DAVID D. MASON
North Carolina State College

Chapter 5

Functional Models and Experimental Designs for Characterizing Response Curves and Surfaces

THE yield of a particular crop is a function of many possible factors, as has been pointed out in Chapters 1 and 2. The climate, variety, management practices, and soil factors are, in fact, broad categories which in themselves contain a number of subfactors, each of which may be modifying or limiting. This chapter is concerned primarily with the functional relationship between yield and a portion of the soil factor, that relating to the nutrient status of the soil.

Background

Even a superficial examination of the numbers and types of factors affecting crop yield will reveal that any function completely describing the relationship would be extremely complex. It is small wonder that widely different hypotheses have been developed and supported, since one may find almost any pattern of response, varying from strong positive linear relations to strong negative linear relations. From a statistical standpoint, the failure of hypotheses, purporting to have general application, to agree arises from failure of the experimenters to adequately sample the population to which inferences are made.

Russell (25) gives an excellent review of the historical development of the concepts of plant nutrition, and of the attempts to obtain rational explanations of various phenomena. Liebig, with his first publication in 1840 and subsequent papers and books on the subject, together with his heavy ridicule of the efforts of his predecessors and contemporaries, contributed much, particularly in the way of stimulating controversy and subsequent research. His law of the minimum, which he stated as "by the deficiency or absence of *one* necessary constituent, all the others being present, the soil is rendered barren for all those crops to the life of which *that one* constituent is indispensable," is perhaps his best remembered contribution.

The field experiment approach to the problems of plant nutrition and response initiated by Boussingault (about 1834) and Lawes and Gilbert in 1843 furnished positive evidence of the response of crops to natural and artificial manures. However, Russell reports that the controversy regarding the use of "chemical manures" went on for many years before their general acceptance was indicated. Even today a remnant of this controversy is evidenced by the "organic gardening" school of thought.

Mitscherlich's contributions, beginning in the first decade of this century, marked the first major attempt to formulate a general functional model. His experiments were made with plants grown in sand cultures supplied with "excess" of all nutrients excepting the one under investigation. His expression is commonly known by the descriptive term, "law of diminishing returns," and has the mathematical properties outlined in Chapter 1. Mitscherlich's work, like Liebig's, produced controversy and has both ardent supporters and critics. His function, together with modifications and contributions by other workers, will be given more quantitative expression in the following section. Spillman (26) later, but independently, developed the same function (in the algebraic form of equation 2 in Chapter 1) and extended the methodology to computation of economically optimum rates of fertilization. Spillman, as did Mitscherlich, suggested optimum experimental designs for obtaining data necessary for the estimation of the parameters of the model.

Anderson has adequately outlined, in Chapter 3, the procedural developments from the standpoint of the statistical approach of developing empirical polynomial functions to characterize the response. The development of the factorial experiment and appropriate methods of statistical analysis led to the definition and characterization of interaction between factors (also called complementarity, or joint effects). This, in turn, has led to the geometrical concept of a response surface as the realistic expression of the contribution of two or more nutrients to yield.

With the increased interest of production economists in the application of quantitative methods in the past several years, several papers have been concerned with the choice of a proper functional model for the characterization of input-output relationship in plant growth. Johnson (17), Heady (11), McPherson (18), Ibach and Mendum (16), Paschal (22), Hutton (14), Hutton and Elderkin (15), and Heady, Pesek, and Brown (12) have set forth, in varying degrees, bases of comparison and procedures for evaluation.

Functional Models for Single-Variable Response Curves

Two general approaches have been used in developing mathematical expressions for the relationship between the amounts of the various factors present, and the amounts of plant growth. They are:

1. Attempts to define a model which expresses basic laws of plant behaviour, and fitting the experimental data to this more or less rigid model.
2. The experimental data are studied by statistical methods and an empirical polynomial equation of "best fit" is developed, with no assumption or hypothesis as to the underlying causes.

The first approach is logically and intuitively more appealing. It has its counterpart in the simple physical and chemical systems where deterministic models are common and useful. However, even the simplest of biological systems is relatively complex, and together with errors of technique and measurement, exact relationships are to be viewed askance. Some of the more common functional models for which some biological justification has been claimed are first considered in the following paragraph.

The Mitscherlich Function

Expressing quantitatively the statement that the increase of crop produced by unit increment of the lacking factors is proportional to the decrement from the maximum, one has:

$$\frac{dy}{dx} = (A - y)c \tag{1}$$

where y is the yield obtained when x = the amount of the factor present, A is the maximum yield obtainable if the factor were present in excess, this being computed from the equation, and c is a constant. Upon integration, and assuming that y = O when x = O,

$$y = A(1 - e^{-cx}). \tag{2}$$

Mitscherlich maintained that the "c" values in his expression were constant for a given nutrient over different crops and growing conditions. Most of the early controversy about his work centered around his hypothesis concerning the "c" values. The workers subsequently mentioned as using the Mitscherlich-type equation have assumed that "c" is a parameter to be estimated from the data. This function is expressed in other algebraic forms by Spillman (26) and Stevens (27), and has been widely used by many workers. Ibach and Mendum (16) have detailed instructions for computations, together with examples, using the Spillman form. Monroe (19), Pimentel-Gomez (24), and Stevens (27) give simplified least squares procedures for estimation of parameters for solution, when the X levels are equally spaced. Also, standard errors may be computed for the estimated parameters.

Prior to the comparatively recent publication of the three references mentioned above, and a paper by Hartley (10), least squares estimates involved such heavy labor that they were seldom made. An interesting example of the application of the Mitscherlich model is given by Crowther and Yates (6), in summarizing all published results of one-year fertilizer experiments conducted in Great Britain and the northern European countries since 1900, in order to formulate a wartime fertilizer distribution policy. Economic analyses, in terms of setting out optimum rates for maximum profit, were made of the data.

One of the other early criticisms of Mitscherlich's equation was that no allowance was made for possible yield depression by harmful excesses of the factor. Mitscherlich, after extensive study of his experimental data, introduced a modification of the following form to allow for such depressions:

$$y = A(1 - 10^{-cx})10^{-kx^2}, \tag{3}$$

with the constant "k" being called the "factor of injury." He felt that this would apply mainly to the response of grain crops to nitrogen. He provided estimates of "k" for several crops.

The Logistic Function

The logistic is the commonly used function for fitting growth curves in biological populations, and may be expressed in the form:

$$y_t = \frac{k}{1 + b\,e^{-at}}, \tag{4}$$

where a, b, and k are parameters to be estimated from the observed data, and y_t is the value of the growth character studied at point of time, t. For yield response models, x, for increment of fertilizer, would be substituted for t.

This curve has a lower asymptote of O, and an upper asymptote at k, and the point of inflection is at $y = \frac{k}{2}$, a point midway between the two asymptotes. Thus, we have the familiar S-shaped or sigmoid curve. Such a model would be useful to characterize the initial "lag" that may occur when the amount of the factor in the soil is very low, and small increments are applied in the low range. In the usual situation this initial lag is not observable. Nair (21) gives an extensive discussion of the logistic function together with methods and illustrations of fitting.

The Power Function (Cobb-Douglas)

The power function,

$$Y = a\,X^b, \tag{5}$$

has been employed as the model in various economic investigations. In this equation, Y is the yield, a and b are constants, with X as the level of the factor. The equation may be written in the linear form as

$$\text{Log } Y = \text{Log } A + b \,.\, \log X \,. \tag{6}$$

Hutton et al. (15) discuss the general characteristics of the Cobb-Douglas function, and suggest methods of analysis. Heady (11) and McPherson (18) also describe the various characteristics of this function and modified forms of the power function. If $b > O$, as would be the case in the yield response curve, y continues to increase as X increases.

The Polynomial

The terms in a polynomial equation may vary from one to n-1 where n is the number of levels of the factor X. In the single variable case, the number of terms and the degree of the equation are normally (but not necessarily) parallel. The first degree (or linear) equation describes a straight line, while the second degree (or quadratic) describes a monotonic curve. The degree less one indicates the number of times the curve may change direction. The usual forms are:

Linear : $Y = \beta_0 + \beta_1 X$

Quadratic: $Y = \beta_0 + \beta_1 X + \beta_{11} X^2$

Cubic : $Y = \beta_0 + \beta_1 X + \beta_{11} X^2 + \beta_{111} X^3$

General : $Y = \beta_0 + \beta_1 X + \ldots\ldots \beta_{\cdot\cdot(n-1)} X^{\cdot\cdot(n-1)}.$

The X may be transformed to the square root, logarithm, reciprocal, or other form, with the same general process of fitting applied. Methods of fitting such curves are straightforward. Discussion of fitting procedures, with examples, is given by Anderson and Bancroft (2) and other texts.

Discussion of Application of Exponential, Power, and Polynomial Models

The functions mentioned above are only a few of the better known of a large number of possible functions. Within the polynomial class alone an almost infinite number of possibilities exist. The problem, therefore, of choosing the "best" function is not soluble from a simple set of rules. By the use of least squares procedures the value of the constants for the equations may be computed. These procedures give the "best" fit for the particular form of functional model, in the sense of describing a curve from which the mean of the squares of the deviations of the individual points from that curve are a minimum.

It cannot be claimed that any of the functions represent fundamental biological laws of growth, although one may rationalize the form of a particular function in a particular situation. One procedure of choosing the "best" function, mentioned by Heady (11) and by Hutton et al. (15), is to examine possible applicable functions, and select the one that best fits the data. A useful procedure, where data are being examined from a replicated experiment (more than one observation at each increment), is to examine the size of the "lack of fit" term, as given in the analysis of variance. The following data, from Veits, Nelson, and Crawford (28), serves to illustrate the procedure.

If the lack of fit term is of the same order of magnitude as the experimental error, then the function is characterizing the data adequately. A significant lack of fit mean square indicates that the model is inadequate to describe the functional relationship.

TABLE 5.1. Observed and Predicted Yields by Three Functions For Corn Yields, 1952

Nitrogen Level (Lbs/A)	Observed Yields (Bu/A)	Estimated Yields Polynomial Equation	Estimated Yields Mitscherlich Equation	Estimated Yields Cobb-Douglas Equation
0	125.8	124.2	126.5	124.2
40	140.2	145.6	145.7	152.3
80	166.8	160.11	156.9	158.1
120	164.3	167.6	163.3	161.6
160	168.5	167.8	167.1	164.1
200	161.8	161.6	169.2	166.2

Source: Veits, et al. (28)

Analysis of Variance:

Source	d.f.	Polynomial M. S.	Mitscherlich M. S.	Cobb-Douglas M. S.
Treatments	5	1219.82	1219.82	1219.82
Due to regression	2*	2880.82	2673.74	2510.28
Lack of fit	3	112.48	250.53	269.63
Experimental error	21	152.6	152.6	152.6

*1 d.f. for regression for Cobb-Douglas, and 4 d.f. for deviation from regression.

Johnson (17) and Heady et al. (12) have examined the three functional models considered in table 5.1, for fitting response curves, and conclude that the quadratic polynomial model generally gives the better fit. Heady, Pesek, and Brown found that fit was improved by using a square root transformation of the X variate, in the quadratic model.

The Mitscherlich and Cobb-Douglas functions obviously give a poor fit when yield is depressed by the higher rates of application. This depression appears to be fairly common, particularly with higher rates of nitrogen. For example: a recent report by Hunter and Youngen (13), on a series of experiments on corn, shows that in six of seven experiments, where N rates were carried to 200 lbs. per acre, a yield depression resulted. This type of response is compatible with biological theory, although depression is more marked in cases of excesses of some of the minor elements.

An alternative that might be followed would be to discard those observations which fall beyond the maximum yield, and fit the exponential or power function, using the rationale that one obviously is not interested in that area of the curve. This would appear to be a poor practice statistically, since one is discarding information and introducing a degree of subjectiveness into the analysis.

One particular advantage claimed for the Mitscherlich function is that it gives plausible results when extrapolated for high values of X. Stevens (27) sharply warns against such extrapolations, and points out that the standard error of the predicted value becomes large as the asymptote is approached. It is necessary to have established the absolute generality of the formula, either by sound theoretical justification, extensive observation, or both, in order to extrapolate with confidence. Stevens points out that this is feasible in some physics models (e.g., Newton's law of cooling), but difficult in biology. Also, in viewing the general trend of economic-agronomic cooperation in experiments, it appears that the need for extrapolation will lessen.

Two more or less ulterior advantages may be claimed for the polynomial. First, it is easier to fit by least squares procedures and easier to provide estimates of standard errors of the parameters. Second, it is the most flexible of the three functions. This carries the added advantage of therefore being more generally applicable to a series of individual experiments conducted at a number of locations and years.

Functional Models for Characterizing Response Surfaces

The three functions considered in table 5.1 may be generalized to give a mathematical expression of the geometrical configuration of a response surface when two or more factors are considered. The principal points of contention regarding the relative suitability of the functions again centers on the restrictions placed on the form of the surface. The Cobb-Douglas and the Mitscherlich functions, as previously mentioned, do not have a declining phase and do not permit the reflection of changing ratios of nutrients for the optimum treatment when the level of yield is changed. Heady (11) illustrates this relationship diagrammatically by showing that the isoclines (a line connecting all points on the same slope of successive isoquants) are required to be linear.

Sufficient data have been accumulated from factorial experiments with fertilizers to give some indication of the nature of the interaction (complementarity) between nutrients. For example, Dumenil and Nelson (7) report on the results of 164 factorial experiments carried out in Iowa on corn, oats, and hay crops with N, P, and K, or two of the three nutrients in combination. Out of these, 62 showed some type of interaction significant. (It is likely that a greater number would have been found significant had individual degrees of freedom associated with particular coefficients been examined). Commonly the interaction between N and P was positive, while negative interactions were found between N and K. The authors conclude: "In view of the number and size of interactions encountered, the use of the factorial design, wherein the different fertilizer elements and rates are used in all possible combinations, appears highly desirable. The value of certain nonfactorial designs now in common use may lead to erroneous conclusions whenever interaction between the fertilizer elements occurs."

The Mitscherlich-Baule Function for Response Surfaces

Baule, according to Russell (25), generalized Mitscherlich's function while retaining the fundamental assumptions. He supposes that each of the factors influencing plant growth is in accordance with Mitscherlich's assumption and that the final yield is the product of the separate expressions. The equation then becomes:

$$y = A(1-e^{-c_1x_1})(1-e^{-c_2x_2})(1-e^{-c_3x_3}). \tag{7}$$

The equation requires, for example, that if two factors, L and M, vary simultaneously, each should produce its own effect independent of the other. Russell (25) illustrates this with the following reasoning and with data adapted from Mitscherlich's publication. If y and y′ represent the yields when x, x′ are the quantities of factor L, the quantity M remaining constant, then

$$y = A(1-e^{-cx})$$
$$y' = A(1-e^{-cx'})$$

where A is the maximum yield obtainable with any quantity of factor L at the given value of M. Now

$$\frac{y}{y'} = \frac{(1-e^{-cx})}{(1-e^{-cx'})} .$$

This ratio is therefore independent of the value of A; that is, it is independent of the level at which M was taken.

TABLE 5.2. Yield of Oats in Pot Experiments with Varied Phosphate Dressings and Varied Water

Calcium Phosphate (x)	Water 1 dose (y)	Water 2 doses (y′)	Ratio $\frac{y'}{y}$
0	6.4	11.0	1.72
1	14.6	25.6	1.75
2	22.6	36.6	1.62
4	29.7	53.1	1.79
8	41.3	70.5	1.71
16	50.8	77.5	1.53
32	55.7	88.5	1.59

Adapted from Mitscherlich.

It seems clear from this illustration that the Baule function accounts for "interaction" in the sense that the interaction arises from the failure of the *differences* between y and y′ to remain constant over the different levels of x. Another criterion of the applicability of the Mitscherlich-Baule function is a test of the deviations from parallelism of the curves when y is plotted as the logarithm of the yield increase, due to the i-th increment, and x as the log of the dose.

The Maskell "Resistance" Formula

Balmukand (3), not satisfied with the Mitscherlich function when applied to field data, critically examined the general problem of relating nutrient level to crop yield. He applies Maskell's formula, which by electrical analogy has been called the Resistance Formula. It may be stated as: Each activity of the plant (yield, etc.) is determined by a potential and a set of resistances, each of which represents one of the external factors. Maskell expresses the effect of nutrient supplies on yield as

$$\frac{1}{y} = F(N) + F'(P), + F''(K) + \dots + C, \tag{8}$$

y being the yield, and F(N), F′(P) and F″(K) being functions of these nutrients supplied, and they have the form,

$$\frac{a_n}{n+N}, \quad \frac{a_p}{p+P}, \quad \frac{a_k}{k+K},$$

where N, P, and K represent the amounts of these nutrients added; n, p, and k represent the amounts of the nutrients available in the soil, and a_n, a_p, and a_k are constants expressing the importance of the nutrients to the crop.

This expression, like Mitscherlich's, assumes each factor acts independently of all the others but fixes the difference of $\frac{1}{y} - \frac{1}{y}$ as constant.

Balmukand (3) illustrates the application of the function to data from replicated factorial experiments and gives least squares procedures for estimating the constants, together with appropriate estimates of the standard errors of the estimated constants. He obtains satisfactory fits of the response surface, using the magnitude of the lack of fit mean square as the criterion. However, the computations involved are heavy, compared to other functions considered. This may be the primary reason why this function has not been used.

The Cobb-Douglas Function

The Cobb-Douglas function may be generalized to

$$Y = aX_1^b \; X_2^c \dots X_n^n \tag{9}$$

where n is the number of factors considered. As mentioned previously, it may be put into the form:

$$\text{(10)} \qquad \text{Log } Y = \log a + b \log X_1 + c \log X_2 + \ldots + n \log X_{n-1}$$

for solution by least squares.

The Polynomial Function

The polynomial may take a wide variety of forms for a given number of factors, depending on the degree (highest exponent or products of exponents in a given term), and the scale in which the X variates are expressed. Some experience has been accumulated in the past few years on this model, both in the biological field and in industrial and engineering applications. Box (4) and Anderson (1) have reviewed the general approach to defining response surfaces and defining optimum operating conditions, using the general polynomial equation. Hanson, Hutton, and Robertson (9) have examined data from a $5^3 \times 2$ factorial experiment, with N, P, K, and lime as factors, and indicate the second degree polynomial equation is generally satisfactory. Heady et al. (12), after a detailed examination of possible functions, concluded that the generalized second degree polynomial with the X's scaled by a square root transformation was most satisfactory.

The generalized polynomial equation for two variables is

$$\text{(11)} \qquad y = b_0 + (b_1x_1 + b_2x_2) + (b_{11}x_1^2 + b_{22}x_2^2 + b_{12}x_1x_2) + (b_{111}x_1^3 + b_{222}x_2^3 + b_{112}x_1^2x_2 + b_{122}x_1x_2^2) + \text{etc.}$$

This corresponds to representing the function by its Taylor series. The brackets enclose the terms containing respectively the first, second, and third order terms. Thus, an equation containing all first order terms only defines a plane; one containing both the first and second order terms is a second degree equation and defines a quadratic surface, and so on. The number of constants to be fitted for functions of various numbers of factors for varying degrees is given in table 5.3, taken from Box (4).

TABLE 5.3. Number of Constants To Be Fitted for Equations of Varying Degree

Number of	Degree of Fitted Equation			
Factors	1	2	3	4
2	3	6	10	15
3	4	10	20	35
4	5	15	35	70
5	6	21	56	126

The equation is then fitted by the method of least squares (multiple regression). The procedure is in essence the application of multiple regression methods to observations in which the values for the independent variates have been planned. By proper choice of these levels the calculations may be simplified, particularly when the x variates may be coded, with the origin (O) at the center of the design.

Example of a Fitted Second Degree Response Surface

An example was chosen from summary tables presented by Hunter and Youngen (13), from a 4 x 4 factorial experiment, with nitrogen level and spacing (number of plants per acre) as the two factors. The experiment was run as a randomized block, with three replications. The above-mentioned authors have kindly supplied the necessary additional information about the experiment, including treatment totals and the analysis of variance, in order to allow a complete analysis of the data. The treatment means are given in table 5.4 below.

TABLE 5.4. Yields, Bushels per Acre, As Influenced by Variation in Plant Stand and N Levels

Nitrogen	Plant Population		(No. Plants per Acre)		
Lbs./A	15,400	17,000	17,900	21,500	Mean
0	106.1	96.7	94.6	98.0	98.8
50	121.9	120.4	126.4	110.8	119.8
100	128.9	129.3	134.2	130.4	130.7
150	119.4	134.5	138.4	140.2	133.1
Mean	119.0	120.2	123.4	119.8	

Second degree (or quadric) equations of the following forms were fitted:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 \tag{12}$$

$$y' = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1 + \beta_{22} X_2 + \beta_{11} X_1 X_2 \ . \tag{13}$$

The following equations were obtained:

$$y = 128.19 + 5.6825\, X_1 + .272864\, X_2 - 1.1624\, X_1^2 - .08675\, X_2^2 + .345165\, X_1 X_2 \tag{14}$$

(a unit of X_1 = 25 lbs. N; unit of X_2 = 500 plants/A)

$$y' = 123.34 + 19.4913\, X_1 + 1.3692\, X_2 - 2.68875\, X_1 - 2.30082\, X_2 + 5.25987\, X_1 X_2 \tag{15}$$

(a unit of X_1 = 50 lbs. N; unit of X_2 = 100 plants/A).

These two equations have their origin at the mean levels of N and spacing. An analysis of variance of the data is given in table 5.5. The F test of the lack of fit term for both equations shows that neither are significant at the 5 percent significance level, although the function with the X's in linear form indicates slightly better fit. No generalization, however, should be made from such a single comparison, particularly in view of other workers, Heady et al. (12) having indications to the contrary.

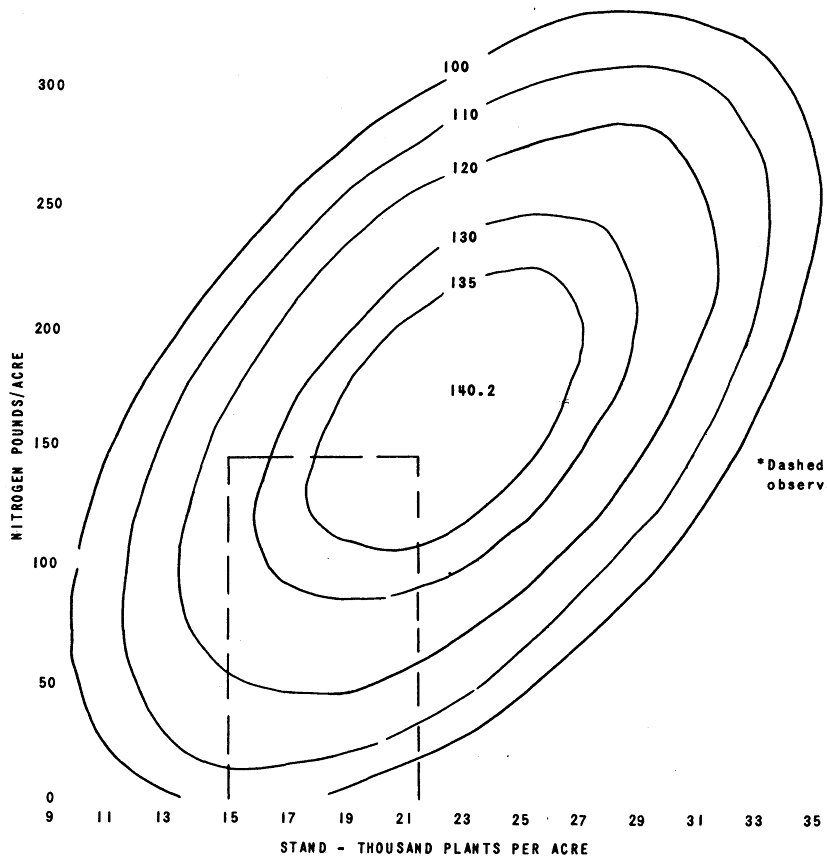
TABLE 5.5. Analysis of Variance of 4 x 4 Factorial Experiment (N Levels x Spacing)

Source of Variation	d.f.	M.S.	M.S. for $\sqrt{\ }$ Transform of X Variate
Replications	2	372.22	372.22
Treatments	15	683.16	683.16
Due to regression	5	1888.44	1823.97
Lack of fit	10	80.52	112.76
Experimental error	30	53.99	53.99
Total	47		

Figure 5.1 is given to illustrate the general picture of the joined yield contours as computed from equation 14. These joined contours show the symmetry required by the function used for fitting. However, the size of the standard errors of predicted points that are much removed from the area of the experimental observations clearly shows that such extrapolation is of little practical value. The importance of having the experimental points in the region of interest is indicated by considering the size of the seven standard errors of predicted yields (Y's) listed with the figure.

Figure 5.2 shows the portion of the yield contours (also computed from equation 14) within the area of experimental observations, together with the observed mean yields for the 16 treatment combinations. Figure 5.3 shows similar contours but computed from the "square-root" equation 15.

This example, incidentally, illustrates the importance of the consideration of subsidiary factors in fertilizer experiments. The data and figures 5.1 and 5.2 show that the response to nitrogen is appreciably modified by the choice of plant population.



STANDARD ERRORS OF SELECTED POINTS

	X_1	X_2	$\hat{Y}$	S.E.$\hat{Y}$	
1.	0	15.40	102.31	3.18	a
2.	75	17.95	128.19	2.15	b
3.	170	22.52	140.20	5.94	c
4.	250	25.40	131.78	19.91	d
5.	100	22.14	131.78	3.24	d
6.	100	17.32	131.78	1.89	d
7.	250	26.01	131.78	21.74	d

a. Lowest treatment combination in Expt.
b. Mean of treatment levels.
c. Maximum of surface.
d. Selected points on 131.78 yield contour.

*Dashed line encloses area of experimental observations. See Figure 5.2 for yields.

Fig. 5.1 — Yield contours (bushels per acre, corn) for 4 x 4 factorial experiment. Hunter and Youngen (13).

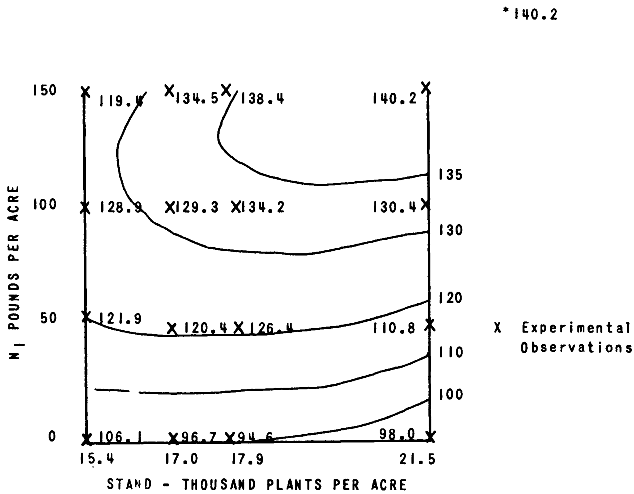


Fig. 5.2 —Yield contours, 4 x 4 factorial experiment with corn. (Data are in bushels per acre.) X's on linear scale.

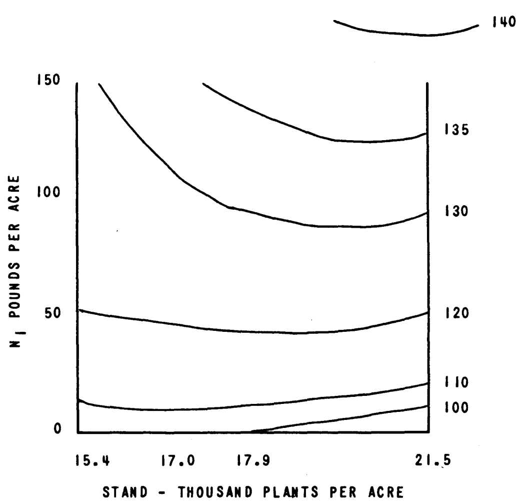


Fig. 5.3 — Yield contours, 4 x 4 factorial experiment, with corn, for square root transformation of X's.

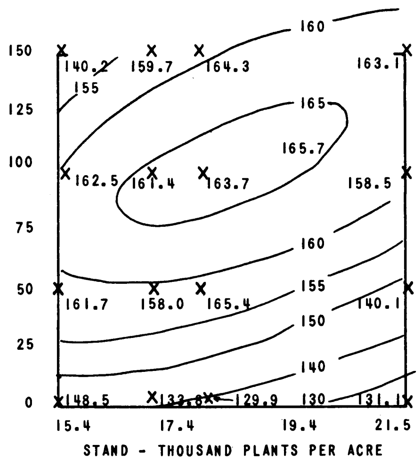


Fig. 5.4 — "Value" contours, after adjustment for cost of fertilizer and stand. Assumed constants: Corn = \$1.40 per bushel; nitrogen = 18 cents per pound; stand = \$1.00 per 1000 plants per acre.

Equation:

$$Y = 143.229 + 21.5042X_1 - 0.6609X_2 - 6.5014X_1^2 - 0.4834X_2^2 + 1.9245X_1X$$

where Y = predicted value (dollars per acre) above cost of fertilizer and stand.

$$X_1 = \frac{\text{lbs. N}}{50}$$

$$X_2 = \frac{\text{stand thousands per acre}}{1000}$$

β_0 in the equation is the expected yield at N = O, and stand = 15,400

The maximum point is:

X_1 = 110 pounds nitrogen
X_2 = 19,100 plants
$\hat{Y}$ = \$154.67, corresponding to a yield of 135.13 bushels per acre.

In figures 5.1 and 5.2 it may be seen that the predicted maximum yield was estimated to occur with a stand of 22,520 plants per acre and nitrogen application of 170 pounds N per acre. If no cost were attached to the N application and increased stand (an obviously rare situation), this point would be in the center of the region of maximum interest. However, consideration of the price of the N applied and the cost of increased stand would normally be expected to shift this region. Figure 5.4 shows the "value" contours, after accounting for the cost of N as 18 cents per pound and the cost of an additional 100 plants per acre as \$1.00. The value of the corn was computed as \$1.40 per bushel. (No overhead costs were considered; if constant overhead charges were assumed, the contour surface would not be changed, but the value attached to each contour would be decreased a constant amount). The "maximum return" of 165.67 is now estimated to occur at 19,100 plants per acre, with 110 pounds N. This corresponds to a yield of 135.1 bushels per acre. This point occurs well within the region of the experimental observations, a situation which is desirable.

Figure 5.4 and the accompanying computations are for illustrative purposes only. Heady et al. (12) and Chapters 1, 6, and 10 outline direct methods for estimating the economic optima, given a function, and input and output costs.

TABLE 5.6. Rates and Coded Values Used in Potato Fertility Experiment

Nutrient	Rates of Fertilizer Element for Coded Value				
	-2	-1	0	+1	+2
N	0	50	100	150	200
P_2O_5	0	60	120	180	240
K_2O	0	50	100	150	200

Table 5.7. Treatments, and Treatment Means for Three Replications, Potato Fertility

Treatment Number	X_1 N	X_2 P_2O_5	X_3 K_2O	Treatment Means (Lbs. U. S. No. 1's per 2-row, 25-ft. Plot)
1	-2	+1	+1	31.9
2	-1	+1	+1	41.5
3	0	+1	+1	54.2
4	+1	+1	+1	57.1
5	+2	+1	+1	51.9
6	+1	-2	+1	56.2
7	+1	-1	+1	56.0
8	+1	0	+1	49.8
9	+1	+2	+1	61.6
10	+1	+1	-2	45.6
11	+1	+1	-1	50.5
12	+1	+1	0	55.4
13	+1	+1	+2	53.2
14	-1	-1	-1	42.6
15	+1	-1	-1	52.3
16	-1	+1	-1	39.3
17	-1	-1	+1	31.4
				$\overline{X} = 48.6$

Example of a Response Surface for Three Factors Estimated from a Multifactor Design

The following data are from a multifactor NPK fertility experiment conducted in Watauga County, North Carolina, in 1954, by M. E. Harward. Table 5.6 gives the rates of N, P_2O_5, and K_2O used, together with their coded values given in the preceding table and following figure. The basic arrangement of the treatment combinations is that of having five levels of each of the nutrients, at a constant rate of the remaining two, with additional points added to form a 2^3 factorial design. Had the constant rate been in the center of the design (e.g., $P_2O_5 = o$, $K_2O = O$, for the rates of N) then it would come in the category of the central composite designs described by Box (4). With this point on one of the corners of the cube, Box describes this as a "noncentral" composite. Table 5.7 lists the treatment means (in terms of pounds of U. S. No. 1's per 2-row, 25-foot plot), together with the coded treatment combinations given in table 5.6. Table 5.8 gives the prediction equation together with the standard errors of the coefficients and the analysis of variance.

TABLE 5.8. Regression Coefficents and Their Standard Errors, and the Analysis of Variance for Second Degree Surface for Data in Table 5.7

Designation	Regression Coefficient ± Standard Error		
b_0	48.59	±	2.103
b_1	6.690	±	.943
b_2	1.254	±	.943
b_3	-.270	±	.943
b_{11}	-2.628	±	.832
b_{22}	1.832	±	.832
b_{33}	-1.231	±	.832
b_{12}	-1.640	±	.858
b_{13}	.786	±	.858
b_{23}	1.850	±	.858

Analysis of Variance:

Source of Variation	d.f.	M.S.
Replications	2	8.14
Treatments	16	232.88
Linear	3	882.08
Quadratic	6	145.52
Lack of fit	7	29.53
Experimental error	32	32.86

Figure 5.5 shows the geometrical configuration of the design. Figure 5.6 illustrates the yield contours for $X_1(N)$ and $X_2(P_2O_5)$, with $X_3(K_2O)$ fixed at the +1 (150 lb.) level. Thus, these contours apply to the front of the cube, and the extended points of N and P_2O_5, and the yields plotted on figure 5.6 may be located with respect to their position in figure 5.5.

Figure 5.6 illustrates the contours of a surface which is obtained when b_{11} and b_{22} are of opposite signs, and is termed a "saddlepoint"

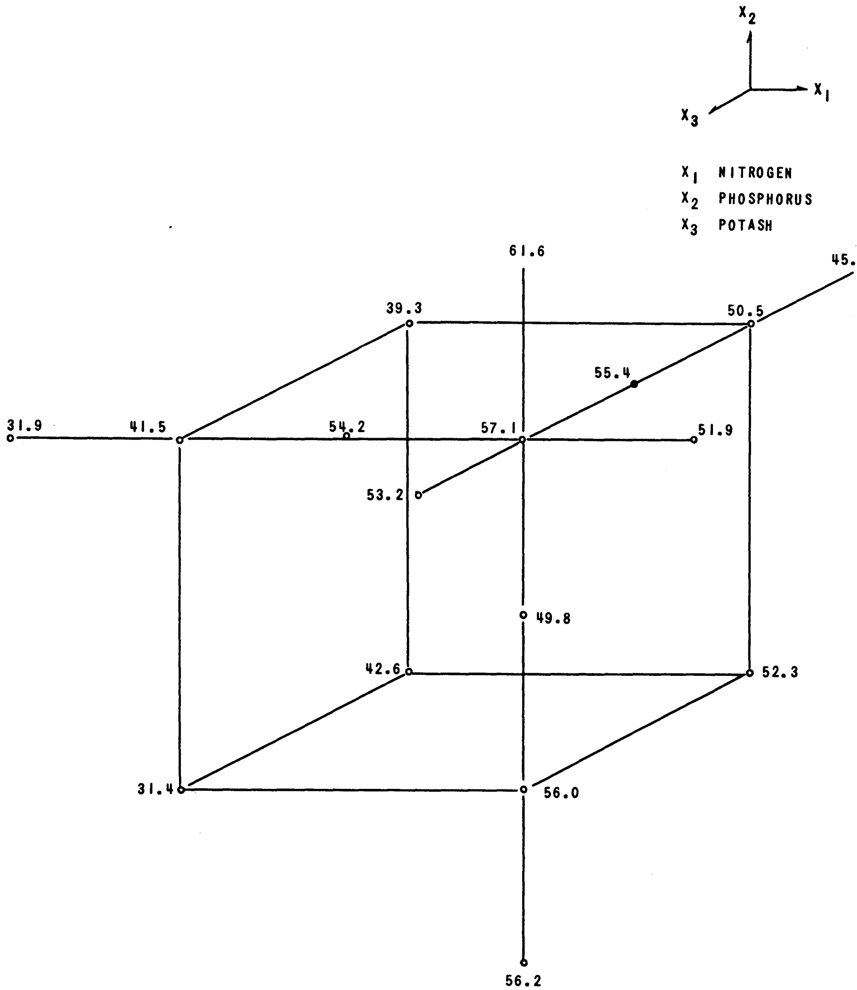


Fig. 5.5 — Design configuration and treatment means for multifactor potato yield experiment (yields are pounds, U.S. No. 1's, per 2 row, 25-foot plot).

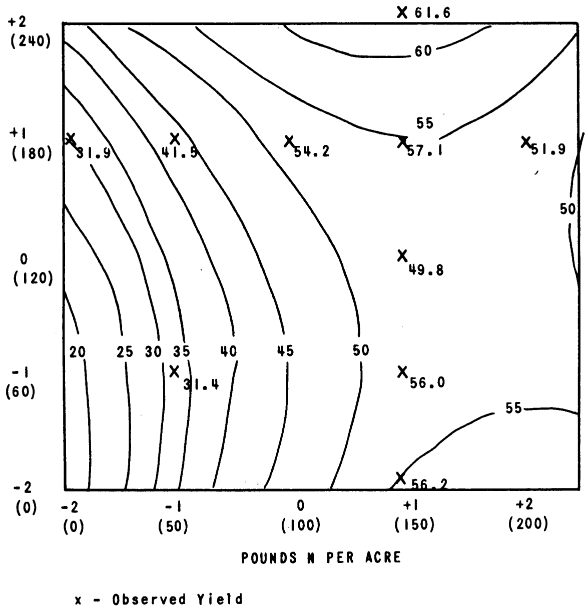


Fig. 5.6 — Yield contours for U. S. No. 1 potatoes, pounds per plot, for variations in X_1 (N) and X_2 (P_2O_5), with X_3 (K_2O) held at +1 (50 pounds per acre). X_1 (nitrogen)

(sometimes referred to as a "col" or "minimax"). Such a surface would appear to be difficult to interpret agronomically. One would certainly like some substantiation of this type of pattern before extending its application too far. A more complete sampling by observation points in the critical region is perhaps in order.

An illustration of the use of the logarithmic scale for the X variates is given in figure 5.7. These data are from one of a series of experiments to be reported on by Moore et al. (20). This example has little

TABLE 5.9. Analysis of Variance of Yield (Gms. Dry Wgt. of Lettuce Tops per 3 Plants)

Source of Variation	d.f.	M.S.
Linear terms	3	197.2
Quadratic	6	168.6
Lack of fit	5	6.2
Error	3	7.0

Source: Moore, et al. (20)

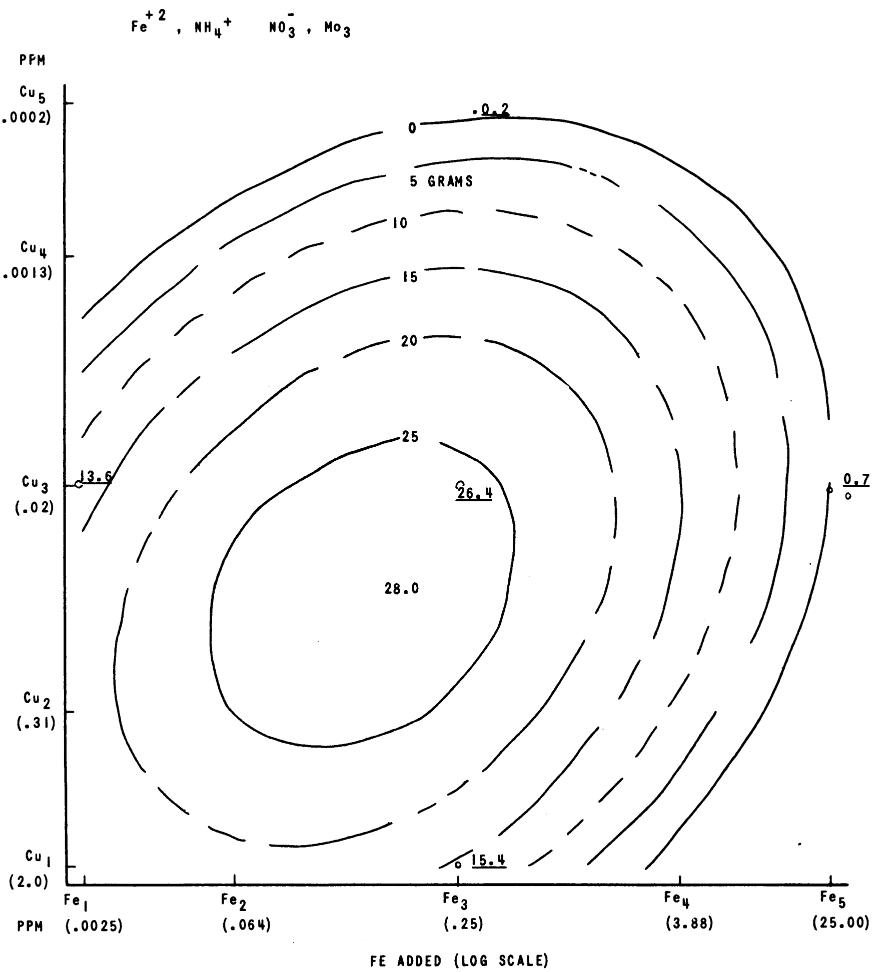


Fig. 5.7 — Yield contours of lettuce tops (gms. dry wt.) as affected by additions of Cu and Fe to nutrient solutions containing Fe^{+2}, NH_4^+ + NO_3^- and the middle level of Mo. Observational points and yields are underlined. The point at the center of the contours is the predicted maximum yield.

direct application regarding economics of fertilizer use. It is part of a greenhouse solution culture study set up with the objective of obtaining a general perspective of the relationship of certain minor elements to the yield response of lettuce, and to obtain information upon which to base more detailed studies. The experimental design used here was one of the "rotatable" designs developed and reported by Box and Hunter (5). The actual levels and the coded values for Cu and Fe are given in figure 5.6. The third factor, molybdenum, had the same levels as copper, ranging from .0002 to 2.0 ppm.

The summary analysis of variance in table 5.9 indicates a reasonably good fit of the second degree surface to the actual data. However, it is realized that the experimental error mean square is poorly estimated with only 3 d.f.

Discussion of Functional Model and Design of Multifactor Experiments

Anderson, in Chapter 3, has stated that if the response surface can be approximated by a simple mathematical function then it seems logical to estimate the parameters of this function instead of main effects and interactions. Spillman (26) recognized this in his suggested treatment arrangements necessary for estimating the constants in the Mitscherlich-Baule function. In the approach used by Box (4) and associates, that of using the general polynomial function of the degree necessary to adequately fit the surface, designs of treatment combinations may be developed that have desirable properties compared to the complete factorial arrangement. They point out two disadvantages of the complete factorial arrangement: (a) estimation of the pure quadratic (β_{11}, β_{22}, etc.), with less precision than the mixed quadratic terms (β_{12}, β_{13}, etc.),; (b) complete factorial arrangements for estimation of many higher order mixed terms which ordinarily are of little interest, and which do not have the corresponding "pure" effects to go along with them. Box states, "To attempt to interpret two factor interactions without the corresponding quadratic effects is precisely analogous to considering covariances without the corresponding variances."

The composite designs also have been discussed by Anderson in Chapter 3. These designs appear to have good possibilities. Additional field experiments are now under way in North Carolina using the second order, composite designs, and it is expected that evaluation of these experiments will provide some measure of their utility. Composite designs have been used in industrial research for estimating cubic or 3rd degree surfaces, as indicated by Pike (23).

In fertilizer response studies it seems desirable to have replication of the treatments, both for providing the necessary precision for the individual points in the design and for checking on the adequacy of the model in characterizing the surface. Although the work by Heady et al. (12), and by Hanson et al. (9), and the illustrations already given in this chapter indicate that a second degree function seems to give an adequate fit of the surface, this needs to be further studied under a wider range of conditions.

References Cited

1. ANDERSON, R. L., 1953. Recent advances in finding best operating conditions. Jour. Amer. Stat. Assn. 48:789-98.

2. ———, and BANCROFT, T. A., 1952. Statistical Theory in Research. McGraw-Hill, New York.

3. BALMUKAND, BH., 1928. Studies in crop variation. V. The relation between yield and soil nutrients. Jour. Agr. Sci. 18:602-27.

4. BOX, G. E. P., 1954. The exploration and exploitation of response surfaces: some general considerations and examples. Biometrics 10:16-60.

5. ———, and HUNTER, J. S., 1954. Multifactor experimental designs. Mimeo Series No. 92, Inst. of Stat., University of North Carolina.

6. CROWTHER, E. W., and YATES, F., 1941. Fertilizer policy in wartime: fertilizer requirements of arable crops. Empire Jour. Exp. Agr. IX:77-97.

7. DUMENIL, L., and NELSON, L. B., 1948. Nutrient balance and interaction in fertilizer experiments. Proc. Soil Sci. Soc. Amer. 13:335-42.

8. HADER, R. J., HARWARD, M. E., MASON, D. D., and MOORE, D. P., 1955. An investigation of some of the relationships between copper, iron and molybdenum in the growth and nutrition of lettuce. I. Exp. design and statistical methods for characterizing the response surface. Proc. Soil Sci. Soc. Amer. (in press).

9. HANSON, W. D., HUTTON, C. E., and ROBERTSON, W. K., 1954. Four years of soil fertility data from a 5 x 5 x 5 x 2 factorial experiment on corn and peanuts in rotation on Red Bay fine sandy loam. I. Statistical analysis. Paper presented at Soil Sci. Soc. Meetings, St. Paul, Minn. (in press).

10. HARTLEY, H. O., 1948. The estimation of nonlinear parameters by "internal least squares." Biometrika 35:32-45.

11. HEADY, E. O., 1954. Choice of functions in estimating input-output relationships. Proceedings of 51st Annual Meeting of the Agricultural Economics and Rural Sociology Section of the Association of Southern Agricultural Workers.

12. ———, PESEK, J. T., and BROWN, W. G., 1955. Crop response surfaces and economic optima in fertilizer use. Iowa Agr. Exp. Sta. Bul. 424.

13. HUNTER, A. S., and YOUNGEN, J. A., 1955. The influence of variations in fertility levels upon the yield and protein content of field corn in eastern Oregon. Proc. Soil Sci. Soc. Amer. 19:214-19.

14. HUTTON, R. F., 1955. An appraisal of research on the economics of fertilizer use. Report No. T 55-1, Tennessee Valley Authority, Division of Agricultural Relations, Agricultural Economics Branch.

15. ———, and ELDERKIN, J. D., 1954. Use of a particular function in investment income studies. Proceedings of 51st Annual Meeting of Agricultural Economics and Rural Sociology Section of the Association of Southern Agricultural Workers.

16. IBACH, D. B., and MENDUM, S. W., 1953. Determining profitable use of fertilizer. F. M. 105, Bureau of Agricultural Economics, USDA.

References Cited

17. JOHNSON, P. R., 1953. Alternative functions for analyzing a fertilizer-yield relationship. Jour. Farm Econ. 35:519-29.

18. MCPHERSON, W. W., 1954. Mathematical functions used in estimating production functions. A paper presented to the Southern Farm Management Research Committee, Memphis, Tennessee.

19. MONROE, R. J., 1949. On the use of nonlinear systems in the estimation of nutritional requirements of animals. Ph.D. thesis, North Carolina State College, Raleigh.

20. MOORE, D. P., et al., 1955. An investigation of some of the relationships of copper, iron, and molybdenum in the growth and nutrition of lettuce. II. Response surfaces of growth and accumulation. Proc. Soil Sci. Soc. Amer. (in press).

21. NAIR, K. R., 1954. The fitting of growth curves, Statistics and Mathematics in Biology, edited by Kempthorne, et al., Iowa State College Press, pp. 119-32.

22. PASCHAL, J. L., 1953. Economic analysis of alfalfa yield response to phosphate fertilizer at three locations in the West. F. M. 104, Bureau of Agricultural Economics, USDA.

23. PIKE, F. P., 1954. The flooding capacity of a pulse column on the benzene-water system. Progress Report No. 4 under Contract No. AT-(40-1)-1320. Dept. Eng. Res., North Carolina State College, Raleigh.

24. PIMENTEL-GOMEZ, FREDERICO, 1953. The use of Mitscherlich's law in the analysis of experiments with fertilizer. Biometrics 9:498-517.

25. RUSSELL, SIR E. J., 1937. Soil Conditions and Plant Growth. Longmans, Green and Co., New York.

26. SPILLMAN, W. J., 1933. Use of the exponential yield curve in fertilizer experiments. Tech. Bul. 348. USDA.

27. STEVENS, W. L., 1951. Asymptotic regression. Biometrics 7:247-68.

28. VEITS, F., NELSON, C. E., and CRAWFORD, C., 1954. The relationships among corn yields, leaf compositions and fertilizers applied. Proc. Soil Sci. Soc. Amer. 18:297-301.

29. YATES, F., 1937. The design and analysis of factorial experiments. Imperial Bureau of Soil Sci. Tech. Comm. No. 35, Harpenden, England.

PART III

Agronomic and Related Considerations in Experiments and Fitting Functions to Existing Data

- ► **Size and Type of Experiment**
- ► **Soil and Moisture Conditions**
- ► **Soil Test Data**
- ► **Standard Curves**