

Kansas State University Libraries

New Prairie Press

---

Conference on Applied Statistics in Agriculture

1999 - 11th Annual Conference Proceedings

---

## NONLINEAR MODELS FOR MULTI-FACTOR PLANT NUTRITION EXPERIMENTS

R. D. Landes

W. W. Stroup

E. T. Paparozzi

M. E. Conley

*See next page for additional authors*

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

---

### Recommended Citation

Landes, R. D.; Stroup, W. W.; Paparozzi, E. T.; and Conley, M. E. (1999). "NONLINEAR MODELS FOR MULTI-FACTOR PLANT NUTRITION EXPERIMENTS," *Conference on Applied Statistics in Agriculture*.  
<https://doi.org/10.4148/2475-7772.1263>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact [cads@k-state.edu](mailto:cads@k-state.edu).

---

**Author Information**

R. D. Landes, W. W. Stroup, E. T. Pappozzi, and M. E. Conley

## NONLINEAR MODELS FOR MULTI-FACTOR PLANT NUTRITION EXPERIMENTS

R.D. Landes<sup>1</sup>, W.W. Stroup<sup>2</sup>, E.T. Pappozzi<sup>3</sup>, and M.E. Conley<sup>3</sup><sup>1</sup> Division of Biometry, University of Arkansas for Medical Sciences<sup>2</sup> Department of Biometry, University of Nebraska-Lincoln<sup>3</sup> Department of Horticulture, University of Nebraska-Lincoln**Abstract**

Plant scientists are interested in measuring plant response to quantitative treatment factors, e.g. amount of nutrient applied. Response surface methods are often used for experiments with multiple quantitative factors. However, in many plant nutrition studies, second-order response surface models result in unacceptable lack of fit. This paper explores multi-factor nonlinear models as an alternative. We have developed multi-factor extensions of Mitscherlich and Gompertz models, and fit them to data from experiments conducted at the University of Nebraska-Lincoln Horticulture department. These data are typical of experiments for which conventional response surface models perform poorly. We propose design selection strategies to facilitate economical multi-factor experiments when second-order response surface models are unlikely to fit.

**1. Introduction**

Whether it is an attractive flower, a high-yielding grain, or an outstanding cut of meat, quality in agricultural products depends on a number of factors. The focus of much agricultural research is to manipulate those factors that can be controlled in the production process to obtain the highest quality product. Often, the factors under study are quantitative, e.g. amount of fertilizer, amount of irrigation, amount of herbicide, etc. Of interest to producers is to optimize their “bottom line.” That is, they want to produce the most marketable product possible using the least expensive combination of input levels. Of interest to researchers is to find answers to producers’ questions within the constraints of their resources. Researchers must be able to

provide accurate information using the least expensive possible experiments.

This paper discusses statistical issues that have arisen from plant nutrition research conducted at the University of Nebraska Department of Horticulture under the direction of Professor Ellen T. Paparozzi. She has directed several M.S. and Ph.D. research projects looking at the effects of the nutrients, nitrogen and sulfur, on the physiology and performance of a number of ornamental plant species. The crucial statistical issue that has emerged is the absence of suitable statistical methods for studying nitrogen and sulfur in the context of other nutrients or other inputs. In this paper, we review:

- methods currently used for these types of experiments,
- methods proposed for use with more than 2 input factors
- why proposed methods often are not appropriate
- some alternative methods using non-linear models that we are exploring.

While the context and examples for this paper are from plant nutrition, the statistical content generalizes to a wide range of applications in agriculture involving quantitative factorial treatment structures.

## 2. Background

Plant nutrition experiments are often limited to studying one or two nutrients at a time. This is primarily a result of the statistical methods -- the design of experiments as well as the analysis -- typically used. To gain a perspective of the methods historically used, a review is in order. We divide the review into three parts: single-factor experiments, two-factor factorials, and multi-factor experiments.

With single-factor experiments, e.g. a study of the effects of one nutrient only, two types of analyses are often used: regression and mean separation procedures. Regression can be very effective when the regression model used adequately describes the effect of nutrient level *and* is consistent with the researcher's objectives. However, for a variety of reasons, regression is often not used. Part of the reason stems from the fact that statistical methods courses tend to emphasize polynomial regression, which is frequently inappropriate for plant nutrition studies (see, for

example, Anderson and Nelson, 1975). Many authors, including Anderson and Nelson, discuss linear-plateau, or spline, or segmented regression. However, researchers prefer to use mean-separation procedures instead, as they are considered easier to perform and interpret for the one nutrient case.

For studies with two nutrient factors (e.g. nitrogen and sulfur) researchers almost always use full factorial treatment designs and either regression or mean-separation procedures. In exploratory research, there is often a fair amount of uncertainty concerning the experimental region. Thus, there are usually at least 4 levels per factor, and factorials as large as 7x8 are not uncommon. Regression, specifically second order response surface regression is often seen as a sensible approach to the two nutrient case, provided that the assumptions – linear or quadratic main effects, at most linear-by-linear interaction – are met. Often, they are not. As in the one-factor case, mean separation procedures are the “standard fall back” analysis. However, these procedures are cumbersome, both to implement and to interpret. For example, a 4 x 4 factorial of nitrogen and sulfur implies that there are  ${}_{16}C_2 = 120$  possible pair-wise comparisons among the nutrient levels. Though a competent researcher would only look at a subset rather than all 120 comparisons, the interpretation can still be confusing. Moreover, the mean separation approach obstructs insight into interaction, and can lead to an inflated Type I error rate.

Plant nutritionists tend to resist studies with three or more factors. In part, this is because of the reliance on full-factorial designs and mean separation procedures discussed above. For example, using methods considered “conventional” in plant nutrition, a three-factor experiment implies a minimum of  $4^3 = 64$  treatment combinations, but, more likely,  $5^3 = 125$  treatment combinations given the uncertainty about the experimental region usually present. These experiments must be adequately replicated: assuming the “usual” 4 replications, this means 256-500 plants, which is unmanageable in a typical university experiment station setting. If mean separation procedures are used, then the number of possible of comparisons becomes staggering - at least  ${}_{125}C_2 = 7750$  - and then the proper interpretation of the comparisons becomes the proverbial Gordian knot.

Alternatives are clearly needed. At first glance, the multi-factor experiment seems to call for response surface methods. The efficiency of central-composite, Box-Behnken, and related

designs, and the relative simplicity of interpreting second order polynomial regression models associated with standard response surface methods (see, for example, Myers, 1976, or Khuri and Cornell, 1996) would seem to be obvious for such research. Alas, things are not so simple. Mead and Pike (1975) reviewed the use of response surface methods in agriculture. They found that while there are many situations in agricultural research that seem to call for response surface methods, these methods are actually used only in a small minority of cases. They concluded that they were three main explanations for this:

- tradition: journal articles tend to use statistical methods that have been used in previous publications.
- education: response surface methods are not usually taught in standard statistical methods sequences for agricultural researchers, hence they tend to be unaware of or uncomfortable using such methods.
- statistical: standard response surface models do not adequately describe the plant response under study or otherwise fail to address the researcher's objectives.

Things have not changed much in the past twenty-plus years, in that the use of response surface methods is still relatively uncommon in agricultural research. The third of Mead and Pike's explanations – that response surface models fall short of researchers' needs – warrants further attention. Response surface methods have been applied with great success in a variety of engineering and quality improvement settings. In these settings, researchers usually have the option of running a number of relatively short experiments to "hunt" for a small subset of the experimental region where a second order polynomial regression provides a reasonable approximation of the response surface. The key phrases here are "number of short experiments" and "small subset." Most agricultural experiments also have relatively small subsets of the experimental regions near the stationary point where second order models fit quite well. The problem is that in many agricultural settings, e.g. plant nutrition, experiments must last as long as it takes to grow a plant, usually months, and it is not possible *a priori* to limit the range of nutrients because the probable stationary point is not well known in advance. Whereas an engineer can run several experiments to hunt the experimental region over a period of a few months, the plant scientist can only run one. This is a major reason why second order polynomial

regression cannot be depended upon as a method of analysis.

The questions are these:

- how justified are the concerns about the validity of second order polynomial regression, and
- are there alternative models that might be used to better describe the responses typically observed, and that might allow efficient designs, possibly in some modified form, to be used?

We looked at data from a number of experiments conducted by Paparozzi, et. al. over the past several years. The two cases described in the next section are typical.

### 3. Two-Factor Examples

In this section, we consider two response variables from an experiment conducted as M.S. thesis research by Macz (1997). One is typical of responses for which second order response surface methods are appropriate. One typifies response variables for which alternative models are clearly essential.

The experiment is a 4 x 5 factorial. Although not pertinent to this discussion, it was conducted using a row-column design. The factors (nutrients) are nitrogen (N) at 50, 100, 150, 200 units, and sulfur (S) at 0, 5, 10 20, 80 units. The experimental units are individually potted chrysanthemums, with the experiment taking place in a greenhouse. There were eight replications per treatment, but due to the nature of the sampling (e.g. destructive sampling) there are three to five replications depending on the response variable. Several response variables are evaluated. The skeleton ANOVA for this experiment is given in Table 1. The response variables reasonably modeled with second order response surfaces are days to bloom, and flower longevity in interior condition, measured in days. Others – height, size of the flower, area of the leaves at 8 weeks, days to bud set, elemental concentration of N and S in the leaves at 2, 4, 6, and 8 weeks – share similar mean profiles which are not adequately modeled by second order polynomial regression. In this discussion, we consider flower longevity and leaf area at 8 weeks.

First, consider flower longevity. Figure 1a shows a plot of the least squares means of flower longevity from an ANOVA of the aforementioned row-column design; from initial observation

the plotted surface seems a likely candidate for response surface regression. The negative of the least squares means are also presented (Figure 1b) to give a clearer view of the surface. Using PROC MIXED (SAS version 6.12, SAS Institute, 1997) an initial analysis of the experimental data is performed (Table 2). The lack of fit term (LOF) is not found to be statistically significant at the 0.10 level. As a precaution, in a secondary analysis (Table 3), the lack of fit term is broken down into its three main parts: higher order ( $\geq 3$ ) main effects of N and S, and higher order interactions between the two. These three lack-of-fit components are all non-significant even at the 0.10 level. There is no evidence that the response surface regression is inappropriate. The estimated regression equation is

$$y = 40.4 - 0.13*N + 0.0004*N^2 - 0.0069*S + 0.0005*S^2 + 0.0001*N*S$$

with an MSE of 2.91 days<sup>2</sup>. Solving for the stationary point yields N at 170 units and S at 62 units. The estimated response surface is given in Figure 1c. Visually comparing the estimated surface to observed least squares mean profile (Figures 1a and 1b), the fit does not seem unreasonable.

Of the response variables from the second group, the least squares means from leaf area (cm<sup>2</sup>) best typifies the general shape of this group. They are plotted in Figure 2a. Inspection of the plots suggests that the second order response surface model is unlikely to provide a useful fit. An initial analysis (Table 4) is performed and the lack of fit term produces an F-statistic with a p-value of 0.0001. Statistically significant lack of fit tends to confirm visual inspection. How badly does the second order polynomial fit? A response surface is estimated and is given in Figure 2b. Visually comparing the observed profile (Figure 2a) to the estimated surface, the fit does not begin to capture what is happening with leaf area. Were a researcher to use the results of conventional response surface methods, seriously, perhaps catastrophically, erroneous conclusions about optimum N and S levels to be recommended and likely responses to be realized would be the result.

For the response variables that behave as leaf area, a possible solution is to search for a non-linear model that provides a better fit. For this group of responses, the Mitscherlich model, logistic model, and Gompertz model were reasonable candidates. We present the Gompertz model here. The two-factor model is as follows:

$$\eta = \theta_{\max} \exp\{-\theta_{\text{mg}} \exp(\theta_N N + \theta_S S + \theta_{NS} NS)\}$$

where  $\eta$  is the expected leaf area,  $\theta_{\max}$  is the maximum area,  $\theta_{\text{mg}}$  is the range of the responses, and  $\theta_N$ ,  $\theta_S$ , and  $\theta_{NS}$  are the rates of growth due to N, S, and their interaction.

The Gompertz regression equation was estimated using PROC NLIN (SAS version 6.12, SAS Institute, 1997). For starting values we used the maximum value of the response, 1025 cm<sup>2</sup>, for  $\theta_{\max}$ , the range of the responses, 766 cm<sup>2</sup>, for  $\theta_{\text{mg}}$ , and a grid search on the interval -0.05 to 0.05 for  $\theta_N$ ,  $\theta_S$ , and  $\theta_{NS}$ . The estimated parameters are given in Table 5. The resulting surface is shown in Figure 2c. The MSE was 8843.0 (cm<sup>2</sup>)<sup>2</sup>. This model provides a better fit to the actual data than does the second order response surface model, especially with respect to what is happening with the response as a function of S. But the model seems to miss the behavior of the response as a function of N

Looking at the least squares mean profile, the leaf area responds linearly to nitrogen. With this in mind, we altered the Gompertz model so that the N term is linear and the S term remains non-linear. The resulting regression equation is

$$\eta = (\theta_{\min} + \theta_N N) \exp\{-\theta_{\text{mg}} \exp(\theta_S S)\}$$

where  $(\theta_{\min} + \theta_N N)$  acts as an asymptote that is a function of N,  $\theta_N$  is the rate of growth due to N,  $\theta_{\text{mg}}$  is a function of the range, and  $\theta_S$  is the rate of growth due to S. Note that  $\theta_N$  can be regarded as the parameter that characterizes the interaction between N and S.

The initial estimate of  $\theta_{\min}$  is the mean response of the combination of the lowest levels of the nutrients.  $\theta_N$  is determined by the rise in the least squares means divided by the range in the N values, which is equivalent to the mean slope of N with respect to the least squares means. Initial estimates for  $\theta_{\text{mg}}$  and  $\theta_S$  are the result of a grid search in PROC NLIN. The final estimates of the parameters from PROC NLIN are given in Table 6 with the resulting surface in Figure 2d. The MSE is 4278.2 (cm<sup>2</sup>)<sup>2</sup>, which is less than half of the original Gompertz model.

Comparing all three of the fitted models for leaf area at 8 weeks, the latter of the three models best describes the data. This model, with appropriate initial estimates, performs similarly with the other response variables in this group.

#### 4. Implication for Design of Experiments

The results in section 4 show that response variables fall into two categories: those that can be adequately modeled by second order polynomial regression (and hence could be handled using standard response surface methods) and those that require more sophisticated methods, such as non-linear models. As mentioned previously, if researchers are going to do multi-factor research, they cannot use conventional, full-factorial designs. They need more efficient alternatives.

For response surface problems for which second order polynomial regression is appropriate, there is a rich literature of efficient designs – central-composite, face-balanced cube, Box-Behnken, etc. Designs suitable for plant nutrition experiments would combine treatment designs for second order response surface regressions with treatment designs for non-linear situations. Box and Lucas (1959) present the methods necessary for determining the optimal design points when the response variables cannot be adequately described with a linear model. With prior knowledge of how a response variable behaves in the presence of a particular nutrient, appropriate non-linear models can be suggested. Next, with information from past experiments, or reasonable educated guesses about anticipated treatment effects, suitable initial estimates may be obtained for the parameters. Then, using the methods of Box and Lucas, optimal design points may be obtained. Finally, combining these with the design points suggested by such designs as the central composite, face-balanced cube, or Box-Behnken, a viable treatment structure may well emerge.

To illustrate, consider a single nutrient case for which some response variables may be modeled by second order polynomial (quadratic) regression, and the rest may be modeled by the Gompertz. To fit the quadratic model,  $\eta_{Lin} = \beta_0 + \beta_1 X + \beta_2 X^2$ , one would use the coded design points  $\{-1, 0, 1\}$ , where -1 corresponds to the lowest level of the treatment factor, 1 corresponds to the highest level, and 0 corresponds to the middle level. That is, the design points are equally spaced. On the other hand, to fit the Gompertz model,  $\eta_{Gom} = \theta_1 \exp[\theta_2 \exp(\theta_3 X)]$ , Box and Lucas show that the design points  $\{-1, g, 1\}$  are D-optimal, where -1 and 1 are as before, and  $g$  is the middle design point obtained using the methods of Box and Lucas. Note that  $g$  is not mid-way between levels -1 and 1.

One idea for an optimal design robust to model (e.g. able to provide useful estimates for either quadratic or Gompertz models) would be to combine the designs, that is, use the points  $\{-1, g, 0, 1\}$  or  $\{-1, 0, g, 1\}$  if  $g$  happens to be greater than 0. Other strategies might include finding a compromise point between  $g$  and 0, say  $c$ , and using  $\{-1, c, 1\}$ . Expanding the idea to three or more factors – which is the setting where these design ideas are most needed – one might consider identifying 0 and  $g$  for each treatment factor, then overlaying two Box-Behnken or two face-balanced cubes, one using 0 as the center point, the other using  $g$ . Or one might use only the compromise center point,  $c$ . Or one might use the fact that the central composite design actually has 5 design points along each treatment factor axis, and use it, either with 0,  $g$ , or the compromise value as a center point.

## **5. Summary**

Much of what is known in plant nutrition and other agricultural areas has come from experiment with one or two quantitative factors. However, experiments involving three or more quantitative factors with multi-levels present serious problems to researchers because of the nature of the response variables and limitations in the availability of resources. With the recent increase in computing power, combining treatment selection strategies for non-linear models with those for second order polynomial models to account for the differing behaviors of the response variables appears to be a promising approach to addressing these problems. The examples shown in this paper, which are typical of plant nutrition research, clearly illustrate the usefulness of non-linear models for two-factor response surface experiments. We expect that we will have similar success applying these methods to three and higher-factor experiments.

## 6. References

- Anderson, R.L. and L.A. Nelson. 1975. "A family of models involving intersecting straight lines and concomitant experimental designs used in evaluating response to fertilizer nutrients." *Biometrics* **31**: 303-318.
- Box, G.E.P. and H.L. Lucas. 1959. "Design of experiments in non-linear situations." *Biometrika* **46**: 77-90.
- Khuri, A.I. and J.A. Cornell. 1996. *Response Surfaces: Designs and Analyses*. New York: Marcel Dekker, Inc.
- Macz, O. 1997. *The Effect of Nitrogen and Sulfur Applications on Pot Chrysanthemum Production and Post-Harvest Performance*. M.S. Thesis in Horticulture, University of Nebraska-Lincoln.
- Mead, R. and D.J. Pike. 1975. "A review of response surface methodology from a biometric viewpoint." *Biometrics* **31**: 803-851.
- Myers, R.H. 1976. *Response Surface Methodology*. Blacksburg, VA: author.
- SAS institute, Inc. 1997. *SAS/STAT Software: Changes and Enhancements through Release 6.12*. Cary NC: SAS Institute.

Table 1. Sources of Variation and degrees of freedom for 4x5 N-S example.

ANOVA	
Source	DF
Row	3
Col	3
Nitrogen	3
Sulfur	4
Nitrogen x Sulfur	12
Error	*
Total	*

\* The Error and Total degrees of freedom depend upon the response variable.  
 Range for Error df: 18 – 54; range for Total df: 53 – 79.

Table 2. Initial PROC MIXED Analysis for Flower Longevity

Tests of Fixed Effects				
Source	Numerator DF	Denominator DF	Type I F	Pr > F
ROW	3	37	2.32	0.0915
COL	3	37	11.02	0.0001
N	1	37	29.69	0.0001
N*N	1	37	19.69	0.0001
S	1	37	0.30	0.5899
S*S	1	37	0.67	0.4194
N*S	1	37	0.85	0.3614
LOF	10	37	0.49	0.8844

Table 3. Secondary PROC MIXED Analysis for Flower Longevity

Tests of Fixed Effects				
Source	Numerator DF	Denominator DF	Type I F	Pr > F
ROW	3	37	2.32	0.0915
COL	3	37	11.02	0.0001
N	1	37	29.69	0.0001
N*N	1	37	19.69	0.0001
S	1	37	0.30	0.5899
S*S	1	37	0.67	0.4194
N*S	1	37	0.85	0.3614
NIT	1	37	0.64	0.4281
SULF	1	37	0.45	0.5070
NIT*SULF	8	37	0.48	0.8634

Table 4. Initial PROC MIXED Analysis for Leaf Area

Source	Tests of Fixed Effects		Type I F	Pr > F
	Numerator	Denominator		
	DF	DF		
ROW	1	18	0.84	0.3704
COL	1	18	0.72	0.4067
N	1	18	303.27	0.0001
N*N	1	18	11.61	0.0031
S	1	18	52.92	0.0001
S*S	1	18	168.58	0.0001
N*S	1	18	10.44	0.0046
LOF	14	18	13.05	0.0001

Table 5. Estimates for the Gompertz Model fit to Leaf Area

Parameter	Estimate	Asymptotic s.e.	Asymptotic 95 % Confidence Interval	
			Lower	Upper
theta_max	846.1	24.1	797.3	895.0
theta_rng	0.762	0.136	0.486	1.037
theta_N	-0.001	0.002	-0.004	0.002
theta_S	-0.220	0.055	-0.332	-0.108
theta_NS	0.004	0.001	0.002	0.007

Table 6. Estimates for the "Hybrid" Gompertz model fit to leaf area

Parameter	Estimate	Asymptotic s.e.	Asymptotic 95 % Confidence Interval	
			Lower	Upper
theta_max	321.2	28.36	263.7	378.7
theta_N	3.27	0.208	2.85	3.70
theta_rng	0.883	0.076	0.728	1.038
theta_S	-0.537	0.127	-0.794	-0.280

Figure 1a. Plot of Least Square Means of Flower Longevity for 4 x 5 Nitrogen-Sulfur Experiment

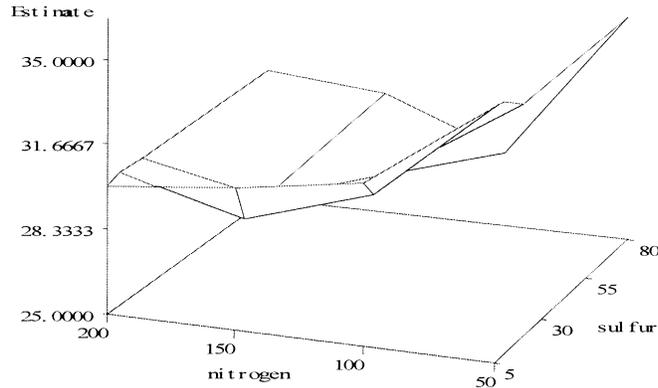


Figure 1b. Plot of Negative values of Flower Longevity Least Square Means from Fig. 1a

1c. Predicted Response Surface for Fig. 2b Using 2nd Order Polynomial Regression Model

Figure 1b

Figure 1c

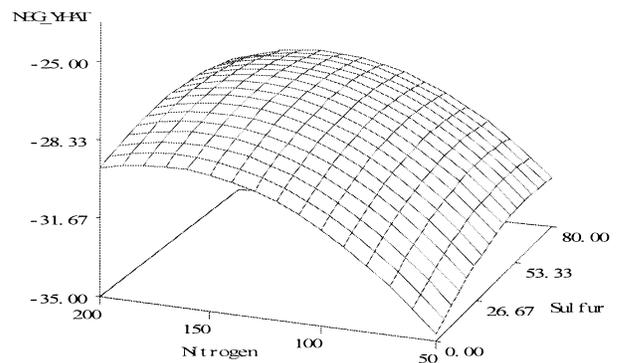
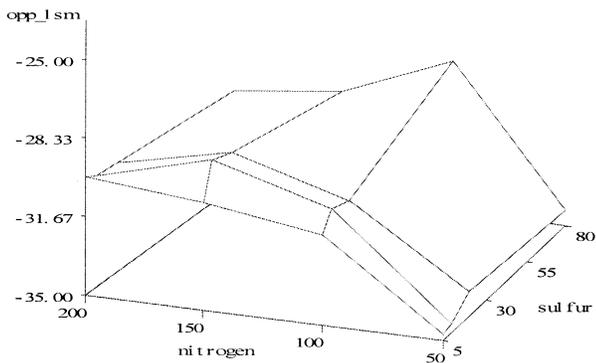


Figure 2a. Plot of Least Square Means of Leaf Area for 4 x 5 Nitrogen-Sulfur Experiment

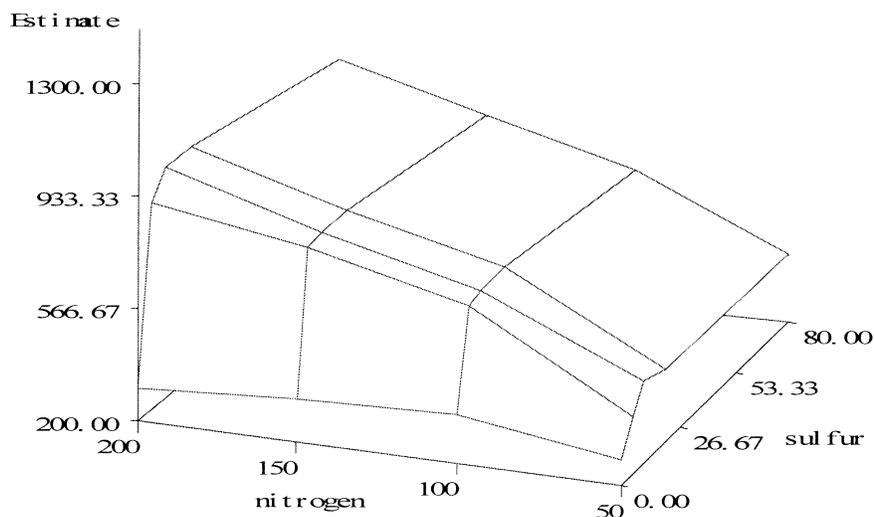


Figure 2b. Predicted Response Surface for Leaf Area Using 2nd Order Polynomial Regression Model

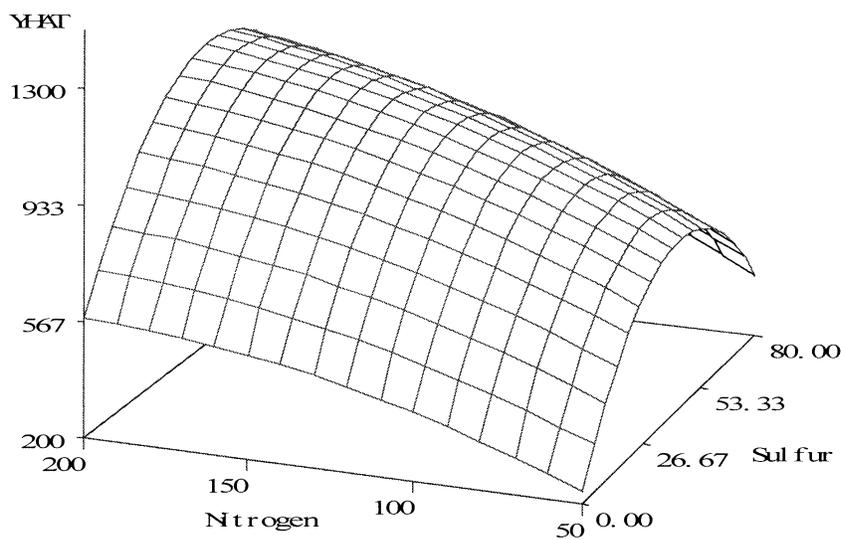


Figure 2c. Predicted Response Surface for Leaf Area Using Gompertz Model

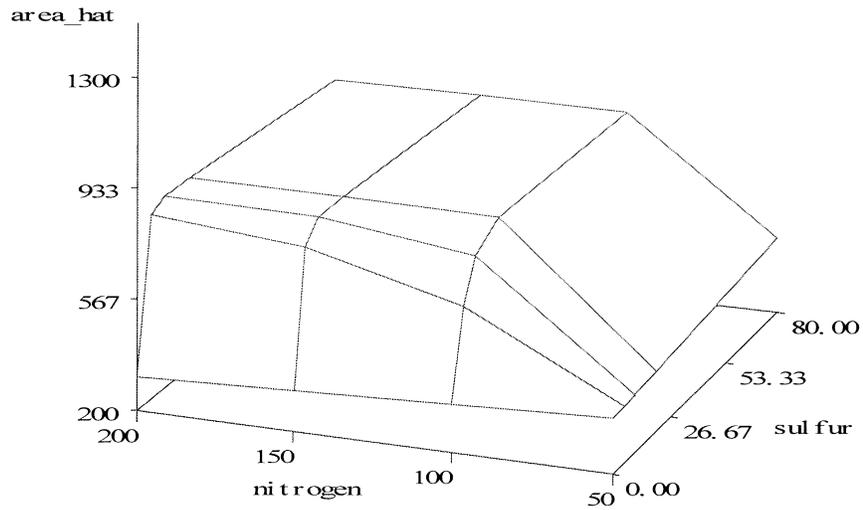


Figure 2d. Predicted Response Surface for Leaf Area Using Modified Gompertz Model (linear in N, non-linear in S)

