A RESPONSE SURFACE OUTLOOK FOR OPTIMALITY IN RICE PRODUCTION IN THE NORTHERN NIGERIAN STATE OF KANO

Adehi M. U.[1*], Dikko H. G.[2], David I. J.[2]

[1]Department of Mathematical Sciences, Keffi, Nasarawa State, Nigeria.
[2]Department of Mathematics, Ahmadu Bello University, Zaria, Kaduna State, Nigeria.
***Corresponding Author's E-mail: maryadehi@yahoo.com**

Abstract

With the aid of response surface methodology a response surface model was developed for a mixed level two factors (irrigation at 3 levels and fertilizer at 4 levels) experimental design on rice yield in Kano State which is situated in the northern part of Nigeria. A quadratic surface was fitted and a canonical analysis was carried out to determine the surface in the region of maximum. Goodness of fit tests was performed in assessing the model where the R^2, adjusted-R^2, and residual plots were done and it was revealed that the model is adequate.

Keywords*:* Response surface, Factors, Canonical analysis, Quadratic surface, Residual analysis, Goodness of fit test, Contour plots, ANOVA.

Introduction

The question of obtaining optimal results is relevant in every aspect of life but it is even more so in the production of food because it is a basic need which also enhances an economy. Every farmer engages in farming in order to maximize profit. The yield is therefore essential. Some farmers stabilize and become successful, while others fail even before they begin. Is the success of the farmer a result of chance or can it be enhanced through the application of certain procedures? In statistics, response surface methodology explores the relationship between several explanatory variables and one or more response variables. It is an easy way to estimate polynomial models through the use of factorial experiment or a fractional factorial design. Response Surface Methodology (RSM) has an effective track record of helping researchers improve products and services. There are many methodological approaches to response surface design in order to obtain the point of optimum among applied treatments within designed experimental units. The approximation of the response function $y = f(x_1, x_2, \ldots, x_n) + e$ is called *Response Surface Methodology*. RSM is a branch of design which involves finding out the characteristics of the factors influencing the variables analyzed and it also emphasizes finding out the particular treatment combinations that cause the optimum response. In addition, RSM involves investigating the response surface near the optimum yield, (Anderson and Mclean, 1974).

RSM is a collection of mathematical and statistical techniques that are useful for the modeling and analysis of problems in which a response of interest is influenced by several variables with the objective to optimize the response, (Montgomery, 2005).

Response surface is sometimes called "optimum" according to (Anderson and Mclean, 1974), and it was developed mainly after world-war II.

Throughout the period in which RSM was being developed, there were two parallel developments;

1. Box approach: - It is concerned with developing methods for tackling applied problems in RSM than the general mathematical theory.

2. Keifer approach: - It is concerned with developing methods for tackling the general mathematical theory.

According to Silvey, (1980), both approaches overlap at some point in the application of theoretical methods. Three general problems were presented for the distribution of an observable random variable y which depends on:

1. A column vector $\mu = (\mu_1, \mu_2, ..., \mu_r)^T$ of real variables called control variables because they can be chosen by the experimenter.

2. A column vector $\theta = (\theta_1, \theta_2, ..., \theta_k)^T$ of parameters which are fixed but unknown to the experimenter; these or certain functions of them are of interest to him.

3. A column vector $\tau = (\tau_1, \tau_2, ..., \tau_l)^T$ of nuisance parameters, these also are fixed and unknown but they are not of primary interest.

These stated problems are posed on full decision theoretic approach. Silvey (1980) further stated that the experimenter is allowed to take n independent observations on y at vectors $\mu_1, \mu_2, ..., \mu_n$ chosen from the set Q. Such a choice of n vectors, not necessarily all distinct, is referred to as an n–observation design. The basic problem is in the choice of n–observation design. The main concern is therefore in selecting a design or designs that minimize the variance associated with the fitted model, y.

The first step in RSM is to find a suitable approximation for the true functional relationship between y and the set of independent variables. Usually, a low-order polynomial in some region of independent variables is employed. If the response is well modeled by a linear function of the independent variables, then the approximating function is the first-order model.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \varepsilon \quad (1)$$

If there is curvature in the system, then a polynomial of higher degree must be used, such as the second-order model.

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \beta_{ii} x_i^2 + \sum \sum \beta_{ij} x_i x_j + \varepsilon$$

(2)

The method of least squares is then used to estimate the parameters; the response surface analysis is then performed using the fitted surface. The researcher then seeks the path of the optimum response (Montgomery, 2005). Box and Draper (1987) stated that the main aim of the optimality theory is in selecting an optimum experimental design which in most cases needs to be multifaceted, therefore the problem of selecting a suitable design is thus a formidable one.

A response surface can behave in either of three main functions namely; linear, quadratic or cubic.

Many researchers have been carried out using different response surface methodologies. Often, the composite design developed by Box and Wilson (1951) is applied instead of the full factorial design; the main reason is as a result of reduction in treatment combination that is achieved with the central composite design. But in this research the full factorial design is been applied to determine the point of maximum for the yield of a particular NERICA rice variety. Also, the response surface and contour plots were plotted in order to understand how the response changes in a given direction by adjusting the design variables.

Materials and Methods

The data used for this research work was obtained from Irrigation Research Station of the Institute for Agricultural Research, located in Kadawa, Kano State. (Latitude. 11.6500^0, Longitude. 8.4500^0). Field trials were conducted during the third quarter of the 2008 rainy seasons in Kano. The plot size was 4m x 3m ($12m^2$) and net plot was 3m x 3m ($9m^2$) excluding the border row. The treatments that were applied in the course of experimentation consisted of 3 irrigation intervals (7, 14, and 21 days) and four fertilizer rates (30, 60, 90, and 120 kg N ha^{-1}) to check their combinatorial effect on the yield of a particular rice variety. Hence, the experiment was based on a mixed level complete factorial design.

Complete Factorial Design

The complete factorial design is adopted for this research work due to the fact that the levels of the factors are mixed and each factor level exceed two while their treatment combination is not far from that of the fractional factorial and the composite design which is based on two levels of each factor.

This research is of a mixed level design with a total treatment combination of 12 that is 3 x 4, while for a central composite design the total treatment combination will be $2^2 + 2(2) + 1 = 9$. Hence, the complete mixed-level factorial will be able to estimate parameters with equal variances and adequate degree of freedom for error.

Quadratic Surface for the Two Factors

$$Y = b_0 + b_1 x_1 + b_2 x_2 + b_{11} x_1^2 + b_{22} x_2^2 + b_{12} x_1 x_2 + e \qquad (3)$$

Differentiating (8) partially and equating to zero, with respect to x_1, x_2 we have,

$$\frac{\partial y}{\partial x_1} = b_1 + 2b_{11}x_1 + b_{12}x_2 = 0 \qquad (4)$$

$$\frac{\partial y}{\partial x_1} = b_2 + 2b_{22}x_2 + b_{12}x_1 = 0 \qquad (5)$$

The solutions of these equations give the factorial combination at which Y is a local maximum or minimum, or a local stationary value.

Canonical Analysis

Canonical analysis is useful to determine whether the stationary point is a maximum, minimum or saddle point.

$$\begin{vmatrix} b_{11} - \lambda & b_{12}/2 \\ b_{21}/2 & b_{22} - \lambda \end{vmatrix} = 0$$

Surfaces and Contours

Contours and surfaces are pictorial representation of quadratic surfaces which aid interpretation of result; a given contour shows all pairs of values of x_1 and x_2 in a two factor experiment for which the response has a specific value. With three factors, the contour surfaces can be built up from contour lines of two dimensions by superposition. Contour surfaces respond by the signs of the coefficients of λ_i, as can be seen in table 2.

Table 2: 2 – Factors with Coefficients λ_1 and λ_2.

$(-,-)$	The surface contours are ellipses
$(\|\lambda_1\| < \|\lambda_2\|)$ and the contour has the major axes along the x_1 direction	The surface is a hill
$(+,+)$	The surface will be a basin
$(-,+)$	The contours are hyperbolas
$(\|\lambda_1\| < \lambda_2)$	The surface is a saddle
$(-,0)$	Contours become straight lines and the surface is called a stationary ridge
(-1, near 0)	The contours are parabolas and the surface is a rising ridge

Source: Peng, 1967.

Contour plots examine the response surface to determine whether the stationary point is a maximum, minimum or saddle point through transformation of the fitted response surface to canonical form.

For $Y = Y_s + \lambda_1 W_1^2 + \lambda_2 W_2^2 + \ldots \lambda_k W_k^2 + \in$ (6)

where $\{W_i\}$ are the transformed independent variables of $x_i\,'s$ and $\{\lambda_i\}$ are constants and the $\{W_i\}$ direction for which $|\lambda_i|$ is the greatest, Montgomery (2005).

Goodness of Fit Test

In checking the adequacy of the response surface model the use of coefficient of determination, R^2, adjusted-R^2, Coefficient of Variation (CV) and residual plot were employed. A residual plot may be used to determine the goodness or lack of fit of a model. It tells us how adequate the response function is. To carry out this check, the

are just the Eigen-values or characteristic roots the sign of the $\{\lambda_i\}$ is useful for interpreting the contour surface plot. If $\{\lambda_i\}$ are all positive, x_s is a point of minimum response; if $\{\lambda_i\}$ are all negative, x_s is a point of maximum response; and if $\{\lambda_i\}$ have different signs, x_s is a saddle point. Furthermore, the surface is steepest in

residual values are plotted against the fitted values or against the levels of the independent variables, if there is no systematic departure from the horizontal zero band, then the response function is accepted as a good fit.

The R^2 and adjusted-R^2 are measures of how much variation in the response variable *y* is explained by the variables (x_1, x_2) included in the model.

The coefficient of determination R^2 takes values within $0 \le R^2 \le 1$.

It should be noted at this point that, a small value of R^2 means that $X_1 + X_2 + \cdots + X_k$

contribute very little information for the prediction of Y, a value of R^2 near 1 means that $X_1 + X_2 + \cdots + X_k$ provide almost

all the information necessary for the prediction on Y.

However, a high R^2 does not necessarily mean that the model is adequate for prediction. This is because R^2 always increase as we add terms to the model, so, trying to find a regression model where R^2 is as close to 1 as possible is not a good practice.

To remedy this problem, many researchers prefer the use of adjusted coefficient of determination, R^2_{adj} statistic defined as

$$R^2_{adj} = 1 - \frac{SSE * df_{Total}}{SST * df_{Error}} \qquad (7)$$

R^2_{adj} is a modification of R^2 that adjusts for the number of explanatory terms in a model. Unlike R^2, R^2_{adj} increases only if the new term improves the model more than would be expected by chance. Thus, R^2_{adj} penalizes when poor predictor variables are been added and it rewards when good ones are added.

The coefficient of variation, (CV) is used to measure the precision with which an experiment has been carried out, and it is given by;

$$CV = (SD / MEAN) * 100 \qquad (8)$$

Results and Discussions

Fitted Model for 2008 Rice Yield:

The full model that was obtained for rice yield in 2008 is:

$$\hat{Y} = 26.5 + 0.1I - 0.02I^2 + 0.54N - 0.01N^2 + 0.000013N^3 - 0.098IN + 0.0014IN^2 - 0.000004IN^3 + 0.004I^2N - 0.0001I^2N^2 + 0.00000021I^2N^3 \qquad (9)$$

where, $\hat{Y}$ is the response estimate for yield; I is the irrigation linear effect; I^2 is the irrigation quadratic effect; N is nitrogen linear effect; N^2 is nitrogen quadratic effect; N^3 is nitrogen cubic effect.Equation (9) is the response surface polynomial function for 2008 rice yield.

The ANOVA table below helps us to select the parameters that are significant and needed for the response surface model.

TABLE 3: ANOVA Table for Rice yield.

SOURCE	DF	SS	MS	F	P-VALUE
REP	1	17.17	17.17	2.10	0.1508
I	2	742.95	371.47	45.52	
I	1	741.84	741.84	90.91	0.0001
I^2	1	1.11	1.11	0.136	0.7134
N	3	198.15	66.05	8.09	
N	1	193.40	193.40	23.70	0.0001
N^2	1	4.75	4.75	0.58	0.4479
N^3	1	0.0002	0.0002	0.000025	0.9957
IN	6	156.62	26.10	3.198	
IN	1	38.15	38.15	4.67	0.0335
IN^2	1	24.15	24.15	2.96	0.0892
IN^3	1	2.51	2.51	0.31	0.5805
I^2N	1	85.50	85.50	10.47	0.0017
I^2N^2	1	5.55	5.55	0.68	0.4119
I^2N^3	1	0.76	0.76	0.09	0.7613
ERROR	83	677.66	8.16		
TOTAL	95	1792.54			

CV	9.834
R^2	0.622
Adjusted-R^2	0.5673

From the table above it can be seen that at 5% level of significance the parameters *I*, *N*, *IN* and I^2N are significant since their p-values are less than α = 5% significance level and our final response surface function is;

$$\hat{Y} = 26.5 + 0.25I + 0.54N - 0.098IN + 0.004I^2N \quad (10)$$

It is observed from the ANOVA table above that I^2, N^2, I^2N^2, N^3, N^2, IN^2, IN^3 and I^2N^3 are not significant at 5% level of significance, therefore, we concluded that their effects on rice yield are not significant. The levels of interaction of the variables that are not significant do not account for variability in the yield response of upland rice, and so, the practice of applying too much fertilizer over irrigation should be discouraged among the farmers.

Canonical analysis for the fitted model

Here we will determine whether the stationary point is a point of maximum, minimum or a saddle point.

The fitted model to be used is equation (10)

$$\hat{Y} = 26.5 + 0.25I + 0.54N - 0.098IN + 0.004I^2N$$

Differentiate equation (17) partially w.r.t. *I* and *N* and equate to zero to find the optimum point, we have;

$$\frac{\partial \hat{Y}}{\partial I} = 0$$

$$0.098N - 0.008IN = 0.25 \quad (11)$$

$$\frac{\partial \hat{Y}}{\partial N} = 0$$

$$0.004I^2 - 0.098I + 0.54 = 0 \quad (12)$$

Solving simultaneously equation (11) and (12), we have to solve equation (12) quadratically to have;

I = 16.13 or 8.37

Substituting I = 16.13 and 8.369 into equation (11) to obtain values for N, we have;

N = - 8.051 and 8.054 when I = 16.13 and 8.369 respectively; these are the optimum points, that is, *I* = 16, *N* = 0 or *I* = 8, *N* = 8.

Substituting I and N using I = 16, N = 0 and I = 8, N = 8.05 in equation (10) to obtain the optimum response Y_M as;

$Y_M = 26.5 + 0.25(16) + 0.54(0) – 0.098(16)(0) + (0.004*0)(16)^2$

$Y_M = 30.5$ is the optimum response for I = 16 and N = 0.

$Y_M = 26.5 + 0.25(8) + 0.54(8) – 0.098(8)(8) + (0.004*8)(8)^2$

$Y_M = 33.3$ is the optimum response for I = 8 and N = 8.

We therefore constructed a determinant matrix to determine the coefficients in the canonical form;

$$\begin{vmatrix} b_{11} - \lambda & (b_{12})/2 \\ (b_{21}/2) & b_{22} - \lambda \end{vmatrix}$$

Determine the Eigen-values or characteristic roots and equate the determinant to zero, we have;

$$\begin{vmatrix} 0 - \lambda & -0.049 \\ -0.049 & 0 - \lambda \end{vmatrix} = 0$$

solving the determinant matrix we have;

$-\lambda(-\lambda) – (-0.049*-0.049) = 0$

$\lambda^2 = 0.002401$

$\lambda = \pm 0.049$

taking the values of λ we have two set of canonical equations as

$$Y – 30.5 = 0.049I^2 – 0.049N^2 \qquad (13)$$

$$Y – 33.3 = 0.049I^2 – 0.049N^2 \qquad (14)$$

Equation (13) is obtained when *I* = 16 and *N* = 0; while equation (14) is obtained when *I* = 8 and *N* = 8. We observed that the contours of both equation (13) and equation (14) are hyperbolas with saddle surfaces because their coefficients β_{11} and β_{22} are positive and negative respectively, which means that there will be a rapid increase in yield if *I* is increased. This shows that there will be an increase in yield in the direction of *I* axis from the optimum response *M*.

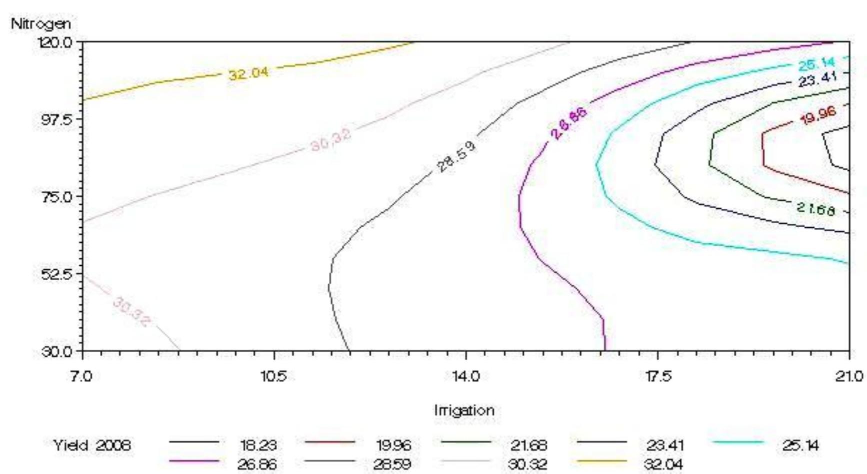

Figure 1: A Contour plot for Rice Yield

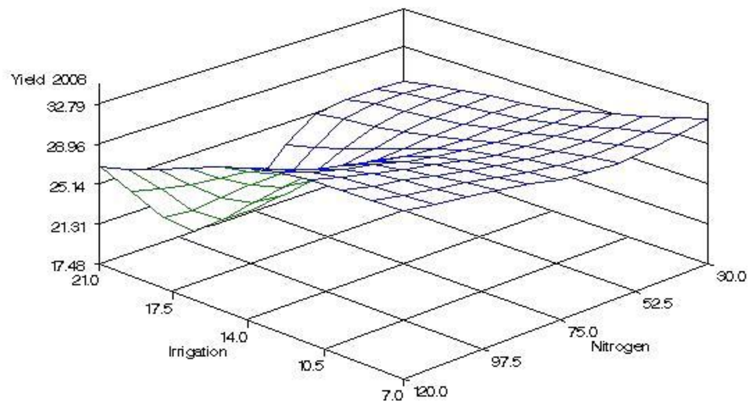

Figure 2: Surface plot for Rice yield

Figure 1 and 2 show the contour and surface plots for 2008 rice yield which confirms the result obtained from the canonical analysis that the contour is a hyperbola and the surface is a saddle.

From table II, the coefficient of variation value of 9.834% is far below 30%, it indicates that the experiment was well managed and that the data collected was adequate. The R^2 value of 0.622 explain above 60% of the total variation of the model but the R^2- adjusted is above 55% which support the conclusion that the model is fairly adequate. Graphical analysis for further model adequacy check is given below:

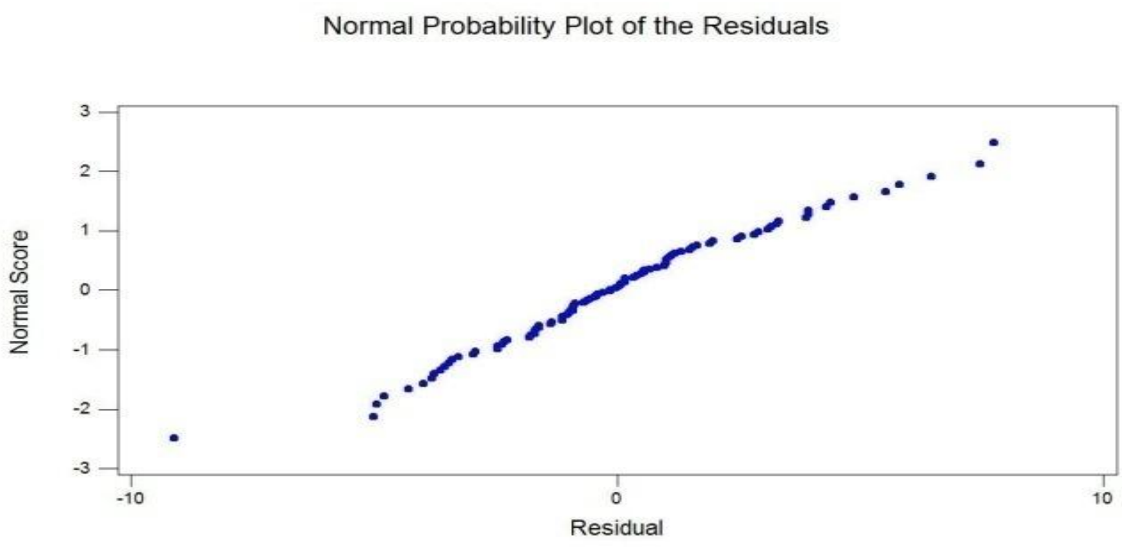

Figure 3: Normal Probability plot for Rice yield.

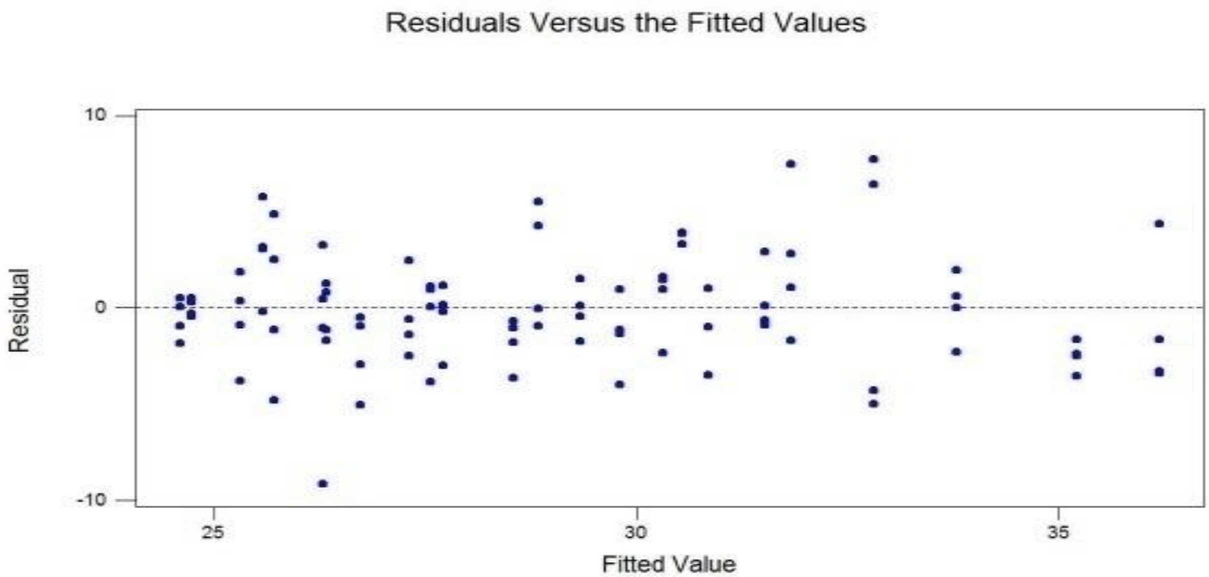

Figure 4: Residual Plot for Rice Yield.

Figure 3 above is the normal probability plot for 2008 rice yield data to check for normality of the data and it appears that the graph follows an *S*-shape which confirms the normality of the data. Figure 4 above is a graph showing the residuals versus the fitted values to check for consistency of the predicted response and it can be observed that the data points are approximately on the horizontal band.

Conclusion

The essence of this research basically is to develop a response surface model from a complete factorial experiment where all the levels of each factor are completely combined. Most of the time, response surface designs are practically conducted with techniques like the composite design, steepest ascent etc, where models developed are from reduced treatment combinations or factors with not more than two levels. In this research, we made use of two factors only, irrigation at three levels and nitrogen fertilizer at four levels.

From the result of the analysis, it was found that the quadratic response surfaces were not significant. This forms the focal point of the research. However, the quadratic surface for irrigation in interaction with the linear surface for nitrogen fertilizer was significant, as such; we did not conclude that the response surface model developed is completely linear in so far as a quadratic surface exists. From the canonical analysis derived for the model. It was observed that there was an increase in yield no matter which way one goes from the point of maximum, M because the response surface models are basins. However, the model showed that in the maximum response region, there is an increase in yield when irrigation and nitrogen are combined at quadratic and linear effect levels respectively. Hence, it’s recommended that the production of Kano upland rice should involve interaction of quadratic irrigation effect and a linear nitrogen effect. That is, the application of the combination of irrigation should be doubled with nitrogen at its single rate. The rice farmer should irrigate more than he applies fertilizer to the growing upland rice to harvest bountifully.

REFERENCES

Anderson, V. L. & Mclean, R. A. (1974). *Design of Experiments: A realistic approach.* Marcel Dekker, Inc. New York and Basel p. 326-387.

Box, G. E. P. & Draper N. R. (1987). *Empirical model building and response surfaces*: New York: John Wiley & Sons, Inc. 424 pp.

Box, G. E. P. & Wilson, K. B. (1951). Experimental Attainment of Optimum conditions (with discussion). Journal of Royal Statistics Society, series B 13(1): 1-45.

Montgomery, D. C. (2005). *Design and Analysis of Experiments*, 5th Ed. John Wiley and Sons New York..

Peng, K. C. (1967). *The Design and Analysis of Scientific Experiments*. Reading, mass; Addison – Wesley publications, p. 151 – 185.

Silvey, S. D. (1980). *Optimal Designs. An Introduction to Theory for Parameter Estimation.* Chapman and Hall, USA, p. 1 – 8.