flat prior on $\boldsymbol{\beta}$ was set as described in Appendix A.

Figure 5 shows a plot of the $\mathbf{w}_i^{*'}\boldsymbol{\Sigma}_w\mathbf{w}_i^*$ obtained from the means of the posteriors of these parameters. Assuming the random effects $\mathbf{w}_i^*$ are normal, this statistic is distributed as a χ^2_p, where p is the dimension of $\mathbf{w}^*$, equal to the number of θ parameters we use in phase 1 (so $p = 5$ in this case). If any random effect $\mathbf{w}_i^*$ is abnormally large, that is evidence profile i is an outlier. From the figure, profile number 12 is an outlier. Figure 5 shows also a plot of all profiles, highlighting the one found as an outlier. We hence deleted this outlier profile and proceeded to refit the model using MCMC. After deleting the outlier, the residuals and the random effects (using their posterior means) pass Royston's test for multivariate normality, as seen in Table 1. In addition to these residual diagnostics, it can be useful to plot the predicted responses, obtained from the mean of the posterior density of $\mathbf{y}_i$ at each setting $\mathbf{x}_i$. We do not show these for this example, as the differences between the predictions and the observations are too small to the eye in this example. An additional type of plot useful in profile responses are 'effect' plots, which can be obtained as follows. The expected posterior response can be written as a sum of q terms according to Equation (7), with each of the terms equal to $x_j\mathbf{SB}_{\bullet,j}$, where x_j $(j = 1, \ldots, q)$ is the jth term in the stage 2 model (either a constant, a main effect, or higher order effects) and $\mathbf{B}_{\bullet,j}$ is the jth column of matrix $\mathbf{B}$ in Equation (4). Each of these terms is a

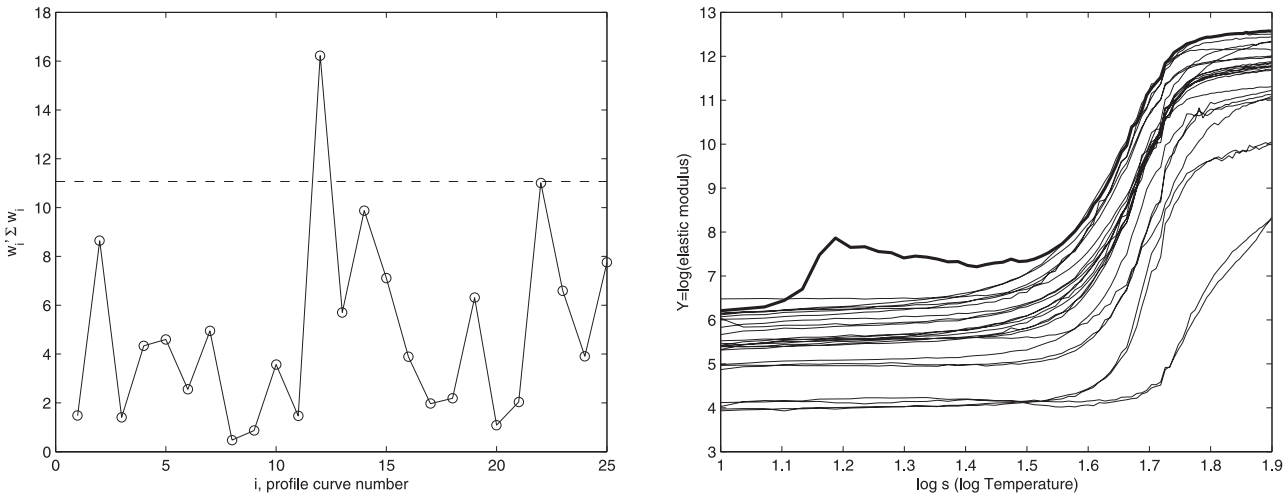


FIGURE 5. Left: χ^2 Statistics $\mathbf{w}_i^{*\prime}\mathbf{S}\mathbf{w}_i^*$ from Metal Injection Process, All 25 Profiles. Horizontal line is at $\chi^2_{0.95,5}$. Right: observed profiles with the outlier in bold, thicker line.

$J \times 1$ vector that measures the contribution of that term to the posterior mean profile $E[\mathbf{y}_i \mid \mathbf{x}_i, \text{data}]$, so the terms involving main effects only can each be plotted as a function of s twice, for $x_j = +1$ and $x_j = -1$, to visualize the effects *along the shape of the profile*, i.e., over the locations. Note how the first term on the right of Equation (7), $(\mathbf{SB}_{\bullet,1})$, is a vector that gives the overall mean (constant) effect because it is not a function of the factors. For the metal-injection example, the constant and main effect plots over the locations are shown in Figure 6. Similar plots were provided by Govaerts and Noel (2005), who used different models than us. To further analyze the effects of each of the controllable factors, we suggest plotting the posteriors of the β stage 2 parameters. For instance, in the metal-injection example, Figure 7 shows the posteriors of the β parameters associated with x_1, namely, $\beta_{01}, \ldots, \beta_{41}$. As can be seen, the largest effects are due to β_{31} (< 0) and β_{41} (> 0). The former is a coefficient that changes parameter θ_3 in the stage 1 term $\theta_3(\log(s) - 1.7374)_+$, and the later is a coefficient that affects parameter θ_4 in the stage 1 term $\theta_4(\log(s) - 1.7374)_+^2$. These terms control the profile shape for large values of s. Because the diagnostics look adequate, we proceed with the optimization. The bounds and the best predicted profile are shown in Figure 8. These were found after 20 runs of the optimizer. The optimal solution found is $x_1^* = 1.21$ and $x_2^* = -1.11$ (in coded units), with a global probability of conformance of $P(\boldsymbol{L} \le \mathbf{y} \le \boldsymbol{U} \mid \text{data}, \mathbf{x}^*) = 0.8338$. This is in accordance to what Govaerts and Noel (2005) reported: high xanthan concentration makes the strength at low temperatures go up, as desired (see also Figure 6). By a happy coincidence, the optimal solution calls for a low use of chromium (x_2), achieving in this way the stated secondary goal of the experiment. Low values of x_2 increase the strength at low temperatures (see Figure 6). The solution found is for the bounds shown in Figure 8 and provides the optimal predicted profile in that same figure.

Covariance-Model Selection

The p_2 random effects determine the within-profile covariance structure. This value can be selected by choosing p_2 such that it provides the lowest Akaike information criterion (AIC) or (more in tune with the Bayesian approach to estimation/optimization undertaken) the lowest Bayesian information criteria (BIC). These statistics are defined as $\text{AIC} = -2\log L + 2p_t$ and $\text{BIC} = -2\log L + p_t \log(N)$ where L is the restricted maximized likelihood function for model (6) and $p_t = p_2(p_2+1)/2 + 1$ is the number of different parameters assuming a common

TABLE 1. Multivariate Normality Test for Residuals and Estimated Random Effects as Computed by Maximum Likelihood, Example 3, $p_2 = 5$. All appear normal at the 5% significance level.

Statistic	p-value for Royston's test
$\mathbf{r}_2$	0.0501
$\mathbf{r}_3$	0.1250
$\mathbf{w}_i^*$	0.0678

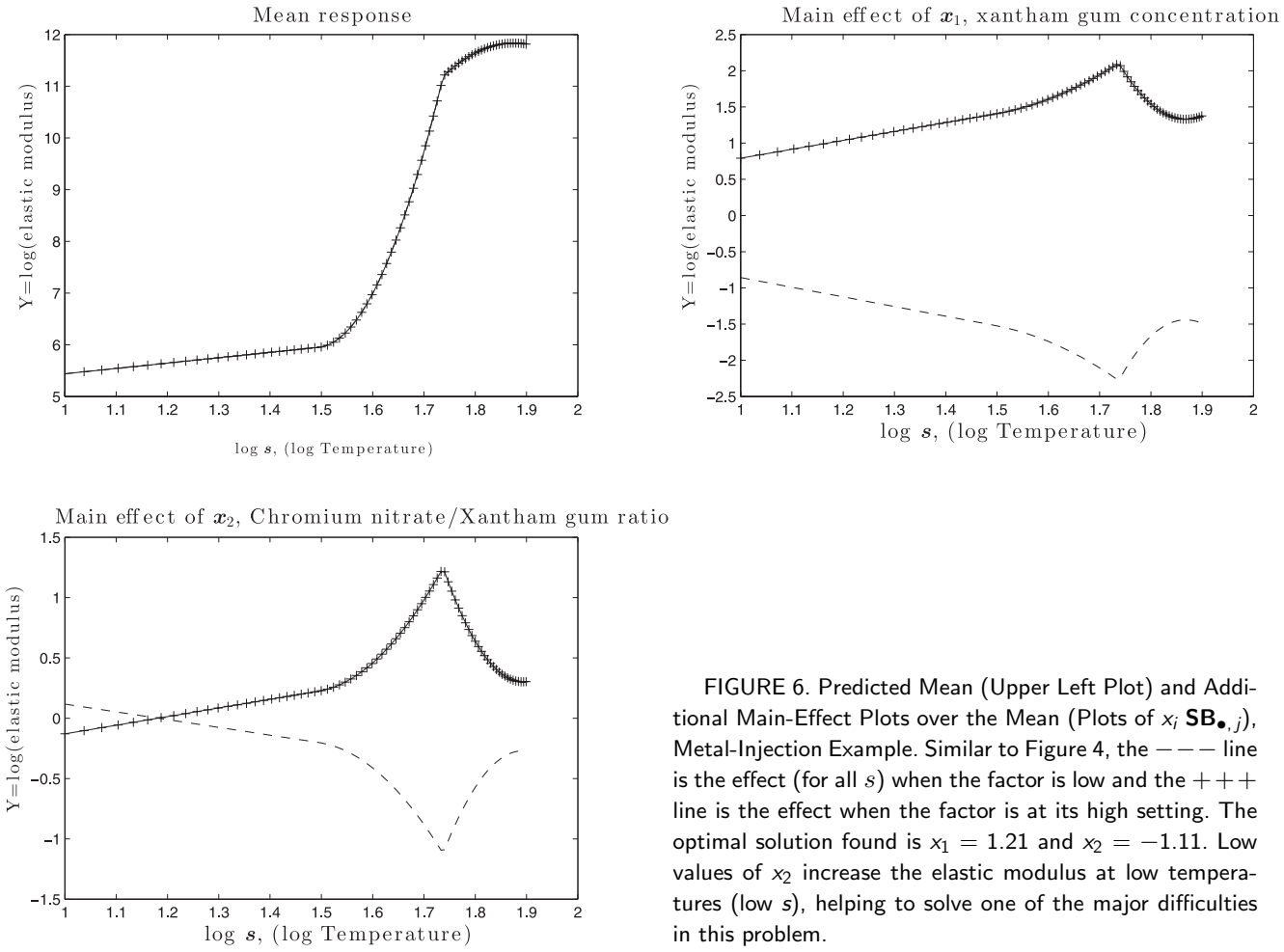


FIGURE 6. Predicted Mean (Upper Left Plot) and Additional Main-Effect Plots over the Mean (Plots of x_i $\mathbf{SB}_{\bullet,j}$), Metal-Injection Example. Similar to Figure 4, the $---$ line is the effect (for all s) when the factor is low and the $+++$ line is the effect when the factor is at its high setting. The optimal solution found is $x_1 = 1.21$ and $x_2 = -1.11$. Low values of x_2 increase the elastic modulus at low temperatures (low s), helping to solve one of the major difficulties in this problem.

model structure for the mean (hence, we are only selecting the best covariance model). See Appendix C for the computer implementation of this method.

Example 3 (Metal Injection, continued)

For the metal-injection data, Table 2 shows the AIC, BIC, and restricted log likehoods for various values of p_2. From the table, both the AIC and the BIC statistics indicate the value $p_2 = 5$ to be best. This was the value used in the preceding analysis of this data.

Sensitivity to the Prior Distributions Used

The solution obtained, $\mathbf{x}_c^*$, and corresponding optimal probabilities of conformance to specifications, $p(\mathbf{x}_c^*)$, depend on the assumed priors. We recommend a sensitivity analysis be conducted on the prior parameters and, in particular, on the parameters of the priors of σ^2 and $\mathbf{\Sigma}_w$. If the solutions $\mathbf{x}_c^*$ do not change much as we change the priors, this is stronger evidence in favor of the solution found. In addition, if for vague priors the probability of conformance is low, a 'preposterior analysis' (Peterson (2004)) can be conducted to determine if collecting more data would improve the models and hence increase $p(\mathbf{x}_c^*)$ or if the problem is simply that the specifications are too demanding. Note, however, that we cannot use *very* vague priors on both $\mathbf{\Sigma}_w$ and σ^2 because the data provides information in a joint manner about these parameters, which would otherwise become unidentifiable (for a discussion on this matter, see Chib and Carlin (1999)).

Example 4 (Metal Injection, continued)

We repeat the model fitting and optimization of the metal-injection problem for various values of the prior parameters (see Appendix A for a description of the type of priors used). In all cases, we fixed $\nu_0 = \max(N/20, p_2)$ and used a flat prior on $\boldsymbol{\beta}$, but tried the scenarios shown in Table 3. The different

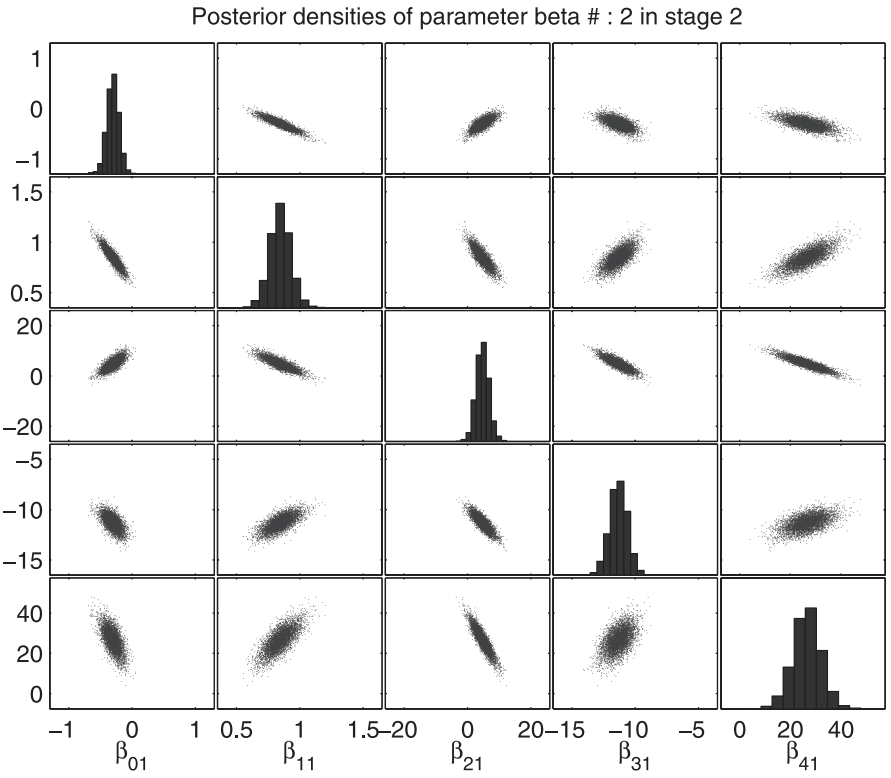


FIGURE 7. Posterior Distributions of the Coefficients β Associated with Factor x_1. Each distribution corresponds to the effect of this factor on the stage 1 parameters θ in the fitted model (9). The largest effects of factor x_1 are on θ_3 and θ_4 (last 2 distributions on the right).

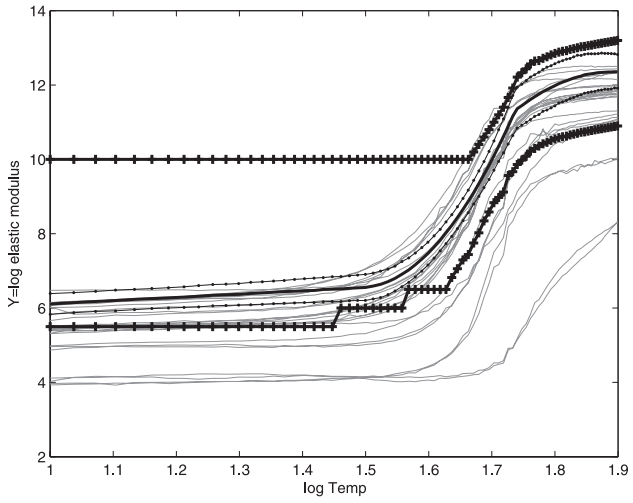


FIGURE 8. Observed Profiles (Grey), Desired Specifications (+), and Predicted Mean Profile at the Optimum Solution Found (Bold Continuous Line), Metal-Injection Example. The two thin, black lines with dots represent the 10th and 90th percentiles of the predictive density of $y_{ij} \mid$ data, $\mathbf{x}^*$.

solutions agree that a high level of xanthan concentration ($x_1 \approx 1$) and a low chromium/xanthan concentration ratio ($x_2 \approx -1$) produces the desired high elastic modulus profile response.

Table 3 indicates relatively low probabilities of conformance when the prior on $\mathbf{\Sigma}_w$ is less informa

TABLE 2. Covariance Model Selection for Metal-Injection Problem, Example 3. From the table, $p_2^* = 5$

p_2	AIC	BIC	$-2\log L$
1	13917	13920	13913
2	13201	13205	13193
3	12779	12787	12765
4	12082	12095	12060
5	**11945**	**11964**	11913
6	11949	11975	11905
7	11959	11993	11901
15	12038	12181	11796

TABLE 3. Sensitivity Analysis on the Parameters of the Priors of σ^2 and $\mathbf{\Sigma}_w$, Metal-Injection Example ($N = 24$). All solutions agree that the optimal solution is close to $x_1 \approx 1$ and $x_2 \approx -1$ regardless of the priors used

Scenario	r_1	r_2	r_3	r_4	r_5	λ_1	λ_2	$E[\sigma^2]$	$\sqrt{\mathrm{Var}(\sigma^2)}$	x_1^*	x_2^*	$p(\mathbf{x}_c^*)$
Baseline	8/5	1	1	1	1	2.001	3.33	0.3	9.48	1.21	−1.10	0.833
Flatter on $\mathbf{\Sigma}_w$	10	5	5	5	5	2.001	3.33	0.3	9.48	0.92	−0.97	0.539
Flatter on σ^2	8/5	1	1	1	1	2.001	0.5	2.0	63.1	1.32	−1.26	0.831
Flatter on $\mathbf{\Sigma}_w, \sigma^2$	10	5	5	5	5	2.001	0.5	2.0	63.1	1.03	−1.04	0.520

tive. This could be due to little information available in prior plus data rather than due to a set of profile-response specifications that are hard to meet, and therefore we proceeded to conduct a preposterior analysis (Peterson (2004)). We increased the amount of data by simply replicating the original data a certain number of times, and repeated the analysis as if the new data were real. From Table 4, $p(\mathbf{x}_c^*)$ increases as the data available increases, and this is an indication that, if vague priors on the variances are necessary due to lack of information, running more experiments than the original 24 is worth the effort. The experimenter could also repeat the sensitivity analysis on the priors with the simulated data to determine if the robustness to the priors could be also enhanced with more data. If experimental resources for more data collection are limited, more data could instead be collected in an observational manner at $\mathbf{x}_c^*$ (and elsewhere, if possible), during the 'Phase I' monitoring stage of the process being developed.

TABLE 4. Preposterior Analysis, Metal-Injection Example. Solutions obtained after replicating the initial data set of 24 observations 2, 3, and 5 times. 'Flatter priors' on $\mathbf{\Sigma}_w$ and σ^2 equal to those used in Table 3 were used. The notable increase in probabilities indicates it is worth collecting more data and reoptimizing

N	Replicates	x_1^*	x_2^*	$p(\mathbf{x}_c^*)$
48	2	1.07	−1.18	0.7598
	2	1.21	−1.17	0.7314
	2	1.09	−1.06	0.7850
72	3	1.10	−1.16	0.7880
	3	1.12	−1.23	0.8500
	3	1.30	−1.43	0.8876
120	5	1.45	−1.38	0.9218
	5	1.44	−1.30	0.9228
	5	1.45	−1.27	0.9172

Robustness to the Normality Assumption

The mixed-effects model is based on an assumption of multivariate normality. Diagnostics to detect lack of normality were described in a previous section. The predicted probabilities of conformance to the specifications $p(\mathbf{x}_c) = p(\mathbf{y} \in A(s) \mid \mathbf{x}_c, data)$ evidently depend on the assumed multivariate normal probability model. If the actual distribution of the errors or of the random effects is a severely skewed, nonnormal distribution, the probabilities and resulting optimal solutions obtained will be considerably biased. In the less severe case of heavy tailed but symmetric error distributions, the behavior of the method is better. To assess the behavior of the approach, a simulation was conducted based on the alternator design data from Nair et al. (2002), discussed in the introduction. Recall there are eight controllable factors and two noise factors and the electric-current profiles were observed over seven values of RPMs. The following stage 1 model was fit to the original alternator data and later on used to generate new profiles in our simulations:

$$Y(s) = \theta_0 + \theta_1 \log(s) + \theta_2(\log(s) - 3.25911)_+ + \theta_3(\log(s) - 3.25911)_+^2 + \varepsilon, \quad (10)$$

where each θ_k depended on the 10 factors via the stage 2 model (4). The experimental design was a Taguchi L_{18} orthogonal array replicated three times ($N = 108$ runs in total). This allows estimating all main effects and all the control × noise interactions, but the only estimable control × control interaction is $x_1 \times x_2$. A multivariate standard t_ν distribution with ν degrees of freedom was used to generate the random effects $\mathbf{w}_i^*$ and the random errors $\boldsymbol{\varepsilon}_i$ in the mixed-effects model. Three ($= p_2$) independent random effects were simulated each with variance $\sigma_w^2 = 100$ and the error variance was set at $\sigma^2 = 10$. The value of p_2 was selected by looking at the AIC and BIC statistics (as described earlier). The degrees of freedom ν were varied from 100 (prac-

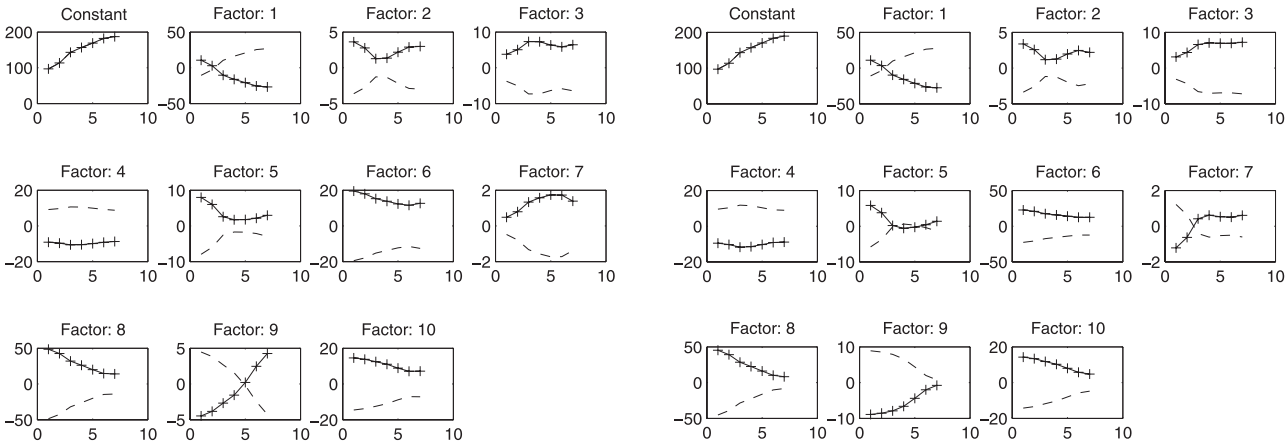


FIGURE 9. Main Effects of the 10 Factors Plotted over the Locations, Alternator Problem. Left: $\nu = 100$ (approx. normal), right: $\nu = 1$ (Cauchy, very fat tailed).

tically a multivariate normal) to the very fat-tailed Cauchy distribution ($\nu = 1$). This fat-tailed distribution can model adequately the presence of severe outliers in the data. The MCMC estimation algorithm was applied for 5000 iterations using the priors in Appendix A with parameters $r_1 = 40$, $r_2 = 16$, and $r_3 = 4$ (anticipating larger variance in θ_0 than in θ_1, etc.) and $\lambda_1 = 4$ and $\lambda_2 = 0.03$, so that $E[\sigma^2] = 11.1$ and $\sqrt{\text{Var}(\sigma^2)} = 7.8$, which is in the order of magnitude of the specifications used in this illustration.

As can be seen from Figure 9, the main effects do not appear to be affected much by the different distributions, and this is reflected in optimal solutions that were not very different (Table 5). However, the probabilities of conformance to the specifications (set at the 70% to 95% percentiles of the simulated profile data) change considerably, with smaller probabilities when the data has fat tails.

Table 6 shows the p-values of Royston's test of multivariate normality for the residuals $\mathbf{r}_1$, $\mathbf{r}_2$ and the random effects $\mathbf{w}^*$. Also shown are the posterior estimate of σ^2 (true value was 10). The MCMC Bayesian estimates started from relative noninformative priors (see Appendix) resulted in unbiased es-

TABLE 5. Optimal Solutions for Simulated Alternator Data for Two Error Distributions. Table shows average, standard deviation, min and max computed over 10 MCMC simulations each 10 K long. The optimization was performed over 30 starting points

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	$p(\mathbf{y} \mid \mathbf{x}_c, \text{data})$
dof $= \nu = 1$ (Cauchy)									
Average	−0.90	0.66	−0.69	0.48	0.10	0.53	−0.35	1.00	0.46
Std. dev	0.16	0.70	0.38	0.61	0.47	0.44	0.69	0.00	0.07
Min	−1.00	−1.00	−1.00	−0.87	−0.76	−0.35	−1.00	1.00	0.35
Max	−0.52	1.00	0.16	1.00	0.79	1.00	0.77	1.00	0.54
dof $= \nu = 100$ (Approx. normal)									
Average	−0.93	0.76	−0.05	0.53	−0.10	0.42	−0.66	1.00	0.58
Std. dev.	0.10	0.62	0.62	0.56	0.78	0.37	0.58	0.00	0.04
Min	−1.00	−1.00	−0.94	−0.65	−0.99	−0.17	−1.00	1.00	0.47
Max	−0.75	1.00	0.94	1.00	0.92	0.83	0.52	1.00	0.61

TABLE 6. *p*-Values for Royston's Normality Tests for Two Error Distributions, Simulated Alternator Data. Table shows average, standard deviation, min and max computed over 10 MCMC simulations each 10 K long. All tests detected lack of normality for $\nu = 1$, whereas only one (false) detection was observed for $\nu = 100$. Also shown are the posterior mean of σ^2 and its MLE estimator (true $\sigma^2 = 10$)

	p for $H_0 : \mathbf{r}_1 \sim N$	p for $H_0 : \mathbf{r}_2 \sim N$	p for $H_0 : \mathbf{w}^* \sim N$	$E[\sigma^2 \mid \text{data}]$	$\widehat{\sigma}^2_{\text{MLE}}$
			dof $= \nu = 1$ (Cauchy)		
Average	0.00	0.00	0.00	10.18	10.52
Std. dev	0.00	0.00	0.00	2.13	2.31
Min	0.00	0.00	0.00	8.03	8.10
Max	0.00	0.00	0.01	15.50	16.27
			dof $= \nu = 100$ (Approx. normal)		
Average	0.17	0.76	0.13	9.91	10.26
Std. dev.	0.10	0.30	0.13	0.39	0.45
Min	0.11	0.01	0.03	9.72	10.06
Max	0.43	0.93	0.49	10.91	11.41

timates of σ^2 regardless of the nonnormality of the errors. The results indicate some degree of robustness of the methods to nonnormal but symmetric errors in the data. If an experimenter is faced with a heavy-tailed distribution, a transformation to normality as suggested by Trimm ((2002), pp. 118–132) could be applied. Congdon ((2006), p. 171) shows how to use gamma weights for Bayesian regression models with t-distributed errors. See also Rajagopal et al. (2005) and Peterson et al. (2009) for Bayesian regression model examples with nonnormal errors.

Conclusions and Further Work

In this paper, a Bayesian approach for the optimization of functional response systems obtained in a designed experiment was presented. The goal is to achieve a desired shape of the functional response that is insensitive to variation in the noise factors. The assumed hierarchical mixed-effects model allows one to control the shape of the function via certain parameters θ_k that are modifiable depending on the values of the controllable factors. We discussed the case when all profiles are observed over the same number of locations (J), but this is an assumption easy to modify; in fact, the Chib–Carlin MCMC estimation routine allows having different numbers of observations J_i over each profile i, $i = 1, \ldots, N$.

An alternative modeling approach for the case where the number of locations is large, not discussed in this paper, is to use a spatial–temporal process model $Y(\mathbf{x}, s)$ where the 'time' coordinate consists of the locations s and the 'spatial' coordinates are made up of the controllable factors. This was suggested by Fang et al. (2006). A spatio–temporal covariance could then be specified to model the linear association between profile responses $\mathbf{y}_i$ and $\mathbf{y}_j$ obtained under experimental conditions $\mathbf{x}_i$ and $\mathbf{x}_j$. The covariance matrix should account also for within-profile correlation, i.e., 'temporal' correlation, where the measure of linear association between $Y(s_i)$ and $Y(s_{i'})$ is modeled. It is possible, in principle, to specify a full Bayesian approach to such a spatio-temporal model, but performing the MCMC estimation of all parameters is a difficult task.

Acknowledgments

We thank Winson Taam for providing the alternator data from their paper (Nair et al. (2002)) and Bernadette Govaerts for providing the metal-injection moulding data from their paper (Govaerts and Noel (2005)). Thanks are due also to Ying Hung and Roshan Joseph for making available their unpublished paper (Hung et al. (2011)), to the editor, and to two anonymous referees for their many comments. This work was funded by the National Science Foundation grant CMI-0825786 and by the Italian Ministry of Education, University and Research (MIUR) grant FIRB RBIP069S2T 005.

Appendix A
Gibbs Sampling for the Linear Mixed-Effects Model

Lange et al. (1992) give the full conditionals for the parameters $(\boldsymbol{\beta}, \{\mathbf{w}_i^*\}, \{\sigma_i^2\}, \boldsymbol{\Sigma}_w^*)$ in the linear mixed model (8), where the σ_i^2's model different variances among the observed profiles. Chib and Carlin (1999) considered the same case as we do, where $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_J$, showed how the Lange et al. procedure suffers from slow convergence, and proposed two alternative algorithms: a pure Gibbs sampling approach, and an algorithm which has a Metropolis step (a rejection sampling step). Because the convergence properties of these two algorithms appear to be about the same, in particular for the $\boldsymbol{\beta}$ parameters, we chose the first algorithm ('algorithm' 2 in their paper) because it does not require a Metropolis step. The Chib–Carlin paper contains some hard-to-detect errors in the full conditionals of $\boldsymbol{\beta}$; the correct formulation is shown below. We also provide the full conditionals of the other parameters because they were not explicitly given by these authors.

Specification of Prior Distributions

In all examples in this paper, the priors used were

- $\boldsymbol{\beta} \sim N_{pq}(\boldsymbol{\beta}_0, \mathbf{B}_0)$, with $\boldsymbol{\beta}_0 = \mathbf{0}$ and $\mathbf{B}_0 = 10000\mathbf{I}$ (noninformative for $\boldsymbol{\beta}$).
- $\sigma^2 \sim IG(\lambda_1, \lambda_2)$ (inverse-gamma distribution). In our parametrization of the IG, this implies that $E[\sigma^2] = 1/(\lambda_2(\lambda_1 - 1))$ and $\mathrm{Var}(\sigma^2) = 1/(\lambda_2^2(\lambda_1 - 1)(\lambda_1 - 2))$ for $\lambda_1 > 2$. Parameters λ_1 and λ_2 can be chosen such that $\sqrt{E[\sigma^2]} \approx \min(U(s) - L(s))$, that is, select the prior standard deviation in accordance with the specifications for the response. The specifications for a process are, in a sense, prior information, because an engineer would design the specifications of a system with some prior knowledge of what is technologically possible. If the process-error standard deviation is much larger than the specification's band "width", it makes no sense to pretend to optimize the functional response and make it fit inside such a band. Alternatively, these parameters can be chosen by making $\sqrt{E[\sigma^2]}$ in the order of the observed variability in the plot of the profiles $Y(s)$, but this would not be a pure Bayesian approach, of course.
- $\boldsymbol{\Sigma}_w^{-1} \sim$ Wishart$((\nu_0 \mathbf{R}_0)^{-1}, \nu_0)$, so in our parametrization, $E(\boldsymbol{\Sigma}_w) \approx \mathbf{R}_0$. Carlin and Louis ((2000), p. 377) suggest setting $\mathbf{R}_0$ such that

  $$\mathbf{R}_0 = \mathrm{diag}((r_1/8)^2, (r_2/8)^2, \ldots, (r_p/8)^2),$$

  where r_k is the total range of anticipated plausible values for parameter $\theta_{k-1}, k = 1, \ldots, p$ in the stage 1 model (Equation (3)). Because $E(\boldsymbol{\Sigma}_w) \approx \mathbf{R}_0$, this gives approximately a ± 2 standard deviation range for θ_{k-1} that is one quarter of the range of its possible values (because $\pm 2\sigma_{\theta_{k-1}} = r_k/4$). These authors also suggest setting $\nu_0 = N/20$ because ν_0 can be understood as a measure of how much weight the prior has and this expression relates such weight to a percentage of the experiments observed. We adopted this, but because $\nu_0 \geq p$ for a proper prior, we set $\nu_0 = \max(p, N/20)$. An alternative prior formulation, not used in this paper, is given by Gelman and Hill ((2007), pp. 284–287). Their approach provides marginal U$[-1, 1]$ priors on the correlation parameters in $\boldsymbol{\Sigma}_w$.

The different values of λ_1, λ_2 and r_k parameters used in this paper are given in each of the corresponding examples.

Gibbs Sampling Scheme

Let $\mathbf{Y}$ be an $N \times J$ matrix containing all the observed functional observations at each of the N experimental runs. The Gibbs sampling scheme is

1. Sample $\boldsymbol{\beta}$ from $\boldsymbol{\beta} \mid \mathbf{Y}, \sigma^2, \boldsymbol{\Sigma}_w$:

$$N\Bigg(\left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{V}_i^{-1} \boldsymbol{X}_i + \mathbf{B}_0^{-1} \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{V}_i^{-1} \mathbf{y}_i + \mathbf{B}_0^{-1} \boldsymbol{\beta}_0 \right), \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{V}_i^{-1} \boldsymbol{X}_i + \mathbf{B}_0^{-1} \right)^{-1} \Bigg),$$

   where $\boldsymbol{V}_i = \mathbf{S}_i^* \boldsymbol{\Sigma}_w \mathbf{S}_i^{*'} + \sigma^2 \mathbf{I}$, $\boldsymbol{X}_i = \mathbf{x}_i^{(m)'} \otimes \mathbf{S}$, and $\mathbf{S}_i^*$ equals the first p_2 columns of $\boldsymbol{X}_i$.

2. Sample the random effects $\mathbf{w}_i^*$, $i = 1, \ldots, N$, from $\{\mathbf{w}_i^*\} \mid \mathbf{Y}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\Sigma}_w$:

$$N\Bigg(\left(\frac{\mathbf{S}_i^{*'} \mathbf{S}_i^{*'}}{\sigma^2} + \boldsymbol{\Sigma}_w^{-1} \right)^{-1} \frac{\mathbf{S}_i^{*'} \boldsymbol{R}_i^{(w)}}{\sigma^2}, \left(\frac{\mathbf{S}_i^{*'} \mathbf{S}_i^{*'}}{\sigma^2} + \boldsymbol{\Sigma}_w^{-1} \right)^{-1} \Bigg),$$

where $\boldsymbol{R}_i^{(w)} = \mathbf{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$. Note how the mean is the empirical Bayes estimate of $\mathbf{w}_i^*$.

3. Sample $\boldsymbol{\Sigma}_w^{-1}$ from $\boldsymbol{\Sigma}_w^{-1}|\{\mathbf{w}_i^*\}$:

$$\text{Wishart}\left(\left(\sum_{i=1}^{N}\mathbf{w}_i^*\mathbf{w}_i^{*'} + \nu_0\mathbf{R}_0\right)^{-1}, N+\nu_0\right).$$

4. Sample σ^{-2} from $\sigma^{-2} \mid \mathbf{Y}, \boldsymbol{\beta}, \{\mathbf{w}_i^*\}$:

$$\text{Gamma}\left(\lambda_1 + \frac{JN}{2}, \left(\frac{1}{\lambda_2} + \frac{1}{2}\left(\mathbf{R}^{(\sigma^2)'}\mathbf{R}^{(\sigma^2)}\right)\right)^{-1}\right),$$

where $\mathbf{R}^{(\sigma^2)} = \mathcal{Y} - \mathcal{X}\boldsymbol{\beta} - \text{vec}(\mathbf{S}_i^*\mathbf{w}_1^*, \ldots, \mathbf{S}\mathbf{S}_i^*\mathbf{w}_N^*)$ (a $NJ \times 1$ vector).

The MCMC sampling of the mixed-effects model parameters needs not be conducted within the optimization routine necessary to solve Equation (2); otherwise, this would imply a tremendous computational burden. The reason for this is given by the model written as in Equation (6). Given realizations of the posterior of $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \{\mathbf{w}_i^*\}, \boldsymbol{\Sigma}_w, \sigma^2)$, $p(\boldsymbol{\Theta} \mid \text{data})$, we can simulate draws of the posterior predictive density by composition (Gelman et al. (2004)) as follows:

$$\begin{aligned} p(\mathbf{y} \mid \text{data}, \mathbf{x}) &= \int p(\mathbf{y}, \boldsymbol{\Theta} \mid \text{data}, \mathbf{x}) d\boldsymbol{\Theta} \\ &= \int p(\mathbf{y} \mid \text{data}, \boldsymbol{\Theta}, \mathbf{x}) p(\boldsymbol{\Theta} \mid \text{data}) d\boldsymbol{\Theta}. \end{aligned}$$

Thus, we conduct a simulation of the Gibbs sampling chain until convergence, approximating in this way $p(\boldsymbol{\Theta} \mid \text{data})$ and simply sample from it. We then substitute the sampled values into the marginal likelihood $p(\mathbf{y} \mid \text{data}, \boldsymbol{\Theta}, \mathbf{x})$ whenever a $\mathbf{y} \mid \text{data}, \mathbf{x}$ vector is needed in the optimization routine. Hence, the MCMC computations are only run once, before performing the optimization. This represents a common random-numbers strategy used in the optimization routine for the posterior parameters, similar to that recommended for the noise factors. The convergence of the MCMC iterations was verified by time plots and computation of the autocorrelation function of all parameters. For instance, for the metal-injection data, the lag 1 autocorrelation after 10,000 iterations was 0.039 for σ^2 and varied between -0.04 and 0.03 for all of the $p \times q$ parameters in $\boldsymbol{\beta}$. For more information on MCMC methods, see Gelman et al. (2003).

Appendix B
An Alternative Multivariate-Regression Model

In case the number of locations J along the profiles is small, a simpler multivariate linear regression model can be used. The idea is to model each of the values of the functional response at the different locations s_j, $Y(s_j)$ $(j = 1, 2, ...J)$, as J different responses and use a standard Bayesian multivariate linear regression model to compute the predictive density for a new set of J "responses" that make up a complete profile. This strategy makes sense when the number of locations J is small because it requires the estimation of a $J \times J$ covariance matrix of general structure. The model is then $\mathbf{Y} = \mathbf{X}\boldsymbol{\Gamma} + \mathbf{E}$, where $\mathbf{X}$ is an $N \times q$ design matrix, with all the controllable and noise factors expanded in model form according to some specified q-parameter model, $\boldsymbol{\Gamma}$ is a $q \times J$ matrix containing all the q parameters in the model for the response at each location $j = 1, 2, \ldots, J$ (the same model form is assumed at each location), and $\mathbf{E} \equiv [\varepsilon_{ij}]$ is an $N \times J$ matrix of random errors assumed such that $\boldsymbol{\varepsilon}_{i\bullet} \sim N_J(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\varepsilon}_{\bullet j} \sim N_N(\mathbf{0}, \sigma^2\mathbf{I}_N)$. For this model, a new $J \times 1$ vector of observations $\mathbf{y}$, containing a single not-yet-observed profile, for given levels of the controllable and noise factors $\mathbf{x}$, is assumed to be equal to $\mathbf{y} = \boldsymbol{\Gamma}'\boldsymbol{f}(\mathbf{x}) + \boldsymbol{\varepsilon}$, where $\boldsymbol{f}(\mathbf{x})$ is as in Equation (4) and $\boldsymbol{\varepsilon}$ has the same distribution as a row of $\mathbf{E}$, i.e., a $N(\mathbf{0}, \boldsymbol{\Sigma})$.

Under the classical noninformative joint prior for $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$, it is well known (see, e.g., Press (1982), pp. 136) that the predictive density for a new profile vector response $\mathbf{y}$ is given by a J-dimensional t distribution with $\nu = N - q - J + 1$ degrees of freedom. This is denoted $\mathbf{T}_J^\nu(\mathbf{a}, \mathbf{b})$, where $\mathbf{a}$ is the mean vector and $\mathbf{b}$ is the covariance matrix. Thus,

$$\mathbf{y} \mid \mathbf{x}, \text{data} \sim \mathbf{T}_J^\nu\left(\widehat{\boldsymbol{\Gamma}}'\boldsymbol{f}(\mathbf{x}), \frac{\nu}{\nu-2}\mathbf{H}^{-1}\right),$$

where $\widehat{\boldsymbol{\Gamma}}$ is the ordinary least squares (OLS) estimator of $\boldsymbol{\Gamma}$ and $\mathbf{H}$ is given by

$$\mathbf{H} = \left(\frac{\nu}{N-q}\right)\frac{\widehat{\boldsymbol{\Sigma}}^{-1}}{1 + \mathbf{x}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}}$$

where $\widehat{\boldsymbol{\Sigma}}$ is the MLE of $\boldsymbol{\Sigma}$.

Appendix C
Computer Implementation

All computations were implemented in MATLAB. The program `DriverBayesianRPDOpt.m`, available from the first author, contains all data sets and calls the MCMC and optimization routines used. The MATLAB Statistics and Optimization toolboxes were used. Trujillo-Ortiz et al. (2007) provide a Matlab implementation of Royston's test for multivariate normality and Schoenfeld (2008) provides the MLE estimation of the Laird–Ware random effects model of Equation (6), which we used to compute the AIC, BIC, and likelihood statistics used for covariance model selection.

References

American Supplier Institute (ASI). (1998). *Robust Designs Using Taguchi Methods*. Livonia, MI: ASI Press.

Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd. edition. Boca Raton, FL: Chapman & Hall/CRC Press, Inc.

Chib, S. and Carlin, B. P. (1999). "On MCMC Sampling in Hierarchical Longitudinal Models" *Statistics and Computing* 9, pp. 17–26.

Congdon, P. (2006). *Bayesian Statistical Modeling*, 2nd. edition. New York, NY: John Wiley.

Fang, K.-T.; Li, R.; and Sudjianto, A. (2005). *Design and Modeling for Computer Experiments*. Boca Raton, FL: Chapman & Hall/CRC Press Inc.

Fitzmaurice, G. M.; Laird, N. M.; and Ware, J. H. (2004), *Applied Longitudinal Analysis*. Hoboken, NJ: John Wiley & Sons.

Gelfand, I. M. and Fomin, S. V. (1963). *Calculus of Variations*. Englewood Cliffs, NJ: Prentice Hall.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York, NY: Cambridge University Press.

Gelman, A.; Carlin, J. B.; Stern, H. S.; and Rubin, D. B. (2004). *Bayesian Data Analysis*, 2nd edition. Boca Raton, FL: Chapman & Hall/CRC Press Inc.

Govaerts, B. and Noel, J. (2005). "Analyzing the Results of a Designed Experiment when the Response Is a Curve: Methodology and Application in Metal Injection Moulding". *Quality and Reliability Engineering International* 21, pp. 509–520.

Henderson, V. H. and Searle, S. R. (1979). "Vec and Vech Operators for Matrices, with some Uses in Jacobians and Multivariate Statistics". *The Canadian Journal of Statistics* 7(1), pp. 65–81.

Hung, Y. V.; Joseph, R.; and Melkote, S. N. (2011). "Analysis of Computer Experiments with Functional Response". Technical report, Dept. of Industrial & Systems Engineering, Georgia Institute of Technology, GA (submitted for publication).

Joseph, V. R. and Wu, C. F. J. (2002). "Robust Parameter Design of Multiple-Target Systems". *Technometrics* 44(4), pp. 338–346.

Kim, K.; Mahmoud, M. A.; and Woodall, W. H. (2003). "On the Monitoring of Linear Profiles". *Journal of Quality Technology* 35(3), pp. 317–328.

Laird, N. M. and Ware, J. H. (1982). "Random Effects Models for Longitudinal Data". *Biometrics* 38(4), pp. 963–974.

Lange, N.; Carlin, B. P.; and Gelfand, A. E. (1992). "Hierarchical Bayes Models for the Progression of HIV Infection Using Longitudinal CD4 T-Cell Numbers". *Journal of the American Statistical Association* 87, pp. 615–626.

Lesperance, M. L. and Park, S.-M. (2003). "GLMs for the Analysis of Robust Designs with Dynamic Characteristics". *Journal of Quality Technology* 35(3), pp. 253–263.

Lindley, D. V. and Smith, A. F. M. (1972). "Bayes Estimates for the Linear Model". *Journal of the Royal Statistical Society, Series B (Methodological)* 34(1), pp. 1–41.

Miller, A. (2002). "Analysis of Parameter Design Experiments for Signal-Response Systems". *Journal of Quality Technology* 34, pp. 139–151.

Miller, A. and Wu, C. F. J. (1996). "Parameter Design for Signal-Response systems: A Different Look at Taguchi's Dynamic Parameter Design". *Statistical Science* 11(2), pp. 122–136.

Miro, G.; Del Castillo, E.; and Peterson, J. J. (2004). "A Bayesian Approach for Multiple Response Surface Optimization in the Presence of Noise Variables". *Journal of Applied Statistics* 31(3), pp. 251–270.

McCaskey, S. D. and Tsui, K.-L. (1997). "Analysis of Dynamic Robust Design Experiments". *International Journal of Production Research* 35(6), pp. 1561–1574.

Nair, V. J.; Taam, W.; and Ye, K. Q. (2002). "Analysis of Functional Responses from Robust Design Studies". *Journal of Quality Technology* 34(4), pp. 355–370.

Peterson, J. J. (2004). "A Posterior Predictive Approach to Multiple Response Surface Optimization". *Journal of Quality Technology* 36(2), 139–153.

Peterson, J. J.; Miro-Quesada, G.; and del Castillo, E. (2009). "A Bayesian Reliability Approach to Multiple Response Surface Optimization with Seemingly Unrelated Regression Models". *Journal of Quality Technology and Quantitative Management* 6(4), pp. 353–369.

Press, S. J. (1982). *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Analysis*. Malabar, FL: R. E. Krieger Pub. Co.

Qian, P. Z. G. and Wu, C. F. J. (2008). "Bayesian Hierarchical Modeling for Integrating Low-Accuracy and High-Accuracy Experiments". *Technometrics* 50(2), pp. 192–204.

Rajagopal, R.; del Castillo, E.; and Peterson, J. J. (2005), "Model and Distribution-Robust Process Optimization with Noise Factors". *Journal of Quality Technology* 37, pp. 210–222; "Corrigendum". *Journal of Quality Technology* 38, p. 83).

Royston, J. P. (1995). "Remark AS R94: A Remark on Algorithm AS 181: The W-Test for Normality". *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 44(4), pp. 547–551.

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

Schoenfield, D. (2008). "Random Effects: Estimation of the Laird-Ware Random Effects Model". A MATLAB file. http://www.mathworks.com/matlabcentral/fileexchange/20276.

Taguchi, G. (1987). *System of Experimental Design*. New York, NY: Unipub/Kraus International Publications.

Trimm, N. H. (2002). *Applied Multivariate Analysis*. New York, NY: Springer.

Trujillo-Ortiz, A.; Hernandez-Walls, R.; Barba-Rojo, K.; and Cupul-Magana, L. (2007). "Roystest: Royston's Multivariate Normality Test". A MATLAB file. http://www.mathworks.com/matlabcentral/fileexchange/l17811.

Tsui, K.-L. (1999). "Modeling and Analysis of Dynamic Robust Design Experiments". *IIE Transactions* 31(12), pp. 1113–1122.

Ware, J. H. (1985). "Linear Models for the Analysis of Longitudinal Studies". *The American Statistician* 39(2), pp. 95–101.

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