STATISTICAL MODELS FOR ANALYSIS OF DOSE-RESPONSE DATA

Karan P. Singh

Recommended Citation

STATISTICAL MODELS FOR ANALYSIS OF DOSE-RESPONSE DATA

Karan P. Singh*
Department of Mathematics
Central Michigan University
Mt. Pleasant, MI 48859

ABSTRACT

In this paper we propose three families of functional models for analysis of dose-response data. The first family is for modeling data which have a steep sloping line for an ascending portion of the response curve and a plateau representing maximum response or a sloping line representing little response at higher application levels. The second family is for modeling data which represent a steep sloping line on the ascending portion of the response curve and a declining curvature for declining response at higher application levels. The third family is for fitting data which show an initial plateau followed by increased or decreased response, and finally a plateau representing maximum response. One of the advantages of using these families for modeling the dose-response data is that the join points of the response curves are not considered parameters to be estimated, nor their estimates considered random variables in the estimation process. The uses of the families are illustrated using them in fitting the fertilizer response data of single nutrient experiments. The problem of modeling the multi-nutrient response data is addressed and recently developed methods are briefly discussed.

Key Words and Phrases: Families of functional models, Linear regression; Modified quadratic models; Linear-plateau models; Threshold; Plateau-linear-plateau models; Bias in optimal rates; Least squares; Isotonic regression.

I. INTRODUCTION

The basic purpose of evaluating fertilizer response is to provide information on the nutrient status of soil and to predict the response to added nutrient. Also, to obtain maximum yield with economy, it is essential to make a very judicious and efficient use of chemical fertilizer, the absence of which may not only result in wastage of fertilizer but also may have a detrimental effect upon the environment. Furthermore, with the spiraling price of fertilizer-nutrients and procedures which tend to produce upward biases in estimating optimal doses, such doses of chemical fertilizers cannot be recommended, particularly to farmers in developing countries who have great difficulty in obtaining capital to purchase them. In practice, problems are encountered in attempting to apply complex fertilizer response models to existing data (see Waugh, et al., 1983; Anderson and Nelson, 1975; Singh and Singh, 1981, 1982). One such problem is difficulty in interpretation at high rates of application since continuous curvilinear response models are very sensitive to fluctuations in yield at these rates. The most important curvilinear models given in the literature are quadratic, square root, modified quadratic, logarithmic functions, linear-plateau and

---

inverse quadratic.

By reviewing the models proposed by Anderson and Nelson (1975, 1984), Singh (1979), Singh (1981, 1982) and others, three families of functional models for modeling the single nutrient experimental data are proposed. These families for agricultural experiments are of great practical importance since they treat problems which we have with the usual quadratic models. The families are fitted to the real data sets. Also, recently developed methods to deal with the multi-nutrient response data are briefly discussed.

2. THE FAMILY OF MODIFIED QUADRATIC MODELS

Singh and Singh (1981) proposed modified quadratic response models for modeling data having sharp rise at initial increment applications and declining yields at higher rates of application. Prediction is usually the primary purpose for building response models. In developing nations like India, farmers have great difficulty in obtaining capital to purchase fertilizers and do not use the optimal recommended doses on a large scale, so it is initial to predict future responses at initial levels of application of fertilizers with higher precision. For such situations, it seems that the increasing phase of the response pattern is more important than the declining phase. The family of modified quadratic models is probably more appropriate models to describe and interpret fertility response in field trials. Furthermore, the variability in yield observations obtained at higher rates of nutrient application is large and other continuous curvilinear response models are very sensitive to fluctu-ations in yield at higher rates. Anderson and Nelson (1975) suggested the use of linear-plateau models discarding data showing declining yields. Discarding such data does not seem to be the solution and by doing this, use of a linear-plateau model will yield an underestimate of optimal dosage. Singh and Singh (1982) discussed modified quadratic (MQ) models at length and showed their efficiency over the quadratic (Q) models. A typical graph of model (2.1) is given in Figure 1 (also see Singh and Singh, 1982). Suppose there are k equally spaced levels of nutrient to which the line with positive slope is fitted. Let values of the levels be denoted by a, a+d,...,a+(k-1)d. Then the family of modified quadratic models is given by

$$Y = \beta_0 X_0 + \beta_1 X_1 - \beta_2 X_2^2 + e \qquad\qquad (2.1)$$

where: Y is the average yield; β_0 is the intercept; β_1 is the linear component parameter; β_2 is the quadratic component parameter; and e is the experimental error; X_0, X_1, and X_2 are nutrient variates ($X_0 = 1$, X_1 and X_2 are called the design points). X_1 takes values 0, 1, 2, ... , k-1 and X_2 takes values zero up to which responses are steep (sharp rise), say t is the nutrient level number up to which responses are steep, and X_2 takes values 1, 2, ... , k-t beyond sharp rise. Since we are dealing with the region where the yield starts to decline, we need to have the negative sign in front of β_2.

The prediction equation is

$$\hat{Y} = b_0 X_0 + b_1 X_1 - b_2 X_2^2 \qquad\qquad (2.2)$$

where b_i is the experimental estimate of $\beta_i (i = 0, 1, 2)$. The absolute optimal value of nutrient applied, i.e. the value of X for which $\hat{Y}$ in (2.2) will be maximized, is given by

$$X_{opt} = (t + \frac{b_1}{2b_2}) \, d \qquad\qquad (2.3)$$

where d is the interval between two adjoining levels of applications and t is the nutrient level number up to which responses give sharp rise or t is the X-value of the point of start of plateau. The economic optimal value, i.e. the value of X for which the profit is maximum, is given by

$$X_{eco} = (t + \frac{b_1 - r}{2b_2}) \, d \qquad\qquad (2.4)$$

where r is the cost of d units of nutrient divided by the price of a unit of crop produced. The expressions (2.3) and (2.4) are derived by using the rules for the addition of amount (X) of nutrient for the linear plateau models given by Anderson and Nelson (1975, 1984) and Singh and Singh (1982) and the expressions for the absolute and economic optimal values derived from the quadratic model. The choice of t can be made by inspection of the scatter diagram and varying t by several steps in both directions from the position suggested by the scatter diagram. Selection of the final model in terms of t is based on minimum mean square error for the model. The given examples in section 6 will be very helpful to the reader.

3. THE FAMILIES OF LINEAR-PLATEAU AND PLATEAU-LINEAR-PLATEAU MODELS

We now consider cases where there is a marked response to first increment application followed by little and/or no response at higher application rates. For this category Anderson and Nelson (1975) proposed a family of linear-plateau models. The name linear-plateau implies a region of linear response and a plateau. Singh (1979) supplemented these results and worked out the design matrices for all possible sub-models for experiments having four to seven equi-spaced levels of applications. He also proposed an additional mode (VIII) for the situations where there are three intersecting straight lines. Singh and Singh (1981, 1982) discussed some of these models at length and gave design matrices. The linear-plateau models are given by

$$Y = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e \qquad\qquad (3.1)$$

where Y is the mean yield; β_1 and β_2 are the slopes of the sloping lines; β_3 is the slope of the third sloping line or the distance of the plateau above the value of the yield at the last design point before the start of the plateau for sub-models exhibiting plateau; β_0 and e are as before. As before $X_0 = 1$ at every coded level of the nutrient and X_1, X_2, X_3 are the design points (nutrient variates) respectively, i.e. the value of the coded X at which two sloping lines are intersecting. The prediction equation is given by

$$\hat{Y} = b_0 X_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 \qquad (3.2)$$

where b_i is the experimental estimate of B_i ($i = 0,1,2,3$). As explained in section 2, we are more interested in the increasing phase, it is therefore required in general that $b_1 > b_2 > b_3 > 0$; if b_2 or b_3 is negative, the family of MQ response models discussed in Section 2 are suggested. For selecting the most appropriate models or sub-models the scatter diagram may be inspected and likely models or sub-models may be fitted using a least squares method. The model yielding the minimum mean square error among all competitors may be selected as the working model for the given data set. However, Anderson and Nelson (1984) proposed an alternative procedure based on "isotonic regression" (see Barlow et al., 1972). The procedure is to compute a successive moving average (MA), starting with the last observation; the plateau terminates just before MA begins to decrease monotonically. This procedure will determine the number of observations (m) on a plateau. Thus, those models for which there are m observations on the plateau are the models to be used and this way we can choose the best model based on minimum mean square error for a given data set. A general strategy for choosing the appropriate model from the family of linear-plateau models is given in Anderson and Nelson (1984) and in Singh and Singh (1982) and in order to economize the space we do not reproduce here. In this section we now discuss a family of plateau-linear-plateau models.

There are some biological response situations where a plateau occurs within the range of the lowest treatment levels as well as the highest treatment levels. Anderson and Nelson (1984) proposed models called plateau-linear-plateau (PLP) models for such situations. The intersection point of the first plateau and a sloping line is a threshold input. The predicted linear regression line for a PLP model is given by

$$\hat{Y} = b_0^* + b_1^* X_1^* + b_2^* X_2^* + b_3^* X_3^* \qquad (3.3)$$

Here, the PLP models can easily be constructed from the LP models by a special coding of the original data. Also, the design-matrices worked out by Singh (1979) and given in Singh and singh (1982) may be used to obtain the design-matrices for a complete family of sub-models for experiments having four to seven levels of application of nutrient. X*'s are used to distinguish the PLP models from the LP models. However, the X*'s will be the same as the X's for an appropriate LP model. It is assumed that $b_2^* > b_3^* \geq 0$. However, b_1^* may be less than b_2^* ($b_1^* > 0$).

The procedures for the determination of economic optimum differ from model to model and from sub-model to sub-model. Since the number of models and sub-models is large, it is considered to be advisable to avoid the reproduction of the procedures for the determination of economic optimum. The reader is referred to Anderson and Nelson (1975, 1984) and Singh and Singh (1982) for the procedures.

4. GRAPHICAL PROCEDURE FOR MULTI-NUTRIENT EXPERIMENTS

Singh and Singh (1981, 1982) illustrated using real data sets that the use of a quadratic model yields upwards bias in economic does of nutrients. They suggested an alternative procedure for determination of the optimal combination of nutrients outlined below (also see Anderson and Nelson, 1975).

For experiments in which the (two or more) nutrients are varied factorially, the LP and PLP models briefly discussed in Section 3, for a single nutrient, may be applied individually to means if the preliminary analysis of variance shows no statistically significant two-factor interactions and the choice of which model to use for means of a nutrient depends upon the response pattern for means. In the case of some statistically significant two-factor interactions from the analysis of variance, it is necessary to make estimates of the optimal rates on each nutrient at each level of the other nutrient assuming only two nutrients, a procedure which is not too efficient. The optimal values would then be connected by contours connecting the set of optima for a nutrient at each level of the other nutrient for the two nutrients. The points where the two contours intersect is taken for the optimal combination. (See Figure 4).

5. INVERSE QUADRATIC MODELS FOR MULTI-NUTRIENT EXPERIMENTS

In this section an idea of using inverse quadratic models for analysis of multi-nutrient experiments is explored. This discussion is based on the work being done by Singh and Chaudhary (1990). The relationship between crop yield and response to nutrients in compound fertilizers seems to be a mixture of parabolic and asymptotic functions. The quadratic models are symmetric about their stationary values and this stationary constraint is overcome by modifying the quadratic to square root models. An inverse quadratic is an intermediate type, which can have behavior depending upon parameter values and also has optimal values. With the application of nitrogen-phosphorus-potassium (NPK) fertilizer, it can be assumed that the crop yield includes both asymptotic and parabolic responses. Rationale of such an assumption is contained by the analysis of trends of response to nutrient and is shown in Figure 2. With sufficiently high levels of fertilizer application, generally it is observed that there are parabolic trends for the response to the nutrient N, and asymptotic form for the two responses to the nutrients P and K.

The inverse quadratic model for n nutrients can be written as

$$(\prod_{i=1}^{n} X_i)/Y = \text{polynomial in } (X_1, X_2, \ldots, X_n) \qquad (5.1)$$

where Y is the average response and X_1, X_2, ... , X_n are the nutrient factors. The form (5.1) will have parameters indicating main effects and iterations. Generally Y tends to an asymptotic value if either of X_i increases and others are held constants and/or if they are increased so the ratios being held constant. Similarly, Y can be approximated by retaining first degree terms

for small X_i corresponding to the factor. This work has just begun and a complete mathematical treatment will appear elsewhere.

6. ILLUSTRATIONS

The average yield data for different crops with different locations in India are taken as examples to compare the relative performance of the proposed family of modified quadratic (MQ) models with the quadratic models. These data were collected by Singh (1979) when he extended the results of Anderson and Nelson (1975). Most of the discussion here has resulted from the findings of Singh (1979) and Singh and Singh (1981, 1982). The data are related to the situations where there is a marked response to the initial increment applications followed by little and/or no response and/or subsequently declining yields at higher rates application. The variability in yield observations obtained at high rates of nutrient application is considerably high and continuous curvilinear response models, especially parabolic curves, are not flexible enough to accommodate the sharp initial rise and subsequent flattening. The application of modified quadratic (MQ) models is first illustrated using the two data sets given in Table 1.

A quadratic (Q) prediction function $\hat{Y} = b_0 + b_1 X - b_2 X^2$ is also fitted to the above data sets. The estimates of β_i's along with their standard errors (in parentheses) are given in Table 2. The models are fitted by the method of least squares. The design matrix for the MQ model for location 1 is taken as:

$$X = \begin{array}{ccc} X_0 & X_1 & X_2^2 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 1 \\ 1 & 4 & 4 \end{array}$$

Similarly, the design matrix for the MQ model for location 2 can easily be calculated. In Table 2, the absolute optimal dose (N_{opt}), the economic optimal dose (N_{eco}), mean square error (MSE) and the coefficient of determination (R^2) are also given. R^2 is quite high in each case and it does not provide an exact picture of the adequacy of the models. Limitation of R^2 in the present case is that for each residual sum of squares, R^2 increases with the steepness of the regression surface. This also suggests that it is more appropriate to stress more on the standard errors of the estimated regression coefficients. Inferences based on them give a better insight about the adequacy of the model. From Figure 3 it is also clear that the MQ models do a better job in prediction than the Q models. It may be pointed out that the quadratic models happened to give more realistic results over other models like square root models for the type of response exhibited by the present experiments.

The example given in Table 3 is on nitrogen-dwarf wheat in split-plot design (Singh, 1979; Singh and Singh, 1981) and illustrates the use of a linear-plateau model. Two prediction functions of the Q and LP (VI_{14}) models are fitted by the usual method and the estimates of β_i's along with their

standard errors (in parentheses) are given in Table 4. The design matrix for the linear plateau model (VI_{14}) is given by

$$
X = \begin{array}{ccc}
X_0 & X_1 & X_2 \\
1 & 0 & 0 \\
1 & 1 & 0 \\
1 & 1 & 1 \\
1 & 1 & 2 \\
1 & 1 & 3 \\
1 & 1 & 3
\end{array}
$$

Using the usual Q model the optimal values are unrealistically high and are biased upward as the yield has already stabilized at $N = 160$. The use of the linear plateau model (VI_{14}) gives very realistic values of optima. It may be added here that the performance of a square root model though not reported here was observed to be considerably inferior to even the Q model. Singh (1979) and Singh and Singh (1982) showed uniformly better performance of linear plateau models over other models for many other sets of measured data exhibiting plateau at high rates of application. The author has not been able to obtain a data set which may be used to illustrate the use of PLP models.

The uses of the graphical approach for the multi-nutrient experiments discussed in Section 4 is illustrated using the two data sets given in Table 5. The first experiment was conducted in randomized block design and the second experiment in split-plot design. More details are given in Singh (1979). The canonical analysis indicates that the second order surface fitted to the first data set cannot give the optimal values since the eigenvalues do not happen to be negative and thus, the requirement for the stationary point is not satisfied. The economic optimal combination using the graphical technique is $(N_{eco}, P_{eco}) = (75, 50)$. For the second experiment, the eigenvalues turn out to be negative, thus using the second order surface $(N_{eco}, P_{eco}) = (25, 96)$ whereas the graphical method gives $(N_{eco}, P_{eco}) = (22, 43)$. For both experiments, the economic optimal combinations determined by graphical method based on linear plateau techniques seem to be realistic and show the usefulness of the graphical procedure. An illustration of the graphical approach for calculating economic optimum combination from the data of Experiment 1 is given in Figure 4.

7. SUMMARY

In practice, a number of problems are encountered in attempting to apply complex fertilizer response models to existing data. Several of these are: (i) lack of sufficient information to apply the fertilizer response model correctly, (ii) bias in the interpretation of the model, (iii) variability in yield observations obtained at high rates of nutrient application, (iv) difficulty in practical analysis if the computers are not available. Thus, in order to overcome some of the aforesaid difficulties, three families of

functional models for analysis of the single nutrient experimental data have been proposed. Also, two methods for analysis of the multi-nutrient experimental data have briefly been discussed. The families proposed here seem to enjoy an edge over other existing response models for describing the dose-yield relationship and prediction of responses at initial level of application of nutrient in this paper. It is expected that lack of fit will play a significant role in selecting a family and thus, selecting a model within the family. It is being under investigation. We end our discussion by remarking that the analysis of the proposed families possesses the following requirements: (i) leads to reasonable estimates of optimal fertilizer rates for various decision rules, (ii) easily adaptable to obtain results based on the average of a number of experiments, (iii) amenable to easy calculation, (iv) permits interpretation and recommendation on the spot, (v) produces a satisfactory goodness of fit to the data (Anderson and Nelson, 1975, 1984), (vi) the joint points are not considered parameters to be estimated, nor their estimates considered random variables. The requirement (vi) makes the estimation process simpler with respect to the joint points. We hope that the families and the methods discussed in this paper will be useful in modeling the data from other areas.

ACKNOWLEDGEMENT

REFERENCES

Anderson, R. L. and Nelson, L. A. (1975). A family of models involving intersecting straight lines and concomitant experimental designs used in evaluating response to fertilizer nutrients. Biometrics, Vol. 31, 303-318.

Anderson, R. L. and Nelson, L. A. (1984). Recent developments in the use of linear-plateau models to estimate response relationship. Proceedings 50th Anniversary Conference, Iowa State Statistical Laboratory, H. A. David and H. T. David (eds.). The Iowa State University Press.

Barlow, R. E., Bartholomew, D. J., Bremmer, J. M., and Brunk, H. D. (1972). Statistical Inferences Under Order Restrictions, Wiley, New York.

Singh, K. P. (1979). Evaluating the fertilizer response. Unpublished M. Sc. thesis, Haryana Agricultural University, Hisar, India.

Singh, K. P. and Singh, U. (1981). A statistical study of fertilizer response data. Haryana Agricultural University Journal of Research, Vol. 10, 3, 323-332.

Singh, U. and Singh, K. P. (1982).   Improved models for fertilizer nutrient
        experiments. American Journal of Mathematical and Management Sciences,
        Vol. 2, 2, 99-121.

Singh, K. P. and Chaudhary, F. S. (1990). Multi-nutrient response analysis with
        inverse quadratic models. Unpublished work.

Waugh, D. L. Cate, R. B. Jr., and Nelson, L. A. (1973). Discontinuous models
        for rapid correlation, interpretation, and utilization of soil analysis
        and fertilizer response data. Technical Bulletin No. 7, NCSU, Raleigh,
        North Carolina.

TABLE 1: Average Yields of Different Crops (quintal/hectare)

Location	Crop	Average Yield						$S_{\bar{y}}$	r
DOSES (N/kg)		0	50	100	150	200			
Hyderabad (1)	Sorghum	34.1	42.8	57.4	71.4	67.5		3.56	0.73
DOSES (N/kg)		0	40	80	120	160	200		
Ludhiana (2)	Wheat	9.25	18.30	30.35	42.40	42.25	40.35	1.50	1.20

$S_{\bar{y}}$: Standard error of mean yield, r: the cost of d units of nutrient divided by the price of a unit of crop produced. A quintal is 1,000 kilograms and a hectare is 10,000 square meters.

TABLE 2: Estimates of Regression Coefficients, Standard Errors (in Parentheses) of Estimates, Mean Square Error (MSE), Coefficient of Determination (R^2), Absolute Optimal Value (N_{opt}) and Economic Optimal Value (N_{eco})

Location	Model	b_0	b_1	b_2	MSE	R^2	N_{opt}	N_{eco}
1	MQ	31.94 (3.31)	13.61 (1.97)	4.53 (1.80)	12.94	.998	175	168
	Q	31.87 (3.35)	16.91 (3.96)	1.84 (.95)	29.16	.996	229	214
2	MQ	8.08 (1.31)	11.69 (0.72)	2.92 (.37)	3.43	.998	159	151
	Q	7.22 (1.41)	16.09 (1.33)	1.85 (.26)	10.26	.995	174	161

TABLE 3: Average Yields of Dwarf Wheat (quintal/hectare)

N(kg/ha)	0	40	80	120	160	200	$S_{\bar{y}}$	r
Yield (q/ha)	21.54	31.05	35.88	45.25	49.50	49.84	1.912	1.20

TABLE 4: Estimates of Regression Coefficients, Standard Errors (in Parentheses) of Estimates, mean Square Error (MSE), Coefficient of Determination (R^2), Absolute Optimal Value (N_{opt}) and Economic Optimal Value (N_{eco})

Model	b_0	b_1	b_2	N_{opt}	N_{eco}	MSE	R^2
LP	21.54 (1.912)	9.21 (2.475)	6.42 (0.328)	160	160	0.72	.999
Q	21.25 (1.733)	10.19 (1.630)	0.86 (.313)	237	209	4.90	.999

TABLE 5: Average Yields (quintal/hectare) of Sunflower and Groundnut

Crop	Sunflower: Experiment 1				Groundnut: Experiment 2				
N(kg/ha)	P(kg/ha)				N(kg/ha)	P(kg/ha)			
	0	25	50	75		0	20	40	60
0	9.73	11.05	12.60	15.62	0	14.40	14.70	16.40	16.50
25	10.00	11.92	13.92	16.87	20	16.80	17.10	17.90	18.30
50	10.37	12.70	14.30	18.57	40	16.80	17.50	18.20	18.20
75	10.72	12.60	14.90	18.80					
*	$r_1 = .215$, $r_2 = .208$					$r_1 = .340$, $r_2 = .332$			

r_1: the cost of d units of N divided by the price of a unit of crop produced.

r_2: the cost of d units of P divided by the price of a unit of crop produced.
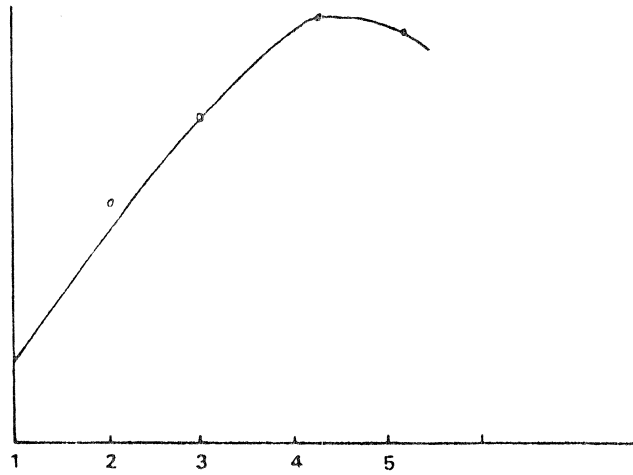
Figure 1: Typical Graph of Modified Quadratic Model.
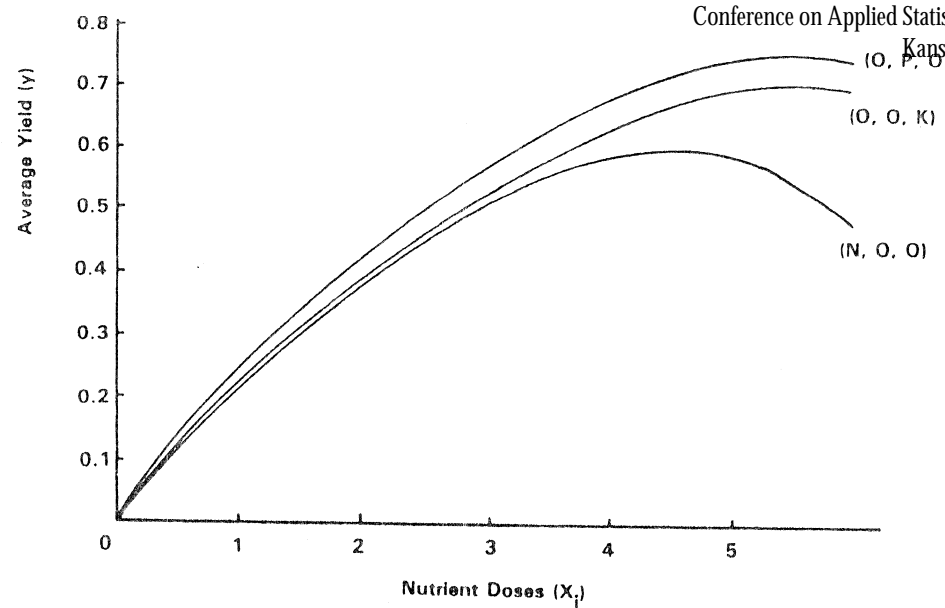

Figure 2: Behavior of Response Function For Multi-nutrient Doses.
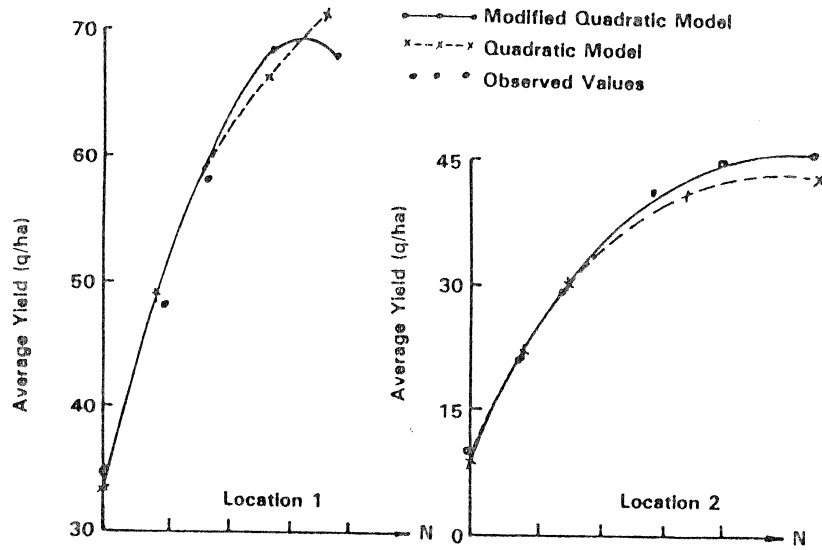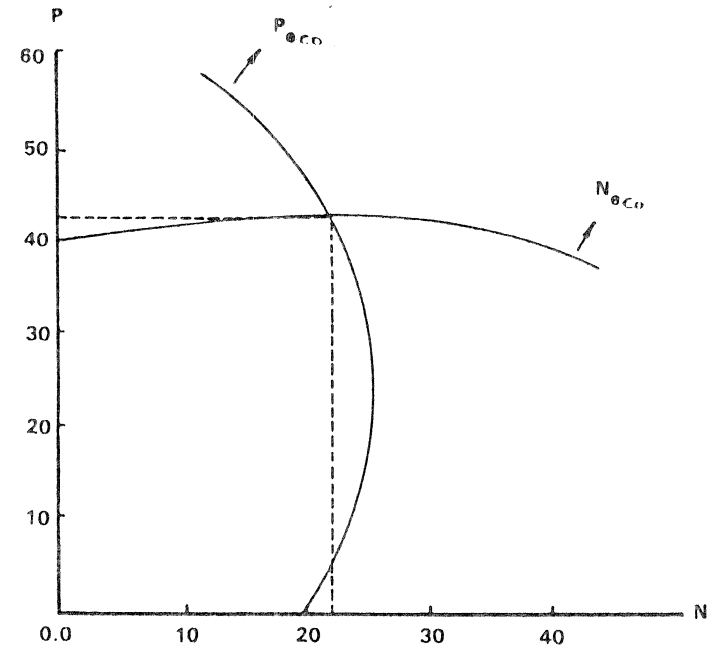

Figure 3: A Comparative Study of Quadratic and Modified Quadratic Models.

Figure 4: Economic Optimum Combination.