SFB 823

Joint optimization of multiple responses based on loss functions

Discussion Paper

Martina Erdbrügge, Sonja Kuhnt, Nikolaus Rudak

Nr. 10/2011

SFB 823

Joint Optimization of Multiple Responses based on Loss Functions

Martina Erdbrügge, Sonja Kuhnt, Nikolaus Rudak

TU Dortmund University, 44221 Dortmund, Germany

Abstract

Most of the existing methods for the analysis and optimization of multiple responses require some kind of weighting of these responses, for instance in terms of cost or desirability. Particularly at the design stage, such information is hardly available or will rather be subjective. Kuhnt and Erdbrügge (2004) present an alternative strategy using loss functions and a penalty matrix which can be decomposed into a standardizing (data-driven) and a weight matrix. The effect of different weight matrices is displayed in joint optimization plots in terms of predicted means and variances of the response variables. In this article, we propose how to choose weight matrices for two and more responses. Furthermore we prove the Pareto optimality of every point that minimizes the conditional mean of the loss function.

1 Introduction

For technical applications, off-line quality control prior to the actual manufacturing often implies optimizing the mean as well as minimizing the variance of multiple responses. As separate analysis of each response can yield valuable information about the process, but often results in conflicting recommendations regarding the optimal parameter setting, robust design methods for multiple responses (Murphy *et al.* (2005)) are needed. Current approaches based on response surface methodology (Khuri and Mukhopadhyay (2010)) are mainly either extensions of the desirability functions approach

(Wu (2009), He *et al.* (2010), Derringer and Suich (1980)) or of the squared error loss approach (Shen *et al.* (2010)).

Pignatiello (1993) and Vining (1998) are beyond the first to extend the loss function approach to multiple responses, which involves the use of a pre-specified cost matrix. Experience, however, shows that this might be a hard task at the design stage, since the engineer would rather be able to assign a relative importance to each response than specify actual costs incurred for specific response values. Then by considering only one cost matrix it can only be assumed that a reasonable compromise is reached, i. e. design factor levels are found that achieve relatively good results for all responses. An alternative is provided by the approach suggested in Kuhnt and Erdbrügge (2004), where the estimated expected loss is minimized for a sequences of possible cost matrices. For each matrix, an optimal design factor combination is derived and the results are graphically displayed in so-called joint optimization plots. By this proceeding, additional insight into the process under study is gained and, even more important, the subjective part of the analysis, namely the choice of an appropriate matrix, is postponed to a stage where the consequences are visible in terms of the predicted responses. Whereas the original paper provides the general strategy behind joint optimization plots, we focus on the choice of the employed sequence of cost matrices and show that the optimal factor settings derived are Pareto optimal. The cost matrix itself is factorized in a standardization and a weight matrix. Different choices of both kind of matrices are discussed and illustrated with examples from mechanical engineering applications.

The paper is organized as follows. In Section 2 we give a short introduction to loss functions for the multi-response case. In Section 3 we show that every point that minimizes the conditional mean of the loss function, is Pareto optimal. Data-driven choices of the standardizing matrix component are proposed in Section 4 that are invariant to scale transformations. For a joint optimization plot which simultaneously considers a range of cost matrices special sequences of weight matrices are discussed, especially in cases of more than two response variables. The use of the resulting strategy is demonstrated in Section 5 for applications on thermal spraying processes and springback compensation. We conclude with a discussion in Section 6.

2 Multivariate loss and risk functions

Let us consider a product or production process, 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 $\mathbf{x} = (x_1, \ldots, x_k)' \in \mathcal{X}$. Note, that noise parameters are omitted here for a clearer presentation, but may well be treated (see Kuhnt, Erdbrügge (2004)). The conditional distribution of $\mathbf{Y}$ given $\mathbf{x}$ 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 we allow both, the response mean and covariance matrix to depend on the design parameters. Commonly it is assumed that $\mathbf{Y}|\mathbf{x}$, or a suitable transformation thereof, follows a multivariate normal distribution. The framework of double generalized linear models (Smyth (1989)) allows for extensions to other continuous distribution from the exponential family, such as the lognormal or exponential distribution.

In case of a single response ($p = 1$), the overall quality of the product is viewed in terms of loss resulting from the deviation of a quality characteristic Y from its target τ. This loss is measured by the quadratic loss function: $loss(Y) = c(Y - \tau)^2$, where c is some constant. As a straightforward extension to multiple quality characteristics, $loss(\mathbf{Y}) = \left(\mathbf{Y} - \tau\right)' C \left(\mathbf{Y} - \tau\right)$ has been proposed by Pignatiello (1993). Here C denotes a $p \times p$ dimensional symmetric cost or penalty matrix. For the expected loss, called risk function, it holds that

$$\begin{aligned} R(\mathbf{x}) &= E\big(loss(\mathbf{Y}|\mathbf{x})\big) = E(\mathbf{Y}-\tau)'C(\mathbf{Y}-\tau) \\ &= \operatorname{trace}\big(C\Sigma(\mathbf{x})\big) + \big(\mu(\mathbf{x})-\tau\big)'C\big(\mu(\mathbf{x})-\tau\big)\,. \end{aligned} \tag{1}$$

Hence, minimizing the risk function will, roughly speaking, simultaneously yield a mean near target and minimal variances. Different strategies have been proposed (see e. g. Pignatiello (1993)) to find combinations of the design parameter values which minimize the resulting estimated risk function

$$\widehat{R}(\mathbf{x}) = \operatorname{trace}\left(C\widehat{\Sigma}(\mathbf{x})\right) + \big(\widehat{\mu}(\mathbf{x})-\tau\big)'C\big(\widehat{\mu}(\mathbf{x})-\tau\big). \tag{2}$$

Estimates for $\widehat{\mu}(\mathbf{x})$ and $\widehat{\Sigma}(\mathbf{x})$ may be achieved by modelling the means, variances and covariances of the responses. These models may for instance be fitted according to the methods given by Grize (1995), Engel and Huele (1996), Chiao and Hamada (2001), McCullagh and Nelder (1989) or Nelder and Lee (1991, 2003).

Since the number of replicates is rather small in our examples, we fit models to the mean and variance, separately for each response, that is we assume independent responses, given the parameter setting. In our examples we fit double generalized linear models with identity link and normal probability assumption for the mean model, hence

$$\widehat{E}(Y_r|\mathbf{x}) = f(x). \tag{3}$$

The variance model is derived by a gamma GLM with log link resulting in

$$\widehat{Var}(Y_r|\mathbf{x}) = \exp\{g(x)\}\,. \tag{4}$$

Here f and g are polynomial functions. In this case of normal response and identity link for the mean, the variance model is based on the squared residuals of the mean model as responses. The iterative fitting procedure for both models (3) and (4) alternates between fitting the mean and variance model, at each step using the actual estimates.

Minimizing the resulting estimated risk (2) with respect to $\mathbf{x}$, leads to Pareto optimal parameter settings in case of a diagonal cost matrix C, as is shown next.

3 Calculation of Pareto optimal points

The concept of Pareto optimality allows to compare two different points in $\mathbb{R}^n$ and to decide whether or not a point is optimal in a specific sense. It is based on the following order relation in $\mathbb{R}^n$, see Hillermeier (2001).

Definition 3.1 (Order relation $\leq$ in $\mathbb{R}^n$) *Let $\leq$ denote an order relation in $\mathbb{R}^n$, i.e. a special subset of the set $\mathbb{R}^n \times \mathbb{R}^n$ of all ordered points of elements in $\mathbb{R}^n$. Instead of $(y^1, y^2) \in \leq$ one customarily uses the infix notation $y^1 \leq y^2$. Let the order relation be defined as follows:*

$$y^1 \leq y^2 \qquad \Leftrightarrow \qquad y^2 - y^1 \in \mathbb{R}^n_+,$$

where $\mathbb{R}^n$ denotes the non-negative orthant of $\mathbb{R}^n$.

This order relation does not allow us to compare all vectors in $\mathbb{R}^n$. Thus it is only a partial order. Based on the above order relation an efficient and Pareto optimal point can be defined as follows.

Definition 3.2 (Efficient point, Pareto optimal point) *Let $f : \mathcal{R} \subset \mathbb{R}^n \rightarrow f(\mathcal{R}) \subset \mathbb{R}^k$ be a vector-valued function. A point $y^* \in f(\mathcal{R})$ is called efficient with regard to the order relation $\leq$ defined in $\mathbb{R}^n$, if and only if there exists no other $y \in f(\mathcal{R})$, $y \neq y^*$, with $y \leq y^*$. A point $x^* \in \mathbb{R}^n$ with $y^* = f(x^*)$ is called **Pareto optimal**, if and only if y^* is efficient.*

A point $x^* \in \mathbb{R}^n$ with $y^* = f(x^*) \in \mathbb{R}^k$ is called **Pareto optimal** if there exists no other point $x \neq x^* \in \mathbb{R}^n$ with $y^* \neq y = f(x) \in \mathbb{R}^k$ such that

$$y^*{}_i \leq y_i \quad \forall i \in \{1, \ldots, k\}.$$

Thus if we assume that a point $x^1 \in \mathbb{R}^n$ with $y^1 = f(x^1) \in \mathbb{R}^k$ is not Pareto optimal, then there exists a point $x^2 \in \mathbb{R}^n$ with $y^2 = f(x^2) \in \mathbb{R}^k$ such that

$$y^1 \neq y^2 \text{ and } y_i^2 \leq y_i^1$$

and at least for one index $j_0 \in \{1, \ldots, k\}$ it holds that $y_{j_0}^2 < y_{j_0}^1$.

Let us connect the concept of Pareto optimality with the optimization of the risk function introduced before. The considered unconstrained optimization problem writes as follows

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & R(x) \\ \text{s.t.} \quad & x \in \mathcal{R} \end{aligned} \tag{R}$$

where $\mathcal{R} = \{x \in \mathbb{R}^n | \; \underline{x} \leq x \leq \overline{x}\}$ is the feasible region with box constraints. The models for mean and dispersion are commonly estimated on the basis of a designed experiment. The estimated models are reasonable in the operation region prescribed by the operation region of the experiment. Thus we consider a box constrained optimization problem. Furthermore

$$\begin{aligned} R(\mathbf{x}) &= E\big(\text{loss}(\mathbf{Y}|\mathbf{x})\big) = E(\mathbf{Y} - \tau)' C (\mathbf{Y} - \tau) \\ &= \text{trace}\big(C\Sigma(\mathbf{x})\big) + \big(\mu(\mathbf{x}) - \tau\big)' C \big(\mu(\mathbf{x}) - \tau\big) \end{aligned}$$

is the risk function with a positive definite diagonal matrix $C = \text{diag}(c_1, \ldots, c_k) \in \mathbb{R}^{k \times k}$ with positive diagonal entries, a target vector $\tau \in \mathbb{R}^p$ stands for the ideal point to be reached and a diagonal covariance matrix $\Sigma(\mathbf{x}) = \text{diag}({\sigma_1}^2, \ldots, {\sigma_k}^2) \in \mathbb{R}^{k \times k}$. Because both C and $\Sigma(\mathbf{x})$ are diagonal, we can write the risk function as follows:

$$\begin{aligned} R(\mathbf{x}) &= \text{trace}\big(C\Sigma(\mathbf{x})\big) + \big(\mu(\mathbf{x}) - \tau\big)' C \big(\mu(\mathbf{x}) - \tau\big) \\ &= \sum_{i=1}^{p} c_i({\sigma_i}^2(\mathbf{x}) + \big(\mu_i(\mathbf{x}) - \tau_i\big)^2) \\ &:= \sum_{i=1}^{p} c_i f_i(x) \end{aligned}$$

The single valued functions f_i are positive because both the variances ${\sigma_i}^2(\mathbf{x})$ and the quadratic differences $(\mu_i(\mathbf{x}) - \tau_i)^2$ are positive for all $x \in \mathbb{R}^n$ and

$i \in \{1, \ldots, k\}$. Let $F : \mathbb{R}^n \to \mathbb{R}^k$ with $F(x) = (f_1(x), \ldots, f_k(x))$ be a vector valued function. We assume that there exists an optimal point $x^{\text{opt}} \in \mathcal{R} \subset \mathbb{R}^n$ of the optimization problem (R) under appropriate regularity conditions. This is reasonable because if we assume the risk function to be at least continous then there exists a point in the compact feasible region that minimizes the risk function. In the next proposition we show that an optimal solution x^* of problem (R) is Pareto optimal for

$$\begin{aligned} \min_x \quad & F(x) \\ \text{s.t.} \quad & x \in \mathcal{R} \end{aligned} \tag{$\mathcal{R}_{\text{multi}}$}$$

where we minimize a p-dimensional vector in terms of Definition 3.1. The solutions of the problem ($\mathcal{R}_{\text{multi}}$) are Pareto optimal points in terms of Definition 3.2.

Proposition 3.3 *If $x^{opt} \in \mathcal{R} \subset \mathbb{R}^n$ is optimal for the problem (R) then it is Pareto optimal with $y^{opt} := F(x^{opt}) \in \mathbb{R}^k$ for the problem ($\mathcal{R}_{multi}$).*

Proof: Let us assume that $x^{\text{opt}} \in \mathcal{R} \subset \mathbb{R}^n$ with $y^{\text{opt}} := F(x^{\text{opt}}) \in \mathbb{R}^k$ is not Pareto optimal. Then there exists a point $\hat{x} \in \mathbb{R}^n$ with $\hat{y} = F(\hat{x}) \in \mathbb{R}^k$ such that

$$0 \leq \hat{y}_i \leq {y^{\text{opt}}}_i \quad \forall \quad i = 1, \ldots, k$$

and at least for one index j it holds that

$$0 \leq \hat{y}_i < y_i^{\text{opt}}.$$

It follows that

$$\begin{aligned} R(x^{\text{opt}}) &= \sum_{i=1}^{p} c_i f_i(x^{\text{opt}}) = \sum_{\substack{i=1 \\ i \neq j}}^{p} c_i f_i(x^{\text{opt}}) + c_j f_j(x^{\text{opt}}) \\ &= \sum_{\substack{i=1 \\ i \neq j}}^{p} c_i y_i^{\text{opt}} + c_j y_j^{\text{opt}} \geq \sum_{\substack{i=1 \\ i \neq j}}^{p} c_i \hat{y}_i + c_j y_j^{\text{opt}} \end{aligned}$$

$$
\begin{aligned}
&> \sum_{\substack{i=1\\ i\neq j}}^{p} c_i \hat{y}_i + c_j \hat{y}_j = \sum_{\substack{i=1\\ i\neq j}}^{p} c_i f_i(x^{\text{opt}}) + c_j f_j(x^{\text{opt}}) \\
&= \sum_{i=1}^{p} c_i f_i(\hat{x}) = R(\hat{x}) \geq 0
\end{aligned}
$$

We get that $R(x^{\text{opt}}) > R(\hat{x})$ which contradicts the optimality of x^{opt}. Thus x^{opt} is Pareto optimal.

□

Proposition 3.3 guarantees that every optimal point of the optimization problem (R) is a Pareto optimal point $x_{\text{opt}} \in \mathcal{R} \subset \mathbb{R}^n$ with $y_{\text{opt}} := F(x_{\text{opt}}) \in \mathbb{R}^k$. Thus if we minimize the risk function then the optimal point x_{opt} reduces the variances ${\sigma_i}^2(\mathbf{x})$ and gets the mean $\mu_i(\mathbf{x})$ on target τ_i for all $i \in \{1, \ldots, n\}$ simultaneuosly. The risk function is a weighted sum of the functions f_i with strictly positive weights, hence it is a kind of weighting method as in Hillermeier (2001) and Ehrgott (2000). If the image $F(\mathcal{R})$ is nonconvex then there are Pareto optimal points that can not be calcaluted by minimizing the weighted single valued function, compare Hillermeier (2001) and Ehrgott (2000). But the weighting method has the advantage to calculate some Pareto optimal points by means of minimizing a single valued function which can be done by a numerical solver.

4 Joint optimization plots

As in Kuhnt and Erdbrügge (2004) we consider sequences of cost matrices to gain insight into a range of possible optimal parameter settings. For each element of the sequence the estimated risk (2) is minimized with respect to $\mathbf{x}$ and the result in terms of estimated mean responses and variances is displayed in a graph called joint optimisation plot (JOP).

First of all, we define this sequence according to $C_t = AW_tA$, where A is a diagonal standardization matrix and W_t is a sequence of weight matrices.

For the choice of the standardization matrix, we propose two alternatives

$$\begin{aligned} A_1 &= \operatorname{diag}\left(\left[\frac{1}{m}\sum_{i=1}^{m}\widehat{Var}(Y_r|\mathbf{x}^i)\right]^{-1/2}_{r=1,\dots,p}\right) \\ \text{and} \qquad A_2 &= \operatorname{diag}\left(\left[\frac{1}{m}\sum_{i=1}^{m}\widehat{E}(Y_r|\mathbf{x}^i)\right]^{-1}_{r=1,\dots,p}\right), \end{aligned}$$

where m is the number of different design points $\mathbf{x}^i = (x_1^i, \dots x_k^i)'$, $i = 1, \dots, m$ with n_i observations of response vectors of length p and $n = \sum_{i=1}^m n_i$ the total number of observations.

These matrices both ensure that the optimal design parameter values do not depend on the measurement scales used for the quality characteristics. The choice of A_1 in addition ensures translation invariance of the estimated risk and may therefore be preferred. However, A_2 may be useful in particular if most of the target values are zero and we would like to weight the responses according to their ability to "reach" their target values.

We will focus on assigning sequences of diagonal weight matrices W_t, as this is usually sufficient and assures Pareto optimality. However, if non-diagonal weight matrices are wanted, we suggest that off-diagonal elements of W_t are derived from the diagonal entries by a relation $w_{rs} = \widetilde{w}_{rs}\sqrt{w_{rr}w_{ss}}$, for some desired "scaled" setting of off-diagonal elements $\widetilde{w}_{rs} \in (-1, 1)$. We denote the diagonal of W_t by $\mathbf{w}_t = (w_{11}, \dots, w_{pp})'$ and define relative sizes of weights by

$$\log \mathbf{w}_t = \mathbf{d} \log a_t \tag{5}$$

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 $[\log a_{low}, \log a_{high} = -\log a_{low}]$. In $\mathbf{d}$, weight relations of interest can be specified (i.e., proportionality sizes and directions) to ensure that the weight matrices are linear independent. For the special case of two response values, a sequence that low- or high-weights one of the responses, for instance the sequence of weight matrix diagonals from $\mathbf{w}_1 = (10, 1)'$ to $\mathbf{w}_T = (.1, 1)'$ results from $\mathbf{d} = (1, 0)'$ and $a_{low} = 1/10$, $a_{high} = 10$. For higher-dimensional response, the choice of the slope vector $\mathbf{d}$ is not as straightforward and may for instance be based on correlations of predicted response means for grouping responses that are not contradictory.

After specifying the sequence $\{W_t\}_{t=1}^T$ of weight matrices, the estimated risk is minimized for each of the T weight matrices W_t. The resulting optimal factor settings and associated predicted means and variances of the responses are related to $\{\log a_t\}_{t=1}^T$. The computed results are visualized by separately plotting the optimal factor settings and associated predicted response means (± one standard deviation). This combination of graphs is called "joint optimization plot" (JOP). The weight matrix from $\{W_t\}_{t=1}^T$ corresponding to the "best compromise" regarding process requirements is read off the predictions of the JOP and then transmitted to the corresponding optimal factor settings. Hereby, the joint optimization is not influenced by any kind of subjectivity (i. e. in choosing a cost matrix) prior to the statistical analysis.

We now possess the means to handle compromise optimization in cases where the joint optimization should noticeably depend on all responses as well as cases where single responses or groups of responses are more important. The plots are done by the R-package JOP, version 2.0.0.

5 Examples

We first consider an example from sheet metal forming with two responses. Here the focus lies on the general idea of joint optimization plots and the Pareto optimality. The second example treats a thermal spraying process with three responses, where the impact of different choices of the slope vector becomes apparent.

5.1 Example: Springback Compensation

In the following a sheet metal forming process is considered. A drawback in forming sheet metals with high strength is the springback after the manufacturing process. Thus the sheet metal forming process requires strategies for springback compensation. The experimental set up is displayed in Figure 1. The aim is to minimize the springback on the basis of a replicated 3^2 design with a total run size of 42 simulations. The responses Y_1 in **mm**, the side wall curvature, and Y_2 in degrees, the flange angle, describe the springback behaviour as displayed in Figure 1. The controllable parameters are the blank holder force NH in **kN** and the punch stroke ZT in **mm**. Table 1 summarizes the coded parameters.

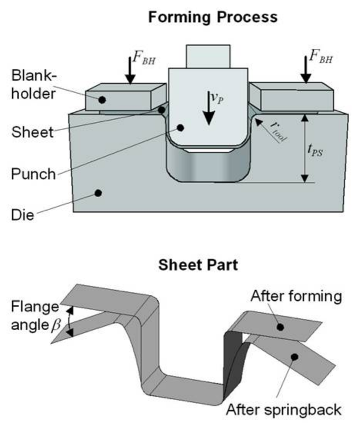

Figure 1: Experimental set up of sheet metal forming

	Level		
Factor	-1	0	1
Blank holder force (NH)	300	600	900
Punch Stroke (ZT)	40	50	60

Table 1: Coded Parameters

We assume normality for both Y_1 and Y_2. Based on a backward selection procedure starting from a model with all interactions and quadratic effects we derive the double generalized linear models:

$$\begin{aligned}
\widehat{\mathrm{E}}(Y_1) &= 46.4390 - 3.2677 \cdot \mathrm{NH} + 10.9302 \cdot \mathrm{ZT} + 1.4479 \cdot \mathrm{NH}^2 - 9.8195 \cdot \mathrm{ZT}^2 \\
\log(\widehat{\mathrm{Var}}(Y_1)) &= 1.6170 - 0.9598 \cdot \mathrm{NH}^2 - 0.7659 \cdot \mathrm{ZT}^2 - 0.4590 \cdot \mathrm{NH} \cdot \mathrm{ZT} \\
\widehat{\mathrm{E}}(Y_2) &= 62.8939 + 3.2706 \cdot \mathrm{NH} + 0.5493 \cdot \mathrm{ZT} + 0.8775 \cdot \mathrm{ZT}^2 \\
\log(\widehat{\mathrm{Var}}(Y_2)) &= 0.3269
\end{aligned}$$

The aim is to calculate parameters that reduce the springback after forming the sheet metal and thus result in small flange angle Y_1 and small sidewall curvature Y_2. In order to achieve this, we set the target values as follows:

$$\tau = \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

We choose only one slope vector

$$d = (0, 1)^T$$

because reverse weighting (i.e. $d = (1, 0)$) just results in opposite behaviour of the responses and parameters. We first use a stretch vector $a = (-20, \ldots, 20)$ with 40 equidistant values between -20 and 20. We use the t-th weight matrix W_t as described in the previous chapter. Now we can calculate optimal parameters with corresponding optimal responses by means of minimizing the risk function for every weight matrix W_t. Figure 2 contains the joint optimization plot.

For each weight matrices (displayed by numbers on the x-axis), the estimated risk function is minimized and the resulting optimal parameter settings are displayed on the y-axis of the plot on the left hand side of Figure 2 and 3. The plot on the right hand side shows the corresponding predictions of the mean model together with a belt with the width of twice the estimated standard deviation from the variance model of the response. Note that the belt width for Y_2 of course is constant due the constant variance model resulting from the model building procedure. The predicted variance for response Y_2 should be as small as possible as a result of minimizing the risk function. This seems to have been achieved to roughly the same extent over the different weightings.

We therefore next concentrate on reaching the target for the mean, hence in case of both Y_1 and Y_2 minimal values. On the right hand side of the plot horizontal lines near the bottom indicate the used target values of zero for both responses. It is quite obvious from the plot that both responses can not be minimized at the same time. The response Y_2 takes minimal values on the left hand side for corresponding low settings of ZT and high values of NH, whereas moderate ZT and low NH values lead to minimal values for Y_1 on the right hand side.

From the plot in Figure 3 it becomes obvious that results stay constant at the lower and upper end of the weight matrices. We therefore reduce the range of the stretch vector to values within -5 and 7. At the same time we increase the precision to 1000 weight matrices.

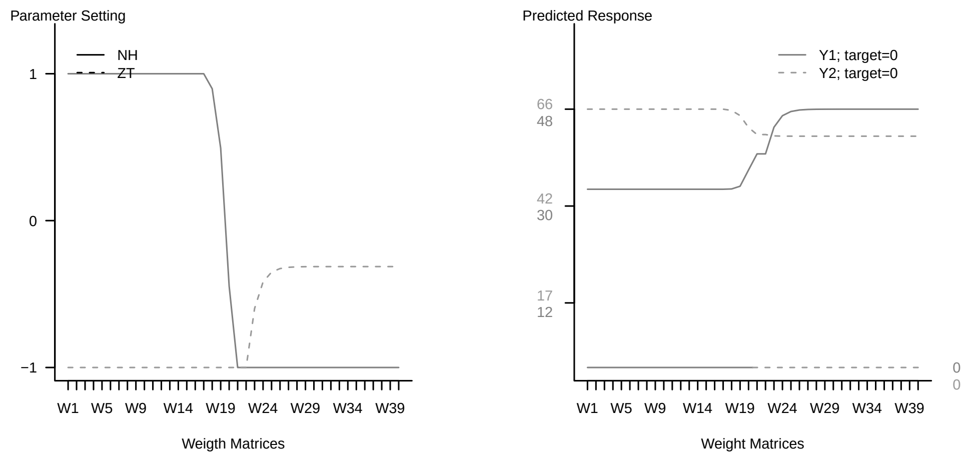

Figure 2: joint optimization plot with 40 weight matrices, $d = (0, 1)$

On the basis of the resulting joint optimization plot the user can now choose a "good" compromise from the right hand plot in Figure 3. This choice requires knowledge of the underlying process. In our example reaching the target for the second response is more important. Thus we choose a point on the right hand side where we get a relatively small flange angle Y_1 and a small side wall curvature Y_2. The corresponding parameters are received from the left hand plot. A possible choice is demonstrated in Figure 3.

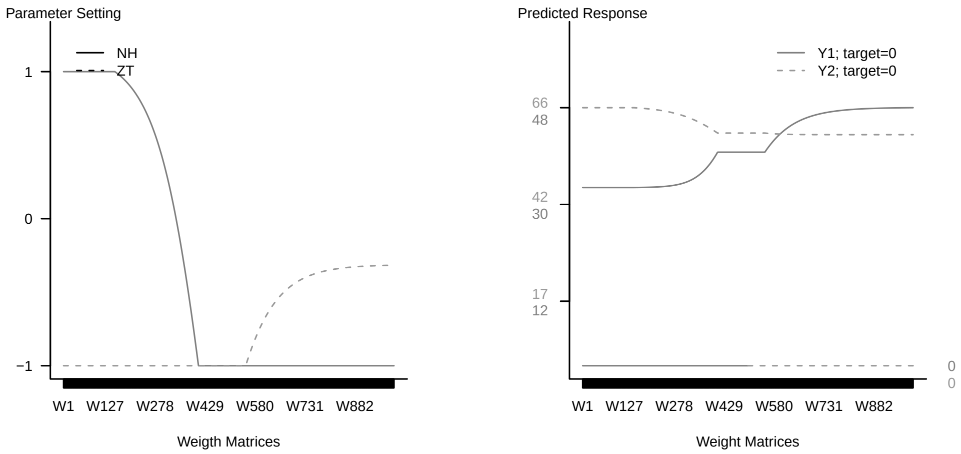

Figure 3: joint optimization plot with 1000 weight matrices, $d = (0, 1)$

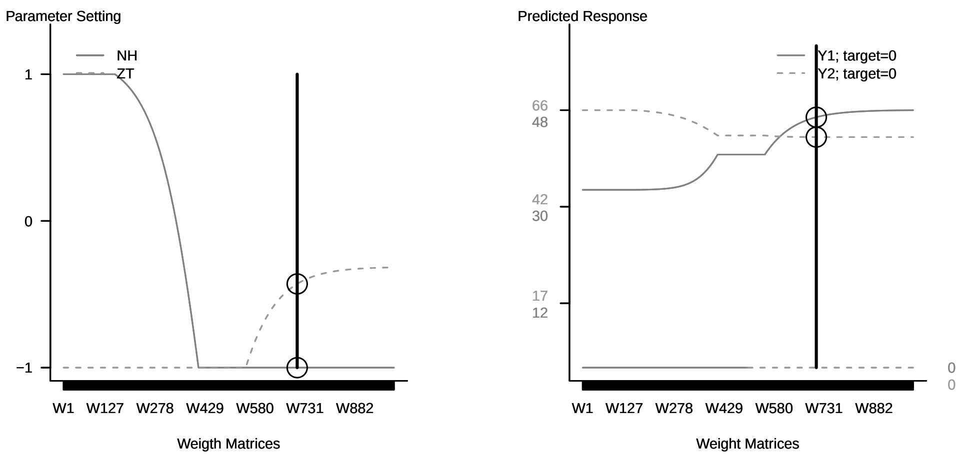

Figure 4: "good" compromise

In Table 2 the chosen "good" compromise is displayed.

Parameters		Responses	
NH	ZT	Y_1	Y_2
-1.00	-0.43	46.31	59.55

Table 2: "good" compromise

Denote by $\mathcal{X}_{\text{spring}}$ the set of all feasible points, $\mathcal{X}_{\text{spring}} = \{(NH, ZT) \in \mathbb{R}^2 \mid -1 \leq NH, ZT \leq 1\}$. Let

$$F_{spring}(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix}$$

be the functions considered in Section 3, where

$$\begin{aligned} f_1(x) &= \widehat{\text{Var}}(Y_1|x) + (\widehat{\text{E}}(Y_1|x) - \tau_1)^2 \\ f_2(x) &= \widehat{\text{Var}}(Y_2|x) + (\widehat{\text{E}}(Y_2|x) - \tau_2)^2. \end{aligned}$$

Figure 5 displays the image of $\mathcal{X}_{\text{spring}}$ under the function $F_{\mathcal{X}_{\text{spring}}}$. The Pareto optimal points for the problem

$$\begin{aligned} \min \quad & F_{\mathcal{X}_{\text{spring}}}(x) \\ s.t \quad & x \in \mathcal{X}_{\text{spring}} \end{aligned}$$

are of interest. Adding the image of x-values from the joint optimization plot in Figure 3 to Figure 5 shows that they cover the complete set of Pareto optimal points.

5.2 Example: thermal spraying process

In this example we want to optimize a thermal spraying process on the basis of a dataset that can be found in Tillmann *et al.*(2010). Tillmann *et al.* (2010) conduct a central composite design with three parameters and three responses with a total run size of 18 experiments. The controllable parameters are

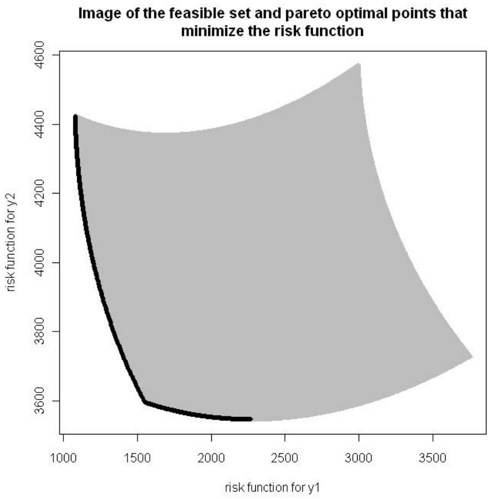

Figure 5: Image $F_{\mathcal{X}_{\text{spring}}}$ and Pareto optimal points

kerosene level (KL) in L/h, stand off distance (SOD) in mm and oxygen level (OL) in L/min. The responses represent coating properties, namely microhardness (MH), deposition efficiency (DE) and roughness in μm (RRa). Figure 6 displays the experimental set up of a thermal spraying process. Table 3 summarizes the coded parameter levels.

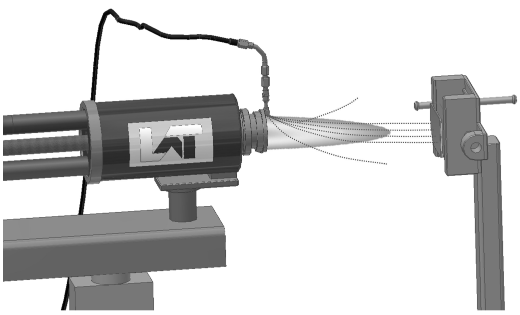

Figure 6: Experimental set up

	Level				
Factor	-2	-1	0	1	2
Kerosene level (KL)	8	9	10	11	12
Oxygen level (OL)	750	800	850	900	950
Stand-off distance (SOD)	100	115	130	145	160

Table 3: Parameter values

We use the linear models as fitted in Tillmann et. al. (2010):

$$\begin{aligned}
\widehat{\mathrm{E}}(\mathrm{MH}) &= 784.57 + 66.77 \cdot \mathrm{KL} + 30.90 \cdot \mathrm{OL}^2 - 60.51 \cdot \mathrm{OL} + 48.07 \cdot \mathrm{SOD}^2 - 46.86 \cdot \mathrm{SOD} \\
\widehat{\mathrm{Var}}(\mathrm{MH}) &= 95.73^2 \\
\widehat{\mathrm{E}}(\mathrm{DE}) &= 55.01 - 4.11 \cdot \mathrm{KL}^2 + 5.54 \cdot \mathrm{KL} - 7.58 \cdot \mathrm{OL} - 4.85 \cdot \mathrm{SOD} - 2.99 \cdot \mathrm{OL} \cdot \mathrm{KL} \\
\widehat{\mathrm{Var}}(\mathrm{DE}) &= 6.64^2 \\
\widehat{\mathrm{E}}(\mathrm{RRa}) &= 2.06 + 0.95 \cdot \mathrm{SOD}^2 - 1.06 \cdot \mathrm{SOD} \\
\widehat{\mathrm{Var}}(\mathrm{RRa}) &= 1.4^2
\end{aligned}$$

Note that in this case the variance is constant.

The aim is to find parameters that produce a coating with high microhardness, high roughness and low porosity. In order to achieve this, we set the target values as follows:

$$\tau = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} = \begin{pmatrix} 1500 \\ 100 \\ 0 \end{pmatrix}$$

With three response variables we may consider the following weighting situations. One variable may get opposite weight to the other two variables, which are treated on an equal footing, i.e. by choosing slope vectors such as $(1,0,0)$, $(0,1,0)$ or $(0,0,1)$. Or two variables are weighted contrary to each other with the third variable beeing inbetween, i.e. slope vectors such as $(1,2,3)$, $(1,3,2)$, $(3,1,2)$ or $(1,2,4)$, $\ldots$, and so on.

We choose the following slope vectors

$$d_1 = (1, 0, 0)^T$$
$$d_2 = (0, 1, 0)^T$$
$$d_3 = (0, 0, 1)^T$$

in order to compare the effect of different weighting of the responses. We choose a stretch vector $a = (-10, \ldots, 10)$ with 100 equidistant values between -10 and 10 for each slope vector for d_2 and d_3. Furthermore we choose a stretch vector $a = (-4, \ldots, 3)$ with 100 equidistant values between -4 and 3 for the slope vector for d_1, because greater weights lead to unreasonable results due to numerical problems. The resulting joint optimization plots are displayed in Figures 7, 8 and 9 together with a vertical line for a chosen optimal result. The slope vector d_1 defines weight of 1 for the coating property MH and weight 0 for DE and RRa. The resulting joint optimization plot leads especially to an improvement of MH compared to RRa on the right hand side of the plot, with the reverse effect on the left hand side, whereas the predicted mean response for DE does not change much with the different weightings. Similar behaviour can be observed for the slope vectors d_2 where hich or low importance is put on DE, resulting in contradictory behaviour with respect to RRa, while MH does not vary much. In Figure 9 RRa is weighted against the other two variables resulting in very low predicted mean values for RRa on the right hand side of the plot. If the engineer has in mind to focus for example on the responses MH regardless the values of DE and RRa, then the slope vector d_1 is a possible choice and a point on the right hand plot is chosen that maximizes MH. The same is true for DE by choosing d_2 and RRa by choosing d_3. In addition the chosen optimal parameters together with optimal responses are summarized in Table 4.

Additionally we consider the slope vector

$$d = (0.45, 0.35, 0.25)^T$$

that corresponds to the weights used for a desirability function approach in

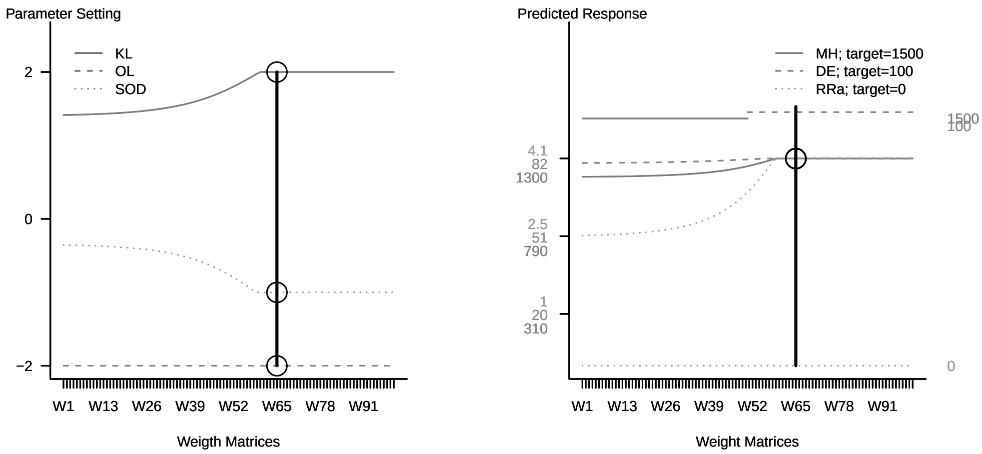

Figure 7: joint optimization plot for d_1 with 100 weight matrices

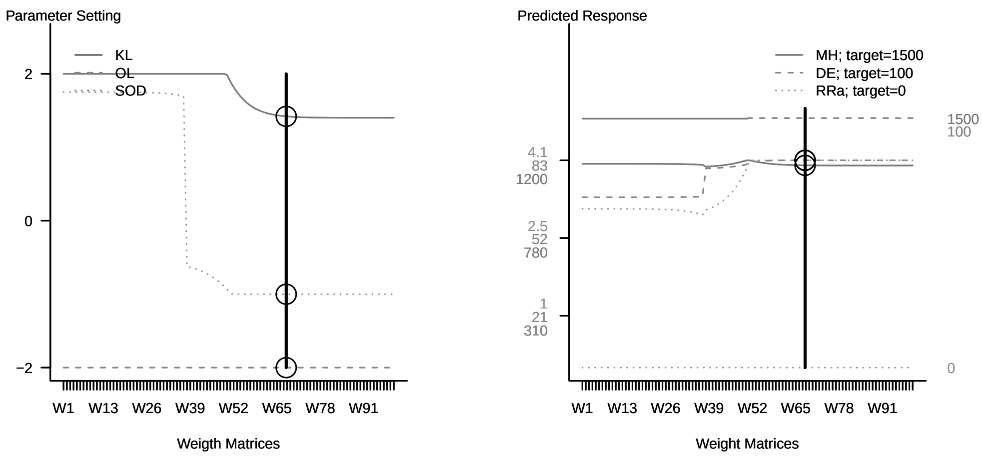

Figure 8: joint optimization plot d_2 with 100 weight matrices

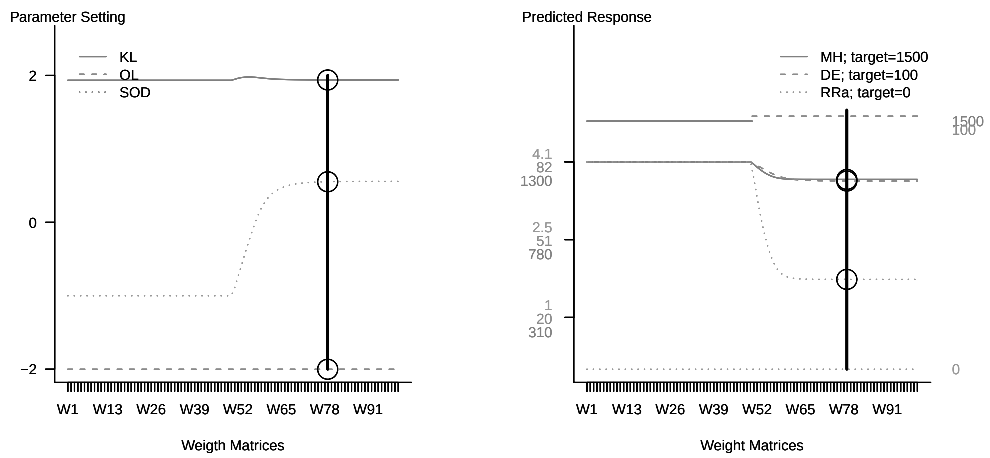

Figure 9: joint optimization plot d_3 with 100 weight matrices

Tillmann *et al.* (2010). They selected these weights in order to ensure an optimum with particularly good MH, which is the most important characteristic for typical applications.

First we choose a stretch vector $a = (-50, \ldots, 20)$ with 20 equidistant values between -50 and 20 in order to identify regions where the responses stay constant or are not reasonable. The resulting joint optimization plots are displayed in Figure 10. We reduce the range of the stretch vector within -25 and 20. At the same time we increase the precision 100 weight matrices. Notice that due constance of the variance functions we only display the responses without standard deviation on the right hand plot. The resulting joint optimization plot is displayed in Figure 11.

We restricted the optimization region by box constraints. In order to compare our results with the optimal parameter setting derived by Tillmann *et al.* (2010) we allow the feasible parameter values to range within -2 and 2. The resulting optimal parameter settings are summarized in Table 4. Notice that the optimal parameter vector () is not in the design region anymore. However, by means of the joint optimization plot we derived a possible further improvement compared to the optimal parameter setting given in Tillmann

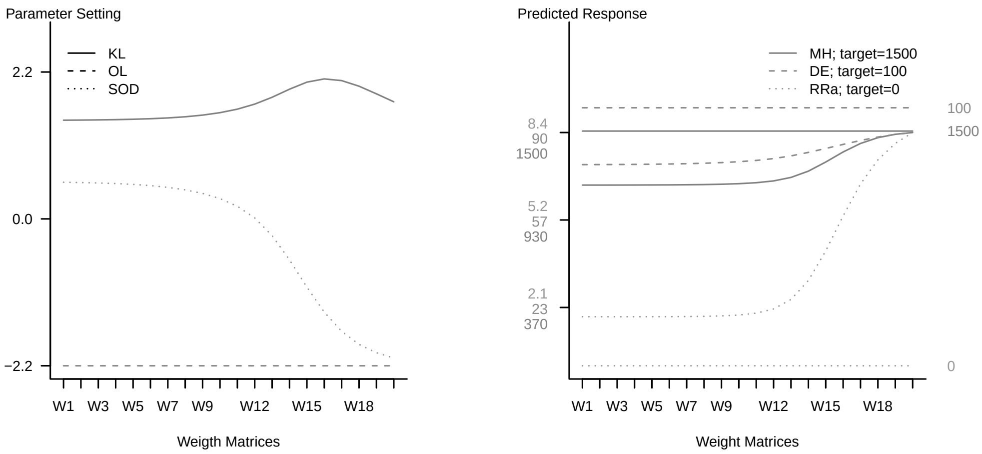

Figure 10: joint optimization plot d with 20 weight matrices

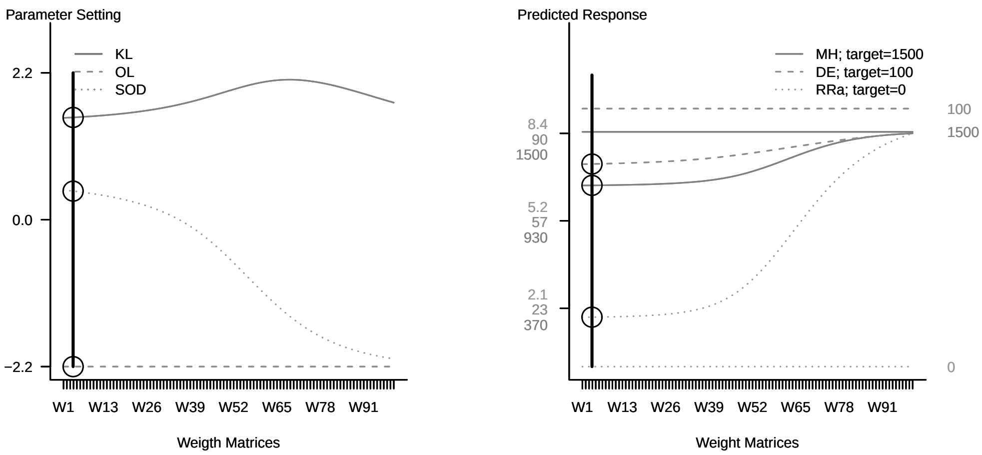

Figure 11: joint optimization plot d with 100 weight matrices

et al. (2010).

Parameters				Responses		
slope vector	KL	OL	SOD	MH	DE	RRa
$d_1 = (1, 0, 0)$	2.00	-2.00	-1.00	1257.66	81.62	4.07
$d_2 = (0, 1, 0)$	1.42	-2.00	-1.00	1219.03	83.09	4.07
$d_3 = (0, 0, 1)$	1.94	-2.00	0.55	1147.45	74.37	1.76
$d = (0.45, 0.35, 0.25)$	1.53	-2.20	0.43	1158.31	78.52	1.78
Tillmann *et al.* (2010)	1.42	-2.2	0.56	1150.90	77.89	1.76

Table 4: "good" compromise

6 Conclusion

In this article we show that the method for multicriterial optimization introduced by Kuhnt and Erdbrügge (2004) gives Pareto optimal points with respect to the considered risk function. We further propose choices of sequences of weight matrices in cases of more than two responses variables. Joint optimization plots display the effect of the resulting optimal parameter setting in terms of the estimated expected response means and variances. We demonstrate the use of joint optimization plots based on two examples. The choice of a "good compromise" with respect to reaching the wanted target values of the responses at small variances then depends on the practical problem.

Acknowledgement

The authors would like to thank the Institute of Forming Technology and Ligthweight Construction at TU Dortmund for the example in Section 5.1. Financial support of the Deutsche Forschungsgemeinschaft (SFB 823, project B1 and SFB 708, project C3) is gratefully acknowledged.

References

[1] Kuhnt S, Erdbruegge M. A strategy of robust paramater design for multiple responses. *Statistical Modelling* 2004; **4**: 249264.

[2] Murphy T, Tsui K-L, Allen J K. A Review of Robust Design Methods for Multiple Responses. *Research in Engineering Design* 2005; **16**: 118-132.

[3] Khuri A I, Mukhopadhyay S. Response Surface Methodology. *Wiley Interdisciplinary Reviews: Computational Statistics* 2010; **2**: 128-149.

[4] Wu F-C. Robust design of nonlinear multiple dynamic quality characteristics. *Computers and Industrial Engineering 2009*; **56**: 1328-1332.

[5] He Z, Wang J, Oh J, Park S H. Robust optimization for multiple responses using response surface methodology. *Applied Stochastic Models in Business and Industry* 2010; **26**: 157-171.

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

[7] Shen L J, Zhao Y, Yang J. Mean Square Error Criteria to Multiple Quality Characteristics Robust Design by the Weighted Tchebycheff Method. *Advanced Materials Research 2010*; **118-120**: 881-885.

[8] Pignatiello J J. Strategies for Robust Multiresponse Quality Engineering. *IIE Transactions* 1993; **25**: 5-15.

[9] Vining, G G. A Compromise Approach to Multiresponse Optimization. *Journal of Quality Technology* 1998; **30**: 309-313.

[10] Smyth G. A Compromise Approach to Multiresponse Optimization. *Journal of Quality Technology* 1989; **51**: 47-60.

[11] Grize Y L. Review: A Review of Robust Process Design Approaches. *Journal of Chemometrics* 1995; **9**: 239-262.

[12] Engel J, Huele A F. (1996). A Generalized Linear Modeling Approach to Robust Design. *Technometrics* 1996; **38**: 365-373.

[13] Chiao C-H, Hamada M. Analyzing Experiments with Correlated Multiple Responses. *Journal of Quality Technology* 2001; **33**: 451-465.

[14] McCullagh P, Nelder J A. *Generalized Linear Models* (2nd edn), Chapman & Hall, London, 1989.

[15] Nelder J A, Lee Y. Generalized Linear Models for the Analysis of Taguchi-type experiments. *Applied Stochastic Models and Data Analysis* 1991; **7**: 107-120.

[16] Lee Y, Nelder J A. Robust Design via Generalized Linear Models. *Journal of Quality Technology* 2003; **35**: 2-12.

[17] Hillermeier C. *Nonlinear Multiobjective Optimization*. Birkhäuser Verlag, Basel, 2001.

[18] Ehrgott M. *Multicriterial Optimization*. Springer Verlag, Berlin Heidelberg, 2000.

[19] Tillmann W, Vogli E, Baumann I, Kopp G, Weihs C. Desirability-Based Multi-Criteria Optimization of HVOF Spray Experiments to Manufacture Fine Structured Wear-Resistant 75Cr3C2-25(NiCr20) Coatings. *Journal of Thermal Spray Technology* 2010; **19**: 392-408.

7 Appendix

The following table contains the data set from example 5.1.

NH	ZT	Y_1	Y_2
−1	−1	39.165	58.48
−1	−1	40.085	60.38
−1	−1	38.975	59.855
−1	−1	39.945	59.16
−1	0	48.805	60.235
−1	0	52.135	61.24
−1	0	51.97	59.95
−1	0	50.905	60.145
−1	0	49.835	59.285
−1	1	59.355	59.91
−1	1	61.305	60.065
−1	1	59.925	62.08
−1	1	62.485	59.23
−1	1	60.97	60.49
−1	1	62.335	59.96
−1	1	62.53	62.095
0	−1	32.36	64.41
0	−1	33.265	62.57
0	−1	35.535	64.345
0	−1	35.9	63.67
0	0	49.01	62.06
0	0	48.165	63.77
0	0	48.455	62.185
0	0	49	62.185
0	1	55.92	63.995
0	1	54.32	64.46
0	1	56.505	64.195
0	1	54.49	66.935

0	1	56.75	66.95
0	1	57.175	63.69
0	1	58.26	64.715
1	−1	33.355	65.695
1	−1	34.16	66.01
1	−1	31.475	65.5
1	−1	32.1	68.59
1	0	43.295	64.935
1	0	43.405	66.86
1	0	44.305	64.065
1	0	46.19	67.435
1	1	53.755	68.04
1	1	55.295	67.485
1	1	55.71	66.075