Statistics and Applications {**ISSN 2452-7395 (online)**}
Volume 17, No. 1, 2019 (New Series), pp 161–178

Response Surface Analysis using Fuzzy Regression Model

Pranesh Kumar[1] **and Jiefei Yang**[2]
Department of Mathematics and Statistics, University of Northern British Columbia, Prince George, BC, V2N 4Z9, Canada

Received: October 11, 2018; Revised: March 03, 2019; Accepted: March 12, 2019

Abstract

Response Surface Methodology is extensively used in applications involving multiple input variables (factors) which influence some output (response) variables, performance measures or certain quality characteristics of a process. These input variables are under the control of the experimenter. The main purpose of response surface methodology is to form a strategy for the experimentation by exploring the space of input variables (process) to develop a suitable functional relationship (model) between response and input variables. This model is then used to find the optimum set of values of the process input variables which are expected to result in the desired optimum response. For identifying and fitting the appropriate response surface model, methodology makes use of fundamentals of experimental design, statistical modelling and optimization techniques. Besides response surface methodology being extensively used in agronomic experiments, it is also commonly used as an automated tool for model calibration and validation especially in modern computational multi-agent large-scale social-networks systems that are used in modeling and simulation of complex social networks. This methodology can be integrated in many large-scale simulation systems such as BioWar, ORA and is currently integrating in Vista, Construct, and DyNet. This paper describes the fuzzy logic approach to fit the response surface model, analysis and the implementation of chosen method by using the agronomic experimental data.

Key words: Response surface methodology; Least squares regression; Fuzzy logics and Regression.

1 Response Surface Model

We consider that the experimenter's interest is to study the response variable (y) which can be observed and is influenced by a set of p-input variables $x_1, \dots, x_p$. To approximate the underlying relationship of response with input variables, we approximate the unknown true

[1]Dedicated to Professor Arun Kumar Nigam, who always has been a source of inspiration and motivation in multiple ways, like, a teacher, guide, co-researcher, more than a friend and a well-wisher, on his 75th Birth Anniversary.
[2] Mitacs Globalink Research Intern from Department of Mathematical Sciences, Dalian University of Technology, Dalian 116024, P. R. China.

Corresponding Author: Pranesh Kumar
Email: Pranesh.Kumar@unbc.ca

relationship with an appropriate empirical model of the form $y = f(x_1, \ldots, x_p) + \varepsilon$, where the term ε represents the observed error in the response. Often, function f is assumed to be a first-order or second-order polynomial. This empirical model is termed as a response surface model. We assume that there is a curvature in the system and, thus, consider a second-order response model:

$$y = a_0 + \sum a_i x_i + \sum a_{ii} x_i^2 + \sum a_{ij} x_i x_j + \varepsilon;\ i, j(\neq i) = 1, \ldots, p. \tag{1}$$

It may be unlikely that a polynomial model will be a reasonable approximation of true functional relationship over the entire space of input variables; however, they work quite well for a relatively small region. The well-known method of ordinary least squares is generally used to fit the model. The response surface analysis is then performed using the fitted model. The model constants can be estimated most effectively by using appropriate experimental designs to collect data. Designs for fitting response surface are termed as response surface designs. The response surface methodology is a sequential procedure. When a point on the response surface is remote from the optimum, there is little curvature in the system and the first-order model will be appropriate. Once the region of the optimum is found, an improved model, such as second-order model, may be employed and analysis may be performed to locate the optimum. To find the levels $(x_{1s}, \ldots, x_{ps})$ of input-variables $x_1, \ldots, x_p$, we optimize the predicted response. This point $(x_{1s}, \ldots, x_{ps})$ is called the stationary point which could represent a point of maximum response, or a minimum response or a saddle point. By generating contour plots using computer software for response surface analysis, we can characterize the shape of the surface and locate the optimum with reasonable accuracy.

2 Characterization of the Response-Surface Model

Mathematically, we may obtain a general solution for the location of the stationary point of the second-order response model. In matrix notation, let fitted model be

$$\hat{y} = \hat{a}_0 + \mathbf{x}^{\mathrm{T}}\boldsymbol{a} + \mathbf{x}^{\mathrm{T}}\boldsymbol{A}\mathbf{x}, \tag{2}$$

where, $\mathbf{x}$ is a p x 1 vector of the input variables, $\boldsymbol{a}$ is a p x 1 vector of the first-order regression coefficients and $\boldsymbol{A}$ is a p x p symmetric matrix having the pure quadratic coefficients (a_{ii}) as main diagonal elements and one-half the mixed quadratic coefficients $(a_{ij}; i \neq j)$ as off-diagonal elements. That means,

$$\mathbf{x}^{\mathrm{T}} = [x_1, \ldots, x_p],\ \boldsymbol{a}^{\mathrm{T}} = [\hat{a}_1, \ldots, \hat{a}_p],\ \boldsymbol{A} = \begin{bmatrix} \hat{a}_{11} & \hat{a}_{12}/2 & \cdots & \hat{a}_{1p}/2 \\ & \hat{a}_{22} & \cdots & \hat{a}_{2p}/2 \\ & & \cdots & \cdots \\ & & & \hat{a}_{pp} \end{bmatrix}. \tag{3}$$

The stationary point is $\mathbf{x_s} = \boldsymbol{A}^{-1}\boldsymbol{a}/2$, and the predicted response at the stationary point $\mathbf{y_s} = \hat{a}_0 + \mathbf{x_s^T}\boldsymbol{a}/2$. We need to determine if the stationary point $\mathbf{x_s}$ leads to a maximum or minimum response or is a saddle point. This may be done easily by constructing and examining the contour plot of the fitted model if there are two or three input variables. Alternatively, we can conduct the canonical analysis. It is useful to transform the model into a new co-ordinate system choosing origin at the stationary point $\mathbf{x_s}$ and then rotating the axes of the system until they are parallel to the principal axes of the fitted response model. Denoting transformed input variables by w_i and constants by λ_i, fitted canonical form of the model is given by

$$\hat{y} = y_s + \sum \lambda_i w_i^2\,,\ i = 1, \dots, p. \tag{4}$$

The constants $\lambda_i{}'$s are also the eigen-values of the matrix $\boldsymbol{A}$. If $\lambda_i{}'$s are all positive, the stationary point $\mathbf{x_s}$ is a point of minimum response; if $\lambda_i{}'$s are all negative, $\mathbf{x_s}$ is a point of maximum response; and if $\lambda_i{}'$s have opposite signs, $\mathbf{x_s}$ is a saddle point. Further the response surface is steepest in the w_i direction for which $|\lambda_i|$ is the largest.

In some applications, it may be necessary to find the relationship between the canonical variables, w_i and the design (input) variables, x_i.

3 Fitting Response-Surface Model: Method of Least Squares

Classical statistical linear regression models are extensively used in almost every field of science wherein interest is in studying the cause-and-effect relationships among the multiple variables. The purpose of regression analysis is to explain the variation of a dependent variable in terms of the variation of explanatory variables. In regression analysis, the degree of contribution of each explanatory variable to the dependent variable is explained by their coefficients. Although, linear regression models and their estimations are well known, for the sake of completeness, we describe them in brevity. Rewriting the response surface model (1) as

$$\mathbf{y} = \mathbf{Xb} + \boldsymbol{\epsilon}, \tag{5}$$

$$\mathbf{y}^{\mathrm{T}} = [y_1, \dots, y_n],\ \mathbf{b}^{\mathrm{T}} = [b_0, b_1, \dots, b_p],\ \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix},\ \boldsymbol{\epsilon}^{\mathrm{T}} = [\epsilon_1, \dots, \epsilon_n],$$

the well-known classical method of least squares results in the least squares estimates of the coefficients of the response-surface model as $\hat{\mathbf{b}} = (\mathbf{X^T X})^{-1}\mathbf{X^T y}$.

4 Fitting Response-Surface Model: Fuzzy Logic and Regression

Crisp data, also known as precise data, are very common in our everyday life. The traditional science and technology pursuit for certainty in all its manifestations and almost all the mathematical theories are developed for handling such kind of data. However, in many cases, data have the characteristic of uncertainty. There are primarily two types of uncertainty. The first is probabilistic uncertainty, which is well developed overtime. The second is what is termed as fuzzy uncertainty. Let us start with fuzzy data, which is a combination of fuzzy variable and random variable and can characterize both fuzziness and randomness. Fuzzy Logic as a superset of conventional (Boolean) logic was first introduced by Zadeh (1965, 1968) to handle the concept of partial truth. Fuzzy Logic is considered as the most powerful tool for dealing with imprecision and uncertainties. Zadeh proposed that fuzzy set can be applied to represent data which is fuzzy and this fuzziness can be represented by the degree of participation to a set called a membership function. Let $\boldsymbol{X}$ be a space of points. A fuzzy set $\mathbf{A}$ in space $\boldsymbol{X}$ is characterized by a membership function, $\boldsymbol{\mu_A}(\boldsymbol{x})$ and the value of $\boldsymbol{\mu_A}(\boldsymbol{x})$ at $\boldsymbol{x}$ representing the grade of membership of $\boldsymbol{x}$ in $\boldsymbol{A}$ where $\boldsymbol{\mu_A}: \boldsymbol{X} \to [\mathbf{0}, \mathbf{1}]$. For traditional bivalent logic, the value of membership function of crisp data can only be 0 or 1, that means, outside the set, or within the set, respectively. However, a fuzzy set allows for its members to have degrees between 0 and 1. Thus, it can describe natural phenomenon more accurately. Further, conventional set theory and binary logic have three elementary binary operations, that is, intersection (and), union (or), and complement set (negation). The rules of binary operations were generalized in order to fitting fuzzy data. The fuzzy logic operations truth table is shown in Table 1. The generalized form of the operators works well for the fuzzy and for the bivalent data as well.

Table 1: The Generalized form of operations; Truth table

x and y	$\min(x, y)$
x or y	$\max(x, y)$
not x	$1 - x$

5 Fuzzy Regression

The classical linear regression has crisp coefficients and is bounded by some strict assumptions about the given data, that is, the unobserved error terms are mutually independent and identically distributed. However, if the data set is too small in size, or, if there is difficulty in verifying that the errors are normally distributed, or, if there is vagueness in the relationship between the dependent and independent variables, or, if there is ambiguity associated with the events, it is well known that the classical linear regression may fail to work satisfactorily. In such cases, alternatively, fuzzy linear regression may be more useful. Fuzzy linear regression (FLR) was first introduced by Tanaka (1982) and then further developed in Tanaka (1987). The FLR model includes a fuzzy output and non-fuzzy input variables and fuzzy coefficients. In this paper, however, our focus is on the type of fuzzy regression model considered by Tanaka (1987). The basic model assumes a fuzzy linear functional form

$$\tilde{y} = \tilde{A}_0 + \tilde{A}_1 x_1 + \cdots + \tilde{A}_p x_p, \tag{6}$$

where $\boldsymbol{x} = \left[x_1, x_2, \dots, x_p\right]^T$ is a vector of input variables, $\widetilde{\mathbf{A}} = \left[\tilde{A}_0, \tilde{A}_1, \dots, \tilde{A}_p\right]$ is a vector of fuzzy coefficients presented in the form of symmetric triangular fuzzy data denoted by $\widetilde{A_j} = (a_j, c_j)$ with its membership function described as

$$\mu_{A_j}(\alpha) = \begin{cases} 1 - \frac{|a_j - \alpha|}{c_j}, \ a_j - c_j \le \alpha \le a_j + c_j; j = 1,2, \dots, p, \\ \qquad\qquad 0, \ otherwise, \end{cases} \tag{7}$$

where a_j is the central value and c_j is the width. The membership function of the fuzzy output can be described as

$$\mu_{Y_i}(y) = \begin{cases} 1 - \frac{|y_i - y|}{e_i}, \ y_i - e_i \le y \le y_i + e_i; i = 1,2, \dots, n, \\ \qquad\qquad 0, \ otherwise. \end{cases} \tag{8}$$

The degree of fitting of the fuzzy regression model to the given data $\boldsymbol{Y}_i = (y_i, e_i)$ is measured by an index $\min_j\left[\bar{h}_j\right]$, where

$$\bar{h}_i = 1 - \frac{|y_i - \mathbf{x}_i^T \alpha|}{\sum_j c_j |x_{ij}| - e_i}. \tag{9}$$

The vagueness of the fuzzy regression model is defined by $J = \sum_j c_j$. The fuzzy parameter $\tilde{A}_j$ is obtained so as to minimize J subject to $\bar{h}_i \ge H$, where H is selected as the degree of fitting the model by the experimenter.

The basic idea is to minimize the fuzziness of the model by minimizing the total support of the fuzzy coefficients subject to including all the given data. As a result, we can obtain the best fitted model for the given data by solving the conventional linear programming problem.

$$\begin{aligned} \min \ \ & J = mc_0 + \sum_{j=1}^{m} \sum_{i=1}^{n} c_i x_{ij} \\ \text{s.t.} \ \ & y_j \ge \sum_{i=1}^{n} a_i x_{ij} - (1 - H) \sum_{i=1}^{n} c_i x_{ij} + (1 - H) e_j, \\ & y_j \le \sum_{i=1}^{n} a_i x_{ij} + (1 - H) \sum_{i=1}^{n} c_i x_{ij} - (1 - H) e_j, \\ & c_i \ge 0, \ i = 0,1, \dots, n. \end{aligned} \tag{10}$$

For the sake of completeness, Matlab 2018 programming codes for fitting the model are included in the appendix.

6 Illustration

For illustration of the fuzzy response surface analysis, we have adapted the data from an experiment which was conducted at the Division of Agronomy, Indian Agricultural Research Institute, New Delhi [Data Source: Design Resources Server, www.iasri.res.in/design]. The main purpose of the experiment was to derive the optimum combination of Nitrogen (N) and Sulphur (S) for maximizing the yield of paddy crop. The experimenter recorded paddy yield (kg/ha) by applying four levels of Nitrogen as 0, 50, 100, 150 kg/ha and four levels of Sulphur as 0, 20, 40, 60 kg/ha. The experiment was conducted using a RCB design in three replications. The experiment data are shown in Table 2.

Table 2: Agronomy experiment data

Yield y	Nitrogen x_1	Sulphur x_2
4121.212	0	0
4678.03	0	20
4742.424	0	40
4727.273	0	60
6083.333	50	0
6041.667	50	20
6223.485	50	40
6715.909	50	60
6761.364	100	0
6916.667	100	20
6852.273	100	40
6810.606	100	60
6174.242	150	0
7022.727	150	20
7003.788	150	40
6943.182	150	60

The analysis of variance of the data revealed that replications were not significantly different, *i.e.*, the replication mean square was smaller in comparison to the error mean square.

For analysis, a second order response surface is considered using the inputs in order to fit the quadratic response surface. We have fitted the response model with coefficients $a_0, a_1, a_2, a_3, a_4,$ and a_5:

$$E(Yield) = a_0 + a_1 N + a_2 S + a_3 N^2 + a_4 S^2 + a_5 NS \tag{11}$$

The co-ordinates of stationary point and the predicted yield at the stationary point are computed. For fuzzy model to obtain fuzzy coefficients of response surface, we have assumed 1% spread of the yield which could be due to the measurement errors or due to other unknown sources. The upper bounds and the lower bounds of the predicted yield are obtained. The observed values are supposed to be in the interval of the computed bounds. For the algorithm, computer codes are prepared and the statistical software Matlab 2018 is used to analyze the data.

6.1 Response Surface: Method of Least Squares

By applying the method of ordinary least squares, the estimated coefficients of the response surface model are described in Table 3.

Table 3: Estimated least squares coefficients, $\hat{a}_0, \hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{a}_4$, and $\hat{a}_5$

$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{a}_4$	$\hat{a}_5$
4266.250	40.906	19.226	-0.175	-0.179	-0.008

The co-ordinates of the stationary point, $\mathbf{x}_s$ = (N=115.8582 kg/ha, S=51.2538 kg/ha) and the predicted yield at the stationary point is $y_{pred} = 7128.60$ kg/ha.

After finding the stationary point, it is necessary to characterize the response surface in the immediate vicinity of this point. Characterizing the response surface means determining whether the stationary point is a point of maximum or minimum response or a saddle point. In order to find the nature of the stationary point, first we have transformed the model into a new coordinate system with the origin at the stationary point x_0 and then rotated the axes of this system until they are parallel to the principal axes of the fitted response surface. This results in the fitted model

$$y = \lambda_0 + \lambda_1 w_1^2 + \lambda_2 w_2^2, \tag{12}$$

where $\{w_i's\}$ are the transformed independent variables (the canonical variables) and $\{\lambda_i's\}$ are constants.

Furthermore, $\{\lambda_i's\}$ are just the eigenvalues of the matrix $\mathbf{A} = \begin{pmatrix} a_3 & 0.5a_5 \\ 0.5a_5 & a_4 \end{pmatrix}$. If the $\{\lambda_i's\}$ are all positive, x_0 is a point of minimum response; if they are all negative, x_0 is a point of maximum response; and if they have opposite signs, x_0 is a saddle point. In this example, the eigenvalues of $\mathbf{A}$ are $\lambda_1 = -0.1811$ and $\lambda_2 = -0.1724$, implying that the stationary point is a point of maximum response. The quadratic response surface is shown in Figure 1. The contour plot is shown in Figure 2.

Sometime, it may not be possible to operate the process at the stationary point. Then, we need to find the relationship between the canonical variables $\{w_i\}$ and the design variables $\{x_i\}$, which can be illustrated by the following equation

$$\mathbf{w} = \mathbf{M}'(\mathbf{x} - \mathbf{x}_0), \tag{13}$$

where $\mathbf{M}$ is a (2×2) orthogonal matrix. The columns of $\mathbf{M}$ are normalized eigenvectors associated with the $\{\lambda_i's\}$. In other words, the ith columns of $\mathbf{M}$ denoted $\mathbf{m_i}$ is the solution to

$$(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{M_i} = \mathbf{0}, \tag{14}$$

for which $m_{1i}^2 + m_{2i}^2 = 1$.

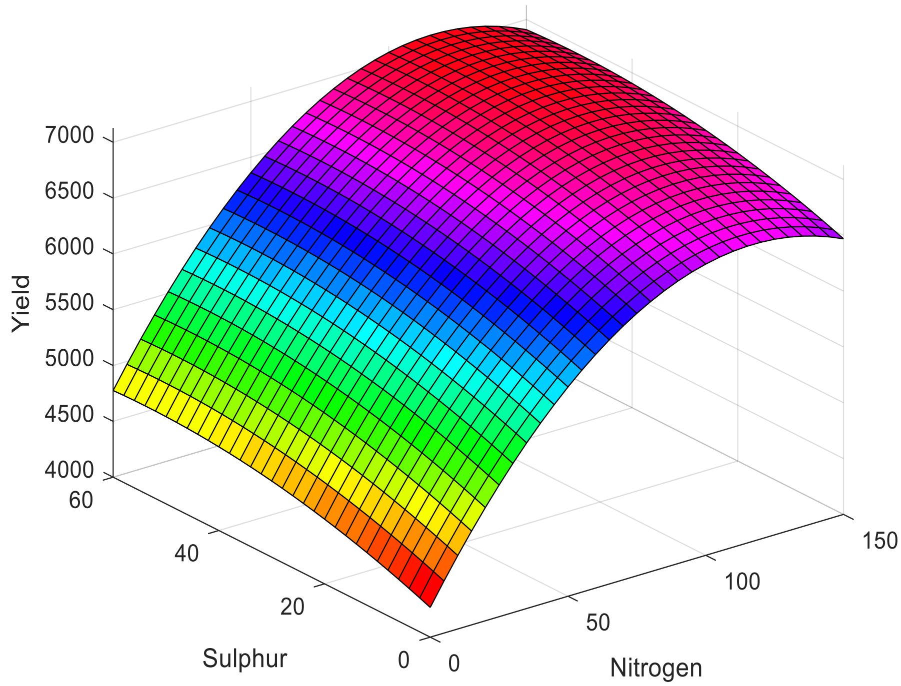

Figure 1: Response surface of agronomy experiment data

Using mathematical software MATLAB 2018, the relationship between the canonical variables $\{w_i\}$ and the design variables $\{x_i\}$ is found to be

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0.5245 & -0.8514 \\ 0.8514 & 0.5245 \end{bmatrix} \begin{bmatrix} x_1 - 115.8582 \\ x_2 - 51.2538 \end{bmatrix},$$

or,

$$w_1 = 0.5245(x_1 - 115.8582) - 0.8514(x_2 - 51.2538),$$

$$w_2 = 0.8514(x_1 - 115.8582) + 0.5245(x_2 - 51.2538).$$

By the above equations, we can determine appropriate points at which to observe in the (w_1, w_2) space and convert these points into the (x_1, x_2) space to further explore the response surface around the stationary point $x_0 = (115.8582, 51.2538)$.

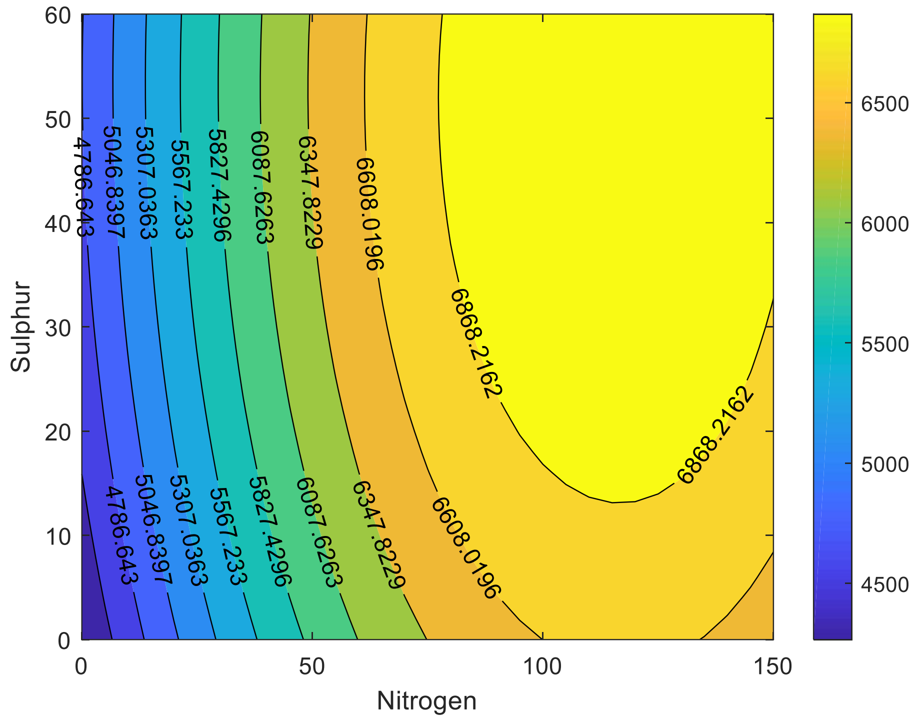

Figure 2: The Contour plot of the response surface of agronomy experiment data

6.2 Response Surface: Method of Fuzzy Regression

To fit the fuzzy response surface, we consider 1% spread for the yield accounting for measurement errors or yield losses due to unknown sources. We apply fuzzy regression model to obtain fuzzy coefficients of the response surface and calculate the upper and the lower bounds of the predicted yield. Estimated fuzzy coefficients of the response surface are given in Table 4.

Table 4: Fuzzy coefficients of response surface model

Fuzzy coefficients	$\widetilde{A_0}$	$\widetilde{A_1}$	$\widetilde{A_2}$	$\widetilde{A_3}$	$\widetilde{A_4}$	$\widetilde{A_5}$
Center a_j	4227.421	43.134	19.802	-0.187	-0.142	-0.038
Width c_j	253.632	2.842	0	0	0.040	0

From the fitted fuzzy response surface using center (a_j) coeffcients, the co-ordinates of stationary point is $x_{0c} = (109.624, 55.098)$ and the predicted yield at the stationary point is $y_{pred,c} = 7137.19$.

Further, let the eigenvalues of $\widetilde{\mathbf{A}}$ be $\{\widetilde{\lambda_i}\}$. We obtain that $\widetilde{\lambda_1} = -0.1941$ and $\widetilde{\lambda_2} = -0.1355$, which indicates that the stationary point of the predicted surface is a point of maximum response.

Besides, the range of co-ordinates of the maximum point is obtained. In order to maximizing the yield of paddy crop, the optimum levels of Nitrogen and Sulphur are expected to be in the interval $[103.191, 115.124]$ (kg/ha) and $[43.579, 75.917]$ (kg/ha), respectively. The interval of predicted yield is [6484.12, 7879.18] (kg/ha).

The predicted yields from the fitted fuzzy response surface based on central values a_j, the upper bound $(a_j + c_j)$, and the lower bound $(a_j - c_j)$ are shown in the Figure 3.

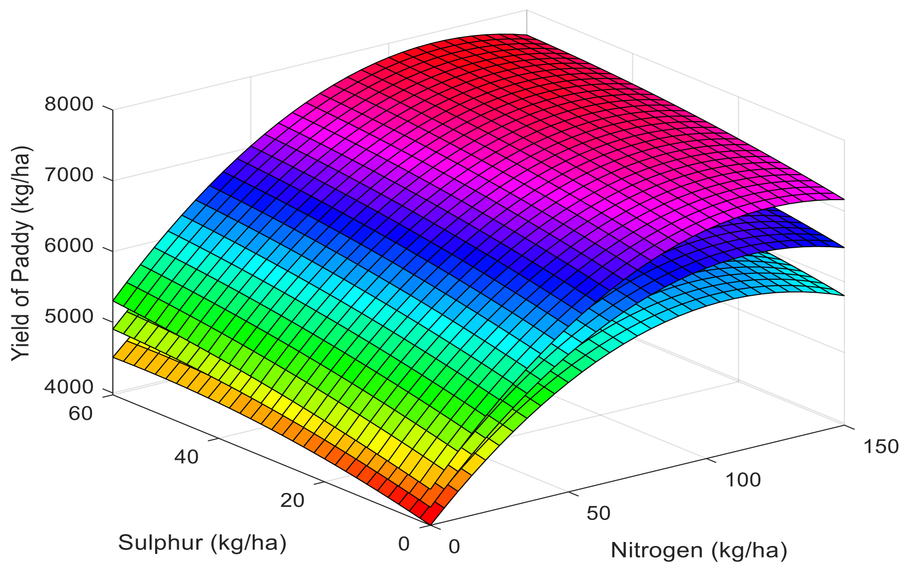

Figure 3: Fuzzy Response Surface

For comparative results of the usual least squares response surface and fuzzy response surface models, we present in Table 5, the observed and predicted yields. The center predicted yields, the upper bounds, and the lower bounds are shown in the Figure 4.

Table 5: Predicted yields

Levels		Observed	Predicted	Predicted	Predicted	Predicted
N	S	$y_{observed}$	y_{center}	y_{lower}	y_{upper}	y_{lse}
0	0	4121.212	4227.422	3973.790	4481.054	4266.250
0	20	4678.030	4566.526	4296.738	4836.314	4579.280
0	40	4742.424	4791.771	4473.513	5110.028	4749.318
0	60	4727.273	4903.156	4504.117	5302.196	4776.364
50	0	6083.333	5915.874	5520.124	6311.625	5874.545
50	20	6041.667	6217.412	5805.505	6629.319	6179.773
50	40	6223.485	6405.090	5944.714	6865.466	6342.008
50	60	6715.909	6478.910	5937.752	7020.067	6361.250
100	0	6761.364	6667.855	6129.987	7205.724	6608.788
100	20	6916.667	6931.826	6377.801	7485.851	6906.212

100	40	6852.273	7081.938	6479.444	7684.433	7060.644
100	60	6810.606	7118.191	6434.915	7801.467	7072.084
150	0	6174.242	6483.364	5803.377	7163.351	6468.977
150	20	7022.727	6709.769	6013.626	7405.912	6758.598
150	40	7003.788	6822.314	6077.702	7566.927	6905.227
150	60	6943.182	6821.001	5995.606	7646.395	6908.864

6.3 Error Analysis

To make performance comparisons, we have calculated prediction errors denoted by $e_f = y_{observed} - y_{center}$ and $e_{lse} = y_{observed} - y_{lse}$ and corresponding standardized errors $e_{fstandardized} = \frac{e_f}{\sigma_{e_f}}$, and $e_{lstandardized} = \frac{e_{lse}}{\sigma_{else}}$, where σ_{e_f} and σ_{else} denote the standard errors. Further, root mean squared errors are calculated as $RMSE = \sqrt{\frac{\sum e^2}{n-10}}$. The coefficient of determination, denoted by R^2, which is the proportion of variance in the dependent variable that could be explained by the regression model, is calculated as $R^2 = 1 - \frac{\sum e^2}{\sum (y-\bar{y})^2}$, where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

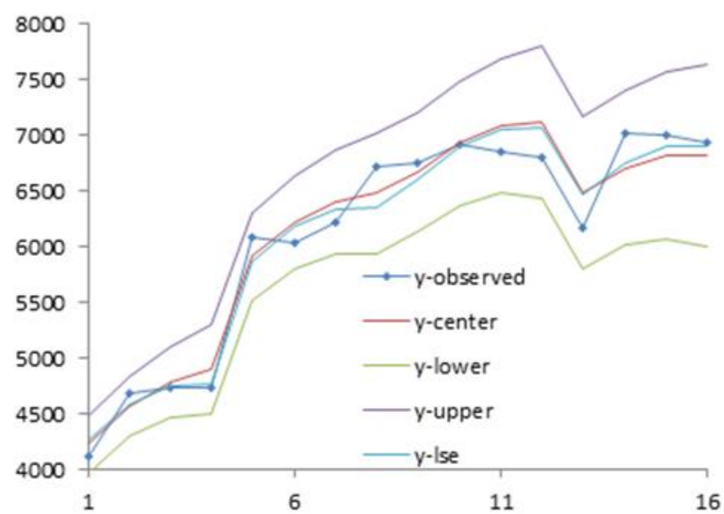

Figure 4: Predicted yields

We have plotted the prediction standardized errors due to fitted fuzzy and least squares response models in Figure 5 and calculated the error summary statistics: root mean square error (RMSE) and R^2 in Table 6.

Table 6: Prediction error statistics of fuzzy and least squares response models

Data Point		Fuzzy RSD		RSD	
x_1	x_2	e_f	$e_{standardize}$	e_{lse}	$e_{standardize}$
0	0	-106.210	-0.550	-145.038	-0.791
0	20	111.504	0.578	98.750	0.538
0	40	-49.347	-0.256	-6.894	-0.038
0	60	-175.883	-0.911	-49.091	-0.268
50	0	167.459	0.867	208.788	1.138
50	20	-175.745	-0.910	-138.106	-0.753
50	40	-181.605	-0.941	-118.523	-0.646
50	60	236.999	1.228	354.659	1.933
100	0	93.509	0.484	152.576	0.832
100	20	-15.159	-0.079	10.455	0.057
100	40	-229.665	-1.190	-208.371	-1.136
100	60	-307.585	-1.593	-261.478	-1.425
150	0	-309.122	-1.601	-294.735	-1.607
150	20	312.958	1.621	264.129	1.440
150	40	181.474	0.940	98.561	0.537
150	60	122.181	0.633	34.318	0.187
RMSE		244.1883		232.0440	
R^2		0.958766		0.963171	

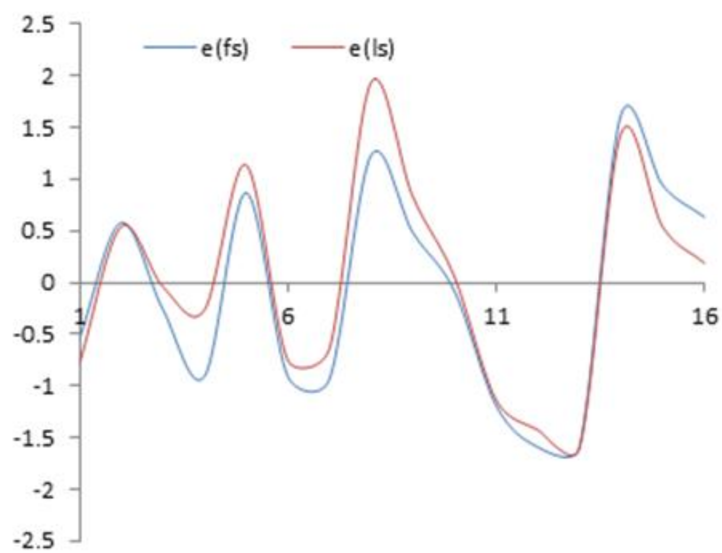

Figure 5: Standardized prediction errors

7 Concluding remarks

In the context of response surface methodology, we have considered fitting the quadratic response surface based on fuzzy regression and have compared it with that based on ordinary least squares method. Thus, we have shown the application of alternate method of fitting the model. By considering an agronomic experiment data on paddy yield and different levels of Nitrogen and Sulphur, the estimated values of the optimum combination and the yield are obtained as fuzzy sets which represent the fuzziness of the system structure. We have calculated the upper and the lower bounds of the predicted yield and carried error analysis which clearly indicates the comparative performance of two fitting methods with possible edge of the fuzzy regression over the ordinary least squares regression.

Acknowledgements

Authors wish to place on record thanks to the reviewer and the Guest Editor for their valuable comments which helped in finalizing the paper.

References

Parsad, R., Gupta, V.K., Umrao, A.K. and Kole, B. *Design Resources Server*, Indian Agricultural Statistics Research Institute (ICAR), New Delhi, India. www.iasri.res.in/design

Tanaka, H., Uejima, S. and Asai, K. (1982) Linear regression analysis with fuzzy model. Systems Man and Cybernetics, *IEEE Transactions*, **12 (6)**, 903-907.

Tanka, H. (1987). Fuzzy data analysis by possibilistic linear models. *Fuzzy Sets and Systems*, **24(3)**, 363-375.

Zadeh, L.A. (1965) Fuzzy sets. *Information and Control*, **8**(**3**), 338–353.

Zadeh, L.A. (1968) Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications*, **23**(**2**), 421-427.

Appendix

A1. Least Squares Response Surface Model

```
%% Least Squares Response Surface Model
% First enter matrix B, which has diagonal elements as
% the estimated regression coefficients of pure quadratic terms and off diagonal elements as
% half of the estimated regression coefficients of cross product terms
% Compute the co-ordinates of Stationary point x0= -0.5*Binv*bL, where bL is the vector
% of estimated regression coefficients of linear terms
% the stationary point is a point of maxima if the eigenvalues of B are -ve;
%                 is a point of minima if the eigenvalues of B are +ve;
%                 is a saddle point if some values are positive and other negative
% compute y-pred at stationary point
clear,clc
data = xlsread('rsd.xls');
yield = data(:,1);
N = data(:,2);
S = data(:,3);
NN = data(:,4);
SS = data(:,5);
NS = data(:,6);
n = length(yield);
%% Fit a second order response surface using the above data
a = regress(yield,[ones(n,1),N,S,NN,SS,NS])
%% plot response surface
N1 = min(N):5:max(N);
```

```
S1 = min(S):2:max(S);
[Nm,Sm] = meshgrid(N1,S1);
y = a(1)+a(2)*Nm+a(3)*Sm+a(4)*Nm.*Nm+a(5)*Sm.*Sm+a(6)*Nm.*Sm;
surf(Nm,Sm,y),colormap HSV
hold on
%% Obtain the co-ordinates of the stationary point
B = [a(4),0.5*a(6);
   0.5*a(6),a(5)];
bL = [a(2);a(3)];
Binv = inv(B);
x0 = -0.5*Binv*bL
y_pred = a(1)+x0'*bL+x0'*B*x0
%% Find the nature of the stationary point
[V,D] = eig(B)
% The eigenvalues of B are -ve.
% Thus, the stationary point is a point of maxima.
```

A2. Fuzzy Response Surface Model

```
%% Fuzzy Response Surface Model
% Considering 1% spread for yield and fuzzy coefficient of N and S.
clear,clc
data = xlsread('rsd.xls');
yield = data(:,1);
N = data(:,2);
S = data(:,3);
NN = data(:,4);
```

```
SS = data(:,5);
NS = data(:,6);
n = length(yield);
format long
%% Fit a second order response surface using the above data
X = [ones(n,1),N,S,NN,SS,NS];
a = regress(yield,X)
%% Considering 1% spread for yield
H = 0.5;
e = yield.*0.01;
f = [0;0;0;0;0;0;n;sum(N);sum(S);sum(NN);sum(SS);sum(NS)];
A1 = [X,(1-H)*X];
A2 = [-X,(1-H)*X];
A = -[A1;A2];
b1 = yield+(1-H)*e;
b2 = -yield+(1-H)*e;
b = -[b1;b2];
lb = [-10000*ones(6,1);zeros(6,1)];
[a,fval,exitflag,output,lambda] = linprog(f,A,b,[],[],lb);
ac = a(1:6)
aw = a(7:12)
format long
%% Obtain the co-ordinates of the central stationary point
Bc = [ac(4),0.5*ac(6);
   0.5*ac(6),ac(5)];
```

```
bLc = [ac(2);ac(3)];

Binvc = inv(Bc);

x0c = -0.5*Binvc*bLc

y_predc = ac(1)+x0c'*bLc+x0c'*Bc*x0c

%% The stationary point of bounds of yield

au = ac+aw;

al = ac-aw;

Bu = [au(4),0.5*au(6);

    0.5*au(6),au(5)];

bLu = [au(2);au(3)];

Binvu = inv(Bu);

x0u = -0.5*Binvu*bLu

y_predu = au(1)+x0u'*bLu+x0u'*Bu*x0u

Bl = [al(4),0.5*al(6);

    0.5*al(6),al(5)];

bLl = [al(2);al(3)];

Binvl = inv(Bl);

x0l = -0.5*Binvl*bLl

y_predl = al(1)+x0l'*bLl+x0l'*Bl*x0l

%% Consider the bound of y_pred

N1 = min(N):5:max(N);

S1 = min(S):2:max(S);

[Nm,Sm] = meshgrid(N1,S1);

yu = au(1)+au(2)*Nm+au(3)*Sm+au(4)*Nm.*Nm+au(5)*Sm.*Sm+au(6)*Nm.*Sm;

surf(Nm,Sm,yu),colormap HSV
```

```
hold on
yl = al(1)+al(2)*Nm+al(3)*Sm+al(4)*Nm.*Nm+al(5)*Sm.*Sm+al(6)*Nm.*Sm;
surf(Nm,Sm,yl)
hold on
%% plot center and bound of the response surface
yc = ac(1)+ac(2)*Nm+ac(3)*Sm+ac(4)*Nm.*Nm+ac(5)*Sm.*Sm+ac(6)*Nm.*Sm;
surf(Nm,Sm,yc)
colormap HSV
hold on
xlabel('Nitrogen (kg/ha)')
ylabel('Sulphur (kg/ha)')
zlabel('Yield of Paddy (kg/ha)')
%% Find the nature of the stationary point
[V,D] = eig(Bc)
% The eigenvalues of B are -ve.
% Thus, the stationary point is a point of maxima.
```