Gaussian Process Modeling and Optimization of Profile Response Experiments

Hussam Alshraideh

Department of Industrial Engineering

Jordan University of Science and Technology, Irbid, 22110, Jordan

Enrique del Castillo

Department of Industrial and Manufacturing Engineering

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

November 1, 2012

Abstract

Experiments where the response of interest is a curve or "profile" arise in a variety of applications in engineering practice. In a recent paper (Journal of Quality Technology, 44, 2, pp. 117-135, 2012) a mixed effects, bayesian approach was proposed for the bayesian optimization of profile response systems, where a particular shape of the profile response defines desired properties of the product or process. This paper proposes an alternative spatio-temporal Gaussian Random Function process model for such profile response systems, which is more flexible with respect to the types of desired profiles shapes that can be modeled, and allows to model profile-to-profile correlation, if this exists. The method is illustrated with real examples taken from the literature, and practical aspects related to model building and diagnostics are discussed.

Keywords: Functional Responses, Gaussian Random Function Processes, Separability, Robust Parameter Design.

1. Introduction

In many areas of manufacturing and engineering design, the response of interest in an experiment does not consist simply of a single observation measured at each of the experimental

conditions. Instead, at each experimental run a curve is measured at different locations, or values, of an independent variable and the goal is to model the observed curve. This kind of response variable is known as a "profile response" in the Statistical Process Control literature (Kang and Albin, 2000) or a functional response in the Statistics literature (Ramsay and Silverman, 2005).

More precisely, consider an experiment where the response of interest $\{y_j\}_{j=1}^{J}$ is observed at J locations $\boldsymbol{s} = \{s_1, s_2, ..., s_J\}$, every time an experimental run is conducted. The locations s_j can refer to instances in time when the profile response is observed, or in general, they can refer to some other scalar variable the observed profiles depend on. Assume the process performance depends on the shape of the sampled profile, where a given target profile shape is desired. The shape of the sampled profile is modifiable through a set of control factors $\boldsymbol{x}_c$ and is also affected by noise factors $\boldsymbol{x}_n$. Similarly to classical Robust Parameter Design (RPD) experiments, noise factors are assumed to be controllable in a carefully designed experiment but are uncontrollable once the product or process under study is in regular use. The goal is to find the best settings of $\boldsymbol{x}_c$ that make the process achieve a desired target profile shape with maximum probability.

An example of a profile response experiment is given by Nair et al. (2002) where the design of an electric alternator was studied. The response of interest in the experiment was the electric current generated at different rotational speeds (in RPM's) at which the alternator operates. A designed experiment was run that consisted of 8 controllable factors and 2 noise factors. Thus, we have $\boldsymbol{x}_c = (x_1, x_2, ..., x_8)'$ and $\boldsymbol{x}_n = (x_9, x_{10})'$. The experimental design used was a Taguchi L_{18} orthogonal array replicated 6 times. At each of the 108 $(= N)$ runs, 7 $(= J)$ electrical current measurements were recorded at different RPM's. Figure 1 shows the observed profiles for each of the 18 control factor combinations in the design. Based on the DOE data, the goal is to find the controllable factor settings that lead to a specified shape of the electric current profile with high probability. We will return to this example in sections 2 and 3.

The structure of the data just described is similar to what in the Spatial Statistics literature is referred to as *spatio-temporal* data (see e.g., Banerjee et al., 2004, section 8.1). In our case, the 'spatial' and 'temporal' spaces correspond to the design factor space and the measurement locations space, respectively. In Spatial Statistics, this kind of data is usually modeled using a Gaussian Random Function (GRF) process (Lifshit, 1995). In this paper, a GRF process model is adopted to model and optimize profile responses.

A recent review of the RPD problem for profile responses was given by Del Castillo et al. (2012) where a full Bayesian two-stage mixed effects regression model was proposed. Their approach allows to estimate the probability of meeting given specifications for the profile response. However, this model assumes no profile-to-profile correlation, and this might not be true specially if profiles are observed at similar design factor settings. Also, their methods

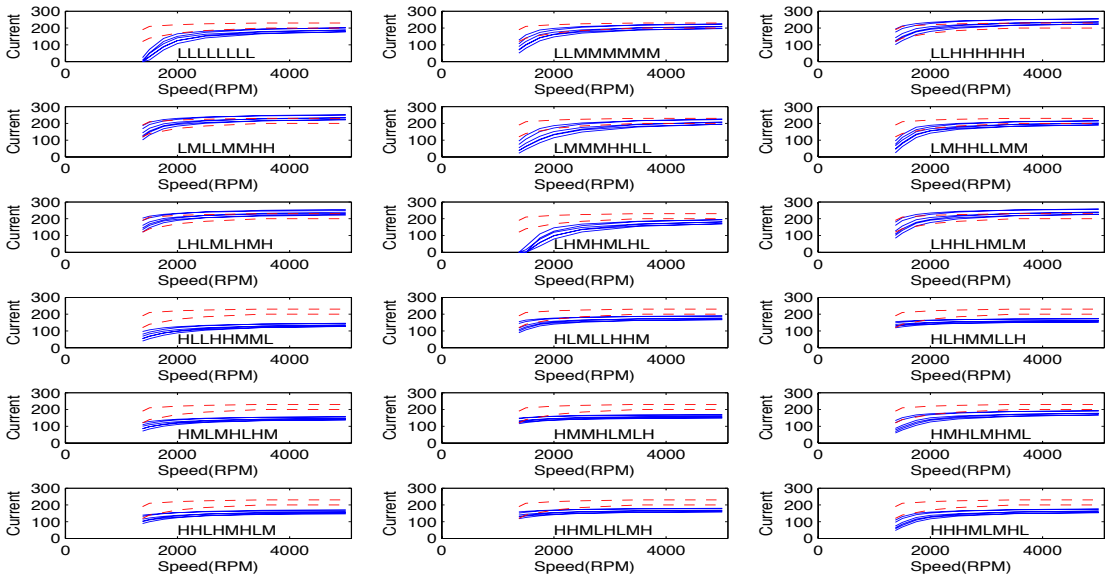


Figure 1: Observed profiles for the electric alternator example at each of the 18 control factors combinations (H=high, M=medium, L=low), after Nair et al. (2002). Each factor combination was replicated six times. The assumed specification limits are shown in dashed lines.

require a linear statistical model in the s_j's for the mean profile that fits the data well, otherwise, model predictions may not be accurate. As an alternative, we propose a more flexible Bayesian modeling approach to model and optimize profile response systems based on a spatio-temporal GRF process.

The rest of the paper is organized as follows. The proposed model is explained in section 2. Section 3 discusses model robustness and validation. A summary and a discussion of areas for further research conclude the paper in section 4. Technical details about the particular Markov Chain Monte Carlo methods utilized are shown in the Appendix.

2. Description of the proposed model

Following common convention in GRF modeling, we assume that the process that generates the observed profiles follows an infinite dimensional normal distribution. Due to the properties of the normal distribution, any finite set of observations will also be normally distributed (see e.g., Mardia and Goodall, 1993). It is also assumed that all N profiles are measured (sampled) at the same J locations. Let y_{ij} be the observed response value of profile i at location j. Also,

let $\boldsymbol{x} = (\boldsymbol{x}_c, \boldsymbol{x}_n)$ and define $\boldsymbol{f}(\boldsymbol{x}_i, s_j)$ to be a function of the design factor settings $\boldsymbol{x}$ under which the j^{th} location in the i^{th} profile was observed. Then the observed response values can be modeled as

$$\begin{aligned} y_{ij} &= \mu(\boldsymbol{x}_i, s_j) + \epsilon_{ij}, \qquad \epsilon_{ij} \sim N(0, \sigma^2(\boldsymbol{x}_i, s_j)) \\ &= \boldsymbol{f}(\boldsymbol{x}_i, s_j)'\boldsymbol{\beta} + \epsilon_{ij} \end{aligned} \tag{1}$$

We note that both the mean and the variance are functions of the design factor settings and of the measurement locations. Let $\boldsymbol{\epsilon}_i$ be the vector of errors ϵ_{ij} observed along the i^{th} profile, i.e., $\boldsymbol{\epsilon}_i = [\epsilon_{i1}, \epsilon_{i2}, ..., \epsilon_{iJ}]'$ and let

$$\boldsymbol{F}_i(\boldsymbol{x}_i, \boldsymbol{s}) = \begin{bmatrix} \boldsymbol{f}(\boldsymbol{x}_i, s_1)' \\ \boldsymbol{f}(\boldsymbol{x}_i, s_2)' \\ \vdots \\ \boldsymbol{f}(\boldsymbol{x}_i, s_J)' \end{bmatrix}.$$

Then the i^{th} profile, $\boldsymbol{y}_i$, can be expressed as

$$\boldsymbol{y}_i(\boldsymbol{x}_i) = \boldsymbol{F}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \qquad \boldsymbol{\epsilon}_i \sim N_J(\boldsymbol{0}, \boldsymbol{\Sigma}_s(\boldsymbol{x}_i, \boldsymbol{s})) \tag{2}$$

where $\boldsymbol{\Sigma}_s$ is the within-profile covariance matrix. If we define the matrices

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1J} \\ y_{21} & y_{22} & \cdots & y_{2J} \\ \vdots & \vdots & \vdots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NJ} \end{pmatrix}, \quad \boldsymbol{F} = \begin{bmatrix} \boldsymbol{F}_1 \\ \boldsymbol{F}_2 \\ \vdots \\ \boldsymbol{F}_N \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_N \end{bmatrix}, \tag{3}$$

then $\boldsymbol{Y}$, the matrix of all N profiles, can be represented as:

$$\begin{aligned} vec(\boldsymbol{Y}') &= \boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N_{NJ}(\boldsymbol{0}, \boldsymbol{\Sigma}(\boldsymbol{x}, \boldsymbol{s})) \quad \text{or} \\ vec(\boldsymbol{Y}') &\sim N_{NJ}\left(\boldsymbol{F}\boldsymbol{\beta}, \boldsymbol{\Sigma}\right) \end{aligned} \tag{4}$$

where $\boldsymbol{F}$ is of size $NJ \times q$, $\boldsymbol{\beta}$ is a $q \times 1$ vector containing regression parameters, and $vec(\cdot)$ is the operator that concatenates matrix columns into one vector.

The covariance structure, $\boldsymbol{\Sigma}$, must capture both the within- and the between-profile correlations. In order to reduce the number of covariance parameters to be estimated one could assume a spatial covariance model over both $\boldsymbol{x}$ and $\boldsymbol{s}$. However, the design factor space

(the x-space) and the measurement location space (the s-space) are usually measured using different scales and hence a single spatial covariance model would be inadequate (Wikle and Berliner, 2005). Therefore, we assume instead two separate spatial covariance models, one for the x-space and another for the s-space, in such a way that:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_s \tag{5}$$

where the $N \times N$ matrix $\boldsymbol{\Sigma}_x$ models the between-profile correlations due to the similarity in control factor settings, the $J \times J$ matrix $\boldsymbol{\Sigma}_s$ models the within-profile correlations due to the proximity of any two locations in $\boldsymbol{s}$, and $\otimes$ denotes the Kronecker product. Assumption (5) is referred to as *separability* in the Spatial Statistics literature (Genton, 2007). The rationale is that if operating conditions $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are close in x-space, then they will tend to result in similar profile responses. Likewise, if responses y_{ij} and y_{ik} are such that j and k are close in s-space, then they will tend to be similar. To illustrate, Figure 2 shows a schematic representation of a profile response experiment in a two dimensional x-space. The two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are relatively close to each other in the x-space, and it is reasonable to expect they generate two profiles $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ in the s-space that are correlated. In contrast, point $\boldsymbol{x}_3$ is relatively farther away (in the x-space) from $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ and should generate a profile response $\boldsymbol{y}_3$ that is not that correlated with $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$. The Kronecker product decomposition of $\boldsymbol{\Sigma}$ makes the model more attractive computationally for large size problems (Genton, 2007), since one deals with separate $N \times N$ and $J \times J$ covariance matrices $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_s$, respectively, instead of a single $NJ \times NJ$ covariance matrix, $\boldsymbol{\Sigma}$. Another reason behind the wide use of separable covariance structures is that they provide an easy way for generating positive definite covariance matrices (Genton, 2007; Gneiting, 2002). We note that this Kronecker product decomposition is not unique since for any scalar c,

$$\boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_s = (c\boldsymbol{\Sigma}_x) \otimes \left(\frac{1}{c}\boldsymbol{\Sigma}_s\right). \tag{6}$$

which results in a non-identifiability problem. To further reduce the number of parameters in the separable covariance structure and to eliminate the non-identifiability problem, we assume the following exponential covariance functions:

$$\boldsymbol{\Sigma}_s = \exp\{-\boldsymbol{D}_s/\phi_s\} \tag{7}$$

$$\boldsymbol{\Sigma}_x = \kappa \exp\{-\boldsymbol{D}_x/\phi_x\} + \psi_x \boldsymbol{I}. \tag{8}$$

where $\boldsymbol{D}_s$ is the $J \times J$ matrix of Euclidean distances between the locations $\{s_j\}$ and $\boldsymbol{D}_x$ is

an $N \times N$ matrix of distances between the design factor settings. Notice that we have only four parameters to determine, $(\phi_s, \kappa, \phi_x, \psi_x)$. The covariance functions (7-8) set the diagonal elements of $\boldsymbol{\Sigma}_s$ to one, restricting the constant c to be equal to one in (6).

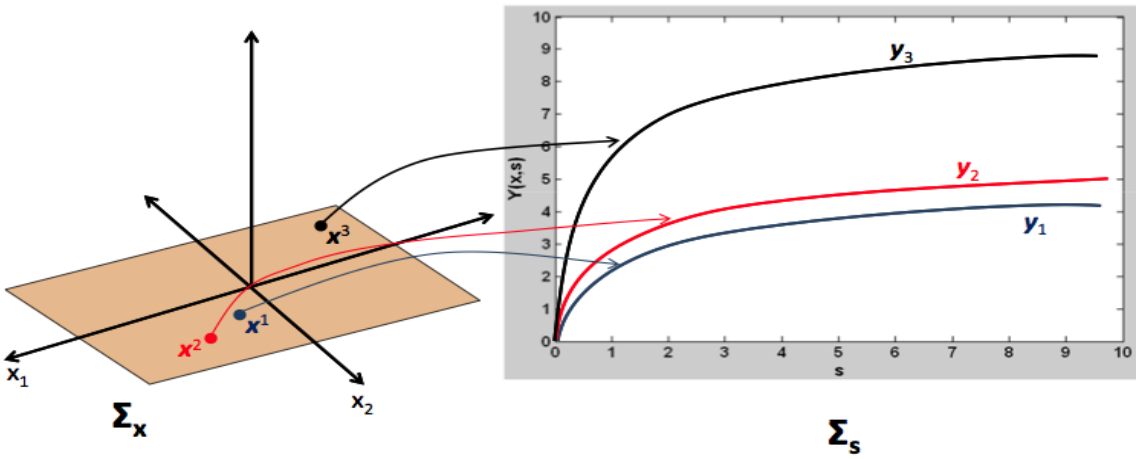


Figure 2: A schematic representation of the spatio-temporal data structure. Points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are close in the x-space and hence they are expected to generate similar profiles, $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$. In contrast, a point $\boldsymbol{x}_3$ in x-space that is farther away from $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is expected to generate a profile response $\boldsymbol{y}_3$ that is not that correlated with $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$. These between-profile correlations are modeled in $\boldsymbol{\Sigma}_x$. Within profile correlations are modeled in matrix $\boldsymbol{\Sigma}_s$.

Letting $\boldsymbol{\theta} = (\phi_s, \psi_x, \phi_x, \kappa)$ and assuming the priors $\pi(\boldsymbol{\theta})$ and $\pi(\boldsymbol{\beta})$ are available, the joint posterior density is:

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\beta} \mid \boldsymbol{Y}, \boldsymbol{F}) \quad \propto \quad & \pi(\boldsymbol{\theta})\pi(\boldsymbol{\beta})|\boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_s|^{-\frac{1}{2}} \\ & \exp\{-\frac{1}{2}(vec(\boldsymbol{Y}') - \boldsymbol{F}\boldsymbol{\beta})'(\boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_s)^{-1}(vec(\boldsymbol{Y}') - \boldsymbol{F}\boldsymbol{\beta})\} \end{aligned} \tag{9}$$

For this model, full conditionals of the parameters are usually hard or even impossible to derive in a closed form (see Banerjee et al., 2004, section 5.1.1), and hence Gibbs sampling is not possible in general. Therefore, a Metropolis-Hastings Markov Chain Monte Carlo (MCMC) algorithm was used to draw samples from the posterior distribution in (9). Full Bayesian estimation of a GRF model is complicated due to convergence problems (see Besag and Green, 1993). We have achieved good convergence behavior with the parametrization (7-8) and the adaptive Metropolis method of Appendix A where the adaptation method of Haario et al. (2001) was used.

If $\boldsymbol{z}$ is a new observed profile at factor settings $\boldsymbol{x}^*$ then, due to the GRF process assumption

we have:

$$\left(\begin{bmatrix} vec(\boldsymbol{Y}') \\ \boldsymbol{z}' \end{bmatrix} \mid \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{F}, \boldsymbol{x}^* \right) \sim N_{(N+1)J} \left(\begin{bmatrix} \boldsymbol{F}\boldsymbol{\beta} \\ \boldsymbol{F}_z(\boldsymbol{x}^*, \boldsymbol{s})\boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right) \tag{10}$$

where $\boldsymbol{\Sigma}_{11} = \boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_s$, $\boldsymbol{\Sigma}_{22} = (\psi_x + \kappa)\boldsymbol{\Sigma}_s$ and $\boldsymbol{\Sigma}_{12}$ is an $NJ \times J$ matrix such that the i^{th} $J \times J$ block is $[\psi_x + \kappa \exp(-d(\boldsymbol{x}_i, \boldsymbol{x}^*)/\phi_x)]\boldsymbol{\Sigma}_s$ for $i = 1, ..., N$. The distribution of $\boldsymbol{z} \mid \boldsymbol{Y}, \boldsymbol{F}, \boldsymbol{x}^*, \boldsymbol{\theta}, \boldsymbol{\beta}$ is easily derived using basic results from Multivariate Normal Theory (see, e.g., Johnson and Wichern, 1998) and is equal to:

$$\boldsymbol{z}' \mid \boldsymbol{Y}, \boldsymbol{F}, \boldsymbol{x}^*, \boldsymbol{\theta}, \boldsymbol{\beta} \sim N\left(\boldsymbol{F}_z(\boldsymbol{x}^*, \boldsymbol{s})\boldsymbol{\beta} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(vec(\boldsymbol{Y}') - \boldsymbol{F}\boldsymbol{\beta}), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right) \tag{11}$$

Using the composition rule (see, e.g., Gelman et al., 2004) we can integrate numerically the posterior predictive density $\pi(\boldsymbol{z} \mid \boldsymbol{Y}, \boldsymbol{F}, \boldsymbol{x}^*)$:

$$\pi(\boldsymbol{z} \mid \boldsymbol{Y}, \boldsymbol{F}, \boldsymbol{x}^*) \;\; \propto \;\; \int\int \pi(\boldsymbol{z}|\boldsymbol{x}^*, \boldsymbol{\beta}, \boldsymbol{\theta})\pi(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \boldsymbol{Y}, \boldsymbol{F})d\boldsymbol{\beta}d\boldsymbol{\theta} \tag{12}$$

The integration in (12) needs to be carried out by first generating samples for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ from their joint posterior density (9). Appendix A provides an adaptive MCMC algorithm to generate $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ samples from the posterior distribution in (9). The prior distributions shown in expressions (14-18) in Appendix A were used for the model parameters in all examples in this paper. The generated $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ samples are then used in turn to generate as many samples as needed from the density $\pi(\boldsymbol{z} \mid \boldsymbol{Y}, \boldsymbol{F}, \boldsymbol{x}^*, \boldsymbol{\theta}, \boldsymbol{\beta})$.

To find the optimal control factor settings that make the process robust to variability in the noise factors, we maximize the "probability of conformance" to the given specification limits, $p(\boldsymbol{x}_c)_{RPD}$, with respect to $\boldsymbol{x}_c$, where

$$p(\boldsymbol{x}_c)_{RPD} \;\; = \;\; E_{\boldsymbol{x}_n}[P(\boldsymbol{z} \in \boldsymbol{T} \mid \boldsymbol{Y}, \boldsymbol{F}, \boldsymbol{x}_c, \boldsymbol{x}_n)] \;\; = \;\; \int P(\boldsymbol{z} \in \boldsymbol{T} \mid \boldsymbol{Y}, \boldsymbol{F}, \boldsymbol{x}_c, \boldsymbol{x}_n)\pi(\boldsymbol{x}_n)d\boldsymbol{x}_n$$

Here, $\boldsymbol{T}$ is a given set of specifications for the desired values for the response (e.g., low and high values of y at each location s_j), and $\pi(\boldsymbol{x}_n)$ is the probability density of the noise factors which we assume known following standard RPD assumptions. This is the Bayesian predictive optimization approach of Peterson (2004) extended to the RPD case by Miro et al. (2004), applied to functional or profile responses, where $p(\boldsymbol{x}_c)_{RPD}$ is estimated by Monte Carlo integration. It is related to the approach in Del Castillo et al. (2012) but with a more flexible GRF model substituting the mixed effects model of that paper. The Bayesian predictive

approach has the advantage of considering the correlation structure of the data, the variability of the noise factors, and the uncertainty in the model parameters.

If the observed data has a trend in the x-space, the s-space or both, then this is modeled via the mean $\mu(\boldsymbol{x}, s)$. Noise×control factor interactions for the mean may be needed in a Robust Parameter Design problem (Myers and Montgomery, 1995) and should be included in $\mu(\boldsymbol{x}, s)$.

Example 1, no noise factors: metal injection moulding process. Govaerts and Noel (2005) report an experiment where 25 profiles of the elastic modulus ($\boldsymbol{y}$) of 'green' parts (products before a sintering operation) were observed in a metal injection moulding process. The elastic modulus was measured at 701 values (locations) of the debinding temperature which ranged from 10 to 80 ^{0}C. The experiment consisted of two controllable factors in the ingredients of the binder, namely, Xanthan concentration (denoted by x_1 and varied at 5 levels from 1 to 5) and Chromium/Xanthan concentration ratio (denoted by x_2 and varied at 4 levels from 1:1 to 4:1). To speed up computations, the number of locations was reduced to 78 locations by sampling every 9^{th} observed value. It was reported by the authors that one of the profiles was a clear outlier, so it was excluded from the analysis. The objective of the experiment as discussed by the authors is to obtain a large elastic modulus at low temperature values while using low Chromium concentration, given it is a pollutant. The specification limits were therefore set considering the range of the observed profiles and the conditions above. Figure 3 shows the observed 24 profiles (after removing the outlier profile) along the 14 distinct design factor settings used.

We assume that the mean structure has a first order form in design factors and locations, i.e., $\mu(\boldsymbol{x}_i, s_j) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 s_j$, $i = 1, ..., 24$ and $j = 1, ..., 78$. One million samples from the posterior distribution were generated and thinned by keeping every 10^{th} sample resulting in 100,000 samples. The thinned samples were checked for convergence through the following plots of the four covariance parameters: a trace plot, an autocorrelation function plot (ACF), a plot for the expected value vs. the sample number, a plot of the MCMC standard error of the posterior mean estimate calculated using the batch means method (see e.g., Flegal et al., 2008), and a plot for the posterior variance vs. the sample number. As it can be seen from Figure 4, the trace plots show convergence to a steady state behavior. In addition, the expected value and variance plots also converge, and the MCMC standard error approaches zero as the sample number increases. The generated MCMC samples were used together with the MATLAB's `ga` genetic optimization algorithm to maximize $p(\boldsymbol{x}_c)_{RPD}$. The optimization was constrained to $1 \leq x_1 \leq 5$ and $1 \leq x_2 \leq 4$. The optimal solution found was at $x_1^* = 4.995$ and $x_2^* = 2.006$ with probability of conformance of 0.574. This solution can be verified by looking at the surface plot of $p(\boldsymbol{x}_c)_{RPD}$ shown in Figure 5(a). In more general multi-factor experiments, either using an algorithm which attempts to find a global optimum (like `ga` does) or using a local non-linear

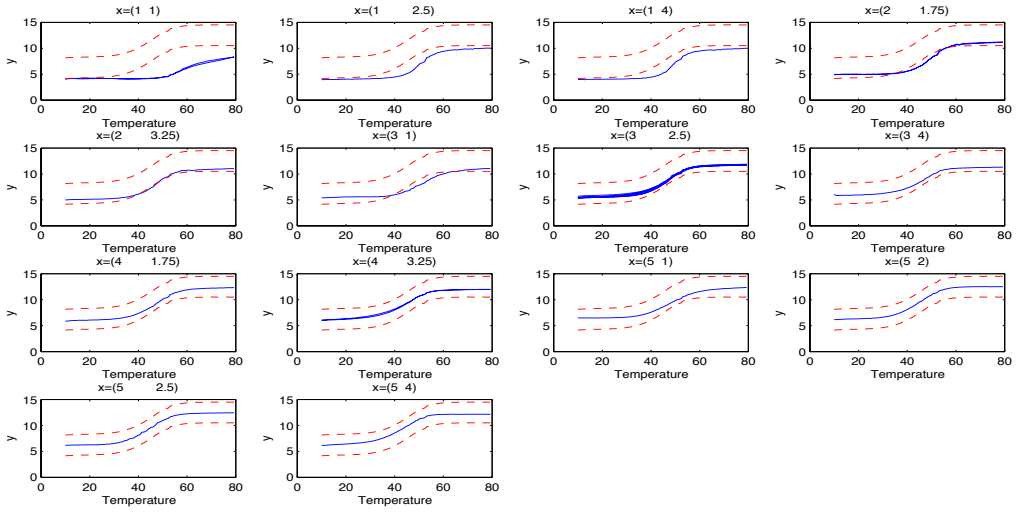


Figure 3: Observed elastic modulus profiles at each of the 14 control factor settings in the metal injection moulding process. The lower and upper specification limits are shown by dashed lines.

optimizer started from a large number of initial points is necessary given $p(\boldsymbol{x}_c)_{RPD}$ is a complex function and not concave in general.

Figure 5(b) shows the mean, 10^{th} and 90^{th} percentiles of the posterior predictive density at the optimal solution found. Luckily, this optimal solution satisfies the criteria for a low Chromium requirement, otherwise, the upper bound for x_2 would need to be lowered. The optimal solution found coincides with that found by Del Castillo et al. (2012), but they report higher conformance probability since they used wider specification limits. It is important to mention that the fitted model needs to be validated before it is used for process optimization, otherwise, optimization results may not be accurate. Model validation is discussed in section 3. We first present an example where noise factors are considered in the experiment. For more examples see Alshraideh (2011).

Example 2. Robust parameter design of an electric alternator. In the electric alternator design example by Nair et al. (2002), presented in the introduction, suppose the goal is to find the controllable factor settings that maximize the probability the electric current profile lies within the specified limits U and L given in Table 1. We assume model (4-8) has a mean model with an intercept, all main effects and all control $\times$ noise interaction terms, i.e., $\mu(\boldsymbol{x}_i, s_j) = \beta_0 + \sum_{k=1}^{10} \beta_k x_{ik} + \sum_{k=1}^{8} \beta_{ik9} x_{ik} x_{i9} + \sum_{k=1}^{8} \beta_{ik10} x_{ik} x_{i10} + \beta_{28} s_j$, $i = 1, ..., 108$ and $j = 1, ..., 7$ (an additive model in $\boldsymbol{x}$ and $\boldsymbol{s}$). Again, one million samples from the posterior distribution were generated using the MCMC algorithm in Appendix A which were thinned

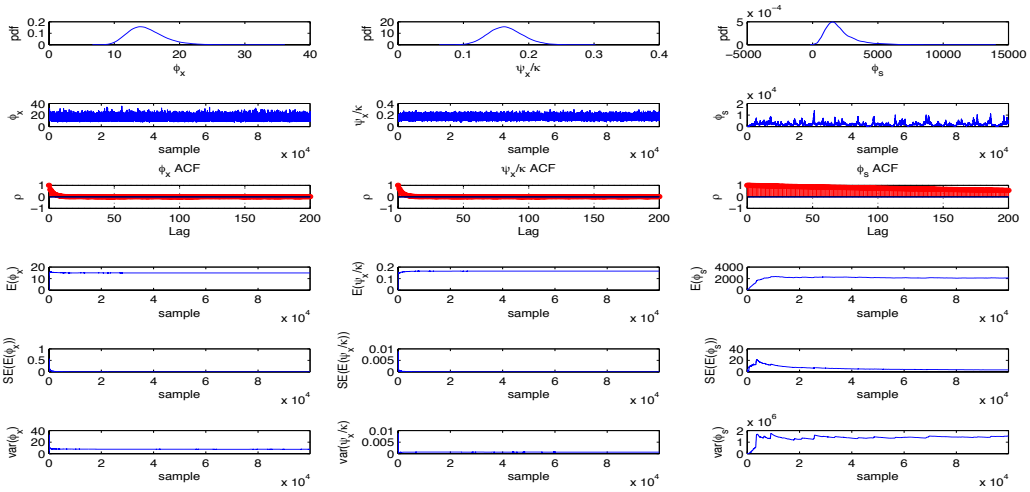
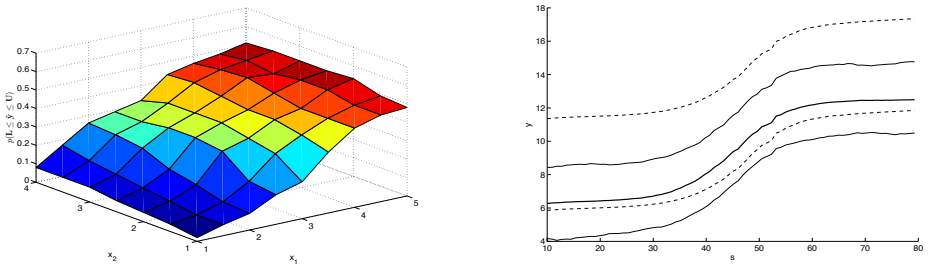


Figure 4: MCMC convergence plots, metal injection moulding example. From top to bottom: posterior densities, trace plots, autocorrelation functions (ACF), expected value plots, MCMC standard error of the estimated posterior mean, and posterior variance plots of the generated covariance parameters $(\phi_x, \psi_x/\kappa, \phi_x)$ (one column of plots per parameter). Plots for the ratio ψ_x/κ are shown since it is known in the Spatial Statistics literature (see e.g., Banerjee et al., 2004) that the ratio of these two parameters converges to its true distribution even if the two parameters do not converge separately.



(a) (b)

Figure 5: Metal injection moulding example: (a) Probability of conformance to specifications as a function of the design factors. (b) Predicted profiles at the optimal design factor settings found. The mean of the posterior predictive density is shown with thick solid line. The other two solid lines are the 10^{th} and 90^{th} percentiles of the posterior predictive density. Upper and lower specification limits are shown as dashed lines.

	s_1	s_2	s_3	s_4	s_5	s_6	s_7
$U(s_j)$	190	210	215	220	225	230	230
$L(s_j)$	120	140	155	170	185	200	200

Table 1: Assumed specification limits at each of the 7 RPM locations for the electric alternator example.

to 100,000 samples by keeping every 10^{th} sample. The posterior densities for the covariance parameters along with the plots used to check for convergence of the MCMC algorithm are shown in Figure 6. The MCMC samples were used with MATLAB's `ga` optimization routine to maximize $p(\boldsymbol{x}_c)_{RPD}$. The optimization was constrained to $-1 \leq x_k \leq 1$ for $k = 1, ..., 8$ and the noise factors were assumed to be independent $U(-1, 1)$ random variables. Joint and marginal probabilities of conformance of the optimal solution found are shown in Table 2. The mean, 10^{th} and 90^{th} percentiles of the posterior predictive density at the optimal solution found are shown in Figure 7. The optimal solution found is similar to that given by Nair et al. (2002) for a robust process. It is noticeable in Table 2 and Figure 7 that the specifications are violated mostly at the last few locations. If the specification limits at these locations can be widened, the conformance probability evidently will improve.

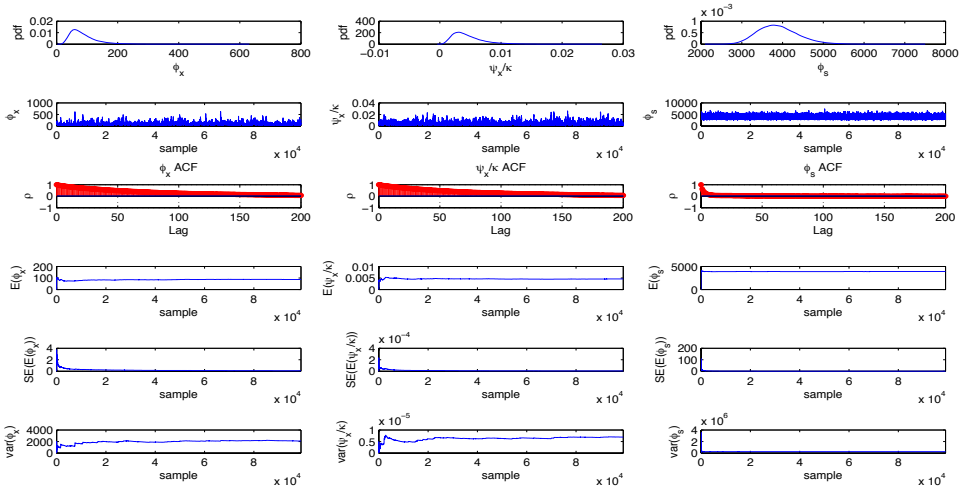


Figure 6: MCMC convergence plots, electric alternator example. From top to bottom: posterior densities, trace plots, autocorrelation functions (ACF), expected value plots, MCMC standard error plots and variance plots of the generated covariance parameters $(\phi_x, \psi_x/\kappa, \phi_x)$ (one column of plots per parameter).

3. Model Robustness and Validation

In this section, we discuss the robustness of the proposed methodology with respect to the covariance separability assumption, followed by a discussion on how to choose the mean structure.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_{opt}	0.198	-0.279	0.997	0.071	0.859	0.983	0.839	0.899
x_{opt} rounded	0	0	1	0	1	1	1	1
$p(L < \boldsymbol{y} < U)$				0.34				
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	
$p(L_i < y_i < U_i)$	0.90	0.92	0.85	0.76	0.75	0.62	0.56	

Table 2: Optimization results for the electric alternator example.

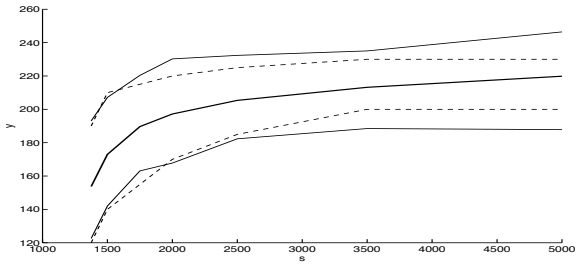


Figure 7: Predicted profiles at the optimal controllable factor settings for the electric alternator example. The mean of the posterior predictive density is shown in solid thick line. The other two solid lines are the 10^{th} and the 90^{th} percentiles of the posterior predictive distribution. The upper and lower specification limits are shown as dashed lines.

3.1. Robustness with Respect to Covariance Separability and Mean Structure Misspecification

Since in the proposed methodology the fitted model is used for process optimization, the reliability of the estimated optimal solutions is important. In this section, we test the robustness of the proposed modeling and optimization approach with respect to

1. misspecification of the model mean structure, and

2. a non-separable covariance structure.

For the simulations in this section, we assume there are two factors x_1 and x_2, each at five levels $\{1, 2, 3, 4, 5\}$, and we assume each profile is measured at 50 locations $\boldsymbol{s} = \{1, 2, \cdots, 50\}$. Three models for the mean structure were assumed. The first one is an intercept only model, i.e., $\mu(\boldsymbol{x}_i, s_j) = \beta_0$. The second mean model is additive in the main effects of factors and locations, that is $\mu(\boldsymbol{x}_i, s_j) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 s_j$. The third mean model has in addition interactions between the x-space and the s-space, i.e., $\mu(\boldsymbol{x}_i, s_j) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 s_j + \beta_4 x_{i1} s_j + \beta_4 x_{i2} + \beta_5 x_{i2} s_j$.

For data generation, we simulated profiles using the model $\boldsymbol{Y} = \boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N_{NJ}(\boldsymbol{0}, \boldsymbol{\Sigma})$ where $\boldsymbol{F}$ is defined as in (3) and the ij^{th} element of $\boldsymbol{\Sigma}$ has the following structure:

$$\boldsymbol{\Sigma}_{ij} = \frac{\sigma^2}{(a(d_s)^{2\alpha} + 1)^{\delta + \gamma d/2}} \exp\left(-\frac{c d_x}{(a(d_s)^{2\alpha} + 1)^{\gamma/2}}\right) \tag{13}$$

In this covariance model, proposed by Genton (2007), d_x is the distance in the x-space between the factor settings at which the i^{th} and j^{th} elements of $vec(\boldsymbol{Y}')$ are observed, d_s is the distance in the s-space between the locations at which the i^{th} and j^{th} elements of $vec(\boldsymbol{Y}')$ are observed, d is the dimension of the $x-$space, and a, c and α are constants that define the smoothness of the covariance function. Notice that $\boldsymbol{\Sigma}$ is the multiplication of two terms. The first is $\sigma^2/(a(d_s)^{2\alpha} + 1)^{\delta + \gamma d/2}$ which is a function of d_s only. The second is $\exp\left(-c\, d_x/(a(d_s)^{2\alpha} + 1)^{\gamma/2}\right)$ which is a function of both d_s and d_x. If $\gamma = 0$, the second term reduces to $\exp(-c\, d_x)$ which depends on d_x only, and hence $\boldsymbol{\Sigma}$ will have a separable structure. Otherwise ($\gamma > 0$), $\boldsymbol{\Sigma}$ has a non-separable structure. The covariance parameters in (13) were set to $\sigma^2 = 0.2$, $a = 1$, $c = 1$, $d = 2$, $\delta = 1$ and $\alpha = 1$ in our simulations. Four values for the constant γ were assumed: $\{0, 0.1, 0.5, 1\}$. The fitted and actual models were then used to find the factor settings that maximize the probability of conformance to the specification limits. The specification limits were fixed at ± 1.5 of the true mean profile evaluated when $x_1 = 2$, $x_2 = 2.5$ and $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, 0.2\boldsymbol{I})$ for all simulations performed.

Table 3 shows the estimated optimal solutions for the 36 combinations tried (3 true mean models, 3 assumed mean models and the 4 values for the non-separability parameter γ) along with the true optimal solution for each case.

As can be seen, when the correct mean structure model is fit (i.e., only the covariance structure separability is assumed and no mean structure misspecification is present), the relative error is smallest. An intercept only model provide acceptable results when the true mean has an intercept only, otherwise, the non-stationarity of the data will not be accounted which results in a poor model performance. If a higher order model than the true one is used for the mean structure, the relative error is higher. This is a similar phenomenon as to that of overfitting in linear regression.

3.2. Selection of the structure of the mean model

The chosen mean structure defines the model residuals, and hence affects the performance of the assumed separable covariance structure model. If the stochastic process that generates the observed profile data is separable (i.e., no significant spatial-temporal interaction), then the covariance structure of that process is separable (Genton, 2007). A separable process

True Mean	γ	True optimal solution x_1^*	x_2^*	$\widehat{p}$	Fitted Mean	Estimated optimal solution x_1^*	x_2^*	$\widehat{p}$	e_x
Intercept	0	2.103	2.622	0.972	Intercept	2.104	2.534	0.99	0.001
					Additive	2.211	2.631	0.99	0.001
					Interaction	2.224	2.862	0.98	0.006
	0.1	3.213	1.250	0.974	Intercept	3.201	1.368	0.99	0.001
					Additive	3.263	1.333	0.99	0.001
					Interaction	3.201	1.458	0.98	0.004
	0.5	2.310	1.473	0.975	Intercept	2.904	1.458	0.98	0.047
					Additive	2.904	1.458	0.98	0.047
					Interaction	2.904	1.458	0.98	0.047
	1	1.722	2.529	0.975	Intercept	2.728	1.893	0.97	0.151
					Additive	1.310	2.868	0.98	0.030
					Interaction	2.201	2.868	0.98	0.037
Additive	0	2.155	2.835	0.971	Intercept	1.977	3.297	0.03	0.019
					Additive	2.098	2.757	0.96	0.001
					Interaction	2.122	2.725	0.95	0.001
	0.1	1.187	2.825	0.971	Intercept	2.967	4.333	0.03	0.058
					Additive	1.405	2.757	0.96	0.006
					Interaction	1.422	2.725	0.95	0.007
	0.5	1.155	2.837	0.964	Intercept	1.155	2.457	0.01	0.015
					Additive	1.422	2.725	0.96	0.009
					Interaction	1.409	2.749	0.96	0.008
	1	1.155	2.873	0.963	Intercept	3.728	1.941	0.02	0.971
					Additive	1.438	2.725	0.96	0.010
					Interaction	1.422	2.725	0.96	0.009
Interaction	0	2.032	2.839	0.959	Intercept	1.832	3.335	0.03	0.023
					Additive	1.739	2.657	0.04	0.010
					Interaction	2.106	2.997	0.97	0.003
	0.1	1.163	2.839	0.961	Intercept	4.637	1.454	0.03	1.486
					Additive	2.132	2.808	0.05	0.100
					Interaction	1.218	2.757	0.97	0.001
	0.5	1.155	2.843	0.965	Intercept	4.477	1.516	0.01	1.359
					Additive	2.405	3.457	0.02	0.206
					Interaction	1.255	2.757	0.96	0.002
	1	1.163	2.839	0.969	Intercept	4.477	1.516	0.01	1.353
					Additive	3.666	1.941	0.01	0.751
					Interaction	2.155	2.108	0.97	0.161

Table 3: True vs. estimated optimal solutions for the 9 model combinations each at 4 levels of the non-separability parameter γ. The relative error is calculated as $e_x = \|\boldsymbol{x}^*_{true} - \boldsymbol{x}^*_{estimated}\|_F / \|\boldsymbol{x}^*_{true}\|_F$. In case the optimizer returns multiple optimal solutions, only one of them is listed.

can be written as the product of two independent processes, one purely spatial and a second purely temporal. Let $\boldsymbol{Y}^*$ be the detrended observed data matrix. Then using singular value decomposition (SVD), $\boldsymbol{Y}^*$ can be written as

$$\boldsymbol{Y}^* = \boldsymbol{U}\boldsymbol{\Delta}\boldsymbol{V}' = \sum_{i=1}^{min(N,J)} \delta_i \boldsymbol{u}_i \boldsymbol{v}_i'$$

where $\delta_i \boldsymbol{u}_i \boldsymbol{v}_i'$ is called the i^{th} Empirical Orthogonal Function (EOF) (see e.g., Banerjee et al., 2004, section 8.1.1). If $\boldsymbol{Y}^*$ can be approximated by its first EOF, $\delta_1 \boldsymbol{u}_1 \boldsymbol{v}_1'$, and $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ are independent, then $\boldsymbol{Y}^*$ is separable (see e.g., Banerjee et al., 2004; Mardia and Goodall, 1993). Therefore, the following three step procedure provides an easy and fast way to check how reasonable the separability assumption is for the observed data:

1. Assume a mean function form and use Least Squares to detrend the data such that $vec(\boldsymbol{Y}^*) = vec(\boldsymbol{Y}) - \boldsymbol{F}\widehat{\boldsymbol{\beta}}$. The mean structure might include an intercept, main effects of design factors and locations, interactions, etc.

2. Construct the singular value decomposition of $\boldsymbol{Y}^*$ and its approximation $\boldsymbol{Y}^*_{approx} = \delta_1 \boldsymbol{u}_1 \boldsymbol{v}_1'$ where δ_1 is the largest eigenvalue, and $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ are the left and right singular vectors, respectively, corresponding to the largest eigenvalue.

3. Graph both $\boldsymbol{Y}^*$ and $\boldsymbol{Y}^*_{approx}$ vs. $\boldsymbol{s}$. If the two plots look similar, then separability is a reasonable assumption . A more precise metric is the relative approximation error defined as $e_{approx} = \|\boldsymbol{Y}^* - \delta_1 \boldsymbol{u}_1 \boldsymbol{v}_1'\|_F / \|\boldsymbol{Y}^*\|_F$ where $\|\cdot\|$ is the Frobenius norm.

This procedure can be repeated for different mean models, and the e_{approx} statistic can be used to select a mean model that is most compatible with the covariance separability model.

To assess the mean model structure, we use cross validation techniques (Hastie et al., 2009). Two different cross validations are shown in what follows, one based on leave-one-out predictions and a second one based on leave-10%-out predictions. In leave-one-out predictions, one of the observed profiles is left out for testing and the remaining $N - 1$ profiles are used for model fitting. The fitted model is then used to predict the left out profile. In leave-10%-out, we instead leave 10% of the observed profiles out and use the remaining 90% for model fitting. Then the fitted model is used to predict the 10% profiles left out. If the fitted model provides acceptable predictions for testing profiles, then it would provide acceptable predictions at other locations in the x-space.

Example 1 (cont.). In the injection moulding example, to check the observed elastic modulus data for separability we fit the same three models as before (intercept only, additive and with interactions). The relative error statistics were 0.48887 for the intercept model, 0.28234 for the additive model and 0.28187 for the interaction model. Looking at the relative error values, it can be seen that an additive mean structure is acceptable, since it provides a simple model with a relative error almost equal to the more complicated interaction model. However, using either model would be adequate in this example, and for the purposes of RPD optimization, if noise factors are present an interaction model that contains the important control $\times$ noise interactions should be preferred.

The additive model fitted for this data was further checked by leave-one-out and leave-10%-out cross validations. Cross validation results are shown in Figures 8 and 9. Based on these figures it can be seen that the model provides acceptable predictions. Hence, the fitted model can be used for process optimization.

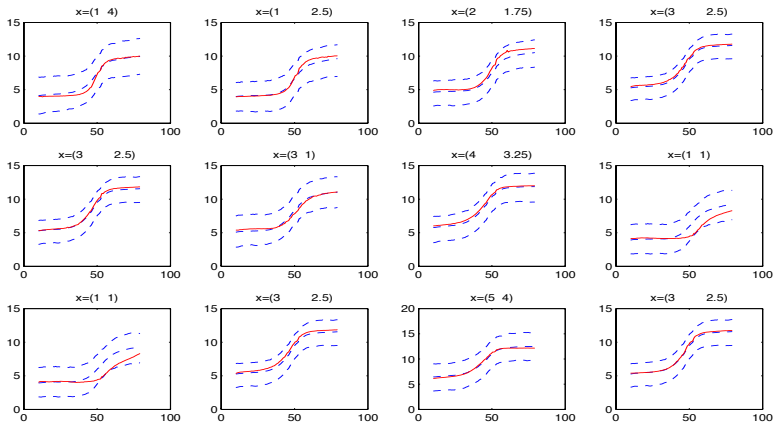


Figure 8: Leave-one-out cross validation results for the testing profiles in the metal injection example. Solid lines are the observed profiles used for testing. The mean, 10^{th} and 90^{th} percentiles of the posterior predictive density are shown in dashed lines. Only 12 of the 24 observed profiles are shown.

Example 2 (cont.). In the alternator design problem of Section 2, we check the observed data of the electric current for separability by fitting again the same three models for the mean profile. The relative error statistics were 0.85403 for the intercept only model, 0.64357 for the additive model (including all two factor control $\times$ noise interactions in the x-space, crucial in RPD optimization) and 0.63703 for the interaction model (including interactions between the x-space and the s-space in addition to all two factor control $\times$ noise interactions). Since the additive model provides a compromised e_{approx} value, then it is the mean structure that is preferred. Cross validation results for the interaction model are shown in Figures 10 and 11.

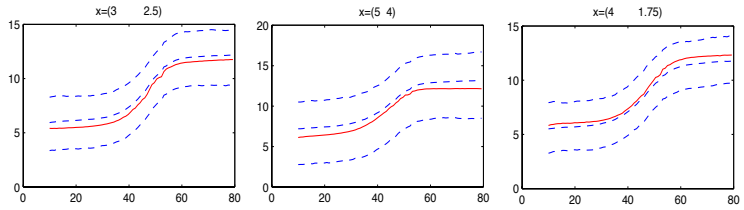


Figure 9: Leave-10%-out cross validation results for the testing profiles in the metal injection example. Solid lines are the observed profiles used for testing. The mean, 10^{th} and 90^{th} percentiles of the posterior predictive density are shown in dashed lines.

Based on cross validation results it can be seen that the model provides acceptable predictions but with high posterior variance due to the large e_{approx} value. The performance of the model can be enhanced by employing a non-separable covariance structure. For illustration purposes, we proceed with the optimization step.

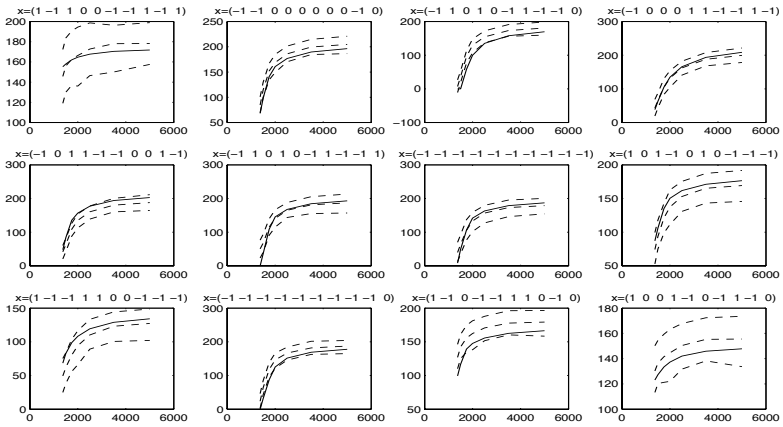


Figure 10: Leave-one-out cross validation results for the testing profiles in the electric alternator example. Solid lines are the observed profiles used for testing. The mean, 10^{th} and 90^{th} percentiles of the posterior predictive density are shown in dashed lines.

4. Conclusions and Future Work

This paper introduces a new approach to solve the RPD problem of profile response systems based on a spatio-temporal Gaussian Random Function. The 'spatial' and 'temporal' spaces correspond to the design factor space and the location measurement space, respectively. In this approach, the observed data is assumed to have a multivariate normal distribution with

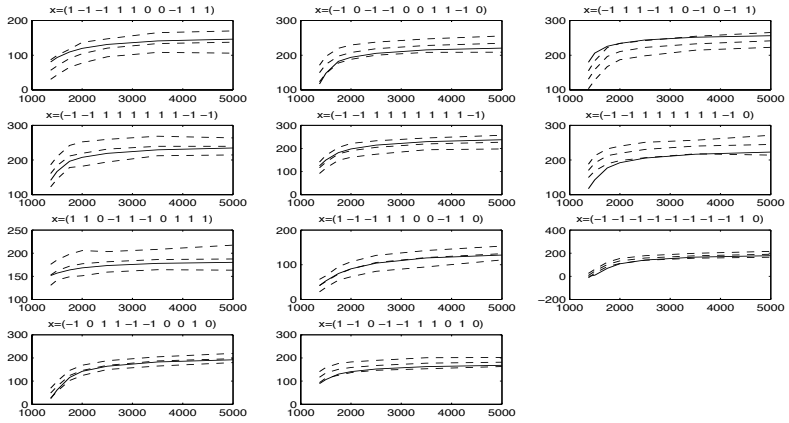


Figure 11: Leave-10%-out cross validation results for the testing profiles in the electric alternator example. Solid lines are the observed profiles used for testing. The mean, 10^{th} and 90^{th} percentiles of the posterior predictive density are shown in dashed lines.

mean structure $\boldsymbol{\mu}(\boldsymbol{x}, s)$ and covariance structure $\boldsymbol{\Sigma}$. The covariance structure is assumed to be the Kronecker product of the between-profiles and the within-profile covariance structures (i.e., we assume covariance separability). As usually done in Spatial Statistics, a parametric model is assumed for both covariance structures, resulting in a notable reduction in the number of covariance parameters to estimate. Like in Universal Kriging practice, the main purpose of the mean structure is to detrend the data and obtain zero mean stationary residuals. In case noise factors are present in the experiment (RPD case), we suggest to include all control $\times$ noise interaction terms (judged important by the experiment) in the mean structure model. The proposed model is flexible enough to provide a good fit even under a specific type of non-separable covariance structure proposed in the literature Genton (2007). Non-separable covariance structures can have many forms, and hence it is impossible to study model performance for general covariance forms. Instead, we propose a methodology to choose the mean structure and then to use cross validations to assess the model performance regardless of the assumed separable covariance structure.

The gaussian random function model allows the experimenter to consider the covariance between profiles. The spatial covariance model modifies the predictions given by the mean model and provides in this way a more flexible approach than the recently proposed mixed-effects model of Del Castillo et al. (2012) which requires in comparison a very good fit for the mean profile.

The assumed separable spatio-temporal covariance might not hold for some profile response data. Hence, using instead a non-separable covariance structure is an area where future research

is needed. In the present model, if a significant interaction between the x-space and the s-space is anticipated (non-separability), this interaction could be included in the mean structure $\boldsymbol{\mu}(\boldsymbol{x}, \boldsymbol{s})$ to ameliorate the effects of the otherwise inadequate separable covariance model we assume.

Acknowledgment. We would like to thank Professor Murali Haran for his assistance in this work. We thank Professor Bernadette Govaerts for kindly providing the metal injection data from their paper (Govaerts and Noel, 2005) and Dr. Winson Taam for providing the alternator data from their paper (Nair et al., 2002). This work was partially funded by the NSF grant CMMI 0825786.

Appendices

A. Adaptive MCMC Sampling Algorithm

The joint posterior distribution for the parameters in model (4-8) is:

$$\begin{aligned}\pi(\phi_s, \psi_x, \phi_x, \kappa, \boldsymbol{\beta} \mid \boldsymbol{Y}, \boldsymbol{F}) \quad \propto \quad & \pi(\phi_s)\pi(\psi_x)\pi(\phi_x)\pi(\kappa)\pi(\boldsymbol{\beta})|\boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_s|^{-\frac{1}{2}} \\ & \exp\{-\frac{1}{2}(vec(\boldsymbol{Y}') - \boldsymbol{F}\boldsymbol{\beta})'(\boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_s)^{-1}(vec(\boldsymbol{Y}') - \boldsymbol{F}\boldsymbol{\beta})\}\end{aligned}$$

Assume the following noninformative priors for model parameters:

$$\pi(\boldsymbol{\beta}) \sim \text{constant} \quad \text{(flat prior)} \tag{14}$$
$$\pi(\psi_x) \sim logN(\mu = 7, \sigma^2 = 1) \tag{15}$$
$$\pi(\phi_s) \sim logN(\mu = 7, \sigma^2 = 1) \tag{16}$$
$$\pi(\phi_x) \sim logN(\mu = 7, \sigma^2 = 1) \tag{17}$$
$$\pi(\kappa) \sim\sim logN(\mu = 7, \sigma^2 = 1) \tag{18}$$

and the following proposal distributions

$$\begin{aligned}q(\psi_x^{new}) &\sim \text{truncated } N(\mu = \psi_x^{current}, \sigma^2_{\psi_x}, 0, \text{inf}) \\ q(\phi_s^{new}) &\sim \text{truncated } N(\mu = \phi_s^{current}, \sigma^2_{\phi_s}, 0, \text{inf}) \\ q(\phi_x^{new}) &\sim \text{truncated } N(\mu = \phi_x^{current}, \sigma^2_{\phi_x}, 0, \text{inf}) \\ q(\kappa^{new}) &\sim \text{truncated } N(\mu = \kappa^{current}, \sigma^2_{\kappa}, 0, \text{inf})\end{aligned}$$

The following function, suggested by (see Haario et al., 2001), is used to adaptively update the proposal distribution variances, $\{\sigma^2_{\psi_x}, \sigma^2_{\phi_s}, \sigma^2_{\phi_x}, \sigma^2_{\kappa}\}$:

$$\sigma^2 = \begin{cases} \sigma_0^2, & k \leq k_0; \\ s_d \, var(s^0, s^1, \cdots, s^{k-1}) + s_d \epsilon, & k > k_0. \end{cases} \tag{19}$$

where s^k is the k^{th} sample of that parameter and σ_0^2,κ_0, s_d (the variance multiplier) and ϵ are constants.

Let $(\phi_s^k, \psi_x^k, \phi_x^k, \kappa^k, \boldsymbol{\beta}^k)$ be the k^{th} sample from the posterior distribution. To get the $k+1$ sample follow the steps:

1. Using the k^{th} sample, calculate $\boldsymbol{\Sigma}_x^k$ and $\boldsymbol{\Sigma}_s^k$ such that

$$\begin{aligned} \boldsymbol{\Sigma}_s^k &= \exp\{-\boldsymbol{D}_s/\phi_s^k\} \\ \boldsymbol{\Sigma}_x^k &= \kappa \exp\{-\boldsymbol{D}_x/\phi_x^k\} + \psi_x^k I \end{aligned}$$

2. **Update ϕ_s:** a. Evaluate the conditional distribution of ϕ_s, $f_{\phi_s}^k$, where

$$\begin{aligned} f_{\phi_s}^k &= \pi(\phi_s^k)|\boldsymbol{\Sigma}_x^k \otimes \boldsymbol{\Sigma}_s^k|^{-\frac{1}{2}} \\ &\quad \exp\{-\frac{1}{2}(vec(\boldsymbol{Y}') - \boldsymbol{F}\boldsymbol{\beta})'(\boldsymbol{\Sigma}_x^k \otimes \boldsymbol{\Sigma}_s^k)^{-1}(vec(\boldsymbol{Y}') - \boldsymbol{F}\boldsymbol{\beta})\} \end{aligned}$$

b. Update $\sigma^2_{\phi_s}$ using (19), propose ϕ_s^{new} from its proposal distribution, and then update $\boldsymbol{\Sigma}_s$ such that

$$\boldsymbol{\Sigma}_s^{new} = \exp\{-\boldsymbol{D}_s/\phi_s^{new}\}$$

c. Evaluate $f_{\phi_s}^{new}$ where

$$\begin{aligned} f_{\phi_s}^{new} &= \pi(\phi_s^{new})|\boldsymbol{\Sigma}_x^k \otimes \boldsymbol{\Sigma}_s^{new}|^{-\frac{1}{2}} \\ &\quad \exp\{-\frac{1}{2}(vec(\boldsymbol{Y}') - \boldsymbol{F}\boldsymbol{\beta})'(\boldsymbol{\Sigma}_x^k \otimes \boldsymbol{\Sigma}_s^{new})^{-1}(vec(\boldsymbol{Y}') - \boldsymbol{F}\boldsymbol{\beta})\} \end{aligned}$$

d. If $u \sim$ uniform(0,1) $\leq \frac{f_{\phi_s}^{new} q(\phi_s^k, \phi_s^{new})}{f_{\phi_s}^k q(\phi_s^{new}, \phi_s^k)}$, then $\phi_s^{k+1} = \phi_s^{new}$ and $\boldsymbol{\Sigma}_s^{k+1} = \boldsymbol{\Sigma}_s^{new}$ otherwise $\phi_s^{k+1} = \phi_s^k$ and $\boldsymbol{\Sigma}_s^{k+1} = \boldsymbol{\Sigma}_s^k$.

3. Update ψ_x: a. Evaluate the conditional distribution of ψ_x, $f^k_{\psi_x}$, where

$$\begin{aligned} f^k_{\psi_x} &= \pi(\psi^k_x)|\boldsymbol{\Sigma}^k_x \otimes \boldsymbol{\Sigma}^{k+1}_s|^{-\frac{1}{2}} \\ &\quad \exp\{-\frac{1}{2}(vec(\boldsymbol{Y}') - \boldsymbol{F\beta})'(\boldsymbol{\Sigma}^k_x \otimes \boldsymbol{\Sigma}^{k+1}_s)^{-1}(vec(\boldsymbol{Y}') - \boldsymbol{F\beta})\} \end{aligned}$$

b. Update $\sigma^2_{\psi_x}$ using (19), propose ψ^{new}_x from its proposal distribution, and then update $\boldsymbol{\Sigma}_x$ such that

$$\boldsymbol{\Sigma}^{new}_x = \kappa^k \exp\{-\boldsymbol{D}_x/\phi^k_x\} + \psi^{new}_x I$$

c. Evaluate $f^{new}_{\psi_x}$ where

$$\begin{aligned} f^{new}_{\psi_x} &= \pi(\psi^{new}_x)|\boldsymbol{\Sigma}^{new}_x \otimes \boldsymbol{\Sigma}^{k+1}_s|^{-\frac{1}{2}} \\ &\quad \exp\{-\frac{1}{2}(vec(\boldsymbol{Y}') - \boldsymbol{F\beta})'(\boldsymbol{\Sigma}^{new}_x \otimes \boldsymbol{\Sigma}^{k+1}_s)^{-1}(vec(\boldsymbol{Y}') - \boldsymbol{F\beta})\} \end{aligned}$$

d. If $u \sim$ uniform(0,1) $\leq \frac{f^{new}_{\psi_x} q(\psi^k_x, \psi^{new}_x)}{f^k_{\psi_x} q(\psi^{new}_x, \psi^k_x)}$, then $\psi^{k+1}_x = \psi^{new}_x$ and $\boldsymbol{\Sigma}^k_x = \boldsymbol{\Sigma}^{new}_x$ otherwise $\psi^{k+1}_x = \psi^k_x$.

4. Update ϕ_x: a. Evaluate the conditional distribution of ϕ_x, $f^k_{\phi_x}$, where

$$\begin{aligned} f^k_{\phi_x} &= \pi(\phi^k_x)|\boldsymbol{\Sigma}^k_x \otimes \boldsymbol{\Sigma}^{k+1}_s|^{-\frac{1}{2}} \\ &\quad \exp\{-\frac{1}{2}(vec(\boldsymbol{Y}') - \boldsymbol{F\beta})'(\boldsymbol{\Sigma}^k_x \otimes \boldsymbol{\Sigma}^{k+1}_s)^{-1}(vec(\boldsymbol{Y}') - \boldsymbol{F\beta})\} \end{aligned}$$

b. Update $\sigma^2_{\phi_x}$ using (19), propose ϕ^{new}_x from its proposal distribution, and then update $\boldsymbol{\Sigma}_x$ such that

$$\boldsymbol{\Sigma}^{new}_x = \kappa^k \exp\{-\boldsymbol{D}_x/\phi^{new}_x\} + \psi^{k+1}_x I$$

c. Evaluate $f^{new}_{\phi_x}$ where

$$\begin{aligned} f^{new}_{\phi_x} &= \pi(\phi^{new}_x)|\boldsymbol{\Sigma}^{new}_x \otimes \boldsymbol{\Sigma}^{k+1}_s|^{-\frac{1}{2}} \\ &\quad \exp\{-\frac{1}{2}(vec(\boldsymbol{Y}') - \boldsymbol{F\beta})'(\boldsymbol{\Sigma}^{new}_x \otimes \boldsymbol{\Sigma}^{k+1}_s)^{-1}(vec(\boldsymbol{Y}') - \boldsymbol{F\beta})\} \end{aligned}$$

d. If $u \sim$ uniform(0,1) $\leq \frac{f^{new}_{\phi_x} q(\phi^k_x, \phi^{new}_x)}{f^k_{\phi_x} q(\phi^{new}_x, \phi^k_x)}$, then $\phi^{k+1}_x = \phi^{new}_x$ and $\boldsymbol{\Sigma}^k_x = \boldsymbol{\Sigma}^{new}_x$ otherwise $\phi^{k+1}_x = \phi^k_x$.

5. Update κ: a. Evaluate the conditional distribution of κ, f_κ^k, where

$$\begin{aligned} f_\kappa^k &= \pi(\kappa^k)|\boldsymbol{\Sigma}_x^k \otimes \boldsymbol{\Sigma}_s^{k+1}|^{-\frac{1}{2}} \\ &\quad \exp\{-\frac{1}{2}(vec(\boldsymbol{Y}') - \boldsymbol{F}\boldsymbol{\beta})'(\boldsymbol{\Sigma}_x^k \otimes \boldsymbol{\Sigma}_s^{k+1})^{-1}(vec(\boldsymbol{Y}') - \boldsymbol{F}\boldsymbol{\beta})\} \end{aligned}$$

b. Update σ_κ^2 using (19), propose κ^{new} from its proposal distribution, and then update $\boldsymbol{\Sigma}_x$ such that

$$\boldsymbol{\Sigma}_x^{new} = \kappa^{new} \exp\{-\boldsymbol{D}_x/\phi_x^{k+1}\} + \psi_x^{k+1} I$$

c. Evaluate f_κ^{new} where

$$\begin{aligned} f_\kappa^{new} &= \pi(\kappa^{new})|\boldsymbol{\Sigma}_x^{new} \otimes \boldsymbol{\Sigma}_s^{k+1}|^{-\frac{1}{2}} \\ &\quad \exp\{-\frac{1}{2}(vec(\boldsymbol{Y}') - \boldsymbol{F}\boldsymbol{\beta})'(\boldsymbol{\Sigma}_x^{new} \otimes \boldsymbol{\Sigma}_s^{k+1})^{-1}(vec(\boldsymbol{Y}') - \boldsymbol{F}\boldsymbol{\beta})\} \end{aligned}$$

d. If $u \sim$ uniform(0,1) $\leq \frac{f_\kappa^{new} q(\kappa^k, \kappa^{new})}{f_\kappa^k q(\kappa^{new}, \kappa^k)}$, then $\kappa^{k+1} = \kappa^{new}$ and $\boldsymbol{\Sigma}_x^{k+1} = \boldsymbol{\Sigma}_x^{new}$ otherwise $\kappa^{k+1} = \kappa^k$ and $\boldsymbol{\Sigma}_x^{k+1} = \boldsymbol{\Sigma}_x^k$.

6. Update $\boldsymbol{\beta}$: Given $\boldsymbol{\Sigma}_x^{k+1}$ and $\boldsymbol{\Sigma}_s^{k+1}$ then

$$\boldsymbol{\Sigma}^{k+1} = \boldsymbol{\Sigma}_x^{k+1} \otimes \boldsymbol{\Sigma}_s^{k+1}$$

Sample $\boldsymbol{\beta}^{k+1}$ from $\pi(\boldsymbol{\beta} \mid \boldsymbol{\Sigma}^{k+1}, \boldsymbol{Y}, \boldsymbol{F})$, where the full conditional of $\boldsymbol{\beta}$ is

$$\pi(\boldsymbol{\beta} \mid \boldsymbol{\Sigma}^{k+1}, \boldsymbol{Y}, \boldsymbol{F}) \sim N\left((\boldsymbol{F}'(\boldsymbol{\Sigma}^{k+1})^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'(\boldsymbol{\Sigma}^{k+1})^{-1}vec(\boldsymbol{Y}'), (\boldsymbol{F}'(\boldsymbol{\Sigma}^{k+1})^{-1}\boldsymbol{F})^{-1}\right)$$

B. Software implementation

All computations reported in this paper were implemented in MATLAB (version 2010a). A graphical user interface has been developed for model building and optimization. The program requires the Statistics and Global Optimization toolboxes.

References

Alshraideh, H. A. (2011), *Analysis and Optimization of Profile and Geometric Shape Response Experiments*, Ph.D dissertation, Penn State University.

Banerjee, S., Carlin, B.P. and Gelfand, A.E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, Chapman & Hall/CRC.

Bastos, L. S., and O'Hagan, A. (2009), "Diagnostics for Gaussian Process Emulators", *Technometrics*, 51(4), 425-438.

Besag, J., and Green, P. J. (1993), "Spatial Statistics and Bayesian Computation", *Journal of the Royal Statistical Society, Series B (Methodological)*, 55(1), 25-37.

Carlin, B. P. and Louis, T. A. (2008), *Bayesian Methods for Data Analysis*, Boca Raton, Chapman & Hall/CRC.

Del Castillo, E., Colosimo, B.M., and Alshraideh, H. (2012), "Bayesian Modeling and Robust Optimization of Functional Responses Affected by Noise Factors". *Journal of Quality Technology*, 44, 2, pp. 117-135.

Flegal, J., Haran, M. and Jones, G. L. (2008), "Markov Chain Monte Carlo: Can We Trust the Third Significant Figure?". *Statistical Science*, 23(2), 250-260.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004), *Bayesian Data Analysis*. 2nd ed., Chapman and Hall CRC, Boca Raton, FL.

Genton, M. G. (2007), "Separable Approximations of Space-time Covariance Matrices", *Environmetrics*, 18, 681-695.

Gneiting, T. (2002), "Nonseparable, Stationary Covariance Functions for Space-Time Data", *Journal of the American Statistical Association*, 97(458), 590-600.

Govaerts, B., and Noel, J. (2005), "Analysing 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, 509-520.

Haario, H., Saksman, E. and Tamminen, J. (2001), "An Adaptive Metropolis Algorithm", *Bernoulli*, 7(2), 223-242.

Hastie, T., Tibshirani, R. and Friedman, J. (2001), *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, New York.

Johnson, R.A. and Wichern, D.W. (1998), *Applied Multivariate Statistical Analysis*, 4th ed., Prentice Hall, New Jersey.

Kang, L., and Albin, S. L. (2000), "On-line Monitoring When the Process Yields a Linear Profile". *Journal of Quality Technology*, 32, 418-426.

Lifshit, M.A. (1995), *Gaussian Random Functions*, Springer, New York.

Mardia, K.V. and Goodall, C. (1993), "Spatio-temporal Analyses of Multivariate Environmental Monitoring Data", In *Multivariate Environmental Statistics*, eds.

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

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

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

Myers, R.H., and Montgomery, D.C. (1995), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments* , John Wiley & Sons.

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

Peterson, J.P. (2004), " Bayesian Reliability Approach to Multiple Response Surface Optimization", *Journal of Quality Technology*, 36(2), 139-153.

Ramsay, J. O. and Silverman, B.W. (2005), *Functional Data Analysis*. 2nd ed., New York, Springer.

Shen, Q., and Xu, H. (2007), "Diagnostics for Linear Models with Functional Responses", *Technometrics*, 49(1), 26-33.

Wikle, C. K., and Berliner, L. M. (2005), "Combining Information across Spatial Scales", *Technometrics*, 47(1), 80-91.

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