SEMINAR REPORT

Optimization through Response Surface Methodology

By

SHIVAKUMAR M.

(2018-19-005)

Presented on 22-11-2019

Submitted in partial fulfillment of requirement of the course

STAT 591: Master's Seminar (0+1)

DEPARTMENT OF AGRICULTURAL STATISTICS

COLLEGE OF HORTICULTURE

KERALA AGRICULTURAL UNIVERSITY, VELLANIKKARA

THRISSUR, KERALA- 680656

CERTIFICATE

This is to certify that the seminar report entitled **“Optimization through Response Surface Methodology”** has been solely prepared by Shivakumar M. (2018-19-005) after going through various references cited herein under my guidance and has not been copied from seminar reports of any seniors, juniors or fellow students.

Vellanikkara
22-11-2019

Dr. Ajitha T. K.
Major advisor
Associate Professor
Dept. of Agricultural Statistics
College of Horticulture, Vellanikkara

DECLARATION

I, Shivakumar M. (2018-19-005) hereby declare that the seminar entitled **"Optimization through Response Surface Methodology"** has been prepared by me after going through various references cited at the end and I haven't copied from any of my fellow students.

Vellanikkara
22-11-2019

Shivakumar M.
(2018-19-005)

CERTIFICATE

This is to Certify that the seminar report entitled **"Optimization through Response Surface Methodology"** is a record of seminar presented by Shivakumar M. (2018-19-005) on 22-11-2019 and is submitted for the partial requirement of the course STAT 591.

Dr. Anil Kuruvila
Professor
Department of Agricultural Economics
College of Horticulture, Vellanikkara

Dr. Reshmy Vijayaraghvan
Assistant Professor
Department of Plant Pathology
College of Horticulture, Vellanikkara

Dr. Sangeetha Kuttty M.
Assistant Professor
Department of Vegetable Science
College of Horticulture, Vellanikkara

Contents

List of figures

List of tables

Introduction

Design of experiments (DOE) is a systematic approach for investigation of a system or process. It plays an important role in several areas of science and industry. It is necessary to observe the process and the operations of the system well. DOE consists of a set of experimental runs, in which each run is defined by the combination of each factor level (variables) and analysis of experiments. DOE helps to make product and processes more robust. It is a proven technique that continues to show the increasing usage in manufacturing and chemical process industries especially for fast, cost saving and accurate results. The DoE aims selection of the most suitable points where the response should be well examined (Grum and Slabe, 2004).

Designed experiments are often carried out in four phases:
planning, screening, optimization, and verification.

Factorial Design

Factorial designs are most efficient designs to study the joint effect of two or more factors on a response. The most important of these designs are the 2^k factorial designs which are widely used in research work and allow to study the joint effect of "k" factors, each at only two levels, on a response. These levels may be quantitative, such as two values of temperature or time; or they may be qualitative, such as the "high" or "low" levels of a factor. However, there is also a special case of 2^k factorial designs with "n" centre points or replicated runs added to the centre of the 2^k design, that is "n" replicated runs at medium levels of factors. When we add these centre runs, we assume that "k" factors are quantitative. Replicated runs at centre point of the design allow the experimenter to check for quadratic effects (curvature) as well as an independent estimate of error to be obtained. The first design in the 2^k series with "n" centre points is the one with two factors, each run at two levels, and usually with five replicated runs at centre point called 2^2 factorial design with five centre points. This design provides the smallest number of runs with which the joint effects of two factors on response, can be studied completely.

Response Surface Methodology – RSM

One of the most commonly used experimental designs for optimization is the response surface methodology (RSM). Because it allows evaluating the effects of multiple factors and their interactions on one or more response variables it is a useful method. The parameters that affect the process are called independent variables, while the responses are called dependent variables. The RSM investigates an appropriate approximation relationship between input and output variables and identify the optimal operating conditions for a system under study.

Response surface methodology is a collection of mathematical and statistical techniques that are useful for the modelling and analysis of problems in which a response of interest is influenced by several variables and the objective is to optimize this response (Aydar, 2018).

For example:

The levels of temperature (x_1) and pressure (x_2) to maximize the yield (y) of a process. In Figure 1 it can be observed that the different level combinations of two factors will give different values of responses. These values of responses are determined by contour lines or response contours on a plane surface, whereas in 3D it is represented by the response surface. It can be seen in the graph that, as the contour lines decreases, the value of responses increases, and at the mid-point, maximum response is achieved. Projecting the mid-point straightly on a response surface will give the maximum value of response which is known as optimum response (Croarkin and Tobias, 2012).

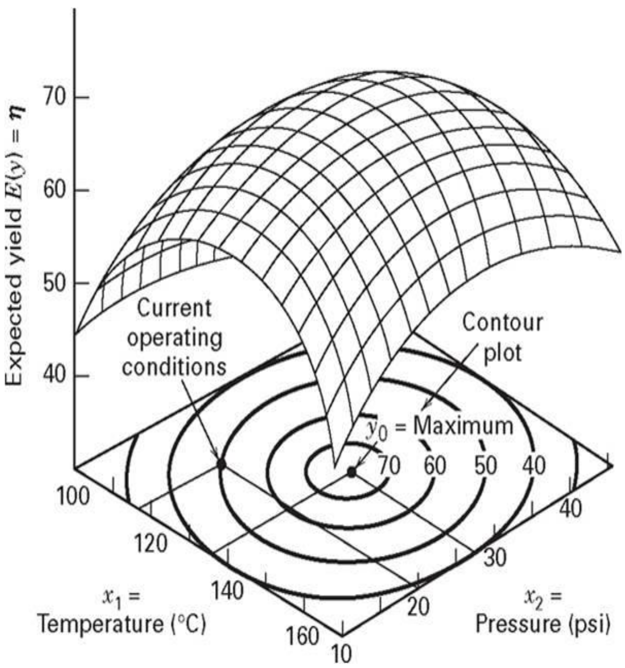

Fig 1: Response surface

Factor levels for Higher Order Designs

Figure 2 through 4 illustrate possible behaviors of responses as functions of factor settings.

If a response behaves as in Figure 2, the design matrix to quantify that behavior need the factors containing only with two levels- low and high. This model is a basic assumption of simple two-level factorial and fractional factorial designs.

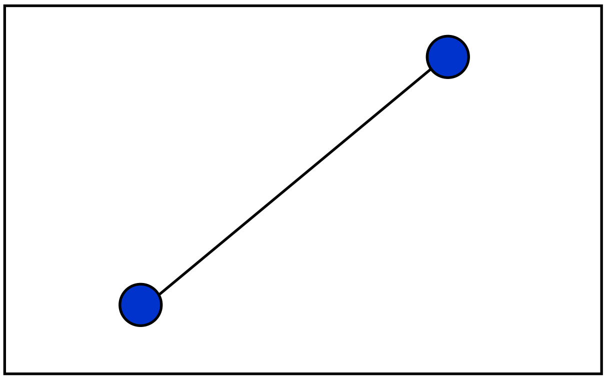

Fig 2: Linear Function

If a response behaves as in Figure 3, the minimum number of levels required for a factor to quantify that behavior is three. While a two-level design with center points cannot estimate individual pure quadratic effects, it can detect them effectively.

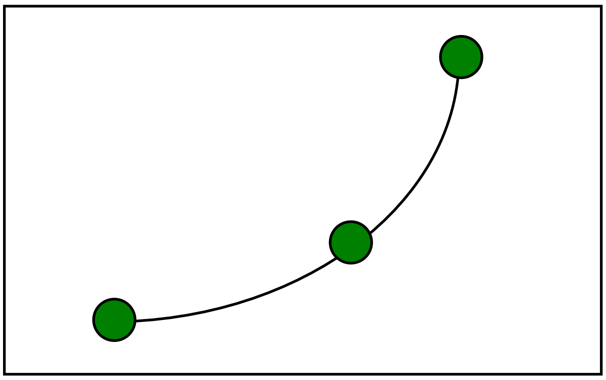

Fig 3: Quadratic Function

Finally, in more complex cases such as illustrated in Figure 4, the design matrix must contain at least four levels of each factor to characterize the behavior of the response adequately.

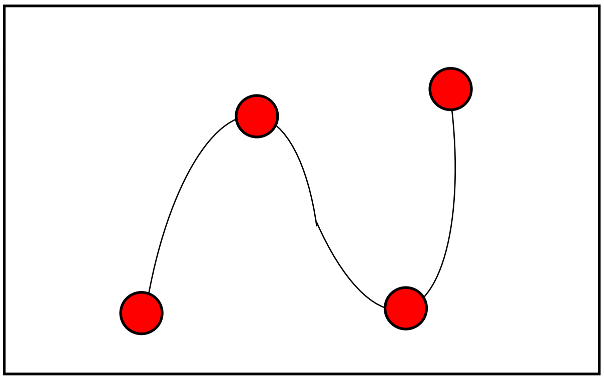

Fig 4: Cubic Function

First order response surface

Usually, a first order regression model is sufficient at the current operating conditions because the operating conditions are normally far from the optimum response settings. The experimenter needs to move from the current operating conditions to the optimum region in the most efficient way by using the minimum number of experiments. This is done using the method of steepest ascent. In this method, the contour plot of the first order model is used to decide the settings for the next experiment, in order to move towards the optimum conditions. Consider a process where the response has been found to be a function of two factors. To explore the region around the current operating conditions, the experimenter fits the following first order model between the response and the two factors.

$$y=\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

The response surface plot for the model, along with the contours, is shown in Fig. 5. It can be seen in the figure that in order to maximize the response, the most efficient direction in which to move the experiment is along the line perpendicular to the contours.

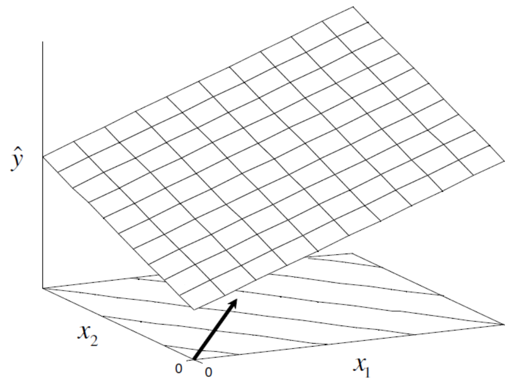

Fig 5: First order response surface

This line, also referred to as the path of steepest ascent, is the line along which the rate of increase of the response is maximum. The steps along this line to move towards the optimum region are proportional to the regression coefficients, β_j of the fitted first order model. Experiments are conducted along each step of the path of steepest ascent until an increase in the response is not seen. Then, a new first order model is fit at the region of the maximum response. If the first order model shows a lack of fit, then this indicates that the experimenter has reached the vicinity of the optimum. RSM designs are then used to explore the region thoroughly and obtain the point of the maximum response. If the first order model does not show a lack of fit, then a new path of steepest ascent is determined and the process is repeated.

Second order response surface

A second order model is generally used to approximate the response once it is realized that the experiment is close to the optimum response region where a first order model is no longer adequate. The second order model is usually sufficient for the optimum region, as third order and higher effects are seldom important. The second order regression model takes the following form for K factors

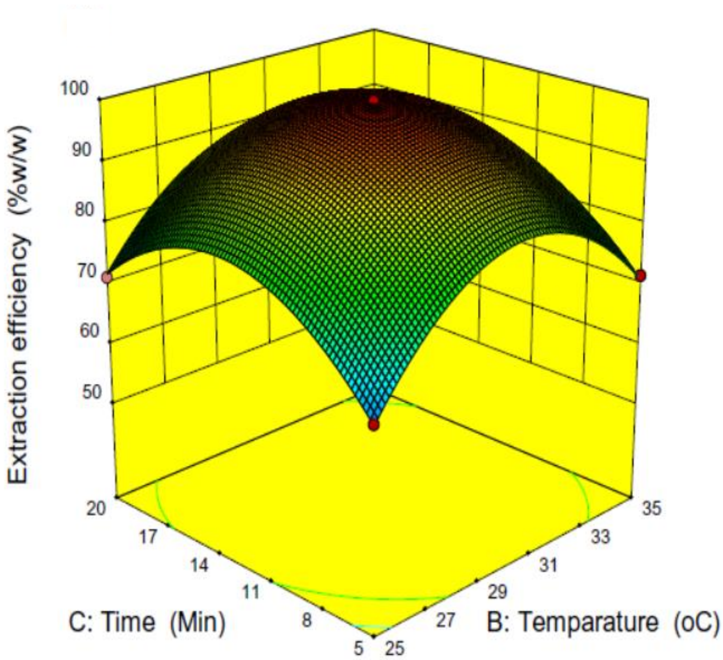

Fig 6: Second order response surface

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i<j} \sum \beta_{ij} x_i x_j + \varepsilon$$

The model contains p=(k+1)(k+2)/2 regression parameters that include coefficients for main effects (β_1,β_2...β_k), coefficients for quadratic main effects (β_{11},β_{22}...β_{kk}) and coefficients for two factor interaction effects (β_{12},β_{23}...β_{k-1k}). A full factorial design with all factors at three levels would provide estimation of all the required regression parameters. However, full factorial three level designs are expensive to use as the number of runs increases rapidly with the number of factors.

Different steps to implement RSM

a. First step: Designing and conducting first order experiment, then modelling it with linear first order regression. The factor levels used for this experiment refer to the current machine operating condition. Usually, two levels are used such as “high” and “low” values

b. Second step: Checking the response surface from the first order design. If there is any lack of-fit for the regression model, then the optimal solution has been found. Otherwise, if there is no lack-of-fit meaning that the experiment should be continued to search new factor level that can optimize the response

c. Third step: Conducting the steepest ascent (or descent) experiment. The factor levels should be shifted onto the various settings along operational machine condition that refers to the path of steepest ascent. This shifting could be stopped while the optimal response indication has been found

d. Fourth step: When the indication of optimal response is found at once, the second order design experiment would be conducted with new factor levels that have shifted from the first order design. Then, it models the second order regression. If there is no lack-of-fit in this second order, it means that the optimal response has been found

Response surface methodology can basically be categorized into two groups namely:
1)Box-Behnken design
2)Central Composite Design

1) Box-Behnken design

Box–Behnken designs are experimental designs for response surface methodology, devised by George E. P. Box and Donald Behnken in 1960.The Box-Behnken design is a spherical design with all points lying on a sphere of radius 2 shown in Figure 7. Also, this does not contain any points at the vertices of the cubic region created by the upper and lower limits for each variable. Each factor or independent variable, is placed at one of three equally spaced values. At least three levels are needed to conduct this design. This could be advantageous, when the points on the corners of the cube represent factor level combinations that are prohibitively expensive or impossible to test because of physical constraints.

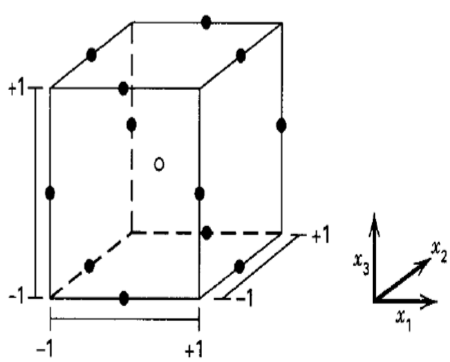

Fig 7: Box-Behnken design

2) Central composite design

A Box-Wilson central composite design, commonly called central composite design (CCD), is frequently used for building a second-order polynomial for the response variables in response surface methodology without using a complete full factorial design of experiments. To establish the coefficients of a polynomial with quadratic terms, the experimental design must have at least three levels of each factor. In CCD, there are three different points, namely factorial points, central points and axial points. Factorial points are vertices of the n-dimensional cube which are coming from the full or fractional factorial design where the factor levels are coded to –1, + 1. Central point is the point at the centre of the design space. Axial points are located on the axes of the coordinate system symmetrically with respect to the central point at a distance α from the design centre.

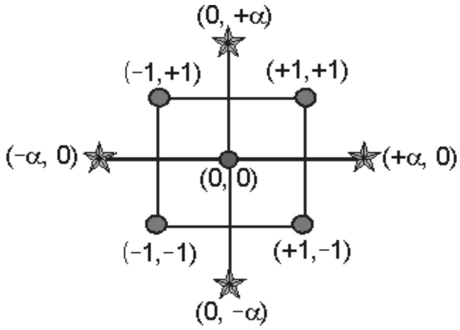

Fig 8: Central composite design

A representation of a classic central composite design for 2 factors is given in Fig.8

Four corners of the square represent the factorial (+/-) design points

Four-star points represent the axial (+/- alpha) design points

Replicated centre point

Table 1: The number of experiments for n number of factors

	n=2	n=3	n=4	n=5
Factorial points	4	8	16	32
Axial points	4	6	8	10
Centre points	5	5	6	6
Total (N)	13	19	30	48

$N = 2^n + 2\times n + n_c$

N - number of runs / experiments

n - number of factors

n_c - number of centre points the designer desire

Components of CCD

Consider an experiment that was conducted to investigate the effects of Nitrogen (kg/ha) and Azotobacter (mg) in optimizing parameters for enhancing plant growth of Pearl millet using CCD (Sunitha *et al.*, 2015).

Table 2: Components of CCD

Factors	$-\alpha$ (1.414) (axial point)	-1 (lower level)	0 (Centre point)	+1 (upper level)	$+\alpha$ (1.414) (axial point)
X_1 - (N) kg/ha	0	7.09	24	40.97	48
X_2 - (Azotobacter) mg	0	5.86	20	34.14	40

Centre point = average between the upper and lower level

To get value for the axial point we need to first get the value for α

Alpha (α) = $2^{k/4}$

For the example above where the number of factors is two

Therefore, $\alpha = 2^{2/4} = 2^{1/2} = \sqrt{2}$ which is equal to 1.414

To get value for the axial point, we apply this equation.

Axial point for Nitrogen factor = $X \pm \alpha\ (\text{Range}/2)$

$= (UL+LL)/2 \pm \alpha\ (UL-LL)/2$

$= (40.97+7.09)/2 \pm 1.414\ (40.97-7.09)/2$

$= 24.03 \pm 24$

For the upper axial point (i.e. $+\alpha$) = 24+24=48

For the lower axial point (i.e. $-\alpha$) = 24-24=0

Applying the same formula for the calculation of axial points for Azotobacter

Table 3: CCD matrix in coded values and responses

Runs	Nitrogen (kg/ha)	Azotobacter (mg)	Shoot height, cm	
			Observed	Predicted
1	-1	+1	90.15	88.18
2	+1	-1	94.93	98.43
3	-1	-1	155.24	154.20
4	+1	+1	62.3	64.86
5	+1.414	0	50.5	46.52
6	-1.414	0	100	102.44
7	0	-1.414	165	163.57
8	0	+1.414	93.25	93.14
9	0	0	140	140
10	0	0	140	140
11	0	0	140	140
12	0	0	140	140
13	0	0	140	140

Note: Runs 1 to 4 is known as factorial runs, runs 5 to 8 is known as axial runs while runs 9 to 13 is known as centre point

Case Studies

Application of Response Surface Methodology for Optimization of Urea Grafted Multiwalled Carbon Nanotubes in Enhancing Nitrogen Use Efficiency and Nitrogen Uptake by Paddy Plants

**Norazlina Mohamad Yatim, Azizah Shaaban, Mohd Fairuz Dimin,
Faridah Yusof, and Jeefferie Abd Razak**

Efficient use of urea fertilizer (UF) as important nitrogen (N) source in the world's rice production has been a concern. Carbon-based materials developed to improve UF performance still represent a great challenge to be formulated for plant nutrition. Advanced N nanocarrier is developed based on functionalized multiwall carbon nanotubes (f-MWCNTs) grafted with UF to produce urea-multiwall carbon nanotubes (UF-MWCNTs) for enhancing the nitrogen uptake (NU) and use efficiency (NUE).The grafted N can be absorbed and utilized by rice efficiently to overcome the N loss from soil-plant systems. The individual and interaction effect between the specified factors of f-MWCNTs amount (0.10–0.60 wt%) and functionalization reflux time (12–24 hrs) with the corresponding responses (NUE, NU) were structured via the Response Surface Methodology (RSM) based on five-level CCD. The UF-MWCNTs with optimized 0.5 wt% f-MWCNTs treated at 21 hrs reflux time achieve tremendous NUE up to 96% and NU at 1363.5 mg/pot. Significant model terms (p value < 0.05) for NUE and NU responses were confirmed by the ANOVA.

The weight % of f-MWCNTs and functionalization reflux time were assessed at five levels: −1.414, −1, 0, +1, and +1.414 as presented in Table 4.

Table 4: Experimental codes and levels of independent variables

Variables	Units	Levels				
		-1.414	-1	0	+1	+1.414
MWCNTs – X_1	Wt%	0.1	0.2	0.3	0.5	0.6
Reflux time – X_2	Hour	12	15	18	21	24

A total of 13 experiments (Table 5) were conducted for the optimization of two chosen design factors.

Table 5: Experimental design matrix

Run	**Factor 1** **MWCNTs (Wt%)**	**Factor 2** **Reflux time (Hours)**
1	0.6	18
2	0.3	12
3	0.3	24
4	0.3	18
5	0.3	18
6	0.3	12
7	0.5	21
8	0.2	15
9	0.2	21
10	0.1	18
11	0.3	18
12	0.5	15
13	0.3	18

Statistical Analysis and Modelling

The experimental results were fitted into the second-order polynomial regression equation and the analysis of variance (ANOVA) was conducted:

$$Y = \beta_\circ + \Sigma_i \beta_i X_i + \Sigma_{ii} \beta_{ii} X^2_{ii} + \Sigma_{ij} \beta_{ij} X_i X_j$$

where Y is the predicted response, X_i and X_j are independent variables, $\beta_\circ$ is the offset term, β_i is the i_{th} linear coefficient, β_{ii} is the i_{th} quadratic coefficient, and β_{ij} is the ij_{th} interaction coefficient.

The statistical software package, *Design Expert 9*, was used for the regression analysis of the experimental data and also, to plot the response surface graphs.

Results of CCD

The optimum conditions for fertilizer NU, and NUE of rice growth were determined by means of the Central Composite Design (CCD) of RSM. The overall results are shown in Table 5. For UF-MWCNTs fertilizer, the obtained responses varied between 71 and 96%, 978 and 1363 mg/pot for NUE, and NU respectively.

Table 6: Optimum levels of NU and NUE for rice growth

Run	Response 1 NUE	Response 2 NU
1	86.29	1182.12
2	86.16	1100.00
3	85.20	1344.75
4	85.00	1100.00
5	86.25	1181.60
6	83.85	1256.30
7	96.35	1363.55
8	71.38	977.84
9	86.50	1185.02
10	81.09	1134.47
11	86.67	1187.41
12	75.46	1087.31
13	90.00	1189.00

Table 7: The modelling equations, in terms of both coded and actual factors with respect to two responses

Responses	Coded factors	Actual factors
Nitrogen use efficiency (NUE) Y_1	$Y_1 = 86.4 + 4.83x_1 + 4.44x_2 + 2.79x_1x_2 - 3.39x_1^2 - 1.724x_2^2$	$Y_1 = 7.23 - 8.92x_1 + 7.06x_2 + 4.65x_1x_2 - 84.64x_1^2 - 0.19x_2^2$
Nitrogen uptake (NU) Y_2	$Y_2 = 1181.23 + 81.47x_1 + 50.81x_2 + 29.90x_1x_2 - 58.94x_1^2 + 28.56x_2^2$	$Y_2 = 1918.97 + 394.43x_1 - 112.27x_2 + 49.83x_1x_2 - 1473.60x_1^2 + 3.17x_2^2$

The analysis of variance for response surface quadratic model of NUE to test its adequacy is shown in Table 8.

Table 8: ANOVA for NUE

Source	Sum of squares	Degree of freedom	Mean square	F- value	P- value
Model	462.90	5	92.58	10.32	0.0040
X_1	119.67	1	119.67	13.34	0.0082
X_2	236.74	1	236.74	26.39	0.0013
X_1X_2	31.25	1	31.25	3.48	0.1042
X_1^2	52.66	1	52.66	5.87	0.0459
X_2^2	73.80	1	73.80	8.23	0.0241
Residual	62.80	7	8.97		
Lack-of-Fit	62.80	3	20.93		
Pure error	0.000	4	0.000		
Corrected Total	525.70	12			

The coefficient of determination value (R^2 = 0.88) indicates that the response model can explain about 88% of the total variations in the response variable

The adjusted coefficient of determination value (Adj R^2 = 0.80) was also high enough to indicate the significance of the selected model.

The model F-value of 10.32 implies that the model is significant. It means that there is only about 0.40% chance that the F-value, this large could occur due to noise.

p values less than 0.05 indicates that the effect of those factors are significant. In this case the individual effects of X_1, X_2, X_{21}, and X_{22} are significant However, insignificant p value was found for the interaction (X_1X_2) between the factors. Even an insignificant p value does not automatically specify that the particular research project has "failed to disprove the null hypothesis", but here, in this case, conclusions should only be made after consideration of other significant factors and primary objectives of the study.

Hence, the amount of f-MWCNTs (X_1) and the functionalization reflux time (X_2) had a significant positive effect individually on NUE response by paddy. NUE response increased with increasing amount of f-MWCNTs (X_1) and the functionalization reflux time.

Furthermore, the ANOVA for response surface quadratic model of NU to test its adequacy is presented in Table 9.

Table 9: ANOVA for NU

Source	**Sum of squares**	**Degree of freedom**	**Mean square**	**F- value**	**P- value**
Model	93968.46	5	18793.69	5.18	0.026
X_1	34088.30	1	34088.30	9.39	0.018
X_2	30984.95	1	30984.95	8.54	0.022
X_1X_2	3576.64	1	3576.64	0.99	0.353
X_1^2	15960.55	1	15960.55	4.40	0.074
X_2^2	19780.29	1	19780.29	5.45	0.052
Residual	25401.24	7	3628.75		
Lack-of-Fit	24231.77	3	8077.26	27.63	0.003
Pure error	11169.47	4	292.37		
Corrected Total	1.194E + 005	12			

Similar to NUE, the proposed model suggests that the amount of f- MWCNTs (X_1) and functionalization reflux time (X_2) had a significant effect on NU by paddy treated with UF-MWCNTs.

Here, the coefficient of determination value (R^2 = 0.79) indicates that the response model can explain about 79% of the total variations which is lower than that for NUE response.

Additionally, the adjusted coefficient of determination value (Adj R^2 = 0.64) was also high enough to indicate the significance of the model.

The model F-value of 5.18 implies that the model is significant since p is < .05. There is only a 2.63% chance that F-value could occur due to noise.

In comparison with NUE, in this case X_1 and X_2 only are significant model terms. Individually, the increase in amount of f-MWCNTs and the functionalization reflux time positively increase the NU response by paddy.

Optimization of responses of paddy under UF-MWCCNTs fertilizer treatment

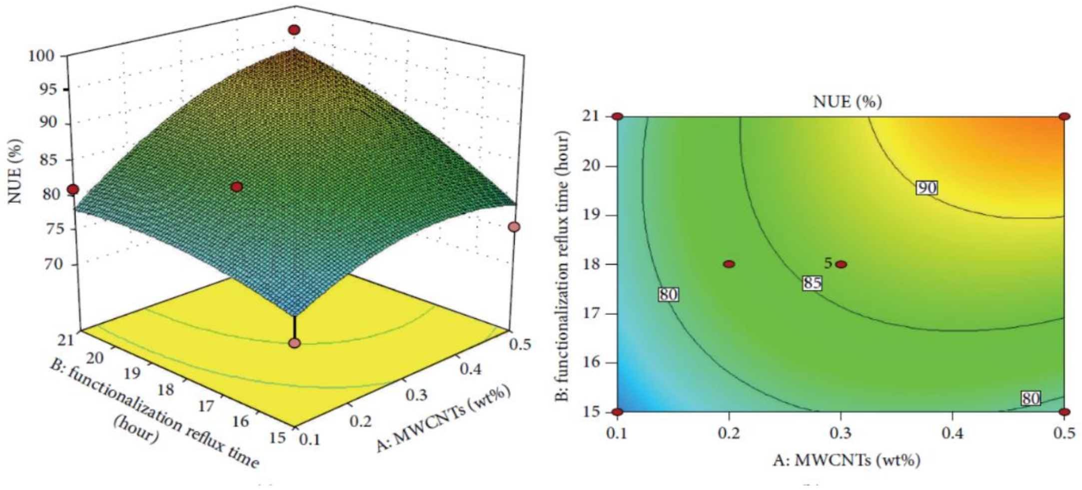

Fig 9: 3D Response surface plot of nitrogen use efficiency

The UF-MWCNTs with optimized 0.5 wt% f-MWCNTs treated at 21 hrs reflux time achieve tremendous NUE up to 96%.

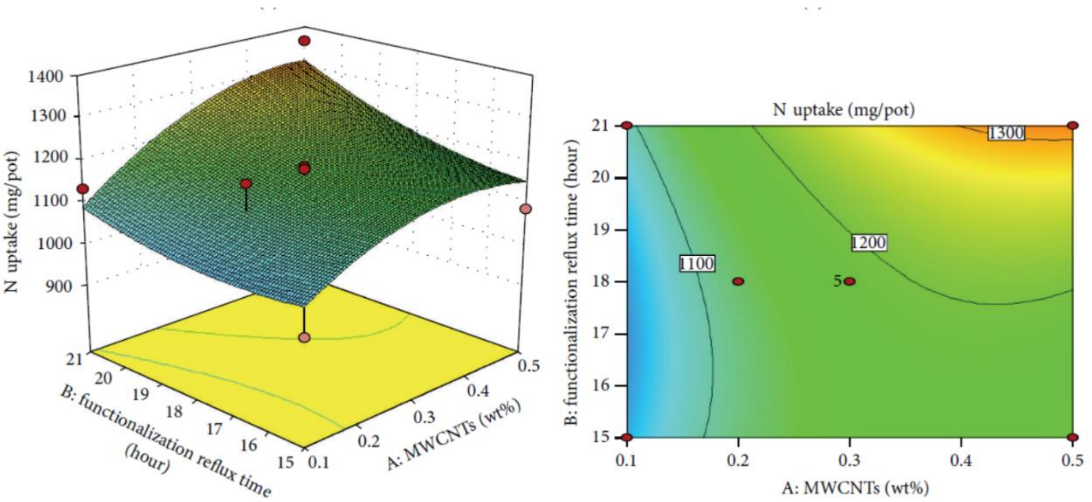

Fig 10: 3D Response surface plot of nitrogen uptake

The 0.5 wt% of MWCNTs treated at 21 hrs reflux time achieve more nitrogen uptake of 1363.5 mg/pot

OPTIMIZATION OF BREAD BAKING CONDITIONS IN SUPERHEATED STEAM OVEN USING RESPONSE SURFACE METHODOLOGY

RASHA MUSA OSMAN, TAJUL ARIS YANG, MUDAWI HASSAN ALI AND SAIFELDIN MOHAMED KHAIR

The objective of the study was to optimize the bread baking condition in superheated steam oven. Independent variables were the baking time (20, 25, and 30 min.) and the baking temperature (180°, 200° and 220°C). The bread quality parameters including moisture content, colour and texture properties were measured. The response surface methodology was used for the optimization. The effect of baking condition on the parameters of bread were investigated using second-order central composite design. Baking temperature and time significantly affect moisture content and colour of bread. Numerical optimization and superimposed contour plots suggested the optimum baking condition of bread to be 180 °C (temperature) and 20.77 minutes (time). The optimum moisture content, L* and hardness value are predicted to be 38.52 %, 76.24 and 13.26 N of the baked bread respectively. Baking bread in these conditions produce high quality bread in terms of moisture content, colour and texture properties.

The experimental design (the coded and actual values) is presented in Table 10.

Table 10: CCD to extract optimal parameters for Bread baking

Runs	**Temperature (°C, X_1) Coded**	**Actual**	**Time (min, X_2) Coded**	**Actual**
1	0	200	0	25
2	1	200	0	30
3	0	200	0	25
4	0	180	-1	25
5	1	220	1	30
6	1	180	-1	30
7	0	200	0	25
8	0	200	0	25
9	-1	180	-1	20
10	-1	200	0	20
11	0	200	0	25
12	0	220	1	25
13	-1	220	1	20

The experimental values of moisture content (MC), colour values (L*, a*, and b*), textural properties of baked bread are given in Table 11

Table 11: Experimental values of different components in bread baking

Runs	Moisture content (%)	L*	a*	b*	Hardness (N)
1	37.25	72.49	2.29	27.20	14.32
2	34.42	70.29	5.76	3.09	13.48
3	37.25	72.49	2.29	27.20	14.32
4	38.27	75.21	0.73	24.27	13.11
5	32.67	63.76	10.27	32.82	13.33
6	35.96	72.09	3.83	28.58	14.42
7	37.25	72.49	2.29	27.20	14.32
8	37.25	72.49	2.29	27.20	14.32
9	38.39	76.25	0.74	23.91	13.34
10	38.72	74.71	2.37	26.85	14.28
11	37.25	72.49	2.29	27.20	14.32
12	37.24	64.13	8.88	33.01	14.28
13	38.77	69.08	4.74	28.60	14.54

The L*, a* and b* colour values stand for lightness, redness and yellowness of the colour components respectively.

Table 12: ANOVA of (p- value) for Moisture content (%), L*, a* and b* colour values and Hardness

Source	Moisture Content (%)	L*	a*	b*	Hardness (N)
Model	<0.0001**	<0.0001**	0.0002**	0.0015**	0.0697
$X_1(^0C)$	0.0002**	<0.0001**	<0.0001**	0.0003**	0.1897
X_2(min)	<0.001**	0.0001**	0.0007**	0.0024**	0.3239
X_1^2	0.1763	0.0009**	0.0129*	0.2849	0.1879
X_2^2	0.0001**	0.5175	0.1035	0.4001	0.5719
X_1X_2	<0.0001**	0.4643	0.1965	0.8410	0.0151*

*Statistically significant at p<0.05 ** Statistically significant at p<0.01

Table 12 shows the analysis of variance of (p-value) result for the tested quality parameters of the quadratic model of response surface methodology. Statistical parameters extracted from ANOVA are presented in Table 12.

The adjusted R^2 were 0.98, 0.96, 0.92 and 0.85 for moisture content, L*, a* and b* colour values respectively. The R^2 were 0.99 for moisture content, 0.98 for L*, 0.95 for a* colour value and 0.91 for b* value in the equation and lack of fit being insignificant ($p>0.05$). The CV values obtained from the statistical analysis were less than 10%.

Numerical optimization and superimposed contour plots suggested the optimum baking condition of bread to be 180 °C for temperature and 20.77 minutes for time.

The optimum moisture content is 38.52 %

The color value L* is 6.24

Hardness value is 13.26 N

Baking bread in these conditions produce high quality bread in terms of moisture content, color and texture properties.

Conclusion

Agricultural experiments are key undertaking in the development of most countries. Over the years the growth of scientific agricultural experiments has transformed the agricultural sector. The development of various forms of scientific experiments on crop improvement has led to considerable advancement in finding the optimal agricultural conditions for crop production. Seeking the optimal input and output setting for crop production is a critical experimental undertaking that solves the food insecurity and hence the decreasing incomes to small scale farmers. Such experimental agenda, give experimenter adequate range of treatment factors (fertilizer, seed among others) to balance for optimum response (yield). This has also been motivated by the need to produce more food to feed the increasing world population and also to strike a balance between producing more food and the food safety concept. In statistics, optimization is the process of making a design as functional and effective as possible. The search for optimal settings in crop production naturally presents the need for more than one treatment factor to be optimized (minimized/maximized).

Response surface methodology is found to be an ideal and time saving approach which overcomes the problems faced in conventional method of optimization. In the present scenario, it is a widely used statistical tool applied in food industry, agriculture, biotechnology, microbiology, genetics, pharmacy *etc.*

References

Aydar, A. Y., 2018. Utilization of Response Surface Methodology in Optimization of Extraction of Plant Materials. Statistical Approaches with Emphasis on Design of Experiments Applied to Chemical Processes. doi:10.5772/intechopen.73690

Croarkin, C. and Tobias, P. 2012. *Engineering Statistics Handbook*. NIST, SEMATECH. U. Department of Commerce. 2333p.

Grum, J. and Slabe, J. M., 2004. The use of factorial design and response surface methodology for fast determination of optimal heat treatment conditions of different Ni–Co–Mo surfaced layers. *Journal of Materials Processing Technology*, 155: 2026-2032.

Osman, R. M., Yang, T. A., Ahmed, A. H. R., Ahmed, K. E., Ali, M. H., Khair, S. M. 2017. Optimization of bread baking conditions in superheated steam oven using response surface methodology. *Int. J. Agri & Env. Res.* 3(3): 290-301.

Sunitha, P., Lakshmi, K. A. and Saibaba, K. N., 2015. Application of response surface methodology for the optimization of growth of pearl millet. *British Micro. Res. J.* 6(3): 154-166.

Yatim, N., Shaaban, A., Dimin, M. F., Yusof, F., and Abd Razak, J. 2016. Application of response surface methodology for optimization of urea grafted multi walled carbon nanotubes in enhancing nitrogen use efficiency and nitrogen uptake by paddy plants. *J. Nanotechnol.* 2016: 14.

Discussions

1. What is Lack of Fit?

Ans. A regression model exhibits lack of fit when it fails to adequately describe the functional relationship between the experimental and the response variable.
Lack of Fit can occur if important terms from the model such as interactions or quadratic terms are not included. The sum of square due to lack of fit is the difference between the total error and pure error sum of square.

2. Why the Adjusted R^2 is preferred over R^2?

Ans. The value of R^2 would not decrease when more variables are added to the model. As a result, there is always a temptation to add more variables in the model, because of which problems such as insignificant regression coefficients etc. can arise therefore in that case adjusted R^2 is used as a measure of predictability The adjusted R^2 takes into account the number of independent variables in the model. As a general thumb rule if adjusted R^2 increases when a new variable is added to the model, the variable should remain in the model. If adjusted R^2 decreases when the new variable is added then the variable should not remain in the model. Other factors may also be considered. Since adjusted R^2 has been adjusted for the number of predictors in the model it can be well utilised for comparing the relative efficiency of different models each with varying number of predictor variables.

3. What is the formula for R^2 and adjusted R^2?

Ans. The formula for R^2 is given by, $R^2 = 1 - \dfrac{S.S_{Residual}}{S.S_{Total}} = 1 - \sum (Y_e - \overline{Y}^2) \Big/ 1 - \sum (Y - \overline{Y}^2)$

While the formula for adjusted R^2 is given by, $R^2_{adj} = 1 - \dfrac{M.S_{Residual}}{M.S_{Total}}$

$= 1 - \dfrac{S.S_{Residual}/(n-k)}{S.S_{Total}/(n-1)}$ n = Total no of observations
k = no of independent variables

4. How do you interpret the optimum response in 3D response surface plot?

Ans. The graphical representation of response for the different level combination of factors is known as response surface plot. At any point on the 3D response surface will give the response, obtained for the respective level combination of factors. The peak point on the graph showing the maximum response is the optimum response obtained for the respective level combination of the factors if our aim is to maximise the response. The 3D response surface plot shows the nature of the response obtained.

5. What is the relationship between the residual sum of square, lack of fit and pure error sum of square?

Ans. The breakdown of residual error ("error sum of squares –SSE) into two components, one component is due to lack of model fit ("lack of fit sum of squares" – SSLF) and the other component is due to pure random error ("pure error sum of squares" – SSPE).

The formula is: SSE = SSLF+SSPE

$$\sum_{i=1}^{c}\sum_{j=1}^{n_i}\left(y_{ij}-\hat{y}_{ij}\right)^2=\sum_{i=1}^{c}\sum_{j=1}^{n_i}\left(\bar{y}_i-\hat{y}_{ij}\right)^2+\sum_{i=1}^{c}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}_i\right)^2$$

6. Which is the conventional method used in optimization of response?

Ans. Canonical analysis technique is the conventional method of approach to determine the optimal conditions for obtaining the optimum response. Analysis of the eigen structure of response surfaces is known as the canonical analysis of response surface model.
Canonical correlation analysis is a multivariate statistical tool that facilitates the study of interrelationships among the sets of multiple dependent variables and multiple independent variables.
A canonical analysis facilitates interpretation of the response surfaces by examining the overall shape of the curve and establishing whether the estimated stationary point is a maximum, a minimum or a saddle point.
Canonical analysis can be used to ascertain
I. the shape of the surface (a hill, a valley, a saddle surface or a flat surface)
II. whether there is a unique optimum combination of factor values
III. to identify which factor or factors represent the most sensitive predicted response

KERALA AGRICULTURAL UNIVERSITY

COLLEGE OF HORTICULTURE, VELLANIKKARA

Department of Agricultural Statistics STAT-591: Masters Seminar

Name : Shivakumar M. **Venue** : Seminar hall

Admission no. : 2018-19-005 **Date** : 22-11-2019

Major Advisor : Dr. Ajitha T. K. **Time** : 9:15 am.

Optimization through Response Surface Methodology

Abstract

In India, agriculture sector is the main stream of economic development. It is the primary source of livelihood for about 58% of India's population. However, still the agriculture sector is faced with several challenges that needs to be solved. Among them one such is the increase in the cost of cultivation due to insignificant use of inputs and decline in the yield.

Experimental design deals with planning, conducting, analyzing and interpreting the results of controlled experiments. If several factors affect simultaneously the characteristic under study, we adopt factorial experiments which involve the study of main effects and the interaction effects among different factors. But the simple factorial experiments becomes inadequate when we want to assess the possible effects of the intervening levels of the factors or their combinations which were not tried in the actual conduct of the experiment.

The response surface methodology (RSM) is an advanced technique used to optimize the response using input factors which are quantitative in nature (Grum and Slabe, 2004). It combines Design of Experiment, Regression analysis and Optimization methods. The polynomials which adequately represent the true independent variable-response relationship are called response surfaces and the design that allow the fitting of response surfaces are called response surface designs. To estimate a first-order response surface a factorial or a fractional factorial experiment can be used. On the other hand to estimate a second order response surface the most widely used tool is the central composite design (Croarkin and Tobias,2012).

Once the second order response surface model is fitted using CCD and the ANOVA is constructed, the significant independent variables can be identified and a second order polynomial regression equation is fitted to predict the dependent variable. In order to verify the model adequacy, the coefficient of determination (R^2), adjusted R^2 *etc.* are noted. A low value of the coefficient of variation indicate good precision and reliability of the model. The computations can be done by using statistical software packages "Design Expert" or "SAS" (Peng *et al.*, 2015). The aim of experiment is to determine the optimal levels of independent variables that gives the Optimum response.

Yatim *et al.* (2016) applied response surface methodology to determine the optimum operating condition for urea fertilizer grafted multiwalled carbon nanotubes (UF-MWCNT) to enhance nitrogen use efficiency (NUE) and nitrogen uptake (NU) by paddy plants. The individual and interaction effect between the specified factors of functionalized MWCNTs amount (0.10-0.60 wt.%) and functionalization reflux time (12-24 hrs.) with the corresponding responses (NUE, NU) were structured via the RSM based on five-level CCD. The UF-MWCNTs with optimized 0.5 wt.% f- MWCNTs treated at 21 hrs. reflux time achieved tremendous NUE up to 96% and NU at 1363 mg/pot.

Osman *et al.* (2017) optimized the conditions for bread baking using response surface methodology taking baking time (20, 25, and 30 min) and the baking temperature (180°, 200° and 220°C) as independent variables. The bread quality parameters including moisture content, color and texture properties were measured. The effect of baking condition on the parameters of bread were investigated using second-order central composite design. The optimum baking condition of bread were determined to be 180°C and 20.77 minutes. The optimum moisture content, color and hardness value were predicted to be 38.52 %, 76.24 and 13.26 N of the baked bread respectively.

Response surface methodology is found to be an ideal and time saving approach which overcomes the problems faced in conventional method of optimization and it is an important tool to optimize the process parameters for improving the final product quality. RSM is the most widely used statistical optimization tool applied in food industry, agriculture, biotechnology, microbiology, genetics *etc.*

References

Croarkin, C. and Tobias, P. 2012. *Engineering Statistics Handbook*. NIST, SEMATECH. U. Department of Commerce. 2333p.

Grum, J. and Slabe, J.M., 2004. The use of factorial design and response surface methodology for fast determination of optimal heat treatment conditions of different Ni–Co–Mo surfaced layers. *J. Mater. Process. Technol.* 155: 2026-2032.

Osman, R. M., Yang, T. A., Ahmed, A. H. R., Ahmed, K. E., Ali, M. H., Khair, S. M. 2017. Optimization of bread baking conditions in superheated steam oven using response surface methodology. *Int. J. Agri & Env. Res.* 3(3): 290-301.

Peng, B., Lei, Y., Zhao, H., and Cui, L., 2015. Response surface methodology for optimization of fermentation process parameters for improving apple wine quality. *J. Food Sci. Technol.* 52(11): 7513-7518.

Yatim, N., Shaaban, A., Dimin, M. F., Yusof, F., and Abd Razak, J. 2016. Application of response surface methodology for optimization of urea grafted multi walled carbon nanotubes in enhancing nitrogen use efficiency and nitrogen uptake by paddy plants. *J. Nanotechnol.* 2016: 14.