CHAPTER 3

RESPONSE SURFACE METHODOLOGY

3.1   Introduction

Response surface Methodology (RSM) is a collection of mathematical and statistical techniques useful for analysing problems where several independent variables influence a dependent variable or response, and the goal is to optimise this response. We denote the independent variables by $x_1, x_2, x_3 \ldots \ldots \ldots x_k$. It is assumed that these variables are continuous and controllable by the experimenter with negligible error. The response, 'y' is assumed to be a random variable. RSM is used for the design and analysis of experiments; it seeks to relate an average response to the value of quantitative variables that effect response. RSM answers different kind of questions, such as the following [68].

  (i)   How is a particular response affected by a given set of input variables over some specified region of interest?

  (ii)   To what level the inputs are to be controlled, to give a product simultaneously satisfying desired specifications?

  (iii) What values of inputs will yield a maximum for a specific response, and what is the nature of response surface close to the maximum?

Figure 3.1 is a flow diagram, showing possible paths that can be taken in response surface studies.

The relationship between the dependent variable and independent variables can be represented as [68, 70].

$$y = f(x_1, x_2, x_3, \ldots\ldots x_k) + \varepsilon \qquad (3.1)$$

where, ε represents the noise or error observed in the response 'y'.

If we denote the expected response by

$$E(y) = f(x_1, x_2, x_3, \ldots\ldots x_k) = \eta \qquad (3.2)$$

then, the surface represented by

$$\eta = f(x_1, x_2, x_3, \ldots\ldots x_k) \qquad (3.3)$$

is called the **response surface**. This surface is drawn between some response such as material removal rate whose levels are denoted 'm', and number of quantitative variables (or factors), whose levels are denoted by $x_1, x_2, x_3 \ldots\ldots x_k$.

The feature of the surface of greatest interest is often the values of variables $x_1, x_2, x_3 \ldots\ldots x_k$, for which m is a maximum or minimum.

In most RSM problems, the form of the relationship between the response and the independent variables is unknown. Thus, the first step in RSM is to find suitable approximation for the true functional relationship between y and the set of independent variables. Usually, a low-order polynomial in some region of the independent variable is employed.

If the response is well modelled by a linear function of the independent variables, then the approximating function is the **first-order model** [68-85].
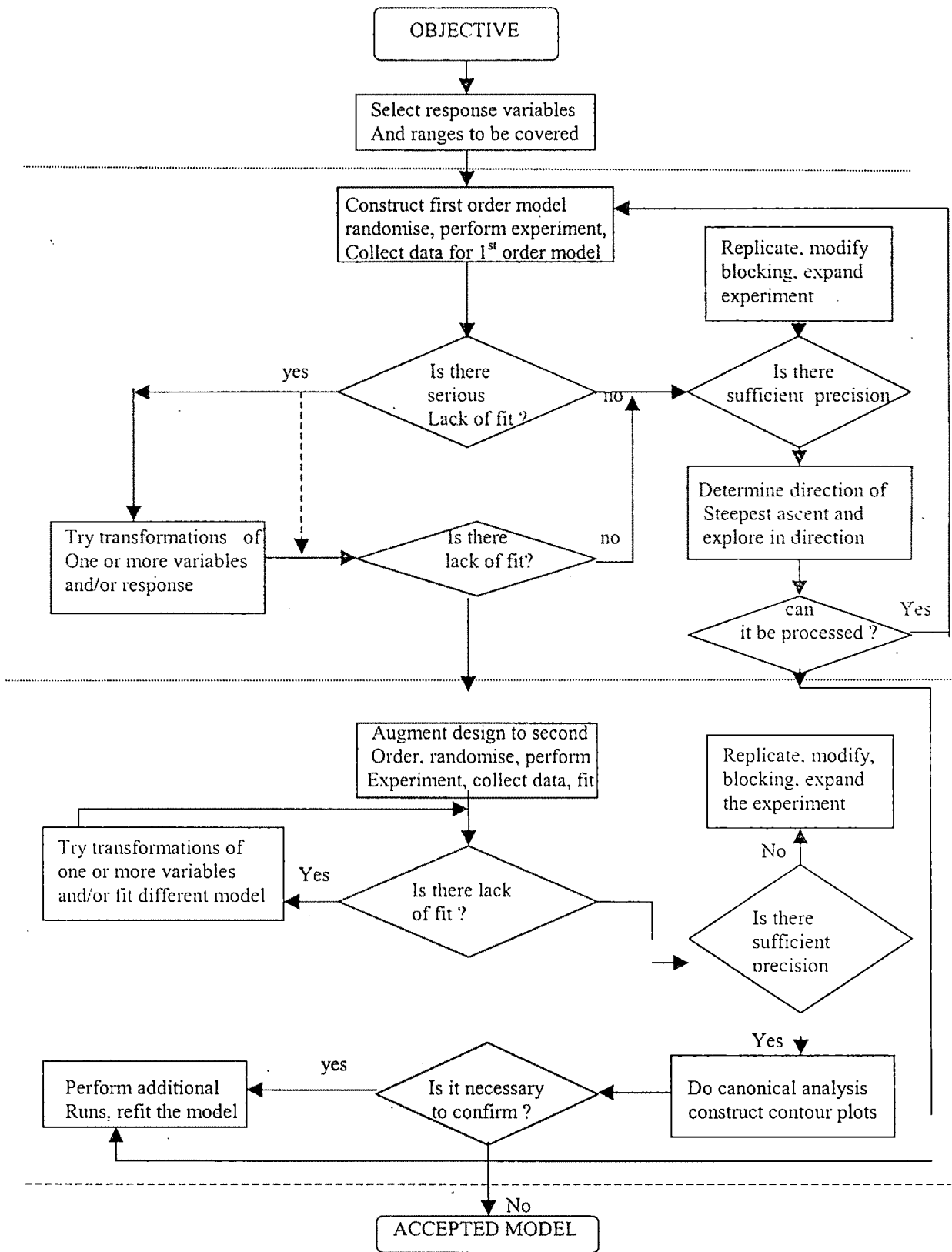
OBJECTIVE

Select response variables
And ranges to be covered

Construct first order model
randomise, perform experiment,
Collect data for 1ˢᵗ order model

Replicate. modify
blocking. expand
experiment

Is there
serious
Lack of fit ?   no   Is there
sufficient precision

yes

Determine direction of
Steepest ascent and
explore in direction

Try transformations of
One or more variables
and/or response

Is there
lack of fit?   no

can
it be processed ?   Yes

Augment design to second
Order. randomise, perform
Experiment, collect data, fit

Replicate. modify,
blocking. expand
the experiment

Try transformations of
one or more variables
and/or fit different model

Yes

Is there lack
of fit ?

No

Is there
sufficient
precision

Perform additional
Runs, refit the model   yes   Is it necessary
to confirm ?   Do canonical analysis
construct contour plots

Yes

No

ACCEPTED MODEL

Figure 3.1 Flow Diagram Showing Different Paths in RSM

$$y = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots\ldots\beta_k x_k + \varepsilon \qquad (3.4)$$

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

$$y = \beta_0 x_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum\sum_{i<j} \beta_{ij} x_i x_j + \varepsilon \qquad (3.5)$$

Almost all RSM problems use one or both of these models. Of course, it is unlikely that a polynomial model will be a reasonable approximation of the true functional relationship over the entire space of the independent variables, but for a relatively small region they usually work.

The method of least squares is used to estimate the parameters in the approximating polynomials. The response surface analysis is then performed using the fitted surface. If the fitted surface is an adequate approximation of the true response function, then analysis of the fitted surface will be approximately equivalent to analysis of the actual system. The model parameters can be estimated most effectively if proper designs are used to collect the data. Designs for fitting response surface are called **response surface designs.**

3.2   First Order Designs

In this case the response surface is fitted with polynomials of first degree.

$$\eta = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots\ldots + \beta_k x_k \qquad (3.6)$$

or $$\hat{y} = b_0 x_0 + b_1 x_1 + b_2 x_2 + \ldots\ldots\ldots + b_k x_k \qquad (3.7)$$

59

Fitting of a polynomial can be treated as a particular case of multiple linear regressions. The 2^k factorial design in single or fractional replication, are convenient in exploratory work, for fitting a linear relation between the response and variables. Box et al [78, 85-86] have discussed the suitable fractional designs for exploring response surfaces.

These designs do not provide any estimate of the experimental error variance. This can be obtained

(i)   by replication of the whole experiment;

(ii)   by the use of an estimate from previous experimentation, if there is convincing evidence that error variance remains stable through time;

(iii)   by adding to the 2^k factorial a number of tests made at the point at which all 'x' have the value '0' in the coded scale.

The linear equation in k (x) variables contains (k+1) regression coefficients that must be estimated. The smallest experiment to which a linear equation can be fitted is one that has (k + 1) observations.

If there is no lack of fit and sufficient precision is obtained, on the basis of this, direction of steepest ascent is determined and exploration is continued. Otherwise try with transformations of one or more variables and response. Careful blocking and expanding the size of design can increase precision. If satisfactory fit and precision is not obtained then second order design are to be resorted.

3.3   Second – Order Design

The general form of second-degree polynomial can be represented as equation 3.8 [68, 70].

$$y = \left(b_0 x_0 + b_1 x_1 + b_2 x_2 + ... + b_k x_k\right) + \left(b_{12} x_1 x_2 + b_{13} x_1 x_3 + ... + b_{(k-1)} x_{(k-1)} x_k\right)$$
$$+ \left(b_{11} x_1^2 + b_{22} x_2^2 + ... + b_{kk} x_k^2\right)$$

(3.8)

The source contains linear terms x_1, x_2 .... x_k; squared terms x_1^2, x_2^2, ... x_k^2 and cross product terms $x_1 x_2$, $x_1 x_3$, ... $x_{k-1} x_k$.

In order to estimate the regression coefficients in this model, each variable must take at least three different levels. Use of factorial designs of 3^k will be necessary in this case. Main disadvantage of a 3^k factorial design is that with more than three variables experiments become large. Further, Box and Wilson pointed out that coefficient b_{11}, b_{22} .... b_{kk} of squared terms are estimated with relatively low precision.

Box and Wilson developed a new design for fitting the second order response surface. The composite designs are constructed by adding further treatment combinations to the first order design. Central composite designs consist of additional (2k + 1) treatments,

(0, 0, 0, ... 0) ; (-α, 0, 0, ... 0);   (α, 0, 0, ... 0); (0, -α, 0, ...0);

(0, α, 0, ...0);   ....   (0, 0, 0, ...α)

Total number of treatment combinations is (2^k + 2k + 1). The value of 'α' can be chosen to make the regression coefficient orthogonal to one another. Central composite design can be fitted into a sequential program of experimentation.

3.3.1 Non-Central Composite Design

This design has k extra points, one for each factor. Non-central Composite Design is used when 2^k factorial experiments have suggested that the point of maximum response is near to one of the factor combinations than to the centre. For three-factor systems central composite and non-central composite designs are illustrated in Figure 3.2.

3.3.2 Rotatable Second Order Design

The design used for fitting a second order response surface should be easy to compute. Box and Hunter proposed the criterion of rotatability. In a rotatable design, standard error is the same for all points.

Box and Hunter showed that a rotatable design is obtained by making test at n_s points equally spaced around circumference of a circle in the x_1, x_2 plane with centre (0, 0), plus one or more tests at the centre. The points on the circumference lie at the vertices of a regular polygon inscribed in a circle. Since there are six regression coefficients to be determined when k = 2, the smallest design consists of a pentagon plus one point at the centre.

The replicated points at the centre have two purposes. They provide *(n_o −1)* degrees of freedom for estimating the experimental error, and they determine the precision of y at the centre. If there are many replications of the centre point, the standard error of y is low at the centre and with a few replications at the centre standard error of y may be greater.

As a compromise, Box and Hunter suggested that the number of centre points be chosen so that standard error of y is approximately the same at the centre as at all points on circle with radius '1' in coded unit.

Box and Hunter have derived rotatable second order design for any number of independent variables x_1, x_2, x_3, ..., x_k. These designs are composed using the vertices of regular figures or combinations of regular figures, with one or more points at the centre of design array.

For rotatability, the axis arms of measure polytope should be $\alpha = 2^{k/4}$. The total number of points required for rotatable central composite design is $2^k + 2^k + n_o$, where n_o equals the number of points at the origin.

3.3.3 Determination of Factor Levels for Optimum Condition

At the outset, the experimenter must decide which factors are to be included in the experiments. Sometimes there are initially as may as a dozen or more factors that might influence the response. Some preliminary weeding out of factors that seem likely to be of minor importance is necessary. The range within which the level of each factor is to be varied must also be selected [68-85].

3.4 The Method of Steepest Ascent

RSM is a **sequential procedure**. Often, when we are at a point on the response surface that is remote from the optimum, such as the current operating conditions in

figure 4.3, there is little curvature in the system and the first order model will be appropriate. The objective is to lead the experimenter rapidly and efficiently along a path of improvement toward the vicinity of the **optimum**. Once the region of the optimum has been found, a more elaborate model, such as the second-order model, may be employed, and an analysis may be performed to locate the optimum. From figure 3.3, we see that the analysis of a response surface can be thought of as "climbing a hill," where the top of the hill represents the point of maximum response. If the true optimum is a point of minimum response, then we may think of "descending into valley".

The method of steepest ascent is a sequential procedure for moving sequentially along the path of steepest ascent, that is, in the direction of the maximum increase in the response. Of course, if minimization is desired, then we call this technique the method of steepest decent [68-85]. The eventual objective of the RSM is to determine the optimum operating conditions for the system or to determine a region of the factor space in which operating requirements are satisfied. When we are remote from the optimum, we usually assume that a first order model is an adequate approximation to the true surface in a small region of the x's.

Box and Wilson proposed the method of steepest ascent/decent. The maximum is located by means of a series of experiments, each planned from the result of the proceeding ones. At the end of each experiment a polynomial approximation to response surface is fitted to the results and is used to determine the nature of the next experiment. The first experiment has two purposes.

(i)   To fit linear equation

(ii) To test whether the linear approximation fits within the limits of experimental

     errors.

The 2^k factorial or functional designs are useful for this purpose.
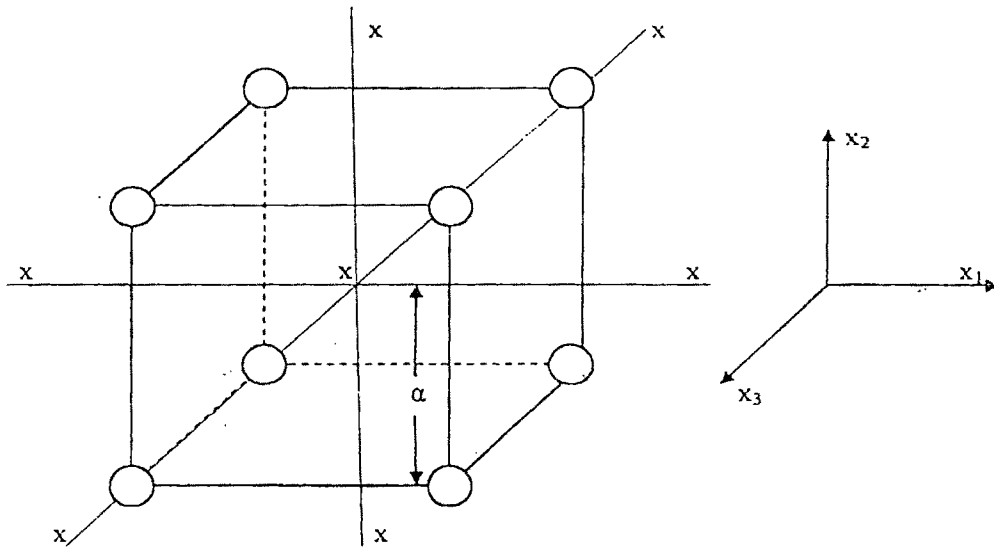
Central Composite Design
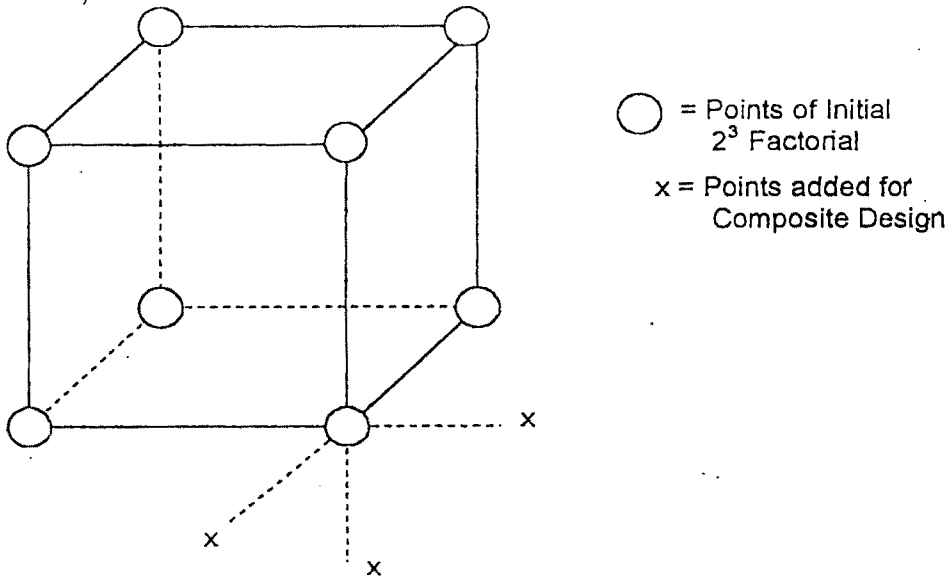
Non Central Composite Design

Figure 3.2 Central and non-Central Composite Design [68, 70]

Table 3.1 Central Composite Rotatable Designs [70].

k	n_c	n_s	n_o	Total	α	Blocking
2	4	4	5	13	1.414	I – Four Points of Squares plus two centre points II – Four Points at star plus two centre points. Total number of points in blocked design = 12
3	8	6	6	20	1.682	I & II – Formed from ½ replicates of 2^3 factorial each with two centre points III – Six points of star plus two centres points. Total number of points in blocked design = 20
4	16	8	7	31	2.000	I & II – Formed from ½ replicates of 2^3 factorial each with two centre points. Blocking is orthogonal. Total number of points in blocked design = 30
5	32	10	10	52	2.378	I to IV – Formed from ¼ replicates of 2^5 factorial design each with two centre V – 10 of the star design plus four centre points. For orthogonal blocking take = 2.336. Total number of points in blocked design = 54
5	16	12	15	91	2.000	I – Formed of the 16 points of the ½ replicates of 2^5 factorial plus six centre points. II – 110 points star design plus one centre point. For orthogonal blocking take = 2.336 Total number of points in blocked design = 32 Blocked design = 90
6	64	12	15	91	2.828	I to VIII – Formed from the eight 1/8 replicates of the 2^6 factorial, each with one centre points. Total number of points in block design = 9 IX – 12 points of star design plus six centre points. For orthogonal blocking take = 2.336 Total number of points = 32.
6	32	12	9	53	2.378	I & II – Formed from ½ fractions of the ½ replicate of 2 factorial design. Each block contains 16 points plus four centre points. III – 12 points of star design plus two centre points. For orthogonal blocking take = 2.336 Total number of points in blocked design = 54
7	128	14	21	163	3.333	I to XVI – Formed from 1/16 replicates of 2factorial, each block containing 8 points plus an additional centre point. XVII – 14 points of star design plus 11 centre points For orthogonal blocking take 2.364. Total number of points in blocked design = 169
7	64	14	14	92	2.828	I to VIII – Formed from 1/8 fractions of ½ replicates of the 2 factorial design. Each block containing 8 points plus one point at the centre. IX – points of the star design plus four centre points. For orthogonal blocking take = 2.364. Total number of points in blocked design = 90

66

When the first experiment is complete the region of experimentation is shifted to another

set of level of x's. This set is to be chosen so that maximum expected increase in

response occurs. If the centre of the first experiment is taken as the origin, the problem

is to move from the origin, with x co-ordinates (0, 0, 0 ......0) to the point, say P, with

co-ordinates $(x'_1, x'_2, x'_3, ..., x'_k)$, so that the response $\Phi(x'_1, x'_2, x'_3, ..., x'_k)$ is

maximized.

For the fitted first order model, given by equation 3.9, the contours of η are a series of

parallel lines such as that shown in figure 3.4. The direction of steepest ascent is the

direction in which η increases most rapidly. This direction is parallel to the normal to the

fitted response surface. We usually take as the path of steepest ascent the line through the

centre of the region of interest and normal to the fitted surface. Thus, the steps along the

path are proportional to the regression coefficients $(\hat{\beta}_i)$. The experimenter based on

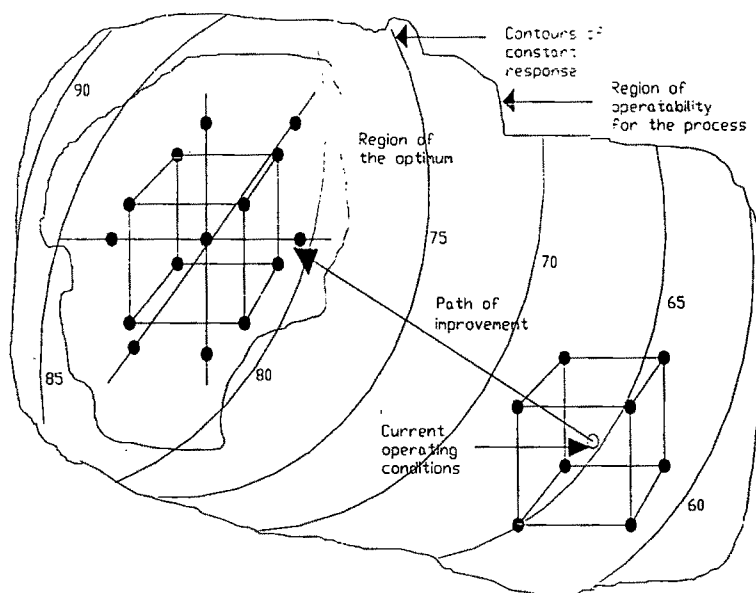process knowledge or other practical considerations determines the actual step size.

Figure 3.3.The sequential nature of RSM [68]

Experiments are conducted along the path of steepest ascent until no further increase in response is observed.

$$\eta = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i \qquad (3.9)$$

The change in the response depends upon the size of the jump that is made from 0 to point P. By geometrical analogy, the 'distance' r from 0 to P is defined as

$$r = \sqrt{\left(x_1'^2 + x_2'^2 + x_3'^2 + \ldots\ldots\ldots + x_k'^2\right)} \qquad (3.10)$$

Path of this steepest ascent is determined. Exploration is continued with new experiments. In course of time a situation is reached in which 2^k factorial designs give one of the following

(i)   The linear equation still appears to fit, but all coefficients bi are small. This is the indication of approach of Plateau.

(ii)   The lack of fit terms shows that the linear approximation is inadequate. This indicates that the experiment is carried out in the region in which curvature of the surface exists.

For further exploration, second order designs are used. Second order designs are reconstructed by adding additional points to the last 2^k factorial experiments.

3.5   Canonical Analysis

Once the stationary point is found, it is usually necessary to characterise the response surface in the immediate vicinity if this point. Characterise means determine

whether the stationary point is a point of maximum or minimum response or a saddle point. We also usually want to study the relative sensitivity of the response to the variables x_1, x_2, x_3, ..., x_k. The most straightforward way to do this is to examine a contour plot of the fitted model. If there are only two or three process variables, the construction and interpretation of this contour plot is relatively easy. However, even when there are relatively few variables, a more formal analysis, called the Canonical analysis, can be useful.
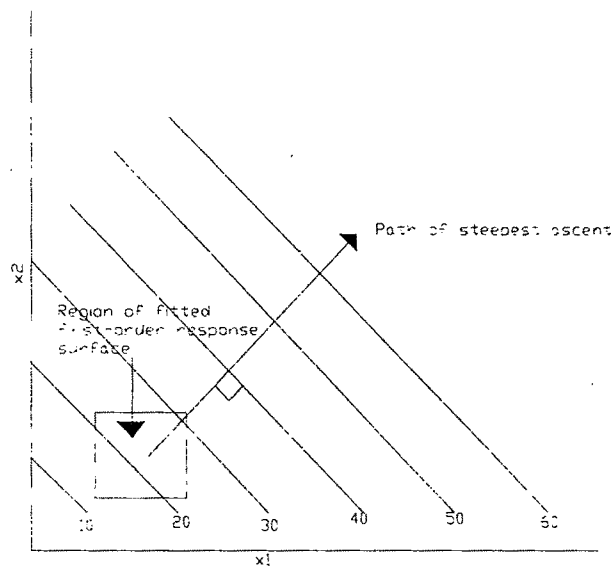

Figure 3.4   First order response surface and path of steepest ascent [68]

To gain an understanding of the nature of the response surface, canonical analysis can be effectively used. Canonical analysis transforms the estimated regression equation into a simpler form and interprets the resulting expression in terms of geometric concepts.

A canonical equation is an equation transformed to a new co-ordinate system by translation of the centre of the old co-ordinates to extreme of the response surface with subsequent rotation of the axis to achieve symmetry. Figure 3.5 illustrates the canonical transformation procedure. The translation corresponds to deletion of the linear terms; the

rotation corresponds to the deletion of the cross product terms. Hence canonical equations consists only quadratic effects.
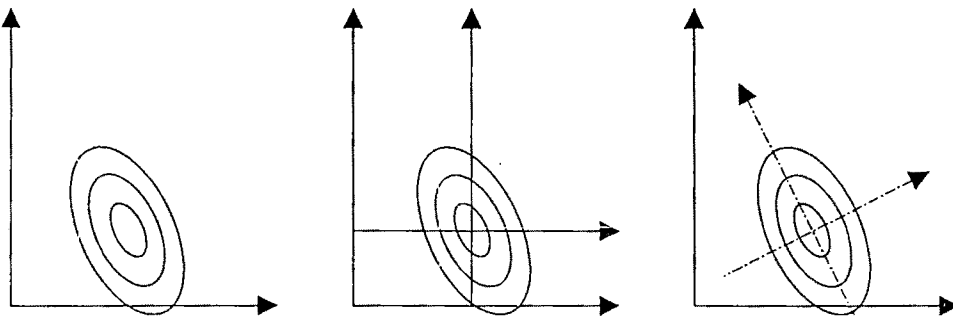

Figure 3.5 Transformation of coordinates in to canonical coordinates [70]

Second-degree equation for the two variable x_1 and x_2 is given by equation 3.11

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_{11} x_1^2 + b_{22} x_2^2 + b_{12} x_1 x_2 \qquad (3.11)$$

The coefficient b_1 and b_2 are called the linear effect; b_{11} and b_{22} the quadratic effects and b_{12} the interactions effect. Canonical equation is

$$\dot{y} = y_s = b_{11} x_1^2 + b_{22} x_2^2 \qquad (3.12)$$

Where y_s = predicted response at the centre of response surface.

Figure 3.6 illustrates the contours for second order models with two independent variables and Table 3.2 gives the interpretation. The second order response surface with three independent variables is represented by equation 3.13.

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{11} x_1^2 + b_{22} x_2^2 + b_{33} x_3^2 + b_{12} x_1 x_2 + b_{13} x_2 x_3 \qquad (3.13)$$

$$\hat{y} = \hat{y}_s = b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 \qquad (3.14)$$

In general, analysis of k dimensional second-degree fitted surface will follow the same lines. From the $(k+1)$ $(k+2)/2$ coefficients of the original equation, calculate;

(i)   The k co-ordinates of the new centre x_{1s}, x_{2s}, ...x_{ks} and value y_s of the response at this point.

(ii)   The canonical form of the equation

$$\hat{y} = \hat{y}_s = b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 + ...+ b_{kk}x_k^2 \qquad (3.15)$$

which contain k coefficients.

(iii) The k equations of the new co-ordinates x's in terms of the old co-ordinates x's.

Table 3.2 Interpretation of the Canonical Equation [70]

| Sr. No.	Case Relations	Coefficient Signs B₁₁ O₁₁		Types of Curves	Geometric Interpretation	Centre	Figure 3.6
1.	$B_{11} = B_{22}$	-	-	Circles	Circular Hill	Maximum	(a)
2.	$B_{11} = B_{22}$	+	+	Circles	Circular Valley	Minimum	(a)
3.	$B_{11} = B_{22}$	-	-	Ellipses	Elliptical Hill	Maximum	(b)
4.	$B_{11} = B_{22}$	+	+	Ellipses	Elliptical Valley	Minimum	(b)
5.	$B_{11} = B_{22}$	+	-	Hyperbola	Symmetrical Saddle	Saddle point	(c)
6.	$B_{11} = B_{22}$	-	+	Hyperbola	- do -	- do -	(c)
7.	$B_{11} = B_{22}$	+	-	Hyperbola	Elongated Saddle	- do -	(d)
8.	$B_{22} = 0$	-	0	Straight lines	Stationery Ridge	None	(e)
9.	$B_{22} = 0$	-	0	Parabolas	Rising Ridge	At Infinity	(f)
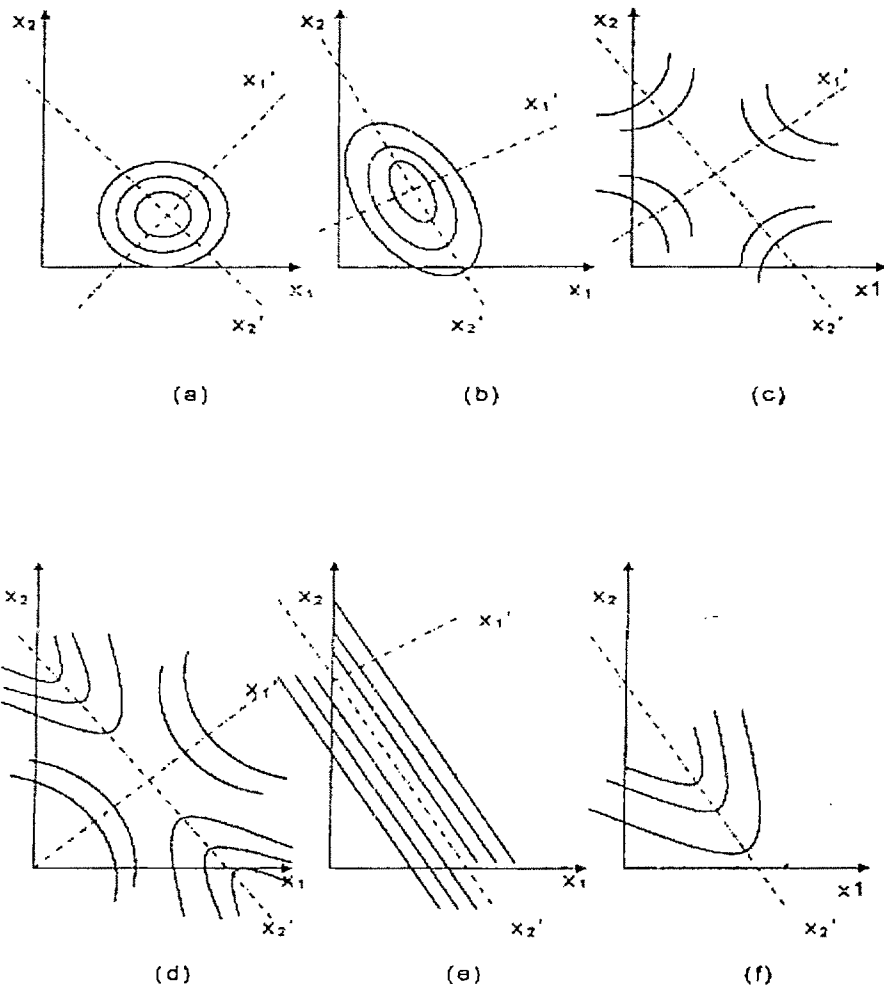
Figure 3.6 Contours for second order models with two independent variables [70]

3.6 Conclusion

Response Surface Methodology is the tool used to obtain the empirical models of the machinability parameters in terms of the variables of electro discharge machining after identifying their relative contribution using Taguchi Technique. The method of Steepest Ascent/Descent can be effectively used to identify the optimum values of the responses.