

TREATMENT DESIGN FOR FERTILIZER USE EXPERIMENTATION

Foster B. Cady
Reggie J. Laird



**CENTRO INTERNACIONAL DE MEJORAMIENTO DE MAIZ Y TRIGO
INTERNATIONAL MAIZE AND WHEAT IMPROVEMENT CENTER**
Londres 40 Apartado Postal 6-641 México 6, D. F., México

CONTENTS

	Page
ABSTRACT	1
BACKGROUND	2
A MODEL FOR FERTILIZER RESPONSE FUNCTIONS	4
PREDICTION ERRORS	8
Variance error	9
Bias error	10
A combined variance and bias error	13
CRITERIA FOR COMPARING TREATMENT DESIGNS	14
EMPIRICAL COMPARISON OF SELECTED DESIGNS	17
SUMMARY	25
RESUME	26
RESUMEN	28
REFERENCES	29

**TREATMENT DESIGN
FOR
FERTILIZER USE EXPERIMENTATION**

Foster B. Cady and Reggie J. Laird ¹

ABSTRACT

Ideally the experimenter would like to use treatment design to minimize both variance and bias problems in estimating yield from fertilizer application. Selecting a treatment design for variance considerations will not give a minimum bias design. Since different criteria of optimality lead to different designs, the most appropriate criterion under a given set of circumstances will be determined by the specific use that is made of the data generated by the design, e.g., in fertilizer use studies the slope of the estimated response surface near the economic optimal could be most important.

For the quadratic polynomial estimating function and a square root true function in two variables, treatment designs are compared empirically using several variance criteria in addition to a volume concept of bias. In general, bias is decreased by restricting the treatment combinations to an area of the factor space not including the boundaries. Increased variance of various estimated parameters may be controlled through replication of the relatively small treatment designs needed for bias considerations.

¹ Professor of Biological Statistics, Biometrics Unit, 337 Warren Hall, Cornell University, Ithaca, New York, 14850 and Soil Scientist, International Maize and Wheat Improvement Center, Apdo. Postal 6-641, Mexico 6, D. F., Mexico.

TREATMEN DESIGN FOR FERTILIZER USE EXPERIMENTATION

Foster B. Cady and Reggie J. Laird

Fertilizer experiments are conducted to obtain information on plant response to applications of the essential nutrients. Experiments generally consist of replicated fertilizer treatments where each treatment is a combination of rates of selected plant nutrients. A treatment design is a listing of treatments and may be visualized as points distributed in space. For example, combinations of several rates of nitrogen and phosphorus may be represented graphically as points in a two-dimensional factor space defined by axes corresponding to the nutrients.

Generally fertilizer experiments are limited to observing plant response to rates of the several nutrients that lie within discrete limits. For example, in studying corn response to fertilization, information may be desired for rates of nitrogen between the limits 0 and 300 kg/ha, rates of phosphorus from 0 to 150 kg/ha, and rates of potassium from 0 to 200 kg/ha. These limits define the factor space within which observations on corn response to applications of nitrogen, phosphorus, and potassium are measured and the most meaningful information is thought to be included.

The central question in constructing a treatment design is how a limited number of points should be distributed within a given factor space. Literally an infinite number of rate combinations lie within the limits established for the different elements. However, the estimate of the plant response to fertilization will depend, in part, on the selected rate combinations. The objective, then, of the investigator is to select that particular treatment design which will best obtain the specific information he desires on plant reaction to fertilization.

BACKGROUND

The selection of a treatment design presented no great problem in the early years of fertilizer use research. The plant nutrient factors were commonly varied, one at a time, with others held constant or not applied, and the experimental treatments generally consisted of a small number of equally or geometrically spaced fertilizer rates.

Following the publication by Yates (1937) of procedures for analyzing factorial experiments, the factorial arrangement of fertilizer treatments gradually became accepted as the most adequate distribution of points within a factor space. The treatments corresponding to a factorial consist of all combinations of the several factor levels. Commonly, the levels of the factors are evenly spaced so that the treatments are uniformly distributed within the space defined by the rate limits selected for study. The factorial arrangement of treatments permits the investigator to examine experimental results for interaction in addition to simple linear, quadratic and higher order main effects.

The total number of treatments required for factorial arrangements becomes prohibitively large as the number of factors and levels increases. With more than five levels of two factors (5^2) or three levels of three factors (3^3), for example, the size of the area needed for a single block containing each of the treatments often becomes so large that the precision of the experiment is greatly reduced. For certain factorials this deficiency can be eliminated by arranging the treatments in two or more blocks with certain higher order interactions partially or completely confounded with block effects. Large factorials have a second limitation in that the number of treatments in a single replication may be excessively large in comparison to the small number of effects to be estimated. Fractional factorials offer one way of reducing the total number of treatments without affecting appreciably the precision of certain estimates of fertilizer effects.

The postulate that factorial arrangement of treatments is most adequate for multi-factor experiments was challenged in 1951 by Box and Wilson. They assumed the adequacy of the second degree polynomial for representing the functional relationship, and proposed designs consisting of a carefully selected number of treatments only moderately greater than the number of effects to be estimated. A composite design for a three-factor experiment, for example, consists of only 15 treatments in comparison with the 27 required by a 3^3 factorial. The distribution of treatment combinations in a central composite design for a three-factor experiment is shown in Figure 8A.1 of Cochran and Cox (1957).

The question of optimal treatment design was explored further by Box and Hunter in a paper published in 1957. They introduced the concept of a variance function for an experimental design and proposed that the variance of a predicted yield should be constant at all points equidistant from the center of the factor space. Designs having this property, called rotatable designs, included a particular scaling of dispersion of the treatments and a spherical factor space.

A different and basically intuitive approach to the selection of a design with fewer treatments than the complete factorial arrangements has been employed in recent years. Generally, the investigator begins with a factorial arrangement and systematically eliminates treatment combinations throughout the factor space. The number of possible partial factorials corresponding to any given complete factorial arrangement is large, and no objective criteria have been established to assure the selection of the most efficient design for the specific objective at hand.

A MODEL FOR FERTILIZER RESPONSE FUNCTIONS

The "goodness" with which the results of a fertilizer experiment accomplish the objectives of an experimenter studying the relationship between crop yield and level of fertilization depends largely on: (i) the assumed model that is used to express the results in the form of a yield function and its "closeness" to the true function, and (ii) the magnitude of various random errors associated with doing the experiment.

Perhaps the first question which arises in considering the selection of a statistical model for representing fertilizer use data is whether the model should be discrete or continuous. If a discrete model is selected, then comparisons are limited to treatment means and the investigator chooses as optimal one of the combinations of fertilizer rates used as a treatment. If a continuous model is selected, the yield data are used to estimate the parameters corresponding to a specific mathematical function, yields may be predicted for any levels of the input variable, and optimal fertilizer rates are calculated directly from the resulting function. This may result in a predicted optimal fertilizer rate that is not a level directly applied in the experiment.

The interpretation given to earlier fertilizer use studies indicates that researchers were generally thinking in terms of a discrete statistical model. In 1909, however, Mitscherlich published a study in which he treated yield as a continuous function of level of fertilization and suggested a "diminishing returns" model for the representation of experimental results. Over the years investigators have continued to consider fertilizer use studies in terms of both discrete and continuous models, although preference for the latter has steadily increased. Basic studies on nutrient uptake and plant growth tend to corroborate the thesis that the true fertilizer response model is indeed continuous. However, since continuous models may assume an infinite variety of forms and laws governing plant growth have only partially been elucidated, it has not been possible through deductive processes to specify the true model characteristic of different production conditions. Consequently, in his choice of a continuous model the investigator accepts the fact that the functional representation of his results will be in error or biased to the extent that his assumed model differs from the true model. If he decides to use a discrete model he is unable to make inferences about rates of fertilization other than those actually employed in the experiment and he can only hope to roughly approximate the true economic optimal level of fertilization.

As discussed by Anderson (1956) and Hildreth (1956) there are several potential advantages in using a continuous model. The precision of the estimated crop yields, produced at different fertilization levels, may be substantially increased. The prediction equation provides a convenient means for calculating the optimal rate of fertilization. The cumulative experience of many years on fertilizer response, expressed in the form of an equation, should eventually be useful in defining the true response functions for specific productivity conditions.

Furthermore, fertilizer use researchers have developed a fairly reliable picture of the general relationship between yield and rate of fertilization for some crops. For example, it is fairly well established that the func-

tional relationship between the yield of corn grain and level of nitrogen may be represented by a distorted sigmoidal curve similar to the one shown in Figure 1. For most productivity conditions a reasonable assumption can be made as to the nature of the fertilizer response function. Consequently, for studies designed to determine optimal levels of fertilization, investigators probably should select the continuous model that appears most appropriate in the light of existing knowledge and take measure to minimize the error due to the inadequacy of the model in representing the true relationship.

The response curve shown in Figure 1 extends from the zero level of nitrogen to an upper level well past that required for the maximum yield. In practice, soil nutrient levels are always greater than zero and commonly exceed a , the level corresponding to the inflection point on the curve. Also, as discussed by Anderson (1957), for those cases where the soil nutrient level is less than a , discarding observations on yield obtained at the zero level of fertilization will lead to a diminishing returns type function which can be more adequately represented by a second degree polynomial. Consequently, in the selection of a representational model, interest is generally restricted to that portion of the response curve that lies to the right of a .

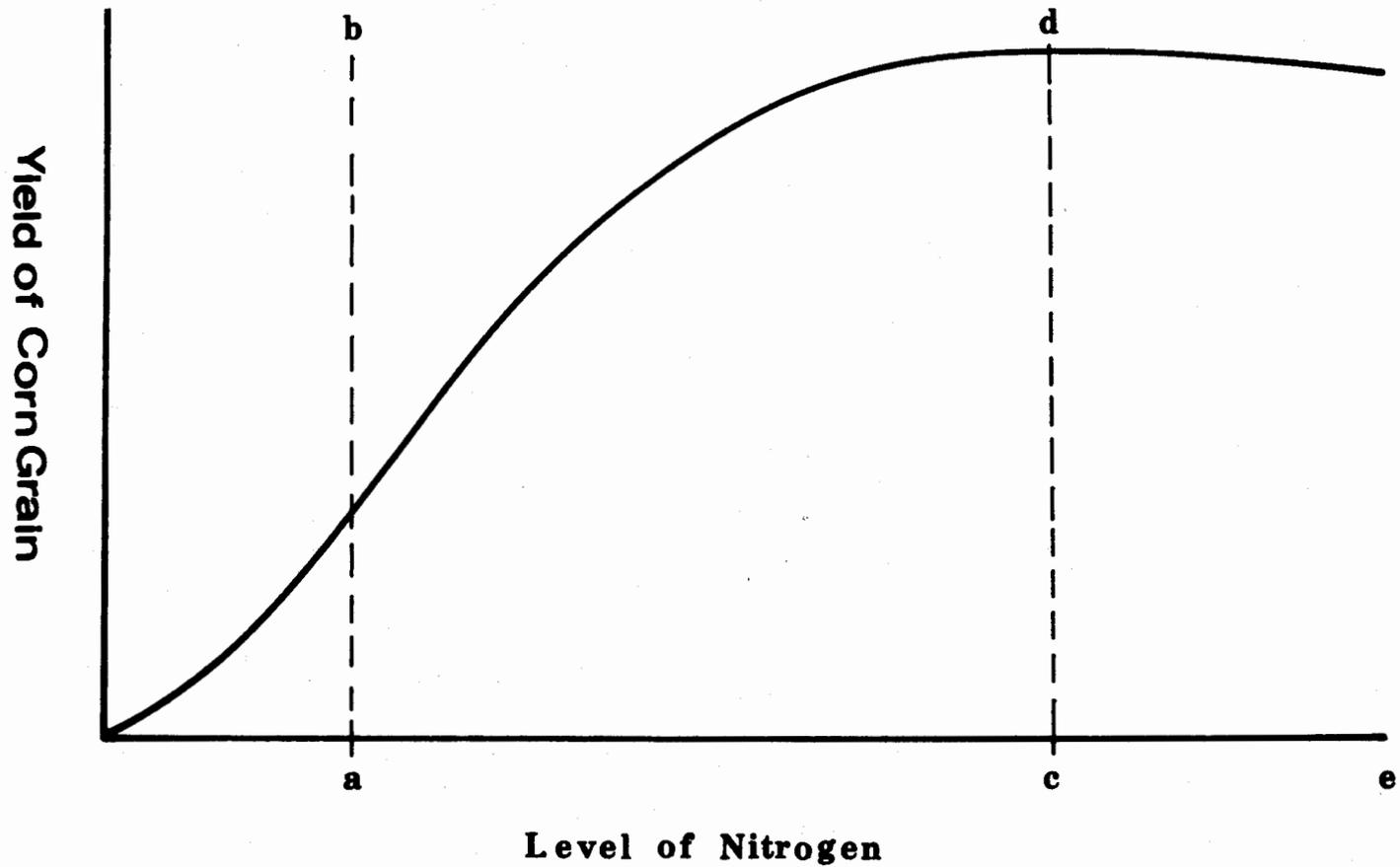
The relatively flat area beyond the maximum of the curve in Figure 1 also represents somewhat of a complication in the selection of a model. However, in the study of economically optimal levels of fertilization, the investigator is generally not concerned with the nature of the response function past the maximum yield. Therefore it seems reasonable to limit the range of interest to that lying between a and c .

The relationship between yield and nitrogen in the range from a to c may be approximated by several algebraic functions. The two forms which have been used most frequently in fertilizer use research are the Mitscherlich-Spillman equation and the second order polynomial (Heady, et al., 1961) The quadratic form of the polynomial has in it the advantage that yields decline at nutrient levels greater than that corresponding to the maximum. A second advantage of the quadratic polynomial model is that its parameters are linear combinations of the observations and may be readily estimated using the least squares procedure. No suitable transformation to linear form is available for the Mitscherlich-Spillman parameters, and they are usually estimated using a more laborious iterative procedure. Thus, for finding optimal level of fertilization, many investigators restrict the range of interest between a and c in Figure 1 and select a polynomial as the estimating function.

The quadratic polynomial function with one factor may be expressed algebraically as:

$$\hat{Y} = b_0 + b_1X + b_2X^2$$

in which \hat{Y} is the predicted yield, b_0 is the estimated yield without any applied nutrient, b_1 is the estimated linear coefficient, b_2 is the estimated quadratic coefficient, and the independent variable X is the level of fertilization. In practice the estimated coefficients can be interpreted only



A hypothetical response function showing the relationship between corn yield and level of nitrogen. There is an inflection point in the curve at a level of nitrogen and the maximum response occurs at c level of nitrogen.

together as a pair. In fertilizer use research a "diminishing return" type of response curve is characterized by a positive linear coefficient and a negative quadratic coefficient. With this combination, the response curve slopes upward for the first increments of X, but at a decreasing rate, reaches a stationary or zero slope point at X equal to $-b_1/b_2$ and then slopes downward.

In addition to the quadratic polynomial, the square root model has been a commonly used representation of fertilizer response data. This function with one factor has the form

$$\hat{Y} = b_0 + b_1X^{.5} + b_2X .$$

This transformation of the independent variable is equivalent to expressing the abscissa of the function in terms of \sqrt{X} instead of X. Thus, for values of X less than one, the square root function rises more rapidly than the quadratic; for values of X greater than one, the slope of the square root function changes more slowly. When the two forms are fitted to the same set of fertilizer response data, the relative magnitude of the quadratic coefficient with respect to the linear coefficient usually is much greater for the square root than for the quadratic equation. The square root function has a larger slope than the quadratic at low levels of application and becomes much flatter than the quadratic at high levels of application. For the square root function to predict diminishing returns with increasing levels of fertilization, b_2 must be negative. If the yields are to increase for values of X less than the maximum and decrease thereafter, then b_1 must also be positive. In this case the maximum value occurs at $X = b_1^2/4b_2^2$.

The quadratic and square root functions can be extended to equations with two or more factors. A quadratic polynomial model with two factors is often expressed as

$$\hat{Y} = b_0 + b_1X_1 + b_2X_1^2 + b_3X_2 + b_4X_2^2 + b_5X_1X_2 .$$

The levels of the two factors or nutrients are denoted as X_1 and X_2 and the estimated linear, pure quadratic and interaction coefficients are b_1, b_2, \dots, b_5 respectively. Usually the signs of the linear terms are positive and the pure quadratics negative. The estimated response surface has a maximum and the calculation of an optimal level of fertilization is practical only if the pure quadratic terms are negative and the product of b_2 and b_4 exceeds b_5^2 . If the range of the observations on yield extends beyond the level of fertilization corresponding to the maximum yield, the estimated response function will normally have a maximum. For experiments in which the observations either fall short of or just reach the maximum yield, the resulting estimated response function may not have a maximum within the experimental range of the fertilizer variables.

The square root estimating function with two unknowns may be written in a form similar to the quadratic model. The efficiency of the quadratic and square root functions in representing fertilizer response data has been compared in several studies. Coefficients of determination (R^2), indicating

the proportion of the total variability in yield which is accounted for by the independent variables, have been reported for quadratic and square root equations fitted to the same data by Heady and Dillon (1961). In their work it appears that the square root prediction equation represents the experimental observations somewhat more closely than the quadratic equation, but differences in the size of the R^2 coefficients are small.

Box and collaborators (1951, 1957), in their development of composite designs, began with the assumption that the quadratic polynomial model can be used to represent the response function within the region of experimentation. Thus, their treatment designs were derived specifically for use in conjunction with a quadratic polynomial. This does not mean that experimental data generated with a composite design cannot be used to fit a square root equation. It is true, however, that the moment requirements for rotatable designs are not satisfied when the X variables are transformed for the square root equation. Investigators who choose to employ composite rotatable designs usually use linear or quadratic polynomial models representing the response surface.

A variation of the polynomial model has been proposed by Nelder (1966). His estimated model has the algebraic form

$$\frac{1}{\hat{Y}} = \frac{b_0}{X} + b_1 + b_2X .$$

Called the "inverse polynomial", it does not have symmetry about the maximum. Also, only positive values of the response can be predicted when the estimated model is used beyond the range of the data. However, in the estimation procedure given by Nelder, it is implicitly assumed that the variance is not constant for different levels of X. A thesis by Tejada (1966) is an example of applying Nelder's inverse polynomial.

In the following discussion of criteria for comparing treatment designs it is assumed that the yield data will be used to estimate a quadratic polynomial model. The selection of the quadratic equation as the estimating function has been made for these reasons: (i) theoretical and empirical evidence indicates that the quadratic is a reasonably adequate representation of the fertilizer response function, if the region of interest is restricted to the range between a and c in figure 1, (ii) computations are simplified by using the quadratic equation, and (iii) economically optimal levels of fertilizer can be easily calculated.

PREDICTION ERRORS

In fertilizer use studies, a functional relationship between yield and the fertilizer variable(s) is envisioned and may be described by a mathematical model. The parameters of the model characterize the form of the functional relationship. The data from an experiment are then used to estimate the model parameters. Usually, the measure of "goodness" of the estimated model is based on the difference between an observed response (Y) and the predicted response (\hat{Y}) calculated from the model with

the estimated parameters and values of the fertilizer variable(s). For example, when a straight line relationship between yield (Y) and applied nitrogen (X) is assumed, the model is

$$Y = \beta_0 + \beta_1 X + \varepsilon, \quad (1)$$

where the betas are the parameters for the intercept and the slope and ε is the random error. There is an unknown value of ε for each value of Y coming from a distribution assumed to have a mean of zero and variance of σ^2 . Each ε is estimated by $Y - \hat{Y}$, the difference between the observed value and that value predicted by the estimated model

$$\hat{Y} = b_0 + b_1 X \quad (2)$$

where b_0 and b_1 are the least squares estimators of β_0 and β_1 respectively.

If the true model, $\eta(X)$, is in fact equal to $\beta_0 + \beta_1 X$, then these differences are estimates of measures of the failure of the model to explain the random error in the n data points. In this case,

$$s^2 = \frac{1}{n-2} [\sum(Y - \hat{Y})^2] \quad (3)$$

is an estimator of σ^2 . That is, the only reason that \hat{Y} is not equal to Y is the random variability of the experiment caused by factors including soil, environmental and measurement variability. The notation of X within the parenthesis after η is used to indicate that η is a function of X.

However, if $\eta(X)$ is not equal to $\beta_0 + \beta_1 X$, then differences between Y and \hat{Y} are induced by a second type of error, called "bias error", in addition to the random error. Bias error arises when the true model is not equal to the expectation of the fitted model. With $Y - \hat{Y}$ now including both random and bias errors, we would expect expression (3) to estimate the sum of σ^2 and a component which is due to the bias error. If the experimenter has replicated at least one of the levels of the fertilizer variable, an independent estimate of σ^2 can be calculated since differences in replicated responses can be due only to random errors. This independent estimate may then be compared with the mean square calculated from the remaining sum of squares as an indication of bias error magnitude.

Variance error

Associated with a statistic arising from response function estimation such as a linear combination of estimated parameters, e.g., the response function, is its variance which we call variance error, having two distinct components. One of these, the experimental error, σ^2 , arises due to random variation among experimental units or plots. In a completely randomized experiment it is recognized as the failure of replicated plots of a given treatment to yield the same. The estimated value, s^2 , when expressed on

the basis of the experimental unit, is an averaged squared deviation of the yield of an individual plot from the treatment mean. It is influenced by such things as variability in the plant, soil, management, and the way in which the treatments are applied to the experimental units.

The second component of the variance error is somewhat more difficult to visualize. It is a function of the distribution of the treatment combinations in the factor space, and thus is influenced by the selection of the treatment design. As an illustration, consider the problem of estimating the slope of a function when it is accepted that the true relationship existing between the two variables is linear. It can be shown that the most efficient way to estimate the slope is to make one-half of the observations at each of the two levels of the X variable corresponding to the extremes of the factor space. This component of the variance error which is related to the distribution of the treatment combinations appears in the regression analysis as particular elements of the inverse matrix. The size of the elements in the inverse matrix depends upon the number of treatment combinations, the distribution of these treatments within the factor space, and the number of times the set of treatments is replicated. The variance error is the product of these two components.

Bias error

As shown in the first part of this section, the quantity, $Y - \hat{Y}$, contains two sources of error if the true model is not equal to the expectation of the estimated model. The difference between the true model and the expectation of the estimated model,

$$\eta(X) - E(\hat{Y}(X)), \quad (4)$$

evaluated at a given point in the factor space, is called the bias of the predicted value of Y. In general, the average bias of $\hat{Y}(X)$ is defined as the difference (4) squared and integrated over a defined factor space. This concept of average bias error was proposed and developed by Box and Draper (1959).

To clarify the various kinds of "bias error" concepts, a formulation, based on polynomial estimated and true models and using matrix notation, is developed. It is convenient to express the true model at each of the N design points as

$$\eta(X) = X_1\beta_1 + X_2\beta_2. \quad (5)$$

The estimated model is shown as

$$\hat{Y}(X) = X_1b_1 \quad (6)$$

where the variable(s) in the estimated or fitted model are included in the treatment design X_1 ($N \times p$ matrix where p is the number of estimated

parameters) and \mathbf{b}_1 is a $p \times 1$ column vector of the estimated coefficients, estimating β_1 . The variable(s) not included in the estimated model but with nonzero coefficients (β_2) are included in X_2 .

Expression (4) can be rewritten as

$$[\mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2] - E(\mathbf{x}, \mathbf{b}_2) \quad (7)$$

where \mathbf{x}_1 is a set of x values of the p variables included in the estimated model to predict Y for a given point in the factor space, i.e. \mathbf{x}_1 is any row vector not necessarily one of the rows of the X_1 matrix. The x values for variables not in the estimated model for the same point in the factor space are included in the row vector \mathbf{x}_2 .

The expectation of $\mathbf{x}_1\mathbf{b}_1$ is not simply $\mathbf{x}_1\beta_1$ because the expectation has to consider the bias of the coefficients in the \mathbf{b}_1 vector. The expectation of \mathbf{b}_1 is $\beta_1 + \mathbf{A}\beta_2$ where $\mathbf{A} = (\mathbf{X}_1\mathbf{X}_1)^{-1}\mathbf{X}_1\mathbf{X}_2$, as shown by Draper and Smith (1966) and $E(\hat{Y}) = \eta(\mathbf{X})$. Therefore the bias of \hat{Y} for a point in the factor space is

$$[\mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2] - \mathbf{x}_1[\beta_1 + \mathbf{A}\beta_2] ,$$

or equivalently,

$$\mathbf{x}_2\beta_2 - \mathbf{x}_1\mathbf{A}\beta_2 . \quad (8)$$

This is a criterion that can be readily considered for comparing alternative treatment designs because the alias matrix \mathbf{A} is affected by the treatment design matrix X_1 . If expression (8) is squared and integrated over the factor space, the Box and Draper (1959) concept of average bias in $\hat{Y}(\mathbf{X})$ can be written.

Another formulation of a bias concept in an estimated function involves the difference of the two terms in expression (8) evaluated at the treatment design points. Now expression (8) can be written as the bias vector

$$\mathbf{X}_2\beta_2 - \mathbf{X}_1\mathbf{A}\beta_2 . \quad (9)$$

This bias vector is associated with calculating lack of fit sum of squares in analysis of variance tables for data from response surface experiments. The squared bias term, a scalar, is the squared length of the bias vector (9) and can be expressed as

$$\beta_2' \mathbf{X}'_2 \mathbf{X}_2 - \mathbf{X}'_2 \mathbf{X}_1 \mathbf{A} \beta_2 . \quad (10)$$

It is the part of the expectation of the sum of squares for lack of fit in an analysis of variance table due to $\mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2$ being the true model

and $X_i b_i$, the fitted or estimated model. It may be seen from (10) that the magnitude of lack of fit may be influenced by the selection of the treatment design, X_i .

It is well at this point to digress momentarily and discuss the term lack of fit, which is not rigorously defined in most textbooks. For example, Cochran and Cox (1957), Chapter 8A, used the term in quotes. Draper and Smith (1966) have five index references to lack of fit but without a clear definition. Draper and Herzberger (1971) give a procedure for dividing the lack of fit into two components in order to gain more insight on the nature of the lack of fit. Generally speaking, lack of fit is associated with the failure of the fitted model to completely describe the data. If a linear polynomial is fitted to data which are actually described by a quadratic polynomial, then the failure of the straight line to show the true shape of the function is attributable to lack of fit. By calculation, lack of fit is a sum of squares based on the data and limited to the discrete experimental points in the factor space. The residual sum of squares, the difference between the total sum of squares and the sum of squares associated with the fitted model, is partitioned into two parts. The first is the experimental error sum of squares, assuming a designed study, which may be calculated directly. Then the experimental error sum of squares is subtracted from the residual sum of squares and the difference called lack of fit. It cannot be calculated directly unless the difference in the degrees of freedom associated with the fitted polynomial model and the treatment degrees of freedom is associated with higher order terms which are then added to the fitted model. The lack of fit sum of squares is then forced to be associated with the extra terms, i.e., the higher order terms are equated with the β_2 vector of expression (10). In this situation the assumed model that is used to force the lack of fit sum of squares to be associated with higher order polynomial terms is not necessarily the true model. Furthermore, the concept of average bias error in \hat{Y} involves integration over the entire factor space, not evaluation at only the treatment design points as done in analysis of variance calculations. Consequently, lack of fit as usually calculated in analysis of variance tables, and average bias, or bias defined by (10) where knowledge of the true model is known, need not be, and usually aren't, equivalent. However, if an experimenter tries to select between two alternative models, then the model with the smallest lack of fit is probably the one with the smallest average bias. An example is helpful in demonstrating the difference. Consider a 3×3 factorial where the two factors are quantitative, e.g., applied nitrogen and phosphorus, and a quadratic polynomial is fitted to the data. If the true model is a quadratic polynomial, then the lack of fit will be an estimate of experimental error. Assume that the true model is a cubic polynomial. A bias exists but it has to be measured using the data from the 3×3 factorial. Lack of fit is calculated by fitting the two linear by quadratic terms and the quadratic by quadratic term. In addition, the fitted model could do very well at estimating the true model at the nine points in the factor space but the average bias could be relatively large.

A combined variance and bias error

In the statistical literature, the variance and bias errors are combined in a function called the mean squared error. Box and Draper (1959) have used minimization of the mean squared error, which is the sum of the variance of $\hat{Y}(X)$ and the square of the bias of $\hat{Y}(X)$, as a criterion for choosing a treatment design.

In the 1959 Box and Draper work, the combined error is standardized by σ^2 on a per observation basis, involves a concept of coding or scaling as explained by Myers (1971), and is integrated over the whole factor space. The mean squared error is defined as

$$J = V + B$$

where V is the average variance of $\hat{Y}(X)$ and B is the average squared bias of $\hat{Y}(X)$. It then seems reasonable to select a treatment design which would minimize J . The advantages of this criterion are not only that J is not affected by linear transformation of the independent variable and that it considers both variance and bias but also that it is integrated over the factor space, not limited to certain points within the factor space. As an example, consider the situation where the experimenter fits

$$\hat{Y} = b_0 + b_1X$$

as his prediction model while the $E(Y)$ or the true model is

$$E(Y) = \eta(X) = \beta_0 + \beta_1X + \beta_2X^2$$

The average variance and bias have been worked out in detail by Myers (1971) and if the design points are selected so that they are symmetric about the center point,

$$V = 1 + \frac{1}{3[11]}$$

$$B = \frac{N\beta_2^2}{\sigma^2} \left\{ ([11] - \frac{1}{3})^2 + \frac{4}{45} \right\}$$

where $[11]$ is called by Box and Hunter (1957) the second moment of the design, i.e., it is the sum of squares of X divided by N , the number of observations.

The term $[11]$ appears in both V and B . Because $[11]$ is in the denominator for the expression of V , variance can be decreased by increasing the sum of squares of X for a given number of design points. For bias, a particular value of $[11]$ equal to $\frac{1}{3}$ is needed to minimize B . However, a problem exists in trying to minimize $J = V + B$ since the value $[11]$ that will minimize J depends on the unknown parameter, β_2 , in the expression for B . If β_2 were relatively small, then values of $[11]$ larger than $\frac{1}{3}$ would minimize J . For example, if β_2 was equal to zero, then V would

be minimized by placing half the design points at each end of the range of X, thereby making [11] as large as possible. On the other hand, if β_2 was relatively large, values of [11] close to or equal to $1/3$ would be best, leading to designs with points between the extremes of the range of X.

Myers (1971) presents a table showing that, if the experimenter did indeed know the values of $N\beta_2^2/\sigma^2$, and could calculate the best second moment for the treatment design to minimize J, the moment selected would be close to that design selected for only bias considerations. This general rule of thumb holds through a range of V to B ratios up to approximately a value of 6. An extension of this simple example to more than one variable and more complex models is given by Myers (1971) with similar results. However, even if $N\beta_2^2/\sigma^2$ were known, this does not determine the best design, only its second moment. Other criteria must be used to select the actual design (Thompson 1971).

CRITERIA FOR COMPARING TREATMENT DESIGNS

It should be clear from the previous discussion that an experimenter has to consider many objectives when choosing a treatment design. Some of these may be summarized as:

- 1) interpretable data without extensive analysis, e.g., for at least two levels of each factor, there should be three of the other factors,
- 2) relatively small number of treatment combinations,
- 3) low variance of estimated coefficients and, if hypothesis testing is important, independent estimates,
- 4) variance of the predicted values small over the central part of the factor space,
- 5) variance of the estimated response function slopes small over the central part of the factor space,
- 6) bias of the predicted values small over at least the central part of the factor space,
- 7) measure of lack of fit,
- 8) inclusion of the "check" plot in the design, and
- 9) exclusion of the combinations of zero of one factor and high level of another factor.

In order to compare two or more treatment designs certain restrictions must be placed on the values which the X variables may assume. Box and Wilson (1951) suggest the following convention for bringing different designs to the same "size". They define the spread, S, of a variable as the average deviation of that variable from its mean, calculated as the square root of the quantity

$$S^2 = \frac{1}{N} \sum_{u=1}^N (X_u - \bar{X})^2$$

where X_u is any value of the variable X , \bar{X} is the mean, and N is the number of observations. Two designs are considered to have comparable size when the spread for each of the variables is the same.

The requirement imposed by the Box and Wilson (1951) work that designs have the same spread in order to be comparable has been questioned by Folks (1958) and Kempthorne (1965). They maintain that it would seem more natural and appropriate for an experimenter to simply define the factor space within which he expects to make inferences. Designs are then evaluated over the same factor space.

The experimenter on the basis of *a priori* information decides that the factor space should either extend to zero or should begin at some level of fertilization above zero. He is also able to state in general terms the upper rate limits which are to appear in the treatment design. The upper limit of the factor space should generally extend slightly beyond the fertilization level corresponding to the maximum yield. This is so that the estimated response function will usually predict a maximum in the region of experimentation. At the same time only a small number of yield observations should be made at fertilization levels beyond the maximum, as observations taken in this part of the factor space may increase the error due to bias in predicted yields in the region near the optimum.

After the factor space has been decided, the treatment design problem is how to place the treatment combinations in the factor space. In order to select the best design among various alternatives, both bias and variance criteria have to be considered.

Stigler (1971) has summarized the work on optimality of design and has discussed several criteria which have been employed to compare designs from the standpoint of minimal variance. Designs that minimize the determinant of the covariance matrix of the estimated coefficients or minimize the maximum variance of \hat{Y} are discussed in detail by Kiefer (1959). Equivalently, the determinant of $X'X$ may be maximized as developed by M. J. Box and Draper (1971). Perhaps the criterion which has been viewed most favorably in past fertilizer use experimentation is the minimization of the variance of the estimated response function coefficients (b 's). Box and associates (1957, 1959) developed the concept of the variance function and considered the mean variance of the predicted yield over the immediate region of interest as the appropriate criterion to employ in comparing designs.

Certainly, as pointed out by Kempthorne (1965), the use of different criteria of optimality leads to different designs. Also, it is equally clear, that the most appropriate criterion under a given set of circumstances will be determined by the specific use that is made of the data generated by the design. In fertilizer use experimentation for the purpose of determining economically optimum levels of fertilization, the investigator is interested primarily in measuring the slope of the response function at some point corresponding to the optimum. He is basically concerned in measuring this slope with the greatest precision possible in accordance with the facilities available. Consequently, for this type of research the

minimization of the variance of the slope of the response function should be a major variance criterion in comparing designs. Using this variance criterion, Ott and Mendenhall (1972) give suggestions for one-factor treatment designs. The procedure for estimating this variance for two factors was developed by Fuller (1962).

If the variance of the slope is calculated for a large number of treatment combinations distributed over the experimental factor space, it becomes apparent that the variance of the slope is minimal at the center of the design and increases toward the limits. The question arises, therefore, as to where the variance of the slope should be evaluated in the comparison of treatment designs. Generally, at the time of selecting a design, the investigator has reasonable assurance that the optimal level will lie fairly close to the center of the design, but he has no information as to the orientation of the optimum with respect to the center. Consequently, it would seem reasonable to select a symmetrical space about the center of the design as the region in which the optimum is expected to occur and minimize the average or maximum variance of a slope within this region.

Complete reliance on variance criteria is justified only if the true model is known. In agronomic research however, the true model is not known and bias error becomes an important consideration. Box and Draper (1959, 1963) suggest mean squared error as a criterion for comparing designs. However, as observed previously, knowledge of the true model is needed for calculating the mean squared error. Furthermore, the design which is best for variance error will not be optimal for bias error. Box and Draper circumvent this dilemma by recommending that a design be selected for bias reasons alone.

Similar conclusions were reached by Draper and Lawrence (1967) and Myers (1971) who conclude that the experimenter should choose minimum bias designs unless it was known that variance was very important relative to bias.

Thompson (1971), working with a class of minimum bias two-factor designs and a square factor space, proposes secondary criteria of maximizing the power to detect the inadequacy of the postulated model and of minimizing the variance function of an estimated higher order polynomial.

Another measure of bias has been proposed by Cady and Laird (1969). With one independent variable, the bias measure is defined as the area between the true and estimated models. With two independent variables, the bias would be a volume. This concept of bias, easy to understand and intuitively appealing, may be expressed as the absolute difference integrated over the factor space, s :

$$\int_s |E\{\hat{Y}(X)\} - \eta(X)| dX$$

It can be seen that several variance and bias criteria are available for comparing designs. For fertilizer use experimentation with inadequate knowledge of the true model, priority will be given to bias considera-

tions in selecting a treatment design, i.e., the experimenter's control of treatment design would then be used primarily for bias. Variance error will be handled through experimental design, e.g., blocking and replication.

EMPIRICAL COMPARISON OF SELECTED DESIGNS

It is apparent from the previous sections that multiple criteria will be necessary to select a treatment design because the experimenter usually has more than one objective in a given experiment. If the individual estimated regression coefficients are of interests, the design minimizing the variances and covariances of the coefficients should be considered. However, if the experimenter is using the estimated regression equation for prediction, other criteria, including bias, become important. In fact, in recent years, bias considerations have become the dominant criterion in design selection.

Some quantitative results of treatment design on bias were obtained in the 1969 work by Cady and Laird. For fertilizer use studies, it was assumed that the nature of the relationship between yield and the independent variable(s) is an increasing function at a decreasing rate, permitting a defined maximum and diminishing marginal products. Then, the true model and any fitted models were selected from a class of equations represented (in one variable) by

$$\eta(X) = \beta_0 + \beta_1 X^c + \beta_2 X^{2c}$$

where $.5 \leq c \leq 1$, i.e., the two extremes are the square root and quadratic polynomial models.

For designs with a uniform spacing of treatments over the factor space, the error due to bias decreases as the number of treatments increases. This decreasing function is not linear and the rate of decrease in bias error per added treatment will become very small as the total number of treatments becomes large. Bias error in subregions of the factor space can be influenced by non-uniform spacing of the treatments. For these asymmetrical designs, error due to bias is minimal in the region of greatest concentration of treatments.

Escobar (1967) has compared factorials, partial factorials and composite designs by calculating the integrated variance and squared bias of $\hat{Y}(X)$ when the fitted model was a quadratic polynomial in two variables, X_1 and X_2 . The true model used was the quadratic polynomial plus two specific cubic terms, $X_1^2 X_2$ and $X_1 X_2^2$. The Escobar results show that the composite designs are superior when bias is considered while the factorials and partial factorials have smaller variances. Designs minimizing the generalized variance, rotatable designs, and designs based on regular and irregular fractions of 2^n factorials are compared using a quadratic nonpolynomial model by Nalimov, et al. (1970).

Integrated bias and variance in $\hat{Y}(X)$ were the major considerations in an empirical study comparing nine treatment designs in two factors

as carried out by Cady and Laird (1972). From the results, an investigator can be guided in the selection of a two-factor treatment design for his specific objectives.

Response functions of the type

$$E(Y) = \beta_0 + \beta_1 X_1^c + \beta_2 X_2^c + \beta_3 X_1^{2c} + \beta_4 X_2^{2c} + \beta_5 X_1^c X_2^c$$

where $.5 \leq c \leq 1.0$ were considered. Reported by Cady and Laird (1972) were results from studies where c of the true function is .5, i.e., a square root model, and the c of the fitted model is 1.0, a quadratic polynomial. The betas of the square root model were chosen to give a representative response function of corn to applied nitrogen (N) and phosphorus (P). The true model where Y , N and P are in kilograms/hectare was:

$$E(Y) = 3000 + 300.5\sqrt{N} + 374.4\sqrt{P} - 15.10 N - 23.33 P + 15.00\sqrt{NP}.$$

The response surface is shown in Figure 2:0 where the ranges for N and P were 0 to 320 and 0 to 240 kilograms/hectare respectively.

The nine designs studied are shown in Figure 3 and the fitted response surfaces (except for the central composite), using a quadratic polynomial as the estimating model, are given in Figure 2:1 through 9. Included were:

- 1) 5×5 factorial, 25 treatments,
- 2) central composite, 9 treatments,
- 3) central composite modification by Myers, 9 treatments,
- 4) 3×3 factorial (rotated), 9 treatments,
- 5) 5×5 partial factorial, 13 treatments,
- 6) 5×5 partial factorial modification by Escobar, 13 treatments.
- 7) 7×7 partial factorial, 17 treatments,
- 8) 3×3 factorial, 9 treatments,
- 9) central composite modification by Thompson, 12 treatments.

The criteria for comparing the designs included:

- 1) bias residuals, $|\eta - \hat{Y}|$,
- 2) index of integrated bias,
- 3) index of variance of \hat{Y} , $V(\hat{Y})$,
- 4) index of variance of the derivative of the response to nitrogen, $V(dY/dN)$,
- 5) index of the variance of estimated regression coefficients.

The true responses were calculated from the true square root model for the treatments in each design, quadratic equations were fitted to the true responses, and predicted responses were then calculated from the fitted quadratic equations. The bias residuals or differences between the true response and the predicted response as summarized in Table 1 give a quick but incomplete impression of the performance of each design. The values in the table are the maximums and averages of the absolute differences between the true yield from the square root model and the

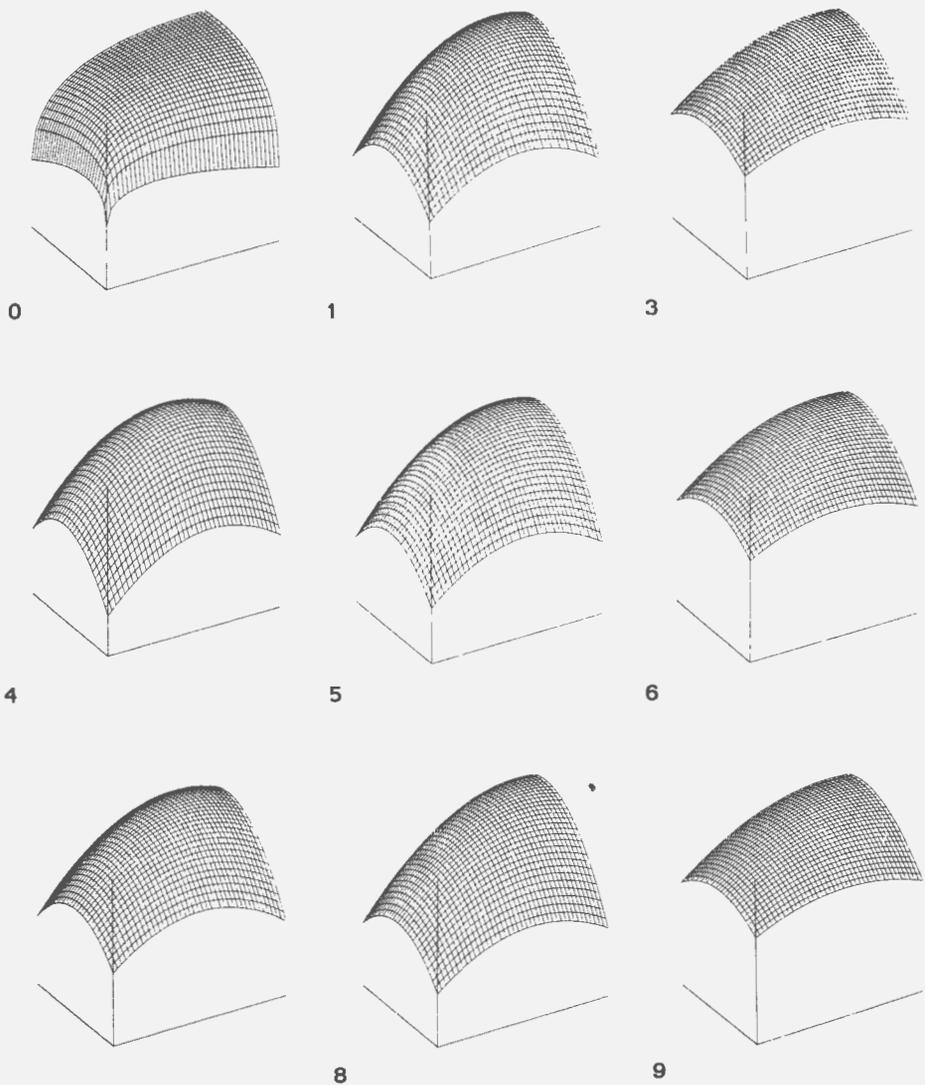


Figure 2.

The response surface, 0, for the square root (true) model used in the Cady and Laird (1972) study, and the fitted response surfaces using a quadratic polynomial as the estimating model for the treatment designs: (1) 5 x 5 factorial, (3) central composite modification by Myers, (4) 3 x 3 factorial (rotated), 5 x 5 partial factorial, (6) 5 x 5 partial factorial modification by Escobar, (7) 7 x 7 partial factorial, (8) 3 x 3 factorial, and (9) central composite modification by Thompson.

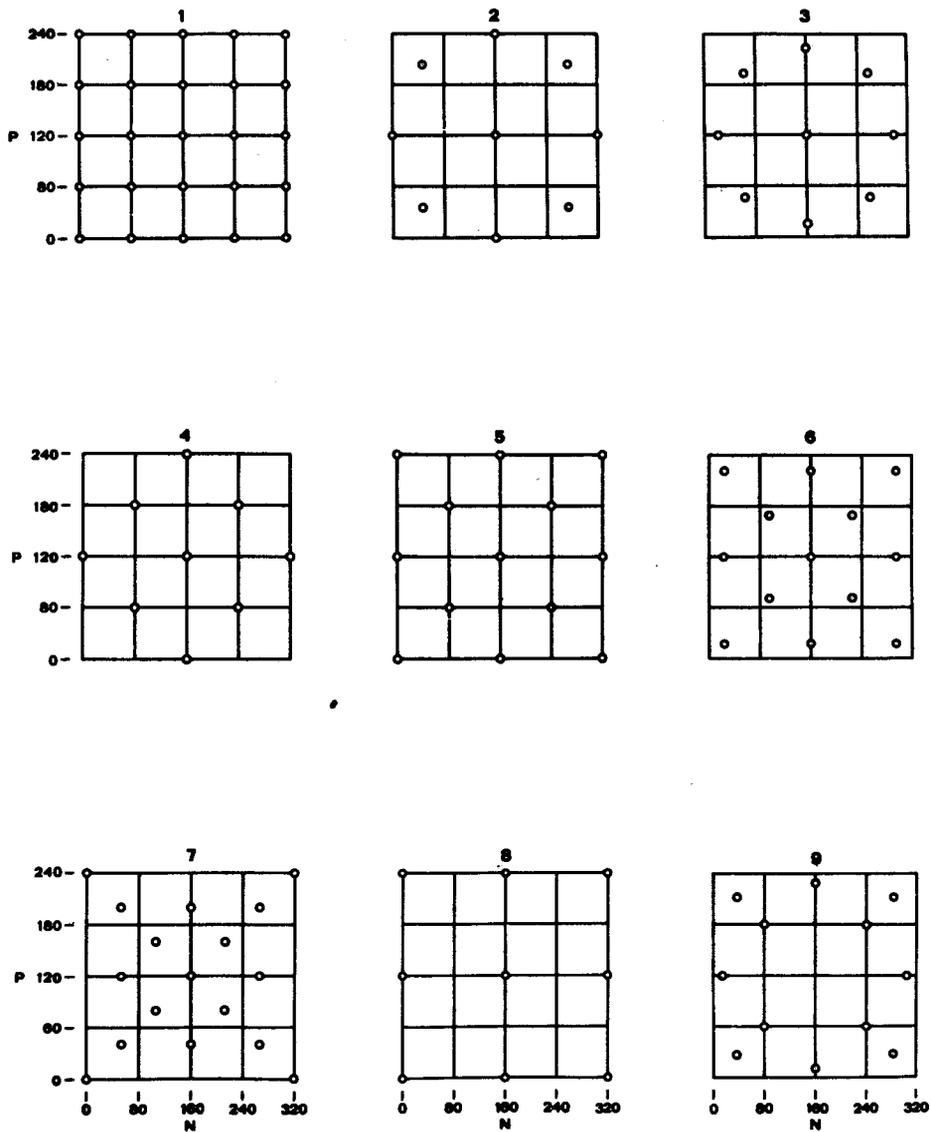


Figure 3.

The treatment designs compared in the Cady and Laird (1972) study: (1) 5 x 5 factorial, (2) central composite, (3) central composite modification by Myers, (4) 3 x 3 factorial (rotated), (5) 5 x 5 partial factorial, (6) 5 x 5 partial factorial modification by Escobar, (7) 7 x 7 partial factorial, (8) 3 x 3 factorial, and (9) central composite modification by Thompson.

predicted yield from the estimated quadratic polynomial, evaluated at the various treatment combinations for each design. The central composite, the 5 × 5 factorials, complete or partial, the 7 × 7 partial or the 3 × 3 rotated perform poorly relative to the designs modified by Myers, Escobar or Thompson. Of surprising note are the small maximum and average residuals of the 3 × 3 factorial. The weakness of looking at the residuals calculated only at the treatment combinations becomes apparent. Eight of the 9 treatment combinations of the 3 × 3 are on the boundaries of the factor space. Hence the apparent good performance of the 3 × 3 might be due to the estimating model fitting relatively well on the boundaries. Hypothesized would be a poor fit across the remaining factor space.

TABLE 1. Bias Residuals, $|\eta - \hat{Y}|$, kilograms/hectare

Design	Number of Design Points	Maximum	Average
1	25	1022	406
2	9	1199	508
3	9	301	135
4	9	996	442
5	13	882	333
6	13	222	84
7	17	766	350
8	9	326	204
9	12	372	161

The concept of integrated bias developed by Cady and Laird (1969) was used to calculate the indices of integrated bias. Values for the indices of integrated bias in Table 2 were calculated by approximating the volume between the true and estimated response surfaces for the central quarter and the entire factor space. The central quarter is defined as the factor space using ranges of 80 to 240 and 60 to 180 for N and P respectively. The calculations are basically those described by Cady and Laird (1969) using the intersections of a 30 × 30 grid placed over the factor space. Finer sized grids were tried in preliminary calculations with differences of less than 4%. Now the largest bias is given by the 3 × 3 factorials followed by the 5 × 5 complete and partial factorials. The modified designs have relatively low biases and the influence of number and distribution of points is shown by the intermediate value of the 7 × 7 partial factorial.

Variance criteria were also considered. The designs are compared in Table 2 by the contribution of the design matrix to the variance of a predicted value, evaluated at 144 common points (intersections of a 12 × 12 grid) over the entire factor space. The maximums and averages given in

TABLE 2. Comparison of Treatment Designs by Integrated Bias and Variance Indexes

Design	Number of Design Points	Index of Integrated Biases		Index of Variance of a Predicted Response			Index of Variance of the Derivative of the Response to Nitrogen, $\left(\frac{\delta Y}{\delta N}\right)$	
		Entire Factor Space ($\times 10^3$)	Central Quarter of Factor Space ($\times 10^3$)	Average $V(\hat{Y})$		Maximum $V(\hat{Y})$ (Central Quarter)	Maximum ($\times 10^{-7}$)	Average ($\times 10^{-7}$)
				Entire Factor Space	Central Quarter Factor Space			
1	25	460	71	0.189	0.134	0.151	136	84
2	9	346	56	0.812	0.732			417
3	9	210	20	1.270	0.674	0.962	1254	692
4	9	485	80	1.310	0.454	0.539	1009	570
5	13	468	71	0.342	0.244	0.269	242	149
6	13	210	21	0.498	0.232	0.253	431	255
7	17	348	58	0.527	0.176	0.185	616	344
8	9	580	88	0.477	0.478	0.547	285	175
9	12	208	20	0.682	0.587	0.800	693	385

Table 2 are based on values calculated by $X'_i(X'X)^{-1}X_i$ as given by Draper and Smith (1966) where X_i is a row vector of values for the columns in the X matrix for a single observation. Also shown in the table are the maximums of the central 36 points. As expected the 5 × 5 factorial has the lowest averages. The Myers modification, the 3 × 3 rotated, and to a lesser degree, the central composite, the Thompson and Escobar modifications and the 7 × 7 partial factorial, have higher values reflecting the scarcity of points on the boundaries. However, these latter designs perform better if only the central quarter of the factor space is considered. In addition, two or three replications of these designs will utilize approximately the same number of plots as one replication of a 5 × 5 factorial. Increasing replication will decrease the variance of \hat{Y} . Consequently two replications of the 5 × 5 partial factorial or three replications for the 3 × 3 factorial will result in favorable variance values compared with one replication of the 5 × 5. If emphasis is placed on the central quarter of the factor space, then several designs compare very favorably with the 5 × 5 factorial, including one of the modified designs with good bias properties, the Escobar modification.

In an economic evaluation of the estimated response surface, the calculation of the optimal quantity of N involves the derivative of the response surface with respect to N. The calculated derivative is an estimated parameter and consequently has a variance. Desired properties of a treatment design would include small values of the variances of the derivatives. The values in Table 2 were calculated by the Fuller (1962) method. The individual variances were evaluated at the intersections of a 5 × 5 grid in the central quarter of the factor space. Again the superiority of the 5 × 5 factorial is seen. However, remembering that two or three replications of the other designs can be used to lower the variances, the other factorials, complete, partial and modified, do well.

TABLE 3. Diagonal Inverse Elements

Design	(x10 ⁻⁴) Nitrogen (N)	(x10 ⁻⁴) Phosphorus (P)	(x10 ⁻⁹) N ²	(x10 ⁻⁹) P ²	(x10 ⁻⁹) NP
1	.4508	.8015	.3487	.1102	.4340
2	2.649	4.717	2.106	.6667	2.743
3	4.624	8.221	3.713	1.173	4.802
4	3.645	6.481	1.907	.6028	10.85
5	.8021	1.426	.6390	.2019	.6382
6	1.488	2.645	1.202	.3800	1.238
7	2.298	4.086	2.112	.6675	.5603
8	.9440	1.678	.7629	.2411	.6781
9	2.540	4.516	2.178	.6883	1.615

If emphasis is placed on estimating and testing the individual regression coefficients, then low and uniform variances and covariances of the estimated coefficients are desirable. The diagonal elements of the inverse of $X'X$ are used in the calculation. As seen previously $X'X$ is the sum of squares and cross products formed from the design matrix X . The diagonal elements of the various inverses are given in Table 3. Of striking note is the large value for the interaction term with a 3×3 rotated, reflecting the absence of treatment points in the corners. As before, replication of a treatment design will decrease the magnitude of the inverse elements.

As expected, the results in Tables 1-3 confirm the inverse relationship between the magnitudes of variance error and bias error associated with the several designs. As argued earlier, however, bias error should be given major importance in selecting a treatment design and variance error should be controlled through experimental design, primarily through replication. Looking again at Table 2 it is seen that the modifications by Myers, Escobar and Thompson (designs 3, 6 and 9) have integrated bias values that are about equal and well below the values for the other designs. The central composite and the 7×7 partial factorial form a second group with higher bias errors and the 5×5 factorial, the 3×3 rotated factorial and the 5×5 partial factorial comprise a third group with still larger bias errors. The 3×3 factorial is in a group by itself with an integrated bias error about three times that of the modified designs.

TABLE 4. The thirteen treatment combinations corresponding to a 5×5 partial factorial modification by Escobar, expressed both as coded values and as actual levels of factor A with a range from 0 to 320 kg/ha and factor B with a range from 0 to 240 kg/ha.

Combination	Coded Levels		Actual Levels	
	Factor A	Factor B	Factor A	Factor B
1	-.85	-.85	24	18
2	0	-.85	160	18
3	.85	-.85	296	18
4	-.40	-.40	96	72
5	.40	-.40	224	72
6	-.85	0	24	120
7	0	0	160	120
8	.85	0	296	120
9	-.40	.40	96	168
10	.40	.40	224	168
11	-.85	.85	24	222
12	0	.85	160	222
13	.85	.85	296	222

Although the bias errors of the Myers, Escobar and Thompson modifications are about equal, the corresponding variance errors (Tables 2 and 3) are quite different. On the average the variance error for the 5×5 partial factorial modification by Escobar is only about one third that of the central composite modification by Myers and two-thirds that of the central composite modification by Thompson.

Another advantage of the Escobar modification is that three levels of each factor are studied at each of three levels of the other factor (Figure 3). This characteristic of the design permits a rapid graphic evaluation of the effects of the two factors. In research programs where a treatment design is used at a number of sites and it is necessary to interpret the data before doing a complete analysis, this spatial characteristic of the design is very desirable.

The 5×5 factorial modification by Escobar, in terms of bias error, variance error and spatial characteristics appears superior to the other designs. The treatment combinations corresponding to the Escobar modification are given in Table 4 in terms of a range from -1 to $+1$ for each factor. They may easily be transformed into actual levels as shown in the last two columns in Table 4 once the range of each factor has been determined. A combination of actual zero levels of both factors (a check plot) may be added as a fourteenth combination, if desired.

The selection of the modified 5×5 partial factorial has been based on the premise that the experimenter has as an objective the estimation of a response surface model. The 5×5 partial factorial would not be used if estimation of one degree of freedom factorial contrasts is desired. The selection of the modified 5×5 partial factorial primarily resulted from the empirical study and consequently is dependent on the parameter values of the square root model used as the true model.

SUMMARY

In the planning of a two-factor study, a decision has to be made on the choice of the levels of each factor that will be used in a combination called a treatment. A factor space for two factors is defined as a two-dimensional area that encompasses all possible combinations of interest. The axes will show the range of interest of each factor. Treatment design construction involves the problem of selecting a relatively small number of points in the factor space.

The best treatment design selected will depend on the specific objectives of the experimenter and his knowledge of the production system. For example, it is well known in the single factor case that if the true relationship between the response variable and the treatment variable is a straight line and if the objective is to minimize the variance of the estimated slope, then one should divide the total number of points equally between the two extremes of the treatment variable range. However, this would be a poor design if the true model was not a straight line and a measure of the inadequacy of the estimated model is desired. In treatment

design then two general considerations are involved: variance of estimated parameters and bias, a measure of the differences between the true model and the estimated model.

Past work has shown that designs which are good for minimizing variance error are not the best for minimizing bias error. Unfortunately, in order to study bias, the true model has to be known, a situation usually not found in practice. Consequently, variance considerations have been emphasized in the past and only more recent studies have shown the importance of bias. In fact, the magnitude of bias is sufficiently large that a recommendation has been made in fertilizer use studies to select a treatment design primarily for bias and to control variance error primarily through replication (Cady and Laird, 1969).

The results are given of an empirical study comparing a number of two-factor treatment designs including designs which theoretically have been shown to be best in protecting against bias. A square root model was selected for the true model and a quadratic polynomial used for the estimated or fitted model.

In general the bias, or a measure of the integrated volume between the estimated quadratic polynomial response surface and the true square root response surface was decreased by restricting the treatment combination to an area of the factor space not including the borders of the factor space. This restriction led to larger deviations at the borders but over most of the factor space, hopefully including that area where the economic optimum will occur, the estimated quadratic polynomial more closely approximated the true model. Not having treatment combinations at the corners of the factor space increased the variance of estimated parameters, particularly the variance of the estimated interaction coefficient. It was concluded that, if a partial factorial which has less bias as compared with the complete factorial is used, then extra replication may be included to control the variance error and the total number of plots will be similar.

Three modified designs were best for minimizing bias. Among these designs, the modified 5×5 partial factorial had lower variances for several variance criteria and, in addition, permits a visual evaluation of the response to the two factors at three levels of each of the other factors. The modified 5×5 partial factorial is recommended as the best two-factor design among the nine alternative designs studied.

RESUME

Pour établir le plan d'une étude bifactorielle, il faut prendre une décision en ce qui concerne le choix des niveaux de chaque facteur qui seront utilisés dans une combinaison nommée traitement. L'espace factoriel pour deux facteurs est défini par une surface bidimensionnelle qui englobe toute combinaison possible. Les axes montreront l'intervalle d'intérêt de

chaque facteur. L'élaboration du plan de traitement implique le problème de sélectionner un petit nombre relatif de points dans l'espace factoriel.

Le meilleur plan de traitement sélectionné dépendre des objectifs spécifiques du chercheur et de sa connaissance du système de production. Par exemple, dans le cas d'un seul facteur on sait bien que si la véritable relation entre la variable dépendante et la variable du traitement est une ligne droite, et si l'objectif est de réduire au minimum la variance de la pente estimée, il n'y aura qu'à diviser le nombre total de points également entre les deux extrêmes de l'intervalle de la variable du traitement. Ceci sera, cependant, un plan incomplet si le vrai modèle n'a pas été une ligne droite et si on veut une mesure de l'insuffisance du modèle estimé. Ensuite, dans le plan des traitements sont comprises deux considérations générales: la variance des paramètres estimés et le biais, une mesure des différences entre le vrai modèle et le modèle estimé.

Le travail antérieur a démontré que les bons plans pour diminuer l'erreur de la variance ne sont pas les meilleurs pour diminuer l'erreur du biais. Malheureusement, pour étudier le biais, le modèle véritable doit être connu, une situation qui ne se rencontre pas dans le pratique. Par conséquent, dans le passé les conditions de la variance étaient soulignées et seules des études plus récentes ont souligné l'importance du biais. En effet, la grandeur du biais est tellement suffisamment importante que dans les études sur l'usage de fertilisants, il est recommandé de sélectionner en primordial un plan de traitements par rapport au biais et pour contrôler l'erreur de la variance principalement au travers de répétitions. (Cady & Laird, 1969).

Dans ce travail sont montrés les résultats d'une étude empirique en ce que sont comparés divers plans de traitement bifactoriels, y compris certains qui théoriquement se sont montrés les meilleurs pour se protéger des biais. On a sélectionné un modèle de racine carrée comme le véritable modèle et on a utilisé un polynomial quadratique comme le modèle estimé ou juste.

En général, le biais ou la mesure d'un volume intégré entre la surface de réponse polynomiale quadratique et la surface de réponse de la racine carrée véritable, a été diminué de façon à restreindre la combinaison de traitement en une surface du facteur de l'espace qui ne comprenne pas les limites de ce dit facteur. Cette restriction a amené à de meilleures déviations dans les limites; cependant, dans la majeure partie du facteur de l'espace, —où on espère inclure l'aire dans laquelle se produit l'optimum économique— le polynomial quadratique estimé s'est rapproché plus près du véritable modèle. Le fait de ne pas avoir de combinaisons de traitement dans les angles du facteur de l'espace a augmenté la variance des paramètres estimés, en particulier la variance du coefficient de l'interaction estimée. On a conclu que si on utilise un facteur partiel avec un biais plus petit pour le comparer avec le factoriel complet, on pourra inclure des répétitions extra afin de contrôler l'erreur de la variance et obtenir un nombre similaire de parcelles qu'avec le factoriel complet.

Trois plans modifiés ont été les meilleurs pour diminuer le biais à son minimum. Entre ces plans, le factoriel partiel de 5 x 5 a eu des variances

mineures pour divers critères de variance, et en plus, a permis une évaluation visuelle de la réponse aux deux facteurs des trois niveaux de chacun des autres facteurs. Le factoriel partiel de 5×5 est recommandé comme étant le meilleur plan bifactoriel des neuf plans étudiés.

RESUMEN

En la planeación de un estudio bifactorial, hay que tomar una decisión con respecto a la escogencia de los niveles de cada factor que se usarán en una combinación llamada un tratamiento. El espacio factorial para dos factores se define como una área bidimensional que abarca toda combinación posible. Los ejes mostrarán el intervalo de interés de cada factor. La elaboración del diseño de tratamiento implica el problema de seleccionar un número relativo pequeño de puntos en el espacio factorial.

El mejor diseño de tratamiento que se seleccione dependerá de los objetivos específicos del investigador y de su conocimiento del sistema de producción. Por ejemplo, en el caso de un solo factor se sabe bien que si la verdadera relación entre la variable dependiente y la variable del tratamiento es una línea recta, y si el objetivo es reducir al mínimo la varianza de la pendiente estimada, habrá que dividir el número total de puntos igualmente entre los dos extremos del intervalo de la variable del tratamiento. Este sería, sin embargo, un diseño deficiente si el modelo verdadero no fuese una línea recta y si se desea una medición de la inadecuación del modelo estimado. Luego, en el diseño de los tratamientos están involucradas dos consideraciones generales: la varianza de los parámetros estimados y el sesgo, una medida de las diferencias entre el modelo verdadero y el modelo estimado.

El trabajo anterior ha mostrado que los diseños buenos para disminuir el error de la varianza no son los mejores para disminuir el error del sesgo. Desafortunadamente, para estudiar el sesgo, el modelo verdadero debe ser conocido, una situación que ordinariamente no se encuentra en la práctica. Por consiguiente, en el pasado se han subrayado, las consideraciones de la varianza y sólo los estudios más recientes han señalado la importancia del sesgo. En efecto, la magnitud del sesgo es tan suficientemente grande que en los estudios sobre uso de fertilizantes se ha recomendado seleccionar un diseño de tratamientos primordialmente con respecto al sesgo y para controlar el error de la varianza principalmente a través de repeticiones (Cady y Laird, 1969).

En este trabajo se muestran los resultados de un estudio empírico en el que se comparan varios diseños de tratamiento bifactoriales, inclusive algunos que teóricamente han mostrado ser los mejores para proteger contra sesgos. Se seleccionó un modelo de raíz cuadrada como el modelo verdadero y se usó un polinomial cuadrático como el modelo estimado o ajustado.

En general, el sesgo —o la medida de un volumen integrado entre la superficie de respuesta polinomial cuadrática estimada y la superficie de respuesta de raíz cuadrada verdadera— fue disminuído al restringir la combinación de tratamiento a una área del factor de espacio que no incluía los límites de dicho factor. Esta restricción llevó a mayores desviaciones en los límites; empero, sobre la mayor parte del factor de espacio —donde se espera incluir el área en la cual ocurre el óptimo económico— el polinomial cuadrático estimado se aproximó más de cerca al modelo verdadero. El hecho de no tener combinaciones de tratamiento en las esquinas del factor de espacio aumentó la varianza de los parámetros estimados, en particular la varianza del coeficiente de interacción estimado. Se concluyó que, si se usa un factorial parcial que tenga menor sesgo al comparársele con el factorial completo, se podrán incluir repeticiones extras a fin de controlar el error de la varianza y obtener un número similar de parcelas que con el factorial completo.

Tres diseños modificados fueron los mejores para disminuir al mínimo el sesgo. Entre estos diseños, el factorial parcial de 5 x 5 tuvo menores varianzas para varios criterios de varianza y, además, permite una evaluación visual de la respuesta a los dos factores a tres niveles de cada uno de los demás factores. El factorial parcial de 5 x 5 se recomienda como el mejor diseño bifactorial entre los nueve diseños que se estudiaron.

REFERENCES

- Anderson, R. L. 1956. A comparison of discrete and continuous models in agriculture production analysis. In Baum, E. L., Heady, E. O. and Blackmore, J., Eds., *Methodological Procedures in the Economic Analysis of Fertilizer Use Data*, pp. 39-61. Iowa State College Press, Ames, Iowa.
- Anderson, R. L. 1957. Some statistical problems in the analysis of fertilizer response data. In Baum, E. L., Heady, E. O., Pesek, J. T. and Hildreth, C. G., Eds., *Economic and Technical Analysis of Fertilizer Innovations and Resource Use*, pp. 187-206. Iowa State College Press, Ames, Iowa.
- Box, G. E. P. and Draper, N. R. 1959. A basis for the selection of a response surface design. *Jour. Amer. Stat. Association* 54:622-654.
- Box, G. E. P. and Draper, N. R. 1963. The choice of a second order rotatable design. *Biometrika* 50:335-352.
- Box, G. E. P. and Hunter, J. S. 1957. Multi-factor experimental designs for exploring response surfaces. *Ann. Math. Stat.* 28:195-241.
- Box, G. E. P. and Wilson, K. B. 1951. On the experimental attainment of optimum conditions. *Jour. Roy. Stat. Soc., Series B* XIII:1-45.
- Box, M. J. and Draper, N. R. 1971. Factorial designs, the $|X'X|$ criterion, and some related matters. *Technometrics* 13:731-742.
- Cady, F. B. and Laird, R. J. 1969. Bias error in yield functions as influenced by treatments design and postulated model. *Soil Sci. Soc. Amer. Proc.* 33:282-286.
- Cady, F. B. and Laird, R. J. 1972. Comparison of treatment designs for two factor response surface experiments. Unpublished manuscript.

- Cochran, W. G. and Cox, G. M. 1957. *Experimental Designs*. Second Edition. John Wiley and Sons, Inc., New York.
- Draper, N. R. and Herzberg, A. M. 1971. On lack of fit. *Technometrics* 13:231-241.
- Draper, N. R. and Lawrence, W. E. 1967. Sequential designs for spherical weight functions. *Technometrics* 9:517-530.
- Draper, N. R. and Smith, H. 1966. *Applied Regression*. John Wiley and Sons, Inc., New York.
- Escobar, J. A. 1967. Consideraciones sobre la comparación de diseños de tratamientos. Unpublished M.S. thesis. Library, Colegio de Postgraduados, Chapingo, Mexico.
- Folks, J. L. 1958. Comparison of designs for exploration of response relationships. Unpublished Ph.D. dissertation. Library, Iowa State University.
- Fuller, W. A. 1962. Estimating the reliability of quantities derived from empirical production functions. *Jour. Farm. Econ.* 44:82-99.
- Heady, E. O., Pesek, J. T. and Brown, W. G. 1955. Crop response surfaces and economic optima in fertilizer use. *Iowa Agr. Exp. Sta. Res. Bul.* 424.
- Heady, E. O. and Dillon, J. L. 1961. *Agricultural Production Functions*. Iowa State University Press, Ames, Iowa.
- Heady, E. O., et al. 1961. Status and methods of research in economic and agronomic aspects of fertilizer response and use. Publ. 918, Nat'l. Acad. Sci., Nat'l. Res. Council, Washington, D. C.
- Hildreth, C. G. 1956. Discrete models with qualitative restrictions. In Baum, E. L., Heady, E. O. and Blackmore, J., eds., *Methodological Procedures in the Economic Analysis of Fertilizer Use Data*, pp. 62-75. Iowa State College Press, Ames, Iowa.
- Kempthorne, O. 1965. Development of the design of experiments over the past ten years. In *Proceedings of the tenth conference on the design of experiments in army research development and testing*. ARO-D Report 65-3. War Office.
- Kiefer, J. 1959. Optimum experimental designs. *Jour. Roy. Stat. Soc., Series B*, 21:273-319.
- Myers, R. H. 1971. *Response Surface Methodology*. Allyn and Bacon, Inc., Boston.
- Nalimov, V. V., Golikova, T. I. and Mikeshina, N. G. 1970. On practical use of the concept of D-optimality. *Technometrics* 12:799-812.
- Nelder, J. A. 1966. Inverse polynomials, a useful group of multifactor response functions. *Biometrics* 22:128.
- Ott, L. and Mendenhall, W. 1972. Designs for estimating the slope of a second order linear model. *Technometrics* 14:341-353.
- Stigler, S. M. 1971. Optimal experimental design for polynomial regression. *Jour. Amer. Stat. Ass'n.* 66:311-318.
- Tejeda, H. R. 1966. Comparison of a rational fraction model with a polynomial to describe a factor-yield relationship. Unpublished M.S. thesis. Library, Iowa State University.
- Thompson, W. O. 1971. An investigation of secondary criteria in the selection of minimum bias designs in two variables. Technical Report 20, Dept. of Statistics, University of Kentucky.
- Yates, F. 1937. The design and analysis of factorial experiments. Tech. Comm. No. 35, Imperial Bureau of Soil Science, Harpenden, England.

