

Single- and multiobjective optimization problems in robust parameter design

AMIT MATHUR and KRISHNA R PATTIPATI

Department of Electrical and Systems Engineering, University of Connecticut,
Storrs, CT 06269, USA
e-mail: [amit,krishna]@sol.uconn.edu; amit@teamqsi.com

Abstract. This paper reviews the evolution of off-line quality engineering methods with respect to one or more quality criteria, and presents some recent results. The fundamental premises that justify the use of robust product/process design are established with an illustrative example. The use of designed experiments to model quality criteria and their optimization is briefly reviewed. The fact that most design-for-quality problems involve multiple quality criteria motivates the development of multiobjective optimization techniques for robust parameter design. Two situations are considered: one in which response surface models for the quality characteristics can be obtained using regression and considered over a continuous factor space, and one in which the problem scenario and the experiment permit only discrete parameter settings for the design factors. In the former scenario, a multiobjective optimization technique based on the reference-point method is presented; this technique also incorporates an inference mechanism to deal with uncertainty in the response surface models caused by finite, noisy data. In the discrete-factors scenario, an efficient method to reduce computational complexity for a class of models is presented.

Keywords. Multiobjective optimization; off-line quality engineering; Taguchi methods; robust product/process design.

1. Introduction

The primary goal of a quality program in the manufacturing industry is to design and implement a production process geared towards satisfying the customer. It concerns itself with all aspects of the process, from conceptualization through design, to production and operations. Besides the need for an environment of *total quality management* built on a concurrent engineering foundation, the *quality engineering* tools needed to evaluate and optimize physical quality characteristics are paramount to the success of a quality program. Development of some of these tools is the focus of this article.

Quality engineering deals with a methodology for maintaining (controlling) and improving product quality. It addresses two main types of problems: on-line quality-control, and off-line design of products and processes for quality. The on-line techniques have generally been researched and applied under the rubric of statistical process control (SPC), where the popular tools are various types of control charts (Montgomery 1991a). The off-line design-for-quality problem has aroused widespread interest in the West only since the eighties with the popularization of Taguchi's methods of parameter design using designed experiments (also known as *robust design*, Phadke 1989). While SPC aims at *maintaining* the quality of a production process at prescribed levels with the tacit assumption that the design has been optimized prior to production and that no further design improvement is necessary, off-line product development uses the tools of experimental design to obtain models for the various quality metrics of interest, which are then optimized with respect to the design variables (factors). In both exercises, however, the ultimate objective is to ensure that the production system performs as close to the targeted levels and with as little variability as possible.

The prerequisites for a robust product/process-design exercise are the identification of suitable metrics or performance measures that can summarize quality, and the specification of targets for these measures. The objective then is to obtain values of the various product and process parameters that bring these measures to their targets. Manufacturing processes almost always require consideration of several individual quality criteria, whose physical relationships to the control parameters or design variables, as well as with each other, are not always known. The problem, therefore, involves modelling all quality characteristics with respect to the control parameters (input variables) and simultaneous optimization of these quality criteria. Since the same settings of control parameters, in general, cannot optimize all objectives, a trade-off or compromise is inevitable. The formal methodology for doing so, viz., *multiobjective optimization* (also termed multiple-criteria decision-making or multiobjective programming), has been a subject of extensive research in the area of resource management and planning for more than two decades (Zeleny 1982; Chankong & Haimes 1983; Steuer 1989). On the other hand, quality engineering has concerned itself mainly with the statistical and modelling aspects of the problem with the underlying expectation that the number of optimization variables would be small and that the concomitant optimization problem would be solvable via *ad hoc* methods. With the need for increasing sophistication in product design and process improvement in modern manufacturing, it is important to incorporate a formal methodology in quality engineering that integrates both statistical modelling and optimization into a single framework. This is especially so as the number of quality criteria and the number of variables affecting the process or product performance increase, rendering visual/graphical tools cumbersome or ineffective. Multiobjective optimization methods can be applied to quality engineering problems with only a few conceptual extensions.

This paper reviews the evolution of off-line quality engineering methods for single and multiple objectives, and introduces some recent results. Since the focus is on the engineering tools (also known as the CAE (computer-aided engineering) tools, Boza *et al* 1994), the issue of how to identify appropriate quality characteristics and the factors of interest in a production system is ignored – those objectives are best accomplished via extensive brainstorming by the quality team (Bendell *et al* 1989; Boza *et al* 1994). In the

remainder of this article, we first give a general overview of experimental design techniques and the notion of robust design attributed to Taguchi. The optimization of quality metrics with respect to continuous factors is addressed first and then the optimization problem with respect to discrete factors is considered. For the case of continuous design factors, we also examine the optimization problem for the single as well as the multiple objective cases in presence of uncertainty in models estimated from experimental data. We present a strategy for design improvements that account for the uncertainty in regression models. All concepts and methods are illustrated by examples.

2. Overview of off-line design

2.1 Quality metrics and experimental design

The goal of Taguchi’s design-for-quality approach is to choose levels (or settings) of factors (or design parameters) that minimize the variability of a product’s function while achieving the target (Taguchi 1987; Phadke 1989; Nair 1992). This is in contrast with the traditional (and uneconomical) approach of ‘accept–reject,’ where products not satisfying the specifications within a tolerance are rejected. Taguchi’s philosophy is based on the premise that quality can be built into a product by a proper choice of design settings to lower the inherent variability, and thus to lower the cost of producing quality products. This cost first needs to be quantified by a *quality loss function*. A simple quality loss function for a quantitative quality characteristic (or response) y , which is targeted to have a value T , is the quadratic loss function

$$L(y) = k_0(y - T)^2.$$

Figure 1 plots the quadratic loss function in relation to the step loss function that corresponds to the tolerance-based approach; δ is the tolerance limit about the target T , and A_0 is the cost incurred in rejecting a product. Thus, the objective with respect to the quadratic loss function is to achieve product response as close to the target as possible by choosing appropriate parameter settings and without modifying the product technology itself. On the other hand, the step loss function attributes no loss of quality to a product whose response lies within $\pm\delta$ of the target T , and hence no attempt is made to bring product response as close to the target as possible. Quality loss can be reduced in this approach only by improving technology – a more expensive alternative to parameter design.

Let $\mathbf{x} = [x_1, x_2, \dots, x_K]$ be a vector denoting the values of K (controllable) factors which affect the product response y , and whose levels can be fixed by design. Since the

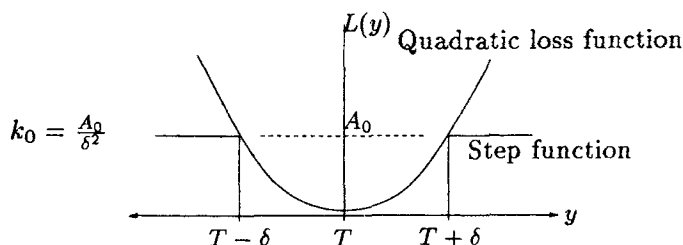


Figure 1. Quality loss functions.

response $y(\mathbf{x})$ at the factor settings \mathbf{x} varies randomly from sample to sample, resulting in a distribution of loss function values, an average (over the entire product population) of the loss function should be considered. For the quadratic loss function, the average loss function (assuming $k_0 = 1$) is the mean squared-error,

$$\text{MSE}(\mathbf{x}) = E[L(y(\mathbf{x}))] = \sigma^2(\mathbf{x}) + (\mu(\mathbf{x}) - T)^2,$$

where E is the expectation operator, and $\mu(\mathbf{x})$ and $\sigma^2(\mathbf{x})$ are, respectively, the mean and variance of the response y parameterized by \mathbf{x} . For the mean squared-error type quality metric, quality is optimized by minimizing $\text{MSE}(\mathbf{x})$ with respect to \mathbf{x} taking values in a set Ξ (factor space).

Taguchi has proposed quality measures, called *signal-to-noise ratios*, for three different types of quality characteristics (e.g., Taguchi 1987; Phadke 1989). For a *smaller-the-better* type quality characteristic, where the target T to be achieved by the characteristic y is zero, it is defined as

$$SN_S(\mathbf{x}) = -10 \log_{10} E[y^2(\mathbf{x})].$$

For a *nominal-the-best* type characteristic, it is defined as

$$SN_T(\mathbf{x}) = 10 \log_{10} [\mu^2(\mathbf{x})/\sigma^2(\mathbf{x})],$$

and for the *larger-the-better* type characteristic, it is defined as

$$SN_L(\mathbf{x}) = -10 \log_{10} E[1/y^2(\mathbf{x})].$$

Taguchi's assertion is that maximization of these signal-to-noise ratios with respect to \mathbf{x} would result in minimization of variability. For the *nominal-the-best* characteristics, he suggests identification of *signal factors* which affect the mean of y , but not its SN_T ; the signal factors can be used to bring the mean to the target following the maximization of SN_T . Statisticians have devoted extensive effort to justify Taguchi's measures and his two-step approach, but have found only very special scenarios in which their use can be validated (León *et al* 1987; Box 1988). Where the quadratic loss function is appropriate, the use of the mean squared-error (MSE) would be ideal, provided it can be empirically modelled using experiments. The fact that signal-to-noise ratios have been reported to have worked in several case studies demonstrates that there is room for significant improvement in many currently used manufacturing processes and that even *ad hoc* methods can realize some of that improvement.

2.2 The scope for robust design

To demonstrate that variability can indeed be reduced by parameter design, consider the following illustrative example (taken from Boza *et al* 1994).

A circuit example: Consider the simple AC circuit shown in figure 2, where the power source \mathcal{E} is known to operate at a tightly toleranced frequency f of either 50 or 60 Hz, and an rms value of 100 VAC with a tolerance of $\pm 10\%$. It is desired that the rms current I be as close to 10 amperes and with as little variability as possible. The design parameters are

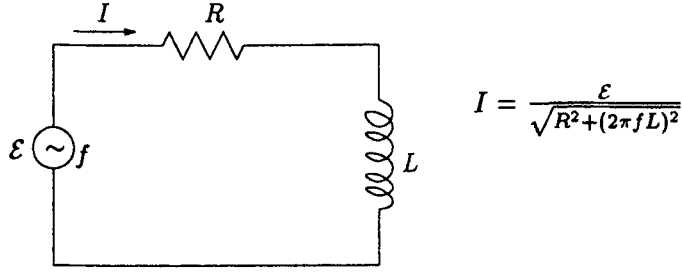


Figure 2. An AC circuit.

the nominal values of the resistance R and the inductance L , which are also tolerated at $\pm 10\%$ about their nominal values. It is assumed that \mathcal{E} , R and L are independent random variables with means at their nominal values and the ($\pm 10\%$) tolerances at 3σ limits from nominal, e.g., the nominal value R_0 of R is such that

$$3\sigma_{R_0} = 0.1R_0.$$

The frequency f is assumed to have zero variance about a nominal of either 50 or 60 Hz.

Using the Taylor series expansion about the nominal values of the circuit parameters, the current I can be expressed as

$$\begin{aligned} I = I_0 + \left. \frac{\partial I}{\partial \mathcal{E}} \right|_0 \Delta \mathcal{E} + \left. \frac{\partial I}{\partial R} \right|_0 \Delta R + \left. \frac{\partial I}{\partial L} \right|_0 \Delta L \\ + \left. \frac{\partial I}{\partial f} \right|_0 \Delta f + \dots \text{(higher-order terms),} \end{aligned} \quad (1)$$

where the subscript zero on I and the partial derivatives denotes their evaluation at the nominal parameter values \mathcal{E}_0 , R_0 , L_0 and f_0 ($\mathcal{E}_0 = 100$ VAC, $f_0 = 50$ or 60 Hz), and

$$\Delta \mathcal{E} = \mathcal{E} - \mathcal{E}_0, \quad \Delta R = R - R_0, \quad \Delta L = L - L_0, \quad \Delta f = f - f_0.$$

Using the assumptions that the parameters are independent and that the terms of third and higher orders are negligible, taking the expectation of I in (1) gives the mean value of the output current as

$$\begin{aligned} \mu_I(\mathcal{E}_0, f_0, R_0, L_0) = I_0 + \frac{1}{2!} \left(\left. \frac{\partial^2 I}{\partial \mathcal{E}^2} \right|_0 \sigma_{\mathcal{E}_0}^2 + \left. \frac{\partial^2 I}{\partial R^2} \right|_0 \sigma_{R_0}^2 \right. \\ \left. + \left. \frac{\partial^2 I}{\partial L^2} \right|_0 \sigma_{L_0}^2 + \left. \frac{\partial^2 I}{\partial f^2} \right|_0 \sigma_{f_0}^2 \right) + \dots \\ \simeq I_0 + \frac{1}{2} \left(\left. \frac{\partial^2 I}{\partial R^2} \right|_0 \sigma_{R_0}^2 + \left. \frac{\partial^2 I}{\partial L^2} \right|_0 \sigma_{L_0}^2 \right). \end{aligned} \quad (2)$$

In obtaining (2), we have also used the facts that $(\partial^2 I / \partial \mathcal{E}^2) \equiv 0$ and $\sigma_f^2 \simeq 0$. For the nominal values being considered, the bias $E(I) - I_0$ is negligible (less than 0.1%), and so the variance of I can be obtained as

$$\sigma_I^2(\varepsilon_0, f_0, R_0, L_0) \simeq \left[\frac{\partial I}{\partial \varepsilon} \right]_0^2 \sigma_{\varepsilon_0}^2 + \left[\frac{\partial I}{\partial R} \right]_0^2 \sigma_{R_0}^2 + \left[\frac{\partial I}{\partial L} \right]_0^2 \sigma_{L_0}^2. \quad (3)$$

With the choice of ε_0 and f_0 not in the designer's control, the mean squared-error

$$\text{MSE}_I(R_0, L_0) = \sigma_I^2(R_0, L_0) + (\mu_I(R_0, L_0) - T_I)^2,$$

(where $T_I = 10\text{A}$ is the target for I) can be minimized with respect to the nominal values R_0 and L_0 of the series resistance and inductance to obtain a robust circuit design. Using the functional relationship between I and the circuit parameters given in figure 2, the following partial derivatives can be derived:

$$\begin{aligned} \frac{\partial I}{\partial \varepsilon} &= \frac{I}{\varepsilon}; & \frac{\partial^2 I}{\partial \varepsilon^2} &= 0; \\ \frac{\partial I}{\partial R} &= -\frac{I^3 R}{\varepsilon^2}; & \frac{\partial^2 I}{\partial R^2} &= \frac{I^3}{\varepsilon^2} \left(\frac{3R^2 I^2}{\varepsilon^2} - 1 \right); \\ \frac{\partial I}{\partial L} &= -(2\pi f)^2 \frac{I^3 L}{\varepsilon^2}; & \frac{\partial^2 I}{\partial L^2} &= (2\pi f)^2 \frac{I^3}{\varepsilon^2} \left(\frac{3(2\pi f)^2 L^2 I^2}{\varepsilon^2} - 1 \right). \end{aligned}$$

The contours of the mean squared-error versus the nominal values of R and L are shown in figure 3 for both nominal frequencies. It can be seen that the surface has ridge-like behaviour, and, hence, for large ranges of R and L values, the mean squared-error is very close to its minimum.

Boza *et al* (1994) assumed an initial design setting of $(R_0, L_0) = (5\Omega, 0.02\text{H})$, and obtained a final design of $(R_0, L_0) = (8\Omega, 0.017\text{H})$. Their objective, however, was only the minimization of the variability σ_I^2 which was a sufficient goal as the customer-specified tolerance of 2.5 amperes on the targeted I was not violated. However, to achieve good robustness at both nominal frequencies, figure 4 suggests a design near $(R_0, L_0) = (9.43\Omega, 0.01\text{H})$, when at both frequencies an $\text{MSE} = 0.21\text{A}^2$ is achieved.

Table 1 gives a summary of the circuit's performance at the initial design, the design obtained by Boza *et al* (1994), the designs at which MSE is optimum at either frequency,

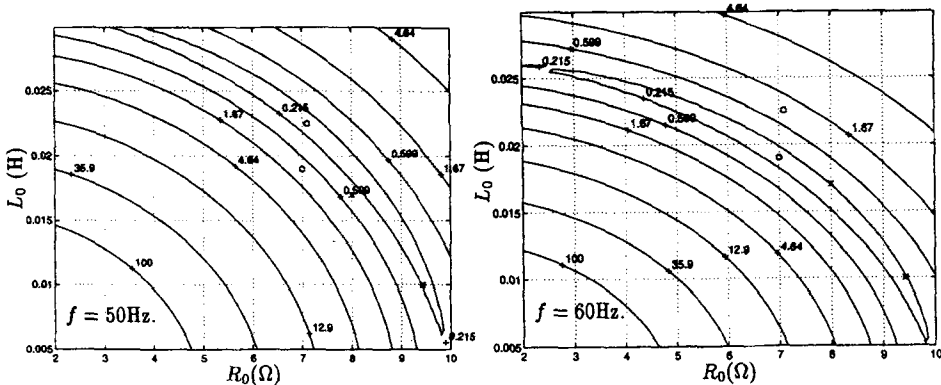


Figure 3. Contour-plots of MSE_I versus R_0 and L_0 at 50 and 60 Hz. The '*' marks our final design, the 'x' marks the design obtained by Boza *et al* (1994), and the 'o's mark the individual optima of MSE_I for the two frequencies.

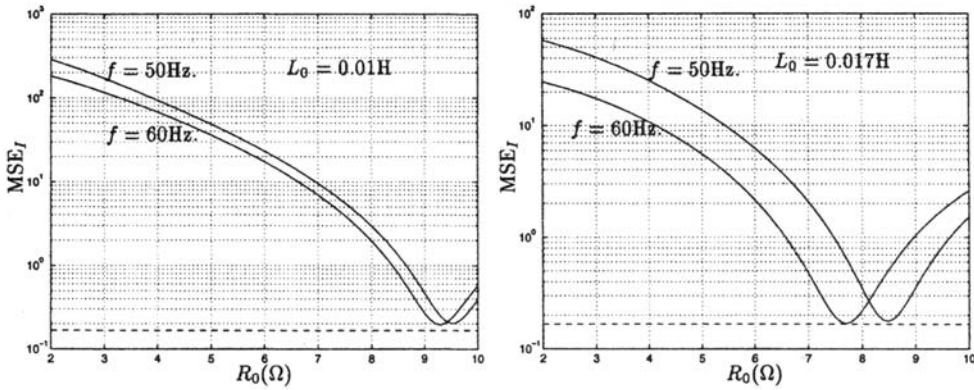


Figure 4. MSE profiles versus R_0 at two fixed values of L_0 ; the dashed line indicates the minimum achievable MSE under either frequency scenario.

and at our suggested design. Table 1 also gives the values of the *capability index* C_{pk} at either frequency for each parameter design. C_{pk} is defined as (e.g., Montgomery 1991a),

$$C_{pk} = \min \left\{ \frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma} \right\},$$

where USL and LSL are respectively the upper and lower specification limits for the response (for this example, $USL_I = 12.5A$ and $LSL_I = 7.5A$). If the above expression returns a negative number, C_{pk} is taken to be zero. For Six-Sigma Quality (pioneered by Motorola Inc; Harry 1988), a C_{pk} value should be greater than or equal to 1.5. For this example, robust design has indeed achieved this goal. Further performance gains are possible only by technological improvement, e.g., by improving the component tolerances.

The above example demonstrates that in order for the product to function equally well under two frequency scenarios, a *compromise* design should be used. While this compromise design was relatively easy to obtain using the contour plots of figure 3 and the MSE profiles in figure 4, a visual approach is not possible if the number of design factors or the number of objectives exceeds two. The problem of multiple objectives is examined in the next section.

2.3 Experimental response modelling

In the example considered in the previous sub-section, a physical relationship between the response of interest and the design factors was available. In a general manufacturing

Table 1. Performance of AC circuit for different parameter designs.

R_0 (Ω)	L_0 (H)	$f = 50$ Hz				$f = 60$ Hz			
		μ_I (A)	σ_I (A)	MSE_I (A^2)	C_{pk}	μ_I (A)	σ_I (A)	MSE_I (A^2)	C_{pk}
5	0.02	12.458	0.513	6.302	0.03	11.058	0.462	1.333	1.04
7.1	0.0225	9.984	0.408	0.166	2.03	9.06	0.371	1.053	1.39
7	0.019	10.874	0.446	0.962	1.22	9.988	0.408	0.166	2.03
8	0.017	10.4	0.435	0.349	1.61	9.759	0.401	0.219	1.88
9.43	0.01	10.069	0.452	0.210	1.79	9.854	0.436	0.211	1.80

process, this would rarely be the case. Consequently, empirical models for the metrics are obtained. An experiment consists of a systematic variation of the factors and recording of the response(s) defining the quality metric(s) to enable reliable estimation of the models. For an experiment of N runs, the procedure involves obtaining (noisy) measurements $y(\mathbf{x})$ of a response at N distinct values $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ in the space Ξ of the design factors. In Taguchi experiments, since the objective of interest is an aggregate measure, such as the signal-to-noise ratio of a product/process characteristic, which summarizes the variability of quality characteristic for a given combination of design-factor settings, an experiment involves variation of not only the *design factors*, but also a systematic variation of *noise factors* that contribute to the variability in the product's response. Noise factors are variables that are uncontrollable in the real environment but can be manipulated in the experimental set-up. Thus, a Taguchi experiment consists of (i) an *inner array* which lists the design factor-level combinations, and (ii) an *outer array* that lists the noise factor-level combinations for each design factor-level combination. If m is the number of combinations of noise factors chosen for an outer array, an experiment yields m replicate measurements for each of the N runs. What the best placement of the points $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ should be depends on the assumed model-type for $\eta(\mathbf{x})$ (e.g., linear, quadratic etc.), and on practical considerations. The problem of design of experiments has been extensively researched during this century by applied statisticians, and there exists a rich body of literature documenting numerous design types and approaches (see, for example, Box *et al* 1978; Montgomery 1991b, and the bibliographies therein). In this article, we shall sidestep this issue by assuming that, for a given problem, an appropriate design has been identified. However, for the purposes of this paper, we mention one useful and frequently used design-type: the *fractional-factorial design*. Most of the *orthogonal-array* designs compiled and recommended by Taguchi are of the fractional-factorial type. For example, consider an experiment involving the study of the effects of three factors x_1, x_2 and x_3 on the response y . If each factor is constrained to take only two fixed levels, denoted by 1 and 2, during the experiment, the maximum number of distinct runs N is 8, so that all combinations of the levels of each factor are tested. This arrangement, shown in the array L_8 in figure 5, is a two-level *full-factorial* design. If only four of the eight possible combinations are run,

Run no.	Factor			y
	x_1	x_2	x_3	
1	1	1	1	y_1
2	1	1	2	y_2
3	1	2	1	y_3
4	1	2	2	y_4
5	2	1	1	y_5
6	2	1	2	y_6
7	2	2	1	y_7
8	2	2	2	y_8

Run no.	Factor			y
	x_1	x_2	x_3	
1	1	1	1	y_1
2	1	2	2	y_2
3	2	2	1	y_3
4	2	1	2	y_4

Figure 5. Two orthogonal factorial designs.

as in the array L_4 in figure 5, the resulting design is of the fractional-factorial type. A useful feature of fractional-factorial designs is their orthogonality, i.e., the analysis of data with respect to each factor can be done using only the column(s) corresponding to that factor. Orthogonal arrays (which include some fractional-factorial designs) also possess the following key features:

- each level of a factor appears in an equal number of runs (factor-level combinations); this is called the balancing property – it ensures unbiased estimation of all column effects;
- the effects measured by any two columns are mutually orthogonal (uncorrelated);
- the effect measured by a column may be aliased (confounded) with the combined effect (interaction) of two or more columns.

Orthogonal arrays are normally used in Taguchi's off-line design procedure for the *inner array* which determines the combinations of the design factor-levels. In addition, their use has also been recommended in the *outer array* which determines the combinations of noise factor-levels, even though the statistical implications of doing are debatable (Montgomery 1991a).

2.4 Data analysis for orthogonal-array experiments

In this sub-section, we briefly review the common data analysis techniques for experimental data: analysis of means, analysis of variances (ANOVA), and multiple linear regression. These techniques are used to estimate the functional relationship between the factors and a response, and to assess the statistical significance of the estimated relationship. While ANOVA and regression are the common tools of choice, the analysis of means is often convenient for measuring factor effects when an orthogonal-array experiment is used. The following results are given with respect to analysis of experimental data using an orthogonal array.

If a trial (or run) corresponds to the level combination (x_1, x_2, \dots, x_K) of the K factors (denoted by \mathbf{x}), then we shall assume that a response y can be expressed as

$$y(\mathbf{x}) = \eta(\mathbf{x}) + \varepsilon,$$

where $\eta(\mathbf{x})$ is an unknown deterministic function — the mean response corresponding to this factor-level combination, and ε is a zero-mean random noise with variance σ^2 . Here, the response y is used to denote measurements of either the quality characteristic or any suitable loss function derived from it, e.g., the MSE or a signal-to-noise ratio. In either case, we assume that ε is independent with constant variance over all experimental runs. The task of experimental modelling is eventually to estimate the functional relationship $\eta(\mathbf{x})$.

2.4a Analysis of means and variances: Consider a balanced experiment in which each factor-level combination appears in an equal number of trials, and, consequently, each of the L_k levels of factor x_k appears in an equal number, n_k , of trials, i.e., $n_k L_k = N$. The number of trials N need not be equal to the total number of possible factor-level combinations, but in a factorial design, for example, N must divide $\prod_{k=1}^K L_k$. We assume

that we take m replicates per trial and, hence, have Nm observations of the response variable: y_{ij} , $i = 1, 2, \dots, N$, $j = 1, 2, \dots, m$ (note that if y denotes an aggregate or summary measure such as an MSE or a signal-to-noise ratio, we would have $m = 1$ since replicated measurements of the quality characteristic would already have been used to compute y). Now, the following sample averages and sums of squares may be defined:

$$\bar{y} = \frac{1}{Nm} \sum_{i=1}^N \sum_{j=1}^m y_{ij}, \quad (4)$$

$$\bar{y}_i = \frac{1}{m} \sum_{j=1}^m y_{ij}, \quad i = 1, \dots, N, \quad (5)$$

$$\bar{y}_{k,l} = \frac{1}{n_k m} \sum_{\{i:l_k=l\}} \sum_{j=1}^m y_{ij}, \quad k = 1, 2, \dots, K; \quad l = 1, 2, \dots, L_k, \quad (6)$$

$$S = \sum_{i=1}^N \sum_{j=1}^m (y_{ij} - \bar{y})^2, \quad (7)$$

$$S_k = mn_k \sum_{l=1}^{L_k} (\bar{y}_{k,l} - \bar{y})^2, \quad k = 1, 2, \dots, K, \quad (8)$$

$$S_e = \sum_{i=1}^N \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2, \quad (9)$$

$$S_r = S - \sum_{k=1}^K S_k = S_e + S_{LOF}, \quad (10)$$

where \bar{y} is the overall mean, \bar{y}_i is the average response in the i th trial, $\bar{y}_{k,l}$ is the average response due to factor k taking level l , $\{i : l_k = l\}$ is the set of indices denoting trial numbers in which factor k takes level l . S is the overall sum of squares (corrected for the mean), S_k is the sum of squares contributed by the variation due to changes in the levels of factor k , and S_e is the sum of squares due to pure random error (it accounts for the within-run variations). S_r , known as the residual sum of squares, accounts for the contribution S_{LOF} to the variability in y due to lack of fit (unaccounted-for effects, e.g., interactions), and due to pure random error (S_e). The first three of the above set of equations constitute the analysis of means (ANOM) which enables measurement of effects of factors. The remaining four equations are used for ANOVA, which enables the identification of significant factor effects that should be included in the model for η . In (6), the mean responses have been computed only for each of the factors taking a given level (main effects) assuming a linear (additive) model for η . But a factorial design and some orthogonal arrays may permit estimation of not just the main effects but also interaction effects that measure the combined effect of two or more factors on the response. A two-level factorial design, for example, can measure as many effects (main or interaction) as the number of columns in the array. The computation of these effects is similar to that given in (6) since exclusive (orthogonal) columns in the array can be assigned to measure these effects provided N is large enough to accommodate their estimation (see Box *et al* 1978; Montgomery 1991b). In general, if p (main and interaction) effects can be measured by an orthogonal array ($p \leq N - 1$), a

prediction equation for η can be obtained as,

$$\hat{\eta}(l_1, l_2, \dots, l_K) = \bar{y} + \sum_{i=1}^p (\bar{y}_{i,l_i} - \bar{y}), \quad (11)$$

where $\hat{\eta}(l_1, l_2, \dots, l_K)$ is the estimated response at the factor-level combination (l_1, l_2, \dots, l_K) of the K factors; the levels of the interaction effects are determined from those of the corresponding interacting factors. The *goodness of fit* of the model can be measured by the R^2 value defined as

$$R^2 = \left(\sum_{i=1}^p S_i \right) / S.$$

The maximum possible R^2 is $(S - S_e)/S$; if the number of replications $m = 1$, the maximum possible R^2 is 1.

2.4b Multiple linear regression: An ANOVA is usually accompanied by least-squares regression to estimate the response-surface model for η . ANOVA gives the significant effects that can be included in a polynomial model for η . That is, $\eta(\mathbf{x})$ is assumed to have the functional form

$$\eta(\mathbf{x}) = \boldsymbol{\theta}' \mathbf{z}(\mathbf{x}),$$

where

$$\mathbf{z}(\mathbf{x}) = [1 \ x_1 \ \dots \ x_k \ x_1^2 \ \dots \ x_k^2 \ x_1 x_2 \ \dots]'$$

is a $p \times 1$ vector containing 1 as the first element, and powers and cross-products of x_1, x_2, \dots, x_k , that are found significant from ANOVA, as the remaining $p - 1$ elements. Assuming $m = 1$, if $\mathbf{z}(\mathbf{x}_i)'$, $i = 1, \dots, N$, corresponding to the N runs, constitute the rows of the $N \times p$ matrix X , and if $\mathbf{y} = [y_1, \dots, y_N]'$ denotes the vector of N independent measurements of the response, least-squares regression gives the estimates (Anderson 1984),

$$\hat{\boldsymbol{\theta}} = (X'X)^{-1} X' \mathbf{y}, \quad (12)$$

$$\hat{\eta}(\mathbf{x}) = \hat{\boldsymbol{\theta}}' \mathbf{z}(\mathbf{x}), \quad (13)$$

for the model coefficients and the response surface.

2.4c M responses: Multiple linear regression can be extended in a straightforward manner to the case of simultaneous modelling of M responses using design matrix X . Now M vectors of coefficients, represented as the M columns of the $p \times M$ matrix Θ in the model,

$$\mathbf{y}(\mathbf{x}) = \Theta' \mathbf{z}(\mathbf{x}) + \boldsymbol{\varepsilon},$$

where $\Theta = [\boldsymbol{\theta}_1 \ \dots \ \boldsymbol{\theta}_M]$ are to be estimated, the measurements \mathbf{y} and the noise $\boldsymbol{\varepsilon}$ are M -vectors, and $\boldsymbol{\varepsilon}$ is assumed to be zero-mean with $M \times M$ covariance matrix Σ . Now, based on the measurement data

$$Y = X\Theta + \mathbf{E},$$

where Y and \mathbf{E} are $N \times M$, and X is $N \times p$, the least-squares estimates are

$$\hat{\Theta} = (X'X)^{-1}X'Y, \quad (14)$$

$$\hat{\eta}(\mathbf{x}) = \hat{\Theta}'\mathbf{z}(\mathbf{x}). \quad (15)$$

The least-squares estimates satisfy the following statistics (Anderson 1984)

$$E(\hat{\theta}_j) = \theta_j \quad (\text{unbiased})$$

$$\text{Cov}(\hat{\theta}_i, \hat{\theta}_j) = \sigma_{ij}(X'X)^{-1}$$

$$E(\hat{\eta}(\mathbf{x})) = \Theta'\mathbf{z}(\mathbf{x}) = \eta(\mathbf{x}) \quad (\text{unbiased}) \quad (16)$$

$$\text{Cov}(\hat{\eta}(\mathbf{x})) = \mathbf{z}(\mathbf{x})'(X'X)^{-1}\mathbf{z}(\mathbf{x})\Sigma. \quad (17)$$

σ_{ij} is the ij th element of Σ ($\sigma_{jj} = \sigma_j^2$). Also,

$$\hat{\Sigma} = [1/(N-p)]Y'[I_N - X(X'X)^{-1}X']Y, \quad (18)$$

where I_N is the $N \times N$ identity matrix, and $\hat{\Sigma}$ is the unbiased estimator of Σ provided Y is of rank M ($M \leq N - p$).

Under normality assumptions, the estimates are distributed according to the following two independent (multivariate) distributions

$$(\hat{\theta}'_1 \hat{\theta}'_2 \cdots \hat{\theta}'_M)' \sim \mathcal{N}((\theta'_1 \theta'_2 \cdots \theta'_M)', \Sigma \cdot (X'X)^{-1}), \quad (19)$$

$$(N-p)\hat{\Sigma} \sim W(\Sigma, N-p), \quad (20)$$

where $\Sigma \cdot (X'X)^{-1}$ is a scalar product of the two matrices (an $mp \times mp$ matrix with each element of Σ multiplied by the matrix $(X'X)^{-1}$). In (20), $W(\Sigma, N-p)$ is a Wishart distribution with $(N-p)$ degrees of freedom, parameterized by the matrix Σ (see Anderson 1984). The diagonal elements of a $W(\Sigma, N-p)$ -distributed matrix are chi-squared distributed random variables with $N-p$ degrees of freedom.

In the rest of this article, the use of prediction models, obtained either from least-squares regression or analysis of means, for optimizing quality criteria will be considered.

3. Multiobjective optimization

Consider now the robust design of a product/process with respect to M quality metrics. Ignoring for now their estimation using experimentation, suppose the response functions $\eta(\mathbf{x}) = [\eta_1(\mathbf{x}) \eta_2(\mathbf{x}) \cdots \eta_M(\mathbf{x})]'$ are perfectly known. The design problem then involves obtaining the process variable settings $\mathbf{x}^* \in \Xi$ that simultaneously result in the most desirable compromise of the m responses. As against single-objective optimization, in which the optimum response is a unique value (maximum or minimum response), 'optimization' with respect to multiple objectives refers to the attainment of any one solution in a set termed the *nondominated solution set*. A solution is said to be nondominated (or Pareto-optimal, or noninferior) if no other solution is at least as good as this solution with respect to every objective and better than this solution with respect to at least one

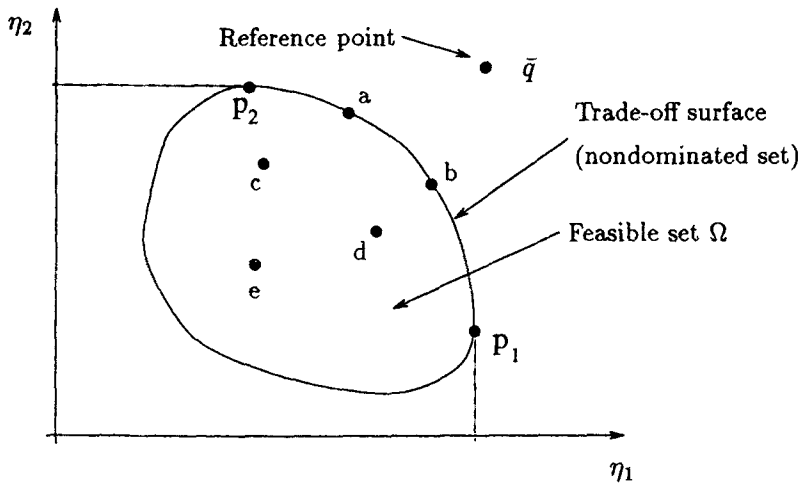


Figure 6. The concept of nondominated solutions (in a two-dimensional objective space).

objective (see Chankong & Haimes 1983). One can use the partial ordering available in the M -dimensional objective space \mathcal{R}^M to identify the nondominated set by eliminating each of those feasible points that is definitely worse than at least one other feasible point (with respect to all M objectives). The concept of nondominated points is illustrated in figure 6 for the case of maximization of two objectives ($M = 2$), where the set of attainable (feasible) objective-pairs is shown as a closed set Ω . Seven points in this set: **a**, **b**, **c**, **d**, **e**, **p**₁ and **p**₂, are also marked for the purpose of illustration. Point **a** dominates **c** and **e** since **a** is better than **c** as well as **e** with respect to both objectives, η_1 and η_2 . Similarly, **b** dominates **d** and **e**. Points **a**, **b**, **p**₁, **p**₂, and all other points on that segment of the boundary of the feasible region Ω between **p**₁ and **p**₂, are nondominated, while **c**, **d**, **e**, and all other points inside the feasible region are dominated.

The concept of domination or partial ordering in an M -dimensional objectives space \mathcal{R}^M can be mathematically defined using the concept of a *positive cone* D (any closed, convex, proper cone):

$$\mathbf{q}_1, \mathbf{q}_2 \in \mathcal{R}^M, \mathbf{q}_1 \leq \mathbf{q}_2 \iff \mathbf{q}_2 - \mathbf{q}_1 \in D.$$

In figure 7, since the point $\mathbf{q}_2 - \mathbf{q}_1$ is in D , \mathbf{q}_2 dominates \mathbf{q}_1 , and since both individually lie in D as well, both dominate the zero-vector $\mathbf{0}$. A nondominated objective (or D -maximal or Pareto-optimal objective) $\check{\mathbf{q}}$ in Ω is defined by

$$\check{\mathbf{q}} \in \Omega \text{ is } D\text{-maximal} \iff \Omega \cap (\check{\mathbf{q}} + \tilde{D}) = \emptyset.$$

The selection of any particular nondominated solution from the set of all nondominated solutions in Ω must be qualified by the preferences of the decision-maker (process- or quality-engineer), for the choice of any one nondominated point over the others implies a *trade-off* of one or more objectives for a gain in another objective. There exist several different, but closely related, methods of incorporating a decision-maker's preferences to search for the final solution (see, for example, Zeleny 1982, Steuer 1989). In one such method, called the *method of reference points* (Wierzbicki 1980), a reference point (a vector of desired objectives) is specified by the decision maker. The method maximizes

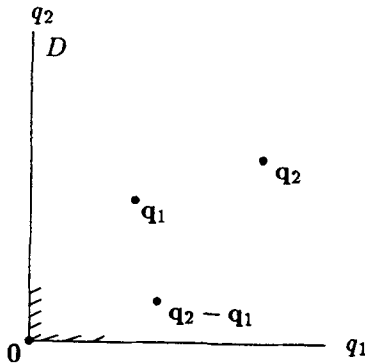


Figure 7. Positive cone D in a two-dimensional ($M = 2$) objectives space \mathcal{R}^2 .

a scalarizing function called the *achievement function* that guarantees a nondominated solution. This is introduced in the next section.

4. Multiobjective optimization of response surfaces

Most designed experiments enable estimation of models for the quality metrics over a continuous region Ξ using least-squares regression. In this case, response surface models (Box & Draper 1987) are available for optimization, and nonlinear programming methods can be used. For multiobjective optimization, the *reference point method* can be applied by maximizing a scalarizing function (called the *achievement function*) of the M response models over the feasible space. One possible scalarizing function is (Lewandowski *et al* 1989)

$$s(\mathbf{q}, \bar{\mathbf{q}}) = \min_{i \in \{1, 2, \dots, M\}} \alpha_i (q_i - \bar{q}_i) + \beta \sum_{j=1}^M \alpha_j (q_j - \bar{q}_j), \quad (21)$$

defined for a general point \mathbf{q} in the objectives space $\Omega \subset \mathcal{R}^M$, and for a reference-point $\bar{\mathbf{q}}$. β and α_i , $i = 1, \dots, M$ are positive constants, whose values are to be fixed by the decision-maker.

Figure 8 illustrates the concept of the achievement function in (21) in a two-objectives space, where the two objectives q_1 and q_2 are to be maximized subject to the constraint that they lie in the set Ω of attainable objective-pairs. The reference point $\bar{\mathbf{q}}$ that is chosen in this figure is unattainable; in general, it could lie inside Ω and the reference-point method would still return a nondominated solution. The ‘ideal’ point (Khuri & Conlon 1981), whose coordinates are the maxima of the individual objectives, is also shown, denoted here by \mathbf{q}_I . (In an example that we have considered later in this section, the ideal point is chosen as the reference point.) The contours of the achievement function in (21) are shown for two values: $s = 0$, and $s = s_2$ ($s_2 < 0$). The shape of these contours is also the shape of the chosen domination cone D_β , e.g., all points right of the contour corresponding to $s = 0$ dominate the reference point $\bar{\mathbf{q}}$. The boundary $\bar{\Omega}$ of the attainable set Ω indicated by the thick line comprises the nondominated set, i.e., no other point in Ω dominates a point in this set with respect to both objectives. The point \mathbf{q}^* is the multiobjective optimum solution obtained by the maximization of the chosen achievement function over $\mathbf{q} \in \Omega$ (the

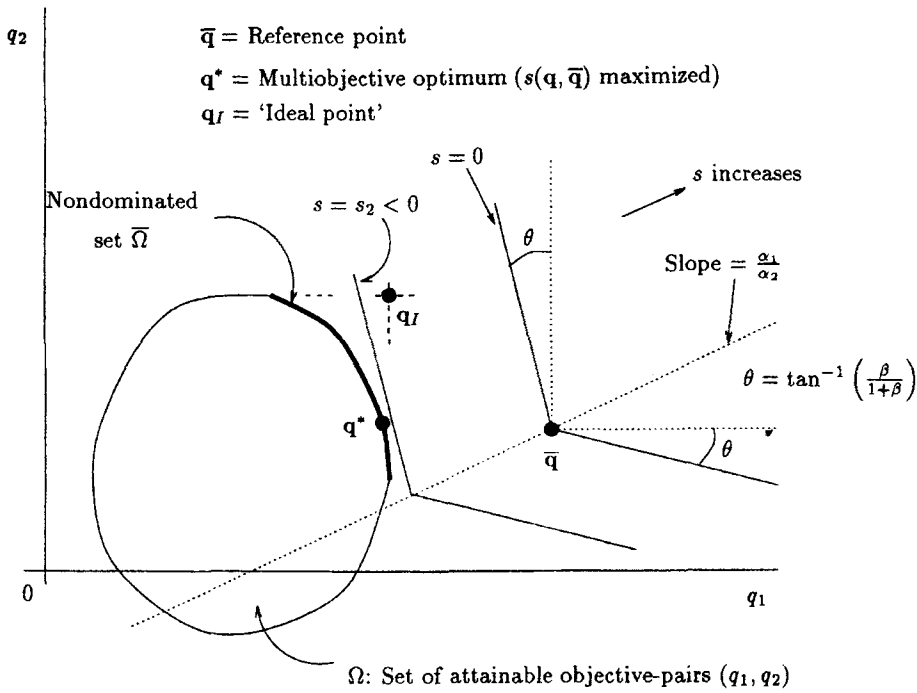


Figure 8. Concepts of reference point and achievement function for a two-objective maximization case.

s -contour is tangent to the set Ω at \mathbf{q}^* . β is related to the angle θ as shown in figure 8. It determines the order that a decision-maker wishes to assign to points that do not dominate one another. The parameters α_1 and α_2 determine the slope of the dividing plane between the two piecewise-linear regions for s .

The optimization problem can now be expressed as

$$\max_{\mathbf{q} \in \Omega} s(\mathbf{q}, \bar{\mathbf{q}}) \quad (22)$$

or, using $\mathbf{q} = \boldsymbol{\eta}(\mathbf{x})$,

$$\max_{\mathbf{x} \in \Xi} s(\boldsymbol{\eta}(\mathbf{x}), \bar{\mathbf{q}}). \quad (23)$$

When a unique, deterministic mapping $\boldsymbol{\eta} : \Xi \rightarrow \mathcal{R}^M$ is not known, and response surface models have to be used, one may instead maximize

$$s(\hat{\boldsymbol{\eta}}(\mathbf{x}), \bar{\mathbf{q}}) = \min_{i \in \{1, 2, \dots, M\}} \alpha_i (\hat{\eta}_i(\mathbf{x}) - \bar{q}_i) + \beta \sum_{j=1}^M \alpha_j (\hat{\eta}_j(\mathbf{x}) - \bar{q}_j). \quad (24)$$

The issue that the objectives are no longer deterministic, but involve modelling uncertainty, can be handled as follows (see Mathur & Pattipati 1995).

The maximum position \mathbf{x}^* of $s(\hat{\boldsymbol{\eta}}(\mathbf{x}), \bar{\mathbf{q}})$ is declared as the 'multiobjective optimum' but, in addition, a region similar to a confidence region is associated with it. This region is obtained via Monte Carlo simulations about the estimated models using the least-squares estimates and their distributions. Since, from (19),

$$(\theta'_1 \theta'_2 \cdots \theta'_M)' - (\hat{\theta}'_1 \hat{\theta}'_2 \cdots \hat{\theta}'_M)' \sim \mathcal{N}(\mathbf{0}, \Sigma \cdot (X'X)^{-1}),$$

the model coefficients $\theta_1, \theta_2, \dots, \theta_M$ may be generated using the normal distribution,

$$\mathcal{N}((\hat{\theta}'_1 \hat{\theta}'_2 \cdots \hat{\theta}'_M)', \hat{\Sigma} \cdot (X'X)^{-1}), \quad (25)$$

which is parameterized by $\hat{\Sigma}$ instead of the (unknown) Σ . Also, since $(N - p)\hat{\Sigma}$ is itself distributed as a Wishart distribution $W(\Sigma, N - p)$ (see, for example, Anderson 1984), one might generate matrices using $W(\hat{\Sigma}, N - p)$ and use these to parameterize the above normal distribution. The multiobjective optima of the models so generated can then be obtained for all the runs and plotted on a scatter plot to visualize the region of distribution of the optima. For cases involving more than two or three input variables, where visualization of such a region is not easy, a Monte Carlo significance test procedure (Barnard 1963) may be used to compute a boundary for this region.

Monte Carlo significance test: Let $\mathbf{x}_i, i = 1, 2, \dots, N_s$, be the positions of the optima obtained from N_s independent Monte Carlo simulations. Let these points be realizations of independent random vectors with distribution $p(\mathbf{x})$ which is unknown. Let $t(\mathbf{x})$ be a test criterion for the hypothesis that the true optimum is distributed according to $p(\mathbf{x})$. We denote these statistics by t_1, t_2, \dots, t_{N_s} for the N_s runs. Now, for a level of significance α ($0 < \alpha < 1$), the region in which the true optimum lies can be obtained as

$$\{\mathbf{x} : t(\mathbf{x}) < t_{(\lfloor N_s(1-\alpha) \rfloor)}\}, \quad (26)$$

where $t_{(j)}$ is the statistic of rank j obtained after sorting $\{t_i\}$ in ascending order, and $\lfloor N_s(1 - \alpha) \rfloor$ is the largest integer less than or equal to $N_s(1 - \alpha)$. We shall refer to the region defined by the set in (26) as the 'significance region.'

The choice of a statistic t depends on the distribution $p(\cdot)$ of the \mathbf{x}_i , which is unknown. The ideal (but intractable) approach would be to obtain a nonparametric estimate of the distribution function based on the N_s measurements \mathbf{x}_i . However, for some applications, it is possible to obtain reasonably good approximate regions based on a statistic that utilizes the sample estimates of the first few higher-order statistics: mean, covariance, (multivariate) skewness, kurtosis, and so on. Let $\bar{\mathbf{x}}$ be the sample mean, and S be the sample covariance matrix. In addition, defining $\mathbf{u}_i = S^{-\frac{1}{2}}(\mathbf{x}_i - \bar{\mathbf{x}})$, where $\mathbf{u}_i = (u_{i1}, u_{i2}, \dots, u_{iK})'$, let $r_{lmn}, l, m, n = 1, 2, \dots, K$, be the lmn th element of the sample third-order cumulant (skewness) tensor:

$$r_{lmn} = \frac{1}{N_s} \sum_{i=1}^{N_s} u_{il} u_{im} u_{in},$$

and let $\kappa_{lmnp}, l, m, n, p = 1, 2, \dots, K$, be the $lmnp$ th element of the sample fourth-order cumulant (kurtosis) tensor:

$$\kappa_{lmnp} = \frac{1}{N_s} \sum_{i=1}^{N_s} u_{il} u_{im} u_{in} u_{ip} - (\delta_{lm} \delta_{np} + \delta_{ln} \delta_{mp} + \delta_{lp} \delta_{mn}),$$

Table 2. Experimental design and measured response data (example).

Run	Design (original)		Design (coded)		Responses			
	x_1 (mM)	x_2 (mM)	x_1	x_2	y_1 (kg)	y_2	y_3	y_4 (mm)
1	8.0	6.5	-1	-1	2.48	0.55	1.95	0.22
2	34.0	6.5	1	-1	0.91	0.52	1.37	0.67
3	8.0	25.9	-1	1	0.71	0.67	1.74	0.57
4	34.0	25.9	1	1	0.41	0.36	1.20	0.69
5	2.6	16.2	-1.414	0	2.28	0.59	1.75	0.33
6	39.4	16.2	1.414	0	0.35	0.31	1.13	0.67
7	21.0	2.5	0	-1.414	2.14	0.54	1.68	0.42
8	21.0	29.9	0	1.414	0.78	0.51	1.51	0.57
9	21.0	16.2	0	0	1.50	0.66	1.80	0.44
10	21.0	16.2	0	0	1.66	0.66	1.79	0.50
11	21.0	16.2	0	0	1.48	0.66	1.79	0.50
12	21.0	16.2	0	0	1.41	0.66	1.77	0.43
13	21.0	16.2	0	0	1.58	0.66	1.73	0.47

mM = millimolar

where $\delta_{ij} = 1$, if $i = j$; otherwise, it is zero. Higher-order cumulants can similarly be computed. If the skewness, kurtosis, and higher-order cumulants are negligible, it may suffice to use the statistic

$$t_1(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}})' S^{-1} (\mathbf{x} - \bar{\mathbf{x}}) = \mathbf{u}' \mathbf{u}. \quad (27)$$

In general, a statistic based on the likelihood function involving the higher-order terms in an Edgeworth expansion (see, for example, Kolassa 1994) could be used:

$$t_2(\mathbf{x}) = -\phi(\mathbf{u}) \left(1 + \frac{1}{6} \sum_{l,m,n=1}^K r_{lmn} H_{lmn}(\mathbf{u}) + \frac{1}{24} \sum_{l,m,n,p=1}^K \kappa_{lmnp} H_{lmnp}(\mathbf{u}) + \dots \right), \quad (28)$$

where $\phi(\mathbf{u})$ is the K -variate standard Gaussian density function (for mean $\mathbf{0}$ and covariance matrix \mathbf{I}). $H_{lmn}(\mathbf{u})$ and $H_{lmnp}(\mathbf{u})$ are, respectively, the third-order and fourth-order K -dimensional Hermite polynomials defined as

$$H_{lmn}(\mathbf{u}) = \frac{(-1)^3}{\phi(\mathbf{u})} \frac{\partial^3 \phi(\mathbf{u})}{\partial u_l \partial u_m \partial u_n} = u_l u_m u_n - (\delta_{lm} u_n + \delta_{mn} u_l + \delta_{nl} u_m),$$

and

$$H_{lmnp}(\mathbf{u}) = \frac{(-1)^4}{\phi(\mathbf{u})} \frac{\partial^4 \phi(\mathbf{u})}{\partial u_l \partial u_m \partial u_n \partial u_p} = u_l u_m u_n u_p - (\delta_{np} u_l u_m + \dots + \delta_{mp} u_l u_n) + (\delta_{lm} \delta_{np} + \delta_{ln} \delta_{mp} + \delta_{lp} \delta_{mn}).$$

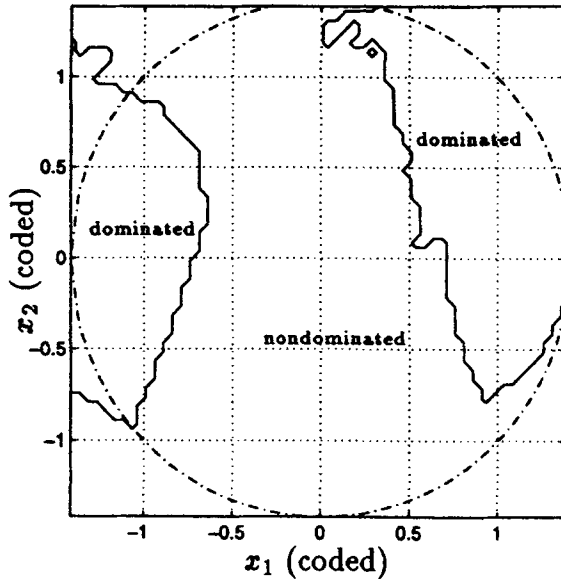


Figure 9. Regions corresponding to non-dominated solutions and dominated solutions with respect to all four responses based on the estimated models (example).

The higher-order polynomials have similar extensions.

To summarize the above, the procedure involves the following steps:

- (1) Using the experimental data for the responses, obtain regression models and estimates of the covariance matrices of the model coefficients using (15) and (18);
- (2) Compute the achievement function (24) over the region Ξ using the predicted responses and the reference point input by the user; obtain the multiobjective optimum settings by maximizing the achievement function over the region Ξ ;
- (3) Obtain several independent sets of models by Monte Carlo simulations using the least-squares estimates of the statistics in the generating distribution (25); for each realization, repeat step 2 to obtain a sample of points $\{x_i\}$;
- (4) Apply the statistic in (28) (or (27) if adequate) to the sample optima $\{x_i\}$ of step 3 to obtain a region of the form (26).

The feasible region Ξ in the factor space is recommended to be (two or three times) larger than the experimental design region in order to ensure that the sample points obtained

Table 3. The second degree regression models (example).

Model term	Regression coefficients			
	y_1	y_2	y_3	y_4
int.	1.526	0.660	1.776	0.468
x_1	-0.575	-0.092	-0.250	0.131
x_2	-0.524	-0.010	-0.078	0.073
x_1^2	-0.171	-0.096	-0.156	0.026
x_2^2	-0.098	-0.058	-0.079	0.024
x_1x_2	0.318	-0.070	0.010	-0.083
R^2	0.95	0.98	0.98	0.95

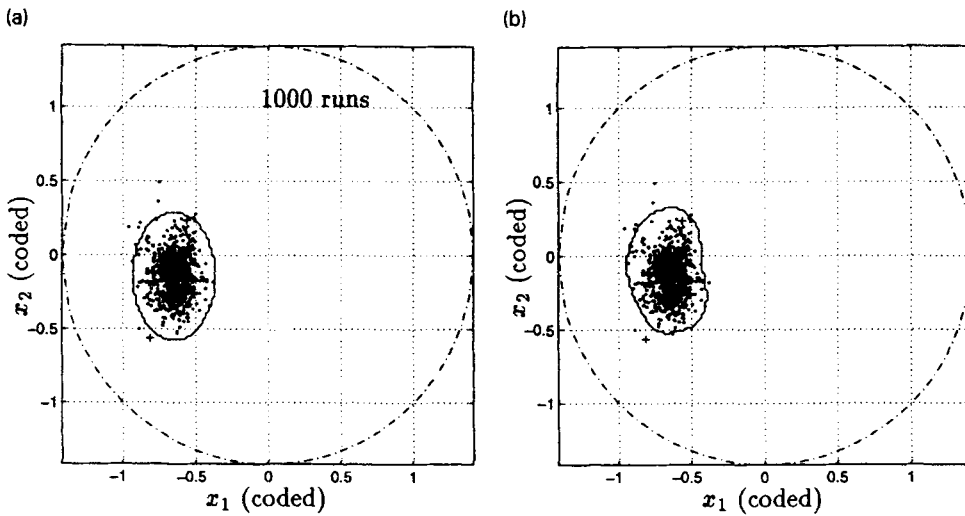


Figure 10. Distribution of the dual-objective optimum (for y_2 and y_3) for the example; the '+'s show the positions of the optima of the individual (estimated) responses. The region of significance is obtained using the statistic t_1 in (a) and t_2 (with the first four orders of moments) in (b).

via simulations lie inside Ξ (see Mathur & Pattipati 1995); for if a large number of such points lie on the boundary of Ξ , the estimation of the moments (and cumulants) used in the statistic t could incur large errors. In the final analysis, all inferences can be confined to the experimental region. The region of significance should be interpreted as follows: if it includes the design centre (current settings), there is not enough evidence to suggest a change in the factor settings; if it does not, but is large compared to the size of the experimental region, or if it lies outside the experimental region, further experimentation is needed.

4.1 Example

The following example was originally discussed by Khuri & Conlon (1981) (also see Mathur & Pattipati 1996). In this case study of a dialyzed whey-protein-concentrate (WPC) gel system, the effects of two inputs: concentrations of cysteine (x_1) and calcium chloride (x_2), on four responses measuring the textural and water-holding characteristics of the WPC gel were studied. The four texture characteristics are: hardness (y_1), cohesiveness (y_2), springiness (y_3), and compressible water (y_4). The goal of the problem is to maximize the measures of all these characteristics. A 13-point rotatable central-composite-design experiment (Montgomery 1991b) with five centre points (for uniform precision) was conducted to vary the input variables. The experiment design and the measured values of the two responses are recorded in table 2, reproduced from Khuri & Conlon (1981). The second-order regression parameters (corresponding to the coded design) for the two responses are tabulated in table 3.

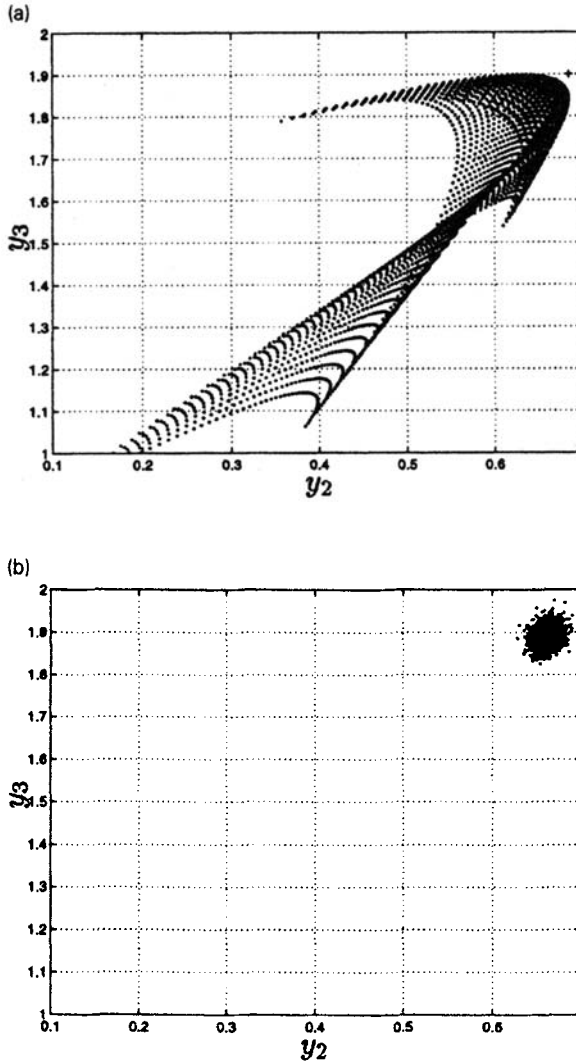


Figure 11. (Example, two-response case) (a). Attainable objective pairs (of y_2 and y_3) based on estimated models. The '+' marks the chosen reference point. (b). Scatter of the dual-objective optima obtained from 1000 Monte Carlo simulations of the models.

Only y_2 and y_3 considered: Considering all four responses, it turns out that the current operating point $\mathbf{x} = (0, 0)$ is efficient (nondominated). That is, no further improvement in any response is possible without trade-off in at least one other response. In fact, a large area of the region $x_1^2 + x_2^2 \leq 2$ maps to the nondominated set in the objective space (see figure 9; this efficient set is based on the predicted responses). So, for illustrative purposes, we first consider only two of the four responses: y_2 and y_3 . The choice of ignoring responses y_1 and y_4 in particular was made because y_1 and y_4 have no global maxima (based on their least-squares models); their stationary points are saddle points which lie in opposite quadrants of the variable space, far from the design region.

For this case, the efficient set (based on the response models) in the factor space is a curve connecting the two individual optima. We maximize the achievement function in (24) with α_j chosen as $1/\hat{\sigma}_j$, $\beta = 0.01$, and the estimated ideal point $\hat{\phi} = [0.68 \ 1.9]'$ as the reference point. While this choice of α_j is arbitrary, and while other choices are possible, it normalizes

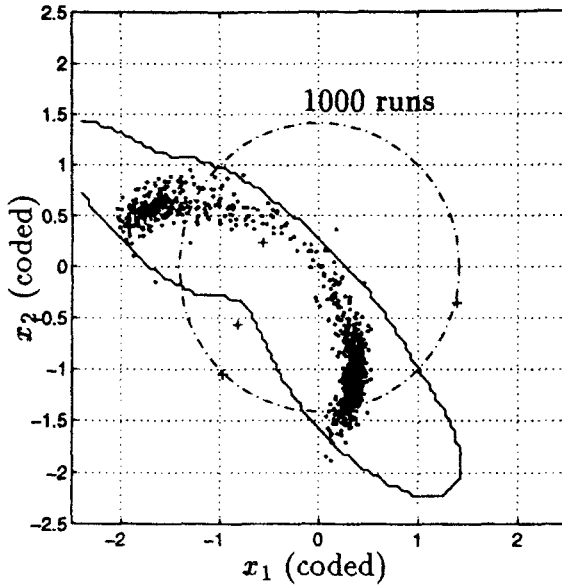


Figure 12. Distribution of the multiobjective optimum (for all four responses) for the example; the '+'s show the positions of the optima of the individual (estimated) responses. The region of significance is obtained using the statistic t_2 with up to sixth-order moments.

the achievement function against differences in the noise powers in the two responses, and also happens approximately to normalize for the different absolute scales. The estimates of σ_2^2 and σ_3^2 are 0.0005 and 0.0025 respectively. For maximization of s , a Simplex search method was used (Nelder & Mead 1965). The model optimum and its 'significance' regions obtained using the statistics of (27) and (28) (using up to fourth-order terms) are shown in figure 10. The multiobjective optimum is obtained at $(x_1^*, x_2^*) = (-0.61, -0.16)$, and the predicted response vector at this point is $(y_2, y_3) = (0.67, 1.88)$. At the individual model optima, the estimated response vectors are $(y_2, y_3) = (0.68, 1.84)$ and $(y_2, y_3) = (0.63, 1.9)$. The spread of the optimum, obtained from 1000 Monte Carlo simulations of the models, indicates the region in which the true optimum could lie assuming the quadratic model assumptions to be correct. The region enclosed by the solid line is that obtained from the Monte Carlo significance test for level 0.01; that is, the ten most extreme values were excluded from the region. For this example, the statistic in (27) seems to be adequate. The statistic t_2 is a little better in following the shape of the scatter of optima as it accounts for the sample skewness and kurtosis. In any case, these regions indicate that both responses y_2 and y_3 can be improved by lowering the variable x_1 to a (scaled) value of approximately -0.6 , while not altering the variable x_2 from its current value. Since we are considering only two responses at this point, it is possible to view the attainable set Ω in the objectives space, shown in figure 11a. The scatter of the multiobjective optima is shown in figure 11b.

All four responses considered: We now apply our method to the multiobjective optimization of all four responses in this example. We again choose the reference point to be the ideal point $[2.69, 0.68, 1.9, 0.72]'$. The α_j 's are again chosen to be $1/\hat{\sigma}_j$'s, and $\beta = 0.01$ ($\hat{\sigma}_1^2 = 0.0399$ and $\hat{\sigma}_4^2 = 0.0017$). The multiobjective optimum for these parameters was found at $(0.37, -1.21)$. The scatter of the multiobjective optima obtained from the Monte Carlo simulations of the models is shown in figure 12. The shape of the region obtained

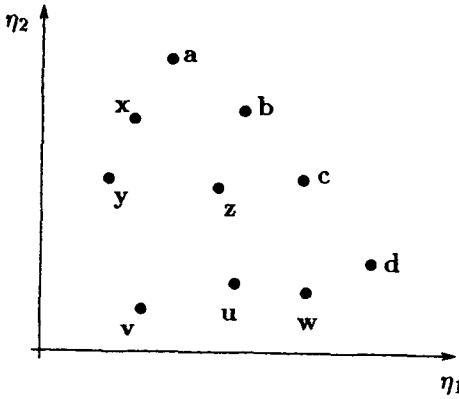


Figure 13. A two-dimensional discrete objectives set.

from the Monte Carlo significance test (for level 0.01) using the statistic t_2 with up to sixth-order terms is also shown by the solid-line curves. The scatter reveals the extent to which uncertainty in the four models can affect the position of the (multiobjective) optimum factor settings: the best factor settings could lie anywhere in a band that stretches across the experimental design region from the vicinity of $(-1, 0.5)$ to the vicinity of $(0, -1.4)$. While maximizing the achievement function in each Monte Carlo run, we have deliberately not constrained the maximum to lie inside the circle $x_1^2 + x_2^2 \leq 2$ (see the remarks following the introduction of our method). If the scatter of the global maximum of the achievement function lies mostly outside the experimental region, there is not much sense in fitting a region of significance, as a better interpretation might be to conduct more experiments in a new region outside the current design region. For fitting a region about the scatter of points in this example, the shape of the scatter demands that higher-order terms beyond the third- and fourth-order terms be included in the Edgeworth-series statistic t_2 . Figure 12 shows the region obtained by including up to sixth-order terms. Since the current operating point (the design centre) also lies inside the significance region, there is not a strong case for changing the current settings for multiobjective improvement of all four responses.

5. Multiobjective optimization w.r.t. discrete factors

When quantitative factors are constrained to take only a finite number of levels, or when the factors are qualitative (or categorical), the multiobjective optimization problem reduces to a combinatorial optimization problem. Here, the assumption is that the k th factor ($k = 1, \dots, K$) can only take one of L_k levels: $x_k \in \{l_{k,1}, \dots, l_{k,L_k}\}$; that is, the feasible set Ξ is finite (with cardinality $\prod_{k=1}^K L_k$), and so is the set Ω of feasible objectives. Using experimentation, a prediction model is obtained at all factor-level combinations. The solution of the discrete (multiobjective) optimization problem (DMOP) requires an exhaustive search in a space of cardinality $\prod_{k=1}^K L_k$ (which we will refer to as the size of the DMOP). However, under the frequently occurring conditions of separability of factor effects, the problem can be decomposed into problems involving much smaller search spaces.

Let the mean η_j of the j th response ($j = 1, \dots, M$) be related to the K factors by a relationship of the form

$$\eta_j(\mathbf{x}) = g_1^{(j)}(x_1) + g_2^{(j)}(x_2) + \cdots + g_K^{(j)}(x_K).$$

Then, it is clear that the set of x_k 's, $k = 1, \dots, K$, which optimize the respective $g_k^{(j)}$ would also optimize η_j . This approach can be extended in a straightforward way to the multiobjective optimization of all M responses, as shown by the following theorem (Song *et al* 1995).

Theorem 1. *Let the sets $B_k \subset \mathcal{R}^M$, $k = 1, \dots, K$ and the set C , defined as*

$$C = \{\boldsymbol{\eta} \mid \boldsymbol{\eta} = \mathbf{g}_1 + \mathbf{g}_2 + \cdots + \mathbf{g}_K, \mathbf{g}_k \in B_k\},$$

have the same domination cone. Then, if $\check{\mathbf{g}}_k$ is dominated in B_k for at least one k , $k = 1, \dots, K$, the point $\check{\boldsymbol{\eta}} = \check{\mathbf{g}}_1 + \cdots + \check{\mathbf{g}}_K$ is dominated in C .

Proof 1. Assume that D is the domination cone, and $\check{\boldsymbol{\eta}} = \check{\mathbf{g}}_1 + \cdots + \check{\mathbf{g}}_K$ is a nondominated point in C . Then

$$(\check{\boldsymbol{\eta}} + D) \cap C = \emptyset. \quad (29)$$

Suppose $\check{\mathbf{g}}_k$ is dominated for some k . Then there exists a $\tilde{\mathbf{g}}_k \in B_k$ such that $\tilde{\mathbf{g}}_k \in \check{\mathbf{g}}_k + D$. Since $\check{\boldsymbol{\eta}} = \check{\mathbf{g}}_1 + \cdots + \check{\mathbf{g}}_K$, we have

$$\tilde{\boldsymbol{\eta}} = \check{\mathbf{g}}_1 + \cdots + \tilde{\mathbf{g}}_k + \cdots + \check{\mathbf{g}}_K \in \check{\boldsymbol{\eta}} + D. \quad (30)$$

Thus $\tilde{\boldsymbol{\eta}} \in C$ and $\tilde{\boldsymbol{\eta}} \in \check{\boldsymbol{\eta}} + D$, which contradicts the assumption that $\check{\boldsymbol{\eta}}$ is nondominated.

The applicability of the above theorem to the discrete factors case can greatly reduce the complexity of the optimization problem, for it enables decomposition of the multiobjective optimization problem into two stages: the first stage consists of K multiobjective optimization problems of sizes L_1, L_2, \dots, L_K : (multiobjective) optimize $\bar{\mathbf{g}}_k$ with respect to x_k . The nondominated sets B_k obtained from the first stage can be used to construct the search space C for the second stage whose cardinality can be much smaller than $\prod_{k=1}^K L_k$. For example, if an experiment involves six three-level factors, the search space contains $3^6 = 729$ combinations. If the low-level (first-stage) optimization problems (each of size 3) reduce the size of the nondominated sets even by one for each factor, the size of the search space for the second stage would be reduced to $2^6 = 64$. Thus, the cost of the high-level optimization would be reduced ten-fold. If the effects of some factors do interact, the decomposition can still be done with respect to those factors that do not interact, and some reduction in complexity attained.

Search for the nondominated set: An explicit enumeration technique called the *technique of dominate approximations* (TDA) (Majchrzak 1989) can be used to obtain the nondominated set from a discrete space Ω . The technique can be illustrated using figure 13. In iteration one, the method maximizes the performance measure η_1 to obtain the point \mathbf{d} , and then rejects all points dominated by \mathbf{d} to generate a dominated approximation $\Omega_1 = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{x}, \mathbf{y}, \mathbf{z}\}$; in the second iteration, the method maximizes component η_2 (excluding \mathbf{d}) to obtain \mathbf{a} , and then rejects all points dominated by \mathbf{a} to generate another

Table 4. Orthogonal design (L_{18}) and summarized data.

	Factors							Responses			
	1	2	3	4	5	6	7	L	W	F	S
1	1	1	1	1	1	1	1	-8.6	-12.2	-2.2	1.0
2	1	1	2	2	2	2	2	-10.3	-9.2	-5.7	1.0
3	1	1	3	3	3	3	3	-15.9	-14.0	-8.5	1.33
4	1	2	1	1	2	2	3	-2.2	-7.2	-1.1	1.0
5	1	2	2	2	3	3	1	-13.0	-13.6	-7.4	1.0
6	1	2	3	3	1	1	2	-10.2	-17.5	-14.7	1.33
7	1	3	1	2	1	3	2	-13.3	-16.3	-7.7	1.5
8	1	3	2	3	2	1	3	-14.3	-16.3	-10.4	1.0
9	1	3	3	1	3	2	1	-16.9	-18.2	-14.2	2.33
10	2	1	1	3	3	2	2	-12.0	-11.7	-2.5	1.0
11	2	1	2	1	1	3	3	-15.8	-18.2	-11.1	1.0
12	2	1	3	2	2	1	1	-17.1	-8.4	-12.2	1.67
13	2	2	1	2	3	1	3	-18.1	-13.7	-9.7	3.0
14	2	2	2	3	1	2	1	-5.6	-15.1	-7.8	1.33
15	2	2	3	1	2	3	2	-16.1	-15.1	-11.8	3.0
16	2	3	1	3	2	3	1	-11.4	-16.7	-12.5	1.83
17	2	3	2	1	3	1	2	-12.1	-16.3	-11.0	1.33
18	2	3	3	2	1	2	3	-6.3	-20.9	-12.5	2.0

Factors 1: Injection pressure; 2: Injection speed; 3: Mould temperature; 4: Melt temperature; 5: Holding pressure; 6: Cool time; 7: Hold time

Responses L : Length SNR; W : Width SNR; F : Flatness SNR; S : Surface quality

dominated approximation $\Omega_2 = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{z}\}$. The process continues until all dominated points are eliminated.

The above arguments are given with respect to perfectly known metric functions $\eta(\mathbf{x})$. When experimentation is used, prediction models for these relationships are obtained and used in place of the true relationships. To identify the significant factor-effects, an analysis of variance (ANOVA) is used. With the use of orthogonal arrays, an analysis of means may also be used for easy computation of factor effects. The problem of dealing with uncertainty in the models in the discrete-factors case would require a different formulation from that for the continuous-factors case of the previous section. In the discrete-factors case, there would obviously be no region of significance about the estimated multiobjective optimum design, but rather a set of probable designs with an associated discrete probability distribution. This problem will be addressed in future research.

5.1 Example

An experiment was conducted (Greenall 1989), for the optimization of a manufacturing process for injection-moulded plastic housings. The experiment involved seven factors: injection pressure, injection speed, mould temperature, melt temperature, holding pressure, cool time, and hold time; the first factor was studied at two levels, and the rest of the factors at three levels each. Four response variables were considered: overall housing length (L), overall housing width (W), flatness (F), and surface quality (S). An $L_{18}(2^1 3^7)$ orthogonal array was used. Table 4 shows the orthogonal design and the summarized data for the four responses; the first three responses have been summarized into signal-to-noise ratios,

Table 5. Results of low-level optimization.

Factors	Levels	Mean effects			
		<i>L</i>	<i>W</i>	<i>F</i>	<i>S</i>
Injection pressure	1	-11.63	-13.83	-7.99	1.28
	2	-12.7	-15.1	-10.1	1.80
Injection speed	1	-13.28	-12.28	-7.03	1.17
	2	-10.87	-13.70	-8.75	1.78
Mould temperature	1	-10.93	-12.97	-5.95	1.56
	3	-13.75	-15.68	-12.32	1.94
Melt temperature	1	-11.95	-14.53	-8.57	1.61
	2	-13.02	-13.68	-9.20	1.70
	3	-11.57	-15.22	-9.40	1.30
Hold pressure	1	-9.97	-16.70	-9.33	1.36
	2	-11.90	-12.15	-8.95	1.58
	3	-14.67	-14.58	-8.88	1.67
Cool time	1	-13.40	-14.07	-10.03	1.56
	2	-8.88	-13.72	-7.30	1.44
	3	-14.25	-15.65	-9.83	1.61
Hold time	1	-12.10	-14.03	-9.38	1.53
	2	-12.33	-14.35	-8.90	1.53
	3	-12.10	-15.05	-8.88	1.56

while surface quality is an average of three independent assessments on a scale of 1 to 3 (see Greenall 1989, for the complete data and analysis).

The ANOVA carried out by Greenall (1989) on the signal-to-noise ratios showed that the additivity assumptions are fairly well satisfied; so, the two-step approach presented above can be applied.

The mean responses can be computed using an analysis of means to obtain a prediction equation of the form (11). From table 4, it can be seen that a total of $2 \times 3^6 = 1458$ combinations (and, therefore, 1458 performance predictions) of different factor-levels are possible. A low-level optimization eliminates dominated points for each factor based on the mean effects (the results are shown in table 5). From table 5, it can be seen that the number of combinations from the remaining points is $2^3 \times 3^4 = 648$. The high-level optimization performed on the set of 648 reduces combinations, based on the predictions of performance measures using (11), resulted in 135 nondominated solutions. The final choice of one operating condition should be made by the design engineer. In practice, the design engineer can use the reference point approach and his/her preferences to find the most preferred solution from this nondominated set in an interactive manner. Let \mathbf{r} be a reference point specified by the user (decision-maker). A simple strategy would be to find the closest nondominated point \mathbf{q} , from the already determined set of nondominated solutions (denoted $P_N \subset \Omega$), to \mathbf{r} such that $\mathbf{q} = \arg\{\min|\mathbf{q} - \mathbf{r}|, \mathbf{q} \in P_N\}$.

For illustrative purposes, each of the seven recommended designs obtained (Greenall 1989) was chosen as a reference point and the nondominated solution closest to it obtained. The results are shown in table 6. ref_i ($i = 1, \dots, 7$) in table 6 denotes the i th solution in Greenall (1989), and sol_i denotes the Pareto-optimal solution found by our algorithm

Table 6. Final results of optimization.

	Predicted performance measures				Levels
	<i>L</i>	<i>W</i>	<i>F</i>	<i>S</i>	
ref ₁	-8.91	-6.12	-0.99	1.03	1 1 1 2 2 2 2
sol ₁	-8.91	-6.12	-0.99	1.03	1 1 1 2 2 2 2
ref ₂	-11.7	-8.83	-7.36	1.42	1 1 3 2 2 2 2*
sol ₂	-10.8	-7.57	-5.92	1.75	1 2 1 2 2 1 1
ref ₃	-10.0	-7.40	-3.12	1.55	2 1 1 2 2 2 2
sol ₃	-10.0	-7.40	-3.12	1.55	2 1 1 2 2 2 2
ref ₄	-12.8	-10.1	-9.49	1.94	2 1 3 2 2 2 2*
sol ₄	-11.8	-8.86	-8.06	2.27	2 2 1 2 2 1 1
ref ₅	-9.83	-7.93	-3.94	0.59	1 1 2 2 2 2 2*
sol ₅	-9.77	-7.08	-3.61	1.54	2 1 1 2 2 2 1
ref ₆	-10.9	-9.22	-6.07	1.10	2 1 2 2 2 2 2*
sol ₆	-10.8	-7.57	-5.92	1.75	1 2 1 2 2 1 1
ref ₇	-10.4	-11.5	-11.2	2.55	2 2 3 2 2 2 2
sol ₇	-10.4	-11.5	-11.2	2.55	2 2 3 2 2 2 2

*Dominated solutions in Greenall (1989).

using ref_{*i*} as the reference point. It can be seen that four of the original solutions, marked by *, are dominated by solutions found by our method. In other words, four of the original solutions in Greenall (1989) are not Pareto-optimal.

It is clear that an *ad hoc* approach, such as that used by Greenall (1989), can run into difficulties, if the search space of the problem is very large. The two-step approach, on the other hand, successfully finds all 135 nondominated solutions and provides a simple method for the engineer to specify preferences. The reference point itself may or may not be attainable; one can enter *any point* that reflects one's preferences, and the algorithm would always find the 'closest' nondominated point. In the case when little knowledge or preference information is available for the problem (as would happen during the initial stage of design and optimization), one can use the *ideal point* as the reference point.

6. Summary

This article has reviewed the use of experimentation to determine the best operating points for a manufacturing process, or the best design for a product's parameters, so as to optimize one or more quality criteria. After illustrating the steps in robust design used to lower variability, it specifically examined the optimization problems arising when several models for quality characteristics are estimated from experimental data. In the case of continuous factors, a new approach was discussed for dealing with the uncertainty associated with the use of response surface models for the quality metrics. This approach prevents faulty inferences from the optimization step and gives the designer or process engineer a means for determining whether to conduct further experiments or to accept the optimization results. In the case of discrete factors, an efficient search technique for the multiobjective optimal

design was presented. The problem of developing a scheme for dealing with uncertainty in the discrete-factors case is a potential subject for future research.

References

- Anderson T W 1984 *An introduction to multivariate statistical analysis* (New York: John Wiley & Sons)
- Barnard G A 1963 Discussion of Bartlett's paper: The spectral analysis of point processes. *J. R. Stat. Soc. B25*: 294–296
- Bendell A, Disney J, Pridmore W A (eds) 1989 *Taguchi methods: Applications in world industry* (Kempston, Bedford: IFS; Berlin, New York: Springer Verlag)
- Box G E P 1988 Signal-to-noise ratios, performance criteria, and transformations. *Technometrics* 30: 1–17
- Box G E P, Draper N R 1987 *Empirical model-building and response surfaces* (New York: John Wiley & Sons)
- Box G E P, Hunter W G, Hunter J S 1978 *Statistics for experimenters, An introduction to design, data analysis, and model building* (New York: John Wiley & Sons)
- Boza L A, Ciaccia T J, Gatenby D A, Muise R W, Ng K K, Yanizeski G M 1994 Achieving robust design through customer satisfaction. *AT&T Tech. J.* 73: 48–58
- Chankong V, Haimes Y 1983 *Multiobjective decision making: Theory and methodology* (New York: North Holland)
- Greenall R 1989 A Taguchi optimization of the manufacturing process for an injection molded housing. In *Taguchi methods: Applications in World Industry* (eds) A Bendell, J Disney, W A Pridmore (Kempston, Bedford: IFS; Berlin, New York: Springer Verlag)
- Harry M J 1988 *The nature of six sigma quality* (Schaumburg, Illinois: Motorola)
- Khuri A I, Conlon M 1981 Simultaneous optimization of multiple responses represented by polynomial regression functions. *Technometrics* 23: 363–375
- Kolassa J E 1994 *Series approximation methods in statistics* (New York: Springer-Verlag)
- León R V, Shoemaker A C, Kacker R N 1987 Performance measures independent of adjustment. *Technometrics* 29: 253–265
- Lewandowski A, Kreglewski T, Rogowski T, Wierzbicki A P 1989 Decision support systems of DIDAS family (Dynamic Interactive Decision Analysis & Support). In *Aspiration based decision support systems: Theory, software, and applications* (eds) A Lewandowski, A P Wierzbicki (New York: Springer Verlag)
- Majchrzak J 1989 A methodological guide to the decision support system DISCRET for discrete alternatives problems. In *Aspiration based decision support systems* (eds) A Lewandowski, A P Wierzbicki (New York: Springer Verlag)
- Mathur A, Pattipati K R 1995 Multiobjective optimization using regression models. *J. Quality Technol.* (submitted)
- Mathur A, Pattipati K R 1996 Multiobjective optimization of process parameters using regression models. *Proceedings of CIMAT '96, International Conference on Computer Integrated Manufacturing and Automation Technology*, Grenoble, France
- Montgomery D C 1991a *Introduction to statistical quality control* (New York: John Wiley & Sons)
- Montgomery D C 1991b *Design and analysis of experiments* (New York: John Wiley & Sons)
- Nair V N 1992 Taguchi's parameter design: A panel discussion. *Technometrics* 34: 127–161
- Nelder J A, Mead R 1965 A simplex method for function minimization. *Comput. J.* 7: 308–313
- Phadke M D 1989 *Quality engineering using robust design* (Englewood Cliffs, NJ: Prentice Hall)

- Song A, Mathur A, Pattipati KR 1995 Design of process parameters using robust design techniques and multiple criteria optimization. *IEEE Trans. Syst. Man, Cybern.* 25: 1437–1446
- Steuer R E 1989 *Multiple criteria optimization: Theory, computation and applications* (Malabar, FL: Robert E Krieger)
- Taguchi G 1987 *System of experimental design* (White Plains, NY: Kraus International) vols. 1 & 2
- Wierzbicki A P 1980 The use of reference objectives in multiobjective optimization. In: *Multiple criteria decision making, theory and applications* (eds) Fandel, Gal (New York: Springer Verlag)
- Zeleny M 1982 *Multiple criteria decision making* (New York: McGraw Hill)