A Semiparametric Technique for the Multi-Response Optimization Problem

WEN WAN[1] and JEFFREY B. BIRCH[2]*

[1] *University of Alabama at Birmingham, AL 35294-3300*
[2] *Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0439*

Multi-response optimization (MRO) in response surface methodology (RSM) is quite common in applications. Before the optimization phase, appropriate fitted models for each response are required. A common problem is model misspecification and occurs when any of the models built for the responses are misspecified resulting in an erroneous optimal solution. The model robust regression technique, a semiparametric method, has been shown to be more robust to misspecification than either parametric or nonparametric methods. In this study, we propose the use of model robust regression to improve the quality of model estimation and adapt its fits of each response to the desirability function approach, one of the most popular MRO techniques. A case study and simulation studies are presented to illustrate the procedure and to compare the semiparametric method with the parametric and nonparametric methods. The results show that model robust regression performs much better than the other two methods in terms of model comparison criteria in most situations during the modeling stage. In addition, the simulated optimization results for model robust regression are more reliable during the optimization stage.

KEY WORDS: Desirability Function; Model Robust Regression (MRR); Monte Carlo (MC); Multi-response Optimization (MRO); Response Surface Methodology (RSM).

1   Introduction

Model misspecification is a common problem in many areas including multi-response optimization (MRO) in response surface methodology (RSM). In industry and in many other areas of science, data collected often contain m responses of interest for a single set of explanatory variables, where the $m > 1$. An objective for the MRO problem is to find an optimal setting or several feasible settings of the explanatory variables that provides the best compromise for the multiple responses simultaneously. Before the optimization is completed, m regression models must be obtained for each of the m response variables.

The MRO problem is illustrated by the minced fish quality example presented by[1]. For this example, a central composite design (CCD) was conducted with three processing factors (washing temperature, washing time, and washing ratio of water volume to sample weight). The four response variables related to minced fish quality are springiness, thiobarbituric acid number (TBA), cooking loss, and whiteness index. The goal of the study is to determine the optimum combination of the levels of the three factors so that springiness and whiteness index are maximized while TBA and cooking loss are minimized. This example will be analyzed in greater detail in Section 5.

If any of the m regression models are misspecified, the resulting estimates may be biased and, consequently, an estimated optimal solution will likely be incorrect and suboptimal. A Bayesian method[2] is to assess the effect of the uncertainty of the parameter estimates for all the responses so that a solution obtained is robust to the uncertainty in the form of the true models. But this method needs to assume that all the models for each response are true and have the same form. The seemingly unrelated regression (SUR) method where different responses can have different model forms may be useful when response variables are highly correlated[1]. But SUR is not preferable in those cases with a small sample size and/or with a low correlation among response variables because in these situations the covariance matrix used by SUR will likely experience increased sampling variability.

[3] point out that traditional parametric methods are often inadequate and suggest the use of nonparametric techniques. Unfortunately, in sparse data situations, which are typically the case with response surface experiments, nonparametric techniques often result in highly variable estimates. The model robust regression (MRR) estimation method, a semiparametric procedure, has been shown in[4] to be more robust to misspecification than either the parametric or nonparametric methods. MRR essentially combines the advantages from the parametric and nonparametric methods and avoids their disadvantages by simultaneously reducing both bias and variance of estimation. For the case of a single response,[5] has demonstrated that the MRR technique can be successfully applied to model the mean response for data from designed experiments.

Once the model building stage is completed, where each regression model built for each of the m responses variables is assumed to be appropriate, the optimization stage starts. One of the most popular and formal approaches to MRO is to use some specific function (an objective function) to combine the estimated responses so that the multiple dimensional problem can be transformed into a one dimensional problem. Examples for such specific functions are the desirability function by[6], the generalized distance measure function by[7], and the weighted squared error loss function by[8]. In this study, we consider only the desirability function method because it is easy to use and to interpret and readily available on standard statistical software.

To deal with the model misspecification problem in the multivariate case, we propose the use of MRR and adapt the MRR fits of each response to the desirability function method. A case study and simulations studies illustrate that MRR performs much better than the other two methods in most situations and its corresponding optimization results are often more accurate.

Since the desirability function method is a highly nonlinear function, we have proposed use of a genetic algorithm (GA), a global optimization tool, to help solve the MRO problem.[9] have developed a modified genetic algorithm (MGA) which has been shown to be more computationally efficient than GA and is therefore used in this study to find an optimal solution for the desirability function.

The remainder of this paper is organized as follows. Section 2 presents an overview the traditional parametric and nonparametric modeling methods and Section 3 summarizes the MRR technique, a semiparametric method. Bandwidth selection and mixing parameter selection associated with MRR are also introduced in this section. Section 4 describes the adaptation of the MRR to the desirability function method. Sections 5 and 6 compare the optimization results for the three modeling methods through a case study and simulation studies.

2 An Overview of Parametric and Nonparametric Approaches

Once the data are collected, our goal is to fit a model to estimate the true relationship between the explanatory variables and each response.

For the multiple response problem, we may use multivariate regression techniques (which are an extension of multiple linear regression for a single response) to model the relationships between the explanatory variables and the multiple responses simultaneously. However, in fact, the fits obtained by the regression techniques in the univariate case are equivalent to the fits obtained by the multivariate regression techniques, including the parametric[10], nonparametric and semiparametric methods, as discussed in[11]. Therefore, in this study, we model each response separately using the modeling techniques for a single response.

For each of the m response variables, suppose the true relationship between the k explanatory variables, $x_{1i}, x_{2i}, ...x_{ki}$, and the response, y_i, $i = 1, ..., n$, is

$$y_i = f(x_{1i}, x_{2i}, ..., x_{ki}) + \varepsilon_i, \tag{1}$$

where the function f represents the true relationship, n is the sample size, and ε_i represents a random error term from the process assumed to be independent, identically distributed, with mean zero and constant variance σ_i^2. Consequently, $E(y_i|x_{1i}, ..., x_{ki}) = \mu_i = f(x_{1i}, ..., x_{ki})$ is the true mean response function. It should be noted that the function f may be a different function of the same k regressors for each of the m response variables.

2.1 Parametric Approach

A parametric approach is to assume that the underlying mean response function can be expressed parametrically. We will consider only the linear model written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon \tag{2}$$

where $\mathbf{X}$ is a $n \times q$ model matrix with q regressors possibly including a column of ones in the model and $\boldsymbol{\beta}$ is a $q \times 1$ vector of unknown parameters. The number of regressors, q, $q \geq k$, consists of k original variables (the "main effects") and those functions of them, such as quadratic and cross-product terms, that are deemed to be essential to the model. Usually, the estimator of $\boldsymbol{\beta}$ can be obtained by the ordinary least squares (OLS) method as $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Thus, the OLS fit at location $\mathbf{x}_0' = (x_{10}, x_{20}, ..., x_{k0})$ is

$$\hat{y}_i^{(OLS)} = \breve{\mathbf{x}}_0'\hat{\boldsymbol{\beta}} = \breve{\mathbf{x}}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{h}_0^{(OLS)'}\mathbf{y}, \tag{3}$$

where $\breve{\mathbf{x}}_0' = (1, x_{10}, ..., x_{k0}, x_{(k+1)0}, ..., x_{(q-1)0})$, $\mathbf{h}_0^{(OLS)'} = \breve{\mathbf{x}}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and $\mathbf{X}' = [\breve{\mathbf{x}}_1, ..., \breve{\mathbf{x}}_n]$. The OLS estimated fits can be further obtained as

$$\hat{\mathbf{y}}^{(OLS)} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}^{(OLS)}\mathbf{y}, \tag{4}$$

where the $n \times n$ matrix $\mathbf{H}^{(OLS)}$ is known as the "HAT" matrix. For more details on the OLS and the HAT matrix, see[12;13].

Certain optimal properties of the OLS estimator are achieved if the observed y's are independent with constant variances across all observations. On the other hand, suppose the observed y's are independent but with different variances. That is, $\text{cov}(\mathbf{y}) = \text{cov}(\boldsymbol{\varepsilon}) =$

$\mathbf{V} = diag(\sigma_1^2, ..., \sigma_n^2) \neq \sigma^2\mathbf{I}$, where the $\mathbf{V}$ is a positive definite diagonal matrix. Then we consider instead that the estimates for $\boldsymbol{\beta}$ may be obtained by the weighted least squares (WLS) method as

$$\hat{\boldsymbol{\beta}}^{(WLS)} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}, \tag{5}$$

where $\mathbf{W} = \mathbf{V}^{-1}$ and the fitted value at $\mathbf{x}_0$ is given by

$$\hat{y}_0^{(WLS)} = \breve{\mathbf{x}}_0'\hat{\boldsymbol{\beta}}^{(WLS)} = \breve{\mathbf{x}}_0'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} \tag{6}$$

The $\mathbf{W}$ in (5) and (6) essentially represents a "global" weight matrix since the weights are unchanged cross all locations of $x_1, ..., x_k$. These global weights are different from "local" weights, which change at different values of $\mathbf{x}_0$, arbitrary settings of $x_1, ..., x_k$, including all values of $\mathbf{x}_i$, the observed settings of $x_1, ..., x_k$, $i = 1, 2, ..., n$. These "local" weights will be discussed in next section.

2.2 Nonparametric Approach

If the parametric function in (2) is incorrect in practice, then the parametric approach becomes inappropriate and a nonparametric approach may be an alternative choice.[14] suggests the use of nonparametric RSM (NPRSM) in the following three scenarios: (i) the main focus of the experiment is on optimization and not on parameter interpretation; (ii) there is less interest in an interpretive function and more interest in the shape of a response surface; or (iii) the functional form of the relationship between the explanatory variables and the response is highly nonlinear and not well behaved.[3;15;5] are some examples for nonparametric applications in RSM.

Unlike parametric approaches, nonparametric regression approaches make no assumptions about the parametric relationship between variables. The form of $\hat{f}(\mathbf{x}_0)$ solely depends on the data itself. There are several smoothness fitting techniques in the nonparametric regression literature such as kernel regression (also called the Nadaraya-Watson estimator), local polynomial regression (LPR), and spline-based regression. For details, see[16;17;18]. LPR is a popular class of nonparametric smoothing techniques and is particularly appealing in response surface applications due to its robustness to biased estimates at the boundary of the design space.

LPR is designed to fit a polynomial locally with a distance-based weighting scheme. Actually, similar to the parametric approaches which use the (global) weighted least squares method to obtain (global) parameters and estimates, LPR uses the local weighted least squares method[16] to obtain local parameters and estimates. The local polynomial may be zeroth-order (kernel regression), 1st- or higher-order. In our study, we focus on the 1st-order polynomial, which is commonly referred to the local linear regression (LLR).

The LLR fit at $\mathbf{x}_0' = (x_{10}, x_{20}, ..., x_{k0})$ is given by

$$\hat{y}_0^{(LLR)} = \tilde{\mathbf{x}}_0'\left(\tilde{\mathbf{X}}'\mathbf{W}_0\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\mathbf{W}_0\mathbf{y} = \mathbf{h}_0^{(LLR)'}\mathbf{y}, \tag{7}$$

where the local weight matrix $\mathbf{W}_0 = diag(h_{01}^{(KER)}, \cdots, h_{0n}^{(KER)})$, $h_{0i}^{(KER)} = \frac{K(\mathbf{x}_0, \mathbf{x}_i)}{\sum\limits_{i=1}^{n} K(\mathbf{x}_0, \mathbf{x}_i)}$ is a kernel weight associated with the distance of $\mathbf{x}_i'$ to $\mathbf{x}_0'$, $\tilde{\mathbf{x}}_0' = (1 \; x_{10} \; ... \; x_{k0}) = (1, \mathbf{x}_0')$. $K(\mathbf{x}_0, \mathbf{x}_i)$ is the kernel function, and the LLR model matrix, $\tilde{\mathbf{X}}$, is defined as $\tilde{\mathbf{X}}' = [\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, ..., \tilde{\mathbf{x}}_n]$, $\tilde{\mathbf{x}}_i' = (1 \; x_{1i} \; ... \; x_{ki})$. We use the so-called Gaussian kernel function in this work, $K(\mathbf{x}_0, \mathbf{x}_i)$, given by

$$K(\mathbf{x}_0, \mathbf{x}_i) \propto K\left(\left\|\frac{\mathbf{x}_0 - \mathbf{x}_i}{b}\right\|\right) \quad or \quad \prod_{j=1}^{k} K\left(\frac{x_{0j}-x_{ij}}{b}\right), \tag{8}$$

where $K\left(\frac{x_{0j}-x_{ij}}{b}\right) = e^{-\left(\frac{x_{0j}-x_{ij}}{b}\right)^2}$, $j = 1, ..., k$. b is is a specific bandwidth, also called the smoothing parameter, utilized to determine the smoothness of the estimates. The choice of the bandwidth is critical and will be discussed later. For details on the types of kernel functions and the multivariate kernel function, see[16;19;5].

The LLR fit in (7) is written in WLS format, except that the weight matrix is localized as it changes with each value of $\mathbf{x}_0$, unlike the global weight matrix $\mathbf{W}$ in (5) that is unchanged across all locations. In matrix notation, the LLR estimated fits may be expressed as

$$\hat{\mathbf{y}}^{(LLR)} = \mathbf{H}^{(LLR)}\mathbf{y}, \tag{9}$$

where $\mathbf{H}^{(LLR)}$, known as the LLR HAT matrix (or smoother matrix), is given by

$$\mathbf{H}^{(LLR)} = \left[\begin{array}{c} \mathbf{h}_1^{(LLR)\prime} \\ \mathbf{h}_2^{(LLR)\prime} \\ \vdots \\ \mathbf{h}_n^{(LLR)\prime} \end{array}\right], \tag{10}$$

where $\mathbf{h}_i^{(LLR)\prime} = \tilde{\mathbf{x}}_i'(\tilde{\mathbf{X}}'\mathbf{W}_i\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{W}_i$. It is easy to see from the formula above the estimate of mean response at any location, either $\mathbf{x}_i'$ (an observed data location) or $\mathbf{x}_0'$ (an unobserved data location) is associated with its special weight matrix, due to the local weighting scheme. The trace of $\mathbf{H}^{(LLR)}$ provides the model degrees of freedom for the LLR fit[18].

2.3 Parametric vs. Nonparametric

As mentioned earlier, both parametric and nonparametric methods have shortcomings. Parametric methods are inflexible in that a parametric function must be specified before fitting and if this model is incorrect, the resulting fits are subject to the consequence of model misspecification error such as bias in estimated mean response. Nonparametric methods are too flexible in that the resulting estimates of mean response completely depend on the observed data itself and often these fits are subject to high variance. In addition, the successful application of the nonparametric approach has usually been limited to fairly large sample sizes and space-filling designs. However, the typical characteristics of traditional RSM experiments, such as small sample size, sparse data, and most of design points on the edge of design space, all restrict the application of nonparametric approach.

3 A Semiparametric Approach

Semiparametric approaches combine a parametric method with a nonparametric method. One semiparametric method, model robust regression (MRR), proposed by[4], was originally developed for situations when there is partial knowledge about the underlying model, a situation very common in real life. Unlike the nonparametric method, MRR does not require a large sample and tends to work very well when the sample size is small[4;20;5]. It has been shown that MRR can improve estimates of mean response by combining both the parametric and nonparametric estimates into one estimate, simultaneously reducing both bias and variance of estimation. MRR essentially combines the advantages from the parametric and nonparametric methods and avoids their disadvantages.[5] has demonstrated that the MRR technique can be successfully applied to model mean response for data from designed experiments for the case of a single response.

MRR combines the parametric fit to the raw data with a nonparametric fit to the residuals from the parametric fit via a mixing parameter, λ. In this study, our MRR combines the parametric fit by the OLS method with the nonparametric fit by the LLR method. However, the MRR approach allows one to specify any other type of parametric and nonparametric methods for some special situations and conditions.

Our MRR fit at $\mathbf{x}_0$ is given by

$$\hat{y}_0^{(MRR)} = \breve{\mathbf{x}}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} + \lambda\tilde{\mathbf{x}}_0'(\tilde{\mathbf{X}}'\mathbf{W}_{r0}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{W}_{r0}\mathbf{r} = \mathbf{h}_0^{(OLS)'}\mathbf{y} + \lambda\mathbf{h}_{\mathbf{r0}}^{(LLR)'}\mathbf{r}, \quad (11)$$

where $\lambda \in [0,1]$, $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}^{(OLS)}$, $\mathbf{h}_{\mathbf{r0}}^{(LLR)'} = \tilde{\mathbf{x}}_0'(\tilde{\mathbf{X}}'\mathbf{W}_{r0}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{W}_{r0}$, and $\mathbf{W}_{r0}$, similar to $\mathbf{W}_0$ in (7), is a diagonal weight matrix for the residuals $\mathbf{r}$ at $\mathbf{x}_0$. In terms of HAT matrices, the equation above may be expressed as

$$\hat{\mathbf{y}}^{(MRR)} = \mathbf{H}^{(OLS)}\mathbf{y} + \lambda\mathbf{H}_r^{(LLR)}\mathbf{r} = \left[\mathbf{H}^{(OLS)} + \lambda\mathbf{H}_r^{(LLR)}(\mathbf{I} - \mathbf{H}^{(OLS)})\right]\mathbf{y} = \mathbf{H}^{(MRR)}\mathbf{y}, \quad (12)$$

where $\mathbf{H}_r^{(LLR)}$ is the LLR HAT matrix for fitting the residuals $\mathbf{r}$ from the parametric fits $\hat{\mathbf{y}}^{(OLS)}$ using the LLR method. The trace of $\mathbf{H}^{(MRR)}$, the MRR hat matrix, is the model degrees of freedom for the MRR fit.

As shown in (11) and (12) and as stated in[4], the amount of misspecification of the parametric model, and the amount of correction needed from the residual fit, is reflected in the size of λ. The philosophy of MRR is that the user, having examined the data and with subject matter knowledge, proposes their best parametric model. Any structure in the data not captured by the user's model is contained in the residuals. A nonparametric fit to these residuals describes this remaining structure. The nonparametric fit, however, may be following these residuals a bit too closely, fitting error variability along with the remaining model structure. The λ parameter is needed to dampen the impact of the extra variability in the nonparametric fit to the residuals.

If the parametric fit is adequate, then λ should be chosen close to zero by some appropriate λ selector (which will be discussed later). On the other hand, if the parametric fit is inadequate, then λ will be chosen large enough (close to one) so that the nonparametric fit to the OLS residuals can be used to make up for the parametric fit's inadequacy. In practical applications, the user does not know the true underlying function and, consequently, does not know the amount of model misspecification. Thus, the MRR method provides an alternative estimation procedure that is robust to the model misspecification that may be present in the user's proposed model and to the high variability that may be present in a nonparametric method.

3.1 Choice of the Smoothing Parameter b

As in all kernel-based nonparametric methods, MRR requires the selection of the smoothing parameter b to be used by the method when fitting the residuals from the parametric model. In this study, since LLR will be used as the nonparametric method or as part of the semiparametric method to fit the residuals, the following discussion on the choice of the bandwidth will be related to LLR using y as the response values. It is easy to extend the data-driven method for the choice of bandwidth to the nonparametric part of MRR by considering residuals as response values.

The choice of bandwidth is crucial in obtaining a "proper" estimate of function f. Any suitable criterion to deal with the trade-off between bias and variance such as the mean squared error (MSE) may be used here to select an appropriate bandwidth. For a thorough discussion of bandwidth selectors, see[16;21]. There are several bandwidth selectors

such as MSE, PRESS (prediction error sum of squares), PRESS* (a modified PRESS with an adjustment in its denominator), PRESS** (another modified PRESS with two adjustments in its denominator), the generalized cross-validation (GCV) and Akaike's Information criterion (AIC). Among them, PRESS** has been shown in[22;20] to be the best choice in terms of minimizing integrated mean squared error of fit across a broad variety of data scenarios (including some used in our simulation study). Consequently, we use PRESS** as a bandwidth selector in this study.

The PRESS** is given by

$$
\begin{aligned}
PRESS^{**}(b) &= \frac{\sum (y - \hat{y}_{i,-i}(b))^2}{n - trace(\mathbf{H}^{(LLR)}(b)) + (n - k - 1)\frac{SSE_{\max} - SSE_b}{SSE_{\max}}} \\
&= \frac{PRESS(b)}{n - trace(\mathbf{H}^{(LLR)}(b)) + (n - k - 1)\frac{SSE_{\max} - SSE_b}{SSE_{\max}}},
\end{aligned}
\tag{13}
$$

where $SSE_{\max}$ is the largest error sum of squares over all possible bandwidth values (essentially, $SSE_{\max}$ is the parametric SSE by OLS that results when b tends to infinity) and SSE_b is the error sum of squares associated with a specific bandwidth value b.

3.2 Choice of the Mixing Parameter λ in MRR

As the sample size n goes to infinity, the estimated asymptotically optimal value (see[4]) of the mixing parameter for MRR is given by

$$
\hat{\lambda} = \frac{\left\langle \hat{\mathbf{r}}^{(LLR)}, \mathbf{y} - \hat{\mathbf{y}}^{(OLS)} \right\rangle}{\left\| \hat{\mathbf{r}}^{(LLR)} \right\|^2},
\tag{14}
$$

where $\langle \rangle$ represents the inner product and $\|\|$ represents the standard L_2 (Euclidean) norm. The examples in[4] show that this method works well for small sample sizes.

4 Desirability Function Method with Modeling Techniques

After the model building stage is completed where each regression model built for each response is assumed to be appropriate, the MRO techniques can then be used. As mentioned in the introduction, in this study, the desirability function method, one of the most popular multi-response optimization techniques, is used to find the best compromise among the multiple responses simultaneously.

The desirability function method, proposed by[6], transforms each response into an individual desirability value through a desirability function and then combines these individual desirabilities into one desirability value using a geometric mean. The desirability function (considered here as an objective function utilized for optimization) at location $\mathbf{x}_0$ is given by

$$
D(\mathbf{x}_0) = [d_1(\hat{y}_1(\mathbf{x}_0)) \times d_2(\hat{y}_2(\mathbf{x}_0)) \times \cdots \times d_M(\hat{y}_M(\mathbf{x}_0))]^{1/M} \times 100\%,
\tag{15}
$$

where M is the number of response variables, $m = 1, ..., M$, and $d_m()$[6] is the mth individual desirability function and transforms the predicted value for the mth response at $\mathbf{x}_0$, $\hat{y}_m(\mathbf{x}_0)$, into a dimensionless individual desirability value, scaled within $[0, 1]$. Obviously, the range of D is $[0, 1]$. The individual desirability function is a discrete function with some

pre-specified parameters, such as L (the lower limit), U (the upper limit), and T (the target value). The individual desirability functions are presented in the appendix. For more details about the use of the desirability function and their parameters, see[6;23]. The predicted value for the mth response, $\hat{y}_m$, could be obtained from either a parametric, nonparametric or semiparametric fit. The goal of the method is to find the maximum of the overall desirability D and its associated optimal location(s).

A case study and a simulation study are used to compare MRR fits with OLS and LLR fits during the modeling stage. Then, these resulting fits are compared during the optimization stage using the desirability function. A modified genetic algorithm[9] is used to find an optimal solution for the desirability function under the three different modeling methods.

To compare the three modeling methods, some numerical criteria are utilized: 1) DF_{error}, the degree of freedom of error, is given by $DF_{error} = n - tr(\mathbf{H})$[18]; 2) s^2, estimate of error variance; 3) R^2, coefficient of determination; 4) R^2_{adj}, adjusted R^2; 5) PRESS** (using (13) with estimated bandwidth for LLR and with estimated bandwidth and lambda for MRR). Criteria 1-4 focus on describing how well a model is fit by the observed data, while Criteria 5 focus on describing the prediction capability associated with the fitted model. Criteria 2-4 are standard criteria for comparing models[12]. The DF_{error} can be used to compute the degrees of freedom for the model as $DF_{model} = DF_{total} - DF_{error}$. DF_{model} represents the complexity of the model. DF_{total} is the total number of degrees of freedom, equal to n, the sample size.

5 The Minced Fish Quality Example

To illustrate our method, we return to the minced fish quality example presented in the Introduction. A CCD was conducted with a total of 17 design points, including 2^3 runs in the factorial region, augmented with six axial runs, and three center runs.

[1] use the second-order polynomial parametric method to model each response to obtain the optimal fitted value $\hat{y}_i(\mathbf{x})$ at location $\mathbf{x}$, for $i = 1, 2, 3, 4$. The final fitted second-order models for the four responses by OLS are given in[1]. For the responses y_1 and y_4, the final fitted models include three terms: intercept, x_1 and x_1^2. The final model for the response y_2 includes five terms: intercept, x_1, x_2, x_1^2 and x_{12}. The model for y_3 has a total of eight terms, including intercept, x_1, x_2, x_3, x_1^2, x_{12}, x_{13}, and x_3^2. The natural independent variables are transformed into the coded variables within the range of [0, 1]. For each response, the design spaces we use for the LLR and the nonparametric part of the MRR are the same as the ones used for the OLS by[1].

5.1 Results on Model Comparisons

During the modeling stage, Table 1 shows the numerical results for model comparisons of OLS, LLR, and MRR for all the responses respectively with respect to the five criteria as previously mentioned. Table 1 shows that MRR has smaller, often substantially smaller, s^2 than OLS and LLR across all of the four responses. Table 1 shows that MRR has smaller, often substantially smaller, s^2 and larger, often substantially larger, R^2_{adj} than OLS and LLR across all of the four responses. MRR has larger R^2 than OLS and LLR in three of the four responses. MRR has smaller PRESS** than OLS and LLR in three of the four cases. Furthermore, in the case where MRR does not perform the best among the three modeling techniques, MRR still is very competitive to the best modeling technique in terms of all the five criteria.

As can be seen from Table 1, the models obtained by OLS use fewer model degrees of freedom than the MRR method. It should be pointed out, however, that the OLS results cannot be improved by adding other first and second order terms. The OLS model is clearly misspecified but this misspecification cannot be captured by using lower order polynomial terms. MRR has improved the OLS fits by utilizing all of the unspecified model structure found by LLR in the OLS residuals.

It also should be noted that using LLR on the raw data results in very poor fits to several of the responses, indicating the difficulty of a nonparametric method in fitting sparse data.

[**Table 1 here**]

5.2 Optimization Results

During the optimization stage, the desirability function method is used to obtain the best compromise of the four responses. In the individual desirability function, the pre-specified parameters (see the appendix for the parameters of the desirable function) are set as follows. The maximum (or minimum) of the observed data is used as the T value for each response since they are not available in[1]. That is, in this example, $T_1 = 1.92$, $T_2 = 20.16$, $T_3 = 16.80$, and $T_4 = 50.98$. The limited boundaries for each response are $L_1 = 1.7$, $U_2 = 21$, $U_3 = 20$, and $L_4 = 45$, for y_1 through y_4, respectively. The weights r, r_1, or r_2 are all 1.0 in this case study, since no priorities are given in[1].

Since a CCD was utilized in the example, the solution vector $\mathbf{x}_s$ shall be constrained to be within the experimental region R, a hyper-circle in this 3-dimensional example. As mentioned previously, the natural independent variables are transformed into the coded variables within the range of $[0, 1]$. Therefore, the solution vector $\mathbf{x}_s$ is defined as $(x_1 - 0.5)^2 + (x_2 - 0.5)^2 + (x_3 - 0.5)^2 \leq 0.5^2$ in the transformed experimental region.

As mentioned earlier,[1] use the second-order polynomial to parametrically model each response to obtain the optimal fitted value $\hat{y}_i(\mathbf{x})$ at location $\mathbf{x}$, for $i = 1, 2, 3, 4$. The final fitted models are given in[1] as well as the location where the simultaneous optimal solution is found. Based on the location they found using Design-Expert, the corresponding fitted values for the four responses are re-calculated by us as well as the desirability value D and given as follows.

The OLS solution: $x_1 = 0.3514$ $x_2 = 0.7973$ $x_3 = 0.7319$
$\hat{y}_1 = 1.9074$ $\hat{y}_2 = 20.2910$ $\hat{y}_3 = 17.6381$ $\hat{y}_4 = 49.8346$ $D = 0.8301$

We use the MGA to find the optimization solutions by the three different modeling techniques. The parametric models we used for each response are exactly the same as the models in[1]. The solutions we find by the OLS, LLR and MRR methods are as follows.

The OLS solution: $x_1 = 0.3857$ $x_2 = 0.9693$ $x_3 = 0.6784$
$\hat{y}_1 = 1.9067$ $\hat{y}_2 = 19.7378$ $\hat{y}_3 = 17.3903$ $\hat{y}_4 = 50.4668$ $D = 0.9149$

The MRR solution: $x_1 = 0.3379$ $x_2 = 0.9422$ $x_3 = 0.7080$
$\hat{y}_1 = 1.8947$ $\hat{y}_2 = 19.5246$ $\hat{y}_3 = 17.7507$ $\hat{y}_4 = 51.3969$ $D = 0.8880$

We do not show the LLR solution, because the desirability function value D = 0 since the LLR fits can be highly inadequate.

It is easy to see that our OLS solution found by the MGA is better than the OLS solution found by Design-Expert in[1]. The OLS solution should not be compared to the MRR solution, because the OLS solution has been shown in Table 1 to be misspecified. Consequently, the MRR solution should be viewed as more reliable than the OLS solution,

because the MRR method provides a superior fit to the data than the OLS method in terms of our model comparison criteria.

6 Simulation Studies

In this section, the MRO problem is simulated and the three methods compared via Monte Carlo methods.

6.1 The MRO Goals and Simulation Process

To simulate the MRO problem, a CCD design with two factors and two responses are simulated using two true underlying response functions. The CCD contains a total of 13 design points, including four axial runs and five center runs for each simulated data set. A similar design is used in the chemical process example in[23].

Each Monte Carlo simulation is based on the following two underlying models:

$$
\begin{aligned}
y_{1i} &= \mu_1(\mathbf{x}_i) + \varepsilon_{1i} \\
&= 66 + 22x_{1i} + 10x_{2i} + 13x_{1i}x_{2i} - 23x_{1i}^2 - 25x_{2i}^2 \\
&\quad + \gamma[-2sin(3\pi x_{i1}) - 2cos(3\pi x_{2i}) + 2sin(2\pi x_{1i}x_{2i})] + \varepsilon_{1i},
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
y_{2i} &= \mu_2(\mathbf{x}_i) + \varepsilon_{2i} \\
&= 70 - 15x_{1i} - 10x_{2i} - 14x_{1i}x_{2i} + 15x_{1i}^2 + 25x_{2i}^2 \\
&\quad + \gamma[2sin(3\pi x_{1i}) - 2cos(3\pi x_{2i}) + 2sin(3\pi x_{1i}x_{2i})] + \varepsilon_{2i}.
\end{aligned}
\tag{17}
$$

where $\mu_1(.)$ and $\mu_2(.)$ are the true mean functions of y_1 and y_2, respectively, $\mathbf{x}_i = (x_{1i}, x_{2i})'$ is the ith design point, the two error terms, $(\varepsilon_{1i}, \varepsilon_{2i})$, are independent normally distributed random variables with means of zero and variances of one, $i = 1, ..., n$, with $n = 13$. In both models (16) and (17), γ represents the model misspecification parameter. That is, the user's models are represented by the second-order polynomial model with $\gamma = 0$. As the value of γ increases, the amount of misspecification increases in the models. Five degrees of model misspecification are studied ($\gamma = 0.00, 0.25, 0.5, 0.75,$ and 1.00), where $\gamma = 0$ represents a correctly specific model. According to these five levels of γ, there are five Monte Carlo simulations respectively, in each of which 500 simulated data sets are generated.

Figure 1 shows the surfaces of the true mean function of the response y_1 when $\gamma = 0.00, 0.25, 0.50, 0.75,$ and 1.00, respectively. Similarly, Figure 2 shows the surfaces of the true mean function of the response y_2. Both figures show that as γ increases, the curvatures of the surfaces becomes more pronounced indicating the increasing nature of the model misspecification in the user's quadratic model.

[**Figure 1 here**]

[**Figure 2 here**]

Two MRO goals are used for each Monte Carlo simulation. Goal 1 is to maximize y_1 and minimize y_2 simultaneously. Goal 2 is to achieve target values for y_1 and y_2 simultaneously. Each goal is solved by the desirability function method with the pre-specified parameters values for T, L, and U (as defined in the Appendix). In Goal 1, we choose $T_1 = 83$ and $L_1 = 65$ for y_1, and $T_2 = 58$ and $U_2 = 75$ for y_2, while in Goal 2, we choose $T_1 = 60$, $L_1 = 55$, and $U_1 = 65$ for y_1 and $T_2 = 75$, $L_2 = 70$, and $U_2 = 80$ for y_2. All these parameters are fixed across all levels of γ.

6.2 Simulation Criterion During The Modeling Stage

After completion of the data generation stage, the modeling stage begins using the OLS, LLR and MRR methods respectively to model the two responses. Similar to the univariate case in[5], for each simulation, we compare the MRR method with the OLS and LLR methods in terms of simulated integrated mean squared error (SIMSE). The SIMSE in the multivariate case is given by:

$$SIMSE = \frac{\sum ASE}{500}, \tag{18}$$

where

$$ASE = \frac{\sum_{l=1}^{s} (\mu(\mathbf{x}_l) - \hat{\mathbf{y}}(\mathbf{x}_l))'(\mu(\mathbf{x}_l) - \hat{\mathbf{y}}(\mathbf{x}_l))}{s}, \tag{19}$$

where ASE denotes the average squared error for the estimates from the true mean functions for each of the 500 simulated data sets, $\mu(\mathbf{x}_l) = (\mu_1(\mathbf{x}_l), \mu_2(\mathbf{x}_l))'$ is the true mean functions of y_1 and y_2 at location $\mathbf{x}_l = (x_{1l}, x_{2l})'$, $\hat{\mathbf{y}}(\mathbf{x}_l) = (\hat{y}_1(\mathbf{x}_l), \hat{y}_2(\mathbf{x}_l))'$ is the fits at $\mathbf{x}_l$ by OLS, LLR, or MRR, $l = 1, ..., s$, and s is the number of locations within the experimental space used for prediction based on the 41×41 uniform grid of points. Those points outside of the experimental space are excluded. In this study, $s = 1257$. SIMSE, a measure we want to be as small as possible, provides an indication of the fit performance of each of the three methods over the entire design space.

6.3 Simulation Results During The Modeling Stage

Table 2 provides a comparison of the OLS, LLR, and MRR based on the SIMSE values for the varying degrees of model misspecification in the simulations based on the CCD. For the scenario in which the researchers correctly specify the form of the underlying models (i.e., $\gamma = 0.00$), we would expect the parametric approach to be superior. The first row of Table 2 shows that OLS has the mininum SIMSE, MRR is a close second, while LLR provides an inferior fit.

[**Table 2 here**]

The remaining rows of Table 2 provide the SIMSE values for the scenario in which the researchers misspecify the models (i.e., $\gamma > 0$). As expected, MRR performs better than OLS and LLR with smaller SIMSE values through each non-zero degree of misspecification.

Table 2 also includes the Monte Carlo errors of the SIMSE values. Tukey's multiple comparison method is used to determine significant differences among the three methods at the overall type I error rate of 0.05. When $\gamma = 0.00$ and 0.25, there are no significant differences between OLS and MRR. When $\gamma = 0.50$, 0.75 and 1.00, MRR is significantly smaller than OLS. In addition, LLR is significantly larger than OLS and MRR through all of the five degrees of model misspecification.

As mentioned in (18), each SIMSE value which indicates a simulated integrated mean squared error for the estimates from the true mean functions for each of the 500 simulated data sets accounts for both bias and variance. A small SIMSE value means that both bias and variance of estimation are small. Thus, the SIMSE results along with the Monte Carlo errors, as shown in Table 2, suggest that the semiparametric approach is highly competitive to the parametric approach when no model misspecification occurs,

and can be superior to both the parametric and nonparametric approaches, simultaneously reducing both bias and variance of estimation, when there exists some moderate model misspecification.

6.4 Simulation Criteria During The Optimization Stage

After completion of the modeling stage, the optimization stage begins using the desirability function approach to find an optimal solution. In practice, the optimization stage is concerned with finding a location where both responses are simultaneously close to their respective optimal values. This may be a location different for where each response is individually optimal. Therefore, there are two criteria used to compare the three methods, both of which are related to an optimal location found by our MGA through the desirability function for each simulated data set.

For each Monte Carlo simulation, the three methods are compared in terms of the average squared error loss (ASEL) from the true target response values. The ASEL is given by

$$ASEL = \frac{\sum_{q=1}^{q=500} (\mu(\mathbf{x}_q^*) - \mathbf{T})'(\mu(\mathbf{x}_q^*) - \mathbf{T})}{500}, \tag{20}$$

where $\mu(\mathbf{x}_q^*)$ is a 2×1 vector of the values of the true mean function at an optimal location $\mathbf{x}_q^*$, the location $\mathbf{x}_q^* = (x_{1q}^*, x_{2q}^*)'$ is obtained by MGA through the desirability function for the qth simulated data set, $q = 1, ..., 500$, and $\mathbf{T}$ is a 2×1 vector of the target values for the responses. For Goal 1, as previously mentioned, $\mathbf{T} = (83, 58)'$, and for Goal 2, $\mathbf{T} = (60, 75)'$.

For each Monte Carlo simulation, the three methods also are compared in terms of average desirability function (AD), which is given by

$$AD = \frac{\sum_{q=1}^{500} D_q}{500}, \tag{21}$$

where

$$D_q = (d_1(\mathbf{x}_q^*)d_2(\mathbf{x}_q^*))^{1/2}, \tag{22}$$

and similar to (20), $\mathbf{x}_q^*$ is an optimal location obtained by our MGA through the desirability function for the qth simulated data set, d_1 and d_2 are individual desirabilities for y_1 and y_2, respectively. Essentially, both ASEL and AD measure the performance of the locations chosen by each modeling method. We measure with ASEL the average Euclidean distance of the mean response vector from the target vector at the chosen locations for each method. And, with AD we measure the average desirability of the chosen functions. Of course, we prefer a method with a small value of ASEL and a large value of AD.

Graphs of the desirability functions, using the two true mean functions shown in (16) and (17), with the different levels of γ, have been obtained but not presented here due to limited space. All the figures show that the surfaces have only a single optimal solution for each level of γ.

6.5 Simulation Results During The Optimization Stage

Table 3 provides a comparison of OLS, LLR, and MRR in terms of the ASEL and AD values for the varying degrees of model misspecification for Goal 1 in the simulations based on CCD during the optimization stage, followed by the estimated Monte Carlo errors of

ASEL and AD in parentheses, respectively. When $\gamma = 0.00$ or 0.25, OLS has the smallest ASEL values and the largest AD values. When γ is 0.50 or 0.75, MRR has the smallest ASEL values and the largest AD values. At γ equal to 1.00, LLR has the smallest and largest values of ASEL and AD, respectively.

Tukey's multiple comparison method is used to determine significant differences among the three methods at the overall type I error rate of 0.05. When $\gamma = 0.00$, 0.25, and 0.50, there are no significant differences between OLS and MRR, and LLR is significantly smaller than OLS and MRR in ASEL and significantly larger than OLS and MRR in AD. When $\gamma = 0.75$, there are no significant differences among the three methods in ASEL and MRR is significantly smaller than the other two in AD. When $\gamma = 1.0$, the three methods are significantly different from each other in both ASEL and AD.

[**Table 3 here**]

As seen from Tables 2 and 3, those methods with the smallest values of SIMSE from the fitting stage tend to have the smallest values of ASEL and the largest values of AD from the optimization stage. It is also observed that in those cases where the SIMSE values for a particular method are very poor, for example, the LLR results when $\gamma = 0.00$, that the ASEL and AD results for that method are very competitive with the best methods. Justification for this phenomenon is provided by plots (not shown here) of the estimated mean functions obtained by OLS, LLR, and MRR for y_1 and y_2 across various values of x_1 and x_2. When compared to the true mean functions, these plots indicate that despite the relatively poor fit by LLR, the locations where the maximum occurs for y_1 and the minimum for y_2 are very close to the appropriate locations for the OLS and MRR fits, resulting in relatively small ASEL and large AD values.

Similar to Table 3, Table 4 provides a comparison of the three modeling methods in terms of the ASEL and AD values for the varying degrees of model misspecification for Goal 2, followed by the Monte Carlo errors of ASEL and AD in parentheses, respectively. When $\gamma = 0.00$, LLR performs better than OLS and MRR since it has the smallest ASEL value and the biggest AD value. But LLR only performs slightly better, since actually, all of the three methods have nearly equal values of ASEL and AD. When $\gamma > 0.00$, however, MRR performs better than the other two, since it has smaller ASEL values and bigger AD values.

[**Table 4 here**]

Tukey's multiple comparison method reveals that at $\gamma = 0.00$, there are no significant differences among the three methods. When $\gamma = 0.25$, MRR and LLR are significantly better than OLS in ASEL. When $\gamma = 0.50, 075$, and 1.0, MRR is significantly better than OLS and LLR in both ASEL and AD.

For both Goals 1 and 2, as shown in Tables 3 and 4, the optimization results by the semiparametric fit are highly competitive to the results by the other two methods when there is no or low model misspecification, and superior or highly competitive to the results for higher levels of misspecification. Thus, we can conclude that the optimization results by the semiparametric approach are more reliable than the ones by the parametric and nonparametric approaches, in general, at least for the situations examined in our simulation.

7 Conclusion and Discussion

RSM has utilized parametric regression techniques to study products and processes since its inception. One drawback, however, is that optimization depends too heavily on the assumption of well-estimated models for the responses of interest, and it is often the case that the user's specified parametric models are not flexible enough to adequately model the process. Nonparametric smoothing has been considered when the user is unable to specify the explicit form for the underlying function. However, in small sample settings, which are customary for response surface experiments, the nonparametric approach often produces estimates that are highly variable. Therefore, we suggest MRR, a semiparametric approach, which combines a parametric method with a nonparametric method. MRR combines the advantages from both the parametric and nonparametric methods and reduces some of their disadvantages (high bias mainly from the parametric method, and large variance mainly from the nonparametric method).

In the minced fish quality example of the MRO problem, the results show that MRR is superior to OLS and LLR in terms of the five model comparison criteria through all of the responses. During the optimization stage, the models by the OLS, LLR, and MRR are assumed to be correct respectively, and the desirability function method has been utilized to find the optimal solutions with the best compromise among the responses. Although the optimal solutions by the three modeling methods are incomparable directly, the optimization results by MRR is more reliable, because the MRR model appears to have less misspecification (i.e., both lower bias and variances) than either the parametric and nonparametric methods based on the model comparison results.

Simulation studies based on a CCD with different degrees of model misspecification were conducted to compare the three approaches more generally. If the user correctly specifies the model for each response of interest, the parametric approach yields the best fit in terms of SIMSE and its corresponding optimization results are the best for Goal 1 and highly competitive to the results by the other methods for Goal 2 in terms of ASEL and AD. If there exists some moderate model misspecification, the semiparametric approach always yields the best fit in terms of SIMSE and its corresponding optimization results are superior or highly competitive to the ones by the other two methods for Goal 1, and are always the best for Goal 2. Thus, we can conclude that the semiparametric approach consistently performs well in general and has less misspecification in terms of SIMSE and its corresponding optimization results are more reliable than the other two.

Since, in practice, one never knows if the form of the underlying model for each response of interest has been correctly specified, we advocate MRR as it is the only method among those considered in this paper which consistently performs well over all degrees of potential misspecification and its corresponding optimization results are more reliable. Although MRR is slightly more complex than OLS, R code, available from authors, is easily implemented and greatly reduces the complexity.

The philosophy of MRR is that if the user's model is a very good descriptor of the data, $\hat{\lambda}$ will be close to zero, indicating that the nonparametric component contributes little toward improving the parametric fit. If the user's model is a very poor descriptor of the data, then $\hat{\lambda}$ will be close to one. This indicates that the MRR fit is substantially improved by using the nonparametric component.

The real benefit of MRR occurs when the user's model provides an adequate, but not wholly satisfactory fit to the some regions of the data. In this case, $\hat{\lambda}$ will be somewhere in the middle between zero and one and the fits obtained via MRR can be far superior to the fits obtained by either the parametric or the nonparametric method.

Of course, if $\hat{\lambda}$ is close to one, nothing prevents the user from adjusting the parametric

model, perhaps by adding higher order terms or by changing the model in some other appropriate way, and recalculating the MRR fit. If the new parametric model is an improvement, the new MRR fit will also be an improvement. Should $\hat{\lambda}$ now be close to zero, the fit from the revised parametric model may be chosen as the final fit. Thus, MRR may be used as an exploratory tool as part of the model building process.

Consequently, as a strategy for using MRR, it is recommended that the user build the parametric component of MRR by using the best regression modeling techniques known to them. MRR can then be used to improve an adequate but not wholly satisfactory parametric model fit or to help confirm a reasonable parametric model.

Several issues remain for further study. Procedures that work well when data are dense and uniformly spaced over the design region may need adjustment for the RSM problem where the design points are sparse with most of the design points on the edge of the design space and where the sample size is small. Examples include; 1) the selection of various parameters for the nonparametric methods (bandwidth selection for kernel-based methods or knot selection for penalized-spline methods); 2) the adjustment of the parametric and nonparametric methods for dealing with outliers in the response variables; 3) adjustment of standard RSM designs, appropriate of correctly specified parametric models, but unlikely appropriate for use with misspecified parametric models nor appropriate in most cases for use with nonparametric regression and, consequently, with MRR.

References

[1] Shah KH, Montgomery DC, Carlyle WM. Response surface modeling and optimization in multiresponse experiments using seemingly unrelated regressions. *Quality Engineering* 2004; 16 (3): 387-397.

[2] Peterson JJ. A Bayesian reliability approach to multiple response surface optimization. *Journal of Quality Technology* 2004; 36: 139-153.

[3] Vining GG, Bohn LL. Response surface for the mean and variance using a nonparametric approach. *Journal of Quality Technology* 1998; 30: 282-291.

[4] Mays JE, Birch JB, Starnes BA. Model robust regression: combining parametric, nonparametric, and semiparametric methods. *Journal of Nonparametric Statistics* 2001; 13: 245-277.

[5] Pickle SM, Robinson TJ, Birch JB, Anderson-Cook CM. A Semi-parametric approach to robust parameter design. *Journal of Statistical Planning and Inference* 2008; 138: 114-131.

[6] Derringer GC, Suich R. (1980). Simultaneous Optimization of Several Response Variables. *Journal of Quality Technology* 1980; 12: 214-219.

[7] Khuri AI, Conlon M. Simultaneous Optimization of Multiple Responses Represented by Polynomial Regression-Functions. *Technometrics* 1981; 23: 363-375.

[8] Vining GG. A Compromise Approach to Multi-response Optimization. *Journal of Quality Technology* 1998; 30: 309-313.

[9] Wan W, Birch JB. An Improved Genetic Algorithm Using a Derivative-free Directional Search. Unpublished 2009.

[10] Johnson DE. *Applied Multivariate Methods for Data Analysis*. Duxbury Press, Pacific Grove, CA. 1998.

[11] Wan W. Semi-Parametric Techniques for Multi-Response Optimization. Ph.D. Dissertation. Department of Statistics, Virginia Polytechnic Institute & State University, Blacksburg, VA. 2007.

[12] Myers RH. *Classical and Modern Regression with Applications* PWS-Kent, Boston, MA, 1990.

[13] Rencher AC. *Method of multivariate analysis.* John Wiley and Sons, Inc., New York, 2002.

[14] Myers RH. Response Surface Methodology—Current Status and Future Directions. *Journal of Quality Technology* 1999; 31(1): 30-44.

[15] Anderson-Cook CM, Prewitt K. Some Guidelines for Using Nonparametric Methods for Modeling Data from Response Surface Designs. *Journal of Modern Applied Statistical Methods* 2005; 4: 106-119.

[16] Härdle W. *Applied Nonparametric Regression.* Cambridge Univ. Press, London, 1990.

[17] Takezawa K. *Introduction to Nonparameteric Regression.* John Wiley and Sons, Inc., New Jersey, 2006.

[18] Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression.* Cambridge Series in Statistical and Probabilistic Mathematics. 2003.

[19] Scott DW. *Multivariate Density Estimation: Theory, Practice, and Visualization.* Wiley, New York, 1992.

[20] Mays JE, Birch JB. (2002). Smoothing for Small Samples with Model Misspecification: Nonparametric and Semiparametric Concerns. *Journal of Applied Statistics.* 2002; 29: 1023-1045.

[21] Härdle W, Muller M, Sperlich S, Werwatz A. *Nonparametric and Semiparametric Models.* Springer, Berlin, 2004.

[22] Mays JE, Birch JB. Smoothing Considerations in Nonparametric and Semiparametric Regression. Technical Report Number 98-2. Department of Statistics, Virginia Polytechnic Institute & State University, Blacksburg, VA, 1998.

[23] Myers RH, Montgomery DC. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments.* John Wiley and Sons, Inc., New York, NY, 2002.

APPENDIX: Individual Desirability

The following details concern the individual desirability functions and their parameters.

As mentioned previously, $\hat{y}_i(\mathbf{x})$ is needed to transform into a dimensionless individual desirability, d_i, in the desirability function. There are two cases for transformation to consider: one-sided and two-sided transformations. One-sided transformations are used when the goal is to either maximize the response or minimize the response. Two-sided transformations are used when the goal is for the response to achieve some specified target value. When the goal is to maximize the ith response, the individual desirability is given by the one-sided transformation

$$d_i = \begin{cases} 0 & \hat{y}_i(\mathbf{x}) < L \\ \left[\frac{\hat{y}_i(\mathbf{x}) - L}{T - L} \right]^r & L \leq \hat{y}_i(\mathbf{x}) \leq T \\ 1 & \hat{y}_i(\mathbf{x}) > T \end{cases} , \qquad (23)$$

where T represents an acceptable maximum value, L represents the acceptable minimum value and r is known as a "weight", specified by the user. Similarly, when the goal is to minimize the ith response, the corresponding individual desirability is written as the one-sided transformation

$$d_i = \begin{cases} 1 & \hat{y}_i(\mathbf{x}) < T \\ \left[\frac{U - \hat{y}_i(\mathbf{x})}{U - T} \right]^r & T \leq \hat{y}_i(\mathbf{x}) \leq U \\ 0 & \hat{y}_i(\mathbf{x}) > U \end{cases} , \qquad (24)$$

where T is an acceptable minimum value and U is the acceptable maximum value. When the goal is to obtain a target value, the individual desirability is given by the two-sided transformation

$$
d_i = \begin{cases}
0 & \hat{y}_i(\mathbf{x}) < L \\
\left[\frac{\hat{y}_i(\mathbf{x})-L}{T-L}\right]^{r_1} & L \leq \hat{y}_i(\mathbf{x}) \leq T \\
\left[\frac{U-\hat{y}_i(\mathbf{x})}{U-T}\right]^{r_2} & T \leq \hat{y}_i(\mathbf{x}) \leq U \\
0 & \hat{y}_i(\mathbf{x}) > U
\end{cases}, \tag{25}
$$

where T is the target value, and L and U are the acceptable minimum and maximum values respectively, and r_1 and r_2 are weights, specified by the users.

This desirability function D offers the user great flexibility in the setting of the desirabilities due to allowing users to chose appropriate values of L, U, and T, and of r, r_1, and r_2, for their different specific situations. For more details on the desirability function, see, for example, Derringer and Suich (1980) and Myers and Montgomery (2002).

Table 1: Results on model comparisons of OLS, LLR, and MRR with two different methods for λ selection for all the responses in the minced fish quality example. Best values in bold.

		b	$\hat{\lambda}$	DF_{error}	s^2	R^2	R^2_{adj}	PRESS**
	OLS	—	—	14.000	1.653E-03	0.9211	0.9099	0.0042
y_1	LLR	0.146	—	12.138	1.039E-03	**0.9570**	0.9433	0.0026
	MRR	0.170	1	12.268	**1.033E-03**	0.9568	**0.9436**	**0.0025**
	OLS	—	—	12.000	7.5417	0.9341	0.9122	19.5097
	LLR	0.436	—	11.212	21.8508	0.8217	0.7456	36.4222
y_2	MRR	0.277	1	8.940	**4.8253**	**0.9686**	**0.9438**	19.6311
	OLS	—	—	9.000	4.5641	0.8408	0.7170	20.2719
y_3	LLR	0.537	—	8.373	9.7990	0.6821	0.3925	17.0554
	MRR	0.542	1	6.596	**2.9031**	**0.9258**	**0.8200**	**13.1264**
	OLS	—	—	14.000	14.2182	0.5407	0.4751	48.9101
y_4	LLR	0.120	—	12.031	1.0197	0.9717	0.9624	**17.1484**
	MRR	0.119	1	12.029	**1.0158**	**0.9718**	**0.9625**	18.6472

Table 2: Simulated integrated mean squared error (SIMSE) values by OLS, LLR, and MRR in the simulations based on CCD and the estimated Monte Carlo (MC) error of SIMSE. Best values in bold.

γ	SIMSE			MC error(SIMSE)		
	OLS	LLR	MRR	OLS	LLR	MRR
0.00	**0.6330**	5.2171	0.6720	0.0120	0.0491	0.0123
0.25	1.4435	5.6521	**1.4161**	0.0181	0.0472	0.0184
0.50	3.8424	7.4220	**3.4809**	0.0293	0.0482	0.0302
0.75	7.8296	10.2053	**6.8348**	0.0417	0.0551	0.0430
1.00	13.4051	14.0769	**11.5172**	0.0545	0.0578	0.0558

Table 3: Average squared error loss (ASEL) and averaged desirability function (AD) values by OLS, LLR, and $MRR_{\lambda 2}$ for Goal 1 in the simulations based on CCD, followed by the estimated Monte Carlo errors of ASEL and AD in parentheses, respectively. (The estimated Monte Carlo errors of AD are in $\times 10^{-4}$.) Best values in bold.

γ	ASEL			AD		
	OLS	LLR	MRR	OLS	LLR	MRR
0.00	**105.3**(0.035)	109.8(0.470)	105.4(0.042)	**0.595**(0.67)	0.587(8.4)	0.595(0.79)
0.25	**90.7** (0.074)	95.0(0.270)	90.8(0.095)	**0.619**(1.5)	0.610(5.3)	0.619(1.8)
0.50	78.5(0.089)	82.9(0.249)	**78.3**(0.148)	0.642(1.9)	0.632(5.4)	**0.643**(3.0)
0.75	69.4(0.092)	69.4(0.263)	**68.8**(0.164)	0.663(2.3)	0.663(6.2)	**0.664**(3.8)
1.00	63.7(0.100)	**57.3**(0.111)	63.0(0.137)	0.681(2.6)	**0.695**(2.5)	0.684(3.9)

Table 4: ASEL and AD values by OLS, LLR, and MRR for Goal 2 in the simulations based on CCD, followed by the Monte Carlo errors of ASEL and AD in parentheses, respectively. Best values in bold.

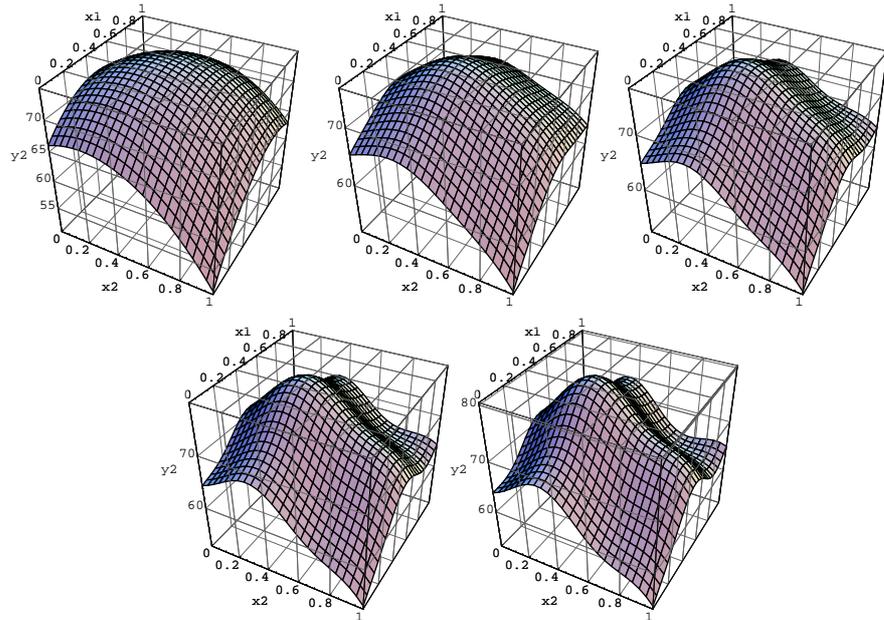| γ	ASEL			AD		
	OLS	LLR	MRR	OLS	LLR	MRR
0.00	2.7(0.068)	**2.5**(0.061)	2.6(0.069)	0.80(0.002)	**0.80**(0.002)	0.80(0.002)
0.25	5.1(0.181)	4.8(0.158)	**4.5**(0.153)	0.72(0.006)	0.73(0.005)	**0.74**(0.005)
0.50	8.3(0.350)	7.4(0.308)	**6.1**(0.264)	0.64(0.010)	0.67(0.009)	**0.70**(0.008)
0.75	13.7(0.518)	12.1(0.447)	**8.0**(0.356)	0.50(0.014)	0.52(0.013)	**0.64**(0.010)
1.00	20.3(0.649)	17.5(0.552)	**10.1**(0.396)	0.40(0.012)	0.43(0.011)	**0.60**(0.009)

Figure 1: Surfaces for the true mean function of the response y_1 when $\gamma = 0.00$ (top left), 0.25 (top middle), 0.50 (top right), 0.75 (bottom left), and 1.00 (bottom right), respectively.
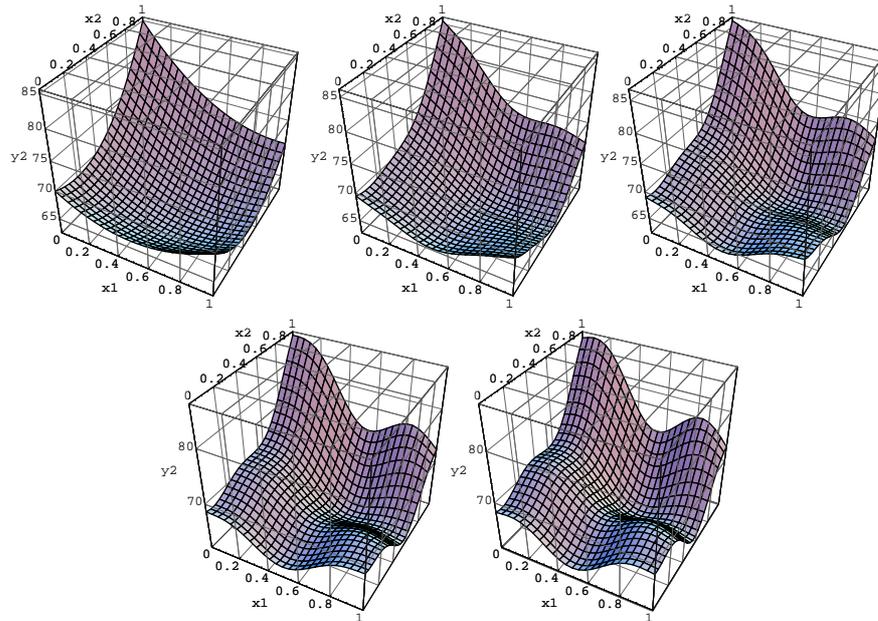
Figure 2: Surfaces for the true mean function of the response y_2 when $\gamma = 0.00$ (top left), 0.25 (top middle), 0.50 (top right), 0.75 (bottom left), and 1.00 (bottom right), respectively.

Authors' biographies

Wen Wan received her PhD in Statistics from the Virginia Polytechnic Institute and State University. She is an Assistant Professor in Biostatistics and Bioinformatics Unit in the Division of Preventive Medicine at the University of Alabama at Birmingham.

Jeffrey B. Birch is Professor of Statistics and Director of Graduate Programs at Virginia Tech, where he has been a faculty member since 1977. Dr. Birch received his M.S. (1970) and Ph. D. (1977) degrees in biostatistics from the University of Washington. His area of primary research includes most aspects of regression analysis including robust regression, nonparametric and semiparametric regression, and profile monitoring. Dr. Birch is a former associate editor of Biometrics (1989-1993).