Santner, Pan:

The Use of Subset Selection in Combined Array Experiments to Determine Optimal Product or Process Designs. (REVISED, June 1997)

Sonderforschungsbereich 386, Paper 51 (1996)

Online unter: http://epub.ub.uni-muenchen.de/

1

The Use of Subset Selection in Combined-Array Experiments to Determine Optimal Product or Process Designs

Thomas Santner[1,2] **and Guohua Pan**[3]
[1] *Department of Statistics*
Ohio State University
Columbus, OH, 43210

[3] *Department of Mathematical Sciences*
Oakland University
Rochester, MI 48309

Abstract.

In the quality control literature, a number of authors have advocated the use of combined-arrays in screening experiments to identify robust product designs or robust process designs [Shoemaker, Tsui, and Wu (1991); Nair *et al.* (1992); Myers, Khuri, and Vining (1992), among others]. This paper considers product design and process design applications in which there are one or more "control" factors that can be modified by the manufacturer, and one or more "environmental" (or "noise") factors that that vary under field or manufacturing conditions. We show how Gupta's subset selection philosophy can be implemented in such a setting to identify optimal combinations of the levels of the control factors [Gupta (1956, 1965)]. By optimal, we mean those settings of the control factors that yield product designs whose performance is the most robust to variations in environmental factors. For process designs, the optimal settings of the control factors yield a fabrication method whose product quality is as nearly independent as possible to variations in the uncontrollable manufacturing factors, for example to daily temperature fluctuations.

Keywords and phrases: Combined-array, Inner array, Maxmin approach, Outer array, Product-array, Quality improvement, Response model, Screening, Subset selection, Variance reduction.

[2]Supported in part by Sonderforschungsbereich 386, Ludwig-Maximilians-Universität München, 80539 München

The authors would like to thank an anonymous referee for comments that helped improve this paper.

1.1 Introduction

In his pioneering work on product and process improvement, Taguchi (1986) emphasizes two types of factors that effect product quality: "control factors" are those variables that can be (easily) manipulated by the manufacturer and "noise factors" are those variables that represent either different environmental conditions that affect the performance of a product in the field *or* (uncontrollable) variability in component parts or raw materials that affect the performance of an end-product. For experiments to improve product or process design, Taguchi advocates using statistical designs that are products of highly fractionated orthogonal arrays in the control and noise factors. In the case of product design, for example, the goal of such experiments is to determine conditions under which the mean product quality is independent of the noise factors. While some of Taguchi's proposals have been controversial [Box (1988)], the basic viewpoint that he advocates has been applied widely and with many successes [Taguchi and Phadke (1984)].

A number of authors have proposed statistical refinements to the Taguchi methodology [Shoemaker, Tsui, and Wu (1991); Nair *et al.* (1992); Myers, Khuri, and Vining (1992), for example]. One of these proposals is to use combined-arrays in the control and noise factors to design quality improvement experiments rather than Taguchi's product-arrays. At the expense of confounding higher-order interactions, carefully chosen combined-arrays allow the experimenter to determine interactions among the control factors and interactions among the noise factors, as well as the critical control factor by noise factor interactions that allow one to minimize the effect of noise factors in product quality. A second proposal is to apply response surface methodology to the combined-array data to identify parsimonious models for the quality characteristic(s) of interest; these models can be used to select the levels of the control factors.

This paper shows how the subset selection philosophy introduced by Gupta (1956, 1965) can be fruitfully used to screen for control factor combinations in quality improvement experiments based on data analytic models. Bechhofer, Santner, and Goldsman (1995) give an overview of this field and present procedures to accomplish other important experimental goals. We focus on the case where the quality control characteristic of interest is to be *maximized*. Section 1.2 introduces a study typical of those in the food industry to improve the quality of a cake mix recipe by manipulating its ingredients, the control variables, when the baking is performed under a variety of time and oven temperature conditions, the noise variables. The procedure is developed based on a model relating the control and noise variables to the results of a taste test from a 2^{5-1} combined-array experiment. Section 1.3 introduces the subset selection procedure proposed for identifying the recipe that maximizes the minimum mean taste test response where the minimum is taken over the levels of the oven

temperature × baking time variables. The critical value required to implement the procedure is determined. Some generalizations and caveats are presented in the final section.

1.2 The data

We now describe an experiment whose goal is to improve the taste of a cake mix that consumers bake under conditions that can vary from the directions printed on the cakebox. There are three control factors in this experiment: the amounts of Flour, Shortening, and Egg used in the cake mix which will be denoted by F, S, and E throughout. There are two noise factors corresponding to variations of the baking temperature, denoted by T, and baking time, denoted by Z ("Zeit" is "time" in German). The response is a taste test, denoted by Y, with values ranging from 1 to 8; larger values correspond to a better tasting product.

The variables considered in this illustrative experiment are the same as those involved in a cake mix example described by Box and Jones (1992). However, the papers have different goals and use different experimental designs. Box and Jones illustrate the use of ANOVA methods based on a split-plot design while this paper developes a screening procedure for the optimal product design based on a completely randomized 2_V^{5-1} (combination-array) fractional factorial. Screening procedures to determine optimal product designs for complete split-plot designs are studied in Pan and Santner (1996).

Specifically, we use the 2_V^{5-1} design with defining contrast $I = FSETZ$. The response and treatment combinations are listed in Table 1.1. Let

$$\mathcal{D} \equiv \{(i,j,k,\ell,m) \text{ in the } 2_V^{5-1} \text{ design of Table 1.1}\}. \tag{1.1}$$

For this experimental design, the main effects are confounded with the 4-way interactions and the 2-way interactions are confounded with 3-way interactions.

The fitted coefficients based on the model with all main effects and 2-way interactions are listed in Table 1.2 and a normal probability plot is shown in Figure 1.1. The normal probability plot suggests that only the shortening × temperature interaction, and the main effects of flour, egg, temperature, and time are important.

Thus we illustrate the procedure of Section 1.3 for the hierarchical model that includes the shortening × temperature interaction and the significant main effects

$$Y_{ijk\ell m} = \mu_{ijk\ell m} + \epsilon_{ijklm} \quad (i,j,k,\ell,m) \in \mathcal{D} \tag{1.2}$$

where we assume

$$\mu_{ijk\ell m} = \mu_0 + F_i + S_j + E_k + T_\ell + Z_m + (ST)_{j\ell} \tag{1.3}$$

Table 1.1: Results of 2_V^{5-1} Cake Tasting Experiment with Defining Fraction $I = FSETZ$

Flour (F)	Shortening (S)	Egg (E)	Temp (T)	Time (Z)	Taste (Y)
0	0	0	1	0	1.6
0	0	0	0	1	1.2
1	0	0	0	0	2.2
1	0	0	1	1	6.5
0	1	0	0	0	1.3
0	1	0	1	1	1.7
1	1	0	1	0	3.5
1	1	0	0	1	3.8
0	0	1	0	0	1.6
0	0	1	1	1	4.4
1	0	1	1	0	6.1
1	0	1	0	1	4.9
0	1	1	1	0	2.4
0	1	1	0	1	2.6
1	1	1	0	0	5.2
1	1	1	1	1	6.0

holds for all observed and unobserved treatment combinations i, j, k, ℓ, and m. The terms F_i, S_j, E_k, T_ℓ, and Z_m are the flour, shortening, egg, temperature, and time main effects, respectively. As usual, the sum of the parameters over any subscript is zero and ϵ_{ijklm} are independent and $N(0, \sigma^2)$ distributed.

For Model 1.3, Table 1.2 lists, in boldface, the estimated coefficients and tests of the null hypotheses that individual coefficients are zero. The estimated σ^2 is $s^2 = (0.4989)^2$ based on 9 degrees of freedom. The procedure of Section 1.3 will be applied to Model 1.3.

1.3 The Procedure

Consider a generic setup in which there are r combinations of the levels of the control factor(s) and c combinations of the levels of the noise factor(s). For the cake tasting example introduced in Section 1.2, $r = 8 = 2^3$ and $c = 4 = 2^2$. In the following, we regard the r and c combinations as the number of levels of a single (composite) control factor and a single (composite) noise factor.

In this setup, suppose that the quality measure at ith level of the control variable and jth level of the noise variable is μ_{ij} which satisfies

$$\mu_{ij} = \boldsymbol{x}_{ij}^\top \boldsymbol{\beta} \quad (1 \le i \le r, 1 \le j \le c) \tag{1.4}$$

Table 1.2: Estimated Effects for the Fully Saturated Quadratic Model Based on the Data in Table 1.1. The Rows in Boldface Give, for the Reduced Model (1.2)–(1.3), the Estimated Effects and P-values for Individual Tests that Coefficients are zero. Based on (1.2)–(1.3), the Estimated Standard Error of each Boldface Coefficient is 0.1247

Term	Estimated Coefficient	P-value
(Intercept)	**3.4375**	**0.0000**
flour	**1.3375**	**0.0000**
shortening	**-0.125**	**0.3424**
egg	**0.7125**	**0.0003**
temp	**0.5875**	**0.0011**
time	**0.45**	**0.0057**
flour:shortening	-0.025	
flour:egg	0.0625	
flour:temp	0.1625	
flour:time	0.075	
shortening:egg	0.025	
shortening:temp	**-0.5**	**0.0031**
shortening:time	-0.2375	
egg:temp	-0.0125	
egg:time	-0.125	
temp:time	0.175	

where $\boldsymbol{x}_{ij}$ is a vector of known covariates. The quantity $\xi_i = \min_{1\le j\le c}\{\mu_{ij}\}$ measures the worst case (mean) performance of the ith level of the control variable against the noise variable. Let

$$\xi_{[1]} \le \cdots \le \xi_{[r]}$$

denote the ordered ξ_i values.

Suppose that we observe

$$Y_{ij} = \mu_{ij} + \epsilon_{ij} \quad (i,j) \in \mathcal{D}^\star \tag{1.5}$$

where $\mathcal{D}^\star$ is a subset of the direct product $\{1,\ldots,r\} \times \{1,\ldots,c\}$ and the $\{\epsilon_{ij}\}$ are independent mean zero normal random variables with common, unknown variance σ^2. We assume that the design is such that the least squares estimator of $\boldsymbol{\beta}$ based on (1.4) exists for data $\{Y_{ij}|(i,j) \in \mathcal{D}^\star\}$. Throughout we subscript (and regard) probabilities involving the Y_{ij} as functions of $r \times c$ matrix of means $\boldsymbol{\mu} = (\mu_{ij})$ since our focus is selection in terms of means; we could equally well subscript them by $\boldsymbol{\beta}$.

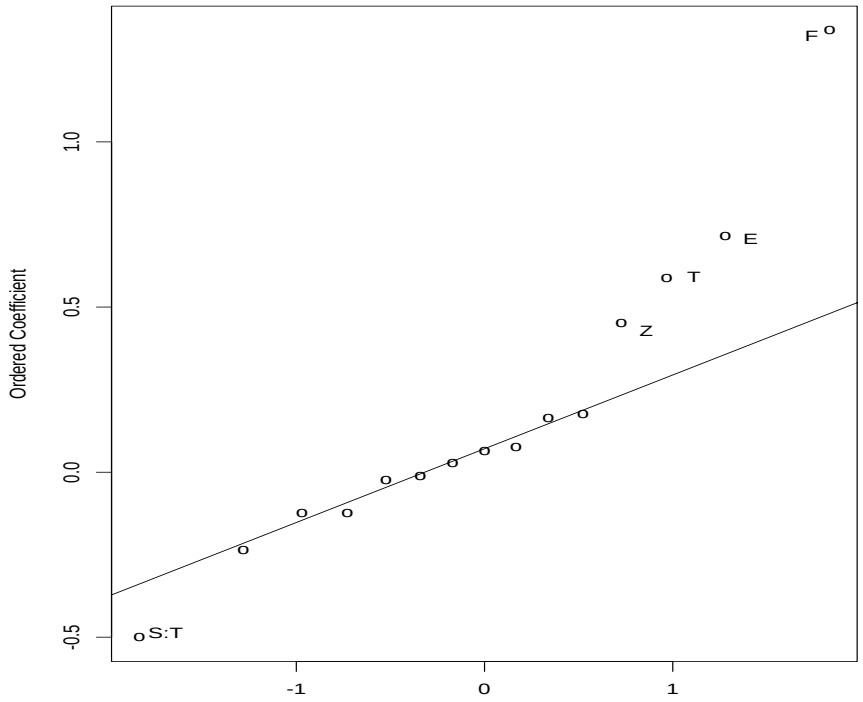

Figure 1.1: Normal Probability Plot of the 15 Estimated Coefficients in Table 1.2

Based on observations from (1.5) for the combined-array experiment, our goal is to screen the levels of the control variable by selecting a subset of the control levels $\{1, \ldots, r\}$ so as to contain the optimum control level, $[r]$. We wish to achieve this goal subject to the following performance requirement of the procedure.

Confidence Requirement: Given α with $0 < \alpha < 1$, we desire that

$$P_{\boldsymbol{\mu}}\{CS\} \geq 1 - \alpha \tag{1.6}$$

for all $\boldsymbol{\mu}$ satisfying Model (1.4) where $[CS]$ denotes the event that the selected subset contains the control level associated with $\xi_{[r]}$.

To achieve the design goal (1.6), we use the LS estimator of $\boldsymbol{\beta}$ based on (1.4)–(1.5), denoted by $\widehat{\boldsymbol{\beta}}$, to form $\widehat{\mu}_{ij} = \boldsymbol{x}_{ij}^{\top}\widehat{\boldsymbol{\beta}}$ $(1 \leq i \leq r, 1 \leq j \leq c)$ and then estimate ξ_i by

$$\widehat{\xi}_i = \min_{1 \leq j \leq c}\{\widehat{\mu}_{ij}\} \quad (1 \leq i \leq r).$$

Let

$$\widehat{\xi}_{[1]} \leq \cdots \leq \widehat{\xi}_{[r]}$$

denote the ordered $\widehat{\xi}_i$. Also suppose that an estimator of σ^2 is available, denoted by S^2, for which $\nu S^2/\sigma^2$ has a chi-square distribution with known degrees of freedom ν and whose distribution is independent of $\{\widehat{\mu}_{ij}\}$. We propose the following procedure to select levels of the control factor.

Procedure $\mathcal{G}$: For given α, $0 < \alpha < 1$, select control level i if and only if

$$\widehat{\xi}_i \geq \widehat{\xi}_{[r]} - h\, S$$

where h is chosen so that

$$\min P_{\boldsymbol{\mu}}\{\widehat{\xi}_{(r)} \geq \widehat{\xi}_{[r]} - h\, S\} \geq 1 - \alpha, \tag{1.7}$$

with $\widehat{\xi}_{(r)}$ denoting the estimator associated with $\xi_{[r]}$, and the minimum being taken over all $\boldsymbol{\mu} = (\mu_{ij})$ that satisfy (1.4).

In general, the choice of the constant h depends on the least favorable configuration, i.e., the $\boldsymbol{\mu}$ that minimizes the probability of correct selection in (1.7), which in turns depends on the specific model used in the experiment. As an example of the general methodology, this paper analyzes the Box and Jones cake mix data based on Model (1.2)–(1.3). As noted above, in this case the generic index i that identifies a control level of the mean is the triple (i, j, k) with $r = 8$ values and the generic j that identifies a noise level of the mean is (ℓ, m) with $c = 4$ values; also

$$\xi_{ijk} = \min\{\mu_{ijk00}, \mu_{ijk01}, \mu_{ijk10}, \mu_{ijk11}\}.$$

Later we will require a more explicit expression for ξ_{ijk} which is straightforward to determine for Model (1.2) as

$$\begin{aligned} \xi_{ijk} &= F_i + E_k + \mu_0 + S_j + \min_{\ell,m}\{T_\ell + Z_m + (ST)_{j\ell}\} \\ &= F_i + E_k + \mu_0 + S_j - |Z_1| + \min_{\ell}\{T_\ell + (ST)_{j\ell}\} \end{aligned} \tag{1.8}$$

where (1.8) holds because $\min_{(a,b)\in A\times B}\{s_a + t_b\} = \min_{a\in A}\{s_a\} + \min_{b\in B}\{t_b\}$ and $\min_m\{Z_m\} = -|Z_1|$.

Some notation will be required to describe and analyze Procedure $\mathcal{G}$ for Model (1.2). Throughout, it will be convenient to let $\boldsymbol{\mu}$ denote the 8×4 matrix with $(i, j, k)th$ row $(\mu_{ijk00}, \mu_{ijk01}, \mu_{ijk10}, \mu_{ijk11})$ and with row elements arranged so that the means corresponding to $S = 0$ are first and lexicographically according to (F, E) within each S level. Every ξ_{ijk} is the minimum of one row of $\boldsymbol{\mu}$. We regard the data to be collected using experimental design $\mathcal{D}$, denoted by $\boldsymbol{Y}_{\mathcal{D}}$, as arranged conformably with $\boldsymbol{\mu}$ but with missing data in the positions where the design of Table 1.1 collects no observations. For example, using lower case notation for observed values, we have

$$
\underset{8\times 4}{\boldsymbol{y}_{\mathcal{D}}} = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \\ (0,0) \quad (0,1) \quad (1,0) \quad (1,1) \\ (T,Z) \end{array}
\begin{array}{c} (F,S,E) \\ \left[\begin{array}{cccc} — & 1.2 & 1.6 & — \\ 1.6 & — & — & 4.4 \\ 2.2 & — & — & 6.5 \\ — & 4.9 & 6.1 & — \\ 1.3 & — & — & 1.7 \\ — & 2.6 & 2.4 & — \\ — & 3.8 & 3.5 & — \\ 5.2 & — & — & 6.0 \end{array}\right] \\ \begin{array}{c} (0,0,0) \\ (0,0,1) \\ (1,0,0) \\ (1,0,1) \\ (0,1,0) \\ (0,1,1) \\ (1,1,0) \\ (1,1,1) \end{array} \end{array}
$$

is the data in Table 1.1 arranged in this manner for $\mathcal{D}$. The phrase "$\boldsymbol{Y}_{\mathcal{D}}$ has mean $\boldsymbol{\mu}$" expresses the fact that the components of $\boldsymbol{Y}_{\mathcal{D}}$ have means given by the corresponding element of $\boldsymbol{\mu}$.

The estimated ξ_{ijk} are

$$\widehat{\xi}_{ijk} = \min\{\widehat{\mu}_{ijk00}, \widehat{\mu}_{ijk01}, \widehat{\mu}_{ijk10}, \widehat{\mu}_{ijk11}\} \tag{1.9}$$

where

$$\widehat{\mu}_{ijk\ell m} = \widehat{\mu}_0 + \widehat{F}_i + \widehat{S}_j + \widehat{E}_k + \widehat{T}_\ell + \widehat{Z}_m + \widehat{(ST)}_{j\ell}$$

and the components are estimated by least squares based on (1.2). Explicitly, we have $\widehat{\mu}_0 = \overline{Y}_{\cdots\cdots}$, $\widehat{F}_i = \overline{Y}_{i\cdots} - \overline{Y}_{\cdots\cdots}$, $\widehat{S}_j = \overline{Y}_{\cdot j\cdots} - \overline{Y}_{\cdots\cdots}$, $\widehat{E}_k = \overline{Y}_{\cdot\cdot k\cdot\cdot} - \overline{Y}_{\cdots\cdots}$, $\widehat{T}_\ell = \overline{Y}_{\cdots\ell\cdot} - \overline{Y}_{\cdots\cdots}$, $\widehat{Z}_m = \overline{Y}_{\cdots\cdot m} - \overline{Y}_{\cdots\cdots}$, and $\widehat{(ST)}_{j\ell} = \overline{Y}_{\cdot j\ell\cdot} - \overline{Y}_{\cdot j\cdots} - \overline{Y}_{\cdots\ell\cdot} + \overline{Y}_{\cdots\cdots}$. *Here the averages are taken over the observations in the design* $\mathcal{D}$. Thus

$$\begin{aligned} \widehat{\mu}_{ijk\ell m} &= -3\ \overline{Y}_{\cdots\cdots} + \overline{Y}_{i\cdots\cdot} + \overline{Y}_{\cdot\cdot k\cdot\cdot} + \overline{Y}_{\cdots\cdot m} + \overline{Y}_{\cdot j\cdot\ell\cdot} \\ &= \mu_{ijk\ell m} - 3\ \overline{\epsilon}_{\cdots\cdots} + \overline{\epsilon}_{i\cdots\cdot} + \overline{\epsilon}_{\cdot\cdot k\cdot\cdot} + \overline{\epsilon}_{\cdots\cdot m} + \overline{\epsilon}_{\cdot j\cdot\ell\cdot} \end{aligned} \tag{1.10}$$

where $\mu_{ijk\ell m}$ is given by (1.3). The estimated $\widehat{\mu}_{ijk\ell m}$ for each control and noise factor combination are listed in Table 1.3 together with $\widehat{\xi}_{ijk}$ based on the four $T \times Z$ baking combinations.

To achieve confidence level $100\times(1-\alpha)\%$, procedure $\mathcal{G}$ selects those (F,S,E) combinations i, j, k for which

$$\widehat{\xi}_{ijk} \geq \max_{i^\star, j^\star, k^\star} \widehat{\xi}_{i^\star j^\star k^\star} - h\, S \tag{1.11}$$

Table 1.3: Estimated Control Variable Means and Row Minimums for each (F, S, E) Combination

	(T,Z)				
(F,S,E)	(0,0)	(0,1)	(1,0)	(1,1)	$\widehat{\xi}_{F,S,E}$
(0,0,0)	0.0	0.9	2.2	3.1	0.0
(0,0,1)	1.4	2.3	3.6	4.5	1.4
(1,0,0)	2.6	3.6	4.8	5.7	2.6
(1,0,1)	4.1	5.0	6.3	7.2	4.1
(0,1,0)	0.7	1.6	0.9	1.8	0.7
(0,1,1)	2.2	3.1	2.3	3.2	2.2
(1,1,0)	3.4	4.3	3.6	4.5	3.4
(1,1,1)	4.8	5.7	5.0	5.9	4.8

where the maximum is over $(i^\star, j^\star, k^\star) \in \{0,1\}^3$ and h is given by Equation (1.12) in Theorem 1.3.1. This result identifies the least favorable configuration (LFC) of the means for Procedure $\mathcal{G}$ and gives an expression for the probability of correct selection at this configuration.

Theorem 1.3.1 The procedure $\mathcal{G}$ attains confidence level $100 \times (1-\alpha)\%$ for Model (1.2)–(1.3) if h is the solution of the equation

$$P_{\boldsymbol{\mu}_{LFC}}\left\{\widehat{\xi}_{111} \geq \max_{i,j,k} \widehat{\xi}_{ijk} - h\sqrt{W/9}\right\} = 1-\alpha \tag{1.12}$$

where $\boldsymbol{Y}_{\mathcal{D}}$ has mean $\boldsymbol{\mu}_{LFC} = \begin{pmatrix} 0 & +\infty \\ 0 & 0 \end{pmatrix} \otimes J_{4\times 2}$ with $\otimes$ denoting Kronecker product, $J_{4\times 2}$ is the 4×2 matrix of 1's, the variance is unity of each $Y_{ijk\ell m}$ with $(i,j,k,\ell,m) \in \mathcal{D}$, and W has a chi-squared distribution with 9 degrees of freedom that is independent of $\boldsymbol{Y}_{\mathcal{D}}$.

Before proving this result, note that $\xi_{ijk} = 0$ for all (i,j,k) under $\boldsymbol{\mu}_{LFC}$ given in the theorem. Hence we also have $\max_{i,j,k} \xi_{ijk} = 0$. This result is similar to many others in the subset selection literature concerning least favorable configurations—the parameter of interest is identical under the LFC.

The minimum of the probability of correct selection corresponds to the event that the sample estimator for the *last row*, $\widehat{\xi}_{111}$, is greater than the remaining seven $\widehat{\xi}_{ijk}$. This is different from many other (simple) problems in which the estimators of the parameters of interest are often independent and identically distributed; in the latter case, one computes the minimum probability over the parameter space as the probability that the estimator of any specific parameter exceeds the remaining parameter estimators by the yardstick for that problem.

In this example, the marginal distributions of the $\widehat{\xi}_{ijk}$ are not all the same. The reason for focusing on $\widehat{\xi}_{111}$ will be demonstrated in the proof.

Proof of Theorem 1.3.1: By Lemma .0.1 in the Appendix, it suffices to determine the infimum over configurations of the form

$$\begin{pmatrix} v_{00} & v_{01} \\ v_{10} & v_{11} \end{pmatrix} \otimes J_{4\times 2} = \boldsymbol{v} \otimes J_{4\times 2}, \tag{1.13}$$

say, where $\boldsymbol{v} \in \mathbb{R}^{2\times 2}$. Thus given $\boldsymbol{\mu}$ with components (1.13), we have $\mu_{ijk\ell m} = v_{j\ell}$ and $\xi_{ijk} = \min\{v_{j0}, v_{j1}\}$. Assuming, without loss of generality, that $v_{10} = \min\{v_{10}, v_{11}\} \geq \min\{v_{00}, v_{01}\} = v_{00}$, we also have $\xi_{111} = \max\{\xi_{ijk}\}$ and thus treatment combination $(1,1,1)$ is best. (Formally the row labeled $(1,1,1)$ is the unique best row corresponding to $\boldsymbol{\mu}$ if we add ϵ to each element in this row and the argument below holds in the limit as $\epsilon \downarrow 0$).

Also we have by the definition of $\widehat{\xi}_{ijk}$ and by expression (1.10) that

$$\begin{aligned}
\widehat{\xi}_{ijk} &= \min_{\ell,m} \widehat{\mu}_{ijk\ell m} \\
&= -3\ \overline{\epsilon}_{\cdot\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{i\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot k\cdot\cdot} + \min_{\ell,m}\{v_{j\ell} + \overline{\epsilon}_{\cdot\cdot\cdot\cdot m} + \overline{\epsilon}_{\cdot j\cdot\ell\cdot}\} \\
&= -3\ \overline{\epsilon}_{\cdot\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{i\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot k\cdot\cdot} + \min_{m}\{\overline{\epsilon}_{\cdot\cdot\cdot\cdot m}\} + \min_{\ell}\{v_{j\ell} + \overline{\epsilon}_{\cdot j\cdot\ell\cdot}\}.
\end{aligned}$$

Then the PCS is given by

$$\begin{aligned}
P_{\boldsymbol{\mu}}\{CS\} &= P_{\boldsymbol{\mu}}\left\{\widehat{\xi}_{111} \geq \widehat{\xi}_{ijk} - h\,S \ \ \forall\ \ (i,j,k)\right\} \\
&= P_{\boldsymbol{\mu}}\{\overline{\epsilon}_{1\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot 1\cdot\cdot} \geq \overline{\epsilon}_{i\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot k\cdot\cdot} - h\,S\ \forall\ (i,k); \\
&\qquad \overline{\epsilon}_{1\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot 1\cdot\cdot} + \min\{v_{10} + \overline{\epsilon}_{\cdot 1\cdot 0\cdot}, v_{11} + \overline{\epsilon}_{\cdot 1\cdot 1\cdot}\} \\
&\qquad \geq \overline{\epsilon}_{i\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot k\cdot\cdot} + \min\{v_{00} + \overline{\epsilon}_{\cdot 0\cdot 0\cdot}, v_{01} + \overline{\epsilon}_{\cdot 0\cdot 1\cdot}\} - h\,S\ \forall\ (i,k)\} \\
&= P_{\boldsymbol{\mu}}\{\overline{\epsilon}_{1\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot 1\cdot\cdot} \geq \overline{\epsilon}_{i\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot k\cdot\cdot} - h\,S\ \forall\ (i,k); \\
&\qquad \overline{\epsilon}_{1\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot 1\cdot\cdot} + \min\{\overline{\epsilon}_{\cdot 1\cdot 0\cdot}, v_{11} - v_{10} + \overline{\epsilon}_{\cdot 1\cdot 1\cdot}\} + v_{10} \\
&\qquad \geq \overline{\epsilon}_{i\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot k\cdot\cdot} + \min\{\overline{\epsilon}_{\cdot 0\cdot 0\cdot}, v_{01} - v_{00} + \overline{\epsilon}_{\cdot 0\cdot 1\cdot}\} + v_{00} - h\,S\ \forall\ (i,k)\} \\
&\geq P_{\boldsymbol{\mu}}\{\overline{\epsilon}_{1\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot 1\cdot\cdot} \geq \overline{\epsilon}_{i\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot k\cdot\cdot} - h\,S\ \forall\ (i,k); \\
&\qquad \overline{\epsilon}_{1\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot 1\cdot\cdot} + \min\{\overline{\epsilon}_{\cdot 1\cdot 0\cdot}, \overline{\epsilon}_{\cdot 1\cdot 1\cdot}\} + v_{10} \\
&\qquad \geq \overline{\epsilon}_{i\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot k\cdot\cdot} + \min\{\overline{\epsilon}_{\cdot 0\cdot 0\cdot}, v_{01} - v_{00} + \overline{\epsilon}_{\cdot 0\cdot 1\cdot}\} \\
&\qquad + v_{00} - h\,S\ \forall\ (i,k)\} \qquad (1.14) \\
&\geq P_{\boldsymbol{\mu}}\{\overline{\epsilon}_{1\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot 1\cdot\cdot} \geq \overline{\epsilon}_{i\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot k\cdot\cdot} - h\,S\ \forall\ (i,k); \\
&\qquad \overline{\epsilon}_{1\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot 1\cdot\cdot} + \min\{\overline{\epsilon}_{\cdot 1\cdot 0\cdot}, \overline{\epsilon}_{\cdot 1\cdot 1\cdot}\} + v_{10} \\
&\qquad \geq \overline{\epsilon}_{i\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot k\cdot\cdot} + \overline{\epsilon}_{\cdot 0\cdot 0\cdot} + v_{00} - h\,S\ \forall\ (i,k)\} \qquad (1.15) \\
&\geq P_{\boldsymbol{\mu}}\{\overline{\epsilon}_{1\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot 1\cdot\cdot} \geq \overline{\epsilon}_{i\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot k\cdot\cdot} - h\,S\ \forall\ (i,k); \\
&\qquad \overline{\epsilon}_{1\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot 1\cdot\cdot} + \min\{\overline{\epsilon}_{\cdot 1\cdot 0\cdot}, \overline{\epsilon}_{\cdot 1\cdot 1\cdot}\} \\
&\qquad \geq \overline{\epsilon}_{i\cdot\cdot\cdot\cdot} + \overline{\epsilon}_{\cdot\cdot k\cdot\cdot} + \overline{\epsilon}_{\cdot 0\cdot 0\cdot} - h\,S\ \forall\ (i,k)\} \qquad (1.16)
\end{aligned}$$

where (1.14) holds since $v_{11} \geq v_{10}$ implies that $\min\{\overline{\epsilon}_{\cdot 1 \cdot 0 \cdot}, v_{11} - v_{10} + \overline{\epsilon}_{\cdot 1 \cdot 1 \cdot}\} \geq \min\{\overline{\epsilon}_{\cdot 1 \cdot 0 \cdot}, \overline{\epsilon}_{\cdot 1 \cdot 1 \cdot}\}$, (1.15) holds because $\overline{\epsilon}_{\cdot 0 \cdot 0 \cdot} \geq \min\{\overline{\epsilon}_{\cdot 0 \cdot 0 \cdot}, v_{01} - v_{00} + \overline{\epsilon}_{\cdot 0 \cdot 1 \cdot}\}$ for all $0 \leq v_{01} - v_{00} < +\infty$, and (1.16) holds because $v_{10} \geq v_{00}$. But

$$(1.16) = P_{\boldsymbol{\mu}_{LFC}} \left\{ \widehat{\xi}_{111} \geq \widehat{\xi}_{ijk} - h\, S \quad \forall \quad (i,j,k) \right\}$$

After division of each $\widehat{\xi}_{ijk}$ by σ, we obtain (1.12) as the infimum of the PCS over all $\boldsymbol{\mu}$ satisfying (1.2). $\square$

The critical point h is computed as the $1-\alpha$ percentile of the distribution

$$V = \sqrt{9}(\widehat{\xi}_{111} - \max_{ijk} \widehat{\xi}_{ijk})/\sqrt{W}$$

where $W \sim \chi^2_9$ and each $\widehat{\xi}_{ijk}$ is calculated from (1.9) based on $\boldsymbol{Y}_{\boldsymbol{\mathcal{D}}}$ with the mean and variance structure of Theorem 1.3.1. Based on this description, we used simulation to calculate that $h = 1.344$ for our application when $\alpha = .05$. Thus the subset selection procedure selects those treatment combinations for which

$$\widehat{\xi}_{ijk} \geq 4.8 - 1.344 \times (.4989) = 4.8 - .67 = 4.13$$

which results in the *single combination* $(F, S, E) = (1, 1, 1)$ being selected (flour, shortening, and egg all at their high levels).

1.4 Discussion

Like many multiple comparison procedures for linear models, the confidence level of the procedure proposed in this paper is conditional based on the validity of the data analytic model determined in the first stage of the analysis. The procedure is fully justified if the model can be determined a priori by physical considerations or from a pilot data set. Its unconditional use with a single data set requires that a (complicated) two-stage stage procedure be developed that describes the probabilistic choice of data analytic model, including the subjective graphical aspects of the process, given a true model as well as the subsequent selection process for each possible model.

This paper develops a subset selection procedure for a specific model that describes the means of a factorial experiment. A general theory that will allow the determination of the least favorable configuration for arbitrary models is under development. Equally important, software that allows experimenters to easily compute critical values for such procedures will be required.

In practice, despite the fact that the combined-array fractional factorial experiments studied in this paper have relatively few runs compared to their full factorial versions, many experimenters may desire to use some form of randomization restriction in conducting the experiment. Typically, the use of randomization restriction will complicate the appropriate model by adding

interaction terms. For example, Pan and Santner (1996) analyze procedures for selecting and screening best control treatments using complete factorial experiments run according to a split-plot design. It would also be possible to analyze such a procedure in this setting.

If the model used for the subset selection procedure does not permit degrees of freedom for estimating σ^2, then the experimenter has to either know σ^2 or have available an independent (chi-square) estimate from other sources. Subset selection cannot be performed unless one of these circumstances holds.

This paper has restricted attention to the case where the experimenter wishes to maximize a quality characteristic. In other applications, it may be desired to move the process to a target value τ_0, say, subject to minimizing the product variability about τ_0. In the latter case it may be more appropriate to screen control treatments to find those treatments i having small values of $s_i^2 = \sum_{j=1}^{n_e}(\mu_{ij} - \tau_0)^2$ or some other measure of spread about the target. For normally distributed data, the former problem amounts to selection in terms of a non-central chi-square random variable and can be solved using techniques similar to those of Alam and Rizvi (1966).

Appendix

Let Ω denote the set of $\boldsymbol{\mu}$ satisfying (1.2) and Ω_0 be those $\boldsymbol{\mu} \in \Omega$ for which $F_i \overset{i}{=} 0$, $E_k \overset{k}{=} 0$, and $Z_m \overset{m}{=} 0$. Thus

$$\Omega_0 \equiv \left\{ \boldsymbol{\mu} = \begin{pmatrix} v_{00} & v_{01} \\ v_{10} & v_{11} \end{pmatrix} \otimes J_{4\times 2} \;\middle|\; v_{ab} \quad \text{are real numbers,} \quad a,b = 0,1 \right\}$$

Lemma .0.1 Suppose $\boldsymbol{Y_{\mathcal{D}}}$ has mean $\boldsymbol{\mu}$ satisfying (1.2). Then Procedure $\mathcal{G}$ satisfies

$$\inf_{\Omega} P\{CS\} = \inf_{\Omega_0} P\{CS\} \tag{17}$$

Proof: Given $\boldsymbol{\mu} \in \Omega$ set

$$Y^*_{ijk\ell m} \equiv Y_{ijk\ell m} - F_i - E_k - Z_m$$

for $(i,j,k,\ell,m) \in \mathcal{D}$. Then $Y^*_{ijk\ell m} \sim N(\mu^*_{ijk\ell m} = \mu_0 + S_j + T_\ell + (ST)_{j\ell}, \sigma^2)$ and $\boldsymbol{\mu}^* = (\mu^*_{ijk\ell m}) \in \Omega_0$. It is straightforward to show that the ξ_{ijk} based on the $\mu_{ijk\ell m}$ and the $\xi^\star_{ijk}$ based on the $\mu^*_{ijk\ell m}$ are related by $\xi_{ijk} = F_i + E_k - |Z_1| + \xi^\star_{ijk}$. Similarly, the $\widehat{\xi}_{ijk}$ based on $\boldsymbol{Y_{\mathcal{D}}}$ and the $\widehat{\xi}^\star_{ijk}$ based on $Y^\star_{\mathcal{D}}$ satisfy $\widehat{\xi}_{ijk} = \widehat{F}_i + \widehat{E}_k - |\widehat{Z}_1| + \min_\ell\{\widehat{\mu}_0 + \widehat{S}_j + \widehat{T}_\ell + \widehat{(ST)}_{j\ell}\} = \widehat{\xi}^\star_{ijk} + F_i + E_k + |\widehat{Z}_1 - Z_1| - |\widehat{Z}_1|$ for all i, j, k, ℓ, and m. We assume without loss of generality that $F_1 = \max_i\{F_i\}$, $E_1 = \max_k\{E_k\}$, and $\min_\ell(S_1 + (ST)_{1\ell}) = \max_j \min_\ell(S_j + (ST)_{j\ell})$ so that

$$\max_{i,j,k}\{\xi_{ijk}\} = F_1 + E_1 - |Z_1| + \max_j \min_\ell\{\mu_0 + S_j + T_\ell + (S\times T)_{j\ell}\} = \xi_{111}$$

and $\xi^\star_{111} \geq \xi^\star_{ijk}$ for all i, j, and k. Then by definition,

$$\begin{aligned}
P_{\boldsymbol{\mu}}\{CS\} &= P_{\boldsymbol{\mu}}\{\widehat{\xi}_{111} \geq \widehat{\xi}_{ijk} - h\,S \;\;\forall\;\; (i,j,k)\} \\
&= P_{\boldsymbol{\mu}}\{\widehat{\xi}_{111} - F_1 - E_1 - |\widehat{Z}_1 - Z_1| + |\widehat{Z}_1| \\
&\qquad \geq\; \widehat{\xi}_{ijk} - F_i - E_k - |\widehat{Z}_1 - Z_1| + |\widehat{Z}_1| \\
&\qquad +\; F_i - F_1 + E_k - E_1 - h\,S \;\;\forall\;\; (i,j,k)\} \\
&\geq P_{\boldsymbol{\mu}}\{\widehat{\xi}^\star_{111} \geq \widehat{\xi}^\star_{ijk} - h\,S \;\;\forall\;\; (i,j,k)\} \qquad (18) \\
&= P_{\boldsymbol{\mu}^\star}\{\widehat{\xi}_{111} \geq \widehat{\xi}_{ijk} - h\,S \;\;\forall\;\; (i,j,k)\} \qquad (19)
\end{aligned}$$

where (18) holds because $F_i - F_1 \leq 0$ for $i = 0,1$ and $E_k - E_1 \leq 0$ for $k = 0,1$. □

References

Alam, K. and Rizvi, M. H. (1966). Selection from multivariate normal populations. *Annals Institute of Statistical Mathematics* **18**, 307-318.

Bechhofer, R. E., Santner, T. J. and Goldsman, D. M. (1995). *Designing Experiments for Statistical Selection, Screening, and Multiple Comparisons*, New York:J. Wiley & Sons.

Box, G. E. P. (1988). Signal-to-Noise ratios, performance criteria, and transformations. *Technometrics*, **30**, 1-17.

Box, G. E. P. and Jones, S. (1992). Split-plot designs for robust product experimentation. *Jour. Applied Statistics*, **19**, 3-26.

Gupta, S. S. (1956). *On A Decision Rule for a Problem in Ranking Mean.* Ph.D. Dissertation, Institute of Statistics, University of North Carolina, Chapel Hill, NC.

Gupta, S. S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics*, **7**, 225-245.

Myers, R. H., Khuri, A. I., and Vining, G. (1992). Response Surface Alternatives to the Taguchi Robust Parameter Design Approach, *American Statistician*, **46**, 131-139.

Nair, V. J. et al. (1992). Taguchi's parameter design: a panel discussion, *Technometrics*, **34**, 127-161.

Pan, G. and Santner, T. J. (1996). Selection and screening procedures to determine optimal product designs, submitted.

Shoemaker, A. C., Tsui, K., and Wu, C.-F. J. (1991). Economical experimental methods for robust design, *Technometrics*, **33**, 415-427.

Taguchi, G. (1986). *Introduction to Quality Engineering: Designing Quality into Products and Processes,* Tokyo: Asian Productivity Organization.

Taguchi, G. and Phadke, M. S. (1984). Quality engineering through design optimization, Atlanta: Conference Record, *GLOBECOM84 Meeting, IEEE Communications Society*, 1106-1113.