A Dual Response Approach to the Multivariate Robust Parameter Design Problem.

Guillermo Miró Quesada and Enrique Del Castillo
Department of Industrial and Manufacturing Engineering
The Pennsylvania State University
University Park, PA 16802

December 5, 2003

Abstract

An extension to the Dual Response approach to Robust Parameter Design for the case of multiple responses is proposed. The methodology provides unbiased estimates of the process covariance matrix and of the vector of expected values using parameter estimates from a multivariate regression fit. There are no restrictions on the types of experimental designs that can be used, apart from their ability to fit the parameter estimates. Conditions for zeroing-out the variance contribution of the noise factors on the responses are also given for an unconstrained problem. For the more practical constrained case a discussion of the possible scalar optimization criteria of the covariance matrix is also given together with two illustrative examples taken from the literature.

1 Introduction

Robust Parameter Design (RPD) has been successfully used to improve the quality of products since the mid eighties (see Taguchi (1986, 1987) and Taguchi and Wu (1985); for a detailed discussion see Nair et al. (1992)). The technique consists in determining the levels of some set of controllable factors that reduces the sensitivity of the process to variations in another set of uncontrollable or noise factors, thus increasing the robustness of the process. The applicability of the technique relies completely in the existence of interactions between elements of both sets and in the ability to control the noise factors during experimentation.

The use of Response Surface Methodology (RSM) for RPD can be traced back to the early nineties (e.g. Box and Jones (1990) Vining and Myers (1990)) when it was proposed as an alternative to Taguchi's crossed array methods and the use of signal-to-noise ratios (see Wu & Hamada (2000) for a recent discussion on RPD and RSM). Since then a great number of articles have been published regarding their use to model the response variance. Perhaps the more frequently used model is the one initially proposed by Box and Jones (1990):

$$Y|\mathbf{z} = \beta_0 + \mathbf{x}'\boldsymbol{\beta} + \mathbf{x}'\mathbf{B}\mathbf{x} + \mathbf{z}'\boldsymbol{\gamma} + \mathbf{x}'\boldsymbol{\Delta}\mathbf{z} + \epsilon \tag{1}$$

where: $Y|\mathbf{z}$ is the value of the response variable for a fixed value of the noise variables $\mathbf{z}$, $\mathbf{x}$ is the vector of controllable factors, ϵ is a normally distributed random variable with zero mean and variance σ_ϵ and β_0, $\boldsymbol{\beta}$, $\mathbf{B}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\Delta}$ are the model parameters.

Although the model in equation (1) does not consider second order effects and interactions due to the noise factors, Box and Jones (1990) showed that any effect due to these terms is irrelevant for RPD purposes.

One of the main advantages of the model in equation (1) is the ease with which the unconditional variance is obtained:

$$Var(Y) = (\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x})'\boldsymbol{\Sigma}_{\mathbf{z}}(\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x}) + \sigma_{\epsilon}^2 \tag{2}$$

where $\boldsymbol{\Sigma}_{\mathbf{z}}$ is the covariance matrix of the noise variables, which is assumed known, possibly, from historic data of the process. Notice that this assumes that the noise factors are measurable. Furthermore, it is frequently assumed that $\boldsymbol{\Sigma}_{\mathbf{z}}$ is a diagonal matrix. A useful analysis based on model (2) consists of finding the operating conditions of zero variance contribution from the noise factors, i.e. find $\mathbf{x}$ such that $Var(Y) = \sigma_{\epsilon}^2$. The vector $\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x}$ is the partial derivative of Y with respect to the noise factors, $\frac{\partial Y}{\partial \mathbf{z}}$. Therefore, to minimize the variance contribution of the noise factors we seek values of the controllable factors where these "slopes" are zero or as flat as possible (Myers and Montgomery, 2002).

Despite all the attention that RPD has received in the literature, very little effort has been made in applying the RPD concept to processes with multiple responses. Chiao and Hamada (2001) propose an interesting and simple methodology where they estimate all the parameters present in a multivariate normal distribution and then use it to find which of the combination of controllable factors maximizes the probability that the vector of responses $\mathbf{Y}$ is in some pre-specified region, possible an acceptance region. However, their methodology assumes the use of a crossed-array replicated design, since the standard deviations are calculated from these replications. See Myers et al. (1997) and Lucas (1994) for other comparisons between response surface and Taguchi methods.

This lack of interest contrasts with the overwhelming attention that the topic of Multivariate Statistical Process Control has received. In this context a suitable set of Multivariate RPD (MRPD) methods could be of great interest as a tool to optimize the same processes for which the Multivariate SPC tools are designed to monitor while they are still at the design stage.

The purpose of this paper is to extend the application of RSM techniques to the MRPD case. In section 2 we develop the basic model for MRPD in conjunction with the models for predicting the mean and variance of the process. In section 3 unbiased estimators of the mean and variance are obtained. Section 4 contains a derivation for the unconstrained minimization of the process variance and some correspondences between the univariate and multivariate models. The application of the models developed for MRPD is proposed in section 5 and section 6 contains examples of actual applications of the proposed methodology. The paper concludes with a discussion of possible extensions of the basic approach in section 7.

2 Basic Model

Let us define a regression model with q responses, k controllable factors and r noise variables in the following way:

$$\underbrace{\mathbf{Y}}_{q\times 1} = \underbrace{\mathbf{\Theta}'}_{q\times p}\underbrace{\mathbf{x}^{(\mathbf{m})}}_{p\times 1} + \underbrace{\mathbf{\Delta}'}_{q\times r(k+1)}\underbrace{\mathbf{z}^{(\mathbf{m})}}_{r(k+1)\times 1} + \underbrace{\boldsymbol{\epsilon}}_{q\times 1} \tag{3}$$

where:

- $\mathbf{\Theta}$ is a $p \times q$ matrix of coefficients for the controllable factors (each column contains all the coefficients for one response);
- $\mathbf{x}^{(\mathbf{m})}$ is a $p \times 1$ vector containing the regressors for the controllable factors in model form (there are $p = \frac{(k+1)(k+2)}{2}$ such factors for a full quadratic model; $(\mathbf{m})$ stands for $\mathbf{x}$ in "model form"):

$$\mathbf{x}^{(\mathbf{m})} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_k \\ x_1x_2 \\ x_1x_3 \\ \vdots \\ x_{k-1}x_k \\ x_1^2 \\ x_2^2 \\ \vdots \\ x_k^2 \end{bmatrix} \tag{4}$$

- $\mathbf{\Delta}$ is a $r(k+1) \times q$ matrix containing first order coefficients for the noise factors as well as coefficients for the controllable factor-noise interactions. These were combined in a single matrix to avoid obtaining a cross-covariance term when applying the variance operator:

$$\mathbf{\Delta} = \begin{bmatrix} \beta_{y_1z_1} & \beta_{y_2z_1} & \cdots & \beta_{y_qz_1} \\ \beta_{y_1z_1x_1} & \beta_{y_2z_1x_1} & \cdots & \beta_{y_qz_1x_1} \\ \beta_{y_1z_1x_2} & \beta_{y_2z_1x_2} & \cdots & \beta_{y_qz_1x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{y_1z_1x_k} & \beta_{y_2z_2x_k} & \cdots & \beta_{y_qz_2x_k} \\ \beta_{y_1z_2} & \beta_{y_2z_2} & \cdots & \beta_{y_qz_2} \\ \beta_{y_1z_2x_1} & \beta_{y_2z_2x_1} & \cdots & \beta_{y_qz_2x_1} \\ \beta_{y_1z_2x_2} & \beta_{y_2z_2x_2} & \cdots & \beta_{y_qz_2x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{y_1z_2x_k} & \beta_{y_2z_2x_k} & \cdots & \beta_{y_qz_2x_k} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{y_1z_r} & \beta_{y_2z_r} & \cdots & \beta_{y_qz_r} \\ \beta_{y_1z_rx_1} & \beta_{y_2z_rx_1} & \cdots & \beta_{y_qz_rx_1} \\ \beta_{y_1z_rx_2} & \beta_{y_2z_rx_2} & \cdots & \beta_{y_qz_rx_2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{y_1z_rx_k} & \beta_{y_2z_rx_k} & \cdots & \beta_{y_qz_rx_k} \end{bmatrix} \tag{5}$$

where $\beta_{y_i x_j z_k}$ is the coefficient for the interaction between x_j and z_k for the i^{th} response. The first order coefficients for the noise factors are the $\beta_{y_i z_k}$, hence the matrix is $r(k+1)\times q$ and compatible with the definition of $\mathbf{z^{(m)}}$ that follows next.

- $\mathbf{z^{(m)}}$ is the $r(k+1)\times 1$ vector containing the noise factors and their interactions with the controllable factors:

$$\mathbf{z^{(m)}} = \begin{bmatrix} z_1 \\ z_1x_1 \\ z_1x_2 \\ \vdots \\ z_1x_k \\ z_2 \\ z_2x_1 \\ z_2x_2 \\ \vdots \\ z_2x_k \\ \vdots \\ z_r \\ z_rx_1 \\ z_rx_2 \\ \vdots \\ z_rx_k \end{bmatrix} \tag{6}$$

note that $\mathbf{z^{(m)}}$ is a function of the controllable factors (i.e. $\mathbf{z^{(m)}} = f(\mathbf{x})$) but we write $\mathbf{z^{(m)}}$ only for notational simplicity.

- ϵ is a $q\times 1$ normally distributed random vector with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}_\epsilon$:

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_q \end{bmatrix} \sim N_q(\mathbf{0}, \mathbf{\Sigma}_\epsilon)$$

Notice that in equation (3) we are assuming that all the responses can be appropriately modelled by functions of the same form, that is by functions containing the same significant parameters. Since in practice this will seldom occur, some type of approximation would typically be required (we discuss this assumption further in the last section of the paper).

In the remaining of this paper we assume that:

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_r \end{bmatrix} \sim N_r(\mathbf{0}, \mathbf{\Sigma_z})$$

where it is assumed that $\mathbf{\Sigma_z}$ is known, possibly from historical data, similarly as in the univariate RPD problem.

Using the previous assumption we can take expectation and variance operators in equation (3):

$$E(\mathbf{Y}) = \mathbf{\Theta}'\mathbf{x}^{(\mathbf{m})}. \tag{7}$$

A multivariate approach equivalent to the one used by Box-Jones (1990) and Myers (1991) for the univariate case would be to take the variance operator in equation (3) to obtain an equation useful to predict the variance of the responses once the noise factors are not tightly controlled as they were during the experiments. In other words we would get the $q \times q$ matrix:

$$Var(\mathbf{Y}) = \mathbf{\Delta}'Cov(\mathbf{z}^{(\mathbf{m})})\mathbf{\Delta} + \mathbf{\Sigma}_\epsilon \tag{8}$$

where $Cov(\mathbf{z}^{(\mathbf{m})})$ is a $[(k+1)r \times (k+1)r]$ matrix given by:

$$Cov(\mathbf{z}^{(\mathbf{m})}) = \left[\begin{array}{cccc|cccc|c|cccc}
\sigma^2_{z_1} & x_1\sigma^2_{z_1} & \dots & x_k\sigma^2_{z_1} & \sigma_{z_1z_2} & x_1\sigma_{z_1z_2} & \dots & x_k\sigma_{z_1z_2} & \dots & \sigma_{z_1z_r} & x_1\sigma_{z_1z_r} & \dots & x_k\sigma_{z_1z_r} \\
x_1\sigma^2_{z_1} & x_1^2\sigma^2_{z_1} & \dots & x_1x_k\sigma^2_{z_1} & x_1\sigma_{z_1z_2} & x_1^2\sigma_{z_1z_2} & \dots & x_1x_k\sigma_{z_1z_2} & \dots & x_1\sigma_{z_1z_r} & x_1^2\sigma_{z_1z_r} & \dots & x_1x_k\sigma_{z_1z_r} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
x_k\sigma^2_{z_1} & x_1x_k\sigma^2_{z_1} & \dots & x_k^2\sigma^2_{z_1} & x_k\sigma_{z_1z_2} & x_1x_k\sigma_{z_1z_2} & \dots & x_k^2\sigma_{z_1z_2} & \dots & x_k\sigma_{z_1z_r} & x_1x_k\sigma_{z_1z_r} & \dots & x_k^2\sigma_{z_1z_r} \\
\hline
\sigma_{z_1z_2} & x_1\sigma_{z_1z_2} & \dots & x_k\sigma_{z_1z_2} & \sigma^2_{z_2} & x_1\sigma^2_{z_2} & \dots & x_k\sigma^2_{z_2} & \dots & \sigma_{z_2z_r} & x_1\sigma_{z_2z_r} & \dots & x_k\sigma_{z_2z_r} \\
x_1\sigma_{z_1z_2} & x_1^2\sigma_{z_1z_2} & \dots & x_1x_k\sigma_{z_1z_2} & x_1\sigma^2_{z_2} & x_1^2\sigma^2_{z_2} & \dots & x_1x_k\sigma^2_{z_2} & \dots & x_1\sigma_{z_2z_r} & x_1^2\sigma_{z_2z_r} & \dots & x_1x_k\sigma_{z_2z_r} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
x_k\sigma_{z_1z_2} & x_1x_k\sigma_{z_1z_2} & \dots & x_k^2\sigma_{z_1z_2} & x_k\sigma^2_{z_2} & x_1x_k\sigma^2_{z_2} & \dots & x_k^2\sigma^2_{z_2} & \dots & x_k\sigma_{z_2z_r} & x_1x_k\sigma_{z_2z_r} & \dots & x_k^2\sigma_{z_2z_r} \\
\hline
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\hline
\sigma_{z_1z_r} & x_1\sigma_{z_1z_r} & \dots & x_k\sigma_{z_1z_r} & \sigma_{z_1z_2} & x_1\sigma_{z_1z_2} & \dots & x_k\sigma_{z_1z_2} & \dots & \sigma^2_{z_r} & x_1\sigma^2_{z_r} & \dots & x_k\sigma^2_{z_r} \\
x_1\sigma_{z_1z_r} & x_1^2\sigma_{z_1z_r} & \dots & x_1x_k\sigma_{z_1z_r} & x_1\sigma_{z_1z_2} & x_1^2\sigma_{z_1z_2} & \dots & x_1x_k\sigma_{z_1z_2} & \dots & x_1\sigma^2_{z_r} & x_1^2\sigma^2_{z_r} & \dots & x_1x_k\sigma^2_{z_r} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
x_k\sigma_{z_1z_r} & x_1x_k\sigma_{z_1z_r} & \dots & x_k^2\sigma_{z_1z_r} & x_k\sigma_{z_1z_2} & x_1x_k\sigma_{z_1z_2} & \dots & x_k^2\sigma_{z_1z_2} & \dots & x_k\sigma^2_{z_r} & x_1x_k\sigma^2_{z_r} & \dots & x_k^2\sigma^2_{z_r}
\end{array}\right]$$

where $\sigma_{z_iz_j}$ denotes the covariance between z_i and z_j for $i \neq j$ and $\sigma^2_{z_i}$ denotes the variance of z_i. The lines are simply visual aids to help noticing the patterns in the matrix. Note how if the noise factors are uncorrelated, $Cov(\mathbf{z}^{(\mathbf{m})})$ is block diagonal.

The matrix $Cov(\mathbf{z}^{(\mathbf{m})})$ can be written as:

$$Cov(\mathbf{z}^{(\mathbf{m})}) = \mathbf{\Sigma_z} \otimes \left[\mathbf{x}^{(\mathbf{l})}\mathbf{x}'^{(\mathbf{l})}\right] \tag{9}$$

where $\mathbf{x}^{(\mathbf{l})}$ is formed by the first $k+1$ elements of $\mathbf{x}^{(\mathbf{m})}$, $\mathbf{\Sigma_z}$ is the covariance matrix of the noise factors and $\otimes$ denotes the Kronecker or direct product. Unfortunately, the following property of the direct product operation:

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{F}) = \mathbf{AC} \otimes \mathbf{BF}$$

cannot be used to simplify equation (9) since the resulting matrices are non-conformable. Therefore, the most simplified expression for $Var(\mathbf{Y})$ found is:

$$Var(\mathbf{Y}) = \mathbf{\Delta}'\left[\mathbf{\Sigma_z} \otimes \left(\mathbf{x}^{(\mathbf{l})}\mathbf{x}'^{(\mathbf{l})}\right)\right]\mathbf{\Delta} + \mathbf{\Sigma}_\epsilon \tag{10}$$

which can be contrasted with its univariate counterpart given by equation (2). In the remainder of the paper the $r(k+1) \times r(k+1)$ matrix $Cov(\mathbf{z}^{(\mathbf{m})}) = \mathbf{\Sigma_z} \otimes \left(\mathbf{x}^{(\mathbf{l})}\mathbf{x}'^{(\mathbf{l})}\right)$ will be denoted by $\mathbf{\Sigma}^{\otimes}_{\mathbf{z}}$.

3 Estimation of the Mean and Variance Models

By simply substituting $\boldsymbol{\Delta}, \boldsymbol{\Theta}$ and $\boldsymbol{\Sigma}_\epsilon$ by their respective unbiased estimators, $\widehat{\boldsymbol{\Delta}}, \widehat{\boldsymbol{\Theta}}$ and $\widehat{\boldsymbol{\Sigma}}_\epsilon$ in equation (7), we obtain an unbiased estimator for $E(\mathbf{Y})$. To see this let:

$$\widehat{E}(\mathbf{Y}) = \widehat{\boldsymbol{\Theta}}'\mathbf{x}$$

Then, we have that:

$$E(\widehat{E}(\mathbf{Y})) = \boldsymbol{\Theta}'\mathbf{x} = E(\mathbf{Y}) \ .$$

The estimator of the variance-covariance matrix of the response, $\widehat{Var}(\mathbf{Y})$, that would be obtained from equation (10):

$$\widehat{Var}(\mathbf{Y}) = \widehat{\boldsymbol{\Delta}}'\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}\widehat{\boldsymbol{\Delta}} + \widehat{\boldsymbol{\Sigma}}_\epsilon \ , \tag{11}$$

is not an unbiased estimator of $Var(\mathbf{Y})$. To find an unbiased estimator of $Var(\mathbf{Y})$, one common approach is to find the expected value of the "naive" estimator in equation (11) and then, if possible, correct it by an unbiased estimator of the bias. To do this, let $\widehat{\boldsymbol{\delta}}_{.j}$ be the j^{th} column of $\widehat{\boldsymbol{\Delta}}$ and let $\widehat{\sigma}_{ij}$ be the $(ij)^{th}$ element of $Var(\mathbf{Y})$. We then have that:

$$\widehat{\sigma}_{ij} = \widehat{\boldsymbol{\delta}}'_{.i}\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}\widehat{\boldsymbol{\delta}}_{.j} + \widehat{\sigma}_{\epsilon_{ij}}$$

Expectation can be easily taken yielding:

$$E(\widehat{\sigma}_{ij}) = \boldsymbol{\delta}'_{.i}\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}\boldsymbol{\delta}_{.j} + trace\left(\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}\boldsymbol{\Sigma}_{\widehat{\delta}_{.i}\widehat{\delta}_{.j}}\right) + \sigma_{\epsilon_{ij}} \tag{12}$$

where $\boldsymbol{\Sigma}_{\widehat{\delta}_{.i}\widehat{\delta}_{.j}}$ is the cross-covariance matrix between the vectors $\widehat{\delta}_{.i}$ and $\widehat{\delta}_{.j}$.

From Press (1982, p. 233-234) it can be seen that this cross-covariance is just:

$$\boldsymbol{\Sigma}_{\widehat{\delta}_{.i}\widehat{\delta}_{.j}} = \sigma_{\epsilon_{ij}}(\mathbf{X}'_{\boldsymbol{\Delta}}\mathbf{X}_{\boldsymbol{\Delta}})^{-1}$$

where $\mathbf{X}_{\boldsymbol{\Delta}}$ is a matrix formed by the columns of the design matrix $\mathbf{X}$ corresponding to the regressors in $\mathbf{z}^{(\mathbf{m})}$. Hence $(\mathbf{X}'_{\boldsymbol{\Delta}}\mathbf{X}_{\boldsymbol{\Delta}})^{-1}$ is just the scaled covariance matrix of any of the columns of $\widehat{\boldsymbol{\Delta}}$. These two matrices are better illustrated numerically in Example 1 below. Note that every column of $\widehat{\boldsymbol{\Delta}}$ has the same "scaled" covariance matrix, they just differ by a constant, i.e. $\sigma_{\epsilon_{ij}}$.

Substituting $\boldsymbol{\Sigma}_{\widehat{\delta}_{.i}\widehat{\delta}_{.j}}$ in (12) we get:

$$E(\widehat{\sigma}_{ij}) = \boldsymbol{\delta}'_{.i}\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}\boldsymbol{\delta}_{.j} + \sigma_{\epsilon_{ij}}trace\left(\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}(\mathbf{X}'_{\boldsymbol{\Delta}}\mathbf{X}_{\boldsymbol{\Delta}})^{-1}\right) + \sigma_{\epsilon_{ij}} \tag{13}$$

Since the argument of the *trace* operator does not depend on i or j the result presented in equation (13) can be extended to matrix form in the following way:

$$E(\widehat{Var}(\mathbf{Y})) = \boldsymbol{\Delta}'\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}\boldsymbol{\Delta} + \left(1 + trace\left(\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}(\mathbf{X}'_{\boldsymbol{\Delta}}\mathbf{X}_{\boldsymbol{\Delta}})^{-1}\right)\right)\boldsymbol{\Sigma}_\epsilon$$

Finally, an unbiased estimator of $Var(\mathbf{Y})$ is given by:

$$\widehat{Var}(\mathbf{Y}) = \widehat{\boldsymbol{\Delta}}'\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}\widehat{\boldsymbol{\Delta}} + \left(1 - trace\left(\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}(\mathbf{X}'_{\boldsymbol{\Delta}}\mathbf{X}_{\boldsymbol{\Delta}})^{-1}\right)\right)\widehat{\boldsymbol{\Sigma}}_\epsilon \tag{14}$$

In the remaining of this paper we will refer to $1-trace\left(\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}(\mathbf{X}'_{\boldsymbol{\Delta}}\mathbf{X}_{\boldsymbol{\Delta}})^{-1}\right)$ as the *bias correction factor*.

Notice that if $trace\left(\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}(\mathbf{X}'_{\boldsymbol{\Delta}}\mathbf{X}_{\boldsymbol{\Delta}})^{-1}\right) > 1$, then the bias correction factor will be negative and if the elements in this matrix are sufficiently large, then the unbiased estimator will be given by a non-positive definite matrix. An equivalent problem is also present in the univariate case where a similar correction is made (see Myers and Montgomery (2002), p. 576).

The possible non-positive definiteness of $\widehat{Var}(\mathbf{Y})$ arises from the fact that equation (10) only takes into account the variance due to the noise factors and the residual noise.

The term $trace\left(\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}(\mathbf{X}'_{\boldsymbol{\Delta}}\mathbf{X}_{\boldsymbol{\Delta}})^{-1}\right)\widehat{\boldsymbol{\Sigma}}_{\epsilon}$ can be considered as the variance component due to the error in the estimation of $\boldsymbol{\Delta}$, hence it is subtracted in equation (14) in order to estimate only the variance due to the noise factors and the residual noise. Then, the non-positive definiteness occurs when the estimate of this variance component is greater than the estimated variance due to the noise factors and residual noise. When this occurs it means that the predictive properties of the model are not appropriate, since the variance component that we are trying to minimize may be negligible with respect to the variance in estimating the model.

One way of preventing this problem at the design phase is to code the noise factors so as to be centered at their mean, μ_{z_j} with the ± 1 levels at $\mu_{z_j} \pm \sigma_{z_j}$. Then the coded version of $\boldsymbol{\Sigma}_{\mathbf{z}}$ will be in correlation form, i.e. it will have ones in the diagonal and the correlations coefficients between the noise variables in the off-diagonals. Using this coding scheme will help in maintaining the second term of the bias correction factor small, avoiding negative values.

However, if the problem does occur a remedial measure could be to perform new experiments at the ± 2 level of the noise factors in order to reduce the entries in $(\mathbf{X}'_{\boldsymbol{\Delta}}\mathbf{X}_{\boldsymbol{\Delta}})^{-1}$, that is to decrease the variance due to the estimation of $\boldsymbol{\Delta}$.

Another situation when an unbiased estimator of a variance component is obtained by taking the difference of other variance components is in the random effects and mixed ANOVA models (see for example Arnold (1981) or Searle et al. (1992)), where the possibility of negative variance component estimates has been discussed extensively.

4 Unconstrained Minimization of $Var(\mathbf{Y})$

Similarly as in the univariate RPD problem, it may be of interest in some applications to determine under which conditions the variability due to the noise factors can be made equal to zero. Ideally, the user would desire to make all the elements in $\boldsymbol{\Delta}'\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}\boldsymbol{\Delta}$ equal to zero. Since this matrix is at least positive semi-definite, to achieve such a goal it should be enough to make all the diagonal elements zero. That is, for this type of matrices, since $|a_{ij}| \leq \sqrt{a_{ii}a_{jj}}$, hence if $a_{ii} = 0$, then $a_{ij} = 0\ \forall j$. Therefore, we should be concerned with the variances of the individual responses, for which it is simpler to work with q univariate noise models (see equation 2):

$$Var(Y_i) = (\boldsymbol{\gamma}_i + \boldsymbol{\Delta}'_i\mathbf{x})'\boldsymbol{\Sigma}_{\mathbf{z}}(\boldsymbol{\gamma}_i + \boldsymbol{\Delta}'_i\mathbf{x}) + \sigma_i^2$$

where $\boldsymbol{\Delta}_i$ and $\boldsymbol{\gamma}_i$ are constructed using elements of $\boldsymbol{\Delta}$ (as defined in equation 5) in the following manner:

$$\boldsymbol{\Delta}_i = \begin{bmatrix} \beta_{y_i z_1 x_1} & \beta_{y_i z_2 x_1} & \cdots & \beta_{y_i z_r x_1} \\ \beta_{y_i z_1 x_2} & \beta_{y_i z_2 x_2} & \cdots & \beta_{y_i z_r x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{y_i z_1 x_k} & \beta_{y_i z_2 x_k} & \cdots & \beta_{y_i z_r x_k} \end{bmatrix}$$

$$\boldsymbol{\gamma}_i = \begin{bmatrix} \beta_{y_i z_1} \\ \beta_{y_i z_2} \\ \vdots \\ \beta_{y_i z_r} \end{bmatrix}$$

Then, to make one individual variance equal to zero we need to solve the following system of r equations:

$$\boldsymbol{\gamma}_i + \boldsymbol{\Delta}_i' \mathbf{x} = \mathbf{0}_{r \times 1} \tag{15}$$

for each of the q response variables. In matrix notation the complete system can be expressed as:

$$\begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \\ \vdots \\ \boldsymbol{\gamma}_q \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Delta}_1' \\ \boldsymbol{\Delta}_2' \\ \vdots \\ \boldsymbol{\Delta}_q' \end{bmatrix} \mathbf{x} \equiv \boldsymbol{\gamma}_M + \boldsymbol{\Delta}_M' \mathbf{x} = \mathbf{0}_{rq \times 1} \tag{16}$$

where $\boldsymbol{\gamma}_M$ and $\boldsymbol{\Delta}_M'$ are constructed by vertically stacking the individual $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Delta}_i'$. To solve the system in equation (16) we need that $\boldsymbol{\Delta}_M'$ be a full rank square matrix. Although the full rank condition could easily be assumed, for a square matrix we need $k = rq$ controllable factors. If this is true, $\boldsymbol{\Delta}_M$ is invertible and the solution to the system can easily be obtained as:

$$\mathbf{x}^* = -\left(\boldsymbol{\Delta}_M'\right)^{-1} \boldsymbol{\gamma}_M \tag{17}$$

Now, it is also of interest to determine if solving the system in equation (16) is the only way of obtaining zero variance contribution from the noise factors, especially considering the great number of controllable factors required for doing this. Towards this goal, notice that if $\mathbf{x}^*$ solves (17) then it also solves each of the individual systems in (15) and therefore we can state that:

$$\boldsymbol{\gamma}_i + \boldsymbol{\Delta}_i' \mathbf{x}^* = \mathbf{0}_{r \times 1} \Leftrightarrow \begin{bmatrix} \boldsymbol{\gamma}_i & \boldsymbol{\Delta}_i' \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{x}^* \end{bmatrix} = \mathbf{0}_{r \times 1}$$

and, since $\boldsymbol{\Sigma}_{\mathbf{z}}$ is positive definite:

$$\Leftrightarrow Var(Y_i) - \sigma_i^2 = \left(\begin{bmatrix} \boldsymbol{\gamma}_i & \boldsymbol{\Delta}_i' \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{x}^* \end{bmatrix} \right)' \boldsymbol{\Sigma}_{\mathbf{z}} \left(\begin{bmatrix} \boldsymbol{\gamma}_i & \boldsymbol{\Delta}_i' \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{x}^* \end{bmatrix} \right) = 0_{1 \times 1} \tag{18}$$

By comparing the elements of $\boldsymbol{\Delta}$, $\boldsymbol{\Delta}_i$ and $\boldsymbol{\gamma}_i$, it can be seen that :

$$\text{vec}\left(\left[\boldsymbol{\gamma}_i \;\; \boldsymbol{\Delta}_i'\right]'\right) = \boldsymbol{\delta}_{.i}$$

where recall that $\boldsymbol{\delta}_{.i}$ is the i^{th} column of matrix $\boldsymbol{\Delta}$. Applying Proposition 1 in the Appendix to the scalar in equation (18) we get:

$$Var(Y_i) - \sigma_i^2 = \left[\begin{array}{cc} 1 & \mathbf{x}^* \end{array}\right] \left[\begin{array}{c} \boldsymbol{\gamma}_i' \\ \boldsymbol{\Delta}_i \end{array}\right] \boldsymbol{\Sigma}_{\mathbf{z}} \left[\begin{array}{c} \boldsymbol{\gamma}_i' \\ \boldsymbol{\Delta}_i \end{array}\right]' \left[\begin{array}{c} 1 \\ \mathbf{x}^* \end{array}\right] = 0 \quad \Leftrightarrow$$

$$\boldsymbol{\delta}_{.i}' \left(\boldsymbol{\Sigma}_{\mathbf{z}} \otimes \mathbf{x}^{(1)*}\mathbf{x}'^{(1)*}\right) \boldsymbol{\delta}_{.i} = 0 \tag{19}$$

where $\left[\begin{array}{cc} 1 & \mathbf{x}^* \end{array}\right]' = \mathbf{x}^{(1)*}$ as defined earlier in equation (9).

Therefore, it has been proved that a zero covariance matrix for the responses can only be obtained by making all the individual variances zero and that for doing this we may as well solve the q univariate cases, since these are easier to solve linear equations. However, this requires the availability of rq controllable factors, which could be a large number in many practical situations. Even if such a number of controllable factors exist, the inclusion of all of them in a single experimental design will probably represent too many experiments. In addition, the solution $\mathbf{x}^*$ may lie outside the experimental region or could imply levels of the controllable factors that are unattainable in practice.

Since the argument stated before is completely reversible, it is straightforward to see that:

$$Cov(Y_i, Y_j) = (\boldsymbol{\gamma}_i + \boldsymbol{\Delta}_i'\mathbf{x})'\boldsymbol{\Sigma}_{\mathbf{z}}(\boldsymbol{\gamma}_j + \boldsymbol{\Delta}_j'\mathbf{x}) + \sigma_{ij} \tag{20}$$

The first term in the right-hand side of equation (20) may be made equal to zero without having to make either the pre-multiplying or post-multiplying vectors equal to zero as in the case of equation (18). Now a possibly more interesting application would be to make the off-diagonal elements $\boldsymbol{\delta}_{.i}'\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}\boldsymbol{\delta}_{.j} = -\sigma_{\epsilon_{ij}}$ for some $i \neq j$, so that, from equation (10), $Cov(Y_i, Y_j) = 0$ and therefore, the i^{th} and j^{th} will be uncorrelated. Since the normality of the noise variables is also being assumed and the basic model in equation (3) implies a linear relationship between the response and the noise variables, $\mathbf{Y}$ is normally distributed even after taking into account the randomness of the noise variables and, hence, zero correlation implies independence. In practice, when $\boldsymbol{\delta}_{.i}'\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}\boldsymbol{\delta}_{.j} = -\sigma_{\epsilon_{ij}}$ the correlation introduced by the noise variables is being used to cancel the correlation in the residual noise. Notice that having control over the correlations in the multivariate case is the logical extension to having control over the variance in the univariate case. An example in which using this control to induce independent responses could be beneficial is for the implementation of Multivariate SPC, since in this case each response could be monitored independently, making the process easier to monitor.

In addition, notice that even when $\boldsymbol{\Sigma}_{\mathbf{z}}$ and $\boldsymbol{\Sigma}_{\epsilon}$ are diagonal matrices, i.e. when the noise factors ($\mathbf{z}$) and the residual noise ($\boldsymbol{\epsilon}$) are each formed by independent random vectors, the correlation between two responses will not be zero if they happen to interact with at least one common noise variable. Therefore, for the multivariate case, an additional *covariance structure* will be generated by the noise factors.

5 Robust Parameter Design for the Multivariate Case

The models described in the preceding sections can be used to extend the work by Myers (1991) and Box and Jones (1990) to the case where there is more than one response. In the univariate case the dual response approach to the RPD problem is usually specified as follows:

$$\begin{aligned} & min && Var(Y) \\ subject\ to: & && \\ & && a \leq E(Y) \leq b \\ & && \mathbf{x} \in X \end{aligned}$$

where a and b are bounds on the predicted expected value, $\mathbf{x}$ represents the set of controllable factors and X is the region of experimentation where the regression model is valid. The case when the expected response is restricted to only one value, i.e. the case of equality constraints is the special case when $a = b$. Now, the logical extension of this approach to the multivariate case would call for the minimization of some scalar function of the covariance matrix.

Scalar functions of matrices are common for the generation of Optimal Designs according to various criteria. D-optimality minimizes the determinant of the scaled covariance matrix (sometimes called the generalized variance) while A-optimality minimizes the trace. Press (1982) mentions that other common scalar measures of internal scatter are $(trace\, Var(\mathbf{Y})^2)^{1/2}$ and the difference between the highest and smallest eigenvalue of $Var(\mathbf{Y})$. As it is well-known, the determinant is proportional to the volume of a joint confidence ellipsoid for $\mathbf{Y}$ (Khuri & Cornell, 1987).

With the exception of the determinant, which is discussed later in this section, the minimization of either of the aforementioned scalar objectives would be significantly impacted by the scaling of $Var(\mathbf{Y})$. From equation (10) we can see that a proper scaled covariance matrix of the responses can be obtained by scaling the regression coefficients in $\mathbf{\Delta}$, which in turn can be obtained by scaling the values of the responses used to fit the regression model. Using L_p norms is a common technique used in multi-criteria optimization to tackle this kind of problems, see for example Steuer (1989) for a discussion on this issue.

Finally we can state the optimization problem in the following way:

$$\begin{aligned} & min && f\left(Var(\mathbf{Y})\right) \\ subject\ to: & && \\ & && \mathbf{l} \leq E(\mathbf{Y}) \leq \mathbf{u} \\ & && \mathbf{x} \in X \end{aligned}$$

where $\mathbf{u}$ and $\mathbf{l}$ are $q \times 1$ vectors containing lower and upper bounds for the response, $f(\cdot)$ is a suitable scalar function and $Var(\mathbf{Y})$ and $E(\mathbf{Y})$ are given by equations (10) and (7) respectively.

For the case of the determinant we can use the following result to state that the optimization problem defined previously is invariant to the scaling of the responses. Let the subscripts u and s refer to the un-scaled and scaled versions of the associated matrices, then:

$$\begin{aligned}
\left|\widehat{Var}_s(\mathbf{Y})\right| &= \left|\widehat{\boldsymbol{\Delta}}_s' \boldsymbol{\Sigma}_{\mathbf{z}}{}^{\otimes} \widehat{\boldsymbol{\Delta}}_s + \left(1 - trace\left(\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}(\mathbf{X}_{\boldsymbol{\Delta}}'\mathbf{X}_{\boldsymbol{\Delta}})^{-1}\right)\right)\widehat{\boldsymbol{\Sigma}}_{\epsilon,s}\right| \\
&= \left|\mathbf{N}'\widehat{\boldsymbol{\Delta}}_u' \boldsymbol{\Sigma}_{\mathbf{z}}{}^{\otimes} \widehat{\boldsymbol{\Delta}}_u \mathbf{N} + \left(1 - trace\left(\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}(\mathbf{X}_{\boldsymbol{\Delta}}'\mathbf{X}_{\boldsymbol{\Delta}})^{-1}\right)\right)\mathbf{N}'\widehat{\boldsymbol{\Sigma}}_{\epsilon,u}\mathbf{N}\right| \\
&= |\mathbf{N}'|\left|\widehat{Var}_u(\mathbf{Y})\right||\mathbf{N}| = |\mathbf{N}'\mathbf{N}|\left|\widehat{Var}_u(\mathbf{Y})\right|
\end{aligned}$$

where $\mathbf{N}$ is a diagonal matrix containing the normalizing constants used for each response. For the case of the L_2 norm this is given by:

$$\mathbf{N} = \begin{bmatrix} \frac{1}{\|Y_{.1}\|} & 0 & \dots & 0 \\ 0 & \frac{1}{\|Y_{.2}\|} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\|Y_{.q}\|} \end{bmatrix}$$

Being invariant to the type of scaling is a strong advantage since it avoids the possibility of having scale-dependent solutions. Therefore, using the determinant as the scalar function has the advantages that it does not requires any type of scaling, it considers the covariance structure in the responses, and it has the practical interpretation of being proportional to the joint confidence ellipsoid of the vector of responses. However, it requires the estimation of the complete covariance matrix as stated in equation (14).

In contrast, using the trace has the advantage that it only requires the diagonal elements of $Var(\mathbf{Y})$, which can be estimated from the univariate models. However, it does not consider the covariance structure and its interpretation depends on the normalizing method used to bring the individual variances to a common scale.

The other two aforementioned objectives, $(trace\, Var(\mathbf{Y})^2)^{1/2}$ and the difference between the largest and smallest eigenvalues, require the estimation of the complete covariance matrix and they lack the simplicity of the trace without having the advantages of the determinant. However, they will be included in the first example for completeness.

6 Examples

The examples presented in this section are based on experiments that were taken from the literature and, therefore, were not designed to be analyzed using the proposed methodology. To use the data sets, therefore, two assumptions need to be made. The first one corresponds to what factors should be considered as noise factors. This decision was made based on the number of significant interactions with the controllable factors, since these interactions are fundamental for the applicability of the method. The second assumption made was regarding the variance-covariance matrix of the factors that were assigned as noise factors, since no historic data were available to estimate it. The approaches taken in this respect will be justified in each example in order to study some interesting behavior.

The first example illustrates the different optimization criterions based on $\widehat{Var}(\mathbf{Y})$. The second example illustrates the potential advantages of the multivariate approach over optimizing $\widehat{Var}(Y_i)$ for each response i.

Table 1: Factor and Response Data for the HPLC Example

Controllable Factors			Responses			
%IPA	Temp.	pH	Rs	Run Time	S/N ratio	Tailing
65	30	0.175	2.14	22	172	0.76
65	50	0.175	1.73	12	311	0.88
65	40	0.050	1.93	16	251	0.80
65	40	0.300	1.95	16	241	0.80
70	40	0.175	2.17	14	278	0.79
70	50	0.050	1.97	11	371	0.86
70	30	0.300	2.38	19	194	0.74
70	50	0.300	1.98	11	360	0.86
70	30	0.050	2.37	18	204	0.74
70	40	0.175	2.20	14	280	0.78
75	40	0.300	2.42	13	314	0.78
75	30	0.175	2.61	17	223	0.73
75	50	0.175	2.14	10	410	0.85
75	40	0.050	2.42	12	324	0.78
70	40	0.175	2.20	14	281	0.79

6.1 HPLC Optimization

The following example was taken from Peterson (2000). It involves the optimization of a high performance liquid chromatography (HPLC) system to detect mixtures of impurities. The performance of the assay was based upon four quantitative response variables, the critical resolution (Rs), total run time, signal-to-noise ratio of the last peak and the tailing factor of the major peak. Three controllable factors affecting the HPLC assay were included: %IPA, temperature and pH. A Box-Behnken design with three center points was used. The data are presented in Table 1.

The steps required for the minimization of the trace of the variance-covariance matrix will be shown in detail. From this the steps required for the other optimization criteria are straightforward and will not be shown in detail.

Initially, a full quadratic model in all the controllable factors was fitted using ordinary least squares. Suppose %IPA is the only noise factor in the process (this was assumed to illustrate the methodology given the interactions of this factor with other factors for the Run Time and S/N ratio responses). No significant interactions were found for the Rs and Tailing responses. Nonetheless, this does not imply that they should be taken out of the problem, since bounds to their predicted expected values are also present in the problem as constraints.

The following matrix contains the L_2 normalized values for the responses (unless otherwise stated, all the intermediate results shown are based on the normalized values of the responses. In addition, the titles of the columns or rows are included in the following matrices to ease the understanding of their contents):

$$\mathbf{Y} = \begin{bmatrix} Rs & Time & S/N & Tailing \\ \hline 0.2528 & 0.3797 & 0.1538 & 0.2461 \\ 0.2043 & 0.2071 & 0.2782 & 0.285 \\ 0.228 & 0.2761 & 0.2245 & 0.2591 \\ 0.2303 & 0.2761 & 0.2156 & 0.2591 \\ 0.2563 & 0.2416 & 0.2487 & 0.2558 \\ 0.2327 & 0.1899 & 0.3318 & 0.2785 \\ 0.2811 & 0.3279 & 0.1735 & 0.2396 \\ 0.2339 & 0.1899 & 0.322 & 0.2785 \\ 0.2799 & 0.3107 & 0.1825 & 0.2396 \\ 0.2599 & 0.2416 & 0.2504 & 0.2526 \\ 0.2858 & 0.2244 & 0.2809 & 0.2526 \\ 0.3083 & 0.2934 & 0.1995 & 0.2364 \\ 0.2528 & 0.1726 & 0.3667 & 0.2753 \\ 0.2858 & 0.2071 & 0.2898 & 0.2526 \\ 0.2599 & 0.2416 & 0.2513 & 0.2558 \end{bmatrix}$$

The methodology developed in previous sections requires that the response functions be of the same form. To satisfy this requirement all the elements in the respective $\widehat{\mathbf{\Delta}}$ and $\widehat{\mathbf{\Theta}}$ matrices must be included. Since in practice some of these estimates will be statistically non-significant, some type of approximation needs to me made. This approximation could be either leaving non-statistically significant estimates in the models or setting their respective values to zero. Since both of them are approximations to the more rigorous approach of considering different functional relationships between the responses, we will use the approach of setting the non-significant terms to zero. This approach is preferred because the models will only consider significant terms and the solution from the optimization problem will be more realistic, especially for the case of the interactions between the noise and controllable factors on which any reduction in variance depends.

After fitting regression models to all the responses with factors coded in the $(-1, 1)$ range, testing for significance and setting all non-significant coefficients to zero, the following $\widehat{\mathbf{\Delta}}$ matrix was constructed as described by equation (5):

$$\widehat{\mathbf{\Delta}} = \left[\begin{array}{l|rrrr} & Rs & Time & S/N & Tailing \\ \hline IPA & 0.0272 & -0.0302 & 0.0331 & -0.0040 \\ IPA \times Temp & 0 & 0.0129 & 0.0107 & 0 \\ IPA \times pH & 0 & 0 & 0 & 0 \end{array}\right]$$

The matrix of estimates for the coefficients of the controllable factors is as follows (this corresponds to equation (4)):

$$\widehat{\mathbf{\Theta}} = \left[\begin{array}{l|rrrr} & Rs & Time & S/N & Tailing \\ \hline Constant & 0.2576 & 0.2456 & 0.2500 & 0.2556 \\ Temp & -0.0248 & -0.0690 & 0.0737 & 0.0194 \\ pH & 0 & 0 & -0.0046 & 0 \\ Temp \times pH & 0 & 0 & 0 & 0 \\ Temp^2 & 0 & 0.0146 & 0 & 0.0045 \\ pH^2 & 0 & 0 & 0.0027 & 0 \end{array}\right]$$

We point out that for the case when there is only one noise variable, with variance denoted by σ_z^2, if the objective is to minimize $trace\left(\widehat{Var}(\mathbf{Y})\right)$, then the objective function can be simplified to:

$$\begin{aligned} trace\left(\widehat{Var}(\mathbf{Y})\right) &= trace\left(\widehat{\mathbf{\Delta}}'\left[\mathbf{x}^{(1)}\mathbf{x}'^{(1)} \otimes \sigma_z^2\right]\widehat{\mathbf{\Delta}}\right) + \left(1 - trace\left(\left[\mathbf{x}^{(1)}\mathbf{x}'^{(1)} \otimes \sigma_z^2\right](\mathbf{X}'_{\mathbf{\Delta}}\mathbf{X}_{\mathbf{\Delta}})^{-1}\right)\right) trace\left(\widehat{\mathbf{\Sigma}}_\epsilon\right) \\ &= \sigma_z^2\left(\widehat{\mathbf{\Delta}}'\mathbf{x}^{(1)}\right)'\left(\widehat{\mathbf{\Delta}}'\mathbf{x}^{(1)}\right) + \left(1 - \sigma_z^2\mathbf{x}'^{(1)}(\mathbf{X}'_{\mathbf{\Delta}}\mathbf{X}_{\mathbf{\Delta}})^{-\mathbf{1}}\mathbf{x}^{(1)}\right) trace\left(\widehat{\mathbf{\Sigma}}_\epsilon\right) \end{aligned}$$

and the problem is sensitive to the value of σ_z^2 only through the bias correction factor. This is because the trace is a linear operator and the value of σ_z^2 is just a multiplying positive scalar that can be factored from all the terms that contain controllable factors in the first term of equation (14). However, this is not the case for the other scalar functions mentioned in section 5, since they are non-linear functions of $\widehat{Var}(\mathbf{Y})$, and therefore the actual value of $\widehat{\mathbf{\Sigma}}_\epsilon$ has an effect on the objective value, even when there is only one noise factor. The consideration of this effect by the methodology, however, can be considered as a positive feature since the user may be interested in further reducing the variances of the responses for which the corresponding residual variance is large.

Suppose it is known that $\sigma_z^2 = 0.01$ (this value was assumed to demonstrate the methodology since it makes the elements of $\widehat{\mathbf{\Delta}}'\mathbf{\Sigma_z}\otimes\widehat{\mathbf{\Delta}}$ of similar magnitude than the elements of $\mathbf{\Sigma}_\epsilon$). The next step is to calculate the variance contribution of the noise factors:

$$\widehat{\mathbf{\Delta}}'\left[\left(\mathbf{x}^{(1)}\mathbf{x}'^{(1)}\right) \otimes 0.01\right]\widehat{\mathbf{\Delta}} =$$

$$10^{-2}\begin{bmatrix} 0.0272 & 0 & 0 \\ -0.0302 & 0.0129 & 0 \\ 0.0331 & 0.0107 & 0 \\ -0.0040 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & x_1 & x_2 \\ x_1 & x_1^2 & x_1x_2 \\ x_2 & x_1x_2 & x_2^2 \end{bmatrix} \times \begin{bmatrix} 0.0272 & -0.0302 & 0.0331 & -0.0040 \\ 0 & 0.0129 & 0.0107 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} =$$

$$10^{-6}\begin{bmatrix} 7.40 & -8.21 + 3.51x_1 & 9.00 + 2.91x_1 & -1.08 \\ -8.21 + 3.51x_1 & 9.12 - 7.80x_1 + 1.66x_1^2 & -10.00 + 1.04x_1 + 1.38x_1^2 & 1.21 - 0.52x_1 \\ 9.00 + 2.91x_1 & -10.00 + 1.04x_1 + 1.38x_1^2 & 10.95 + 7.08x_1 + 1.15x_1^2 & -1.32 - 0.43x_1 \\ -1.08 & 1.21 - 0.52x_1 & -1.32 - 0.43x_1 & 0.16 \end{bmatrix} \quad (21)$$

In this case we have that

$$\mathbf{X_\Delta} = \left[\begin{array}{ccc} IPA & IPA{\times}Temp & IPA{\times}pH \\ \hline -1 & 1 & 0 \\ -1 & -1 & 0 \\ -1 & 0 & 1 \\ -1 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{array}\right] \quad , \quad (\mathbf{X'_\Delta X_\Delta})^{-1} = \left[\begin{array}{ccc} 0.125 & 0.000 & 0.000 \\ 0.000 & 0.250 & 0.000 \\ 0.000 & 0.000 & 0.250 \end{array}\right]$$

and the bias correction factor is given by:

$$1 - trace\left(\mathbf{\Sigma_z^{\otimes}}(\mathbf{X'_\Delta X_\Delta})^{-1}\right) = 0.99875 - 0.0025(x_1^2 + x_2^2)\ .$$

Since the value of this correction factor is close to 1 for $-1 \leq x_i \leq 1$, its effect on the overall value of $\widehat{Var}(\mathbf{Y})$ will be small, and therefore, for the *trace* objective, the actual choice of $\mathbf{\Sigma_z}$ will have a small effect on the solution, as discussed previously.

Finally, we need an estimate of the residual covariance matrix, $\widehat{\mathbf{\Sigma}}_\epsilon$. This was calculated from the matrix of residuals in the following form (see for example Press(1982)):

$$\widehat{\mathbf{\Sigma}}_\epsilon = \frac{1}{N-p}\mathbf{U'U} = 10^{-4}\left[\begin{array}{cccc} 0.1499 & 0.0280 & 0.0259 & -0.0064 \\ 0.0280 & 1.0965 & 0.0510 & 0.0670 \\ 0.0259 & 0.0510 & 0.0199 & -0.0070 \\ -0.0064 & 0.0670 & -0.0070 & 0.0335 \end{array}\right]$$

where $\mathbf{U}$ is the matrix of residuals from the multivariate regression fit, N is the number of experimental observations ($N = 15$, for our current example) and p is the number of parameters fitted ($p = 9$ in this example). With these computations the objective function equals to:

$$trace\left(\widehat{Var}(\mathbf{Y})\right) =$$

$$10^{-6} \times trace\left(\left[\begin{array}{cccc} 7.40 & -8.21 + 3.51x_1 & 9.00 + 2.91x_1 & -1.08 \\ -8.21 + 3.51x_1 & 9.12 - 7.80x_1 + 1.66x_1^2 & -10.00 + 1.04x_1 + 1.38x_1^2 & 1.21 - 0.52x_1 \\ 9.00 + 2.91x_1 & -10.00 + 1.04x_1 + 1.38x_1^2 & 10.95 + 7.08x_1 + 1.15x_1^2 & -1.32 - 0.43x_1 \\ -1.08 & 1.21 - 0.52x_1 & -1.32 - 0.43x_1 & 0.16 \end{array}\right]\right)$$

$$+10^{-4}\left(0.99875 - 0.0025(x_1^2 + x_2^2)\right) \times trace\left(\left[\begin{array}{cccc} 0.1499 & 0.0280 & 0.0259 & -0.0064 \\ 0.0280 & 1.0965 & 0.0510 & 0.0670 \\ 0.0259 & 0.0510 & 0.0199 & -0.0070 \\ -0.0064 & 0.0670 & -0.0070 & 0.0335 \end{array}\right]\right)$$

$$= 10^{-6} \times (157.4 - .7082x_1 + 2.484x_1^2 - 0.3250x_2^2)$$

Peterson (2000) mentions the following as a region of desirable operating conditions for this problem:

$$\begin{array}{rcl} \mathbf{l} & = & [1.8 \quad -\infty \quad 300 \quad 0.75]' \\ \mathbf{u} & = & [\infty \quad 15 \quad \infty \quad 0.85]' \end{array} . \tag{22}$$

The set of constraints, after normalizing the bounds is:

$$\begin{bmatrix} 0.2126 \\ -\infty \\ 0.2683 \\ 0.2429 \end{bmatrix} \leq \begin{bmatrix} 0.2576 & -0.0248 & 0 & 0 & 0 & 0 \\ 0.2456 & -0.0690 & 0 & 0 & 0.0146 & 0 \\ 0.2500 & 0.0737 & -0.0046 & 0 & 0 & 0.0027 \\ 0.2556 & 0.0194 & 0 & 0 & 0.0045 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_1x_2 \\ x_1^2 \\ x_2^2 \end{bmatrix} \leq \begin{bmatrix} \infty \\ 0.2589 \\ \infty \\ 0.2753 \end{bmatrix}$$

$$-1 \leq x_1 \leq 1 \quad -1 \leq x_2 \leq 1$$

The trace of $\widehat{Var}(\mathbf{Y})$ was minimized subject to the constraints above using MATLAB's *fmincon* nonlinear optimization routine. The algorithm was started at 100 different locations inside the experimental region ($-1 \leq x_1 \leq 1 \;\; ; \;\; -1 \leq x_2 \leq 1$). There were only two convergence points and the one with the smallest objective was selected as the optimum. The problem was solved in less than 0.5 minutes using a Pentium 4 processor.

The solution obtained was $x_1 = 0.1491$ and $x_2 = -1.0000$, with a value of $trace\left(\widehat{Var}(\mathbf{Y})\right) = 15.76 \times 10^{-5}$. The expected vector of responses, after reversing the normalization, is:

$$\widehat{E}\mathbf{Y}_u = \begin{bmatrix} 2.1495 \\ 13.6534 \\ 300.0000 \\ 0.7985 \end{bmatrix}$$

where it can be seen that the lower bound on the S/N ratio response is a binding constraint.

After undoing the L_2 normalization, the resulting covariance matrix is given by:

$$\begin{aligned} \widehat{Var}(\mathbf{Y}_u) &= 10^{-3} \begin{bmatrix} 0.529 & -3.77 & 89.2 & -0.0288 \\ -3.77 & 26.8 & -635.4 & 0.205 \\ 89.2 & -635.4 & 15046 & -4.85 \\ -0.0288 & 0.205 & -4.85 & 0.00156 \end{bmatrix} + 0.99 \times 10^{-3} \begin{bmatrix} 1.07 & 1.38 & 24.6 & -0.0168 \\ 1.38 & 368.1 & 330.6 & 1.20 \\ 24.6 & 330.6 & 2482 & -2.42 \\ -0.0168 & 1.20 & -2.42 & 0.0319 \end{bmatrix} \\ &= 10^{-3} \begin{bmatrix} 1.60 & -2.39 & 113.8 & -0.0455 \\ -2.39 & 394.9 & -304.8 & 1.40 \\ 113.8 & -304.8 & 17528 & -7.26 \\ -0.0455 & 1.40 & -7.26 & 0.0335 \end{bmatrix} \end{aligned}$$

where the first term is the covariance contributed by the noise factor and the second term is the estimated covariance from the multivariate regression fit.

It is also interesting to explore the effect of changing the type of scalar function used for the optimization. Table 2 gives the results for the four types of optimization criteria mentioned in section 5. The results presented were calculated from the normalized matrices, in order to preserve optimality for each criterion.

Table 2: Comparison of Scalar Objectives, Example 1.

		Value of the other Functions at Optimality			
Scalar fun.	Opt. Sol. ($\mathbf{x}^*$)	$trace(\widehat{Var}(\mathbf{Y}))$	$\lvert\widehat{Var}(\mathbf{Y})\rvert$	$trace(\widehat{Var}(\mathbf{Y})^2)$	$\lambda_{max} - \lambda_{min}$
$trace(\widehat{Var}(\mathbf{Y}))$	[0.1491 -1.0000]'	15.76e-5	5.41e-20	1.51e-8	1.16e-4
$\lvert\widehat{Var}(\mathbf{Y})\rvert$	[0.1491 -1.0000]'	15.76e-5	5.41e-20	1.51e-8	1.16e-4
$trace(\widehat{Var}(\mathbf{Y})^2)$	[0.8472 1.0000]'	15.90e-5	7.66e-20	1.42e-8	1.11e-4
$\lambda_{max} - \lambda_{min}$	[0.8472 1.0000]'	15.90e-5	7.66e-20	1.42e-8	1.11e-4

It can be seen that the results can vary significantly between different scalar functions. In addition, it is important to mention that for the $trace(\widehat{Var}(\mathbf{Y})^2)$ and $\lambda_{max} - \lambda_{min}$, the vector of expected responses at optimality is given by:

$$\widehat{E}\mathbf{Y}_u = \begin{bmatrix} 2.0029 \\ 11.4508 \\ 347.2584 \\ 0.8500 \end{bmatrix}$$

which indicates that in this case the upper bound in the Tailing response (and not the bound on the S/N response) is a binding constraint. As in many instances of multi-criteria optimization, the final decision as to which is the most desirable solution should be made using knowledge about the actual process under study.

6.2 Whey Protein Concentrates

Khuri & Cornell (1987) report an experiment performed to investigate the effects of heating temperature (x_1), pH level (x_2), redox potential (x_3), sodium oxalate (x_4) and sodium lauryl sulfate (x_5) on foaming properties of whey protein concentrates. Measurements were made on three responses, the whipping time (Y_1), the maximum overrun (Y_2) and percent soluble protein (Y_3). Table 3 contains the experimental design and the multi-response data.

The $\widehat{\mathbf{\Delta}}$ and $\widehat{\mathbf{\Theta}}$ matrices were obtained after normalizing each response by its corresponding L_2 norm, choosing x_1 and x_3 as the noise variables, and setting to zero any non-significant parameter that would not make any response model non-hierarchical. Significance tests used the pure error from the replicated center points. The matrices are as follows:

Table 3: Factor and Response Data for the Whey Protein Example

Controllable Factors					Responses		
x_1	x_2	x_3	x_4	x_5	Y_1	Y_2	Y_3
-1	-1	-1	-1	1	4.75	1082	81.4
1	-1	-1	-1	-1	4.00	824	69.6
-1	1	-1	-1	-1	5.00	953	105.0
1	1	-1	-1	1	9.50	759	81.2
-1	-1	1	-1	-1	4.00	1163	80.8
1	-1	1	-1	1	5.00	839	76.3
-1	1	1	-1	1	3.00	1343	103.0
1	1	1	-1	-1	7.00	736	76.9
-1	-1	-1	1	-1	5.25	1027	87.2
1	-1	-1	1	1	5.00	836	74.0
-1	1	-1	1	1	3.00	1272	98.5
1	1	-1	1	-1	6.50	825	94.1
-1	-1	1	1	1	3.25	1363	95.9
1	-1	1	1	-1	5.00	855	76.8
-1	1	1	1	-1	2.75	1284	100.0
1	1	1	1	1	5.00	851	104.0
-2	0	0	0	0	3.75	1283	100.0
2	0	0	0	0	11.00	651	50.5
0	-2	0	0	0	4.50	1217	71.2
0	2	0	0	0	4.00	982	101.0
0	0	-2	0	0	5.00	884	85.8
0	0	2	0	0	3.75	1147	103.0
0	0	0	-2	0	3.75	1081	104.0
0	0	0	2	0	4.75	1036	89.4
0	0	0	0	-2	4.00	1213	105.0
0	0	0	0	2	3.50	1103	113.0
0	0	0	0	0	3.50	1179	104.0
0	0	0	0	0	3.50	1183	107.0
0	0	0	0	0	4.00	1120	104.0
0	0	0	0	0	3.50	1180	101.0
0	0	0	0	0	3.00	1195	103.0

$$\widehat{\boldsymbol{\Theta}} = \left[\begin{array}{l|rrr} & Y_1 & Y_2 & Y_3 \\ \hline I & 0.1692 & 0.1812 & 0.1813 \\ x_2 & 0.0068 & -0.0031 & 0.0145 \\ x_4 & -0.0068 & 0.0037 & 0.0022 \\ x_5 & -0.0030 & 0.0032 & 0.0032 \\ x_2x_4 & -0.0181 & 0 & 0 \\ x_2x_5 & 0 & 0 & 0 \\ x_4x_5 & -0.0125 & 0 & 0 \\ x_2^2 & 0 & -0.0022 & -0.0071 \\ x_4^2 & 0 & -0.0039 & -0.0019 \\ x_5^2 & 0 & 0 & 0.0040 \end{array}\right]$$

$$\widehat{\boldsymbol{\Delta}} = \left[\begin{array}{l|rrr} & Y_1 & Y_2 & Y_3 \\ \hline x_1 & 0.0460 & -0.0297 & -0.0159 \\ x_1x_2 & 0.0283 & -0.0042 & 0 \\ x_1x_4 & 0 & 0 & 0.004 \\ x_1x_5 & 0.0113 & -0.0062 & 0 \\ x_3 & -0.0159 & 0.0097 & 0.0046 \\ x_2x_3 & -0.0102 & 0 & 0 \\ x_3x_4 & 0 & 0 & 0 \\ x_3x_5 & -0.0091 & 0 & 0.0079 \end{array}\right]$$

The normalized estimated covariance matrix for the residual noise from the complete fitted models is given by:

$$\widehat{\boldsymbol{\Sigma}}_\epsilon = 10^{-3} \left[\begin{array}{rrr} 3.2580 & -0.7132 & -1.3049 \\ -0.7132 & 0.5304 & 0.3697 \\ -1.3049 & 0.3697 & 0.6347 \end{array}\right]$$

Since in this example we are considering 2 noise variables, we need to specify a 2×2 covariance matrix. Suppose this equals to:

$$\boldsymbol{\Sigma}_{\mathbf{z}} = \left[\begin{array}{rr} 1.00 & -0.25 \\ -0.25 & 1.00 \end{array}\right]$$

The bias correction factor is given by:

$$1 - trace\left(\boldsymbol{\Sigma}_{\mathbf{z}}^{\otimes}(\mathbf{X}'_{\boldsymbol{\Delta}}\mathbf{X}_{\boldsymbol{\Delta}})^{-1}\right) = \frac{11}{12} - \frac{1}{8}\left(x_2^2 + x_4^2 + x_5^2\right)$$

We note that the bias correction factor will be negative for any point in the controllable factor space that is at a distance greater than $\sqrt{22/3} \approx 2.71$ from the center. The estimate of the covariance matrix of the responses is given by:

$$
\begin{aligned}
&\widehat{Var}(\mathbf{Y}) = \\
&\widehat{\boldsymbol{\Delta}}' \left[\begin{bmatrix} 1.00 & -0.25 \\ -0.25 & 1.00 \end{bmatrix} \otimes \left(\mathbf{x}^{(\mathbf{l})}\mathbf{x}'^{(\mathbf{l})}\right) \right] \widehat{\boldsymbol{\Delta}} + \left(\tfrac{11}{12} - \tfrac{1}{8}\left(x_2^2 + x_4^2 + x_5^2\right)\right) \widehat{\boldsymbol{\Sigma}}_\epsilon = \\
&10^{-3} \begin{bmatrix}
2.73 + 3.39x_2 + 1.63x_5 & -1.75 - 1.29x_2 - 0.83x_5 & -0.92 + 0.20x_4 - 0.48x_5 \\
+1.05x_2^2 + 1.01x_2x_5 + 0.26x_5^2 & -0.13x_2^2 - 0.25x_2x_5 - 0.084x_5^2 & -0.56x_2 + 0.12x_2x_4 - 0.13x_2x_5 \\
 & & +0.054x_4x_5 - 0.094x_5^2 \\
-1.75 - 1.29x_2 - 0.83x_5 & 1.12 + 0.27x_2 + 0.40x_5 & 0.59 - 0.13x_4 + 0.24x_5 \\
-0.13x_2^2 - 0.25x_2x_5 - 0.084x_5^2 & +0.018x_2^2 + 0.052x_2x_5 + 0.038x_5^2 & +0.07x_2 - 0.017x_2x_4 + 0.008x_2x_5 \\
 & & -0.025x_4x_5 + 0.012x_5^2 \\
-0.92 + 0.20x_4 - 0.48x_5 & 0.59 - 0.13x_4 + 0.24x_5 & 0.31 - 0.14x_4 + 0.13x_5 \\
-0.56x_2 + 0.12x_2x_4 - 0.13x_2x_5 & +0.07x_2 - 0.017x_2x_4 + 0.008x_2x_5 & +0.016x_4^2 - 0.016x_4x_5 + 0.062x_5^2 \\
+0.054x_4x_5 - 0.094x_5^2 & -0.025x_4x_5 + 0.012x_5^2 &
\end{bmatrix} \\
&+10^{-3}\left(\tfrac{11}{12} - \tfrac{1}{8}\left(x_2^2 + x_4^2 + x_5^2\right)\right) \begin{bmatrix} 3.2580 & -0.7132 & -1.3049 \\ -0.7132 & 0.5304 & 0.3697 \\ -1.3049 & 0.3697 & 0.6347 \end{bmatrix}
\end{aligned}
\tag{23}
$$

Comparing equation (23) with equation (21), it can be observed that in this example there is more control over the covariance matrix, since the controllable factors are present in all the elements of the matrix.

The set of bounds selected for the expected responses were:

$$\mathbf{l} = [-\infty \;\; 800 \;\; 100]$$

$$\mathbf{u} = [5.0 \;\; 1100 \;\; \infty]$$

That is, the expected value of the first response is restricted to be less that 5.0 minutes, the expected second response between 800% and 1100% and the the third response should be higher than 100%. We chose a spherical region of radius 2 for controllable factors, that is we added the following constraint:

$$\sqrt{x_2^2 + x_4^2 + x_5^2} \leq 2$$

In this case the scalar objective functions were the determinant, trace and each of the individual variances, i.e. the diagonal elements of $\widehat{Var}(\mathbf{Y})$. We used the same optimization technique as in example 1. The CPU time to perform 100 runs of the optimization of each objective using random initial points was less than 0.5 minutes on a Pentium 4 computer. The results are summarized in Table 4. We included the normalized results, except for the individual variances for which the results are after reversing the normalization.

Note that the results using the trace are substantially different to the ones obtained with the determinant. In fact, the corresponding $\mathbf{x}^*$ points are separated by a distance of 0.91 coded units which is quite significant. The values of the objectives are also substantially different. The trace of the covariance matrix obtained when the determinant is minimized is about 58%

Table 4: Comparison of Scalar Objectives for Whey Protein Example.

			Value of the other Functions at Optimality				
Scalar fun.	$\mathbf{x}^* = [x_2^*\ x_4^*\ x_5^*]$	$\widehat{E}(\mathbf{Y}^*)$	$\lvert\widehat{Var}(\mathbf{Y})\rvert$	$tr(\widehat{Var}(\mathbf{Y}))$	$\widehat{Var}(Y_{1,u})$	$\widehat{Var}(Y_{2,u})$	$\widehat{Var}(Y_{3,u})$
$tr(\widehat{Var}(\mathbf{Y}))$	[0.17 0.23 -1.98]'	[4.9 1036.8 100]'	0.38e-10	0.0024	1.12	2.12e4	103.8
$\lvert\widehat{Var}(\mathbf{Y})\rvert$	[0.59 0.97 -1.65]'	[5.0 1026.8 100]'	0.32e-10	0.0038	2.14	2.73e4	73.4
$\widehat{Var}(Y_1)$	[0.17 0.17 -1.98]'	[4.9 1035.9 100]'	0.40e-10	0.0024	1.11	2.11e4	106
$\widehat{Var}(Y_2)$	[0.17 0.12 -1.99]'	[4.9 1035.1 100]'	0.41e-10	0.0024	1.11	2.11e4	108
$\widehat{Var}(Y_3)$	[0.77 0.99 -1.55]'	[4.9 1021.4 100]'	0.37e-10	0.0046	2.69	2.96e4	72.0

larger than the one obtained when the trace is minimized. Similarly, the determinant obtained when the trace is minimized is about 18% larger than when the determinant is minimized.

Furthermore, notice that the solution obtained using the trace and the variances of the first and second responses are significantly close to each other. Therefore, using the trace as objective may not capture the overall variability of the vector of responses and could be affected by individual responses, especially if more than one of them is minimized close to the same point.

A more interesting result in this example is that the individual variances for the first two responses obtained when minimizing the determinant are higher to the corresponding variances obtained when minimizing the trace. Therefore, the determinant solution could hardly be obtained from considering only the individual variances.

The above observations underscores the advantages of considering the complete covariance matrix instead of the individual variances only, and demonstrates the benefits of the proposed multivariate approach over using the univariate approach (equation 2) q times in parallel.

The following are the resulting optimal covariance matrices for each criterion (after reversing the normalization):

$$\widehat{Var}(\mathbf{Y}_u)_{tr(\widehat{Var}(\mathbf{Y}))} = \begin{bmatrix} 1.1195 & -117.05 & -8.9489 \\ -117.05 & 21231 & 732.87 \\ -8.9489 & 732.87 & 103.79 \end{bmatrix}$$

$$\widehat{Var}(\mathbf{Y}_u)_{|\widehat{Var}(\mathbf{Y})|} = \begin{bmatrix} 2.1418 & -215.72 & -10.277 \\ -215.72 & 27306 & 874.04 \\ -10.277 & 874.04 & 73.486 \end{bmatrix}$$

$$\widehat{Var}(\mathbf{Y})_{\widehat{Var}(Y_1)} = \begin{bmatrix} 1.117 & -116.62 & -9.0185 \\ -116.62 & 21183 & 737.23 \\ -9.0185 & 737.23 & 106.08 \end{bmatrix}$$

$$\widehat{Var}(\mathbf{Y}_u)_{\widehat{Var}(Y_2)} = \begin{bmatrix} 1.1188 & -116.72 & -9.0897 \\ -116.72 & 21170 & 742.13 \\ -9.0897 & 742.13 & 108 \end{bmatrix}$$

$$\widehat{Var}(\mathbf{Y}_u)_{\widehat{Var}(Y_3)} = \begin{bmatrix} 2.6984 & -259.37 & -11.31 \\ -259.37 & 29692 & 948.7 \\ -11.31 & 948.7 & 72.032 \end{bmatrix}$$

7 Discussion

The proposed methodology requires that all the response functions have the same form and be estimated from the same experimental design. Since in practice some coefficients will be statistically non-significant and the true functional relationships will be different, a methodology that allows for the responses to have different functional relationships would avoid the approximation of either setting non-significant terms to zero or including them in the respective models. Extensions of this type have been done for other multiple response methodologies (e.g. Khuri and Valeroso, 1999) without incurring in significant extra complexity. Their approach consisted in modelling the multivariate response using a seemingly unrelated regressions-type of model of the following form (note that Khuri and Valeroso's methodology is not intended for modelling noise variables):

$$\widetilde{\mathbf{Y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \epsilon \tag{24}$$

where :

- $\tilde{\mathbf{Y}} = [\mathbf{Y}_1', \cdots \mathbf{Y}_q']'$, $\mathbf{Y}_j$ is data for j^{th} response.
- $\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_q \end{bmatrix}$, $\mathbf{X}_j$ is design matrix for j^{th} response.
- $\boldsymbol{\beta}$ is $p \times 1$, $p = \sum_{j=1}^{q} p_j$, p_j is number of parameters in model for j^{th} response.
- $\boldsymbol{\epsilon} \sim N_q(\mathbf{0}, \boldsymbol{\Sigma})$.

The estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ is usually done using a two stage Aitken estimator (Zellner, 1962, Khuri and Cornell, 1987). However, these estimators are not unbiased and their distributional properties are not very well understood. Therefore, the development of unbiased efficient estimators for this case is much more complicated than the one presented in this paper.

Although the objective of conducting an RPD study is generally to minimize the variance due to the noise factors, for which the estimator developed in section 3 is unbiased, the resulting estimate can be negative. Probably a more sensible approach would be to minimize the estimated prediction variance $\widehat{Var}(\widehat{\mathbf{Y}})$ which includes the variance of the model parameters estimates. If this is done, the prediction properties of the model used will also be considered together with the variance of the process.

The proposed methodology does not demand the use of a specific type of experimental design, besides having the ability of estimating all the parameters. However, the type of mixed resolution designs (Borkowski and Lucas, 1997, Borror and Montgomery, 2000, Borkowski and Lucas, 1997), that has been developed for the univariate case (equation, 1) will be a suitable choice in most situations because the current methodology assumes that all the responses will be modelled using the same model form. These designs give special attention to estimating the noise×control interactions which are crucial in the RPD problem.

The topic of experimental designs for multiple response has received some attention in the literature (Draper and Hunter, 1966, Fedorov, 1972, and Khuri and Cornell, 1987). However,

very little of this work has been used in practical situations. Even when the model form is the same, most of the techniques require previous knowledge of the covariance matrix of the responses and hence are useful for augmenting a previous design.

Acknowledgements.-

This research was funded by NSF grant DMI 9988563. The authors wish to express his gratitude to an anonymous referee for his/her thorough review and detailed verifications of the numerical results in Section 6.

References

Arnold, S.F. (1981). *The Theory of Linear Models and Multivariate Analysis.* Wiley, New York, NY.

Borkowski, J.J. and Lucas, J.M. (1997). "Design of Mixed Resolution for Process Robustness Studies". *Technometrics*, 39, pp. 63-70.

Borror, C.M. and Montgomery, D.C. (2000). "Mixed Resolution Designs as Alternatives to Taguchi Inner/Outer Array Designs for Robust Design Problems ". *Quality and Reliability Engineering International*, 16, pp. 117-127.

Box, G.E.P. and Jones, S. (1990). "Designing Products That Are Robust To The Environment. *Report Series in Quality and Productivity.* CPQI, University of Wisconsin, Number 56.

Chiao, C. and Hamada, M. (2001). "Analyzing Experiments with Correlated Multiple Responses". *Journal of Quality Technology*, 33, pp. 451-465.

Draper, N.R. and and Hunter, W. G. (1966). " Design of Experiments for Parameter Estimation in Multiresponse Situations" . *Biometrika*, 53, pp. 525-533.

Federov, V.V. (1972). *Theory of Optimal Experiments.* Academic Press, New York, NY.

Hamilton, J. D. (1994).*Time Series Analysis.* Princeton University Press, Princeton, NJ.

Khuri, A.I. and Cornell, J.A. (1987) *Response Surfaces: Design and Analysis.* Marcel Dekker, New York, NY.

Khuri, A.I. and Valeroso, E.S. (1999). "Optimization Methods in Multiresponse Surface Metholodology". In: *Statistical Process Monitoring and Optimization*, Sung H. Park and G.G. Vining, Eds., New York: Marcel Dekker, pp. 411-433.

Lucas, J.M. (1994). "How to Achieve a Robust Process Using Response Surface Methodology". *Journal of Quality Technology*, 26, pp. 248-260.

Myers, R.H. (1991). "Response Surface Methodology in Quality Improvement". *Comm. Stat.:*

Theory and Methods, 20(2),457-476.

Myers, R.H., Kim, Y. and Griffiths, K.L. (1997). "Response Surface Methods and the use of Noise Variables.*Journal of Quality Technology*, 29, pp. 429-440.

Myers, R.H. and Montgomery, D.C. (2002). *Response Surface Methodology : Process and Product Optimization Using Designed Experiments*. Wiley, New York, NY.

Nair, V. N., editor, (1992). "Taguchi's Parameter Design: A Panel Discussion," *Technometrics*, 34, 2, pp. 127-161.

Neudecker, H. (1969). "Some Theorems on Matrix Differentiation with Special Reference to Kronecker Matrix Products". *Journal of the American Statistical Association*, 64, 953-963.

Peterson, J.J. (2000). "A Probability-based Desirability Function for Multiresponse Optimization", *Proceedings of the Section on Quality and Productivity*, American Statistical Association Joint Statistical Meetings, Indianapolis, IN.

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

Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance Components*. Wiley, New York, NY.

Steuer, R.E. (1989). *Multiple Criteria Optimization: Theory, Computation and Application*. R. E. Krieger Pub. Co. Malabar, FL.

Taguchi, G. (1986). *Introduction to Quality Engineering*. Quality Resources, White Plains, NJ.

Taguchi, G. (1987). *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost*. Quality Resources, White Plains, NJ.

Taguchi, G. and Wu, Y. (1985). *Introduction to Off-Line Quality Control*. Central Japan Quality Control Association (available from American Supplier Institute, Dearborn, MI).

Vining, G.G. and Myers, R.H. (1990). "Combining Taguchi and Response Surface philosophies: A dual response approach". *Journal of Quality Technology*, 22, pp. 38-45.

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

Zellner, A. (1962). "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregated Bias," *Journal of the American Statistical Association*, 57, pp. 348-368.

Appendix

The following is a modification of a result due to Neudecker (1969).

Proposition 1 *Let* $\mathbf{a}_{\cdot j}$ *denote the* j^{th} *column of the matrix* $\mathbf{A}$*. Define the* $(mn \times 1)$ *column vector* $\text{vec}(\mathbf{A})$ *of the* $m \times n$ *matrix* $\mathbf{A}$ *as:*

$$\text{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_{\cdot 1} \\ \mathbf{a}_{\cdot 2} \\ \vdots \\ \mathbf{a}_{\cdot n} \end{bmatrix}$$

.

Let $\mathbf{X_1}$*,* $\mathbf{X_2}$*,* $\mathbf{Z_1}$*,* $\mathbf{Z_2}$ *and* $\mathbf{B}$ *be matrices such that the product* $\mathbf{X_1 Z_1 B Z_2}'\mathbf{X_2}$ *is conformable. Then we have that:*

$$\textit{trace}\left(\mathbf{X_1 Z_1 B Z_2}'\mathbf{X_2}\right) = (\text{vec}(\mathbf{Z_2}))'(\mathbf{B}' \otimes \mathbf{X_1}'\mathbf{X_2}')\text{vec}(\mathbf{Z_1})$$

First we need to prove that (Hamilton, 1994, p. 289):

$$\text{vec}\left(\mathbf{ABC}\right) = \left(\mathbf{C}' \otimes \mathbf{A}\right)\text{vec}(\mathbf{B})$$

$$\mathbf{ABC} = \mathbf{A}\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_r \end{bmatrix}\begin{bmatrix} c_{11} & c_{12} & \dots & c_{1q} \\ c_{21} & c_{22} & \dots & c_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ c_{11} & c_{12} & \dots & c_{1q} \end{bmatrix} =$$

$$[\{\mathbf{Ab}_{\cdot 1}c_{11} + \mathbf{Ab}_{\cdot 2}c_{21} + \ldots + \mathbf{Ab}_{\cdot r}c_{r1}\} \quad \{\mathbf{Ab}_{\cdot 1}c_{12} + \mathbf{Ab}_{\cdot 2}c_{22} + \ldots + \mathbf{Ab}_{\cdot r}c_{r2}\} \quad \ldots \quad \{\mathbf{Ab}_{\cdot 1}c_{1q} + \mathbf{Ab}_{\cdot 2}c_{2q} + \ldots + \mathbf{Ab}_{\cdot r}c_{rq}\}]$$

Then, since all of the elements inside the curly brackets are column vectors, applying the vec$(\cdot)$ operator gives:

$$\begin{aligned} \text{vec}(\mathbf{ABC}) &= \begin{bmatrix} \mathbf{Ab}_{\cdot 1}c_{11} + \mathbf{Ab}_{\cdot 2}c_{21} + \ldots + \mathbf{Ab}_{\cdot r}c_{r1} \\ \mathbf{Ab}_{\cdot 1}c_{12} + \mathbf{Ab}_{\cdot 2}c_{22} + \ldots + \mathbf{Ab}_{\cdot r}c_{r2} \\ \vdots \\ \mathbf{Ab}_{\cdot 1}c_{1q} + \mathbf{Ab}_{\cdot 2}c_{2q} + \ldots + \mathbf{Ab}_{\cdot r}c_{rq} \end{bmatrix} \\ &= \begin{bmatrix} c_{11}\mathbf{A} + c_{21}\mathbf{A} + \ldots + c_{r1}\mathbf{A} \\ c_{12}\mathbf{A} + c_{22}\mathbf{A} + \ldots + c_{r2}\mathbf{A} \\ \vdots \\ c_{1q}\mathbf{A} + c_{2q}\mathbf{A} + \ldots + c_{rq}\mathbf{A} \end{bmatrix}\begin{bmatrix} \mathbf{b}_{\cdot 1} \\ \mathbf{b}_{\cdot 2} \\ \vdots \\ \mathbf{b}_{\cdot r} \end{bmatrix} \\ &= \left(\mathbf{C}' \otimes \mathbf{A}\right) \cdot \text{vec}(\mathbf{B}) \end{aligned}$$

We also need the following result due to Neudecker (1969):

$$trace(\mathbf{AB}) = \text{vec}(\mathbf{A}')'\text{vec}(\mathbf{B})$$

where $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times p$.

$$\begin{aligned}
\text{vec}(\mathbf{A}')'\text{vec}(\mathbf{B}) &= \left[\begin{array}{cccc} \mathbf{a}_{1\cdot} & \mathbf{a}_{2\cdot} & \ldots & \mathbf{a}_{n\cdot} \end{array} \right] \left[\begin{array}{c} \mathbf{b}_{\cdot 1} \\ \mathbf{b}_{\cdot 2} \\ \vdots \\ \mathbf{b}_{\cdot n} \end{array} \right] \\
&= \textstyle\sum_{i=1}^{p} \mathbf{a}_{i\cdot}\mathbf{b}_{\cdot i} \\
&= trace(\mathbf{AB})
\end{aligned}$$

Using these two results, the proof of the proposition is trivial:

$$\begin{aligned}
trace\left(\mathbf{X_1 Z_1 B Z_2}'\mathbf{X_2}\right) &= trace\left(\mathbf{Z_2}'\mathbf{X_2 X_1 Z_1 B}\right) \\
&= \text{vec}(\mathbf{Z_2})'\text{vec}(\mathbf{X_2 X_1 Z_1 B}) \\
&= \text{vec}(\mathbf{Z_2})'(\mathbf{B}' \otimes \mathbf{X_1}'\mathbf{X_2}')\text{vec}(\mathbf{Z_1})
\end{aligned}$$