Chapter XIX

Response Surface Methodology: Application and Case Studies

Eldho Varghese[1] and Hemavathi M[2]

[1]ICAR-Central Marine Fisheries Research Institute
[2]Institute of Agricultural Sciences, Banaras Hindu University
Email: eldho.varghese@icar.gov.in

Response Surface Methodology (RSM) is used to examine this relationship between one or more response variables and a set of experimental variables or factors. These methods are often employed after one has identified a “vital few” controllable factor and the factor settings that optimize the response are to be found. Designs of this type are usually chosen when a curvature in the response surface is suspected.

RSM is thus a set of techniques that includes the following:

- Setting up an experiment (designing an experiment) that will yield adequate and reliable estimates of the response of interest.
- Determining a model that best fits the data collected from the design chosen by conducting appropriate tests of hypotheses concerning the model’s parameters.
- Determining the optimal settings of the experimental factors that produce the maximum (or minimum) value of the response.

For example, RSM can be used to determine the optimum dose of Nitrogen (N), Phosphorous (P) and Potash (K) for a crop to be recommended to farmers.

The response surface methodology was employed by many researchers in the past in agricultural field. Some of the applications are as follows;

- Evaluation of enzyme and microwave-assisted conditions on extraction of anthocyanins and total phenolics from black soybean (*Glycine max L.*) (Kumar *et al.*, 2019).

- Valorisation of black carrot pomace: microwave assisted extraction of bioactive phytoceuticals and antioxidant activity using Box–Behnken design. (Kumar *et al.*, 2019).
- Ultrasound-assisted development of stable grapefruit peel polyphenolic nano-emulsion: Optimization and application in improving oxidative stability of mustard oil. Nishad *et al*., (2020)
- Optimization of enzyme assisted extraction of flavonoids from black carrot (*Daucus carota ssp.)* using Response Surface Methodology. (Kumar *et al*., 2016).
- Black Carrot (*Daucus carota ssp*.) and Black Soybean (*Glycine max L*.) Merr.) Anthocyanin Extract: A Remedy to Enhance Stability and Functionality of Fruit Juices by Copigmentation. Waste and Biomass (Kumar *et al*., 2018)
- Valorization of Black carrot marc: Antioxidant properties and enzyme assisted extraction of flavonoids. (Kumar *et al*., 2018)
- Augmenting pentose utilization and ethanol production of native *Saccharomyces cerevisiae* LN using medium engineering and response surface methodology (Sharma *et al*., 2018)

Example 1: The over use of nitrogen (N) relative to Phosphorus (P) and Potassium (K) concerns both from agronomic and environmental perspective. Phosphatic and Potassic fertilizers have been in short supply and farmers have been more steadily adopting the use of nitrogenous fertilizers because of the impressive virtual response. There is evidence that soil P and K levels are declining. The technique of obtaining individual optimum doses for N, P and K through separate response curves may also be responsible for unbalanced fertilizer use. Hence, determining the optimum and balanced dose of N, P and K for different crops has been an important issue. This optimum and balanced dose should be recommended to farmers in terms of doses from the different sources and not in terms of N, P and K alone, as the optimum combination may vary from source to source. However, in actual practice the values of N, P and K are given in terms of kg/ha rather than the combined doses along with the sources of the fertilizers.

Example 2: For value addition to the agriculture produce, food processing experiments is being conducted. In these experiments, the major objective of the experimenter is to obtain the optimum combination of levels of several factors that are required for the product. To be, specific, suppose that an experiment related to osmotic dehydration of the banana slices is to be conducted to obtain the optimum combination of levels of concentration of sugar solution, solution to sample ratio and temperature of osmosis. The levels of the various factors are the following.

	Factors	**Levels**
1.	Concentration of sugar solution	40%, 50%, 60%, 70% and 80%
2.	Solution to sample ratio	1:1, 3:1, 5:1, 7:1 and 9:1
3.	Temperature of osmosis	25^0C, 35^0C, 45^0C, 55^0C and 65^0C

In this situation, response surface designs for 3 factors each at five equispaced levels can be used.

Example 3: In fish hatchery culture techniques, the survival and quality of larvae is very important to increase production. The temperature and salinity are the primary environmental factors that affect growth, survival and immune function of fish larvae. Early stages of development are the most sensitive phase in life cycle of fish and to maximize their survival, larvae should be reared close to optimum condition. For culturing Tilapia, the larvae are reared at temperature ranged from 16-37 ˚C and salinity ranged from 0 to 20%. Suppose larvae acclimated to five different salinities levels 0, 2.9, 10, 17.1 and 20 % (by increasing or lowering salinity at a rate of 2 % h^{-1}) and five different temperatures (16, 19.1, 26.5, 33.9 and 37˚C) at a rate of 2˚C h^{-1}. In such situation, to determine the optimum culture condition that has greater specific growth rate (SGR, %) a two factor Central Composite Design (CCD) can be preferred.

Example 4: In biological and biomedical industries, the production of high yields of soluble recombinant protein is one of the main objectives. Media composition and culture conditions have a significant influence on recombinant protein expression levels. The different concentration of glucose, glycine, KH_2PO_4, trace elements, vitamins, EDTA, histidine concentration affects the production of recombinant human protein. RSM facilitates screening the significant factor followed by identification of optimal factor level for maximum yield. After screening, four significant factors are selected and experiment the composition each at three level as glucose (20, 50, 80 g/l), glycine (2, 11, 20 g/l), KH_2PO_4 (20, 40, 60 g/l) and histidine (0.2, 0.6, 1 g/l). In such cases Box-Behnken Design can be used to find optimum factor level that enhance the production of recombinant protein (mg/l).

Example 5: Optimization is an essential tool in food industry for the efficient operation of processing systems and unit processes yielding a highly acceptable product. The process includes drying, extraction, roasting, drying, blanching, enzymatic hydrolysis, clarification and formulation. Roasting is the most popular method used to improve and alter quality (flavor, colour, aroma, texture, appearance or physiochemical properties) of agro product or snack product (Sweet potato, peanuts, coffee etc.,) to extend the shelf-life of foods, and to improve the processing efficiency of subsequent treatment. To determine the optimal roasting temperature and time for preparing a coffee-like beverage with high yield (g/g) RSM technique is employed. The experiment with five levels of temperature (160, 180, 200, 220, 240 °C) and time (10, 20, 30, 40, 50 min) considers two factor Central Composite Design for product optimization.

Let there be v independent input/ experimental variables/ factors denoted by $x_1, x_2, ..., x_v$ and a response variable y and there are N observations. The response is a function of input factors, *i.e.*,

$$y_u = f(x_{1u}, x_{2u}, x_{3u}, ..., x_{vu}) + e_u \qquad (1)$$

where u=1,2,…,N, x_{iu} is the level of the i^{th} $(i = 1,2,...,v)$ factor in the u^{th} treatment combination, y_u denotes the response obtained from u^{th} treatment combination. The function f describes the form in which the response and the input variables are related and e_u is the random error associated with the u^{th} observation that is independently and normally distributed with mean zero and common variance σ^2.

In practice, the form of f is not known and it is therefore approximated, within the experimental region, by a polynomial of suitable degree in variables. Polynomials which adequately represent the true dose-response relationship are called response surface models and the designs that allow the fitting of response surfaces and provide a measure for testing their adequacy are called response surface designs.

If the function f is a polynomial of degree one, it is called a first order (linear) response surface *i.e.*,

$$f(x_u) = \beta_0 + \sum_{i=1}^{v} \beta_i x_{iu} + e_u . \qquad (2)$$

There are situations where a second degree response surface is more suitable. Such a surface in general is written as follows:

$$f(x_u) = \beta_0 + \sum_{i=1}^{v} \beta_i x_{iu} + \sum_{i=1}^{v} \beta_{ii} x_{iu}^2 + \sum_{i=1}^{v-1} \sum_{i'=i+1}^{v} \beta_{ii'} x_{iu} x_{i'u} + e_u \qquad (3)$$

β_0 is a constant, β_i is the i^{th} linear regression coefficient, β_{ii} is the i^{th} quadratic regression coefficient and $\beta_{ii'}$ is the $(i,i')^{th}$ interaction coefficient.

In many response surface problems, the experimenter is interested in predicting the response or estimating the mean response at a particular point in the variable space. The variance of the prediction is also of interest, because this is a direct measure of the likely error associated with the point estimate produced by the model. The variance of the estimate of the mean response at the point $\mathbf{x}_0$ is denoted by $Var[\hat{y}(\mathbf{x}_0)]$. If the variance is same for all points $\mathbf{x}_0$ that are the same distance from the center of the design, the design property is called **rotatability**.

Stages in RSM

- Fix the Objective
- Screening phase/Screening Experiment
- Regression modeling
- Experimentation
- Model building and validation
- Ooptimization of response
- Verification of results

First Order Response Designs

First order designs are mainly used for screening purposes. Factorial designs are widely used in experiments involving several factors where it is necessary to investigate the joint effects (main effects and interactions) of the factors on a response variable. A very important special case of the factorial design is that where each of the v factors of interest has only two levels. Because each replicate of such design has exactly 2^v experimental trials or runs, these designs are usually called 2^v factorial designs. The class of 2^v factorial designs are very important in response surface work. Specifically, they find applications in three areas:

- The 2^v design (or a fraction of it) is useful at the start of a response surface study where screening experiments should be performed to identify the important process or system variables.
- A 2^v design is often used to fit a first-order response surface model and to generate the factor effect estimates.
- A 2^v design is a basic building block used to create other higher order response surface designs. For example, augmenting a 2^v design with axial runs, a central composite design is obtained which is one of the most important designs for fitting second-order response surface models.

Other first order designs are (i) Fractional replicates of the 2^v factorial, (ii) Simplex designs, (iii) Plackett-Burman designs, (iv) Definitive Screening Designs (DSD) and (v) Custom Design (CD)

Popular Second Order Response Surface Designs

Central composite design (CCD) is the most popular class of second-order designs. It was introduced by Box and Wilson (1992). Much of the motivation of the CCD evolves from its use in sequential experimentation. It involves the use of a two-level factorial or fraction (resolution V) combined with the following 2v axial or star points and some central points.

$$\begin{array}{cccccc}
x_1 & x_2 & . & . & . & x_v \\
-\alpha & 0 & . & . & . & 0 \\
\alpha & 0 & . & . & . & 0 \\
0 & -\alpha & . & . & . & 0 \\
0 & \alpha & . & . & . & 0 \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & .
\end{array}$$

0	0	.	.	.	$-\alpha$
0	0	.	.	.	α

The design involves, F factorial points, 2v axial points and n_c center runs. The factorial points represent a variance optimal design for a first order model or a first order plus two-factor interaction type model. Center runs clearly provide information about the existence of curvature in the system. If curvature is found in the system, the additional of axial points allow for efficient estimation of the pure quadratic terms.

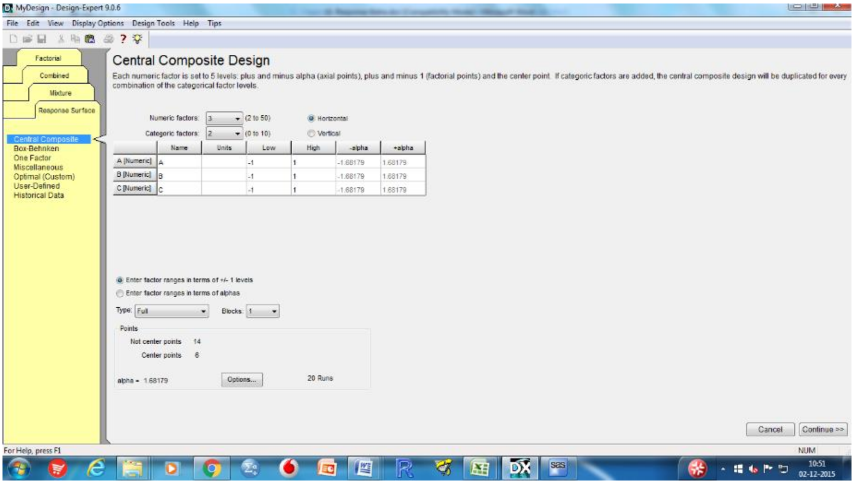
Box and Behnken (1960) developed a family of efficient three level designs for fitting second order response surfaces. The class of designs is based on the construction of Balanced Incomplete Block (BIB) designs. In many RSM situations the study is too large to allow all runs to be made under homogeneous conditions. As a result, it is important and interesting to consider second-order designs that facilitate blocking- that is, the inclusion of block effects. It is important that the assignment of the design points to block be done so as to minimize the impact on the model coefficients. Orthogonal blocking is the property that one seeks and the property implies that the block effects in the model are orthogonal to the model coefficients.

Even though there are several response surface designs available we are restricting the discussion on the generation of second order designs *viz.*, Central Composite Design (CCD) and Box Behnken Design (BBD) using Design Expert Software.

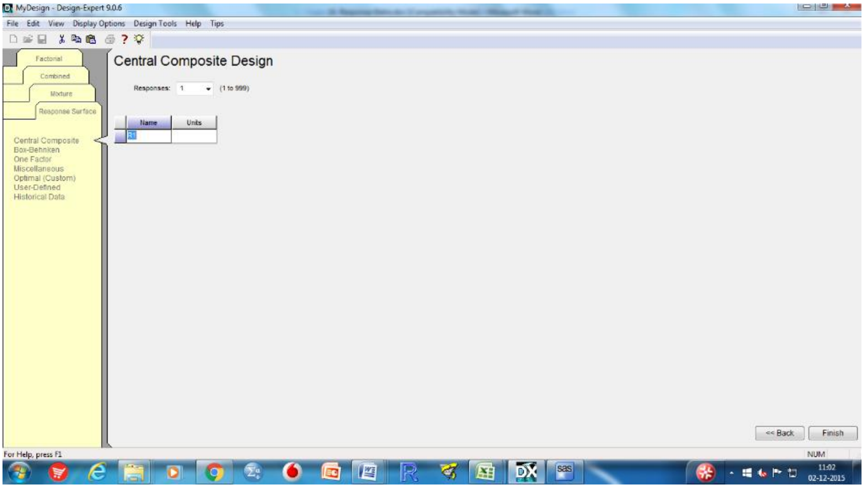
3. Construction of Second Order Response Surface Designs using Design Expert Software

(i) Construction of CCD

Go to main menu and click Central Composite under response surface tab

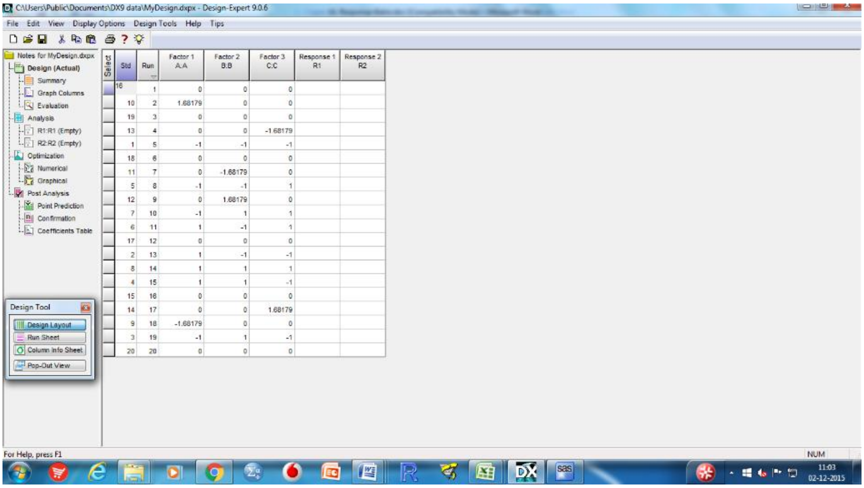


Enter the number of factors and continue



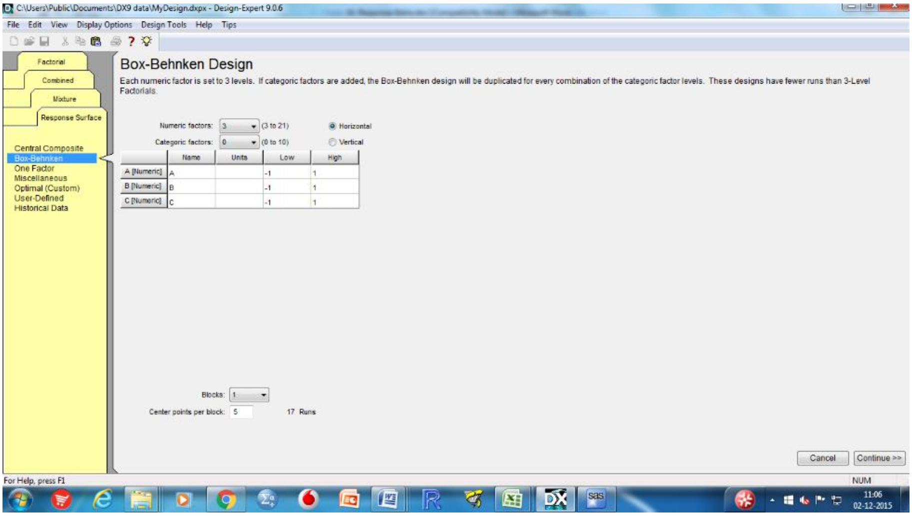
Select the number of response variables and continue

Layout of the design as follows

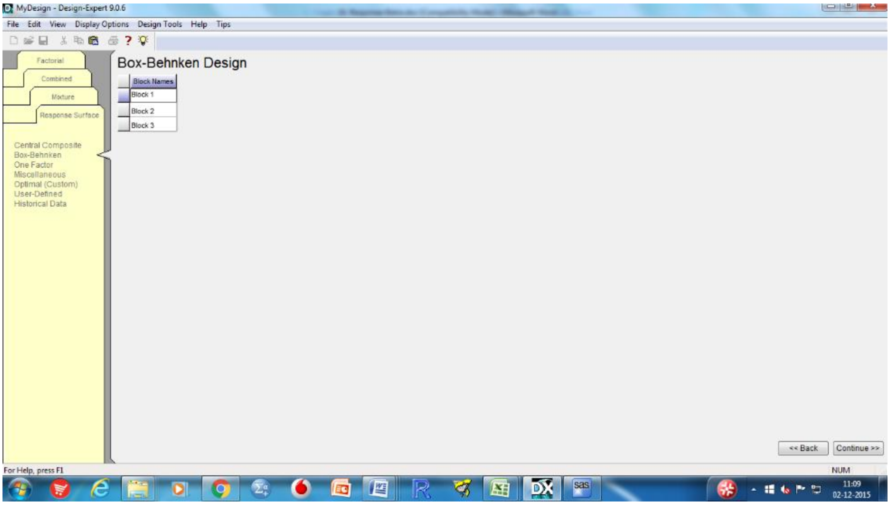


(ii) Construction of Box-Behnken Design

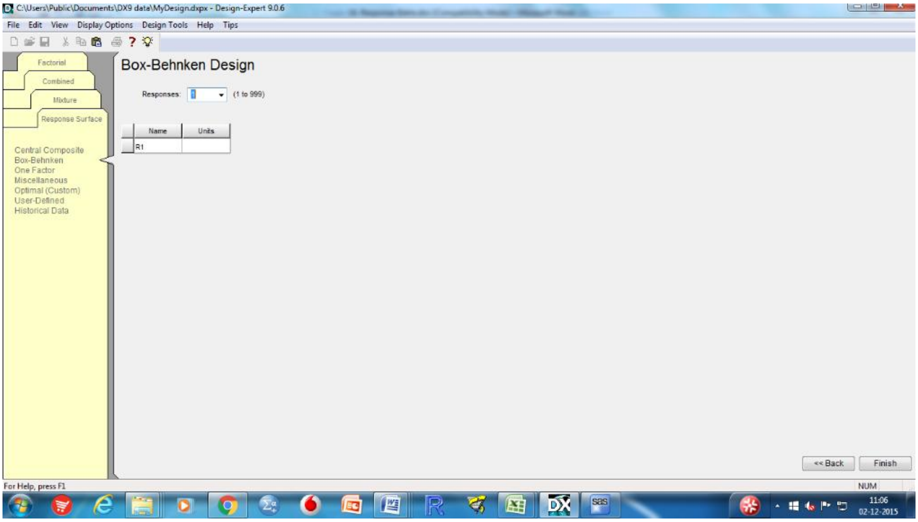
Go to main menu and click Box-Behnken under response surface tab



Enter the number of factors and number of blocks and then continue

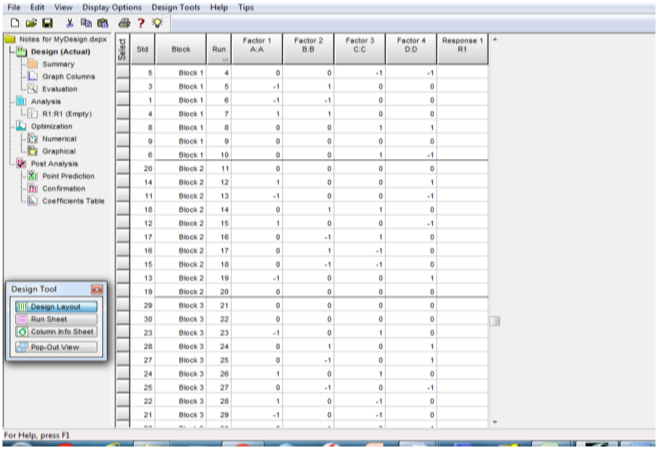


Give the block labels (If required)



Select the number of response variables and continue

Layout of the design as follows



1. Practical Exercise

Exercise 1: A central composite rotatable design was set up to investigate the effects of three fertilizer ingredients on the yield of snap beans under field conditions. The fertilizer ingredients and actual amount applied were nitrogen (N), from 0.89 to 2.83 kg/plot; phosphoric acid (P_2O_5) from 0.265 to 1.336 kg/plot; and potash (K_2O), from 0.27 to 1.89 kg/plot. The response of interest is the average yield in kg per plot of snap beans. The levels of nitrogen, phosphoric acid and potash are coded, and the coded variables are defined as:

$X_1 = (N-1.62)/0.71$, $X_2 = (P_2O_5-0.80)/0.31$, $X_3 = (K_2O - 1.08)/0.48$

The values 1.62, 0.80 and 1.08 kg/plot represent the centres of the values for nitrogen, phosphoric acid and potash, respectively. Five levels of each variable are used in the experimental design. The coded and measured levels for the variables are listed as

	Levels of x_i				
	-1.682	**-1.000**	**0.000**	**+1.000**	**+1.682**
N	0.18	0.91	1.62	2.34	2.83
P_2O_5	0.26	0.48	0.80	1.12	1.33
K_2O	0.27	0.60	1.08	1.57	1.89

Six central point replications were run in order to obtain an estimate of the experimental error variance.

The complete second order model to be fitted to yield values is

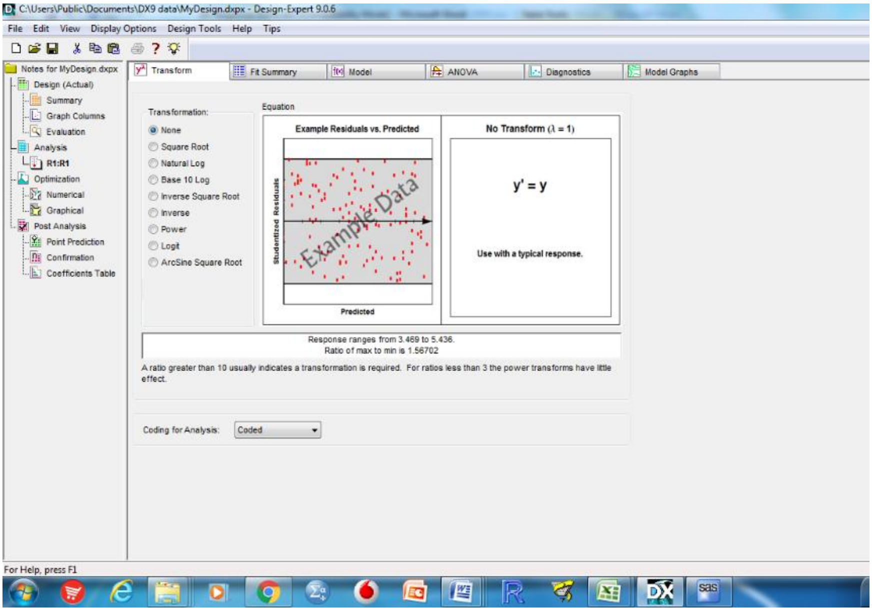
$$Y = \beta_0 + \sum_{i=1}^{3}\beta_i x_i + \sum_{i=1}^{3}\beta_{ii} x_i^2 + \sum_{i=1}^{2}\sum_{i'=2}^{3}\beta_{ii'} x_i x_{i'} + e \tag{4}$$

The following table list the design settings of x_1, x_2 and x_3 and the observed values at 15 design points N, P_2O_5, K_2O and yield are in kg.

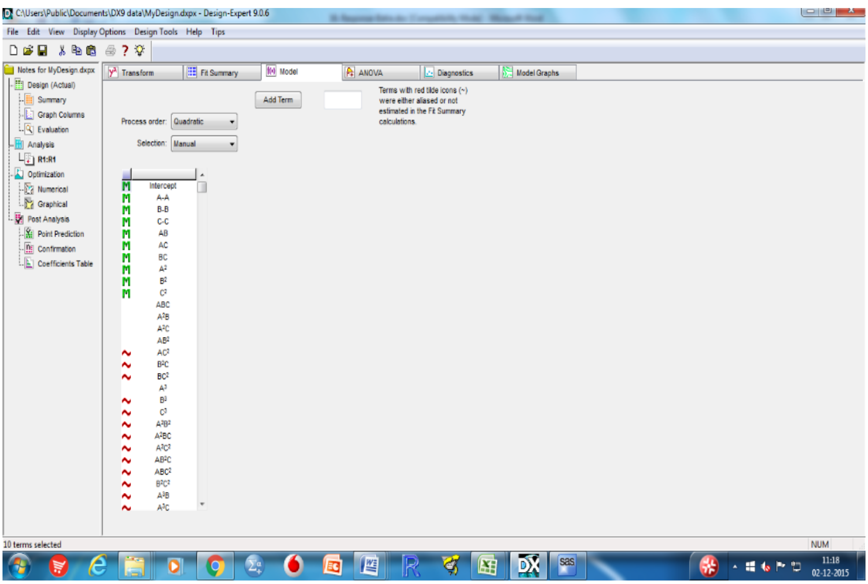
Table 2: Central Composite Rotatable Design Settings in the Coded Variables x_1, x_2 and x_3, the original variables N, P_2O_5, K_2O and the Average Yield of Snap Beans at Each Setting.

x_1	x_2	x_3	**N**	**P_2O_5**	**K_2O**	**Yield**
-1	-1	-1	0.913	0.481	0.607	5.076
1	-1	-1	2.344	0.481	0.607	3.798
-1	1	-1	0.913	1.120	0.607	3.798
1	1	-1	2.344	1.120	0.607	3.469
-1	-1	1	0.913	0.481	1.570	4.023
1	-1	1	2.344	0.481	1.570	4.905
-1	1	1	0.913	1.120	1.570	5.287
1	1	1	2.344	1.120	1.570	4.963
-1.682	0	0	0.423	0.796	1.089	3.541
1.682	0	0	2.830	0.796	1.089	3.541
0	-1.682	0	1.629	0.265	1.089	5.436
0	1.682	0	1.629	1.336	1.089	4.977
0	0	-1.682	1.629	0.796	0.270	3.591
0	0	1.682	1.629	0.796	1.899	4.693
0	0	0	1.629	0.796	1.089	4.563
0	0	0	1.629	0.796	1.089	4.599
0	0	0	1.629	0.796	1.089	4.599
0	0	0	1.629	0.796	1.089	4.275
0	0	0	1.629	0.796	1.089	5.188
0	0	0	1.629	0.796	1.089	4.959

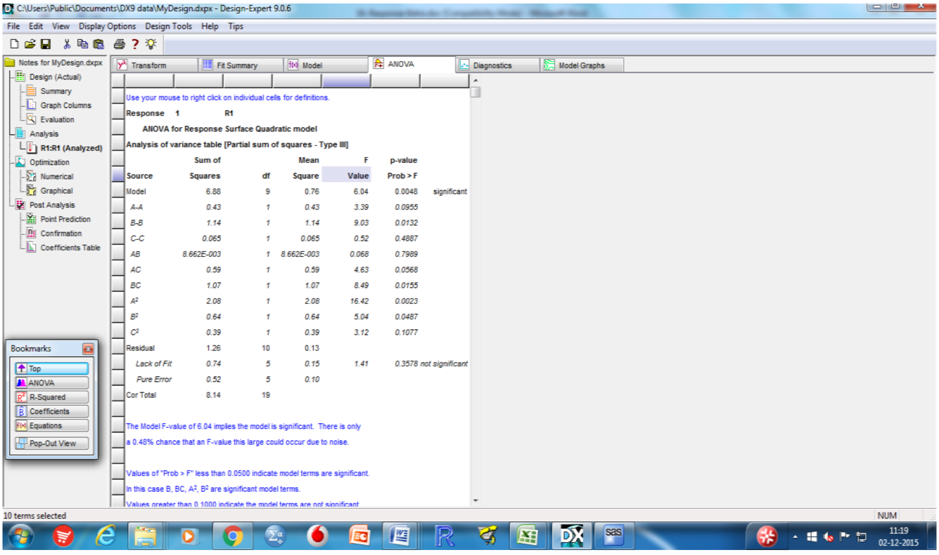
Analysis of data using Design Expert is as follows



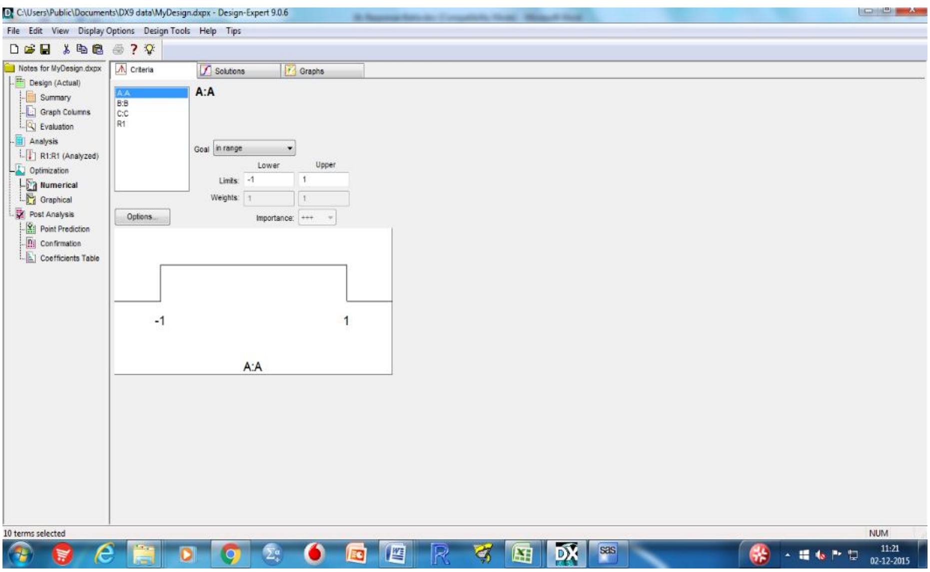
Check whether any transformation is required



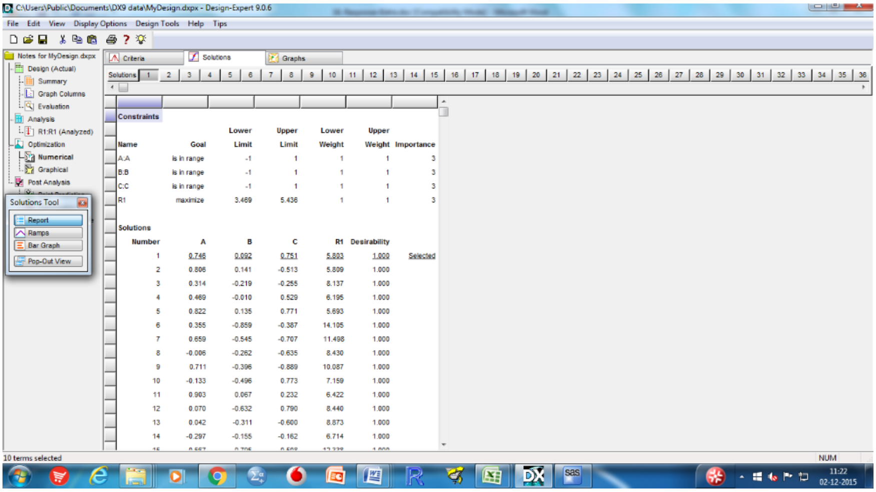
Select the model



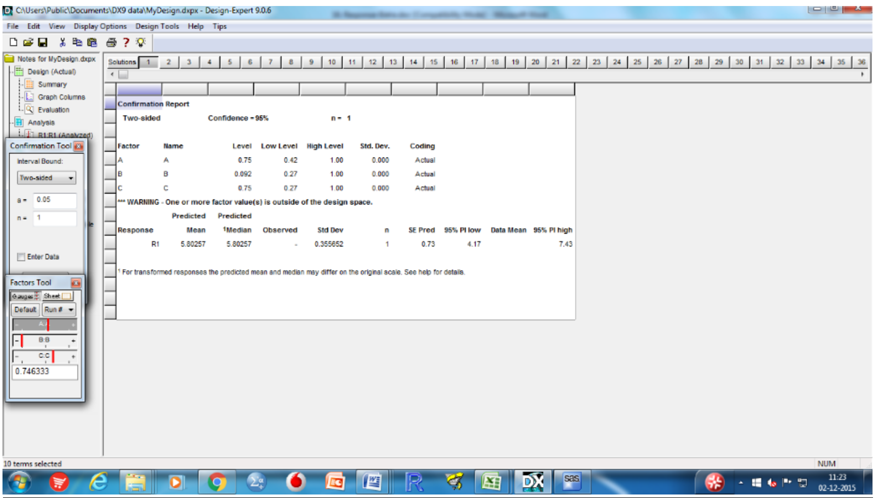
See for significance of the model fitted



Proceed for optimization and specify the target

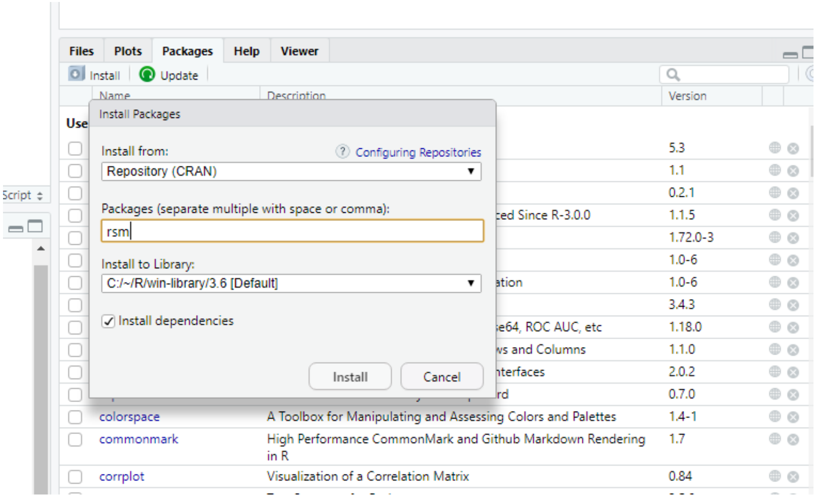


See for desirable solutions

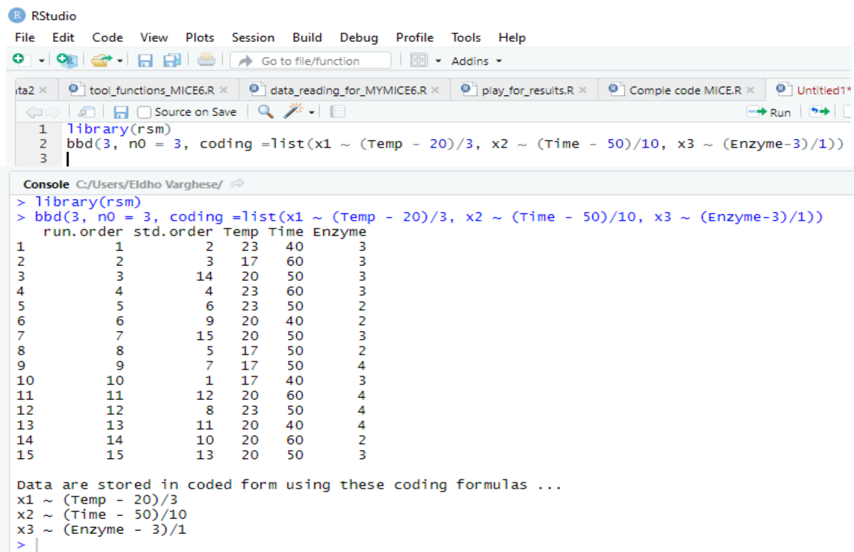


See the confirmation report. Various surface plots can also be generated. There are R packages available for creating response surface designs and performing analysis. Among them, 'rsm' is the most commonly used package.

To install the package named "rsm"



Generation of Box-Behnken design



For fitting the model

A response surface is fitted using the rsm function. This is an extension of lm, and works almost exactly like it: however, the model formula for rsm must make use of the special functions FO, TWI, PQ OR SO (for "first-order", two-way interaction, pure quadratic, and second –order, respectively), because the presence of these specifies the response-surface portion of the model. Other terms that don't involve these functions may be included in the model: often, these terms, would include blocking factors and other categorical predictors.

Analysis output

```
Analysis of Variance Table

Response: Yield
             Df Sum Sq Mean Sq   F value    Pr(>F)
Block         1 69.531  69.531 2611.0950 2.879e-10
FO(x1, x2)    2  9.626   4.813  180.7341 9.450e-07
TWI(x1, x2)   1  0.063   0.063    2.3470    0.1694
PQ(x1, x2)    2 17.791   8.896  334.0539 1.135e-07
Residuals     7  0.186   0.027
Lack of fit   3  0.053   0.018    0.5307    0.6851
Pure error    4  0.133   0.033

Stationary point of response surface:
       x1        x2
0.3722954 0.3343802

Stationary point in original units:
     Time      Temp
 86.86148 176.67190

Eigenanalysis:
eigen() decomposition
$values
[1] -0.9233027 -1.3186949
```

Response Surface Methodology (RSM) is a statistical approach used to explore and optimize relationships between response variables and experimental factors, particularly when curvature in the response surface is suspected. It includes experimental design, model fitting, and optimization to determine ideal factor levels for maximizing or minimizing responses. Widely applied in agriculture, food processing, and biological research, RSM has facilitated advancements such as optimizing fertilizer use, enhancing product quality, and improving efficiency in industrial processes. Designs like Central Composite and Box-Behnken are commonly used for second-order models.

R codes to implement RSM

```
attach(rsd)
library(rsm)
#rsd_fish
attach(rsd_fish)
#rsd_fish$N<-rsd_fish$SD
#rsd_fish$S<-rsd_fish$IM
#rsd_fish$yield<-rsd_fish$WG
rsd<-rsd_fish
# fit second-order (SO) model
analysis= rsm(yield~ SO(N,S), data=rsd)
summary(analysis)
# fit the first-order model
analysis1= rsm(yield~ FO(N,S), data=rsd)
summary(analysis1)
# fit the first-order  with  two-way  interaction  model
analysis2= rsm(yield~ FO(N,S)+TWI(N, S), data=rsd)
summary(analysis2)
#Fit the second-order without interactions model
analysis3= rsm(yield~ FO(N,S)+PQ(N,S), data=rsd)
summary(analysis3)
# compare the reduced first-order model to the full second-order
model
anova(analysis3, analysis)
#detach(rsd)
# conf interval for parameters
confint(analysis)
# Normality of Residuals
library(car)
# externally Studentized residuals
analysis$studres <- rstudent(analysis)
qqPlot(analysis$studres, main="QQ  Plot")
cooks.distance(analysis)
plot(analysis, which =c(4)) #4 is the plot number
# plot diagnostics
plot(rsd$N, analysis$studres, main="Residuals vs N")
# horizontal  line  at  zero
abline(h  =  0,  col  =  "gray75")
plot(rsd$S, analysis$studres, main="Residuals vs S")
# horizontal  line  at  zero
```

```
abline(h = 0, col = "gray75")
# residuals vs order of data
plot(analysis$studres, main="Residuals vs Order of data")
# horizontal line at zero
abline(h = 0, col = "gray75")
# contour plots and surface plots
# first-order model
#par(mfrow=c(2,2))
contour(analysis1, ~ N + S, image = TRUE, main="first-order
model")
persp(analysis1, S ~ N, zlab = "yield", main="first-order model")
# second-order model
contour(analysis, ~ N + S, image = TRUE, main="second-
order model")
persp(analysis, S ~ N, zlab = "yield", main="second-order model")
```

References

Box, G. E., & Wilson, K. B. (1992). On the experimental attainment of optimum conditions. In Breakthroughs in statistics: methodology and distribution (pp. 270-310). New York, NY: Springer New York.

Box, G. E. P., & Behnken, D. W. (1960). Simplex-sum designs: a class of second order rotatable designs derivable from those of first order. The Annals of Mathematical Statistics, 31(4), 838-864.

Kumar, M., Dahuja, A., Sachdev, A., Kaur, C., Varghese, E., Saha, S., & Sairam, K. V. S. S. (2019). Evaluation of enzyme and microwave-assisted conditions on extraction of anthocyanins and total phenolics from black soybean (Glycine max L.) seed coat. International Journal of Biological Macromolecules, 135, 1070-1081.

Kumar, M., Dahuja, A., Sachdev, A., Kaur, C., Varghese, E., Saha, S., & Sairam, K. V. S. S. (2019). Valorisation of black carrot pomace: microwave assisted extraction of bioactive phytoceuticals and antioxidant activity using Box–Behnken design. Journal of Food Science and Technology. https://doi.org/10.1007/s13197-018-03566-9

Nishad, J., Dutta, A., Saha, S., Rudra, S. G., Varghese, E., Sharma, R. R., Tomar, M., Kumar, M., & Kaur, C. (2020). Ultrasound-assisted development of stable grapefruit peel polyphenolic nano-emulsion: Optimization and application in improving oxidative stability of mustard oil. Food Chemistry, 127561.

Kumar, M., Dahuja, A., & Kaur, C. (2016). Optimization of enzyme assisted extraction of flavonoids from black carrot (Daucus carota) using Response Surface Methodology. Indian Research Journal of Genetics and Biotechnology, 8(4), 419–425.

Kumar, M., Dahuja, A., Sachdev, A., Kaur, C., Varghese, E., Saha, S., & Sairam, K. V. S. S. (2018). Black Carrot (Daucus carota ssp.) and Black Soybean (Glycine max (L.) Merr.) Anthocyanin Extract: A Remedy to Enhance Stability and Functionality of Fruit Juices by Copigmentation. Waste and Biomass Valorization. https://doi.org/10.1007/s12649-018-0450-3

Kumar, M., Dahuja, A., Sachdev, A., Kaur, C., Varghese, E., Saha, S., & Sairam, K. V. S. S. (2018). Valorization of Black carrot marc: Antioxidant properties and enzyme assisted extraction of flavonoids. Research Journal of Biotechnology, 13(11), 12-21.

Sharma, S., Varghese, E., Arora, A., Singh, K. N., Singh, S., Nain, L., & Paul, D. (2018). Augmenting pentose utilization and ethanol production of native Saccharomyces cerevisiae LN using medium engineering and response surface methodology.

How to Cite: Varghese, E., and Hemavathi, M. (2023). Response Surface Methodology: Application and Case Studies. In S. Rathod, B. Sailaja, N. Bandumula, S. Arun Kumar, P. A. Lakshmi Prasanna, P. Jeyakumar, A. Waris, P. Muthuraman, and R. M. Sundaram (Eds.), *Statistical Procedures for Analysing Agricultural Data using R* (pp. 243-262). ICAR - Indian Institute of Rice Research, Hyderabad.