Center for Quality and Productivity Improvement
UNIVERSITY OF WISCONSIN
610 Walnut Street
Madison, Wisconsin 53705

(608) 263–2520
(608) 263–1425 FAX

Report No. 63

Robust Product Designs, Part II: Second-Order Models

George Box and Stephen Jones

May 1990

This work was sponsored by the National Science Foundation Grant DDM–8808138 and by the Vilas Trust of University of Wisconsin–Madison.

Robust Product Designs, Part II: Second-Order Models

George Box

Stephen Jones*

Center for Quality and
Productivity Improvement

University of Wisconsin
Madison, Wisconsin

ABSTRACT

Professor Genichi Taguchi has emphasized the use of designed experiments in several novel and important applications. In this paper we focus on the use of statistical experimental designs in designing products to be robust to environmental conditions. The engineering concept of robust product design is very important since it is frequently impossible or prohibitively expensive to control or eliminate sources of variation due to environmental conditions. Robust product design enables the experimenter to discover how to modify the design of the product to minimize the effect due to variation from environmental sources.

In experiments of this kind, Professor Taguchi's total experimental arrangement consists of a cross-product of two experimental designs, an inner array containing the design factors and an outer array containing the environmental factors. Except in situations where both of these arrays are small, this arrangement may involve a prohibitively large amount of experimental work.

In the previous paper (Box and Jones (1990a)) we showed how this amount of work could be reduced. In that paper we developed a robustness measure for this particular experimental situation. Application of this robustness measure led to classes of experimental designs that generally required significantly fewer runs than the designs proposed by Professor Taguchi.

In the present paper we apply the strategy developed in Box and Jones (1990a) to a general second-order model. For this model we derive the robustness measure and give a series of tables of designs that are appropriate for the possible objectives of the experimenter.

KEYWORDS: *Parameter design, robustness, experimental design, response surface methodology, Taguchi*

This work was sponsored by National Science Foundation Grant DDM-8808138 and by the Vilas Trust of the University of Wisconsin–Madison.

* Stephen Jones is now a statistician at Boeing Computer Services, The Boeing Company, P.O. Box 24346, MS 7L–22, Seattle, Washington, 98124.

 This report is part of a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Statistics) at the University of Wisconsin–Madison (1990). Thesis adviser: George Box.

Table of Contents

Second-Order Models

Robust Product Designs, Part II: Second-Order Models

George Box and Stephen Jones

1. Introduction

Much attention has been focused in recent years on the impact of the use of statistics and, in particular, experimental design on the quality of Japanese products and the consequent competitiveness of Japanese industry in the world market-place. A leading quality consultant in that country, who has advocated the widespread use of design of experiments, is Genichi Taguchi. Professor Taguchi has emphasized the use of designed experiments in several novel and important applications. One such application is that of designing a product that is robust to environmental variation.

As an example of such an experiment, consider the set of data given in Table 1, which is typical of tests that have been run in the United States food industry for many years. The manufacturer is seeking the best recipe for a cake mix composed of several ingredients to be sold in a box containing simple instructions. The quality of the cake made by the purchaser of the cake mix will depend not only on the ingredients but also on the extent to which the baking instructions printed on the box are followed.

Table 1. Cake Mix Data

	DESIGN FACTORS			ENVIRONMENTAL FACTORS				
Recipe				T: 0	−	+	−	+
(*k*)	F	S	E	t: 0	−	−	+	+
(0)	0	0	0	6.7	3.4	5.4	4.1	3.8
(1)	−	−	−	3.1	1.1	5.7	6.4	1.3
(2)	+	−	−	3.2	3.8	4.6	4.3	2.1
(3)	−	+	−	5.3	3.7	5.1	6.7	2.9
(4)	+	+	−	4.1	4.5	6.4	5.8	5.2
(5)	−	−	+	5.9	4.2	6.8	6.5	3.5
(6)	+	−	+	6.9	5.0	6.0	5.9	5.7
(7)	−	+	+	3.0	3.1	6.3	6.4	3.0
(8)	+	+	+	4.5	3.9	5.5	5.0	5.4

In this example, experiments are to be run with three design factors, flour (F), shortening (S), and egg powder (E), in a 2^3 factorial design, with a center point, indicated by (0, 0, 0). It is known that the temperature indicator for an oven in a typical domestic stove may be considerably biased up or down. Furthermore, in practice people frequently overcook or undercook a cake. The manufacturer wants a robust recipe so that the cake will taste reasonably good even when the time (t) and temperature (T) of baking differ somewhat from the recommended levels given in the instructions on the box, and indicated here by (0, 0). To supply data on this, for each recipe the temperature and time of baking are varied about the standard conditions, in a 2^2 factorial. The cakes are then baked corresponding to all combinations of these conditions. In this case the ingredients are design variables and the baking conditions given in the instructions printed on the box represent environmental variables. Table 1 gives a typical set of average scores of a taste panel rating the quality of the cakes on a scale of 1 through 7.

Taguchi, in conducting experiments to investigate the robustness of a product to variation from environmental factors, advocates the use of the cross-product of two orthogonal arrays that he calls inner and outer arrays (see Taguchi, 1986; Taguchi and Phadke, 1984; Taguchi and Wu, 1980; Kackar, 1985). The inner array is composed of the factor levels of the variables associated with the design of the product. For each experimental point of the inner array there is an outer array composed of the different levels of the environmental variables. We, however, will use the expressions *design array* instead of inner array and *environmental array* instead of outer array, and refer to the total arrangement as a *cross-product array*, since it is a cross-product of the design array and the environmental array. In general, if there are n_1 runs in the design array and n_2 runs in the environmental array, and the runs are made independently, then the cross-product array will require $n_1 \times n_2$ runs for the total experiment. Thus, except where n_1 and n_2 are both small, this would involve a large amount of experimental work.

An alternative approach to robust product experiments that sought to reduce the amount of experimental work required was described in an earlier paper (Box and Jones (1990a)). In that paper we first considered the possible objectives for an experiment of this kind. From these objectives we developed performance measures that the investigator might seek to optimize. One of the performance measures that we developed is an overall measure of environmental robustness. The approach is to then entertain a class of models which we suppose, tentatively at least, approximates the underlying system over the (product design $\times$ environmental) factor space of interest. We then determine the coefficients of the model that we need to estimate so as to be able to optimize the various performance measures of interest. Having determined the coefficients that it is necessary to estimate, we are then able to choose experimental designs that will yield data which, when the model is true, enable these coefficients to be estimated and so make it possible to optimize the desired performance measures.

In Box and Jones (1990b) we applied this strategy under the assumption that the response of interest, y_{xz}, could be adequately represented by a first-order model in the design and the environmental factors with terms for the interactions between the design and the environmental factors. This model would be too simple for many experimental situations. It might be more reasonable to expect that curved response relationships could occur both for product design and environmental variables and that design variables could interact with other design variables, environmental variables could interact with other environmental variables, and that design and environmental variables could interact with each other. In this paper we will suppose, therefore, that over the region of interest the behavior of the design and environmental variables $(\underline{x}, \underline{z})$ can be represented by a general second-order model. If there are n product design variables and m environmental variables this can be written as

$$\begin{aligned} y_{xz} = \beta_0 &+ \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \beta_{ii} x_i^2 + \sum_{i=1}^{n-1} \sum_{k=i+1}^{n} \beta_{ik} x_i x_k \\ &+ \sum_{j=1}^{m} \gamma_j z_j + \sum_{j=1}^{m} \gamma_{jj} z_j^2 + \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} \gamma_{jk} z_j z_k \\ &+ \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ji} x_i z_j , \end{aligned} \tag{1}$$

or, in matrix notation

$$y_{xz} = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'\mathbf{B}\underline{x} + \underline{z}'\underline{\gamma} + \underline{z}'\mathbf{C}\underline{z} + \underline{z}'\mathbf{D}\underline{x} , \tag{2}$$

where $\underline{\beta}$ is $n\times 1$, $\underline{\gamma}$ is $m\times 1$, $\mathbf{B}' = \mathbf{B}$ is $n\times n$, $\mathbf{C}' = \mathbf{C}$ is $m\times m$, and $\mathbf{D}$ is $m\times n$.

In section 2 we derive the Estimation Table appropriate for the full second-order model given in equation (1). The Estimation Table indicates the constants that need to be estimated to meet the possible objectives of interest of the investigator. The potential objectives that we advocated in Box and Jones (1990a) were: (a) to minimize $M(x)$, a measure of how far the mean response over the environmental variables is from the ideal value, (b) to minimize $V(x)$, a measure of the variability of the response over the environmental variables, (c) to derive the locus of minimum $R(x)$ for different values of λ, where $R(x) = \lambda V(x) + (1-\lambda)M(x)$ is a weighted average of $M(x)$ and $V(x)$, and (d) to minimize $R(x)$ for a specific value of λ. To achieve these various objectives we need to be able to estimate different constants in the model. In section 2 we obtain expressions for $M(x)$, $V(x)$, and $R(x)$ and the coefficients that are required for the various objectives are identified. We show that to minimize $R(x)$ over $\underline{x}$ we need unbiased estimates of β_0, $\underline{\beta}$, $\mathbf{B}$, $\underline{\gamma}$, $tr\,\mathbf{C}$, and $\mathbf{D}$. Notice that we do not need to be able to estimate all of the coefficients in the second-order equation. In particular, we need only the trace of the matrix $\mathbf{C}$, where $tr\,\mathbf{C} = \gamma_{11} + \gamma_{22} + \cdots + \gamma_{mm}$. Thus, for the *environmental* variables we do not need to know the *individual* quadratic terms, nor do we need to know any of their interactions with one another.

In section 3 we construct several classes of designs that can be used to provide estimates of the parameters that are needed to minimize $R(x)$. This objective is referred to as objective (d) in the previous paragraph. These designs are illustrated with examples and tables of designs are given. It will be clear from section 2 and from Table 1 that the designs that are appropriate when the objective is to minimize $R(x)$, that is objective (d), will also be appropriate for describing the locus of minimum $R(x)$, that is objective (c) in the previous paragraph.

In section 4 we consider designs that would be appropriate if the objective of interest was to minimize $V(x)$. This objective is referred to above as objective (b). In this section we describe a method for constructing a suitable class of designs and give a series of tables of these designs. In section 5 we consider designs that would be appropriate if the objective of interest was to minimize $M(x)$. This objective is referred to above as objective (a). In this section we describe a method for constructing a suitable class of designs, and a set of tables of these designs is given.

2. *Derivation of the Estimation Table*

Let us suppose that the response of interest can be represented by the general second-order model given in equation (1). Now let

$$g'(\underline{x}) = \left\{ \left[\frac{\partial y_{xz}}{\partial z_1} \right]_{\underline{z}=\underline{0}}, \left[\frac{\partial y_{xz}}{\partial z_2} \right]_{\underline{z}=\underline{0}}, \ldots, \left[\frac{\partial y_{xz}}{\partial z_m} \right]_{\underline{z}=\underline{0}} \right\}. \tag{3}$$

Then $g(\underline{x}) = \underline{\gamma} + \mathbf{D}\underline{x}$ is a measure of the change in the response in the direction of $\underline{z}$ at $\underline{z} = \underline{0}$.

Let

$$G'(\underline{x}) = \left\{ \left[\frac{\partial y_{xz}}{\partial x_1} \right]_{\underline{z}=\underline{0}}, \left[\frac{\partial y_{xz}}{\partial x_2} \right]_{\underline{z}=\underline{0}}, \ldots, \left[\frac{\partial y_{xz}}{\partial x_n} \right]_{\underline{z}=\underline{0}} \right\}. \tag{4}$$

Then $G(\underline{x}) = \underline{\beta} + 2\mathbf{B}\underline{x}$ is a measure of the change in the response in the direction of $\underline{x}$ at $\underline{z} = \underline{0}$.

Now the mean response over the environmental variables, $\bar{y}_x$, is

$$\bar{y}_x = k \int_{R_z} y_{xz}\, d\underline{z} = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'\mathbf{B}\underline{x} + \frac{1}{3}\, tr\mathbf{C} \tag{5}$$

where $k^{-1} = \int_{R_z} d\underline{z}$ is an integrating constant. Let $\dot{\beta}_0 = \beta_0 + \frac{1}{3}\, tr\mathbf{C}$. Then

$$\bar{y}_x = \dot{\beta}_0 + \underline{x}'\underline{\beta} + \underline{x}'\mathbf{B}\underline{x}. \tag{6}$$

Since $G(\underline{x}) = \underline{\beta} + 2\mathbf{B}\underline{x}$, then $\underline{x} = \frac{1}{2}\, \mathbf{B}^{-1}(G - \underline{\beta})$ and so

$$\bar{y}_x = \dot{\beta}_0 - \frac{1}{4}\,\underline{\beta}'\mathbf{B}^{-1}\underline{\beta} + \frac{1}{4}\,G'\mathbf{B}^{-1}G. \tag{7}$$

Let $\beta_0^* = \dot{\beta}_0 - \frac{1}{4}\,\underline{\beta}'\mathbf{B}^{-1}\underline{\beta}$. Then β_0^* is a constant, independent of $\underline{x}$ and is the maximum of the mean $\bar{y}_x$. We have

$$y_{xz} = \bar{y}_x + \underline{z}'\underline{\gamma} + \underline{z}'\mathbf{C}\underline{z} - \frac{1}{3}\,tr\,\mathbf{C} + \underline{z}'\mathbf{D}\underline{x}. \tag{8}$$

So

$$\begin{aligned} y_{xz} - \bar{y}_x &= \underline{z}'\mathbf{C}\underline{z} + \underline{z}'(\underline{\gamma} + \mathbf{D}\underline{x}) - \frac{1}{3}\,tr\,\mathbf{C} \\ &= \underline{z}'\mathbf{C}\underline{z} - \frac{1}{3}\,tr\,\mathbf{C} + \underline{z}'\,g(\underline{x}). \end{aligned} \tag{9}$$

Let us suppose that the ideal value for the response is τ. One possible objective is to choose the design factor levels, $\underline{x}$, so that the mean response over the environmental variables is as close to τ as possible. A measure of this closeness to the ideal value, τ, is

$$M(x) = k \int_{R_z} (\tau - \bar{y}_x)^2\, d\underline{z} = \left\{ \tau - (\,\beta_0^* + \frac{1}{4}\,G'\mathbf{B}^{-1}G\,) \right\}^2. \tag{10}$$

Then $(\,\tau - \beta_0^*\,)$ is the signed difference between the ideal value and the maximum value obtained by the mean response curve $\bar{y}_x$. Now $G'\mathbf{B}^{-1}G$ is a quadratic form and in the design space, R_x, it defines a set of contours. The shape of these contours depends only on $\underline{\beta}$ and $\mathbf{B}$. The value of the contours also depends on τ, $\dot{\beta}_0$, and tr $\mathbf{C}$, in addition to $\underline{\beta}$ and $\mathbf{B}$.

It would also be desirable to have the variation of the response about the mean as small as possible. A measure of this variation is

$$V(x) = k \int_{R_z} (y_{xz} - \bar{y}_x)^2 \, d\underline{z} = A + \frac{1}{3} g'(\underline{x}) g(\underline{x}), \tag{11}$$

where

$$g(\underline{x}) = \underline{\gamma} + \mathbf{D}\underline{x}$$

and where

$$A = \frac{1}{45} \left\{ 4 \sum_{j=1}^{m} \gamma_{jj}^2 + 5 \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} \gamma_{jk}^2 \right\}$$

is independent of $\underline{x}$.

Now $g'(\underline{x}) g(\underline{x})$ is a quadratic form in the design space, R_x, and it defines a set of contours whose shape depends on $\underline{\gamma}$ and $\mathbf{D}$. Therefore, the shape of the contours of $V(x)$ depends only on $\underline{\gamma}$ and $\mathbf{D}$. The value of the contours of $V(x)$ depends on the elements of $\mathbf{C}$.

Now the relative importance attached to the two measures $V(x)$ and $M(x)$ will vary according to the experimental context. An overall measure of robustness will be a linear combination of $V(x)$ and $M(x)$ and can be written as

$$R(x) = \lambda V(x) + (1 - \lambda) M(x). \tag{12}$$

If the experimenter attaches more weight on being close to the ideal value then he will decrease the value of λ. If he puts more weight on having a small amount of variation then he will increase the value of λ. We see that the appropriate values to choose for $\underline{x}$ will depend on the relative importance that the experimenter gives to the two objectives of being close to the ideal value τ and of having low variation.

Now if we believed that the quadratic loss is an appropriate criterion to minimize, then minimizing

$$L(x) = k \int_{R_z} (\tau - y_{xz})^2 \, d\underline{z} \tag{13}$$

is equivalent to minimizing $R(x)$ with $\lambda = 1/2$. The quadratic loss, therefore, corresponds to giving equal weighting to $M(x)$ and $V(x)$. The quadratic loss is a criterion that has been widely adopted in practice. However, it does not appear that the equal weighting of $V(x)$ and $M(x)$ implied by the adoption of $L(x)$ is the only weighting that should be considered.

The coefficients that need to be estimated depend on the objectives of the experiment. We have shown that the shape of the contours of $M(x)$ depends only on $\underline{\beta}$ and $\mathbf{B}$. However, the value of the contours also depends on τ, β_0, and $tr\ \mathbf{C}$, in addition to $\underline{\beta}$ and $\mathbf{B}$. Similarly, the shape of the contours of $V(x)$ depends only on $\underline{\gamma}$ and $\mathbf{D}$ and the value of the contours of $V(x)$ also depends on $\mathbf{C}$. Thus to determine the locus of points that minimizes $R(x)$ for different values of λ we need to estimate $\underline{\beta}, \underline{\gamma}, \mathbf{B}$, and $\mathbf{D}$. If we are only concerned with minimizing $M(x)$ we only need to estimate $\underline{\beta}$ and $\mathbf{B}$. If we are only concerned with minimizing $V(x)$ we only need to estimate $\underline{\gamma}$ and $\mathbf{D}$.

Thus we see that to minimize $R(x)$ we need to estimate $\underline{\gamma}$, $\mathbf{D}$, β_0, $\underline{\beta}$, $\mathbf{B}$, $tr\ \mathbf{C}$, and to know τ. We do not need to estimate all of the elements of $\mathbf{C}$. In particular, we do not need to be able to estimate the individual quadratic terms for the environmental variables nor the interactions between the environmental factors. Instead all we need to be able to estimate is $tr\ \mathbf{C} = \sum_{j=1}^{m} \gamma_{jj}$. Since minimizing the quadratic loss function, $L(x)$, is equivalent to minimizing $R(x)$ with $\lambda = 1/2$, to minimize $L(x)$ we need to be able to estimate the same coefficients as for minimizing $R(x)$, that is we need to estimate $\underline{\gamma}$, $\mathbf{D}$, β_0, $\underline{\beta}$, $\mathbf{B}$, $tr\ \mathbf{C}$, and to know τ. We do not need to estimate all of the elements of $\mathbf{C}$ but only $tr\ \mathbf{C} = \sum_{j=1}^{m} \gamma_{jj}$.

Therefore, for the second-order model given in equation (1), the constants that we need to estimate for the various objectives of interest are indicated in the Estimation Table shown in Table 2. Since, for the second-order model, we know what these constants are, we can search for experimental designs which estimate them efficiently.

Table 2. Estimation Table for the Second-Order Model

Objective Compute/Minimize	Constants β_0	$\underline{\beta}$	**B**	$\underline{\gamma}$	**C**	**D**	Need to know τ
$R(x) = \lambda_0 V(x) + (1 - \lambda_0) M(x)$	x	x	x	x	*tr*	x	x
$L(x) = V(x) + M(x)$	x	x	x	x	*tr*	x	x
Locus of minimum $R(x)$ for different λ		x	x	x		x	
$V(x)$				x		x	
$M(x)$		x	x				

In table 2, the entry "*tr*" under the column "C" indicates that only the trace of the matrix **C** needs to be known, where trace of **C** is $tr\ \mathbf{C} = \gamma_{11} + \gamma_{22} + \cdots + \gamma_{mm}$.

A method of solution is, therefore, to choose a value for λ (or several candidate values for λ) and then to minimze $R(x) = \lambda V(x) + (1 - \lambda) M(x)$. To minimize $R(x)$ we need to be able to estimate $\underline{\gamma}$, **D**, β_0, $\underline{\beta}$, **B**, tr **C**, and to know τ. We do not need to be able to estimate all of the elements of **C**. Designs which enable the estimation of these parameters will be given in section 3.

An alternative method of solution is to determine an upper bound for $M(x)$, $M(x)^0$ say, and to minimize $V(x)$ subject to the constraint that $M(x) \leq M(x)^0$. To do this we need to be able to estimate the same parameters as above. Note that it is inappropriate to minimize $M(x)$ subject to the constraint that $V(x) \leq V(x)^0$, where $V(x)^0$ is some upper bound on the variation, $V(x)$. This is because to determine $V(x)$ we require all of the elements of **C**.

3. Construction of Designs for the Minimization of R(x)

In the previous section, section 2, we derived the Estimation Table (Table 2) that indicates the coefficients that need to be estimated when the response can be adequately represented by the full second-order model given by equation (2). In that section we showed that to minimize $R(x) = \lambda V(x) + (1-\lambda)M(x)$ for a particular value of λ we need to be able to estimate $\underline{\gamma}$, **D**, β_0, $\underline{\beta}$, **B**, and *tr* C, and we also need to know τ. In sections 3.1 - 3.5 we will construct several classes of experimental designs that will enable us to satisfy this objective and give examples of such designs.

From section 2 we conclude that if the objective of interest is to minimize $R(x)$ for a given value of λ we need to be able to estimate:

(1) all first-order effects in both the design and the environmental variables,

(2) all interactions between the design and the environmental variables,

(3) all interactions among the design variables,

(4) all pure quadratic effects of the design variables,

(5) the sum of the pure quadratic effects of the environmental variables.

In particular, we do not need to estimate the individual pure quadratic effects of the environmental variables nor the interactions among the environmental variables.

Clearly, if the experimenter has an objective of interest other than minimizing $R(x)$ for a given value of λ then it may be possible to run an experiment using a smaller design. Table 2 shows that to determine the locus of points, $R^*(x)$, that minimizes $R(x) = \lambda V(x) + (1-\lambda)M(x)$ for different values of λ $(0 \leq \lambda \leq 1)$, we need to be able to estimate $\underline{\gamma}$, **D**, $\underline{\beta}$, and **B**. In the list given above, this corresponds to being able to estimate (1), (2), (3), and (4), but not (5). From this it is clear that designs that are appropriate if the objective is to minimize $R(x)$ for a particular value of λ will also be appropriate if the objective is to determine the locus of minimum $R(x)$ for different values of λ. However, in some cases it might be possible to use smaller designs for this objective. We will indicate in sections 3.1 - 3.5 how the designs that are constructed for minimizing $R(x)$ might be adapted for the situation when the objective is to determine the locus of minimum $R(x)$.

In section 4 we will construct a class of designs for the situation where the objective of interest is to minimize $V(x)$. In section 5 we will construct a class of designs that are appropriate if the objective is to minimize $M(x)$.

3.1. Augmented Resolution IV and V Designs

One possible class of designs that would be appropriate to use are designs based on the standard Central Composite Designs which were introduced by Box and Wilson (1951) and enable the estimation of all of the coefficients of a full second-order polynomial in k variables, including the pure quadratic terms. These composite designs consist of

(1) $n_c = 2^{k-p} r_c$ points forming the cube portion of the design, with coordinates of the form $(\pm 1, \pm 1, \ldots, \pm 1)$,

(2) $n_s = 2kr_s$ axial points forming the star portion of the design, with coordinates at $(\pm\alpha, 0,..., 0), (0, \pm\alpha,..., 0), (0, 0,..., \pm\alpha)$,

(3) n_0 center points at $(0, 0,..., 0)$.

The value of r_c measures the degree of replication of the cube portion of the design and the value r_s measures the degree of replication of the star portion of the design.

A table of these designs is given in Table 3. In this table we show the total number of factors, $n + m$, that can be accommodated in a design with a given number of runs. The designs would then be augmented by star points for each product design variable and by center points. The designs would be very similar to the standard central composite designs, (see, for example, Box and Wilson (1951), Box and Hunter (1957), Box and Draper (1987)), the only difference being that star points are not added for all of the factors, but only for the n product design factors.

Table 3. Resolution V Designs

Number of Runs	Total Number of Factors ($n + m$)	Factor Representations
2^3	3	A, B, C
2^4	5	A, B, C, D, ABCD
2^5	6	A, B, C, D, E, ABCDE
2^6	8	A, B, C, D, E, F, ABCD, ABEF
2^7	11	A, B, C, D, E, F, G, ABCG, BCDE, ACDF, ABCDEFG

A drawback with using a resolution V design for the cube portion of the design is that these designs can require a large number of runs. For example, a resolution V design with 9 design and environmental factors would require 128 runs for the cube portion of the design.

An alternative to using resolution V designs would be to use a resolution IV design which confounded the interactions of the environmental factors in strings with each other and not with the main effects or the other two-factor interactions. Designs which meet this objective will be ones where, in the defining relation, the only factors that appear in any word of length four are environmental factors. Some designs of resolution IV that can be used to provide unbiased estimates of the main effects and the interactions involving the design factors are given in Table 4.

Let us suppose that the objective of interest is to minimize $R(x)$ for some λ. Then an appropriate experimental arrangement could be obtained by using one of the designs given in Table 4 augmented with star points for each design variable and by center points. The designs would be similar to the standard central composite designs, the contrast being that the particular experimental context means that we can use a resolution IV design for the cube portion of the design and that star points are not added for all of the factors, but only for the n design factors.

Now if the objective of interest is to determine the locus of points that minimize $R(x)$ for different values of λ $(0 \leq \lambda \leq 1)$, we do not need to be able to estimate the sum of the pure quadratic effects of the environmental factors. This sum is estimated from the center points of the design. Thus, for this objective smaller designs might be appropriate by reducing the number of center points in the design.

Table 4. Some Useful Resolution IV Designs

Number of Runs	Number of Design Factors (n)	Number of Environmental Factors (m)	Total Number of Factors	Factor Representations Design Factors ; Environmental Factors
2^5	1	8	9	E ; A, B, C, D, BCDE, ACDE, ABDE, ABCE
	2	7	9	E, ACDE ; A, B, C, D, ABC, ABD, BCD
	3	4	7	A, B, D ; C, E, ABCD, ABDE
2^6	1	16	17	A ; B, C, D, E, F, BCD, BCE, BCF, BDE, BDF, BEF, CDE, CDF, CEF, DEF, BCDEF
	2	15	17	A, ABCDEF ; B, C, D, E, F, BCD, BCE, BCF, BDE, BDF, BEF, CDE, CDF, CEF, DEF
	3	8	11	A, B, ABCDEF ; C, D, E, F, CDE, CDF, CEF, DEF
	4	6	10	A, B, C, ABDE ; D, E, F, DEF, BCEF, BCDF
	5	4	9	B, C, E, F, ABCD ; A, D, ACEF, CDEF
2^7	1	32	33	A ; B, C, D, E, F, G, and all possible odd-order interactions of B - F
	2	16	18	A, B ; C, D, E, F, G, and all possible odd-order interactions of C - F
	3	12	15	E, F, G ; A, B, C, D, ABC, ABD, ACD, BCD, AEFG, BEFG, CEFG, DEFG
	4	9	13	E, F, G, AEFG ; A, B, C, D, ABC, ABD, ACD, BCD, BCEF
	5	8	13	E, F, G, AEFG, BCEF ; A, B, C, D, ABC, ABD, ACD, BCD
	6	6	12	C, D, E, F, G, ABCD ; A, B, ACEF, BCEF, ADFG, BDFG
	7	5	12	C, D, E, F, G, ABCD, ABEF ; A, B, ABCEG, ADFG, BDFG
	8	4	12	C, D, E, F, G, ABCF, ABDG, ABEFG ; A, B, ACDE, BCDE

3.2. Modified Box-Behnken Designs

We have shown in Table 2, that in experiments where the response can be adequately represented by the second-order model in both the design and the environmental factors given by equation (2), that the parameters that need to be estimated to minimize $R(x)$ are all of the main effects, all of the interactions among the design variables, all of the interactions between the design and environmental variables, all of the quadratic effects of the design variables, and the sum of the quadratic effects of the environmental variables. In particular, we do not need to be able to estimate any of the interactions among the environmental factors.

A possible class of designs that can be used to estimate the parameters in this situation can be derived from the Box-Behnken designs. These designs are a class of incomplete three-level factorial designs proposed by Box and Behnken (see Box and Behnken (1960)) that can be used to estimate the all of the coefficients of a second-degree polynomial. The designs either meet, or approximately meet, the criterion of rotatability and possess a high degree of orthogonality. The non-orthogonality is usually the presence of correlated estimates of the mean and the pure quadratic coefficients.

Box-Behnken designs can be used in the experimental context that we are considering, and, since we do not need to estimate the interactions among the environmental variables, certain runs that are used in the standard Box-Behnken designs can be omitted. Thus for some situations we can derive designs that can estimate all of the parameters of interest with fewer runs than a standard Box-Behnken designs. When runs are deleted from the Box-Behnken design it is important to ensure that the design matrix, $\mathbf{X}$, that is then obtained does not yield a singular $\mathbf{X}'\mathbf{X}$ matrix. The method of construction of these modified Box-Behnken designs is best illustrated with an example.

Box and Behnken (1960, p. 460) give a 27-run design that can be used to estimate all the coefficients of a second-order poynomial with 4 factors, each at three levels, denoted by -1, 0, +1. The design is given in Table 5, where in this table $(\pm1, \pm1)$ means that all combinations of plus and minus levels are to be run.

Table 5. Box-Behnken Design in Four Factors

	Factor			
Runs	A	B	C	D
1-4	±1	±1	0	0
5-8	0	0	±1	±1
9-12	0	±1	±1	0
13-16	±1	0	0	±1
17-20	±1	0	±1	0
21-24	0	±1	0	±1
25-27	0	0	0	0

Now if the factors C and D are environmental factors then we can omit runs 5-8. The resulting $\mathbf{X}'\mathbf{X}$ matrix is nonsingular. We can now estimate the 14 coefficients that we require with just 23 runs. The estimate of the variances of the effects depends on whether the effect is associated with a design or an environmental factor. The main effects of the design factors are estimated more precisely than the main effects of the environmental factors. Similary, the quadratic effects of the design factors are estimated more precisely than the quadratic effects of the environmental factors.

In Tables 6 and 7 we give further examples of this modified Box-Behnken approach. In Table 6, if factors D and E are environmental factors and factors A, B, and C are design factors, then we can omit runs 17-20 and will be able to estimate the required 20 coefficients in 42 runs. If factors C, D, and E are environmental factors and factors A and B are design factors, then we can omit runs 5-8, 17-20, and 29-32 and will be able to estimate the required 18 coefficients in 34 runs.

Table 6. Modified Box-Behnken Design in Five Factors

	Factor				
Runs	A	B	C	D	E
1-4	±1	±1	0	0	0
5-8	0	0	±1	±1	0
9-12	0	±1	0	0	±1
13-16	±1	0	±1	0	0
17-20	0	0	0	±1	±1
21-24	0	±1	±1	0	0
25-28	±1	0	0	±1	0
29-32	0	0	±1	0	±1
33-36	±1	0	0	0	±1
37-40	0	±1	0	±1	0
41-46	0	0	0	0	0

In Table 7, if factors A, B, C, and D are design factors and E, F, and G are environmental factors, then it appears that runs 1-8 could be omitted. However, this yields a $X'X$ matrix that is singular. If, instead, we omit runs 1, 4, 6, and 7 then $X'X$ is nonsingular and we will be able to estimate the required 33 coefficients in 58 runs.

Table 7. Modified Box-Behnken Design in Seven Factors

Runs	A	B	C	D	E	F	G
	Factor						
1-8	0	0	0	0	±1	±1	±1
9-16	±1	0	0	±1	0	0	±1
17-24	0	±1	0	±1	0	±1	0
25-32	±1	±1	0	0	±1	0	0
33-40	0	0	±1	±1	±1	0	0
41-48	±1	0	±1	0	0	±1	0
49-56	0	±1	±1	0	0	0	±1
57-62	0	0	0	0	0	0	0

3.3. Koshal Designs

A class of experimental designs that permit the fitting of the full second-order model in a minimal number of runs has been proposed by Koshal (1933). In these designs the number of runs is exactly equal to the number of coefficients that need to be estimated in the full second-order model. Thus for a model equation in k factors, the designs contain $\frac{1}{2}(k + 1)(k + 2)$ points. Fung (1986) identified three types of Koshal designs, classified as Type A, Type B, and Type C. When comparing the three types of designs Fung concluded that "the designs of Type B seem most desirable ... (T)hey are certainly the most convenient to generate and use in calculation; and they are not unduly poor in terms of their information and bias patterns."

To construct the Type B Koshal design suppose that the design for k factors can be represented by rows $(x_1, \ldots, x_k)$. Then, if $i = 1,...,k$ indexes the coordinates of each row, the Type A Koshal design contains

(i) One row with $x_i = 0$ for all $i = 1,...,k$.

(ii) k rows with $x_i = 1$ for $i = j$ and $x_i = 0$ for $i \neq j$; for $j = 1,...,k$.

(iii) k rows with $x_i = -1$ for $i = j$ and $x_i = 0$ for $i \neq j$; for $j = 1,...,k$.

(iv) $\binom{k}{2}$ rows with $x_i = 1$ for both $i = j$ and $i = l$, and $x_i = 0$ otherwise; for all pairs j, l where $j < l$; $j = 1,...,k$ and $l = 1,...,k$.

The designs for $k = 2$ and $k = 3$ factors are given in Figure 1.

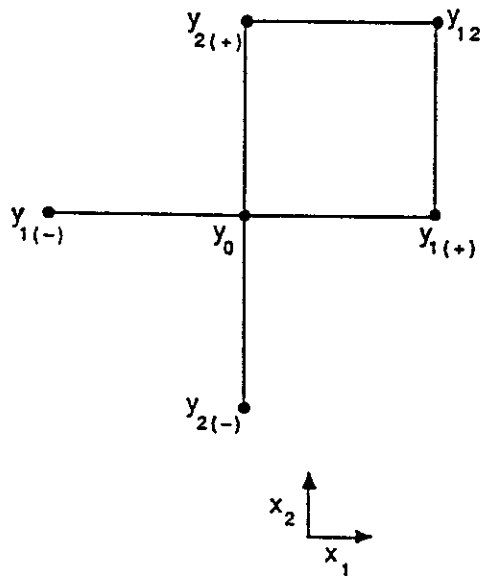
In this section we will look at the use of Koshal designs in experiments where the objective is to minimize the effect of environmental variables. Let us assume, as before, that the response can be adequately represented by the full second order model in the design and environmental factors given in equation (1). We showed earlier that we need to be able to estimate all of the coefficients involving terms that include a design factor but we only need to be able to estimate some of the coefficients involving terms that include an environmental factor. In particular, we do not need to estimate the individual pure quadratic effects of the environmental variables nor the interactions among the environmental variables.

Initially, let us suppose that we are running an experiment which only investigates the effect of design factors on the response. Then we could consider using a Type B Koshal design to enable us to estimate the required coefficients of the design variables. Now consider an experiment where we are only concerned with how the environmental factors affect the response. Then, since we do not need to be able to estimate all of the coefficients, it might be

Figure 1. Type B Koshal Designs for (a) k = 2 and (b) k = 3 factors

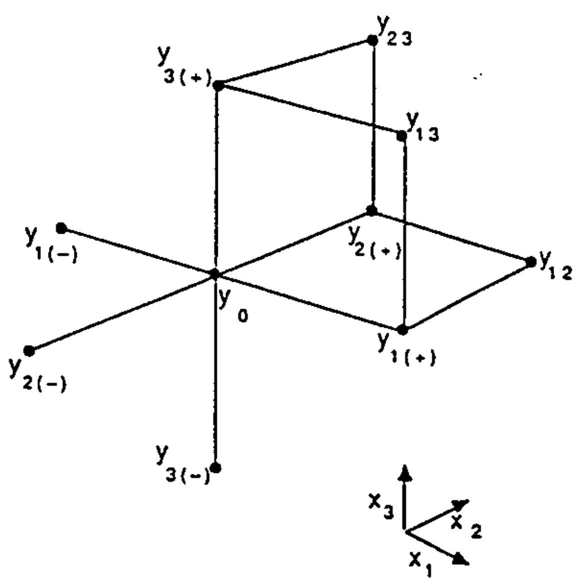
(a) $k = 2$

y	x_1	x_2
y_0	0	0
$y_{1(+)}$	1	0
$y_{2(+)}$	0	1
$y_{1(-)}$	−1	0
$y_{2(-)}$	0	−1
y_{12}	1	1



(b) $k = 3$

y	x_1	x_2	x_3
y_0	0	0	0
$y_{1(+)}$	1	0	0
$y_{2(+)}$	0	1	0
$y_{3(+)}$	0	0	1
$y_{1(-)}$	−1	0	0
$y_{2(-)}$	0	−1	0
$y_{3(-)}$	0	0	−1
y_{12}	1	1	0
y_{13}	1	0	1
y_{23}	0	1	1



conjectured that it is possible to construct a modified Type B Koshal design that contained fewer runs than the standard Type B Koshal design and was such that the required coefficients could be estimated.

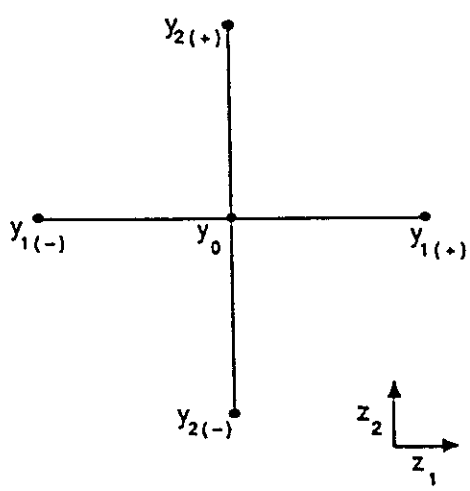
As an example, consider a situation where we have two environmental factors, z_1 and z_2. A full second-order equation in the two factors would be

$$y = \beta_0 + \gamma_1 z_1 + \gamma_2 z_2 + \gamma_{11} z_1^2 + \gamma_{22} z_2^2 + \gamma_{12} z_1 z_2. \tag{14}$$

Then the standard Type B Koshal design with two factors is given in Figure 1(a). But, for the particular experimental context that we are considering, we do not need to be able to estimate the coefficient γ_{12}, the interaction term. This leads us to suggest that an appropriate design would be one that omits the last run in the standard Type B Koshal design. This yields the modified design given in Figure 2, which we will call a Type M Koshal design.

Figure 2. Type M Koshal Design in Two Factors

y	z_1	z_2
y_0	0	0
$y_{1(+)}$	1	0
$y_{2(+)}$	0	1
$y_{1(-)}$	−1	0
$y_{2(-)}$	0	−1



This design has an appealing star-shaped pattern. It enables the five coefficients of interest, $\beta_0, \gamma_1, \gamma_2, \gamma_{11}$, and γ_{22}, to be estimated in only five runs. This modified Type B design can clearly be extended to experimental situations where there are more than two environmental factors. When there are m environmental factors then the Type M Koshal design is constructed by taking rows that contain

(i) One row with $z_j = 0$ for all $j = 1,...,m$.

(ii) m rows with $z_j = 1$ for $j = k$ and $z_j = 0$ for $j \neq k$; for $k = 1,...,m$.

(iii) m rows with $z_j = -1$ for $j = k$ and $z_j = 0$ for $j \neq k$; for $k = 1,...,m$.

The above discussion suggests a method of deriving modified Koshal designs that can be used to estimate the cofficients of interest in a full second-order model. Suppose that we have n design variables, $x_1, x_2, \ldots, x_n$, and m environmental variables, $z_1, z_2, \ldots, z_m$. Then a modified Koshal design (based on the Type M Koshal design) is obtained by constructing rows that contain

(I) One row with $x_i = 0$ for all $i = 1,...,n$ and with $z_j = 0$ for all $j = 1,...,m$.

(II) n rows with $x_i = 1$ for $i = l$ and $x_i = 0$ for $i \neq l$, for $l = 1,...,n$; and $z_j = 0$ for $j = 1,...,m$.

(III) m rows with $z_j = 1$ for $j = k$ and $z_j = 0$ for $j \neq k$, for $k = 1,...,m$; and $x_i = 0$ for $i = 1,...,n$.

(IV) n rows with $x_i = -1$ for $i = l$ and $x_i = 0$ for $i \neq l$, for $l = 1,...,n$; and $z_j = 0$ for $j = 1,...,m$.

(V) m rows with $z_j = -1$ for $j = k$ and $z_j = 0$ for $j \neq k$, for $k = 1,...,m$; and $x_i = 0$ for $i = 1,...,n$.

(VI) $\binom{n}{2}$ rows with $x_i = 1$ for both $i = l$ and $i = k$, and $x_i = 0$ otherwise, for all pairs l, k where $l < k$; $l = 1,...,n$ and $k = 1,...,n$; and $z_j = 0$ for $j = 1,...,m$.

(VII) $n \times m$ rows with $x_i = 1$ for $i = l$ and $z_j = 1$ for $j = k$ for all pairs $l = 1,...,n$ and $k = 1,...,m$; and $x_i = 0$ for $i \neq l$ and $z_j = 0$ for $j \neq k$.

For the modified Koshal designs we see that (I), (II), (IV), and (VI) yield a standard Type B Koshal design in the design variables, and that (I), (III), and (V) yield a Type M Koshal design in the environmental variables. The rows obtained from (VII) are the runs that enable the interactions between the design and the environmental factors, the δ_{ji}, to be estimated.

Table 8 shows the design that is obtained when there are three design factors and three environmental factors. In this design we have 25 runs which is exactly equal to the number of parameters that need to be estimated (given that we estimate the individual quadratic coefficients for the environmental variables, and not just the sum of these quadratic coefficients). The categories given in the table indicate how the particular rows are constructed as detailed above.

3.4. Hartley's Composite Designs

Composite designs were introduced by Box and Wilson (1951) to enable the estimation of all of the coefficients of a full second-degree polynomial in k variables, including the pure quadratic terms. In these designs the cube portion of the design is used to estimate the mean, the main effects, and the interaction terms. This portion of the design is a factorial or fractional factorial design of resolution of at least V. However, the cube portion of the design does not give any information about the pure quadratic coefficients. Estimates of the quadratic terms are obtained by augmenting the cube portion with a star design and center points.

The composite design is particularly useful in the context of sequential experimentation. Hartley (1959) considered the composite designs with regard to their merits as a single (one-shot) experiment. He noted that information on the main effects could be obtained from the star and center points and that the effects that the cube portion needed to be able to estimate were the

Table 8. A Modified Koshal Design with Three Design and Three Environmental Factors

Category	x_1	x_2	x_3	z_1	z_2	z_3
(I)	0	0	0	0	0	0
(II)	1	0	0	0	0	0
	0	1	0	0	0	0
	0	0	1	0	0	0
(III)	0	0	0	1	0	0
	0	0	0	0	1	0
	0	0	0	0	0	1
(IV)	−1	0	0	0	0	0
	0	−1	0	0	0	0
	0	0	−1	0	0	0
(V)	0	0	0	−1	0	0
	0	0	0	0	−1	0
	0	0	0	0	0	−1
(VI)	1	1	0	0	0	0
	1	0	1	0	0	0
	0	1	1	0	0	0
(VII)	1	0	0	1	0	0
	1	0	0	0	1	0
	1	0	0	0	0	1
	0	1	0	1	0	0
	0	1	0	0	1	0
	0	1	0	0	0	1
	0	0	1	1	0	0
	0	0	1	0	1	0
	0	0	1	0	0	1

interactions. He therefore derived composite designs that required fewer runs for the cube portion by using fractional factorial designs that aliased some or all of the two-factor interactions with some of the main effects. Thus, the cube portion of the design could be of resolution lower than V provided that none of the two-factor interactions were aliased with any other two-factor interaction.

Designs of this form have been examined by Westlake (1965) and by Draper and Lin (1988). Draper and Lin (1988) extended this principle and proposed that designs of this type be defined as resolution R^* designs. In general, a two-level fractional factorial design is said to be of resolution R^* if it is of resolution R and there is no word in the defining relation of length $R+1$. By this definition, Hartley's designs are of resolution III^* since they are resolution III designs with no word of length four in the defining relation. Draper and Lin (1988) constructed a table that showed the maximum number of factors that can be accommodated in designs of resolution R^*, for $III \leq R \leq XIII$, and for 2^p runs, for $p \leq 12$.

Now let us consider the applicability of Hartley's composite designs to the experimental context that we are considering, namely that of experimentation to minimize the effect of environmental factors. Let us assume, as before, that the response can be adequately represented by the full second-order model in the design and environmental factors given in equation (1). It was shown in section 2 that to minimize $R(x)$, a measure of non-robustness, we need to be able to estimate all of the interactions involving terms that include a design factor (i.e. β_{ik} and dij.) but that we do not need to estimate the interactions among the environmental variables (the γ_{jk}).

Now Hartley's original designs were constructed by using design generators that did not contain words of length four (but they could contain words of length three; see earlier comments on the definition of designs of resolution R^*). This ensured that within each alias set there was no more than one two-factor interaction term, and so the two-factor interaction coefficients would not be aliased with each other. We can modify Hartley's approach for our particular experimental context by using design generators that do not contain any words of length four that are made up of letters corresponding to the design factors. Thus it is permissible to have words of length four in the design generators, but these words can only contain letters corresponding to the environmental factors. Designs that are constructed using generators of this type will be such that:

(1) interactions involving any one of the n design factors will be in separate alias sets,

(2) interactions among the m environmental factors may be aliased with each other but they will not be aliased with interactions involving any of the design factors.

An example will illustrate the method of construction. Suppose that we wish to conduct an experiment with one design factor and four environmental factors. Ket us, without loss of generality, define A to be the single design factor of interest. Then the cube portion of this design can be run in only eight runs by constructing a fractional factorial design using the generators I = ABC = ADE = BCDE. The alias sets for this design are given below.

I	=	ABC	=	ADE	=	BCDE
AB	=	C	=	BDE	=	ACDE
AC	=	B	=	CDE	=	ABDE
AD	=	BCD	=	E	=	ABCE
AE	=	BCE	=	D	=	ABCD
BC	=	A	=	ABCDE	=	DE
BD	=	ACD	=	ABE	=	CE
BE	=	ACE	=	ABD	=	CD

We see that the interactions of interest, those involving factor A, are in alias sets that contain main effects, and higher-order interactions that are assumed negligible. The interactions among the environmental factors are confounded with each other and not all of them can be separately estimated. For this particular experimental context, since we do not need to estimate the interactions among the environmental factors, the fact that this design confounds these interactions is not of negative consequence.

The following table, Table 9, gives some more useful designs that have been constructed in a similar way. In this table the design factors are designated as A, B,..., according to the number of design factors required, and the complete list of design generators is obtained by taking all generalized interactions of the generators given.

Table 9 can be used to construct experimental designs that are appropriate when the response can be assumed to follow a second-degree model and our objective of interest is to minimize $R(x)$. The complete design would be obtained by augmenting the modified resolution III^* designs by star points for *all* of the factors (both design and environmental factors) and center points. It should be noted that we now need star points for all of the factors because in the resolution R^* designs these star points are used to de-alias the estimates of the two-factor interactions from the main effects.

Table 9. Modified Resolution III* Designs

Number of Factors $n + m$	Number of Design Factors n	Number of Environmental Factors m	Number of Interactions of Interest	Number of Runs	Design Generators
5	1	4	4	8	I = ABC = ADE = ...
7	1	6	6	16	I = ACD = AEF = ABG = ...
8	1	7	7	16	I = ACD = AEF = ABG = ABCEH = ...
8	2	6	13	32	I = ABF = ABCDG = ABCEH = ...
8	3	5	18	32	I = ABF = ABDEG = ABCDEH = ...
8	4	4	22	32	I = ABF = ABEGH = ABCDEH = ...
9	1	9	8	16	I = ABC = ADE = AFG = AHJ = ABDFH = ...
9	2	7	15	32	I = ACD = AEF = AGH = BCEGJ = ...
9	3	6	21	32	I = ADE = AFG = AHJ = BCDFH = ...

Table 9. Modified Resolution III Designs*

Number of Factors $n + m$	Number of Design Factors n	Number of Environmental Factors m	Number of Interactions of Interest	Number of Runs	Design Generators
10	1	9	9	16	I = ABF = ACG = ADH = AEJ = BCDK = FGHJ = ...
10	2	8	17	32	I = ABF = CDG = DEJ
10	1	9	9	32	I = ABF = ACG = ADH = BCDJ = ABCDK = ...
10	2	8	17	32	I = ABF = CDG = DEJ = ABCEH = CDEK = ...
10	3	7	24	32	I = ABC = DEG = DFH = EFJ = DEFK = ...
10	4	6	30	64	I = ABG = CDH = ABCDEJ = ABCDFK = ...
10	5	5	35	64	I = ABG = CDH EFK = ABCDFJ = ...
10	6	4	39	64	I = ABG = CDH EFJ = GHJK = ...

Table 9. Modified Resolution III Designs*

Number of Factors $n+m$	Number of Design Factors n	Number of Environmental Factors m	Number of Interactions of Interest	Number of Runs	Design Generators
11	1	10	10	32	I = ABF = ACG = ADH = AEL = BCDJ = ABCDK = ...
11	2	9	19	32	I = ABF = CDG = CEH = DEJ = CDEK = FGHKL = ...
11	3	8	27	64	I = ABG = CDH = CEJ ABDEK = ABCDEL = ...
11	4	7	34	64	I = ABG = CDH = EFL = ABCDEJ = ABCDFK = ...
11	5	6	40	64	I = ABG = CDH = EFK = ABCDFJ = ABCDEFL = ...
12	1	11	11	32	I = ABF = ACG = ADH = AEK = ABCDJ = ABCEL = ABDEM = ...
12	2	10	21	32	I = ABF = CDG = CEH = DEJ = CDEK = = FGHKL = ABJLM = ...
12	3	9	30	64	I = ABG = CDH = CEJ = CFM = ABDEK = ABCDEL = ...
12	4	8	38	64	I = ABG = CDH = EFL = ABCDEJ = ABCDFK = ABCDEFM = ...

3.5. *Irregular Fractional Factorial Designs*

In section 2 we have shown that if we assume that the response can be represented by a second-order model in the design and environmental factors, then the experimental designs that we construct need to be able to estimate all first-order effects in both the design and the environmental variables, all interactions between the design and the environmental variables, all interactions among the design variables, all pure quadratic effects of the design variables, and the sum of the pure quadratic effects of the environmental variables.

In sections 3.1 - 3.4 we have proposed several alternative experimental designs for this particular experimental situation. In sections 3.1 and 3.4 these designs were based on composite designs, constructed by augmenting a cube design, which is generally some form of a factorial or fractional factorial design, with star and center points.

In this section we will describe an alternative class of experimental designs that, for the cube portion of the design, enables the estimation of all main effects and two-factor interactions but with fewer runs than is required by resolution *V* designs. The designs given in this section differ from regular fractional factorial designs in that the main effects and two-factor interactions are not orthogonal to each other. These effects are grouped together into alias sets. The alias sets are generated by the defining relation of the experiment. Effects in the same alias set are not orthogonal to one another but they are orthogonal to all effects in other alias sets.

We will illustrate the construction of this class of designs by deriving a 3/4 fraction of a 2^4 design. The design is constructed by taking any three quarters of a 2^{4-2} design obtained from some defining relation. Consider the defining relation $I = \pm BD = \pm ABC = \pm ACD$. This defining relation generates four fractions of the 2^4 design, namely the quarters determined by

$$I = +BD = +ABC = +ACD$$

$$I = -BD = +ABC = -ACD$$

$$I = +BD = -ABC = -ACD$$

$$I = -BD = -ABC = +ACD.$$

The 3/4 fraction design is obtained by omitting any one of these quarters. If we omit the first fraction then the three remaining quarters may be combined into three half fractions that are defined by $I = -ACD$ (from combining the second and third quarter), $I = -BD$ (from combining the second and fourth quarter), and $I = -ABC$ (from combining the third and fourth quarter). Assuming that three-factor interactions are negligible, we can estimate A from the half fraction defined by $I = -BD$ since in that fraction A is confounded with $-ABD$. The other main effects and two-factor interactions are found in a similar way as indicated in Table 10.

Table 10. Alias Sets for the 3/4 2^4 Design

Effect	Alias Set	Confounding
A	$I = -BD$	ABD
B	$I = -ACD$	$ABCD$
C	$I = -BD$	BCD
D	$I = -ABC$	$ABCD$
AB	$I = -ACD$	BCD
AC	$I = -BD$	$ABCD$
AD	$I = -ABC$	BCD
BC	$I = -ACD$	ABD
BD	$I = -ACD$	ABC
or	$I = -ABC$	ACD
CD	$I = -ABC$	ABD

It can be shown (John(1962)) that when higher-order interactions are assumed negligible this design yields the least squares estimates of the main effects and two factor interactions. BD and the mean may be estimated from two half fractions and the least squares estimate will be the average of these two estimates.

Designs of this sort, which we will call irregular fractional factorial designs since they are fractional factorial designs where the numerator of the fractions is greater than one, have been examined in some detail elsewhere. (See, for example, John (1961, 1962, 1966, 1969, 1971), Addelman (1961), McLean and Anderson (1984)). A table of useful irregular fractional factorial designs that permit the estimation of all main effects and two-factor interactions is given in Table 11. In this table the complete list of design generators is obtained by taking all generalized interactions of the generators given. All of the designs in this table assume that three-factor and higher-order interactions are negligible. In addition, the second design given also requires the assumption that the EF interaction is negligble. (In the design given, EF is completely confounded with B, AF, and CD.)

The three-quarter fractions can be viewed as being obtained by adding a quarter fraction to a half fraction. John (1971) describes how the addition of other fractions to the half fraction can also yield some useful designs. Using this technique he obtained a design for seven factors in 36 runs by adding to the 2^{6-1} design defined by $I = -ABCDEF$ the four-run fraction defined by $I = +ABCDEF = +ABDEF = -ABC = +BCDEF = \cdots$ and obtaining the seventh factor by putting $G = ABCD$. He also derived a design for nine factors in 80 runs by adding to the 2^{7-1} design defined by $I = -AB$ the 2^{7-3} fraction defined by $I = AB = ACF = \cdots$ and then putting $H = AEFG$ and $J = BCDE$.

Table 11. Useful Irregular Fractional Factorial Designs

Number of Factors	Number of Runs	Fraction	Generators of Defining Relation
4	12	3/4	$I = \pm BD = \pm ABC = \cdots$
6	24	3/8	$I = \pm AE = \pm ABCDE = \pm ABF = \cdots$ *
6	28	7/16	$I = \pm ABC = \pm ABCE = \pm BCDE = \pm ACDEF = \cdots$
7	48	3/8	$I = \pm ABC = \pm DEF = \pm ABDEG = \cdots$
8	48	3/16	$I = \pm ABC = \pm DEF = \pm ABDEG = \pm ACDFH = \cdots$
9	96	3/16	$I = \pm ABCF = \pm CDEG = \pm ABEGH = \pm ADEFJ = \cdots$
10	96	3/32	$I = \pm ABCF = \pm ABEGH = \pm ADEFJ = \pm ACDFGK = \cdots$
11	96	3/64	$I = \pm ABCF = \pm CDEG = \pm ABEGH = \pm ACDFGK = \pm BCDGL = \cdots$

* This design requires the additional assumption that *EF* is negligible

Now let us consider the applicability of these irregular factional factorial designs to the experimental context that we are considering, namely that of experimentation to minimize the effect of environmental factors. The irregular fractional factorial designs that we have given above enable the estimation of *all* main effects and two-factor interactions. It appears that it may be possible to construct irregular fractional factorial designs that estimate all of the parameters of interest, but with fewer runs than required by regular or irregular fractional factorial designs since it is unnecessary for the design to yield estimates of the two-factor interactions among the environmental factors.

As an example, consider the 3/4 fraction of the 2^6 design obtained from the design generators $I = \pm ABCDF = \pm ABCDE = \pm EF$. By putting $G = ABCE$, $H = ADEF$, and $J = BCDEF$, we can obtain a nine factor design in 48 runs, where A, B, and C are product design factors and D, E, F, G, H, and J are environmental factors. The alias sets for this design are given in Table 12. Now we note that three of the rows (2, 3, and 12) each contain four non-negligible effects. For row 2 this is not of importance since the effects in this row are all two-factor interactions of environmental factors, which we do not need to be able to estimate. However, the other two rows contain the main effect of A and the BC interaction, effects that we do need to be able to estimate. Therefore this design can be used provided we assume that two of the two-factor environmental interactions in these two rows (FJ and DJ, say) are negligible. A standard resolution V design in nine factors requires 128 runs, and even the irregular fractional factorial design requires 96 runs whereas this design only requires 48 runs.

The irregular fractional factorial designs described in this section can be used to construct experimental designs that are appropriate for the particular experimental context that we are considering. The complete design would be obtained by augmenting the irregular fractional factorial designs by star points for each of the design factors and center points.

Table 12. Alias Sets for a 48-run Design with Three Design Factors and Six Environmental Factors

Row	Alias Sets				Number of Non-negligible Effects
1	I	ABCDF	ABCDE=DG	EF	3
2	ABC=EG=HJ	DF	DE	ABCE=FG	4
3	A	BCDF=GH=EJ	BCDE=FJ	AEF=DH	4
4	B	ACDF	ACDE	BEF	1
5	C	ABDF	ABDE	CEF	1
6	D	ABCF	ABCE=G	DEF=AH	3
7	E	ABCDEF=AJ	ABCD	F	3
8	AB	CDF	CDE	ABEF	1
9	AC	BDF	BDE	ACEF	1
10	AD	BCF	BCE=AG	ADEF=H	3
11	AE	BCDEF=J	BCD	AF	3
12	BC	ADF=EH=GJ	ADE=FH	BCEF=DJ	4
13	BD	ACF	ACE=BG	BDEF=CJ	3
14	BE	ACDEF=CH	ACD	BF	3
15	CD	ABF	ABE=CG	CDEF=BJ	3
16	CE	ABDEF=BH	ABD	CF	3

4. *Minimization of V(x) for the Second-Order Model*

Table 2 gives the constants that need to be estimated to achieve the possible experimental objectives outlined in Box and Jones (1990a). If we want to minimize $V(x)$ then we only need to be able to estimate $\underline{\gamma}$ and **D**. In this section we will propose a class of designs that are smaller than the designs given in sections 3.1 - 3.5 and can be used to estimate these coefficients. It should be noted that in this section, although we do not need to be able to estimate the design factor main effects, the β_i's, and the two-factor interactions among the design factors, the β_{ik}'s, and among the environmental factors, the γ_{jk}'s, we do not assume that these terms are negligible. Therefore, in constructing the designs we cannot confound these factors with the terms that we do need to be able to estimate, namely the main effects of the environmental factors and the two-factor interactions between the design factors and the environmental factors.

The preceding paragraph implies that we have the following constraints on the identity relationship for the experimental design:

(1) Three-letter words must only contain design factors.

(2) Four-letter words must contain either only design factors or only environmental factors.

(3) Five-letter words can contain any combination of design and environmental factors.

The method of construction that we have followed is to search for designs that satisfy these constraints on the word lengths. The following table, Table 13, gives some designs that have been constructed in this way. These designs use 8, 16, 32, and 64 runs and will provide unbiased estimates of the main effects of the environmental factors and the two-factor interactions between the design and the environmental factors that can be used to minimize $V(x)$. In this table the environmental factors are designated as A, B, ..., according to the number of environmental factors required and the complete list of design generators is obtained by taking all generalized interactions of the generators given.

Table 13. Design Generators for Designs for the Minimization of V(x)

Number of Runs	Number of Environmental Factors	Number of Design Factors	Total Number of Factors	Design Generators
2^3	1	3	4	I = BCD
	2	1	3	
	3	0	3	
	4	0	4	I = ABCD
2^4	1	7	8	I = BCE = BDF = CDG = BCDH = ...
	2	3	5	I = CDE
	3	2	5	I = ABCDE
	4	1	5	I = ABCDE
	5	0	5	I = ABCDE
	6	0	6	I = ABCE = BCDF = ...
	7	0	7	I = ABCE = BCDF = ACDG = ...
	8	0	8	I = BCDE = ACDF = ABCG = ABDH = ...

Table 13. Design Generators for Designs for the Minimization of V(x)

Number of Runs	Number of Environmental Factors	Number of Design Factors	Total Number of Factors	Design Generators
2^5	1	15	16	I = BCF = BDG = BEH = CDJ = CEK = DEL = BCDM = BCEN = BDEO = CDEP = BCDEQ = ...
	2	7	9	I = CDF = CEG = DEH = CDEJ = ...
	3	6	9	I = ABCDE = ABCFH = EFG = ABCGJ = ...
	4	3	7	I = ABCD = EFG = ...
	5	2	7	I = ABCD = ACEFG = ...
	6	2	8	I = ABCD = ABEF = BDFGH = ...
	7	2	9	I = ABCD = ABEF = ACEG = ABGHJ = ...
	8	1	9	I = ABCD = ABEF = ACEG = ADEHJ = ...
	9-16	0	9-16	Resolution *IV* designs

Table 13. Design Generators for Designs for the Minimization of V(x)

Number of Runs	Number of Environmental Factors	Number of Design Factors	Total Number of Factors	Design Generators
2^6	1	31	32	Resolution *III* design in B, C, D, E, F
	2	15	17	Resolution *III* design in C, D, E, F
	3	11	14	I = DEG = DFH = EFJ = DEFK = ABCDEL = ABCDFM = ABCEFN = ABCDEFO = ...
	4	7	11	I = ABCD = EFH = EGJ = FGK = EFGL = ...
	5	5	10	I = ABCFG = ABCHJ = ADEFH = GHK = ...
	6	4	10	I = ABCD = ABEF = GHJ = ACEGK = ...
	7	4	11	I = ABCG = ABDE = ACDF = HJK = ABFHL = ...
	8	3	11	I = ABCG = ABDH = ACDE = ABEF = JKL = ...
	9	3	12	I = ABCG = ABDH = BCDJ = ACDE = BCEF = KLM = ...
	10	3	13	I = ABCG = ABDH = BCDJ ACDK = BCEF = ADEF = LMN = ...

5. *Minimization of M(x) for the Second-Order Model*

Table 2 gives the constants that need to be estimated to achieve the possible experimental objectives outlined in Box and Jones (1990a). If we want to minimize $M(x)$ then we only need to be able to estimate $\underline{\beta}$ and **B**. In this section we will propose a class of designs that are smaller than the designs given in sections 3.1 - 3.5 and can be used to estimate these coefficients. It should be noted that in this section, although we do not need to be able to estimate the environmental factor main effects, the γ_j's, and the two-factor interactions between the design factors and the environmental factors, the δ_{ji}'s, and among the environmental factors, the γ_{jk}'s, we do not assume that these terms are negligible. Therefore, in constructing the designs we cannot confound these factors with the terms that we do need to be able to estimate, namely the main effects of the design factors and the two-factor interactions among the design factors.

The preceding paragraph implies that we have the following constraints on the identity relationship for the experimental design:

(1) Three-letter words must only contain environmental factors.

(2) Four-letter words must contain at most one design factor, that is four-letter words either contain all environmental factors or three environmental factors and one design factor.

(3) Five-letter words can contain any combination of design and environmental factors.

The method of construction that we have followed is to search for designs that satisfy these constraints on word lengths. The following table, Table 14, gives some designs that have been constructed in this way. The designs use 8, 16, 32, and 64 runs and will provide unbiased estimates of the main effects and two-factor interactions of the design factors that can be used to minimize $M(x)$. In this table the design factors are designated as A, B, ..., according to the number of design factors required and the complete list of design generators is obtained by taking all generalized interactions of the generators given.

Table 14. Design Generators for Designs for the Minimization of M(x)

Number of Runs	Number of Design Factors	Number of Environmental Factors	Total Number of Factors	Design Generators
2^3	1	3	4	I = ABCD
	2	1	3	
	3	0	3	
2^4	1	7	8	I = BCE = BDF = CDG = BCDH = ...
	2	3	5	I = ACDE
	3	2	5	I = ABCDE
	4	1	5	I = ABCDE
	5	0	5	I = ABCDE
2^5	1	15	16	Resolution *III* design in B, C, D, E
	2	7	9	I = CDF = CEG = DEH = CDEJ = ...
	3	6	9	I = ABCDE = EFG = ABCFH = ABCGJ = ...
	4	3	7	I = ABCDE = AEFG = ...
	5	1	6	I = ABCDEF
	6	0	6	I = ABCDEF

Table 14. Design Generators for Designs for the Minimization of M(x)

Number of Runs	Number of Design Factors	Number of Environmental Factors	Total Number of Factors	Design Generators
2^6	1	31	32	Resolution *III* design in B, C, D, E, F
	2	15	17	Resolution *III* design in C, D, E, F
	3	7	10	Resolution *III* design D, E, F
	4	7	11	I = AEFG = ABCFH = AEHJ = ABCDK = BCDGHKL = ...
	5	6	11	I = AFGH = ABCDF = EGHJ = AGJK = BCDKL = ...
	6	3	9	I = ABCDG = ABEFH = GHJ = ...
	7	1	8	I = ABCDG = ABEFH = ...
	8	0	8	I = ABCDG = ABEFH = ...

6. Conclusions

In this paper we have considered the problem of designing products that are robust to variation that is due to environmental factors. This is an important issue since it is frequently impossible or inappropriate for the product manufacturer to control the variation that comes from the environment. Thus it is advantageous for the manufacturer to design a product that is insensitive to environmental variation. An issue of concern for the experimenter is the amount of experimental work that is usually required in the cross-product designs that Taguchi and others have advocated for robust product design.

The choice of an appropriate design depends on the experimental circumstances. Box and Draper (1987, p 502, 503) list a series of experimental circumstances that should be considered by the investigator when selecting a response surface design. Many of the same considerations apply to the conducting of experiments to design products that are robust to environmental variation. In this paper we have illustrated the fact that the choice of experimental design should also depend on the objective of the experiment. For example, the investigator who is only interested in obtaining a mean performance close to the ideal value will choose a different design to the investigator whose objective is to achieve low variability about the mean. The designs that we have given in this paper are appropriate when the objective of the experimenter is to minimize a weighted average of two performance measures; the one measures how good is the product design on the average when it is exposed to different environmental conditions, the other measures how much variation there is in the performance at different environmental conditions. Experimental designs that require fewer runs may be appropriate if the objective of interest is different to that of minimizing a weighted average of these two performance measures.

In this paper we have supposed, tentatively at least, that the model that approximates the underlying system over the (design $\times$ environmental) factor space is a second-order model in both the design and the environmental factors. For this model we have determined the coefficients that need to be estimated in order to optimize the various performance measures. We have given experimental designs that will yield data which, when the model is true, enable those coefficients to be estimated and so make it possible to optimize the desired performance measures.

It appears to us that the strategy developed in Box and Jones (1990a) and applied in this paper is more reasonable than the cross-product designs advocated by Taguchi. This approach asks the question, "What do we need to know?" The answer to this question helps determine the experimental designs that should be used. In many cases the designs that are appropriate will require considerably fewer runs than the cross-product designs.

Acknowledgement

Sponsored by the National Science Foundation Grant No. DDM-8808138, and by the Vilas Trust of the University of Wisconsin-Madison.

References

Addelman, S. (1961), "Irregular fractions of the 2^n factorial experiments," *Technometrics*, **3**, 479-496

Box, G. E. P. and D. W. Behnken (1960), "Some new three-level designs for the study of quantitative variables," *Technometrics*, **2**, 455-475.

Box, G. E. P. and N. R. Draper (1987), *Empirical Model-Building and Response Surface Methods*, New York: Wiley.

Box, G. E. P. and J. S. Hunter (1957), "Multifactor experimental designs for exploring response surfaces," *Annals of Mathematical Statistics*, **28**, 195-241.

Box, G. E. P. and S. P. Jones (1990a), "Designing products to be robust to the environment," *(in preparation)*

Box, G. E. P. and S. P. Jones (1990b), "Robust product designs, part I: first-order models with design × environment interactions," *(in preparation)*

Box, G. E. P. and K. B. Wilson (1951), "On the experimental attainment of optimum conditions," *Journal of the Royal Statistical Society*, **B13**, 1-38, discussion 38-45.

Draper, N. R. and D. K. J. Lin (1990), "Connections between two-level designs of resolution III^* and V," *Technometrics*, to appear.

Fung, C. A. (1986), "Statistical topics in off-line quality control," Ph.D. thesis, Department of Statistics, University of Wisconsin-Madison.

Hartley, H. O. (1959), "Smallest composite designs for quadratic response surfaces," *Biometrics*, **15**, 611-624.

John, P. W. M. (1961), "Three-quarter replicates of 2^4 and 2^5 designs," *Biometrics*, **17**, 319-321.

John, P. W. M. (1962), "Three-quarter replicates of 2^n designs," *Biometrics*, **18**, 172-184.

John, P. W. M. (1966), "Augmenting 2^{n-1} designs," *Technometrics*, **8**, 469-480.

John, P. W. M. (1969), "Some non-orthogonal fractions of 2^n designs," *Journal of the Royal Statistical Society*, **B31**, 270-275.

John, P. W. M. (1971), *Statistical Design and Analysis of Experiments*, New York: Macmillan.

Kackar, R. N. (1985), "Off-line quality control, parameter design and the Taguchi method," *Journal of Quality Technology*, **17**, 176-188, discussion 189-209.

Koshal, R. S. (1933), "Application of the method of maximum likelihood to the improvement of curves fitted by the method of moments," *Journal of the Royal Statistical Society*, **A96**, 303-313.

McLean, R. A. and V. L. Anderson (1984), *Applied Factorial and Fractional Designs*, New York: Marcel Dekker.

Taguchi, G. (1986), *Introduction to Quality Engineering: Designing Quality Into Products and Processes*, White Plains, NY: Kraus International Publications.

Taguchi, G. and M. S. Phadke (1984), "Quality engineering through design optimization," *Conference Record*, **3**, IEEE GLOBECOM 1984 Conference, 1106-1113.

Taguchi, G. and Y. Wu (1980) *Introduction to Off-Line Quality Control*, Nagaya, Japan: Central Japan Quality Control Association.

Westlake, W. J. (1965), "Composite designs based on irregular fractions of factorials," *Biometrics*, **21**, 324-336.