Strategies for Multi-Response Parameter Design using Loss Functions and Joint Optimization Plots

M. Erdbrügge, S. Kuhnt

Department of Statistics, University of Dortmund

44221 Dortmund, Germany

Abstract

The development of high-quality products or production processes can often be greatly improved by statistically planned and analysed experiments. Taguchi methods proved to be a milestone in this field, suggesting optimal design settings for a single measured response. However, these often fail to meet the needs of today's products and manufacturing processes, which require simultaneous optimization over several quality characteristics. Current extensions for handling multi-responses assume that all responses are weighted beforehand in terms of costs due to deviations from desired target settings. Such information is usually unavailable, especially with manufacturing processes. As an alternative solution, we propose strategies that use sequences of possible weights assigned to each of the multiple responses. For each weighting a design factor combination is derived, which minimizes a

respective estimated multivariate loss function and is optimal with respect to some compromise of the responses. This compromise can be graphically displayed to the engineer, who can thereby gain much more insight into the production process and draw more valuable conclusions.

1 Introduction

The design of robust products through off-line quality control, before actual manufacturing, has been popularized by the Japanese engineer and statistician G. Taguchi. Robust parameter design tries to prescribe a setting of design parameters that minimizes the effect of hard-to-control factors or noise while still providing a high quality of the product or process. The classical approach by Taguchi has already been widely extended and supplemented (see Grize, 1995, Leon et al., 1987, Taguchi and Phadke, 1984). Most of this research focuses on determining the optimal level settings of process parameters for products with a single measured quality characteristic. In practice, however, for most products more than one quality characteristic is of interest. A separate analysis of each response can yield some information about the process, but this approach often results in conflicting recommendations regarding the parameter setting. Possibly, a setting which simultaneously improves all responses is even missed or overlooked. Recent extensions of robust parameter design strategies to the multi-response problem have involved the analysis of desirabilities (Derringer and Suich, 1980), combining signal-to-noise ratios (Chen, 1997) and applications of principal component analysis (Su and Tong, 1997).

Pignatiello (1993) has provided a straightforward extension to the multi-response case, following the Taguchi philosophy. However, a serious problematic issue in the approaches of Pignatiello (1993), Vining (1998) and others, in practical off-line planning processes, is the assumption of a pre-specified cost matrix. An engineer will, at this stage, rather be able to assign a relative importance to each response than to specify actual costs incurred for specific response values. Tong and Su (1997) use the conversion of linguistic terms, in which relative importances of the responses given by the engineer are translated into corresponding fuzzy numbers. But like most other approaches, the authors consider only one cost matrix from which it can only be assumed that a reasonable compromise is reached (i. e. compromise design factor levels are found that achieve relatively good results for all responses). However, in practice, a single response variable will often dominate the expected loss by its large variability or large deviation from target. This implies that a "real compromise" is attained in relatively few cases, since often a cost matrix used leads to factor levels optimized with respect to only one of the responses. We therefore propose to consider a range of cost matrices simultaneously, instead of specifying one fixed cost matrix. For each of these matrices, a respective optimal design factor combination can be found and, at these combinations, comparisons are then possible between predicted response means and variances.

Following the concept of multivariate robust parameter design, which is summarized in Section 2, a design factor combination is optimal if it minimizes the expected loss. The loss itself, of course, depends again on the used cost matrix. Specific choices are proposed and discussed in Section 3. A better understanding of these matrices is achieved by factoring each into a product of "standardizing" and "weighting" matrices. Data-driven choices of the standardizing matrix component are proposed. The notion of a joint optimization plot is introduced in Section 4, intended to visualize the effect of different weight matrices. In Section 5, the proposed methods are applied to a data set of sheet metal forming. We conclude with a short summary.

2 Multivariate Robust Parameter Design

Let a product or production process be considered, which can be characterized by a vector of quality characteristics $\mathbf{Y} = (Y_1, \ldots, Y_p)'$. We further presume, that a vector of finite target values $\tau = (\tau_1, \ldots, \tau_p)'$ is given. If $\tau_r = \infty$ for any response Y_r, the respective response will be transformed to $\widetilde{Y}_r = 1/Y_r$ with $\widetilde{\tau}_r = 0$. The random vector $\mathbf{Y}$ is assumed to depend functionally on a vector of design parameters $\mathbf{x} = (x_1, \ldots, x_k)' \in \mathcal{X}$ and a noise vector $\mathbf{Z} = (Z_1, \ldots, Z_l)' \in \mathcal{Z}$, that is $\mathbf{Y} = f(\mathbf{x}, \mathbf{Z})$. The distribution of $\mathbf{Y}$ can vary conditioned on the controllable design parameter values $x_1, \ldots, x_k$. For a fixed $\mathbf{x}$, the random vector $\mathbf{Y}$ will be denoted by $\mathbf{Y}|\mathbf{x}$ with expectation $E(\mathbf{Y}|\mathbf{x}) = \mu(\mathbf{x})$ and covariance matrix $Cov(\mathbf{Y}|\mathbf{x}) = \Sigma(\mathbf{x}) = \left[\left(\sigma_{rs}(\mathbf{x})\right)_{\substack{r=1,\ldots,p \\ s=1,\ldots,p}}\right]$. Note that, in the sense of Taguchi, we not only take the response mean to depend on the design parameters but also the covariance matrix. In most cases, it is assumed that $\mathbf{Y}|\mathbf{x}$, or a suitable transformation thereof, follows a multivariate normal distribution for each $\mathbf{x} \in \mathcal{X}$.

The case of a single response ($p = 1$) is treated by the traditional Taguchi approach (see Taguchi and Phadke, 1984). The overall quality of the product is thereby seen in terms of loss resulting from the deviation of quality characteristics from their target values. Taguchi measures this loss for a specific outcome y by the quadratic loss function: $loss(y) = c(y-\tau)^2$, where c is a constant. To derive this constant, it suffices to know the costs occurring for a single deviating value y. The aim of robust parameter design methods is to find combinations of the design parameter values which minimize the risk function $R(\mathbf{x}) = E_{\mathbf{Z}}(loss(Y|\mathbf{x}) = E_{\mathbf{Z}}(f(\mathbf{x}, \mathbf{Z}))$, representing the expectation of the loss taken over the noise. The use of signal-to-noise ratios to achieve this aim, as propagated by Taguchi, has been supplemented by Leon et al. (1987). They describe distributional situations for which the maximization of signal-to-noise ratios equals the minimization of the expected loss and add the notion of performance measures independent of adjustment.

As an extension of the quadratic loss function to multivariate quality characteristics the quadratic form

$$loss(\mathbf{y}) \quad = \quad \big(\mathbf{y}-\tau\big)' C \big(\mathbf{y}-\tau\big) \tag{1}$$

has been proposed by Pignatiello (1993), where C denotes a $p \times p$ dimensional symmetric cost matrix. For the resulting risk function it holds that

$$\begin{aligned} R(\mathbf{x}) \quad &= \quad E_{\mathbf{Z}}\big(loss(\mathbf{Y}|\mathbf{x})\big) = E_{\mathbf{Z}}((f(\mathbf{x}, \mathbf{Z}) - \tau)' C (f(\mathbf{x}, \mathbf{Z}) - \tau)) \\ &= \quad \mathrm{trace}\big(C\Sigma(\mathbf{x})\big) + \big(\mu(\mathbf{x}) - \tau\big)' C \big(\mu(\mathbf{x}) - \tau\big)\,. \end{aligned} \tag{2}$$

Hence, minimizing the risk function will, roughly spoken, simultaneously bring the mean on target and minimize variances-covariances. On the basis of suitably designed experiments, different strategies have been proposed to solve the parameter design problem (see e. g. Pignatiello, 1993, p. 8-13). In a direct approach, empirical means and covariances are derived for every design point i based on $n_i > 1$ replications or noise factor combinations. The resulting risk

$$\bar{R}(x_i) = \operatorname{trace}(CS(x_i)) + \big(\bar{\mu}(x_i) - \tau\big)' C \big(\bar{\mu}(x_i) - \tau\big)\,, \quad i = 1, \ldots, m\,,$$

$\bar{\mu}(x_i) = (\bar{y}_{1i}, \ldots, \bar{y}_{pi})$, where $\bar{y}_{ri} = \frac{1}{n_i}\sum_{u=1}^{n_i} y_{riu}$ and $S(x_i) = \left[\big(s_{rs}(x_i)\big)_{\substack{r=1,\ldots,p\\ s=1,\ldots,p}}\right]$ with elements $s_{rs}(x_i) = \frac{1}{n_i - 1}\sum_{u=1}^{n_i}(y_{riu} - \bar{y}_{ri})(y_{siu} - \bar{y}_{si})$ is fitted by a linear model in the design factors. This approach has a major disadvantage in that we do not really gain information about the multivariate process but rely solely on the chosen weight matrix and the optimal factor setting achieved by minimizing the associated expected loss. Furthermore, losses achieved across different weight matrices C are not comparable.

In addition, Pignatiello (1993) has suggested a response model based strategy of relating the expected values of the responses to the design variables through fitting a response model. From this, the predicted expected loss is derived by

$$\widehat{R}(\mathbf{x}) = \operatorname{trace}\Big(C\widehat{\Sigma}(\mathbf{x})\Big) + \big(\widehat{\mu}(\mathbf{x}) - \tau\big)' C \big(\widehat{\mu}(\mathbf{x}) - \tau\big)\,, \qquad (3)$$

where the fitted mean $\widehat{\mu}$ results from the response model. However, it remains rather unclear how an estimated covariance matrix $\widehat{\Sigma}(\mathbf{x})$, which varies in $\mathbf{x}$, can be derived from a response model as proposed by Pignatiello (1993).

We therefore suggest to fit separate models for the mean and variance of the responses and minimize the expected loss resulting from these models. For a single response, such model building has been addressed by the dual response approach which can be used for replicated experiments (possibly replicates over noise factors), see Grize (1995). Namely, (after suitable transformation) separate linear models for the mean and variance of each response variable $Y_r,\ r = 1, \ldots, p$, are used. For the mean, a weighted linear model is fitted to the sample means $\bar{y}_{ri}$ resulting in

$$\widehat{E}(Y_r|x_i) \quad = \quad \alpha_{r0} + \sum_{j=1}^{k} \alpha_{rj}\, x_{ij}\,, \tag{4}$$

with weights being the reciprocal of the estimated variances. These can be obtained for instance by fitting a linear model to the sample variances $\widehat{\sigma}_{rr}(x_i) = \widehat{\sigma_r^2}(x_i)$ resulting in

$$\widehat{Var}(Y_r|x_i) \quad = \quad \gamma_{r0} + \sum_{j=1}^{k} \gamma_{rj}\, x_{ij}$$

or by a multiplicative variance model resulting in

$$\widehat{Var}(Y_r|x_i) \quad = \quad \exp\left\{\gamma_{r0} + \sum_{j=1}^{k} \gamma_{rj}\, x_{ij}\right\}. \tag{5}$$

The latter can be fitted by using a gamma GLM with log link (compare Engel and Huele, 1996). When using the first model, the parameter space needs to be restricted in order to yield positive variance estimates. For this reason and the interpretive ease of the second model, (5) is more commonly used for variances. Factors x_j with non-significant influences may be discarded from these models by setting $\alpha_{rj} = 0$ or $\gamma_{rj} = 0$, respectively.

If the number of replicates is small, as in the extreme case of unreplicated experiments, variance model (5) may also be based on the squared residuals of the fitted mean model $e_{iu}^2 = \left(y_{riu} - \widehat{E}(Y_r|x_i)\right)^2$. The fitting procedure for both models (4) and (5) then involves iteratively reweighted least squares (IRLS), which alternates between fitting the mean and variance model, always using the actual estimates (see Engel and Huele, 1996).

For correlated responses, estimates for the correlation or covariance structure are also needed, at least if non-diagonal cost matrices are to be applied. Chiao and Hamada (2001) propose the model

$$\widehat{Corr}(Y_r, Y_s|x_i) = \tanh\left\{\delta_{rs0} + \sum_{j=1}^{k} \delta_{rsj}\, x_{ij}\right\}$$

for fitting models to the empirical correlation between two responses.

Joint mean and variance estimation can further employ double generalized linear models if dependencies between means and variances cannot be removed by data transformation. Generalized linear model approaches as proposed by McCullagh and Nelder (1989) or Nelder and Lee (1991) are powerful tools to handle such dependencies as well as arbitrary distributions from the exponential family.

All previously described techniques can be used to derive an estimated risk function $\widehat{R}(\mathbf{x})$ which can then be minimized with respect to design parameters $\mathbf{x}$ after a cost matrix C has been specified. The question of how to find a sensible matrix C will be addressed in the subsequent section.

3 Choice of the matrix C

In case of one quality characteristic, the constant c from the quadratic loss function is traditionally determined by a cost associated with a specific response value. Similarly Pignatiello (1993) suggests a cost related matrix C, where the choice of the entries of C is based on the loss suffered for a specific number of artificial responses. However, such an approach has certain disadvantages. For higher dimensional response vectors $\mathbf{Y}$, costs must be known for a large number of response vectors. Moreover, it is quite unrealistic to assume that such costs can always be specified beforehand. This is especially true in the development stage of a product. The choice of the matrix C will therefore often be quite arbitrary and consequently, optimal design parameter values are derived, for which it is rather obscure what they optimize. Hence, we derive and suggest alternative ways to determine a matrix C in two steps. First we investigate the influence of specific entries of C on the loss function. After which, we partition the matrix C into a product of "standardization" and "weight" matrices.

3.1 Effect of the entries of C on the loss

Every entry c_{rs} of the matrix C can have a major influence on the multivariate loss in (1), as can easily be seen from

$$\begin{aligned} loss(\mathbf{Y}) &= (\mathbf{Y}-\tau)'C(\mathbf{Y}-\tau) = \sum_{r=1}^{p}\sum_{s=1}^{p}(Y_r-\tau_r)c_{rs}(Y_s-\tau_s) \\ &= \sum_{r=1}^{p}(Y_r-\tau_r)^2 c_{rr} + \sum_{r\neq s}^{p}(Y_r-\tau_r)c_{rs}(Y_s-\tau_s). \end{aligned}$$

Obviously the diagonal elements of C give weight to the deviation of the respective responses from their target values. The off-diagonal elements distinguish between a pure additive structure, consisting of weighted losses of the single responses and the integration of combined deviations. In particular, entries $c_{rs} < 0$ penalize deviations of the two responses Y_r and Y_s from their respective target values τ_r and τ_s in opposite directions (one positive, one negative) while entries $c_{rs} > 0$ penalize deviations in the same direction (both positive or both negative). Depending on the weight one wishes to assign, these values should be chosen so that $c_{rs} \in \left(-\sqrt{c_{rr}c_{ss}}, \sqrt{c_{rr}c_{ss}}\right)$ for

$s \neq r$. This restriction assures that C is positive definite, i.e. the case of negative losses ($(\mathbf{Y} - \tau)' \, C' \, (\mathbf{Y} - \tau) \leq 0$) is excluded.

To demonstrate the effect of nonzero, off-diagonal entries c_{rs}, we will consider a simple example with two cases. Let Y_1 and Y_2 represent two quality characteristics with target values $\tau_1 = \tau_2 = 0$ and suppose we assign the same weight to both individual deviations: $c_{11} = c_{22} = 1$. The off-diagonal elements of C are chosen so that $c_{12} = c_{21}$, where both equal 0.5 in the first case and -0.5 in the second case. We can then compare the two corresponding loss functions:

$$loss_1((y_1, y_2)') = y_1^2 + y_2^2 + y_1 y_2$$

and

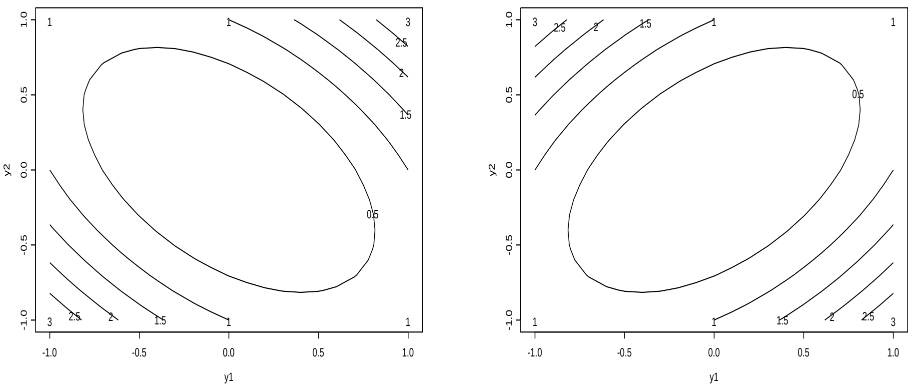
$$loss_2((y_1, y_2)') = y_1^2 + y_2^2 - y_1 y_2.$$

The effect of the two choices of off-diagonal elements can be seen in Figure 1. The positive weighting $c_{12} = 0.5$ leads to increasing losses if both responses simultaneously deviate in the same direction from the target zero. Contrary to this, the loss increases if both responses deviate in different directions for the negative weight $c_{12} = -0.5$.

This simple example already indicates that non-zero off-diagonal elements of C have a rather complex influence on the loss function and should only be selected after serious considerations.

Figure 1: Two example loss functions $loss((y_1, y_2)')$

$y_1^2 + y_2^2 + y_1 y_2$ $(c_{12} = 0.5)$ $y_1^2 + y_2^2 - y_1 y_2$ $(c_{12} = -0.5)$



3.2 Partitioning of C and data-driven choice of the standardisation part

We now consider the problem of choosing a useful matrix C. Quality characteristics are often measurable on different scales. But whether one of the quality characteristic is measured in grammes or in kilo grammes, for example, optimal design specifications should remain the same.

We therefore consider the possibility that the response vector $\mathbf{y}$ may also be measured on a different scale $\tilde{\mathbf{y}} = V\mathbf{y}$, where V denotes a diagonal $(p \times p)$ matrix. We assume that the target value is in this case also determined on the alternative scale, $\tilde{\tau} = V\tau$. The resulting change in the loss function can be seen from

$$loss(\tilde{\mathbf{y}}) \;=\; \big(V\mathbf{y} - V\tau\big)'C_{\tilde{\mathbf{y}}}\big(V\mathbf{y} - V\tau\big) = \big(\mathbf{y} - \tau\big)'V'C_{\tilde{\mathbf{y}}}V\big(\mathbf{y} - \tau\big), \quad (6)$$

where the cost matrix is labeled by $\tilde{\mathbf{y}}$ to denote the use of transformed responses. Comparison with the loss function (1) shows that the loss will be invariant to a scale transformation if $V'C_{\tilde{\mathbf{y}}}V = C_{\mathbf{y}}$, for any diagonal matrix V. To find cost matrices with this invariance property, we suggest a factorization of $C_{\mathbf{y}}$ into a product of a diagonal standardization matrix $A_{\mathbf{y}}$ and an actual weight matrix W according to $C_{\mathbf{y}} = A'_{\mathbf{y}}WA_{\mathbf{y}}$. One possible and natural choice for W would be the identity matrix $W = I$.

It follows then from

$$V'C_{\tilde{\mathbf{y}}}V = V'A'_{\tilde{\mathbf{y}}}WA_{\tilde{\mathbf{y}}}V$$

that the condition $VA_{\tilde{\mathbf{y}}} = A_{\mathbf{y}}$ ensures the loss function will be unaffected by a scale transformation. Consequently, we further examine selection of the standardization matrix, $A_{\tilde{\mathbf{y}}}$. As a first possible choice and example consider a diagonal matrix with the inverses of the target values as entries, which will be denoted by $A_{1,\mathbf{y}}$. Namely, assuming non-zero target values, we define:

$$A_{1,\mathbf{y}} = \text{diag}\left(\left[1/\tau_r\right]_{r=1,\dots,p}\right). \quad (7)$$

For the transformed response $\tilde{\mathbf{y}}$ and target $\tilde{\tau}$ it holds that

$$\begin{aligned} V\mathrm{A}_{1,\tilde{\mathbf{y}}} &= V\,\mathrm{diag}\left([1/\tilde{\tau}_r]_{r=1,\dots,p}\right) = \mathrm{diag}\left([v_r/v_r\tau_r]_{r=1,\dots,p}\right) \\ &= \mathrm{diag}\left([1/\tau_r]_{r=1,\dots,p}\right) = \mathrm{A}_{1,\mathbf{y}}, \end{aligned}$$

so that the use of $A_{1,\mathbf{y}}$ prevents the loss and the expected loss from being affected by a scale transformation of the responses. If, however, one of the target values τ_r, $r = 1, \dots, p$, equals zero, as will often be the case, $A_{1,\mathbf{y}}$ can not be applied.

We next consider data-driven standardization matrices, $\mathrm{A}_{\mathbf{y}}$. Assume that we have a sample $(\mathbf{y}_1, \dots, \mathbf{y}_n)$ of n observed response vectors, each of length p, $\mathbf{y}_l = (y_{1l}, \dots, y_{pl})'$, $l = 1, \dots, n$. We suppose, that the experimental design is based on m different design points $\mathbf{x}_i = (x_{i1}, \dots x_{ik})'$, $i = 1, \dots, m$. For each design point the number of available replications in the experiment is denoted by n_i, adding up to $n = \sum_{i=1}^{m} n_i$ total observations.

As noted in Section 2, direct estimates or predictions from models of the mean and covariance can be substituted into the risk function (2) to achieve an estimated risk. Since any sensible moment estimator or prediction model will possess linear operator properties (as true expectations do), it can be assumed that $\widehat{E}(\tilde{\mathbf{Y}}|\mathbf{x}) = V\,\widehat{\mu}(\mathbf{x})$ and $\widehat{Cov}(\tilde{\mathbf{Y}}|\mathbf{x}) = V\,\widehat{\Sigma}(\mathbf{x})V'$. In which cases the estimated expected loss function (3) of a transformed response is given by

$$\widehat{R}_{\tilde{\mathbf{Y}}}(\mathbf{x}) \quad = \quad \operatorname{trace}\big(V'C_{\tilde{\mathbf{y}}}V\widehat{\Sigma}(\mathbf{x})\big) + \big(\widehat{\mu}(\mathbf{x}) - \tau\big)'V'C_{\tilde{\mathbf{y}}}V\big(\widehat{\mu}(\mathbf{x}) - \tau\big),$$

and it is obvious then that the invariance condition $V'C_{\tilde{\mathbf{y}}}V = C_{\mathbf{y}}$ transfers automatically to the estimated risk. As a straightforward data-driven alternative to $A_{1,\mathbf{y}}$ we propose the use of

$$A_{2,\mathbf{y}} = \operatorname{diag}\left(\left[\frac{1}{m}\sum_{i=1}^{m}\widehat{E}(Y_r|x_i)\right]^{-1}_{r=1,\dots,p}\right)$$

for the minimization of the estimated risk. As a third possibility, we introduce

$$A_{3,\mathbf{y}} = \operatorname{diag}\left(\left[\frac{1}{m}\sum_{i=1}^{m}\widehat{Var}(Y_r|x_i)\right]^{-1/2}_{r=1,\dots,p}\right).$$

For the two data-based choices above, we have

$$\begin{aligned} VA_{2,\tilde{\mathbf{y}}} \quad &= \quad V\left[\operatorname{diag}\left(\left[\frac{1}{m}\sum_{i=1}^{m}\widehat{E}(\tilde{Y}_r|x_i)\right]^{-1}_{r=1,\dots,p}\right)\right] \\ &= \quad \operatorname{diag}\left(\left[\frac{v_r}{v_r m}\sum_{i=1}^{m}\widehat{E}(Y_r|x_i)\right]^{-1}_{r=1,\dots,p}\right) = A_{2,\mathbf{y}} \end{aligned}$$

and

$$\begin{aligned} VA_{3,\tilde{\mathbf{y}}} \quad &= \quad V\operatorname{diag}\left(\left[\frac{1}{m}\sum_{i=1}^{m}\widehat{Var}(\tilde{Y}_r|x_i)\right]^{-1/2}_{r=1,\dots,p}\right) \\ &= \quad \operatorname{diag}\left(\left[\frac{v_r^2}{v_r^2 m}\sum_{i=1}^{m}\widehat{Var}(Y_r|x_i)\right]^{-1/2}_{r=1,\dots,p}\right) = A_{3,\mathbf{y}}\,. \end{aligned}$$

Hence, the choice of either standardization matrix $A_{2,\mathbf{y}}$ or $A_{3,\mathbf{y}}$, will result in the same estimated loss function for $\mathbf{y}$ or $V\mathbf{y}$.

If we want the estimated risk function to be scale and translation invariant, only standardization matrix $A_{3,\mathbf{y}}$ meets the requirement. This can easily be verified by substituting $\tilde{\mathbf{y}} = Vy + b$, $b \in \mathbb{R}^p$, in the previous argument.

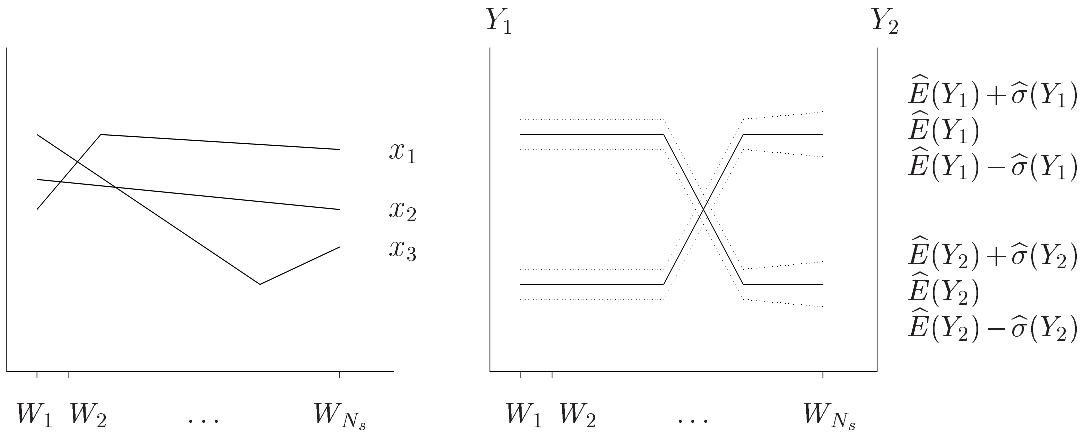
4 Joint optimization plot (JOP)

After choosing an experimental design, estimating mean and covariance models for the p response variables, and selecting a standardization matrix for the partitioning of C as described in the last section, we can now minimize the resulting estimated risk for every desired weight matrix W. However, our objective is to provide "optimal" factor settings for different possible matrices W, such that the final choice of optimal factor combinations rests on knowledge about the effect of a weight matrix selection on the resulting predicted response means and variances. We will first introduce a graphical tool, called a joint optimization plot, for visualizing the optimal design settings and consequent response predictions associated with a given sequence of weight matrices. Subsequently, the issue of defining such weight matrices will be addressed.

We assume, for now, that a sequence of weight matrices is available, say $\{W_t\}_{t=1}^{N_s}$. For each weight matrix W_t, the minimization of the corresponding estimated risk (3) is then restricted to a k-dimensional sphere covering the experimental region, where k denotes the number of design factors (the dimension of an x-vector). Every point within this sphere represents a possible design factor combination. A determination of optimal factor settings uses k-dimensional polar coordinates, so that the optimization routine can in fact be performed on a k-dimensional cube, which is supported by common statistical software packages. It might be necessary though to use different starting points for this minimization procedure to ensure a global optima is obtained.

Performing the optimization procedure for each considered weight matrix $W_t, t = 1, \ldots, N_s$, results in N_s optimal factor settings and N_s associated predicted means and variances. The computed results can be visualized by separately plotting the optimal factor settings and associated predicted response means ($\pm$ one standard deviation) by the corresponding sequence of weight matrices $\{W_t\}_{t=1}^{N_s}$. Figure 2 illustrates the joint optimization plot in the special case of two response variables and three design factors. The weight matrix from $\{W_t\}_{t=1}^{N_s}$ corresponding to the "best compromise" can be read off the right figure of the JOP by the engineer and transmitted to the corresponding optimal factor settings displayed in the left figure. Thereby, we assure that the joint optimization is really based on process requirements.

Figure 2: Joint optimization plot for two response variables and three design factors



For higher-dimensional response vectors it can be advisable to produce several JOPs, for instance by dividing the right figure into two or more plots on top of each other. The responses could for example be arranged by their measurement scale to avoid too many vertical axes in one plot.

However, we still have to state the choice of a reasonable sequence of weight matrices. For simplicity of exposition, we will focus on prescribing sequences of diagonal weight matrices. However, if non-diagonal weight (and cost) matrices are applied, we shall assume that off-diagonal elements of a weight matrix can be derived from the diagonal entries by a relation like $w_{rs} = \widetilde{w}_{rs}\sqrt{w_{rr}w_{ss}}$, for some desired "scaled" setting of off-diagonal elements $\widetilde{w}_{rs} \in (-1, 1)$, specified beforehand. However, as mentioned previously (see Section 3.1), off-diagonal elements of the weight matrix W should be used with care.

Due to the above assumption, we only need to deal with the diagonal of W which is denoted by $\mathbf{w} = (w_{11}, \dots, w_{pp})'$. Furthermore, note that the scale of the assigned weights is inconsequential because the use of $a\mathbf{w}$, $a \in \mathbb{R}\backslash\{0\}$ or $\mathbf{w}$ will lead to identical optimization results. Only the relative sizes of the weights are important.

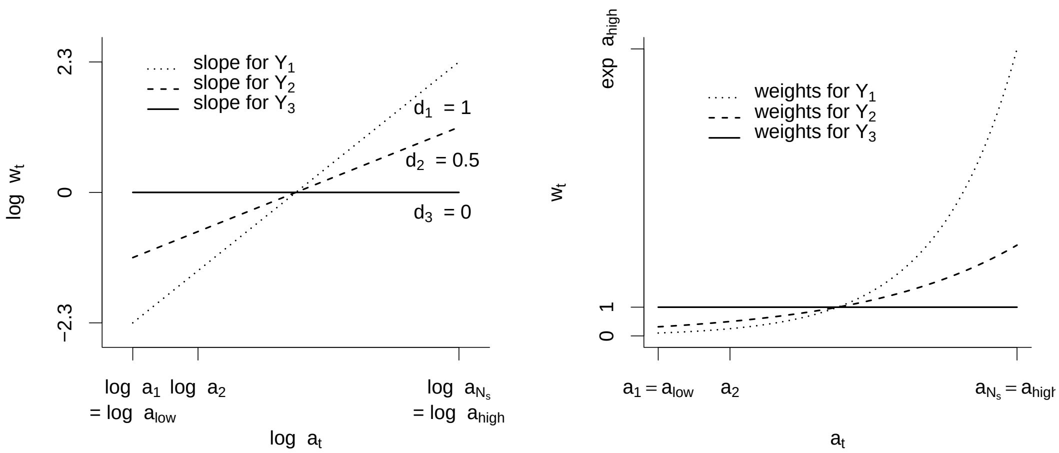
In general, defining a sequence of vectors $\mathbf{w}$ near $\mathbf{1}_p$ seems to be a good starting point since the entries of $\mathbf{w}$ are relative weights due to the standardization discussed in section 3.2. Therefore, we define the sequence $\{\mathbf{w}_t\}_{t=1}^{N_s} \in \mathbb{R}^p$ by

$$\log \mathbf{w}_t = \mathbf{d} \log a_t \tag{8}$$

with a "slope" vector $\mathbf{d} \in \mathbb{R}^p$ and $\{\log a_t\}_{t=1}^{N_s} \in \mathbb{R}$ an increasing equidistant sequence within the interval $[\log a_{low}, \log a_{high}]$ where $a_{low}, a_{high} > 0$, usually $a_{low} = a_{high}^{-1}$. In $\mathbf{d}$, weight relations of interest can be specified (i.e., proportionality sizes and directions). The results in the JOP can be directly related to $\{\log a_t\}_{t=1}^{N_s}$. Optimal factor settings achieved for $\mathbf{w}_t$ are denoted by $\mathbf{x}_t = (x_{t1}, \dots, x_{tk})'$, $t = 1, \dots, N_s$.

Let us consider some examples in order to illustrate the definition in (8) and its consequences for different numbers of responses. For the special case of two response values, there is only one ordering of interest, namely low- or high-weighting one of the responses. For instance, the sequence resulting from $\mathbf{d} = (1,0)'$ and $a_{low} = 1/10$, $a_{high} = 10$ yields weight matrix diagonals from $\mathbf{w}_1 = (10,1)'$ to $\mathbf{w}_{N_s} = (.1,1)'$. If three responses are considered, conceivable choices are, for example the case of only one varying weight $\mathbf{d} = (1,0,0)'$ as well as choices with two varying weights of different "slopes" like $\mathbf{d} = (1,.5,0)'$ or even something like $\mathbf{d} = (-1,1,0)'$ (e.g. anti-proportional weighting). With a_t defined as above, the three slopes result in weights $\mathbf{w}_1 = (10,1,1)'$, $\mathbf{w}_1 = (10,\sqrt{10},1)'$ or $\mathbf{w}_1 = (.1,10,1)'$ to $\mathbf{w}_{N_s} = (.1,1,1)'$, $\mathbf{w}_{N_s} = (.1,\sqrt{.1},1)'$ or $\mathbf{w}_{N_s} = (10,.1,1)'$, respectively. Figure 3 visualizes the effect of slope vector $\mathbf{d} = (1,.5,0)'$ on associated weights.

Figure 3: Sequence of weights in log scale and natural scale for "slope" vector $\mathbf{d} = (d_1,d_2,d_3)' = (1,.5,0)'$



For more than two responses, a series of JOPs with weight matrix diagonals defined iteratively based on the experience of weights used in a previous

JOP can be created. Using fitted mean and variance models, it seems also reasonable to group responses which are not contradictory by using the same slope in $\mathbf{d}$ for each such group. In addition, if there is reason to use a "basis weight" different from 1 for some of the responses, $\mathbf{w}_t$ may be defined by $\log \mathbf{w}_t = \log \mathbf{e} + \mathbf{d} \log a_t$ with desired vector of basis weights $\mathbf{e} \in \mathbb{R}^p$.

With the defined scheme for weight matrices, we possess the means to handle "compromise" optimization cases for which the joint optimization noticeably depends on all responses as well as cases where single responses or groups of responses are more important. The benefit to be gained from the suggested procedure will be elucidated in the following numerical example.

5 Example

The aim of the considered experiment is the optimization of a process of high-pressure sheet metal forming (see Kleiner et al., 1999 and Skotarzik, 2000). A central composite design in two design variables (`K`: blank holder force and `D`: pressure of working media) is used, which has been combined with the factor level combinations of the third design variable (`A`: initial blank thickness) and the one noise factor (`R`: friction). Both of the latter factors assume only two levels: low and high. For this example we have two response variables: `Area`, the area between the workpiece and the desired workpiece contour along a cutting; and RBT, the relative blank thinning. Target values are 0 for Area and 0.05 for RBT. The response values are derived from numerical simulations of the forming process by means of Finite Element Analysis. The data are given in Table 1.

Table 1: Data of the FEA-simulation experiment

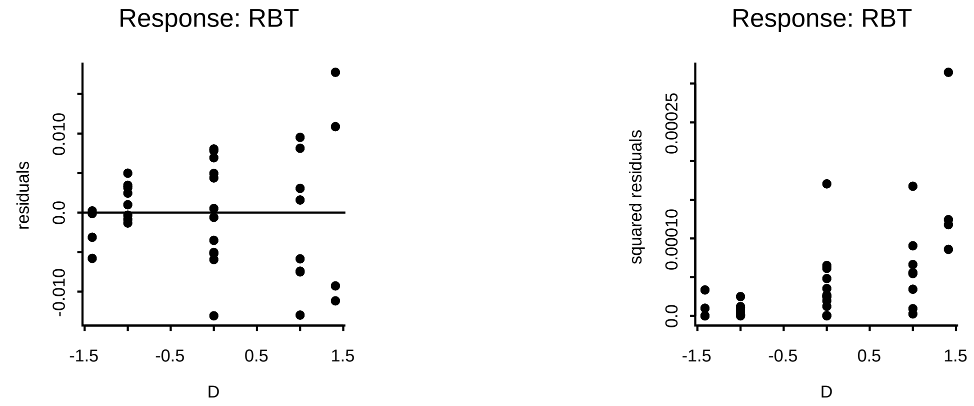
	K	D	A	Area		RBT	
				R= -1	R= 1	R= -1	R= 1
1	1	1	-1	7.9966	20.3601	0.067	0.083
2	1	1	1	19.1287	31.0451	0.067	0.081
3	1	-1	-1	38.6362	38.6033	0.061	0.057
4	1	-1	1	36.7798	44.5056	0.039	0.038
5	-1	1	-1	0.9748	9.1252	0.061	0.078
6	-1	1	1	8.5072	21.3566	0.054	0.063
7	-1	-1	-1	33.6568	37.6792	0.063	0.064
8	-1	-1	1	35.1122	43.9494	0.047	0.043
9	-1.41	0	-1	7.2792	21.7991	0.054	0.074
10	-1.41	0	1	19.1302	36.3412	0.050	0.063
11	1.41	0	-1	23.8342	31.6159	0.066	0.080
12	1.41	0	1	35.0138	36.5504	0.065	0.055
13	0	1.41	-1	1.3413	10.7754	0.065	0.092
14	0	1.41	1	9.0730	21.8355	0.058	0.080
15	0	-1.41	-1	35.0273	45.1586	0.052	0.046
16	0	-1.41	1	46.9272	42.3516	0.030	0.033
17	0	0	-1	15.4458	30.3393	0.066	0.070
18	0	0	1	29.7673	35.1129	0.057	0.062

For the dual response approach, models for the mean and variance of both responses are fitted separately by using the additive mean model and the multiplicative variance model suggested in (4) and (5). Models for the means are fitted first. For the response variable Area, only the three main effects are significant ($\alpha = 0.05$). For RBT, the main effect for K, though statistically insignificant, has not been removed from the model since the interaction term KD seems to be important. Initial mean models are

$$\begin{aligned}
\widehat{E}(\texttt{Area}) &= 26.7 + 3.34\,\texttt{K} - 11.6\,\texttt{D} + 3.97\,\texttt{A} \\
\widehat{E}(\texttt{RBT}) &= 0.064 + 0.0017\,\texttt{K} + 0.01\,\texttt{D} - 0.006\,\texttt{A} - 0.003\,\texttt{D}^2 \\
&\quad + 0.004\,\texttt{KD} + 0.0024\,\texttt{DA}\,.
\end{aligned}$$

Since we only have two replications which result from the two noise factor settings, it seems to be more reasonable to fit a dispersion model for the squared residuals, rather than for the standard deviation of two runs. This results in constant variances for Area and a model only in D for the variance of RBT. The variance model for RBT is justified by Figure 4.

Figure 4: Plot of residuals / squared residuals of the mean model for RBT versus values of D



Using iteratively reweighted least squares (IRLS) methods for fitting the RBT model, we finally get

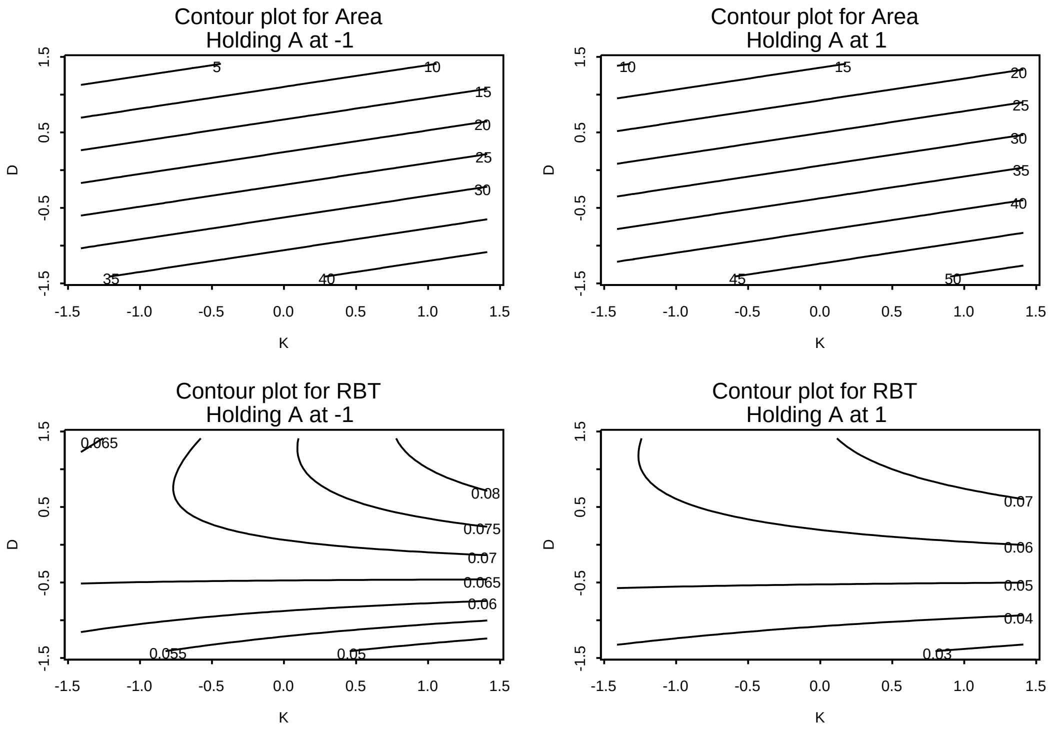
$$\begin{aligned}
\widehat{E}(\mathtt{Area}) &= 26.7 + 3.34\,\mathtt{K} - 11.6\,\mathtt{D} + 3.97\,\mathtt{A} \\
\widehat{Var}(\mathtt{Area}) &= 34.94 \\
\widehat{E}(\mathtt{RBT}) &= 0.065 + 0.0019\,\mathtt{K} + 0.01\,\mathtt{D} - 0.006\,\mathtt{A} - 0.005\,\mathtt{D}^2 \\
&\quad + 0.0045\,\mathtt{KD} + 0.0027\,\mathtt{DA} \\
\widehat{Var}(\mathtt{RBT}) &= \exp\{-10.4 + 1.15\,\mathtt{D}\}\,.
\end{aligned}$$

Contour plots showing the effect of the design variables on the means are given in Figure 5. The response variable Area will be optimized by choosing both K and A small and D large, but the target is not reached within the experimental region. On the other hand, the target value of 0.05 for RBT is attained for different factor combinations within the experimental region. However, obviously it is not possible to reach both targets simultaneously.

For the joint optimization of both responses, there is in this case no clearly justifiable reason for a cost matrix with off-diagonal elements different from zero. Therefore, no model for the covariance of both responses is needed for the simultaneous optimization.

We restrict the search for optimal design factor combinations to a sphere around the center of the experimental region in order to get comparable results, i.e. to avoid ending up in "corners" of the experimental region. To cover

Figure 5: Contour-Plots for estimated mean surfaces

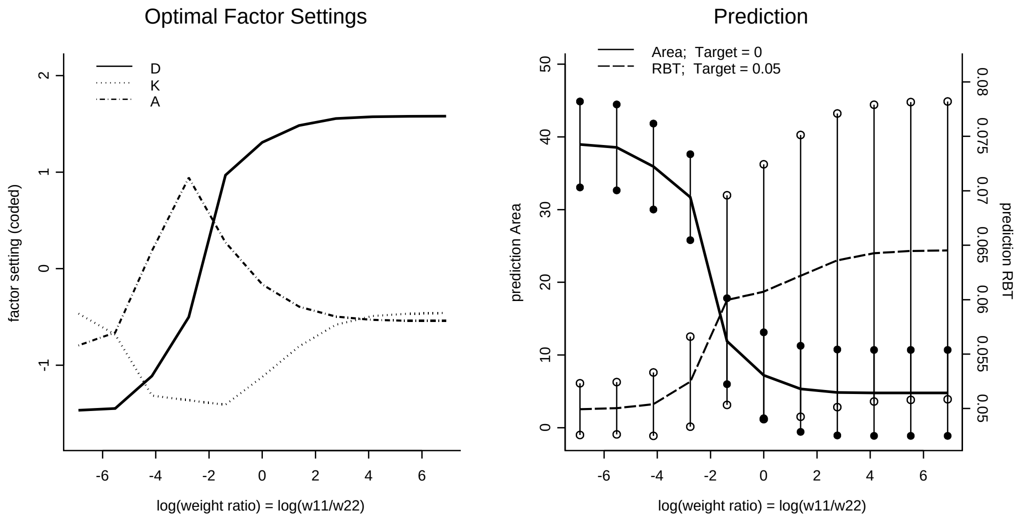


all experimental points, the radius has been chosen equal to $\max ||x_i|| = \sqrt{3}$, where $x_i = (\mathtt{K}_i, \mathtt{D}_i, \mathtt{A}_i)'$, $i = 1, \ldots, m$. For suggesting confirmatory experiments, it is also reasonable to use various increasing radii.

For carrying out the optimization procedure, the first suggested standardization matrix A_1 cannot be applied to this data set since one of the target values equals zero. The results when using standardization matrices A_2 or A_3 are identical but shifted, if $\mathbf{d} = (1, 0)'$ is applied. To see this, consider the use of arbitrary weights w_{11}, w_{22} with A_2 and w_{11}^*, w_{22}^* with A_3. This results in $c_{rr} = w_{rr} \cdot [A_2]^2_{rr}$, $r = 1, 2$ and $c^*_{rr} = w^*_{rr} \cdot [A_3]^2_{rr}$, $r = 1, 2$. Since optimization results only depend on the ratio c_{11}/c_{22} for 2×2 diagonal cost matrices, we get the same result for both matrices, as long as $w_{11}/w_{22} = w^*_{11}/w^*_{22} \cdot ([A_3]^2_{11}[A_2]^2_{22} \,/\, [A_3]^2_{22}[A_2]^2_{11})$ holds. Using (8) and $\mathbf{d} = (1, 0)'$ implies $w_{22} = w^*_{22} = 1$ and therefore $\log(a_t) = \log(w_{11}) = \log(w^*_{11}) + \log([A_3]^2_{11}[A_2]^2_{22} \,/\, [A_3]^2_{22}[A_2]^2_{11}) = \log(a^*_t) + \text{shift}$. Based on these findings, we will only use standardization matrix A_3 due to its invariance properties derived in Section 3.2.

Figure 6 displays the joint optimization plot for $a_{low} = 1/1\,000$, $a_{high} = 1\,000$ resulting in $\log a_{low} \approx -6.91$ and $a_{high} \approx 6.91$ and $N_s = 11$.

Figure 6: JOP for A_3, standardization by the standard deviation, $r = \sqrt{3}$ and "slope" vector $\mathbf{d} = (1, 0)'$



Since $\log a_t = \log(w_{11}/w_{22})$, the logarithm of the weight ratio is displayed on the horizontal axis. If $\log(w_{11}/w_{22})$ equals zero, this corresponds to the use of $W = I$, i. e. the identity matrix. The left hand side summarizes the optimal factor settings for the three design variables while the right hand side shows predicted response means (plus / minus one predicted standard deviation) for the corresponding factor settings. In both parts of Figure 6, negative values of $\log a_t$ imply a greater weighting of $Y_2 = \texttt{RBT}$, while positive values correspond to weight ratios greater than one and thus a higher weighting of the variable $Y_1 = \texttt{Area}$.

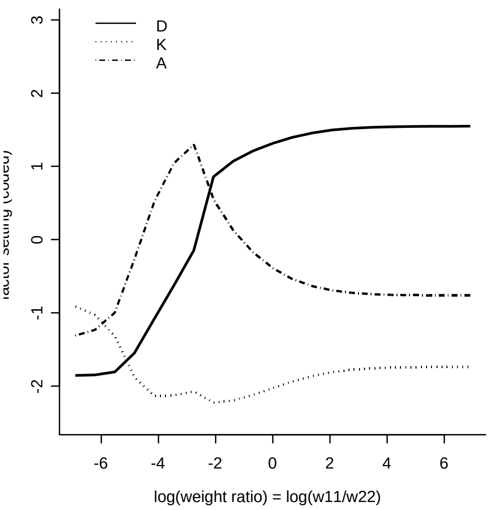
From the JOP, it is now possible to decide which available compromise or design setting is most desirable and should therefore be declared as being the point of joint optimisation within the experimental region. Of course, this decision has to be based on knowledge of the process and the purpose of the manufactured pieces. In this example a reasonable choice might be the compromise attained by using $\log(w_{11}/w_{22})_t = 0$, if for example the response `Area` should not exceed the value 10 and therefore the larger expectation and standard deviation for `RBT` have to be tolerated. Whereas, if the response `RBT` has to be lower than, say, 0.055, irrespectively of the `Area`, factor levels associated with $\log(w_{11}/w_{22})_t = -2.7$, might be considered. The optimal factor settings and predictions for both choices are given in Table 2.

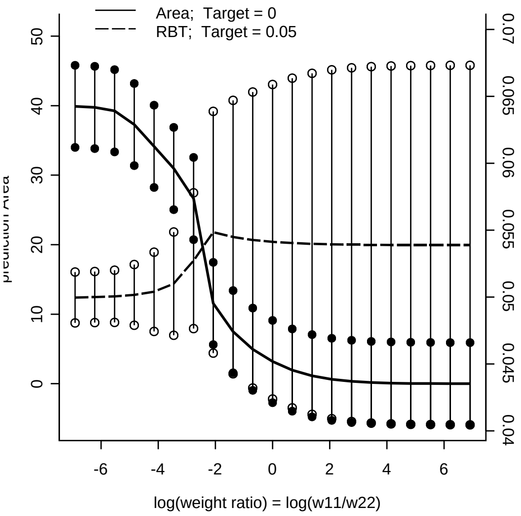
Table 2: Summary of optimal factor settings $\mathbf{x}_t$ and predictions for mean and variance (radius of sphere $= \sqrt{3}$)

$\log\left(\frac{w_{11}}{w_{22}}\right)$	$\mathbf{x}_t' = (\mathtt{D}_t,\ \mathtt{K}_t,\ \mathtt{A}_t)$	$\widehat{E}(Y_1\|\mathbf{x}_t)$	$\widehat{E}(Y_2\|\mathbf{x}_t)$	$\widehat{Var}(Y_1\|\mathbf{x}_t)$	$\widehat{Var}(Y_2\|\mathbf{x}_t)$
0.00	1.309, -1.123, -0.164	7.201	0.0607	34.94	0.000137
-2.76	-0.501, -1.364, 0.943	31.705	0.0525	34.94	0.000017

We can further see from the previous figures that all optimal factor settings are reached on the boundaries of the considered spheres, i. e. $\sum_{j=1}^{k} x_{jt}^2 = \mathtt{D}_t^2 + \mathtt{K}_t^2 + \mathtt{A}_t^2 = 3, \quad \forall\ t = 1, \ldots, N_s$. Therefore it might be possible to reach further optimization by increasing the radius of the sphere, if the fitted models also hold outside the experimental region. Results for the radius $r = \sqrt{6}$ are given in Figures 7 ($N_s = 21$).

Figure 7: JOP for A_3, standardisation by the standard deviation, $\mathbf{d} = (1, 0)'$, $r = \sqrt{6}$ (prediction outside experimental region)





From the right tail of Figure 7, the predicted mean of `RBT` is below 0.055, while the response `Area` approaches zero. Using the same matrices W, a further optimization seems possible. Results are given in Table 3.

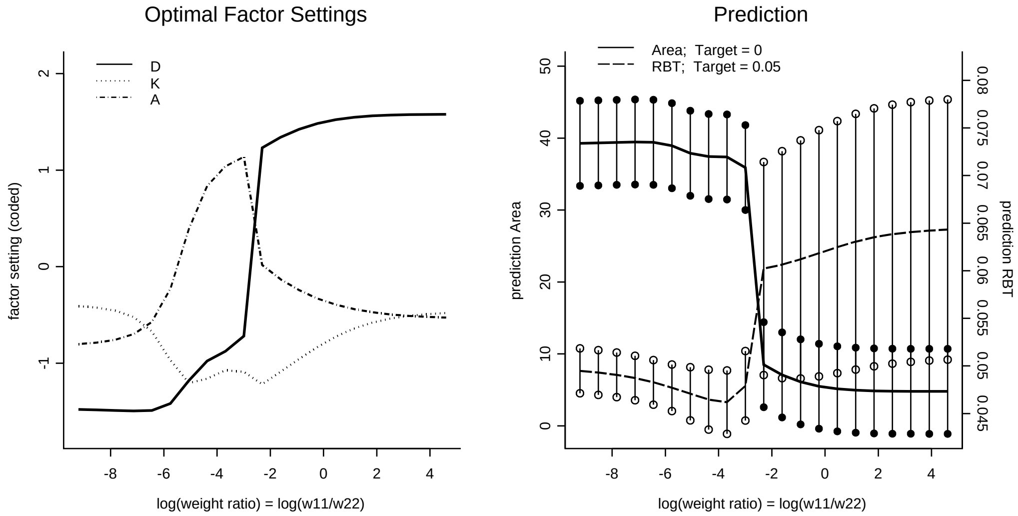
Table 3: Summary of optimal factor settings $\widetilde{\mathbf{x}}$ and predictions (outside experimental region) for mean and variance (radius of sphere $= \sqrt{6}$)

$\log\left(\frac{w_{11}}{w_{22}}\right)$	$\widetilde{\mathbf{x}}' = (\mathtt{D}, \mathtt{K}, \mathtt{A})$	$\widehat{E}(Y_1\|\widetilde{\mathbf{x}})$	$\widehat{E}(Y_2\|\widetilde{\mathbf{x}})$	$\widehat{Var}(Y_1\|\widetilde{\mathbf{x}})$	$\widehat{Var}(Y_2\|\widetilde{\mathbf{x}})$
0.00	1.316, -2.029, -0.391	3.195	0.0541	34.94	0.000138
-2.76	-0.147, -2.073, 1.297	26.642	0.0527	34.94	0.000026

However, confirmatory runs are needed to support the used univariate models as well as the suggested optimal factor settings.

In order to show possible effects of off-diagonal elements in W, Figure 8 displays the joint optimization plot, again for the standardization matrix A_3, and weight matrix W with $w_{12} = w_{21} = 0.5\sqrt{w_{11}w_{22}}$. Since we only have two replications for each design point, estimation and modeling of correlations will not be appropriate, since the correlation will always be either $+1$ or -1. Therefore, $\Sigma(\mathbf{x})$ has been assumed to be diagonal. The use of $\widetilde{w}_{12} = 0.5$ ensures that deviations from target of `RBT` and `Area` in the same direction are penalized. This is the reason for the unexpected behaviour of the optimization plot, namely the choice of design factor combinations that lead to expected values for `RBT` below the target $\tau_2 = 0.05$. The use of diagonal cost matrices as above seems to be advisable.

Figure 8: JOP for A_3, $r = \sqrt{3}$, $\mathbf{d} = (1,0)'$ and $\widetilde{w}_{12} = 0.5$



Overall it can be concluded that joint optimization plots provide valuable information about response values, achieved by optimal choices of design factor settings. It can be expected that equally good results can be achieved for problems other than the described hydroforming process.

6 Summary

In this paper we proposed a data-driven method to choose cost matrices for the joint optimization of multiple responses in statistically designed experiments. We further suggested suitable graphical tools to allow comparison of different factor choices and their implications for predictions of the means and variances of the considered responses. Our method allows for compromise optimization without requiring cost matrices that are difficult to determine and heavily influence the result of the analysis. On the contrary, our method allows for great flexibility.

Acknowledgement

Financial support of the Deutsche Forschungsgemeinschaft (Collaborative Research Centre 475, Research Group 366) is gratefully acknowledged.

References

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

Chen, L.-H. (1997). *Designing robust products with multiple quality characteristics* Computers & Operations Research 10, pp. 937-944

Engel, J. and Huele, A. F. (1996). *A Generalized Linear Modeling Approach to Robust Design.* Technometrics, Vol. **38**, pp. 365-373.

Grize Y.L. (1995). *Review: A Review of Robust Process Design Approaches.*

Journal of Chemometrics, Vol. **9**, pp. 239-262.

Kackar R.N. (1985). *Off-Line Quality Control, Parameter Design, and the Taguchi-Method.* Journal of Quality Technology, Vol. **17**, pp. 176-188

Kleiner M., Hellinger V., Homberg W., Klimmek C. (1999). *Trends in Sheet Metal Hydroforming.* Konferenz-Einzelbericht: Hydroumformung von Rohren, Strangpressprofilen und Blechen. Internat. Konf. Hydroumformung, 12-13. Okt. 1999 Stuttgart, Bd. **1**, pp. 291-303.

Leon R.V., Shoemaker A.C., Kacker R.N. (1987). *Performance Measures Independent of Adjustment.* Technometrics, Vol. **29**, No. 3, pp. 253-265.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* Chapman & Hall.

Nelder, J. A. and Lee, Y. (1991). *Generalized Linear Models for the Analysis of Taguchi-type experiments.* Applied Stochastic Models and Data Analysis, Vol. **7**, pp. 107-120.

Pignatiello J.J. (1993). *Strategies for Robust Multiresponse Quality Engineering.* IIE Transactions, Vol. **25**, pp. 5-15.

Skotarzik M. (2000). *Statistische Methoden der multivariaten Qualitätsplanung.* Diplomarbeit, Fachbereich Statistik, Universität Dortmund.

Taguchi G., Phadke M.S. (1984). *Quality Engineering Through Design Optimization.* Proceedings of GLOBECOM 84 Meeting, Piscataway. NJ: IEEE Communication Society, pp. 1106-1113.

Tsui K.L. (1999). *Robust Design Optimization for Multiple Characteristic Problems.* International Journal of Production Research, Vol. **37**, pp. 433-445.

Tong L.-I., Su, Ch.-T. (1997). *Optimizing Multi-Response Problems in the Taguchi Method by Fuzzy Multiple Attribute Decision Making.* Quality and Reliability Engineering International, Vol. **13**, pp. 25-34.

Vining G.G. (1998). *A Compromise Approach to Multiresponse Optimization.* Journal of Quality Technology, Vol. **30**, pp. 309-313.