

BIOMATHEMATICS TRAINING PROGRAM

MODIFICATIONS OF RESPONSE SURFACE  
TECHNIQUES FOR BIOLOGICAL USE

by

DAVID CHARLES HURST

and

ROBERT J. HADER

Institute of Statistics  
Mimeograph Series No. 330  
June 1962

Partially Supported by  
NIH Grant No. RG7143

TABLE OF CONTENTS

	Page
LIST OF TABLES AND FIGURE . . . . .	vi
PREFACE	
INTRODUCTION. . . . .	1
REVIEW OF LITERATURE AND RELATED RESEARCH . . . . .	3
SINGLE FACTOR: ADDITIVE ERROR CASE . . . . .	16
SINGLE FACTOR CASE: COMBINING INFORMATION FROM SEVERAL LOCATIONS. . . . .	23
SINGLE FACTOR: VARIABLE NUTRIENT CASE. . . . .	36
MULTIPLE FACTOR: VARIABLE NUTRIENT CASE. . . . .	44
CONCLUSIONS AND SUMMARY . . . . .	68
LIST OF REFERENCES. . . . .	69

## LIST OF TABLES AND FIGURE

Tables	Page
1. The effects of adjusting for residual nutrients on the error terms and coefficients of variation . . . . .	12
2. Parameter estimates and their variances as estimated on 6 Norfolk-like soils . . . . .	14
3. Mean yields in bu./A. and coded applied nitrogen rates for 4 locations. . . . .	28
4. Quadratic coefficients for the data in Table 3 with alternate representations of parameter estimates . . . . .	29
5. Treatment combinations and yields for 6 selected locations in bushels/Acre from Baird (1958). . . . .	60
6. Individual and average regressions for the data in Table 5. . . . .	61
7. First approximation to $\delta_{i\lambda}$ for 6 locations of Table 5. . . . .	62
8. The estimated $\delta_{i\lambda}$ for the 6 locations of Table 5 . . . . .	63
9. Average soil tests for the 6 locations of Table 5 . . . . .	65
 Figure	
1. A graphical presentation of the results of 9 iterations using method 2 with the successive estimates plotted against the iteration number. . . . .	34

## PREFACE

In many biological fields experimenters have been faced with problems of estimating and predicting the quantitative effects on "yield" or "response" variables of specified amounts of "input" or "causative" variables.

A statistical technique commonly used in this endeavor has been the Analysis of Variance. In the framework of this method of analysis one of the most vexing problems has been the existence of so-called "treatment by location" interactions, "dosage by animal" interactions, and similar failures of the model to fit the biological facts.

It behooves statisticians, therefore, to reexamine these problems and the proposed techniques more closely to see if better explanations (i.e., deeper understanding) and hopefully, better techniques are possible. To this end, then, we direct this dissertation.

## INTRODUCTION

Low order polynomial models for describing the response to one or more quantitative factors have proved very useful to experimenters in diverse fields for many years. Within this general mode of experimentation, a group of useful designs, techniques and statistical analysis procedures which have enabled experimenters to arrive at a description of the quantitative phenomenon under investigation efficiently and economically, have come to be called Response Surface Techniques.

Response Surface Techniques have shown a great deal of promise in the chemical industry, (for example), in finding regions of interest such as local maxima, (or minima), cols or saddle points, rising ridge systems, and the like, as well as describing these regions in terms of contours of equal response. Since many of these same results are desired in many agronomic-economic experiments where the quantitative variables are chemical fertilizer constituents, much interest has been shown in suitable modifications to these techniques to allow their application to be profitable to this case also. It is recognized, however, that many of the desirable conditions often existing in the chemical industry, such as low error variance, possibility of sequential observations, and unique underlying phenomena do not exist in biological experimentation.

Fertilizer experiments are characterized by highly variable experimental material, necessity of delaying subsequent experimentation to a new growing season and location by treatment interactions.

The sequential problem is probably the least worrisome in that a generally satisfactory positioning of the experimental treatments can be arranged without too much trouble and sufficient treatments can be included to cover the presumed region of interest. There remain the large variations commonly encountered and the very troublesome treatment by location interaction problems. These must be taken into account in a more precise modification of existing procedures. Variations in weather factors are also an important aspect of agricultural field experiments but will not be included within the scope of the present investigation.

In summary, the problem to be considered is the development and justification of modifications that seem appropriate to linear response surface analysis to obtain a more generally applicable form and to more realistically deal with the variation present in the agronomic-economic fertilizer experiment context.

## REVIEW OF LITERATURE AND RELATED RESEARCH

Almost from the beginning of modern scientific designed experimentation, the problems associated with quantitative fertilizer experiments have stimulated statisticians to develop appropriate experimental designs and analyses. The early publications of R. A. Fisher and Frank Yates undoubtedly helped set the pattern for at least thirty years of agronomic research.

The concepts which proved to be far-reaching in their consequences and incisive in exposing the underlying phenomena were those of randomization, objective hypothesis testing, small plot technique coupled with statistical designs which gave unconfounded comparisons, and the introduction of factorial treatment combinations which allowed "hidden replication" and a much broadened inferential base on which to draw conclusions.

With the advent of these techniques, much information was gained which enabled agronomists to recommend more realistically and scientifically to the agricultural population. It was early recognized, however, that to continue in the same vein of research, experiments identical in scope and kind should be tried at different locations so that comparability of results could be assessed and generalizations could be made where appropriate. This extension did not result in allowing very much in the way of sweeping generalizations because identical experiments at different

locations gave decidedly different results. These differences were not of an easily explainable kind and are known to workers in this area as significant treatment by location interactions.

The presence of treatment by location interactions stimulated research into their causes and control. Weather factors and variable initial soil nutrients seemed to be the most obvious general causes; this identification led to elaborate meteorological and soil testing efforts. The problems associated with proper utilization of this wealth of auxiliary information are still largely unsolved.

With the development of a body of knowledge in this area, agronomists were able to increase the objectivity and definiteness of their recommendations. It became apparent, however, that another aspect of the problem had to be specifically dealt with to make the recommendations more realistic. This is the economic aspect which gave rise to cooperative agronomic-economic experiments. From the economic theory viewpoint, the key element was the development of a "production function" by the agronomists which then became one of the cornerstones for the full economic analysis of the situation.

At this stage, three problem areas became relatively critical. The economists needed a production function that both included the diminishing returns concept and was biologically acceptable; something had to be done with the initial nutrients problem; and finally it seemed highly

desirable to have a continuous production function so that marginal costs could be equated to marginal returns, (i.e., the function should possess a derivative).

Several writers investigated the first problem, the model, among them Heady, Pesek, and Brown (1955); Heady, Doll, and Pesek (1956); Anderson (1956, 1957); Stemberger (1956) and Mason (1956). The general results of these investigations were that estimates of economic optima based on different models, e.g., the Mitscherlich, Cobb-Douglas, quadratic, square root quadratic and Hildreth's (1954) model were widely different when applied to the same data. Also no one model fit reasonably well over several different experiments. Some insight on this problem resulted from Anderson's (1957) demonstration that the various models agreed fairly well if the check plot data were omitted. Thus a quadratic model became a desirable form because of its simplicity and ease of application in economic analyses. The check plot yields, of course, contain much information on initial nutrient levels and through this phenomenon can exert considerable influence on the fit of all models. Thus we are again led to the problem area of how to incorporate a realistic formulation of the differing initial nutrient levels into the model.

Mitscherlich (1909) seems to have been the first writer to include explicit use of the concept of initial nutrient levels. The form proposed for the response to a

single nutrient was:

$$y = A[1 - e^{-c(x+b)}]$$

where

A = maximum attainable yield

b = initial nutrient level

c = constant .

Serious objection to this formulation was raised on three counts: The model did not allow the yield to diminish at high levels of application, c was considered by Mitscherlich to be a general constant and, finally, the equation did not fit data at all well.

Many other writers felt the need for a more realistic formulation and there followed the group of alternate forms as mentioned previously. Heady, et al. (1961) give a general review of these efforts. Among the more recent are Sukhatme (1941), who fit quadratic models to rice experiments noting variation in the parameters presumably caused by varying initial nutrient levels and Panse, et al. (1951), who divided some cotton experiments in India into low, medium and high initial fertility, noting a linear trend in the coefficients which he then estimated. Anderson (1956) discussed the problem at some length for the quadratic case and a discussion of that paper follows. (It should be noted that the notation used in the Anderson paper differs somewhat from that used subsequently in the present investigation.)

The first point made in the Anderson paper was that a method was needed to adjust for nutrients available in the soil. The justification of this point is illustrated with the quadratic model

$$E(X) = \beta_0 + \beta_1 X + \beta_{11} X^2$$

where  $X$  is the added amount of fertilizer. If the actual amount of nutrient is designated by  $N$  where

$$N = X + d$$

$$X = N - d \quad \text{and } d \text{ is the amount available,}$$

then  $E(N) = (\beta_0 - \beta_1 d + \beta_{11} d^2) + (\beta_1 - 2\beta_{11} d)N + \beta_{11} N^2$ .

If the  $E(X)$  model is used to estimate the yield at a farm with initial nutrient level  $d_0$  and added amount  $X$  we shall predict the yield to be:

$$F(X) = (\beta_0 - \beta_1 d + \beta_{11} d^2) + (\beta_1 - 2\beta_{11} d)(X + d_0) + \beta_{11} (X + d_0)^2$$

instead of  $E(X)$ . The bias in the prediction will then be:

$$E(X) - F(X) = (d - d_0)(\beta_1 + 2\beta_{11} X) - (d - d_0)^2 \beta_{11}$$

(which will only be zero when the initial nutrient levels available at the farm and experimental plots are equal).

Anderson then goes on to point out that even differences in yield resulting from applying different rates are not estimated unbiasedly. Further discussion reveals that combining results from experiments with different initial nutrient levels only aggravates the problem. The loss in efficiency by not combining when certain structural

parameters are common is then discussed and in summary several general points are made, the first of which are:

"(i) Statisticians have not developed easy and efficient estimation procedures for the more complicated models.

(ii) Procedures for determining available nutrients are not too well developed.

(iii) It is often difficult to calibrate available and applied nutrients."

Following this work of Anderson (1956), Hildreth (1957), and Hurst and Mason (1957) independently proposed alternative formulations to take the varying initial levels into account. Yniguez (1957) subsequently working with Hurst and Mason subjected data generated by a cooperative experiment between North Carolina State College and T.V.A. to an analysis different from, but in the spirit of, the Hildreth formulation by using a modification of multiple covariance analysis. Therefore, we will discuss these efforts in their logical order: Hildreth (1957), Yniguez (1957), and then Hurst and Mason (1957). The following then is a discussion of Hildreth's formulation:

Hildreth also considers the single factor quadratic with knowledge of the initial levels presumed known. Thus the model is:

$$y = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + u \quad \text{or}$$

$$y = \alpha_0 + \alpha_1 n + \alpha_1 \lambda w + \alpha_2 n^2 + \alpha_2 \lambda^2 w^2 \\ + 2\alpha_2 \lambda n w + u$$

written as:

$$y = \beta_0 + \beta_1 n + \beta_2 w + \beta_3 n^2 + \beta_4 w^2 + \beta_5 nw + u$$

where  $y =$  yield

$X = n + \lambda w =$  total nitrogen made up of an applied amount plus a proportionate multiple of an initial amount present

$w =$  initial nitrogen

$n =$  nitrogen added

$\lambda =$  an unknown factor of proportionality

$u =$  a random disturbance.

He points out that estimating the parameters in the last equation would be an ordinary multiple regression problem except for the two restrictions on the  $\beta$ 's, namely:

$$\frac{\beta_2}{\beta_1} = \frac{\beta_4}{\beta_3} = \frac{\beta_5}{2\beta_3} = \lambda .$$

With  $\lambda$  known, however, the first equation could be used in an ordinary multiple regression on  $X = n + \lambda w$ .

Therefore he proposes to estimate  $\lambda$  by maximum likelihood and then use this value to find  $X$  which will then be used in the first equation. The expression he desires to maximize to estimate  $\lambda$  is:

$$L^* = \frac{\sum_{i=0}^6 C_i \lambda_i}{\sum_{i=0}^6 D_i \lambda_i} ,$$

where the  $C_i$  and  $D_i$  are functions of the moments of  $n$  and

w . Hildreth then proposes several ways of computing initial approximations to  $\lambda$  which would then be improved by an iterative scheme until  $L^*$  is maximized.

Brown, et al. (1962) approached the problem from the Hildreth (1957) viewpoint on some alfalfa-lime fertilizer experiments in Oregon. They took two soil test measurements, i.e., before and after fertilization and thus were able to get a more direct calibration of soil test measurements and with the paired observations, an estimate of  $\lambda$  . Unfortunately some soil nutrient tests, e.g., nitrogen, have not so far yielded to satisfactory calibration and in these cases we need some alternative procedures. This is shown by the calculations of Yniguez (1957).

Yniguez (1957) approached the problem from the same point of view as Hildreth using data from the T.V.A.-North Carolina State College cooperative experiment. The variables that were measured initially for each plot were pH, calcium, magnesium, potassium, phosphorus, and organic material. Now if these variables sufficiently characterize the initial nutrient complex existing in the soil, it should be possible to "explain" the variability in yield measurement for constant treatment combination by means of a multiple regression of yield on these variables for each treatment combination separately. Success in achieving this would then be measured by whether or not the residual mean square error was reduced to that achieved by the

pooled separate analysis error encountered. The experiment was of randomized block nature with 4 reps and 18 treatments per block at each location and had error variance of approximately 100. In Table 1 we have the results of one such form, namely a six variable linear regression, linear in all the initial nutrient variables. This model produced the smallest residual errors of the several alternative forms tried. (Portsmouth locations were deleted from the 1955 data because the organic matter variable was incorrectly measured. The situation was corrected in 1956.) The result of the Yniguez investigation was that the characterizing variables chosen did not sufficiently "explain" the variability of the yield figures. It was then concluded that these measurements do not adequately represent the effective available nutrient levels.

It is seen, then, that there is still need for a biological assay of the initial soil nutrient, a need in part met by the Hurst-Mason (1957) paper. Following is a summary of the author's section of that paper.

The single factor quadratic is again used for illustration where the expected yield  $E(Y)$  is related to the applied nutrient  $X_1$  in presence of available nutrient  $\delta_1$  as follows:

$$\begin{aligned} E(Y) &= \beta_0 + \beta_1(X_1 + \delta_1) + \beta_2(X_1 + \delta_1)^2 \\ &= \beta_0 + \beta_1\delta_1 + \beta_2\delta_1^2 + (\beta_1 + 2\beta_2\delta_1)X_1 + \beta_2X_1^2 \\ &= \beta_0^* + \beta_1^*X_1 + \beta_2^*X_1^2, \end{aligned}$$

Table 1. The effects of adjusting for residual nutrients on the error terms and coefficients of variation

Treatment Number	Levels <sup>a</sup> N P K	Data for 1955 <sup>b</sup>			Data for 1956 <sup>c</sup>		
		Mean Bu./A	M.S.E.	C.V.%	Mean Bu./A	M.S.E.	C.V.%
1	111	62.27	61.81	12.6	76.92	187.46	17.8
2	113	63.81	135.21	18.2	77.31	233.85	19.8
3	131	63.06	141.88	18.9	78.35	166.85	16.5
4	133	61.66	132.52	18.7	79.41	201.08	17.9
5	311	69.01	121.39	15.9	93.28	298.52	18.5
6	313	68.87	287.82	24.5	93.71	272.26	17.6
7	331	72.56	279.55	23.1	95.25	273.84	17.4
8	333	68.82	289.12	24.7	97.25	268.52	16.8
9	222	68.99	81.51	13.1	90.84	250.86	17.4
10	022	30.34	163.76	42.7	47.20	251.23	33.6
11	422	66.76	257.81	24.1	95.46	295.46	18.0
12	202	69.18	148.42	17.6	86.76	368.62	22.1
13	242	71.08	267.42	23.0	92.88	228.88	16.3
14	220	72.10	250.68	22.0	88.96	326.68	20.3
15	224	67.70	154.08	18.4	92.59	232.70	16.5
16	044	30.12	142.19	39.6	48.88	229.34	51.4
17	404	65.34	214.87	22.5	91.49	445.41	23.1
18	440	72.05	283.28	23.4	91.31	373.26	21.2

<sup>a</sup> Unit of N = 62.5, P = 37.5, K = 37.5, e.g., 111 is 62.5 lbs. N/A, 37.5 lbs. P/A, 37.5 lbs. K/A .

<sup>b</sup> Calculated by pooling 10 locations on Norfolk-like soil.

<sup>c</sup> Calculated by pooling 8 locations on Portsmouth-like soil and 13 locations of Norfolk-like soil.

where:

$$\beta_0^* = \beta_0 + \beta_1 \delta_1 + \beta_2 \delta_1^2$$

$$\beta_1^* = \beta_1 + 2\beta_2 \delta_1$$

$$\beta_2^* = \beta_2$$

It is noted that the customary regression coefficients are estimates of the starred parameters and are biased as Anderson (1956) points out. Then a check on the formulation is made using some of the T.V.A.-North Carolina data, because it can be seen that if the postulated  $\delta_1$  act as indicated, the estimated values of  $\beta_0^*$  and  $\beta_1^*$  should vary from location to location more than predicted by theory while the estimated values of  $\beta_2^*$  should not. That these observations are substantiated by the data can be seen from Table 2, a modification of one of the tables in the paper.

The paper goes on to justify deleting  $\beta_0$  from the formulation on the ground that the response should be zero for zero total nitrogen. For the modified model, then, the expressions are:

$$\beta_0^* = \beta_1 \delta_1 + \beta_2 \delta_1^2$$

$$\beta_1^* = \beta_1 + 2\beta_2 \delta_1$$

$$\beta_2^* = \beta_2$$

and it is proposed to estimate the starred parameters by least squares. The parameter estimates are then derived and are:

Table 2. Parameter estimates and their variances as estimated on 6 Norfolk-like soils<sup>a</sup>

Parameter	Location Number						Observed $s^2(b_{ij})$	Theoretical $V(b_{ij})$	Test F(5, 252)
	51	52	53	55	57	58			
$\beta_{00}$	81.32	86.45	93.22	92.26	59.69	51.54	310.24	19.67	15.8**
$\beta_{10}$	9.56	4.30	9.91	14.96	3.44	.18	28.99	1.58	18.4**
$\beta_{20}$	.27	.11	- 1.14	1.98	1.08	-.43	1.21	1.58	.8N.S.
$\beta_{30}$	-.16	.78	1.27	-.71	- 1.43	- 2.10	1.66	1.58	1.1N.S.
$\beta_{11}$	- 8.00	- 4.82	- 7.57	- 9.12	- 6.01	- 3.66	4.28	2.28	1.9N.S.
$\beta_{22}$	- 1.85	- .58	- .03	- 1.41	1.09	- 1.76	1.32	2.28	.6N.S.
$\beta_{33}$	- 1.66	- .92	- .30	- .80	.65	-.42	.59	2.28	.3N.S.
$\beta_{12}$	1.18	- 1.07	.64	.95	2.48	2.26	1.64	3.16	.5N.S.
$\beta_{13}$	1.17	- .85	.78	1.23	- 3.77	.85	3.83	3.16	1.2N.S.
$\beta_{23}$	-.95	- .92	-.51	- 3.04	- 1.14	- 1.10	.80	3.16	.3N.S.
M.S. Error	100.12	104.17	89.11	91.58	117.63	104.44	Ave. MSE	101.18	(252 d.f.)

<sup>a</sup> The  $b_{ij}$  are calculated from the 15 points of the central composite design as given by Box and Wilson (1951) using coded applied nutrient levels.

<sup>b</sup>  $\beta_{00}$  is the intercept  $\beta_{10}$ ,  $\beta_{20}$ ,  $\beta_{30}$  are the coefficients for linear N, P, and K respectively,  $\beta_{11}$ ,  $\beta_{22}$ ,  $\beta_{33}$  are coefficients for quadratic N, P, and K and  $\beta_{12}$ ,  $\beta_{13}$ ,  $\beta_{23}$  are the interaction quadratic coefficients.

$$b_1 = \pm \sqrt{b_1^{*2} - 4b_0^*b_2^*}$$

$$d = \frac{b_1^* \mp \sqrt{b_1^{*2} - 4b_0^*b_2^*}}{2b_2^*}$$

$$b_2 = b_2^*$$

where the Latin letters are estimates of the corresponding Greek letter parameters.

It should be noted that normally one would take the positive sign in the first equation and the negative sign in the second equation for biological reasonability.

This concludes the Review of Literature. In the subsequent chapter the single nutrient case as just formulated will be examined in greater detail, and as other references are needed, they will be discussed where appropriate.

## SINGLE FACTOR: ADDITIVE ERROR CASE

An important special case arises when it is either known a priori and/or confirmed experimentally that only one nutrient factor has any measurable effect on yield. In this case we must conclude that coefficients in a functional form or response curve for the other non-effective factors are zero or at least small enough in magnitude to be obscured by the variation present in the experimental material. The Hurst and Mason (1957) paper illustrates a case in point where the yield response was to nitrogen only. The formulation presented there has some statistical aspects not pointed out and worthy of further elucidation. Recall the total nutrient model:

$$\begin{aligned} Y &= \beta_1 N + \beta_2 N^2 + \varepsilon \\ &= \beta_0^* + \beta_1^* X + \beta_2^* X^2 + \varepsilon \end{aligned}$$

where:

$$\beta_0^* = \beta_1 \delta + \beta_2 \delta^2$$

$$\beta_1^* = \beta_1 + 2\beta_2 \delta$$

and

$$\beta_2^* = \beta_2 \quad .$$

If  $\varepsilon$  is a random normally distributed error with expectation zero and variance  $\sigma^2$  we have the following properties:

The  $b_1^*$  are, of course, ordinary least squares estimates of  $\beta_1^*$  and are obtained by solving the normal equations:

$$\begin{bmatrix} N & \Sigma X & \Sigma X^2 \\ \Sigma X & \Sigma X^2 & \Sigma X^3 \\ \Sigma X^2 & \Sigma X^3 & \Sigma X^4 \end{bmatrix} \begin{pmatrix} b_0^* \\ b_1^* \\ b_2^* \end{pmatrix} = \begin{pmatrix} \Sigma Y \\ \Sigma XY \\ \Sigma X^2 Y \end{pmatrix}$$

with  $SS \text{ Reg} = b_0^* \Sigma Y + b_1^* \Sigma XY + b_2^* \Sigma X^2 Y$  and the variance of the  $\epsilon_i$  in the model estimated by:

$$\hat{\sigma}^2 = SSE/(n - 3) = s^2$$

where  $SSE = \Sigma Y^2 - SS \text{ Reg}$ .

From well known least squares theory, then, (e.g., see Graybill (1961)), the equation

$$F(3, n-3) = \frac{\begin{pmatrix} b_0^* - \beta_{00}^* \\ b_1^* - \beta_{10}^* \\ b_2^* - \beta_{20}^* \end{pmatrix}' \begin{bmatrix} n & \Sigma X & \Sigma X^2 \\ \Sigma X & \Sigma X^2 & \Sigma X^3 \\ \Sigma X^2 & \Sigma X^3 & \Sigma X^4 \end{bmatrix} \begin{pmatrix} b_0^* - \beta_{00}^* \\ b_1^* - \beta_{10}^* \\ b_2^* - \beta_{20}^* \end{pmatrix}}{s^2}$$

is a simultaneous test of the null hypothesis

$$H_0: \beta_1 = \beta_{10}, \beta_2 = \beta_{20}, \delta = \delta_0$$

through substitution in the previously presented equations for  $\beta_1^*$ .

The preceding joint test of significance can, of course, be regarded as also defining a simultaneous confidence region for  $\beta_1$ ,  $\beta_2$ , and  $\delta$ , by replacing the  $F$  with an  $F_\alpha$  from tabulated critical points of the  $F$  distribution. This region is then formally of probability content  $1 - \alpha$ , but not very usable.

If individual standard deviations for the parameters are desired, we may compute approximate ones for  $b_1$  and  $d$  with that for  $b_2$ , of course, already exact and known, thus:

Let:

$$\begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} n & \Sigma X & \Sigma X^2 \\ \Sigma X & \Sigma X^2 & \Sigma X^3 \\ \Sigma X^2 & \Sigma X^3 & \Sigma X^4 \end{bmatrix}^{-1}$$

then the variance-covariance matrix of  $b_0^*$ ,  $b_1^*$ ,  $b_2^*$  is:

$$\sigma^2 \begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix} \quad \text{with}$$

$\sigma^2$  estimated as before. A "t" test of  $H_0: \beta_2 = \beta_{20}$  is:

$$t = \frac{b_2^* - \beta_{20}}{\sqrt{C_{22} s^2}} \quad \text{with } n - 3 \text{ degrees of freedom}$$

and C.L. for  $\beta_2$  are:  $b_2^* \pm t_{\alpha/2} \sqrt{C_{22} s^2}$ .

For the standard deviation of  $b_1$  we have:

$$s_{b_1} = \sqrt{\widehat{V}(b_1)}$$

$$\begin{aligned} \text{where: } \widehat{V}(b_1) &\approx \left(\frac{\partial b_1}{\partial b_0^*}\right)^2 \widehat{V}(b_0^*) + \left(\frac{\partial b_1}{\partial b_1^*}\right)^2 \widehat{V}(b_1^*) + \left(\frac{\partial b_1}{\partial b_2^*}\right)^2 \widehat{V}(b_2^*) \\ &+ 2\left(\frac{\partial b_1}{\partial b_0^*}\right)\left(\frac{\partial b_1}{\partial b_1^*}\right) \widehat{\text{cov}}(b_0^*, b_1^*) + 2\left(\frac{\partial b_1}{\partial b_0^*}\right)\left(\frac{\partial b_1}{\partial b_2^*}\right) \widehat{\text{cov}}(b_0^*, b_2^*) \\ &+ 2\left(\frac{\partial b_1}{\partial b_1^*}\right)\left(\frac{\partial b_1}{\partial b_2^*}\right) \widehat{\text{cov}}(b_1^*, b_2^*) \end{aligned}$$

Since:

$$b_1 = \sqrt{b_1^{*2} - 4b_0^*b_2^*}$$

$$\frac{\partial b_1}{\partial b_0^*} = \frac{-2b_2^*}{\sqrt{b_1^{*2} - 4b_0^*b_2^*}}$$

$$\frac{\partial b_1}{\partial b_1^*} = \frac{b_1^*}{\sqrt{b_1^{*2} - 4b_0^*b_2^*}}$$

$$\frac{\partial b_1}{\partial b_2^*} = \frac{-2b_0^*}{\sqrt{b_1^{*2} - 4b_0^*b_2^*}}$$

Thus:

$$\widehat{v}(b_1) \approx \frac{s^2}{(b_1^{*2} - 4b_0^*b_2^*)} [4b_2^{*2}C_{00} + b_1^{*2}C_{11} + 4b_0^{*2}C_{22}$$

$$- 4b_1^*b_2^*C_{01} + 8b_0^*b_2^*C_{02} - 4b_0^*b_1^*C_{12}]$$

and we may take the square root of this for  $s_{b_1}$ .

Similarly for  $d$  we have:

$$d = \frac{b_1^*}{2b_2^*} - \frac{\sqrt{b_1^{*2} - 2b_0^*b_2^*}}{2b_2^*}$$

so that

$$\frac{\partial d}{\partial b_0^*} = \frac{1}{\sqrt{b_1^{*2} - 4b_0^*b_2^*}}$$

$$\frac{\partial d}{\partial b_1^*} = \frac{1}{2b_2^*} \left[ 1 - \frac{b_1^*}{\sqrt{b_1^{*2} - 4b_0^*b_2^*}} \right]$$

$$\frac{\partial d}{\partial b_2^*} = \frac{-b_1^*}{2b_2^{*2}} + \frac{\sqrt{b_1^{*2} - 4b_0^*b_2^*}}{2b_2^{*2}} + \frac{b_0^*}{b_2^*\sqrt{b_1^{*2} - 4b_0^*b_2^*}}$$

It seems simpler computationally to calculate these quantities as they stand and then substitute into:

$$\begin{aligned} \widehat{V}(d) \approx s^2 & \left[ \left( \frac{\partial d}{\partial b_0^*} \right)^2 c_{00} + \left( \frac{\partial d}{\partial b_1^*} \right)^2 c_{11} + \left( \frac{\partial d}{\partial b_2^*} \right)^2 c_{22} \right. \\ & + 2 \left( \frac{\partial d}{\partial b_0^*} \right) \left( \frac{\partial d}{\partial b_1^*} \right) c_{01} + 2 \left( \frac{\partial d}{\partial b_0^*} \right) \left( \frac{\partial d}{\partial b_2^*} \right) c_{02} \\ & \left. + 2 \left( \frac{\partial d}{\partial b_1^*} \right) \left( \frac{\partial d}{\partial b_2^*} \right) c_{12} \right] \end{aligned}$$

then

$$s_d = \sqrt{\widehat{V}(d)}$$

The "confidence limits" on the regression line and predicted values are still the usual formulae using  $b_0^*$ ,  $b_1^*$  and  $b_2^*$  and their variances and covariances.

Alternatively, since our estimates of  $\beta_1$ ,  $\beta_2$ , and  $\delta$  are maximum likelihood, we may compute the asymptotic variance-covariance matrix. For this case the log likelihood, (L), is:

$$L = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{\sum (Y_j - \beta_1 N_j - \beta_2 N_j^2)^2}{2\sigma^2}$$

We then form the Hessian of this function with respect to  $\beta_1$ ,  $\beta_2$ ,  $\delta$ , and  $\sigma^2$  in that order and take the negative expectation of this matrix to obtain the information matrix, thus:

$$I = - E \begin{bmatrix} \frac{\partial^2 L}{\partial \beta_1^2} & \frac{\partial^2 L}{\partial \beta_1 \partial \beta_2} & \frac{\partial^2 L}{\partial \beta_1 \partial \delta} & \frac{\partial^2 L}{\partial \beta_1 \partial (\sigma^2)} \\ \frac{\partial^2 L}{\partial \beta_2 \partial \beta_1} & \frac{\partial^2 L}{\partial \beta_2^2} & \frac{\partial^2 L}{\partial \beta_2 \partial \delta} & \frac{\partial^2 L}{\partial \beta_2 \partial (\sigma^2)} \\ \frac{\partial^2 L}{\partial \delta \partial \beta_1} & \frac{\partial^2 L}{\partial \delta \partial \beta_2} & \frac{\partial^2 L}{\partial \delta^2} & \frac{\partial^2 L}{\partial \delta \partial (\sigma^2)} \\ \frac{\partial^2 L}{\partial (\sigma^2) \partial \beta_1} & \frac{\partial^2 L}{\partial (\sigma^2) \partial \beta_2} & \frac{\partial^2 L}{\partial (\sigma^2) \partial \delta} & \frac{\partial^2 L}{\partial (\sigma^2)^2} \end{bmatrix}$$

Working this out we have:

$$I = \begin{bmatrix} \Sigma N^2 / \sigma^2 & \Sigma N^3 / \sigma^2 & (\beta_1 \Sigma N + 2\beta_2 \Sigma N^2) / \sigma^2 & 0 \\ \Sigma N^3 / \sigma^2 & \Sigma N^4 / \sigma^2 & (\beta_1 \Sigma N^2 + 2\beta_2 \Sigma N^3) / \sigma^2 & 0 \\ (\beta_1 \Sigma N + 2\beta_2 \Sigma N^2) / \sigma^2 & (\beta_1 \Sigma N^2 + 2\beta_2 \Sigma N^3) / \sigma^2 & \Sigma (\beta_1 + 2\beta_2 N)^2 / \sigma^2 & 0 \\ 0 & 0 & 0 & (n/2) / \sigma^4 \end{bmatrix}$$

where  $N = X + \delta$ .

Then the asymptotic variance-covariance matrix of  $\hat{\beta}_1$ ,  $\hat{\beta}_2$ ,  $\hat{\delta}$  respectively is:

$$\frac{\sigma^2}{D} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$$

where:

$$D = \beta_1^2 \left[ (n \Sigma N^2 - (\Sigma N^2)^2 / (\Sigma N^4) + 2(\Sigma N)(\Sigma N^2)(\Sigma N^3) - n(\Sigma N^3)^2 - (\Sigma N^2)^3 \right]$$

$$f_{11} = \beta_1^2 [n\Sigma N^4 - (\Sigma N^2)^2] + 4\beta_1\beta_2 [(\Sigma N)(\Sigma N^4) - (\Sigma N^2)(\Sigma N^3)] \\ + 4\beta_2^2 [(\Sigma N^2)(\Sigma N^4) - (\Sigma N^3)^2]$$

$$f_{12} = f_{21} = \beta_1^2 [(\Sigma N)(\Sigma N^2) - n\Sigma N^3] + 2\beta_1\beta_2 [(\Sigma N^2)^2 \\ - (\Sigma N)(\Sigma N^3)]$$

$$f_{13} = f_{31} = \beta_1 [(\Sigma N^2)(\Sigma N^3) - (\Sigma N)(\Sigma N^4)] + 2\beta_2 [(\Sigma N^3)^2 \\ - (\Sigma N^2)(\Sigma N^4)]$$

$$f_{22} = \beta_1^2 [n\Sigma N^2 - (\Sigma N)^2]$$

$$f_{23} = f_{32} = \beta_1 [(\Sigma N)(\Sigma N^3) - (\Sigma N^2)^2]$$

$$f_{33} = (\Sigma N^2)(\Sigma N^4) - (\Sigma N^3)^2 .$$

An estimate of this matrix can, of course, be obtained by replacing  $\sigma^2$  by  $s^2 = \frac{SSE}{n-3}$  and replacing  $\beta_1$ ,  $\beta_2$ , and  $\delta$ , respectively, by  $b_1$ ,  $b_2$ ,  $d$ . The formulae obtained previously as well as these give only approximate variances for our parameter estimates. Due to the complexity of both formulations it does not seem feasible to attempt a mathematical comparison. A possible method of comparison would be an empirical sampling study. This, however, will be regarded as outside the scope of the present investigation.

SINGLE FACTOR CASE: COMBINING INFORMATION  
FROM SEVERAL LOCATIONS

When several locations are included in an experiment, it becomes desirable to pool the information about  $\beta_1$  and  $\beta_2$  efficiently as well as estimate the  $\delta_i$ . Thus we are led to consider the model:

$$Y_{ij} = \beta_1 N_{ij} + \beta_2 N_{ij}^2 + \varepsilon_{ij}, \text{ where}$$

$N_{ij} = X_j + \delta_i$  refers to the  $j$ th treatment on location  $i$  in the total nutrient form and there are  $\ell$  locations in the experiment, i.e.,

$$i = 1, 2, \dots, \ell$$

and

$$j = 1, 2, \dots, n.$$

Now the least square estimates (or Maximum Likelihood in the normal error case) are obtained from minimizing

$$\begin{aligned} L = \sum_i \sum_j \varepsilon_{ij}^2 &= \sum_i \sum_j (Y_{ij} - \beta_1 N_{ij} - \beta_2 N_{ij}^2)^2 \\ &= \sum_i \sum_j [Y_{ij} - f_{ij}(\delta_1, \delta_2, \dots, \delta_\ell, \beta_1, \beta_2)]^2 \text{ say.} \end{aligned}$$

Thus:

$$\frac{\partial L}{\partial \delta_i} = -2 \sum_j [Y_{ij} - f_{ij}] \left( \frac{\partial f_{ij}}{\partial \delta_i} \right) \quad i = 1, 2, \dots, \ell$$

$$\frac{\partial L}{\partial \beta_1} = -2 \sum_i \sum_j [Y_{ij} - f_{ij}] \left( \frac{\partial f_{ij}}{\partial \beta_1} \right)$$

$$\frac{\partial L}{\partial \beta_2} = -2 \sum_i \sum_j [Y_{ij} - f_{ij}] \left( \frac{\partial f_{ij}}{\partial \beta_2} \right).$$

These are implicit equations in the parameter estimates which do not yield our estimates as simply as before. However we may, c.f. Stevens (1951), Fisher (1950), or Scarborough (1958), use the estimated asymptotic variance-covariance matrix to improve on inefficient estimates by an iterative procedure.

The Hessian of the log likelihood, neglecting  $\sigma^2$ , is obtained from:

$$\begin{aligned} -E\left(\frac{\partial^2 L}{\partial \delta_i^2}\right) &= -E\left[-\sum_{j=1}^n \left(\frac{\partial f_{ij}}{\partial \delta_i}\right)^2 + \sum_j (Y_{ij} - f_{ij}) \left(\frac{\partial^2 f_{ij}}{\partial \delta_i^2}\right)\right] \\ &= \sum_{j=1}^n \left(\frac{\partial f_{ij}}{\partial \delta_i}\right)^2 = \sum_{j=1}^n (\beta_1 + 2\beta_2 N_{ij})^2, \end{aligned}$$

for  $i \neq i'$  we have:

$$\begin{aligned} -E\left(\frac{\partial^2 L}{\partial \delta_i \partial \delta_{i'}}\right) &= 0 \\ -E\left(\frac{\partial^2 L}{\partial \beta_1 \partial \delta_i}\right) &= -E\left(\frac{\partial^2 L}{\partial \delta_i \partial \beta_1}\right) = \sum_j \left(\frac{\partial f_{ij}}{\partial \delta_i}\right) \left(\frac{\partial f_{ij}}{\partial \beta_1}\right) \\ &= \sum_j (\beta_1 N_{ij} + 2\beta_2 N_{ij}^2) = \sum_j (\beta_1 + 2\beta_2 N_{ij}) N_{ij} \\ -E\left(\frac{\partial^2 L}{\partial \beta_2 \partial \delta_i}\right) &= -E\left(\frac{\partial^2 L}{\partial \delta_i \partial \beta_2}\right) = \sum_j (\beta_1 + 2\beta_2 N_{ij}) N_{ij}^2 \\ -E\left(\frac{\partial^2 L}{\partial \beta_1 \partial \beta_2}\right) &= -E\left(\frac{\partial^2 L}{\partial \beta_2 \partial \beta_1}\right) = \sum_i \sum_j N_{ij}^3 \\ -E\left(\frac{\partial^2 L}{\partial \beta_1^2}\right) &= \sum_i \sum_j N_{ij}^2 \end{aligned}$$

and

$$-E\left(\frac{\partial^2 L}{\partial \beta_2^2}\right) = \sum_i \sum_j N_{ij}^4.$$

The procedure then is to substitute estimates  $d_i$ ,  $b_1$  and  $b_2$  into these formulae and compute  $\Delta d_i$ ,  $\Delta b_1$  and  $\Delta b_2$ , additive corrections to the inefficient estimates using the following, (which is essentially a multiple regression procedure):

Call:

$$w_{ij} = b_1 + 2b_2 \hat{N}_{ij} \quad \text{where}$$

$$\hat{N}_{ij} = X_j + d_i$$

and

$$Y'_{ij} = Y_{ij} - b_1 \hat{N}_{ij} - b_2 \hat{N}_{ij}^2$$

Then we have:

$$\begin{bmatrix} \sum_j w_{1j}^2 & 0 & 0 & \dots & 0 & \sum_j w_{1j} \hat{N}_{1j} & \sum_j w_{1j} \hat{N}_{1j}^2 & \Delta d_1 & \sum_j w_{1j} Y'_{1j} \\ 0 & \sum_j w_{2j}^2 & 0 & \dots & 0 & \sum_j w_{2j} \hat{N}_{2j} & \sum_j w_{2j} \hat{N}_{2j}^2 & \Delta d_2 & \sum_j w_{2j} Y'_{2j} \\ 0 & 0 & \cdot & & & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \sum_j w_{\lambda j}^2 & \sum_j w_{\lambda j} \hat{N}_{\lambda j} & \sum_j w_{\lambda j} \hat{N}_{\lambda j}^2 & \Delta d_{\lambda} & \sum_j w_{\lambda j} Y'_{\lambda j} \\ \sum_j w_{1j} \hat{N}_{1j} & \sum_j w_{2j} \hat{N}_{2j} & \sum_j w_{\lambda j} \hat{N}_{\lambda j} & \sum_{ij} \hat{N}_{ij}^2 & \sum_{ij} \hat{N}_{ij}^3 & \Delta b_1 & \sum_{ij} \hat{N}_{ij} Y'_{ij} \\ \sum_j w_{1j} \hat{N}_{1j}^2 & \sum_j w_{2j} \hat{N}_{2j}^2 & \sum_j w_{\lambda j} \hat{N}_{\lambda j}^2 & \sum_{ij} \hat{N}_{ij}^3 & \sum_{ij} \hat{N}_{ij}^4 & \Delta b_2 & \sum_{ij} \hat{N}_{ij}^2 Y'_{ij} \end{bmatrix} =$$

Solving for  $\Delta d_i$ ,  $\Delta b_1$ ,  $\Delta b_2$  we obtain the improved efficient estimates  $\hat{d}_i$ ,  $\hat{b}_1$ ,  $\hat{b}_2$  by

$$\hat{d}_i = d_i + \Delta d_i, \quad \hat{b}_1 = b_1 + \Delta b_1, \quad \hat{b}_2 = b_2 + \Delta b_2$$

One iteration is sufficient to obtain efficient estimates, however if we repeat the process until  $\Delta d_i, \Delta b_1, \Delta b_2$  are zero we will have the least squares or maximum likelihood estimates.

The inverse of the coefficient matrix for the corrections has intrinsic interest, for multiplied by  $\sigma^2$  and with parameters instead of estimates it would become the asymptotic variance-covariance matrix of our estimates. Thus the present form multiplied by an estimate of  $\sigma^2$ , say the residual sums of squares divided by  $n\ell - \ell - 2$ , is a consistent estimate of the asymptotic variances and covariances. Let us then inquire into a feasible computational procedure for the inverse:

Denote the coefficient matrix by  $\begin{bmatrix} D & B \\ B & A \end{bmatrix}$  and its inverse by  $\begin{pmatrix} M & N \\ N & P \end{pmatrix}$  where D represents the diagonal matrix of  $\Sigma w_{ij}^2$ 's, B the  $(\ell \times 2)$  matrix of crossproducts and A the matrix

$$A = \begin{bmatrix} \Sigma \hat{N}_{ij}^2 & \Sigma \hat{N}_{ij}^3 \\ \Sigma \hat{N}_{ij}^3 & \Sigma \hat{N}_{ij}^4 \end{bmatrix}$$

with the same partitioning in the inverse.

After some manipulation it becomes clear it is necessary to invert only the  $2 \times 2$  matrix  $[A - B'D^{-1}B] \equiv R^{-1}$ , for the inverse of D is immediate. Then we have:

$$M = D^{-1} - N B' D^{-1}$$

$$N = - D^{-1} B P$$

$$P = [A - B' D^{-1} B]^{-1} ,$$

where the forming of  $D^{-1} B$  say, is direct namely, divide every row of  $B$  by the corresponding diagonal element of  $D$ . Thus no matter how many locations are involved no matrix larger than  $2 \times 2$  need be inverted. This should be quite important from the computational accuracy standpoint.

The remaining question for the implementation of this technique of improving estimates is a systematic procedure of obtaining starting values  $d_1$ ,  $b_1$ , and  $b_2$ . Let us, then, illustrate the method with some of the data from the T.V.A. -N.C. State experiments.

In Table 3 are listed mean yields in bu./A. of 4 replications for 4 locations on the Norfolk soil type from the 1956 growing season. Also listed are the coded applied nitrogen rates coded as follows:

Applied nitrogen lb./A.	0	62.5	125	187.5	250
Coded value	-2	-1	0	1	2

Quadratics of the form

$$\hat{Y} = b_0^* + b_1^*x + b_2^*x^2 \quad \text{were fitted for each loca-}$$

tion using the coded values and are given in Table 4, together with their alternate representations.

Table 3. Mean yields in bu./A. and coded applied nitrogen rates for 4 locations

applied N (coded)	Location			
	1	2	3	4
-2	46.5	41.8	45.4	24.9
-2	46.5	41.5	37.9	33.4
-1	87.0	69.9	83.8	71.0
-1	92.0	72.4	72.1	71.4
-1	88.1	70.7	80.4	76.0
-1	91.4	71.4	78.4	70.2
0	112.4	88.2	106.5	92.9
0	100.8	77.5	108.5	95.3
0	115.9	79.1	102.8	99.2
0	115.1	87.0	106.1	99.6
0	110.1	86.1	99.3	96.5
1	107.6	83.4	97.2	102.2
1	110.8	92.6	106.0	98.6
1	108.2	81.7	106.6	98.2
1	113.4	87.7	114.1	106.0
2	112.2	82.3	106.0	101.4
2	111.9	81.4	106.4	100.2
2	109.0	88.5	108.6	103.4

Table 4. Quadratic coefficients for the data in Table 3 with alternate representations of parameter estimates.

Location	$b_0^*$	$b_1^*$	$b_2^*=b_2$	$b_1$	d coded	d decoded lbs./A.
1	109.0	14.43	-7.53	59.09	2.96	60.0
2	83.7	9.72	-5.11	42.49	3.21	75.6
3	101.5	15.63	-7.02	55.63	2.85	53.1
4	95.3	17.10	-7.50	56.14	2.60	37.5
Average			-6.79	53.3375		

Table 4 indicates the sort of results that are often obtained from experiments of this type. The linear coefficients are usually positive with the quadratic usually negative. This is what would generally be expected, indicating the response goes through a maximum and thereafter declines.

Occasionally  $b_2$  is estimated as positive. When this happens, it usually means that there is either little response to the factor or that the range of applied nutrient levels has not been sufficiently great to expose the diminishing returns portion of the response. If the latter is the case, the normal recommendation would be to experiment at additional levels before proceeding with the agro-economic

analysis. Thus in either case the locations with unusual signs would not be considered appropriate to be included in the present proposed analysis.

The d's are decoded by subtracting 2 from each and multiplying by 62.5. It is probably easier to work with the parameters in the semi-coded form, however, with units of nitrogen of 62.5 lbs./A. and the origin at zero.

For the starting values of  $b_1$  and  $b_2$  the average is a convenient approximation. The average  $b_1$  and  $b_2$  can be used moreover to obtain "smoothed" starting values for the  $d_i$  by the formula

$$d_i \text{ smoothed} = \frac{b_{1i}^* - \bar{b}_1}{2\bar{b}_2} .$$

We thus have for the starting values:

$$\begin{array}{lll} \bar{b}_1 = 53.3375 & d_1 = .8652 & d_3 = .7768 \\ \bar{b}_2 = -6.79 & d_2 = 1.2120 & d_4 = .6686 . \end{array}$$

Using these in the computing formulae we form the matrices:

$$D = \text{diag}(8096.7385, 6343.8887, 8689.1926, 9486.3321)$$

$$A = \begin{bmatrix} 759.2894 & 2951.1452 \\ 2951.1452 & 12232.0230 \end{bmatrix}$$

$$B = \begin{bmatrix} 312.5240 & 181.6201 \\ 115.1743 & -624.0633 \\ 355.3038 & 345.0616 \\ 402.2125 & 519.4281 \end{bmatrix}$$

with the right-hand side having the elements

$$\Sigma w_{1j} Y_{1j} = 2486.2239$$

$$\Sigma w_{2j} Y_{2j} = -2315.8961$$

$$\Sigma w_{3j} Y_{3j} = 1425.3303$$

$$\Sigma w_{4j} Y_{4j} = 101.2194$$

$$\Sigma \Sigma \hat{N}_{ij} Y_{ij} = -14.9404$$

$$\Sigma \Sigma \hat{N}_{ij}^2 Y_{ij} = -318.9638$$

Then:

$$D^{-1}B = \begin{bmatrix} .03860 & .02243 \\ .01816 & -.09837 \\ .04089 & .03971 \\ .04240 & .05476 \end{bmatrix}$$

and

$$B \cdot D^{-1}B = \begin{bmatrix} 45.7372 & 31.8145 \\ 31.8145 & 107.6091 \end{bmatrix}$$

$$P^{-1} = A - B \cdot D^{-1} B = \begin{bmatrix} 713.5522 & 2919.3307 \\ 2919.3307 & 12124.4139 \end{bmatrix}$$

$$P = \begin{bmatrix} .09405 & -.022646 \\ -.022646 & .005535 \end{bmatrix}$$

$$N = \begin{bmatrix} -.003122 & .0007500 \\ -.003936 & .0009557 \\ -.002946 & .0007062 \\ -.002747 & .0006571 \end{bmatrix}$$

$$M = \begin{bmatrix} .0002272 & .0001305 & .0000979 & .0000913 \\ .0001305 & .0003231 & .0001230 & .0001146 \\ .0000979 & .0001230 & .0002075 & .0000862 \\ .0000913 & .0001146 & .0000862 & .0001859 \end{bmatrix}$$

The matrix

$\begin{pmatrix} M & N \\ N^T & P \end{pmatrix}$  is then multiplied by the vector of

right-hand side elements giving the corrections:

$$\Delta d_1 = .2188 \quad \Delta d_3 = .0818 \quad \Delta b_1 = 2.6944$$

$$\Delta d_2 = -.4829 \quad \Delta d_4 = -.0653 \quad \Delta b_2 = -.7027$$

These corrections are then added to the d's and b's

to give:

$$\begin{array}{lll} d_1 = 1.0888 & d_3 = .8618 & b_1 = 56.0319 \\ d_2 = .7271 & d_4 = .6047 & b_2 = -7.4927 \end{array} .$$

If the process is terminated at one iteration the estimates are efficient but, of course, neither maximum likelihood nor least squares estimates are thereby obtained. These are the results only of repeating the iteration in either one of two ways until convergence is reached.

The first method is to construct the matrix anew with the corrected estimates and proceed as before until the corrections are zero or equivalently the right-hand side vector is zero. The second method is to recalculate only the right-hand side vector anew until convergence is reached. The second method generally involves more iterations but considerably less calculation per iteration.

The results of 9 iterations by the second method are presented graphically in Figure 1. The final estimates of the parameters are then:

$$\begin{array}{lll} d_1 = 1.326 & d_3 = 1.086 & b_1 = 49.654 \\ d_2 = .745 & d_4 = .789 & b_2 = -5.843 \end{array} .$$

The estimated quadratic is then

$$\hat{Y} = 49.654N - 5.843N^2$$

and the decoded  $d_i$  are in lbs. N/A.,

$$\begin{array}{ll} d_1 = 82.9 & d_3 = 67.9 \\ d_2 = 46.4 & d_4 = 49.3 \end{array} .$$

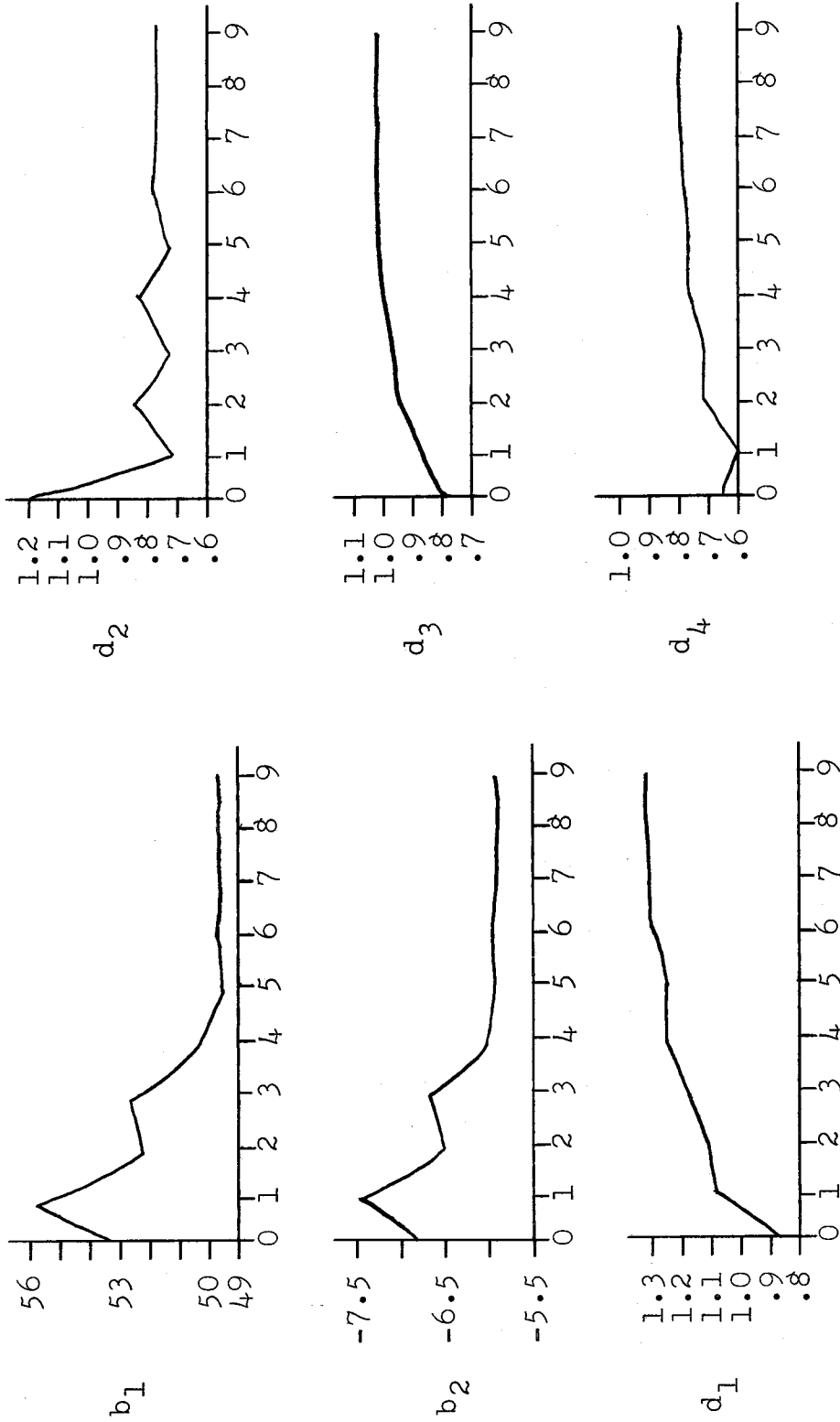


Figure 1. A graphical presentation of the results of 9 iterations using method 2 with the successive estimates plotted against the iteration number.

Thus we see that economic recommendations for the different locations would be rather strongly affected by the differing initial nutrient levels estimated. The estimated biological maximum is 105.5 bu./A. with application of enough N to result in 4.24 units of total nitrogen or 265.6 lbs. N/A.

## SINGLE FACTOR: VARIABLE NUTRIENT CASE

Looking back on the results and philosophy of our approach so far, we see need for another alteration in the model. If locations are different because of soil nutrient factors, then differences between well defined blocks at a single location and, indeed, between plots treated alike within a block must be different at least in part due to initial nutrient differences in the plots.

The practice of some statisticians and agronomists of using soil tests and visual site inspections to select plots to be used in the same block are evidences of at least implicit recognition of this. Thus our basic source of variation must be organic to the experimental material and not merely an added measurement error unrelated to the material.

In the usual analysis of variance techniques we can alleviate this situation because in that case we are dealing with a model containing a sum of constants and random variables so that any "unexplainable" variation will be absorbed by the error term affixed to the end or amalgamated into the overall mean.

For most practical cases in the analysis of variance, then, we introduce various components of variance by our errors of specification, see e.g., Kempthorne and Zyskind (1960). When the errors are in the arguments of other than

linear functions the situation alters in kind, for then the functional assumption comes into play, usually, it must be admitted, in a very unhandy fashion.

To elucidate a little further the nature of the alteration proposed, we may draw attention to the use of "block x treatment interactions" to test treatment comparisons instead of "between duplicate measurements" as common good practice among statisticians. In the context we are using here, "between duplicates" is really measuring an additive error of measurement external to the variability in experimental material or procedural repeatability, (which is in the "block x treatment interaction").

For definiteness consider the model  $Y = f(A)$  where  $f$  is the quadratic response to factor  $A$ , as before, and

$$A_j^\dagger = X_j + \delta_j^\dagger$$

where  $X_j$  = applied level of factor  $A$  and

$$\delta_j^\dagger = \text{random variable, an existing level of}$$

factor  $A$ . The random variable is now an integral part of the response function and not a tacked on afterthought.

Since measurement error is customarily small relative to variation among experimental units, we feel that this formulation properly accounts for the largest part of the variation, whereas the additive error case concentrates attention on the almost negligible additive measurement error. If sufficient information is at hand to identify an additive error as well it should of course be included.

Non-linearity in this form presents great mathematical difficulties, so that one should not expect definitive solutions but rather first approximations. There is something to be said however for an attempt on the right problem in contrast to elaborate and definitive solutions to irrelevant problems. With this in mind consider the following stratagem:

Let us take expectations of  $b_0^*$ ,  $b_1^*$ ,  $b_2^*$  assuming the variable nutrient model in the form:

$$Y_j^* = \alpha_0 + \alpha_1(X_j + \delta_j^*) + \alpha_2(X_j + \delta_j^*)^2 . \quad \text{Then}$$

$$E \begin{pmatrix} \Sigma Y \\ \Sigma XY \\ \Sigma X^2 Y \end{pmatrix} = E \begin{pmatrix} n & \Sigma X_j + \Sigma \delta_j^* & \Sigma X_j^2 + 2 \Sigma X_j \delta_j^* + \Sigma \delta_j^{*2} \\ \Sigma X_j & \Sigma X_j \delta_j^* + \Sigma X_j^2 & \Sigma X_j^3 + 2 \Sigma X_j^2 \delta_j^* + \Sigma X_j \delta_j^{*2} \\ \Sigma X_j^2 & \Sigma X_j^3 + \Sigma X_j^2 \delta_j^* & \Sigma X_j^4 + 2 \Sigma X_j^3 \delta_j^* + \Sigma X_j^2 \delta_j^{*2} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$$

$$= \begin{bmatrix} n & \Sigma X_j + n \mu_1^* & \Sigma X_j^2 + 2 \mu_1^* \Sigma X_j + n \mu_2^* \\ \Sigma X_j & \Sigma X_j^2 + \mu_1^* \Sigma X_j & \Sigma X_j^3 + 2 \mu_1^* \Sigma X_j^2 + \mu_2^* \Sigma X_j \\ \Sigma X_j^2 & \Sigma X_j^3 + \mu_1^* \Sigma X_j^2 & \Sigma X_j^4 + 2 \mu_1^* \Sigma X_j^3 + \mu_2^* \Sigma X_j^2 \end{bmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$$

where:  $\mu_1^* = E(\delta_j^*)$

$$\mu_2^* = E(\delta_j^{*2}) .$$

Now rewrite the matrix as:

$$\left\{ \begin{bmatrix} n & \Sigma X & \Sigma X^2 \\ \Sigma X & \Sigma X^2 & \Sigma X^3 \\ \Sigma X^2 & \Sigma X^3 & \Sigma X^4 \end{bmatrix} + \begin{bmatrix} n & n & n \\ \Sigma X & \Sigma X & \Sigma X \\ \Sigma X^2 & \Sigma X^2 & \Sigma X^2 \end{bmatrix} \begin{bmatrix} 0 & \mu_1' & \mu_2' \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right. \\ \left. + \begin{bmatrix} \Sigma X & \Sigma X & \Sigma X \\ \Sigma X^2 & \Sigma X^2 & \Sigma X^2 \\ \Sigma X^3 & \Sigma X^3 & \Sigma X^3 \end{bmatrix} \begin{bmatrix} 0 & 0 & 2\mu_1' \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right\} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = E \begin{pmatrix} \Sigma Y \\ \Sigma XY \\ \Sigma X^2 Y \end{pmatrix} .$$

Multiplying both sides by:

$$\begin{bmatrix} n & \Sigma X & \Sigma X^2 \\ \Sigma X & \Sigma X^2 & \Sigma X^3 \\ \Sigma X^2 & \Sigma X^3 & \Sigma X^4 \end{bmatrix}^{-1}$$

we have

$$\left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & \mu_1' & \mu_2' \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 2\mu_1' \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right\} \\ \text{(times)} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = E \begin{pmatrix} b_0^* \\ b_1^* \\ b_2^* \end{pmatrix}$$

Therefore

$$E \begin{pmatrix} b_0^* \\ b_1^* \\ b_2^* \end{pmatrix} = \begin{bmatrix} 1 & \mu_1^* & \mu_2^* \\ 0 & 1 & 2\mu_1^* \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} .$$

If we make the identification:

$$\begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 0 \\ \beta_1 \\ \beta_2 \end{pmatrix} , \quad \mu_1^* = \delta , \quad \mu_2^* = \sigma_\delta^2 + \delta^2$$

we have:

$$E(b_0^*) = \delta\beta_1 + (\sigma_\delta^2 + \delta^2)\beta_2 = \delta\beta_1 + \delta^2\beta_2 + \beta_0^* \quad \text{say}$$

$$E(b_1^*) = \beta_1 + 2\delta\beta_2$$

$$E(b_2^*) = \beta_2$$

and we see the very interesting point that the variance of the initial nutrient can bias the parameter estimates. It will be noted that when  $\sigma_\delta^2 = 0$  we have our earlier situation. Substituting in our formulae for  $b_1, b_2, d$  we have:

$$E(b_1) = \beta_1 \sqrt{1 - \frac{4\beta_2^2\sigma^2}{\beta_1^2}}$$

$$E(b_2) = \beta_2$$

$$E(d) = \delta + \frac{\beta_1}{2\beta_2} \left( 1 - \sqrt{1 - \frac{4\beta_2^2\sigma^2}{\beta_1^2}} \right) ,$$

where  $\bar{E}$  is to denote the fact that we have substituted the expected values of  $b_0^*$ ,  $b_1^*$ , and  $b_2^*$  into the formulae for  $b_1$ ,  $b_2$ , and  $d$  and that these equations are not valid expectations.

About the only conclusion one can draw at this very tentative stage is that great care should be taken to achieve homogeneity among plots since variability can possibly upset to some degree at least the estimation of parameters.

Note that a very curious parameter  $\beta_0'$  makes its appearance. It is chimerical in that if there is no variability in  $\delta_j'$  it disappears. It is a mathematical consequence of our random initial nutrient formulation in a model not possessing a constant term, i.e.,

$$Y_j' = \beta_1 N_j' + \beta_2 N_j'^2 .$$

Further it developed in taking the expectation of a linear function of random variables.

Unfortunately our equations are underidentified in the single factor case (except for  $b_2$ ), however there is a slight alleviation of this situation in the multiple factor generalizations. Before we proceed further, however, let us have a closer look at the underlying sampling situation by considering the randomization development of this model.

If we regard the  $n$  plots in an individual experiment as a sample from a location, and designate the totality of possible samples of size  $n$  from an individual location  $k$ ,

then we may consider the way an experiment is customarily performed as double sampling from a finite universe. Taking expectations, then, will be two-stage, i.e., over the randomizations of a sample of size  $n$  to  $n$  fixed plots, then over the  $k$  sample configurations.

Let us designate the observations, then, as:

$$Y'_{ij} = \beta_1(X_j + \delta'_{ij}) + \beta_2(X_j + \delta'_{ij})^2,$$

where:  $j = 1, 2, \dots, n$

$i = 1, 2, \dots, k$

and the primed  $\delta$ 's denote variables over which the randomization is to be performed.

Now:

$$Y'_{ij} = \beta_1 X_j + \beta_1 \delta'_{ij} + \beta_2 X_j^2 + 2\beta_2 X_j \delta'_{ij} + \beta_2 \delta'^2_{ij}.$$

Then:

$$E(Y'_{ij}) = \beta_1 X_j + \beta_1 \frac{1}{n} \sum_{j=1}^n \delta'_{ij} + \beta_2 X_j^2 + 2\beta_2 X_j \frac{1}{n} \sum_{j=1}^n \delta'_{ij} + \beta_2 \frac{1}{n} \sum_{j=1}^n \delta'^2_{ij}.$$

The quantities:

$$\frac{1}{n} \sum_{j=1}^n \delta'_{ij} = \overline{\delta'_i}$$

and  $\frac{1}{n} \sum_{j=1}^n \delta'^2_{ij} = \overline{\delta'^2_i}$ , say

are then the first two non-central moments of the random variable in the finite space of a given  $i$ th sample of size  $n$ .

Since the  $b_i^*$  are linear functions of this typical random variable we may proceed with the expectation to either of two cases. One, we may have only one experiment at a given location and choose to call  $\overline{\delta_i^*}$  and  $\overline{\delta_i^{*2}}$  the corresponding characterizations of that  $i$ th location. Two, we may regard the experiment at location  $i$  as one of the  $k$  possible and consider the  $k$  samples of  $n$  from location  $i$ .

In the first case it is evident that the results derived previously are valid with  $\overline{\delta_i^*}$  and  $\overline{\delta_i^{*2}}$  replacing  $\mu_1^*$  and  $\mu_2^*$  respectively.

In the second case we have:

$$E[E(Y_{ij}^*)] = \beta_1 X_j + \beta_1 \frac{1}{k} \sum_{i=1}^k \overline{\delta_i^*} + \beta_2 X_j^2 + 2\beta_2 X_j \frac{1}{k} \sum_{i=1}^k \overline{\delta_i^*} + \beta_2 \frac{1}{k} \sum_{i=1}^k \overline{\delta_i^{*2}}$$

and the results previously derived are valid if we let

$$\mu_1^* = \frac{1}{k} \sum_{i=1}^k \overline{\delta_i^*}$$

and 
$$\mu_2^* = \frac{1}{k} \sum_{i=1}^k \overline{\delta_i^{*2}} .$$

It should be noted, then, that the randomization derivation tells us two things. One, it does not invalidate the previous results. Two, it does render quite obviously the treatment assignments independent of the plots, i.e., the soil nutrient random variables.

## MULTIPLE FACTOR: VARIABLE NUTRIENT CASE

The next logical generalization of our model is to the multiple factor situation with random variation in the initial levels of each of the factors on a plot-to-plot basis. This is probably the most realistic formulation of all from the agronomic standpoint, for although many fertilizer experiments reduce to one effective nutrient response, many others do not, but, rather, show response to several factors such as N, P, K, etc.

Let us, then, examine the biological structure of the situation in some detail.

First of all, it seems clear that the latent initial nutrient levels of the responding factors in our models must be distributed in some multivariate form. That is to say, the initial nutrient levels of the several factors are not independent but condition one another in their presences and effects.

Next consider the following situation: A completely unproductive field has an important nutrient, nitrogen, say, added to it and crop response occurs. We do not think it reasonable to conclude that total nitrogen alone was zero and that the response observed was a function of only applied nitrogen, without any corollary effects due to the changed chemical complex that is the soil. We think it more reasonable to consider all soil nutrients as zero in response effect in the absence of a "threshold" amount of

nitrogen, and that following the fertilizer application the chemical nature of the soil becomes altered so that latent conditional nutrient levels are released on all effective factors. These then jointly act in an observable result as an initial level of nitrogen. Any interpretation of initial level terms in the model should take this into account by recognizing the conditional nature of the postulated  $\delta_i$ . Thus we formulate a version of Liebig's (1855) Law of the Minimum as a reasonable biological requirement.

As before  $\delta$  is to be regarded as the complex of physical and chemical states that act in their effect on yield as an amount of a latent nutrient factor measured in the units and form in which this nutrient would normally be applied.

Another concept the above discussion leads us into is that of substitution. It is, of course, not to be considered as stating that a plant can use nitrogen, say, interchangeably in place of phosphorus in all its nutritive requirements. Rather the varying rates of applied nutrients act to alter the chemical complex of the soil which then either releases or inhibits the release of nutrients. Therefore different combinations of applied nutrient factors can produce similar net yield effects, within limits, through alteration of the chemical complex. (This implies a possible need for interaction terms in the model.)

Overwhelming evidence demands a curvilinear model with the possibility of diminishing returns to applied nutrients, either when considered separately or jointly. Negative "pure" quadratic coefficients would allow for this in the simplest possible way with the partial derivatives of the function with respect to each variable separately yielding straight lines with negative slopes.

Finally we want our generalization to be consistent with the more restricted single nutrient models considered previously. Thus a single nutrient model such as:

$$Y_j^i = \beta_0^i + \beta_1 N_j^i + \beta_2 N_j^{i2}$$

should follow simply from our generalized form. Just as this single total nutrient model yields a complete quadratic in the applied nutrient function, we should desire our generalized form to lead to complete quadratics in the multiple factor applied nutrients forms. Hopefully also we can limit redundancies in our generalized model to the minimum in order to identify as many effects as possible.

The proposed form will be stated in terms of two factors purely for convenience in expression. The generalization to  $k$  factors will be immediate. The total nutrient formulation for two factors is then:

$$Y_j^i = \beta_{00}^i + \beta_{10} N_j^i + \beta_{20} P_j^i + \beta_{11} N_j^{i2} + \beta_{22} P_j^{i2} \\ + \beta_{12} N_j^i P_j^i + \epsilon_j \quad ,$$

where:

$$N_j^* = X_{1j} + \delta_{1j}^* , \quad E(\delta_{1j}^*) = \delta_1 , \quad V(\delta_{1j}^*) = \sigma_{\delta_1}^2$$

$$P_j^* = X_{2j} + \delta_{2j}^* , \quad E(\delta_{2j}^*) = \delta_2 , \quad V(\delta_{2j}^*) = \sigma_{\delta_2}^2$$

$$\text{cov}(\delta_{1j}^* , \delta_{2j}^*) = \rho_{12} \sigma_{\delta_1} \sigma_{\delta_2}$$

$$E(\varepsilon_j) = 0 , \quad V(\varepsilon_j) = \sigma_{\varepsilon}^2 ,$$

and  $\beta_{00}^*$  is the corresponding ephemeral parameter we obtained in the single factor variable nutrient case.

Thus the total nutrient is composed of an applied amount  $X_{ij}$  and a random initial level  $\delta_{ij}^*$  where the prime on the delta denotes a random variable. The  $\varepsilon$  is a random additive error term independent of the  $\delta_{ij}^*$ .

We may write the applied nutrient formulation of this model by taking expectations of  $Y_j^*$  as follows:

$$E(Y_j^*) = \beta_{00}^* + \beta_{10}^* X_{1j} + \beta_{20}^* X_{2j} + \beta_{11}^* X_{1j}^2 + \beta_{22}^* X_{2j}^2 + \beta_{12}^* X_{1j} X_{2j} .$$

To see the relationship between the starred and unstarred parameters in our formulations we note that:

$$\begin{pmatrix} \beta_{00}^* \\ \beta_{10}^* \\ \beta_{20}^* \\ \beta_{11}^* \\ \beta_{22}^* \\ \beta_{12}^* \end{pmatrix} = E \begin{pmatrix} b_{00}^* \\ b_{10}^* \\ b_{20}^* \\ b_{11}^* \\ b_{22}^* \\ b_{12}^* \end{pmatrix} = A^{-1} E \begin{pmatrix} \Sigma Y^* \\ \Sigma X_1 Y^* \\ \Sigma X_2 Y^* \\ \Sigma X_1^2 Y^* \\ \Sigma X_2^2 Y^* \\ \Sigma X_1 X_2 Y^* \end{pmatrix}$$

where

$$A = \begin{bmatrix} n & \Sigma X_1 & \Sigma X_2 & \Sigma X_1^2 & \Sigma X_2^2 & \Sigma X_1 X_2 \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1 X_2 & \Sigma X_1^3 & \Sigma X_1 X_2^2 & \Sigma X_1^2 X_2 \\ \Sigma X_2 & \Sigma X_1 X_2 & \Sigma X_2^2 & \Sigma X_1^2 X_2 & \Sigma X_2^3 & \Sigma X_1 X_2^2 \\ \Sigma X_1^2 & \Sigma X_1^3 & \Sigma X_1^2 X_2 & \Sigma X_1^4 & \Sigma X_1^2 X_2^2 & \Sigma X_1^3 X_2 \\ \Sigma X_2^2 & \Sigma X_1 X_2^2 & \Sigma X_2^3 & \Sigma X_1^2 X_2^2 & \Sigma X_2^4 & \Sigma X_1 X_2^3 \\ \Sigma X_1 X_2 & \Sigma X_1^2 X_2 & \Sigma X_1 X_2^2 & \Sigma X_1^3 X_2 & \Sigma X_1 X_2^3 & \Sigma X_1^2 X_2^2 \end{bmatrix}$$

Thus we can proceed by taking the last indicated expectation and then multiply by the inverse coefficient matrix of the  $b_{ij}^*$ .

Now taking the expectations and summing we have:

$$E \begin{pmatrix} \Sigma Y' \\ \Sigma X_1 Y' \\ \Sigma X_2 Y' \\ \Sigma X_1^2 Y' \\ \Sigma X_2^2 Y' \\ \Sigma X_1 X_2 Y' \end{pmatrix} = \begin{bmatrix} n & \Sigma X_1 + n\delta_1 & \Sigma X_2 + n\delta_2 \\ \Sigma X_1 & \Sigma X_1^2 + \Sigma X_1 \delta_1 & \Sigma X_1 X_2 + \Sigma X_1 \delta_2 \\ \Sigma X_2 & \Sigma X_1 X_2 + \Sigma X_2 \delta_1 & \Sigma X_2^2 + \Sigma X_2 \delta_2 \\ \Sigma X_1^2 & \Sigma X_1^3 + \Sigma X_1^2 \delta_1 & \Sigma X_1^2 X_2 + \Sigma X_1^2 \delta_2 \\ \Sigma X_2^2 & \Sigma X_1 X_2^2 + \Sigma X_2^2 \delta_1 & \Sigma X_2^3 + \Sigma X_2^2 \delta_2 \\ \Sigma X_1 X_2 & \Sigma X_1^2 X_2 + \Sigma X_1 X_2 \delta_1 & \Sigma X_1 X_2^2 + \Sigma X_1 X_2 \delta_2 \end{bmatrix}$$

$$\Sigma X_1^2 + 2\Sigma X_1 \delta_1 + n(\sigma_{\delta_1}^2 + \delta_1^2)$$

$$\Sigma X_2^2 + 2\Sigma X_2 \delta_2 + n(\sigma_{\delta_2}^2 + \delta_2^2)$$

$$\Sigma X_1^3 + 2\Sigma X_1^2 \delta_1 + \Sigma X_1 (\sigma_{\delta_1}^2 + \delta_1^2)$$

$$\Sigma X_1 X_2^2 + 2\Sigma X_1 X_2 \delta_2 + \Sigma X_1 (\sigma_{\delta_2}^2 + \delta_2^2)$$

$$\Sigma X_1^2 X_2 + 2\Sigma X_1 X_2 \delta_1 + \Sigma X_2 (\sigma_{\delta_1}^2 + \delta_1^2)$$

$$\Sigma X_2^3 + 2\Sigma X_2^2 \delta_2 + \Sigma X_2 (\sigma_{\delta_2}^2 + \delta_2^2)$$

$$\Sigma X_1^4 + 2\Sigma X_1^3 \delta_1 + \Sigma X_1^2 (\sigma_{\delta_1}^2 + \delta_1^2)$$

$$\Sigma X_1^2 X_2^2 + 2\Sigma X_1 X_2 \delta_2 + \Sigma X_1^2 (\sigma_{\delta_2}^2 + \delta_2^2)$$

$$\Sigma X_1^2 X_2^2 + 2\Sigma X_1 X_2^2 \delta_1 + \Sigma X_2^2 (\sigma_{\delta_1}^2 + \delta_1^2)$$

$$\Sigma X_2^4 + 2\Sigma X_2^3 \delta_2 + \Sigma X_2^2 (\sigma_{\delta_2}^2 + \delta_2^2)$$

$$\Sigma X_1^3 X_2 + 2\Sigma X_1^2 X_2 \delta_1 + \Sigma X_1 X_2 (\sigma_{\delta_1}^2 + \delta_1^2)$$

$$\Sigma X_1 X_2^3 + 2\Sigma X_1 X_2^2 \delta_2 + \Sigma X_1 X_2 (\sigma_{\delta_2}^2 + \delta_2^2)$$

$$\Sigma X_1 X_2 + \Sigma X_2 \delta_1 + \Sigma X_1 \delta_2 + n(\rho_{12} \sigma_{\delta_1} \sigma_{\delta_2} + \delta_1 \delta_2)$$

$$\Sigma X_1^2 X_2 + \Sigma X_1 X_2 \delta_1 + \Sigma X_1^2 \delta_2 + \Sigma X_1 (\rho_{12} \sigma_{\delta_1} \sigma_{\delta_2} + \delta_1 \delta_2)$$

$$\Sigma X_1 X_2^2 + \Sigma X_2^2 \delta_1 + \Sigma X_1 X_2 \delta_2 + \Sigma X_2 (\rho_{12} \sigma_{\delta_1} \sigma_{\delta_2} + \delta_1 \delta_2)$$

$$\Sigma X_1^3 X_2 + \Sigma X_1^2 X_2 \delta_1 + \Sigma X_1^3 \delta_2 + \Sigma X_1^2 (\rho_{12} \sigma_{\delta_1} \sigma_{\delta_2} + \delta_1 \delta_2)$$

$$\Sigma X_1 X_2^3 + \Sigma X_2^3 \delta_1 + \Sigma X_1 X_2^2 \delta_2 + \Sigma X_2^2 (\rho_{12} \sigma_{\delta_1} \sigma_{\delta_2} + \delta_1 \delta_2)$$

$$\Sigma X_1^2 X_2^2 + \Sigma X_1 X_2^2 \delta_1 + \Sigma X_1^2 X_2 \delta_2 + \Sigma X_1 X_2 (\rho_{12} \sigma_{\delta_1} \sigma_{\delta_2} + \delta_1 \delta_2)$$

$$\begin{bmatrix} 0 \\ \beta_{10} \\ \beta_{20} \\ \beta_{11} \\ \beta_{22} \\ \beta_{12} \end{bmatrix}$$

$$= A + \begin{pmatrix} n \\ \Sigma X_1 \\ \Sigma X_2 \\ \Sigma X_1^2 \\ \Sigma X_2^2 \\ \Sigma X_1 X_2 \end{pmatrix} 1' \begin{pmatrix} 0 & \delta_1 & \delta_2 & (\sigma_{\delta_1}^2 + \delta_1^2) & (\sigma_{\delta_2}^2 + \delta_2^2) & (\rho_{12} \sigma_{\delta_1} \sigma_{\delta_2} + \delta_1 \delta_2) \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$+ \begin{pmatrix} \Sigma X_1 \\ \Sigma X_1^2 \\ \Sigma X_1 X_2 \\ \Sigma X_1^3 \\ \Sigma X_1 X_2^2 \\ \Sigma X_1^2 X_2 \end{pmatrix} 1' \begin{pmatrix} 0 & 0 & 0 & 2\delta_1 & 0 & \delta_2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$+ \begin{pmatrix} \Sigma X_2 \\ \Sigma X_1 X_2 \\ \Sigma X_2^2 \\ \Sigma X_1^2 X_2 \\ \Sigma X_2^3 \\ \Sigma X_1 X_2^2 \end{pmatrix} 1' \begin{pmatrix} 0 & 0 & 0 & 0 & 2\delta_2 & \delta_1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ \beta_{10} \\ \beta_{20} \\ \beta_{11} \\ \beta_{22} \\ \beta_{12} \end{pmatrix}$$

where  $1' = (1, 1, 1, 1, 1, 1)$  .

Multiplying by  $A^{-1}$  we have

$$\begin{pmatrix} \beta_{00}^* \\ \beta_{10}^* \\ \beta_{20}^* \\ \beta_{11}^* \\ \beta_{22}^* \\ \beta_{12}^* \end{pmatrix} = E \begin{pmatrix} b_{00}^* \\ b_{10}^* \\ b_{20}^* \\ b_{11}^* \\ b_{22}^* \\ b_{12}^* \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \delta_1 & \delta_2 & (\sigma_{\delta_1}^2 + \delta_1^2) & (\sigma_{\delta_2}^2 + \delta_2^2) & (\rho_{12}\sigma_{\delta_1}\sigma_{\delta_2} + \delta_1\delta_2) \\ 0 & 1 & 0 & 2\delta_1 & 0 & \delta_2 \\ 0 & 0 & 1 & 0 & 2\delta_2 & \delta_1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \beta_{10} \\ \beta_{20} \\ \beta_{11} \\ \beta_{22} \\ \beta_{12} \end{pmatrix}$$

Then we obtain:

$$\begin{aligned} \beta_{00}^* &= \beta_{10}\delta_1 + \beta_{20}\delta_2 + \beta_{11}\delta_1^2 + \beta_{22}\delta_2^2 + \beta_{12}\delta_1\delta_2 \\ &\quad + (\sigma_{\delta_1}^2\beta_{11} + \sigma_{\delta_2}^2\beta_{22} + \rho_{12}\sigma_{\delta_1}\sigma_{\delta_2}\beta_{12}) \end{aligned}$$

$$\beta_{10}^* = \beta_{10} + 2\beta_{11}\delta_1 + \beta_{12}\delta_2$$

$$\beta_{20}^* = \beta_{20} + \beta_{12}\delta_1 + 2\beta_{22}\delta_2$$

$$\beta_{11}^* = \beta_{11}$$

$$\beta_{22}^* = \beta_{22}$$

$$\beta_{12}^* = \beta_{12}$$

The quantity in the parenthesis is  $\beta'_{00}$ , which is a consequence of the variable initial nutrient level formulation. An interpretation of  $\beta'_{00}$  is in terms of the "threshold" response phenomenon noted earlier. Thus, when a nutrient is below its threshold level, all nutrients are prevented from showing response and the function formulated passes through the origin. Above this threshold value the function appears to have a non-zero intercept; since the intercept is a combination of quadratic coefficients, one would expect it to be estimated as negative in most experiments.

There are several rather interesting statistical aspects to this model. One is the addition of a known  $X_{ij}$  to the random variable  $\delta'_{ij}$  to form the "total nutrient" variable. We have, it will be recognized, a case of "controlled independent variables" as first considered by Berkson (1950). There are several differences, however. The first is that the model is non-linear and thus is more similar to the cases considered by Geary (1953). Geary also obtained variances of the estimates of the parameters in a somewhat similar fashion to our development. A two-fold difference exists between these models and those considered by Geary, however. Here we assume the model is a function of more than one variable and furthermore the random initial nutrient variables do not have zero expectations. This last point allows relative estimation of the

$\delta_{ij}$  and  $\beta_{i0}$  if more than one location is available in the experiment. It will be seen, however, that without further information or assumptions we can estimate only a linear function of the variances and covariances of the initial nutrient levels, namely  $\beta'_{00}$ .

To see the effect of the present formulation of multiple factor models let us neglect the additional complications that the random initial nutrient levels give us and try to estimate the parameters by assuming only an additive normally distributed error  $\varepsilon$ . For two locations the expression to be minimized for the two factor model is (neglecting multiplicative constants):

$$\begin{aligned} L &= \sum \sum \varepsilon_{ij}^2 = \sum \varepsilon_{1j}^2 + \sum \varepsilon_{2j}^2 \\ &= \sum (Y_1 - \beta_{10}N_1 - \beta_{20}P_1 - \beta_{11}N_1^2 - \beta_{22}P_1^2 - \beta_{12}N_1P_1)^2 \\ &\quad + \sum (Y_2 - \beta_{10}N_2 - \beta_{20}P_2 - \beta_{11}N_2^2 - \beta_{22}P_2^2 - \beta_{12}N_2P_2)^2, \end{aligned}$$

where:

$$\begin{aligned} N_{1j} &= X_{1j} + \delta_{11} && \text{for the 1st location} \\ N_{2j} &= X_{1j} + \delta_{12} && \text{for the 2nd location} \\ \text{and} \quad P_{1j} &= X_{2j} + \delta_{21} && \text{for the 1st location} \\ P_{2j} &= X_{2j} + \delta_{22} && \text{for the 2nd location} . \end{aligned}$$

Then calling:

$$\begin{aligned} S_1 &= \sum (Y_1 - \beta_{10}N_1 - \beta_{20}P_1 - \beta_{11}N_1^2 - \beta_{22}P_1^2 \\ &\quad - \beta_{12}N_1P_1) \\ \text{and} \quad S_2 &= \sum (Y_2 - \beta_{10}N_2 - \beta_{20}P_2 - \beta_{11}N_2^2 - \beta_{22}P_2^2 \\ &\quad - \beta_{12}N_2P_2) , \end{aligned}$$

we have:

$$1 \quad -\frac{1}{2} \frac{\partial L}{\partial \beta_{10}} = S_1 N_1 + S_2 N_2$$

$$2 \quad -\frac{1}{2} \frac{\partial L}{\partial \beta_{20}} = S_1 P_1 + S_2 P_2$$

$$3 \quad -\frac{1}{2} \frac{\partial L}{\partial \beta_{11}} = S_1 N_1^2 + S_2 N_2^2$$

$$4 \quad -\frac{1}{2} \frac{\partial L}{\partial \beta_{22}} = S_1 P_1^2 + S_2 P_2^2$$

$$5 \quad -\frac{1}{2} \frac{\partial L}{\partial \beta_{12}} = S_1 N_1 P_1 + S_2 N_2 P_2$$

$$6 \quad -\frac{1}{2} \frac{\partial L}{\partial \delta_{11}} = S_1 (\beta_{10} + 2\beta_{11} N_1 + \beta_{12} P_1)$$

$$7 \quad -\frac{1}{2} \frac{\partial L}{\partial \delta_{21}} = S_1 (\beta_{20} + \beta_{12} N_1 + 2\beta_{22} P_1)$$

$$8 \quad -\frac{1}{2} \frac{\partial L}{\partial \delta_{12}} = S_2 (\beta_{10} + 2\beta_{11} N_2 + \beta_{12} P_2)$$

$$9 \quad -\frac{1}{2} \frac{\partial L}{\partial \delta_{22}} = S_2 (\beta_{20} + \beta_{12} N_2 + 2\beta_{22} P_2) \quad .$$

Setting these equal to zero and replacing parameters by their estimated values we see there are some redundancies. For, adding the equations obtained from 6 and 8 and using the relations obtained from 1 and 2 we have

$$S_1 + S_2 = 0 \quad .$$

Similarly adding the equations obtained from 7 and 9 and again using the relations obtained from 1 and 2 we have:

$$S_1 + S_2 = 0 \quad .$$

Thus the 9 equations are not independent.

We can, however, obtain estimates of the quadratic coefficients and relative estimates of the  $\delta_{i\ell}$  and  $\beta_{i0}$  by the following procedure:

For each location we can fit the equation

$$Y_{\ell j} = \beta_{00\ell}^* + \beta_{10\ell}^* X_{1j} + \beta_{20\ell}^* X_{2j} + \beta_{11\ell}^* X_{1j}^2 + \beta_{22\ell}^* X_{2j}^2 + \beta_{12\ell}^* X_{1j} X_{2j} + \varepsilon_{\ell j}$$

by minimizing  $\sum_j \varepsilon_{\ell j}^2$ . We obtain

$b_{00\ell}^*$ ,  $b_{10\ell}^*$ ,  $b_{20\ell}^*$ ,  $b_{11\ell}^*$ ,  $b_{22\ell}^*$ , and  $b_{12\ell}^*$  from this procedure. Now the quadratic coefficients  $b_{11\ell}^*$ ,  $b_{22\ell}^*$ , and  $b_{12\ell}^*$  are not affected by the different  $\delta_{i\ell}$  at the various locations and are independent from location to location and possibly quite good estimates of  $\beta_{11}$ ,  $\beta_{22}$ , and  $\beta_{12}$  respectively. The linear coefficients  $b_{10\ell}^*$  and  $b_{20\ell}^*$  are estimates of functions of parameters and we indicate this using the previously derived expectations as follows:

$$b_{10\ell}^* = b_{10} + 2b_{11}d_{1\ell} + b_{12}d_{2\ell}$$

$$b_{20\ell}^* = b_{20} + b_{12}d_{1\ell} + 2b_{22}d_{2\ell} .$$

If the same experiment is performed at all locations then the overall regression equation using applied nutrient levels will give us  $\bar{b}_{00}^*$ ,  $\bar{b}_{10}^*$ ,  $\bar{b}_{20}^*$ ,  $\bar{b}_{11}^*$ ,  $\bar{b}_{22}^*$ , and  $\bar{b}_{12}^*$  where the bars denote average regression coefficients. Then we obtain

$$\bar{b}_{10}^* = b_{10} + 2b_{11}\bar{d}_1 + b_{12}\bar{d}_2$$

$$\bar{b}_{20}^* = b_{20} + b_{12}\bar{d}_1 + 2b_{22}\bar{d}_2$$

$$\bar{b}_{11} = b_{11}$$

$$\bar{b}_{22} = b_{22}$$

$$\bar{b}_{12} = b_{12}$$

as pooled estimating equations. Subtracting the average linear regression coefficients from the individual linear regression coefficients we have

$$b_{10\ell}^* - \bar{b}_{10}^* = 2\bar{b}_{11}(d_{1\ell} - \bar{d}_1) + \bar{b}_{12}(d_{2\ell} - \bar{d}_2)$$

$$b_{20\ell}^* - \bar{b}_{20}^* = \bar{b}_{12}(d_{1\ell} - \bar{d}_1) + 2\bar{b}_{22}(d_{2\ell} - \bar{d}_2)$$

using the average quadratic coefficients obtained from the overall regression. We can now solve for

$$d_{1\ell} - \bar{d}_1 \quad \text{and} \quad d_{2\ell} - \bar{d}_2$$

from these equations. Thus we have estimated to a first approximation the differences of the  $d_{i\ell}$  from their mean values.

The smallest  $d_{i\ell} - \bar{d}_i$ , (i.e., the most negative), will indicate the  $d_{i\ell}$  to be taken as zero for each nutrient, (thus each  $d_{i\ell}$  will be measured from the smallest namely  $d_{i0} = 0$ ) thus giving  $d_{i\ell}^!$  to add to  $X_{ij}$  as the first approximation. Since the estimates affect the estimation of  $b_{i0}$  we must reestimate our linear coefficients using the first approximations to improve our estimation of  $d_{i\ell}^!$  and  $b_{i0}$ . If we use the  $d_{i\ell}^!$  and hold the quadratic

coefficients constant for each location we will obtain additional corrections to the  $d_{i\ell}^1$  from our average pooled fit and our linear coefficients. That is we fit:

$$Z_{\ell j} = Y_{\ell j} - \bar{b}_{11}(X_{1j} + d_{1\ell}^1)^2 - \bar{b}_{22}(X_{2j} - d_{2\ell}^1)^2 - \bar{b}_{12}(X_{1j} - d_{1\ell}^1)(X_{2j} - d_{2\ell}^1) \quad \text{in the form:}$$

$$Z_{\ell j} = \gamma_{00\ell} + \gamma_{10\ell}X_{1j} + \gamma_{20\ell}X_{2j} + \eta_{\ell j}$$

estimating the  $\gamma$ 's individually by minimizing  $\sum \eta_{\ell j}^2$  for each location as well as the pooled equation for all locations. Then using the equations:

$$g_{10\ell} - \bar{g}_{10} = 2\bar{b}_{11}\Delta d_{1\ell}^1 + \bar{b}_{12}\Delta d_{2\ell}^1$$

$$g_{20\ell} - \bar{g}_{20} = \bar{b}_{12}\Delta d_{1\ell}^1 + 2\bar{b}_{22}\Delta d_{2\ell}^1$$

where the  $g_{i0\ell}$  are estimates of  $\gamma_{i0\ell}$ , we may calculate  $\Delta d_{1\ell}^1$  and  $\Delta d_{2\ell}^1$  and obtain

$$d_{1\ell}'' = d_{1\ell}^1 + \Delta d_{1\ell}^1$$

$$d_{2\ell}'' = d_{2\ell}^1 + \Delta d_{2\ell}^1 .$$

A second iteration using  $d_{i\ell}''$  will give convergence, i.e., the individual linear regression coefficients will be equal and we have then obtained our relative estimates of  $\delta_{i\ell}$ , relative in the sense that they are measured from the lowest estimated one. The  $d_{i\ell}''$  are then added to the  $X_{ij}$  for each location and the full quadratic fitted over all locations. This will then be the estimated structural equation with the iteration being completed in this final stage.

Thus the procedure can be summarized as follows: We use the overall general quadratic and the individual quadratics at the several locations to obtain initial approximations to our estimates of initial nutrient levels. By holding the quadratic coefficients constant the solution for the  $\delta_{i\ell}$  converges in two iterations because for constant quadratic coefficients we have a quadratic function of our  $\delta_{i\ell}$ . Then holding these estimated  $\delta_{i\ell}$  constant, our overall structural equation is fitted in one final step because for constant  $\delta_{i\ell}$  our equation is then linear in all coefficients. The estimated  $\delta_{i\ell}$  and  $\beta_{i0}$  are then relative in the sense that they are conditional on the lowest estimated  $\delta_{i\ell}$  being taken as zero. Thus our set of locations in the experiment should be as diverse as possible if we wish to obtain some location for each nutrient factor with a nearly zero initial level. Whether we achieve this or not, however, we still obtain comparisons between the locations as to differences in initial nutrient levels and can rank the locations with respect to each of the initial nutrient factors separately.

Since the objective of this paper has been to propose biologically more realistic models and consequently to alter where necessary existing techniques to achieve this end we decided to put the model to the "acid test" by trying it on some of the data from the N.C. State - T.V.A. project mentioned earlier. Accordingly we selected 6 locations out of the 60 possible with as diverse fitted forms and response

patterns as could be obtained. Hopefully, the coefficients will be biologically reasonable and the estimated  $\delta_{i\ell}$  will be reasonable and consistent with our formulations. In Table 5 are the data and treatment combinations in semicoded units for these locations. The experiment is 3 factor, N, P, K respectively.

The individual regressions using the semicoded levels indicated in Table 5 and the overall or average regression are displayed in Table 6.

Using the average regression quadratic coefficients as coefficients in the following set of linear equations and the differences of the individual linear coefficients from the average linear coefficients as right hand sides we have the 6 sets of equations:

$$2\bar{b}_{11}(d_{1\ell} - \bar{d}_1) + \bar{b}_{12}(d_{2\ell} - \bar{d}_2) + \bar{b}_{13}(d_{3\ell} - \bar{d}_3) = b_{10\ell}^* - \bar{b}_{10}^*$$

$$\bar{b}_{12}(d_{1\ell} - \bar{d}_1) + 2\bar{b}_{22}(d_{2\ell} - \bar{d}_2) + \bar{b}_{23}(d_{3\ell} - \bar{d}_3) = b_{20\ell}^* - \bar{b}_{20}^*$$

$$\bar{b}_{13}(d_{1\ell} - \bar{d}_1) + \bar{b}_{23}(d_{2\ell} - \bar{d}_2) + 2\bar{b}_{33}(d_{3\ell} - \bar{d}_3) = b_{30\ell}^* - \bar{b}_{30}^*$$

whose solution is in Table 7. (We have added the lowest to each of the other  $d_{i\ell} - \bar{d}_i$  and the estimated  $d_{i\ell}^!$  thus no longer sum to zero.) The units of the  $d_{i\ell}^!$  are those given respectively in Table 5.

Adding the  $d_{i\ell}^!$  given in Table 7 to the  $X_{ij}$  given in Table 5 and using the average quadratic coefficients of Table 6 we calculated the quadratic portion of the model and subtracted this from the observations. Using these

Table 5. Treatment combinations and yields for 6 selected locations in bushels/Acre from Baird (1958)

Treatment Combination N* P* K*	Location					
	1	2	3	4	5	6
1 1 1	42.2	23.1	25.4	128.4	50.2	66.3
1 1 3	43.0	19.9	31.3	140.6	57.2	68.0
1 3 1	35.7	37.4	39.0	141.9	55.8	67.0
1 3 3	38.0	41.1	45.9	136.7	58.9	70.0
3 1 1	40.4	31.8	30.1	128.3	54.5	63.5
3 1 3	34.5	29.5	38.4	135.1	57.7	74.9
3 3 1	34.0	38.5	58.0	136.8	58.5	77.1
3 3 3	32.2	45.2	43.5	129.7	60.9	79.5
2 2 2	43.2	32.9	37.1	143.7	59.1	77.9
0 2 2	9.9	20.8	23.9	126.9	37.4	41.9
4 2 2	49.7	36.4	48.1	132.6	59.0	78.8
2 0 2	41.9	2.1	10.6	126.7	41.1	64.9
2 4 2	36.8	55.7	40.3	142.5	58.3	72.0
2 2 0	39.9	34.6	11.2	136.1	49.0	73.8
2 2 4	38.9	41.6	40.3	137.1	57.6	64.0
0 4 4	13.4	25.2	32.0	134.9	37.3	32.2
4 0 4	31.5	2.1	13.0	130.9	32.6	62.4
4 4 0	38.4	45.5	15.3	137.9	46.5	83.4

\* 1 unit of N is 62.5 lbs/A

1 unit of P is 37.5 lbs/A

1 unit of K is 37.5 lbs/A

Table 6. Individual and average regressions for the data in Table 5

Parameter Estimated	Average Regression	Individual Regression					
		1	2	3	4	5	6
$\beta_{00}^*$	-40.04	-15.76	-35.29	-76.78	-43.67	-41.82	-26.93
$\beta_{10}^*$	15.42	16.72	10.91	17.97	15.40	17.42	14.11
$\beta_{20}^*$	14.07	.69	17.40	30.05	17.03	14.70	5.93
$\beta_{30}^*$	9.92	4.01	-1.02	26.01	12.53	12.17	5.79
$\beta_{11}^*$	-2.63	-2.93	-2.50	-1.74	-2.30	-3.14	-3.14
$\beta_{22}^*$	-2.05	.54	-2.43	-4.38	-1.09	-2.76	-1.11
$\beta_{33}^*$	-1.40	.53	.13	-4.30	.59	-1.86	-1.00
$\beta_{12}^*$	.05	.58	.86	-1.96	-1.78	.14	1.83
$\beta_{13}^*$	.65	-1.05	.60	-1.39	-1.61	.95	.48
$\beta_{23}^*$	.87	.14	.45	.66	-3.08	.39	-1.40

Table 7. First approximation to  $\delta_{i\ell}$  for the 6 locations of Table 5

Location	$d_{i\ell}^1$	$d_{2\ell}^2$	$d_{3\ell}^3$
1	0.00	6.22	6.05
2	.76	1.11	9.25
3	.58	0.00	0.00
4	.53	2.30	4.10
5	.13	2.87	4.15
6	.55	4.65	5.78

differences and 6 linear models in total nutrient form we fit 7 models, 1 overall, 6 individually to locations 1-6. This overall linear regression when subtracted from the individual linear regressions can be used to calculate the  $\Delta d_{i\ell}^1$ , (i.e., the corrections to the  $d_{i\ell}^1$ ). Thus:

$$2\bar{b}_{11}\Delta d_{i\ell}^1 + \bar{b}_{12}\Delta d_{2\ell}^2 + \bar{b}_{13}\Delta d_{3\ell}^3 = \bar{g}_{10\ell} - \bar{g}_{10}$$

$$\bar{b}_{12}\Delta d_{i\ell}^1 + 2\bar{b}_{22}\Delta d_{2\ell}^2 + \bar{b}_{23}\Delta d_{3\ell}^3 = \bar{g}_{20\ell} - \bar{g}_{20}$$

$$\bar{b}_{13}\Delta d_{i\ell}^1 + \bar{b}_{23}\Delta d_{2\ell}^2 + 2\bar{b}_{33}\Delta d_{3\ell}^3 = \bar{g}_{30\ell} - \bar{g}_{30}$$

After performing this calculation we added  $\Delta d_{i\ell}^1$  and  $d_{i\ell}^1$  to obtain  $d_{i\ell}''$  and adjusted all figures so that the lowest was zero. The results are given in Table 8.

Table 8. The estimated  $\delta_{i\ell}$  for the 6 locations of Table 5

Location	$d''_{1\ell}$	$d''_{2\ell}$	$d''_{3\ell}$
1	.730	2.758	1.578
2	.809	0.000	2.048
3	1.254	1.320	0.000
4	1.403	1.876	.860
5	1.114	1.890	1.103
6	0.000	1.927	2.296

The estimated  $\delta_{i\ell}$  from Table 8 were added to the  $X_{ij}$  of Table 5 and, using the average quadratic coefficients of Table 6 the quadratic portion was subtracted from the observations as before and individual linear functions were again fitted to investigate the two-step conversion noted earlier. All linear coefficients agreed to 6 significant figures which was the limit of accuracy of the calculations. Thus the estimated  $\delta_{i\ell}$  in Table 8 should provide a diagnosis of the 6 locations, at least in comparison to each other. Looking at Table 8, then, we see that location 1 had next to the lowest initial nitrogen level, a very high initial phosphorus level, and an intermediate amount of potassium. The data in Table 5 bear out this

diagnosis with low yields at zero applied nitrogen and some indications of depression for phosphorus application. The low nitrogen seems to be definitely the limiting factor.

Location 2 is by Table 8 very deficient in phosphorus, an observation also consistent with the data of Table 5 where a yield of only 2.1 bushes/Acre was realized on zero phosphorus treatments. An indication of deficient nitrogen can also be noticed.

Location 3 is the low potassium location. The data of Table 5 seem to justify this diagnosis also. Location 4 was the high yielding location and is estimated to have the most nitrogen, (the most effective nutrient in the experiment). It has an estimated 88 lbs/Acre more than the lowest location. Location 5 is an intermediate type of location that might be much better with just a little more nitrogen. Location 6 is estimated as seriously lacking in nitrogen and we note from Table 5 the tremendous response resulting when even a little nitrogen is added. Thus the 6 locations do seem to give reasonable estimates of  $\delta_i$  at least as judged by the yield response.

For all locations in the experiment soil tests were made on each plot. The averages of the various soil tests for these locations are given in Table 9.

Table 9. Average soil tests for the 6 locations of Table 5

Location	Soil Test						
	pH	Organic Matter %	P ppm	K me/100 g.	Ca me/100 g.	Mg me/100 g.	Cation Ex. Cap. me/100 g.
1	5.2	1.2	32	.201	.56	.14	2.4
2	4.8	16.0	2	.108	2.48	.22	38.2
3	5.4	5.2	1	.047	1.10	.52	9.9
4	5.3	8.2	82	.179	6.19	.94	13.2
5	5.1	1.4	18	.105	.43	.33	2.5
6	5.4	1.7	37	.169	.73	.38	3.6

There are several interesting comparisons to be made from Table 8 and Table 9. The contention that organic matter is a poor measure of nitrogen seems to be borne out. The two low locations on phosphorus turn out as in our analysis to be numbers 2 and 3. The low potassium location is 3 as in our analysis. Hazardous much beyond these observations is more properly the province of an agronomist but it seems that possibly a study relating the predicted initial nutrient levels to soil test measurements might be a promising line of endeavor. Perhaps the predictive power of soil test results could be improved thereby.

We have seen thus far that the postulated  $\delta_{ij}$  and their mode of entering the response function seem to have passed the test of biological interpretability and reasonableness. The final step in the example is the calculation of the structural equation. To do this we add the estimated  $\delta_{ij}$  to the  $X_{ij}$  for each location separately and then calculate, the usual regression coefficients, which will be those of the total nutrient model formulation using all locations. For our example we get the estimated equation to be:

$$\begin{aligned} \hat{Y} = & - 105.22 + 21.61N + 41.97P + 28.68K \\ & - 1.607N^2 - 4.046P^2 - 2.669K^2 \\ & - .499NP - 1.265NK - 1.442PK \end{aligned}$$

The linear coefficients are positive and of about the expected size while the quadratics are small and negative. The constant term is estimated as negative and all in all as far as can be judged the function looks biologically reasonable. Thus the multiple factor variable nutrient model gives the economist a way of combining data from several locations reasonably and the agronomist a diagnostic tool to investigate initial nutrient levels solely from yield data which possibly will throw new light on the calibration of soil test information. The units of the predicted initial levels are in the units of his corrective measures, i.e., applied fertilizer applications.

## CONCLUSIONS AND SUMMARY

In this paper we have considered alterations to the usual response surface techniques so that this useful formulation can have a wider and more proper application in the field of quantitative fertilizer factor experiments. We have succeeded in formulating an alteration in the usual model which will give the agronomist direct biological assay of levels of inherent nutritive material in the soil. This should prove very useful. With this altered formulation we have also provided statistical techniques to estimate these new parameters and assess their reliabilities, in the single nutrient case as well as perform certain significance tests. Finally, a reexamination of the probabilistic structure suggests many new lines of statistical research into more realistic introductions of probability into biological problems.

## LIST OF REFERENCES

- Anderson, R. L., 1956. Comparison of discrete and continuous models in agricultural production analysis. Chapter 3 of Methodological Procedures in the Economic Analysis of Fertilizer Use Data, Iowa State College Press, Ames, Iowa.
- Anderson, R. L., 1957. Some statistical problems in the analysis of fertilizer response data. Chapter 17 in Fertilizer Innovations and Resource Use, Iowa State College Press, Ames, Iowa.
- Baird, B. L., 1958. Factors affecting the nutrient content and yield response of corn to nitrogen, phosphorus, and potassium fertilization. Ph.D. Thesis, N.C. State College, (University Microfilms, Ann Arbor).
- Bartlett, M. S., 1949. Fitting a straight line when both variables are subject to error. *Biometrics* 5:207-212.
- Berkson, J., 1950. Are there two regressions? *J.A.S.A.* 45:164-180.
- Box, G. E. P., 1954. The exploration and exploitation of response surfaces: some general considerations and examples. *Biometrics* 10:16-60.
- Box, G. E. P., and Wilson, K. B., 1951. On the experimental attainment of optimum conditions. *Jour. Roy. Stat. Soc. Series B* 13:1-45.
- Brown, W. G., Harward, M. E., Petersen, R. G., and Ko, W.C. 1962. A method for incorporating soil test measurements into fertilizer response functions. Submitted to *Agronomy Journal*.
- Davies, O. L., 1954. Design and Analysis of Industrial Experiments. Oliver and Boyd, Ltd., London.
- Fisher, R. A., 1950. Contributions to Mathematical Statistics. Wiley, New York.
- Geary, R. C., 1953. Non-linear functional relationship between two variables when one variable is controlled. *J.A.S.A.* 48:94-103.
- Graybill, F. A., 1961. An Introduction to Linear Statistical Models. McGraw-Hill, New York.

- Heady, E. O., 1961. Status and methods of research in economic and agronomic aspects of fertilizer response and use. National Academy of Science - National Research Council publication 918.
- Heady, E. O., Doll, J., and Pesek, J. T., 1956. Alternative production functions and isoclines. (Typed manuscript.) Iowa State College, Ames, Iowa.
- 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.
- Hildreth, C. G., 1954. Point estimates of ordinate concave functions. J.A.S.A. 49:598-619.
- Hildreth, C. G., 1957. Possible models for agronomic-economic research. Chapter 16 of Fertilizer Innovations and Resource Use, Iowa State College Press, Ames, Iowa.
- Hurst, D. C., and Mason, D. D., 1957. Some statistical aspects of the TVA-North Carolina cooperative project on determination of yield response surfaces for corn. Chapter 18 of Fertilizer Innovations and Resource Use, Iowa State College Press, Ames, Iowa.
- Keeping, E. S., 1956. Note on Wald's method of fitting a straight line when both variables are subject to error. Biometrics 12:445-448.
- Kendall, M. G., 1952. Regression, structure and functional relationship. Biometrika parts I and II 39:96-108.
- Koopmans, T. C., and Reiersol, O., 1956. The identification of structural characteristics. Ann. Math. Stat. 27:165-181.
- Liebig, J. V., 1855. Die grundsätze der agriculturchemie mit riicksicht aufdie in England angestellten untersuchungen. Frederick Viewig und sohn, Braunschweig.
- Lindley, D. V., 1953. Estimation of a functional relationship. Biometrika 40:47-49.
- Mason, D. D., 1956. Functional models and experimental designs for characterizing response curves and surfaces. Chapter 5 of Methodological Procedures in the Economic Analysis of Fertilizer Use Data, Iowa State College Press, Ames, Iowa.

- Mitscherlich, E. A., 1909. Das gesetz des minimums und das gesetz des abnehmenden bodenertrages. Landw. Jahrb. 38:537-552.
- Quenouille, M. H., 1952. Associated Measurements. Butterworths Scientific Publications, London.
- Scarborough, J. B., 1958. Numerical Mathematical Analysis. 4th Edition, The Johns Hopkins Press, Baltimore.
- Scott, E. L., 1950. Note on consistent estimates of the linear structural relation between two variables. Ann. Math. Stat. 21:284-288.
- Spillman, W. J., 1933. Use of the exponential yield curve in fertilizer experiments. Tech. Bull. 348, U.S.D.A.
- Stemberger, A., 1956. Economic implications of using various production functions to express nitrogen-corn production relationships. Unpublished doctoral thesis, North Carolina State College Library.
- Wald, A., 1941. The fitting of straight lines if both variables are subject to error. Ann. Math. Stat. 11:284-300.
- Yniguez, A. D., 1957. Investigation of alternative models for utilization of measures of residual nutrients in soil fertility field experiments. Unpublished M.S. thesis, North Carolina State College Library.
- Zyskind, G., and Kempthorne, O., 1960. Treatment errors in comparative experiments. WADC Technical note 59-19.