CHAPTER 10

STATISTICAL OPTIMIZATION

STATISTICAL OPTIMIZATION

10.1 INTRODUCTION

The application of statistics[1] has become popular in the research laboratories of the pharmaceutical companies, making their research work authentic. Part of this recent upsurge has been due to the recognition by research scientists that the application of statistics can be a useful tool in the design, analysis and interpretation of experiments. US Food & Drug Administration (FDA) and several other government agencies have made it regulated through a set of stringent rules, which has made the application of statistical techniques, virtually a necessity.

Theories of statistics are applied in sampling and testing for quality control, stability studies, process validation and design of pre-clinical protocols. The resulting data of the statistical analysis are routinely applied by the pharmaceutical industries to meet both FDA recommendations and their in house standards. Applied statistics, in the field of pharmaceutical research, are considered to be more authentic. Recent publications involving statistics in optimization and experimental design indicates its importance in formulation research.

Statistical experiments[2] are performed by researchers, virtually, in all fields of science, to discover the trend and significance of a particular process of a system. An experiment can be defined as a test or a series of tests, where intentional changes are made in the input variables of a process or a system so as to observe and identify the reasons for changes in the output response.

Great statisticians, namely, Sir Ronald A. Fisher and F. Yates have contributed a great deal in the subject of **Design of Experiments**. The contribution of R.A.Fisher in the logic of scientific method and experimentation leads to the development of the subject of **Design and Analysis of Experiments**[3]. The purpose of the theory of the "Design of Experiments' is to ensure that the researchers obtains data relevant to his hypothesis in an economical way. The researchers are always in a position of being able to spend only a limited amount of time, labor, money and other resources on his investigations. These limitations can be solved in an efficient way using suitable statistical methods. The experimental data obtained from a research are liable to errors which can be minimized with the help of statistical applications[3].

The subject of **Experimental Design** is a vast and complicated field of mathematics and statistics with extensive practical application. According to R.A.Fisher, "The design of experiment is, however, too large a subject, and of too great importance to the general body of scientific workers for any incidental treatment to be adequate".[3] The principle of experimental design can be understood with a sound background of statistical procedures. With that knowledge the experimenter can pick up a design, which have been widely successful in many field, and suits his own experimental requirements. The subject of Experimental Design has been dealt in great depths in various classic books[2-6].

Experimental Design[2] methods are widely applied in various branches of sciences and technology to learn about how systems or processes work. It is a critically important tool in the Engineering world with extensive application in the development of new processes and improvement in the performances of a manufacturing process. The application of experimental design techniques early in process development can result in[2]

i. Improved process yields.
ii. Reduces variability and closer conformance to nominal or target requirement.
iii. Reduces development time.
iv. Reduces overall cost.

Experimental design methods also play a major role in engineering design activities, where new products are developed and existing ones improved. Some applications of experimental design in engineering design include

1. Evaluation and comparison of basic design configurations.
2. Evaluation of material alternatives.
3. Selection of design parameters so that the product will work well under a wide variety of field conditions, that is, the product is robust.
4. Determination of key product design parameters that impact product performance.

The use of experimental design in these areas can result in products that are easier to manufacture, products that have enhanced field performance and reliability, lower product cost and shorter product design and development time.

10.2 Types of Experimental Design[2]

The general approach to planning and conducting the experiment is called **The Strategy of Experimentation.** There are several strategies available, having their own advantages and limitations. The need for these strategies arises from the frequent practices of so-called **Best Approach** used by engineers and scientists[2]. It often works reasonably well because the experimenters usually have a great deal of technical, theoretical and practical knowledge of the system they are investigating. However, this method is often time-consuming and does not guarantee the best possible solution.

Following is a list of various experimental designs that could be used to get the best possible optimal solution in reasonably less time than the best guess approach[2].

1. Simple Comparative Experiments.
2. Experiments with a Single Factor(ANOVA)
3. Randomized Complete Block Design.
4. Latin Square Design.
5. Graeco – Latin Square Design.
6. Balanced Incomplete Block Design.
7. Factorial Design.
    i. Full Factorial Design.
    ii. Factorial Design Blocking and Confounding.
    iii. Mixed level Factorial Design.
    iv. Fractional Factorial Design.
    v. Mixed Level Fractional Factorial Design.
    vi. Factorial Design with Random Factor.
8. Multistage Nested Design.
9. Design with both Nested and Factorial Factors.
10. Split – Plot Design.
11. Split – Split – Plot Design.
12. Mixture Experiments.

10.2.1 Factorial Design

Many experiments involve the study of the effects of two or more factors. In general, Factorial Designs are most efficient for this type of experiments. Factorial designs are used in experiments where the effects of different factors, or conditions, on experimental results are to be elucidated. By a factorial design we mean that in each complete trial or replication of the experiment, all possible combinations of the factors are investigated. When factors are arranged in a factorial design, they are often said to be crossed[2,4].

In earlier times separate experiments were usually being devoted to each factor studying their effect one by one. Fisher[3] pointed out that important advantages are gained by combining the study of several factors in the same factorial experiments. Factorial experimentation is highly efficient because every observation supplies information about all the factors included in the experiment. In addition, this is a workman like method of investigating the relationships between the effects of different factors.[5]

In the field of Pharmaceutical Sciences, some practical examples where factorial design are optimized are experiments to determine the effect of pressure and lubricant on the hardness of a tablet formulation, to determine the effect of disintegrant and lubricant concentration on tablet dissolution or to determine the efficacy of a combination of two active ingredients in an OTC cough preparation. The advantages[6] of factorial design may be cited as below:-

- In the absence of interaction factorial design have maximum efficiency in estimating main effects. (The main effect is the effect of a factor averaged over all levels of the other factor.)
- If interaction (i.e. lack of effects of additives) exists, factorial designs are necessary to reveal and identify the interaction to avoid misleading conclusions.
- Since factors are measured over varying levels of other factor, conclusions apply to a wide range of conditions.
- Maximum use is made of the data since all main effect and interaction are calculated from all of the data.
- Factorial designs are orthogonal. All estimated effects and interactions are independent of effects of other factors.

10.2.2 Analysis of Factorial Experiments[4,7]

Factorial designs usually involve selection of critical factors, which are expected to affect the outcome of experiments. The factors are usually taken with at least two or higher level. The levels are coded by some standard mathematical transformations for simplifying the subsequent calculation. The results obtained are analyzed by various statistical and mathematical procedures to get an insight into the characteristics of the system.

A **factor** is an assigned variable such as concentration, temperature, lubricating agent, drug treatment or diet. The choice of factors to be included in an experiment depends on experimental objective and is predetermined by the experimenter. A factor can be quantitative or qualitative. Single factor design fit the category of one-way ANOVA design. The **levels** of a factor are values or designations assigned to the factor. The **runs** or **trials** that comprise factorial experiment consist of all combination of all levels of all factors. The **effects** of a factor are the change in response caused by varying the level(s) of the factor. The **main effect** is the effect of a factor averaged over all levels of the other factors.

Before starting any factorial design, it is customary to write down all the possible factors that constitute the different runs of the experiment. Factorial Design Experiments can determine the effect of several factors and their interactions simultaneously. The factorial design may be useful for screening purpose or an aid in identifying individual effects of complex systems. It offers a good degree of efficacy and possibility of detecting interactions between factors. Factor effectiveness can be expressed with a mathematical model which explains the influence numerically. In our case we took the help of 3^2 factorial designs to optimize the release data. In 2^2 factorial designs there are two variables each at two levels i.e. at low and high level and total number of experiments is 4. In 3^2 factorial designs there are two variables each at three levels i.e. at low, medium and high level and the total number of experiments is nine. This is usually presented in a tabular form as shown in Table 10.1 and 10.2 for a 2^2 and 3^2 Factorial Design respectively.

Table 10.1 2^2 Factorial Design

A	B
-	+
-	-
+	+
+	-

Where, A and B are two variables and (-) indicates factor at low level and (+) factor at high level.

Table 10.2 3^2 Factorial Design

A	B
-	--
-	0
-	+
0	-
0	0
0	+
+	-
+	0
+	+

Where, A and B are two variables and (-) indicates factor at low level, (0) indicates factor at medium level and (+) indicates factor at high level.

The experiments are done as laid in the contingency table and desired are then subjected to various statistical analysis procedures. Basically the analysis is done to either prove the Null Hypothesis (H_0) or to disprove it, that is, in other words to prove the Alternative Hypothesis (H_a). Generally, a linear statistical model is selected for testing of hypothesis, such as: $\mathbf{x_{ij} = \mu + \alpha_i + e_{ij}}$ ..... **(10.1)** where,

$\mathbf{x_{ij}} \rightarrow$ the j-th observation in the i-th group ; $\boldsymbol{\mu} \rightarrow$ the general effect ; $\boldsymbol{\alpha_i} \rightarrow$ special effect to the i-th population ; $\mathbf{e_{ij}} \rightarrow$ error component.

The analysis of results can be done from various angles or their combinations, by using

1. Analysis of Variance (ANOVA)
2. Regression Analysis.
3. Response Surface Methodology (RSM).

Of particular importance is the fitting of data to appropriate mathematical or statistical model, which can adequately and reliably describe the behavior of the system under investigation. In this regard, the regression analysis and RSM is of particular importance since these not only help in testing the hypothesis but also keep open scope of further improvement of the models and system optimization.

10.3 Analysis of Variance (ANOVA)

Analysis of Variance[1] (ANOVA) is inextricably connected to experimental design. Experiments that are conceived to compare, estimate and test such effects as drug treatments, formulation differences and analytical methods can be designed to yield an optimal return for effort expended. The experimental results may then be analyzed by ANOVA techniques. A good experiment speaks for itself; the conclusions are often obvious without complicated mathematical treatment. If designed properly most experiments can be easily analyzed. In a poorly designed experiment on the other hand, more than one factor may contribute to an experimental result with no way of untangling the effects of the factors. Analysis of Variance separates the total variation in the data into parts, each of which represents variation caused by factors imposed on the experiment. A properly designed experiment allows clear estimate of such variation or, at least, can identify the confounding factors, if present.

In statistics, analysis of variance (ANOVA) is a collection of statistical models, and their associated procedures, in which the observed variance is partitioned into components due to different explanatory variables, usually called factors in Design of experiments. The initial techniques of the analysis of variance were pioneered by the statistician and geneticist R. A. Fisher in the 1920s and 1930s, and are sometimes known as Fisher's ANOVA or Fisher's analysis of variance, due to the use of Fisher's F-distribution as part of the test of statistical significance.

10.3.1 Overview

There are three conceptual classes[8] of such models:

* Fixed-effects model assumes that the data come from normal populations which may differ only in their means. (Model 1)

* Random-effects models assume that the data describe a hierarchy of different populations whose differences are constrained by the hierarchy. (Model 2)

* Mixed effects models describe situations where both fixed and random effects are present. (Model 3)

In practice, there are several types of ANOVA depending on the number of treatments and the way they are applied to the subjects in the experiment:

One-way ANOVA is used to test for differences among three or more independent groups.

One-way ANOVA for repeated measures is used when the subjects are subjected to repeated measures; this means that the same subjects are used for each treatment. Note that this method can be subject to carryover effects.

Factorial ANOVA is used when the experimenter wants to study the effects of two or more treatment variables. The most commonly used type of factorial ANOVA is the 2x2 (read: two by two) design, where there are two independent variables and each variable has two levels or distinct values. Factorial ANOVA can also be multi-level such as 3×3, etc. or higher order such as 2×2×2, etc. but analyses with higher numbers of factors are rarely done because the calculations are lengthy and the results are hard to interpret.

When one wishes to test two or more independent groups subjecting the subjects to repeated measures, one may perform a factorial Mixed-Design ANOVA, in which one factor is independent and the other is repeated measures. This is a type of mixed effect model.

Multivariate analysis of variance (MANOVA) is used when there is more than one dependent variable.

10.3.2 Different Models

A. Fixed-effects model

The fixed-effects model of analysis of variance applies to situations in which the experimenter has subjected his experimental material to several treatments, each of which affects only the mean of the underlying normal distribution of the "response variable".

B. Random-effects model

Random effects models are used to describe situations in which incomparable differences in experimental material occur. The simplest example is that of estimating the unknown mean of a population whose individuals differ from each other. In this case, the variation between individuals is confounded with that of the observing instrument.

10.3.3 Assumptions

Independence of cases - this is a requirement of the design.

Normality - the distributions in each of the groups are normal (use the Kolmogorov-Smirnov and Shapiro-Wilk normality tests to test it). Some say that the F-test is extremely non-robust to deviations from normality (Lindman, 1974) while others say the opposite (Ferguson & Takane 2005: 261-2).

Homogeneity of variances - the variance of data in groups should be the same (use Levene's test for homogeneity of variances).

10.3.4 Logic of ANOVA

The fundamental technique lies in partitioning the total sum of squares into components related to the effects in the model used. For example, we show the model for a simplified ANOVA with one type of treatment at different levels. (If the treatment levels are quantitative and the effects are linear, a linear regression analysis may be appropriate.)

SS Total = SS Error + SS Treatments, where SS signifies Sum of Squares

The number of degrees of freedom (abbreviated df) can be partitioned in a similar way and specifies the chi-square distribution which describes the associated sums of squares.

df Total = df Error + df Treatments.

10.3.5 Degrees of freedom

Degrees of freedom indicate the effective number of observations which contribute to the sum of squares in an ANOVA, the total number of observations minus the number of linear constraints in the data. The degrees of freedom are the number of participants (for each group) minus 1. This removes the error otherwise produced by the differences in variance of such groups to account for the difference in sample and population variance.

10.3.6 ANOVA on Ranks

As first suggested by Conover and Iman in 1981, in many cases when the data do not meet the assumptions of ANOVA, one can replace each original data value by its rank from 1 for the smallest to N for the largest, and then run a standard ANOVA calculation on the rank-transformed data. "Where no equivalent nonparametric methods have yet been developed such as for the two-way design, rank transformation results in tests which are more robust to non-normality, and resistant to outliers and non-constant variance, than is ANOVA without the transformation. (Helsel & Hirsch, 2002, Page 177)." However Seaman et al. (1994) noticed that the rank transformation of Conover and Iman (1981) is not appropriate for testing interactions among effects in a factorial design as it can cause an increase in Type I error (alpha error). Furthermore, if both main factors are significant there is little power to detect interactions.

10.3.7 Mathematics of ANOVA[2]

In the present study author adopted 3^2 factorial designs without replication that is one observation per cell.

If there are two factors and only one observation per cell, the linear statistical model is

$x_{ij} = \mu + \alpha_i + e_{ij}$ ..... **(10.1)** where,

$x_{ij} \rightarrow$ the j-th observation in the i-th group ; $\mu \rightarrow$ the general effect ; $\alpha_i \rightarrow$ special effect to the i-th population ; $e_{ij} \rightarrow$ error component.

The analysis of variance for this situation is shown in Table 10.3

Table 10.3 ANOVA Scheme for Two-factor Factorial Model without replication

Sources of Variation	Sum of Squares (SS)	Degrees of Freedom (dF)	Mean Square (MS)	Expected mean square
A	SS_A	a - 1	$MS_A = SS_A/(a-1)$	$\sigma^2 + (b \Sigma \tau^2)/(a-1)$
B	SS_B	b -1	$MS_B = SS_B/(b-1)$	$\sigma^2 + (a \Sigma \beta_j^2)/(b-1)$
Residual or (Interaction)	Subtraction	(a-1)(b-1)	$MS_{Residual}$	$\sigma^2 + \Sigma \Sigma (\tau \beta)_{ij}^2$ / (b-1)(a-1)
Total	SS_T	ab - 1		

If we consider two factors for analysis of variance and denote,

$i = i^{th}$ level of factor A

$j = j^{th}$ level of factor B

y_i = Total of all observations under the i^{th} level of factor A

y_j = Total of all observations in the j^{th} level of factor B

y_{ij} = Total of all observations in the ij^{th} cell

$y_{..}$ = Grand total of all observations

a = Number of levels of factor A

b = Number of levels of factor B

SS_T = Total Sum of Squares

$$= \sum_{i=1}^{a} \sum_{j=1}^{b} y^2_{ij} - (y^2/ab) \qquad (10.2)$$

$$SS_A = (1/b) \sum_{i=1}^{a} y^2 i.. - (y^2/ab) \qquad (10.3)$$

$$SS_B = (1/a) \sum_{j=1}^{b} y^2_j - (y^2../ab) \qquad (10.4)$$

$$SS_{Residual} = SS_T - SS_A - SS_B \qquad (10.5)$$

From examining the expected mean squares, it is seen that the error variance σ^2 is not estimable, that is, the two-factor interaction effect $(\tau\beta)_{ij}$ and the experimental error can not be separated in any obvious manner. Consequently, there are no tests on main effects unless the interaction effect is zero. If there is no interaction present, the $(\tau\beta)_{ij} = 0$ for all I and j, and a plausible model is

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \qquad (10.6)$$

Where, I = 1,2,.....,a and j = 1,2,.....,b

If the model (equation 10.6) is appropriate, then the residual mean square in Table 10.3 is an unbiased estimator of σ^2, and the main effects may be tested by comparing MS_A and MS_B to $MS_{Residual}$.

A test developed by Turkey is helpful in determining if interaction is present. The procedures assumes that interaction term is of particularly simple form; namely

$$(\tau\beta)_{ij} = \gamma\tau_i\beta_j \qquad (10.7)$$

where, γ is an unknown constant. By defining the interaction term this way we may use a regression approach to test the significance of the interaction term. The test partitions the

residual sum of squares into a single degree of freedom component due to non-additivity (interaction) and a component for error with [(a-1)(b-1)-1] degrees of freedom. Computationally, we have

$$SS_N = \frac{\left[\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}\, y_i\, y_j - y\,(SS_A + SS_B + y^2/ab)\right]^2}{ab\, SS_A\, SS_B} \qquad (10.8)$$

with one degree of freedom, and $SS_{Error} = SS_N$ with (a-1) (b-1) -1 degrees of freedom. To test for the presence of interaction we compute

$$F_0 = \frac{SS_N}{SS_{Error}/\,[(a-1)\,(b-1)-1]} \qquad (10.9)$$

If $F0 > F_{\alpha,\,1,\,(a-1)\,(b-1)\,-1}$, the hypothesis of no interaction must be rejected.

$$SS_{Error} = SS_{Residual} - SS_N \qquad (10.10)$$

$$SS_{Model} = SS_A + SS_B + SS_N \qquad (10.11)$$

Regression coefficient: $R^2 = (SSModel / SS_T)$ (10.12)

10.4 Regression analysis[2,9]

In statistics, regression analysis examines the dependence of a random variable, called dependent or response variable, on other random or deterministic variables, called independent variables or predictors. The mathematical model of their relationship is known as the regression equation. Well known types of regression equations are linear regression for continuous responses, the logistic regression for discrete responses (both generalize in the generalized linear model) and nonlinear regression.

Besides the dependent and independent variables, the regression equations usually contain one or more unknown regression parameters (constants), which are estimated from given data. Applications of regression include curve fitting, forecasting of time series, modeling of causal relationships and testing scientific hypotheses about relationships between variables.

10.4.1 Introduction

Regression analysis estimates the strength of a modeled relationship between one or more response variables (also called dependent variables, explained variables, predicted variables, or regressands) (usually named Y), and the predictors (also called independent variables, explanatory variables, control variables, or regressors, usually named). These strengths of the relationships given that the model is correct are parameters of the model, which are estimated from a sample. Other parameters which are sometimes specified include error variances and covariances of the variables. The theoretical population parameters are commonly designated by Greek letters (e.g. β), their estimated values by a "hatted" Greek letter (e.g.), and the sample coefficients by a Latin letter (e.g. b). This stresses the fact that the sample coefficients are not the same as the population parameters, but the distribution of those parameters in the population can be inferred from the estimates and the sample size. This allows researchers to test for the statistical significance of estimated parameters and to measure goodness of fit of the model.

Still more generally, regression may be viewed as a special case of density estimation. The joint distribution of the response and explanatory variables can be constructed from the conditional distribution of the response variable and the marginal distribution of the explanatory variables. In some problems, it is convenient to work in the other direction: from the joint distribution, the conditional distribution of the response variable can be derived. Regression lines can be extrapolated, where the line is extended to fit the model for values of the explanatory variables outside their original range. However extrapolation may be very inaccurate and can only be used reliably in certain instances.

10.4.2 History of regression[9]

The term "regression" was used in the nineteenth century to describe a biological phenomenon, namely that the progeny of exceptional individuals tend on average to be less exceptional than their parents, and more like their more distant ancestors. Francis Galton studied this phenomenon and applied the slightly misleading term "regression towards mediocrity" to it. For Galton, regression had only this biological meaning, but his work was later extended by Yule and Karl Pearson to a more general statistical context.

10.4.3 Definitions and notation used in regression

The measured variable, y, is conventionally called the "response variable". Other terms include "endogenous variable," "output variable," "criterion variable," and "dependent variable." The controlled or manipulated variables are called the explanatory variables. Other terms include "exogenous variables," "input variables," "predictor variables" and "independent variables."

10.4.4 Types of regression

Several types of regression analysis can be distinguished; all of these can be seen as special cases of the Generalized Linear Model.

 # Linear or First order regression model

Linear regression is a method for determining the parameters of a linear system. The empirical model relating the response variable to the independent variables are described by the following equation

Linear Model: $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ (10.13)

where, y represents the response, X_1 and X_2 represent the two independent variables. The parameters β_0 signifies the intercept of the plane. β_1 and β_2, called partial regression coefficients, where β_1 measures the expected change in 'y', the response, per unit change in X_1 when X_2 kept constant and vice versa for β_2 .This equation can be rewritten in a general form as:

$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ldots\ldots \beta_k X_k + \varepsilon$ (10.14)

The model is a multiple linear regression model with 'k' regressor variables. The model describes a hyperplane in the k-dimensional space.

Further complex model (eqn.10.15) are often analyzed by multiple linear regression technique by adding interaction terms to the first order linear model.

$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \varepsilon$ (10.15)

where, $X_1 X_2$ is the interaction effect of two variables acting simultaneously.

 # Quadratic Model or Second order regression model

$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X^2_1 + \beta_{22} X^2_2 + \beta_{12} X_1 X_2 + \varepsilon$ (10.16)

If we put, $X^2_1 = X_3$, $X^2_2 = X_4$, $X_1 X_2 = X_5$ and $\beta_{11} = \beta_3$, $\beta_{22} = \beta_4$, $\beta_{12} = \beta_5$, then the above equation gets reduces to a linear model. Any model is linear if the (β) coefficients are linear, regardless of the shape of the response surface that it generates.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \qquad (10.17)$$

The explanatory and response variables may be scalars or vectors. In the case, where both the explanatory and response variables are scalars, then the resulting regression is called simple linear regression. When there are more than one explanatory variable, then the resulting regression is called multiple linear regression. It should be noted that the general formulae are the same for both cases.

Two common techniques for solving linear regression models are using least squares analysis or robust regression.

 # Non linear regression models

A number of nonlinear regression techniques may be used to obtain a more accurate regression. It should be noted that an often-used alternative is a transformation of the variables such that the relationship of the transformed variables is again linear.

 # Non-continuous variables

If the variable is not continuous, specific techniques are available. For binary (zero or one) variables, there are the probit and logit model. The multivariate probit model makes it possible to estimate jointly the relationship between several binary dependent variables and some independent variables. For categorical variables with more than two values there is the multinomial logit. For ordinal variables with more than two values, there are the ordered logit and ordered probit models. An alternative to such procedures is linear regression based on polychoric or polyserial correlations between the categorical variables. Such procedures differ in the assumptions made about the distribution of the variables in the population. If the variable is positive with low values and represents the repetition of the occurrence of an event, count models like the Poisson regression or the negative binomial model may be appropriate.

 # Other models

Although these three types are the most common, there also exist supervised learning and unit-weighted regression.

 # Non parametric regression

The models described above are called parametric because the researcher must specify the nature of the relationships between the variables in advance. Several non-parametric techniques may be also used to estimate the impact of an explaining variable on a dependent variable. Nonparametric regressions, like kernel regression, require a high number of observations and are computationally intensive.

Before selecting a model for regression analysis, model adequacy checking and test of significance of the individual regression parameters and analysis of covariance, lack of fit test etc. are done to check the robustness and suitability of the model.

10.5 Response Surface Methodology (RSM)[2]

The development of pharmaceutical formulations depends on several factors and process parameters. The response variables relating to effectiveness, safety and usefulness must be optimized through a factorial relationship by combining the casual factors. However, this effort addresses a multiobjective optimization problem since it has to circumvent many difficulties in the quantitative approach, like the understanding of the actual relationship between casual factors and individual pharmaceutical responses or the prediction of those formulations that are desirable for as many as possible drug properties.

Due to complex nature of the development of pharmaceutical formulations, some computer based optimization techniques have been proposed in the literature. Among them, factorial design (FD) and response surface methodology (RSM) are the most widely used, and several research efforts have adopted either FD followed by an RSM or solely RSM. FD is a technique that contributes to the structure of data collection process. Through a designed experiment, FD is capable of characterizing the relationship between important and unimportant factors. Nevertheless, it is obvious that FD has no prediction possibilities of best formulation. Regarding the RSM procedure, it consists of

- Composite statistical experimental designs that prepare systemic model of formulations.

- Modeling among factors and response variables of these formulations.
- Parameter prediction by predicting the final responses or by keeping them in the desired ranges needed to be obtained using a polynomial multiple regression analysis, and
- Mathematical optimization algorithms for deciding the best formulation under a set of constrained equations.

The aim of RSM is to find out the optimum operating conditions for a given system, or the way in which a particular response is affected by a set of variables over some specific regions of interest. The first step in the RSM is to find a suitable approximation of the true functional relationship between the dependent variable and the set of independent variables (factors). If knowledge concerning the shape of the true response surface is insufficient, generally first attempts try to approximate the shape by fitting the first order model to the response values. When the first order model suffers from lack of fit arising from the existence of surface curvature, the first order model is upgraded to the second order model.

$$Y = A_0 + \Sigma^{k}_{i=1} Ai + \Sigma^{k}_{i=1} A_{ii}xi^2 + \Sigma^{k}_{i=1} A_{ij}x_ix_{ij} + \varepsilon \qquad (10.18)$$

Where x_i, x_2, ..., x_k are the input variables, which influence on the response Y; A_i (i = 1,2,.....,k), A_{ij} (i = 1,2,....,k; j = 1,2,.....,k) are unknown parameters, and ε is random error.

In this study a second order polynomial model was applied to the response values.

The Response Surface Methodology (RSM) is a collection of mathematical and statistical techniques that are useful for the modeling and analysis of problems in which a response of interest is influenced by several independent variables and the objective is to optimize this response. For a two factor design the response (Y) may be a function of the levels of the factors ($x_1$and x_2) represented by

$$Y = f(x_1, x_2) + \varepsilon \qquad (10.19)$$

where, ε represents the noise or error observed in the response Y. If we denote the expressed response by $E(Y) = f(x_1, x_2) = \eta$, then the surface represented by

$$\eta = f(x_1, x_2) \qquad (10.20)$$

shall be called as Response Surface. To help visualize the shape of a response surface, we often plot the contours of the response surface where each contour corresponds to constant response drawn in the x_1, x_2 plane.

In most RSM problems, including its application to factorial design, the form of the relationship between the response and independent variables is unknown. Therefore, a suitable approximation for the true functional relationship between Y and set of independent variable is found by regression analysis as discussed already. Usually, in a factorial design, a low order polynomial is some region of the independent variable is employed. The polynomial may be a linear or higher order polynomial. Almost all RSM problems utilize one or both of these models.

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.

Since the Response Surface Methodology requires a proper experimental design for the proper estimation of the parameters, coupling of RSM to factorial design is a very efficient and scientific method of data analysis and system optimization.

10.6 Statistical Software

All major statistical software packages, e.g. SAS System, SPSS or Stata, perform various types of regression analysis correctly and in a user-friendly way. Simpler regression can be done in spreadsheets like MS Excel or OpenOffice.org Calc. Complex types of regression runs on special programming languages like Mathematica, R programming language or Matlab. There are many minor softwares specialized in a niche form of regression. Investigator has used 'Analyse – It + General 1.73' for calculation of 2 way ANOVA and Linear Regression. Response surface methodology (RSM) and Factorial analysis was done using 'Design Expert version 7.1.2'.

10.7 Statistical Optimization of the Study Design of Frusemide micropellets

In order to optimize the formulation design in producing calcium alginate micropellets of frusemide the investigator studied the effect of the primary polymer, Sodium alginate, and copolymer, Acrycoat E30D, as two important variables, on the nature and performance of the microparticulate drug delivery system. To get an estimate of the performance of the drug delivery system, the following dissolution parameters were considered as responses of the variables.

i) Zero order release rate constant – Ko (mg / hr)
ii) Amount of drug released in 2 hours (Burst effect) – X_{120} (mg)
iii) Time required for 80% drug release - t_{80} (hr)
iv) Peppas Diffusion coefficient -- n

Referring to the previous (**chapter 9**), covering the *in vitro* dissolution study of the formulations, it has been seen that on a span of overall release of 0 – 9 hours the release mechanism of the drug followed predominantly Zero order release model. Hence, the Zero order release rate constant **Ko** was selected as viable response to be studied statistically. The values of Ko of the nine different formulations were analyzed by 2-way ANOVA and linear regression analysis. The regression equation obtained from the analysis was used to predict the value of Ko with uninvestigated concentrations of Sodium alginate and Acrycoat E30D. Formulations with the said concentrations were then prepared and the actual Ko value obtained was compared with the predicted value.

X_{120}, the amount in mg of drug released in 120 minutes or 2 hours was taken into account so as to have an estimate of burst release mechanism and initial therapeutic dose being made available by the formulations so as to elicit fast onset of action.

t_{80}, the time required for the drug to release 80% of its actual content gives an indirect estimation of the dissolution efficiency of the formulations as calculations of 100% release would have been too ideal to study.

n, the diffusion coefficient factor of Korsmeyer – Peppas[10] equation signifies either the diffusion behavior of the drug from the polymer matrix follows Fickian or Non- Fickian mechanism. From the result obtained in Chapter 9, it is evident that the value of 'n' is not significantly affected by the variation in concentration of the polymers. Statistical

optimization was necessary to justify this fact that 'n' remains insignificant irrespective of the concentration of the polymers.

Response surface methodology (RSM) was employed to study the effect of the polymers on the above four variable.

The concentrations of the independent variables and the corresponding dependent factors obtained from experiments of *in-vitro* dissolution study (Chapter 9) were enlisted in Table 10.4.

Table 10.4 Independent Formulation Variables and their Responses

Formulation Code	Sodium alginate conc.%(w/w)	Acrycoat E30D conc. %(w/w)	K_O (mg/hr)	X_{120} (mg)	t_{80} (hr)	n
F1	1	0	10.865	34.95	4.5	1.7429
F2	1	2	10.209	37.34	5.2	1.7309
F3	1	4	10.299	37.341	6.1	1.4378
F4	2	0	9.3247	16.988	7.2	1.8142
F5	2	2	9.0388	16.183	7.6	2.3807
F6	2	4	7.9428	8.001	7.8	1.8064
F7	4	0	7.5642	23.173	8	1.6173
F8	4	2	6.9212	17.382	8.4	1.6074
F9	4	4	6.1069	11.194	9.3	1.5608

10.7.1 Study on the Effect of Polymers on the Release Rate (Ko) by ANOVA

Analysis of variance (ANOVA) was done using the software 'Analyse-it + General 1.73', the results of which are shown in Table 10.5.

Table 10.5 2-way ANOVA table for the release rate (Ko) of Frusemide from micropellets

Test	2-way between subjects ANOVA				analysed with: Analyse-it + General 1.73
	Effect on zero order				
Comparison	Rate constant (Ko) by Alginate, Acrycoat E30D			DATE	
Performed by	Amitava Ghosh			26th May 2007	
n	9				
Rate constant by Alginate	n	Mean	SD	SE	
1	3	11.346	0.283	0.1634	
2	3	11.270	0.172	0.0995	
4	3	9.074	0.719	0.4152	
Rate constant by Acrycoat	n	Mean	SD	SE	
0	3	10.960	1.058	0.6106	
2	3	10.528	1.184	0.6833	
4	3	10.201	1.634	0.9433	
Source of variation	SSq	DF	MSq	F	p
Alginate	9.994	2	4.997	52.03	0.0014
Acrycoat E30D	0.870	2	0.435	4.53	0.0939
Within cells	0.384	4	0.096		
Total	11.247	8			

Where, n = number of measures, SD – Standard deviation, SE – Standard error, SSq – Sum of Squares, MSq – Mean Square, DF – Degree of freedom, F- Fischer's value, p - Probability

10.7.2 Study on the Effect of Polymers on the Release Rate (Ko) by Linear Regression Analysis

Linear regression analysis was done using the software 'Analyse-it + General 1.73', the results of which are shown in Table 10.6.

Table 10.6 Regression Analysis for the release rate (Ko) of Frusemide from micropellets

Test	Linear regression			analysed with: Analyse-it + General 1.73	
	zero order regression				
Fit	ko v alginate, acrycoat				
Performed by	AMITAVA GHOSH			Date	26 May 2007
n	9				
R^2	0.89				
Adjusted R^2	0.85				
SE	0.4631				
Term	Coefficient	SE	p	99% CI of Coefficient	
Intercept	12.8235	0.3781	<0.0001	11.4217	to 14.2253
Alginate	-0.8061	0.1238	0.0006	-1.2649	to -0.3472
Acrycoat E30D	-0.1898	0.0945	0.0915	-0.5402	to 0.1607
Source of variation	SSq	DF	MSq	F	p
Due to regression	9.961	2	4.980	23.22	0.0015
About regression	1.287	6	0.214		
Total	11.247	8			

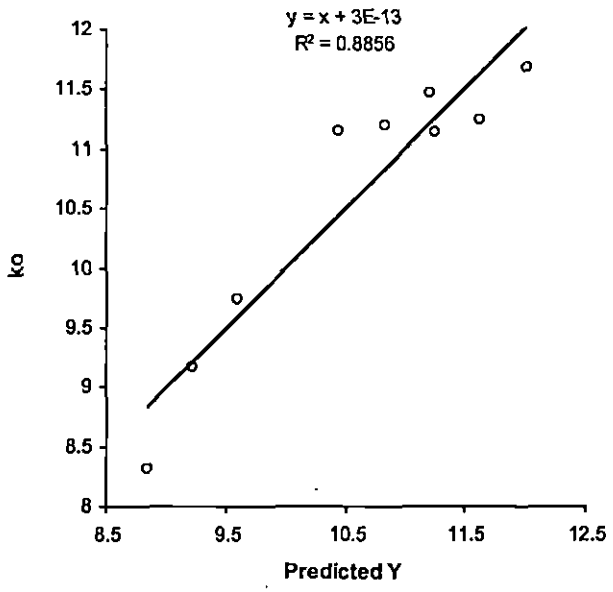

Figure 10.1 Graphical representation of Regression analysis showing the correlation between the actual release rates (Ko) of Frusemide with predicted values.

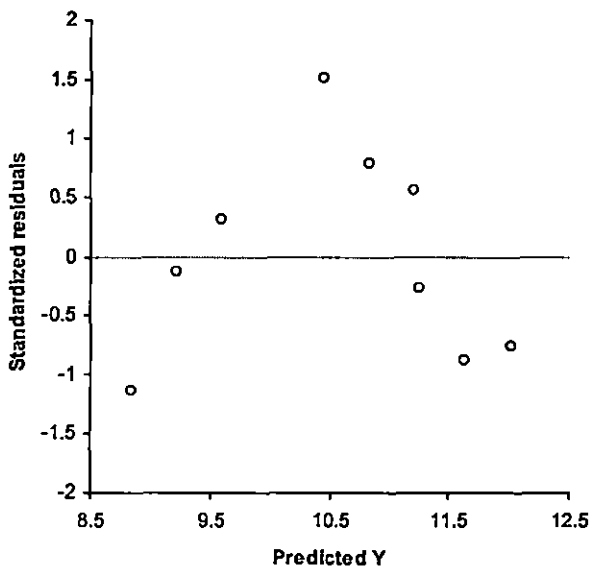

Figure 10.2 Graphical representation of Regression analysis showing the correlation between the standardized residuals with predicted values of Ko.

10.7.3 Interpretation of the Statistical Data

From the results of ANOVA (Table 10.5), effect of Sodium alginate and Acrycoat E30D on the zero order release rate constant (Ko) was estimated. Considering the p-value, it is observed that at 95% confidence level (5% error to be tolerated), former polymer had significant effect on Ko with a p-value = 0.0014 and F = 52.03. The latter polymer, Acrycoat E30D do not have significant effect on the release rate as evident for its p-value = 0.0939 and F= 4.53.

The Linear Regression Analysis (Table 10.6), further justifies the ANOVA results that individual effect of sodium alginate is significant even at 99 % confidence interval as evident from its p-values = 0.0006. Copolymer Acrycoat E30D remained insignificant. Coefficients obtained for both the polymers were negative. For both the cases, interaction term signifying the combined effect of the polymers has not been considered. To consider the interaction effect, Response Surface Methodology (RSM) was employed.
The effect of Ko was related to the concentration (% w/w) of the polymers by the following equation:

$$\mathbf{Ko = 12.8235 - 0.8061 \times Sodium\ Alginate - 0.1898 \times Acrycoat\ E30D} \qquad (10.21)$$

From Fig.10.1 and 10.2, it is evident that there is a clear correlation between the predicted value and actual value of Ko.

10.8 Response Surface Methodology (RSM) and Experimental Design

10.8.1 Experimental Design

The experimental design was done using 'Design Expert version 7.1.2'. A full 3^2 factorial analysis using 2FI design model and 9 runs were performed. The summary of the design is presented in the Table 10.7.

Table 10.7 Summary of Experimental Design

Study Type	**Factorial**	**Runs**	**9**
Initial Design	**Full Factorial**	**Blocks**	**No Blocks**
Center Points	**0**		
Design Model	**2FI**		

Factor	Name	Units	Type	Low Actual	High Actual	Levels			
A	Alginate	% w/w	Categoric	1 %	4 %	3			
B	Acrycoat E30D	% w/w	Categoric	0 %	4 %	3			
Response	**Name**	**Units**	**Observn.**	**Analysis**	**Min.**	**Max.**	**Mean**	**Std.Devn**	**Ratio**
Y1	Ko	mg/hr	9	Factorial	8.89	11.47	10.64	0.82	1.29
Y2	X_{120}	mg	9	Factorial	8.00	37.34	22.51	10.7	4.67
Y3	t_{80}	hr	9	Factorial	4.50	9.30	7.12	1.47	2.07
Y4	n		9	Factorial	0.60	1.79	1.01	0.37	3.00

The Design Matrix Evaluation for Factorial 2FI Model

Degrees of Freedom (dF) for Evaluation

Model	8
Residuals	0
Lack of Fit	0
Pure Error	0
Total	9

A minimum of 3 dF is recommended for a valid lack of fit test and 4 dF for pure error. Since the model has 8 dF, it can be said that it gives a valid lack of fit test.

10.8.2 Use of Response Surface Methodology (RSM) and Experimental Design to study the Effect of Polymers on the Zero order Release rate (Ko)

Table 10.8 Analysis of Variance (ANOVA) table for Response Surface Quadratic Model

[Partial sum of Squares – Type - III]

Source	Sum of Squares	Degree of Freedom dF	Mean Square	F value	p-value prob> F	Effect
Model	5.91	5	1.18	40.21	0.0060	**significant**
A-Alginate %w/w	3.13	1	3.13	106.59	0.0019	
B-Acrycoat E30D %w/w	0.83	1	0.84	28.70	0.0127	
AB	0.23	1	0.23	7.77	0.0685	
A^2	0.97	1	0.97	32.96	0.0105	
B^2	1.875E-004	1	1.875E-004	6.379E-003	0.9414	
Residual	0.088	3	0.029			
Correction total	6	8				

Std. Dev.	0.17	R-Squared	0.9853
Mean	10.64	Adj R-Squared	0.9608
C.V. %	1.61	Pred R-Squared	0.8231
PRESS	1.06	Adeq Precision	18.468

Results:

The Model F-value of 40.21 implies that the model is **significant**. There is only a 0.60% chance that a large "Model F-Value" could occur due to noise. Values of "Probability > F" less than 0.0500 indicate model terms are significant. In this case A, B and A^2 are significant model terms. Values greater than 0.1000 indicate the model terms are not significant. If there are many insignificant model terms (not counting those required to support hierarchy), model reduction may improve the model.

The "Pred R-Squared" of 0.8231 is in reasonable agreement with the "Adj R-Squared" of 0.9608."Adeq Precision" measures the signal to noise ratio. A ratio greater than 4 is desirable. the ratio of 18.468 indicates an adequate signal. This model can be used to navigate the design space.

Table 10.9 Coefficient estimation of Ko

Factor	Coefficient Estimate	dF	Standard error	95 % CI Low	95 % CI High
Intercept	11.11	1	0.140	10.66	11.56
A- Alginate %w/w	-0.72	1	0.070	-0.95	-0.50
B-Acrycoat E30D %w/w	-0.28	1	0.071	-0.6	-0.15
AB	-0.23	1	0.840	-0.05	0.033
A^2	-0.80	1	0.14	-1.24	-0.36
B^2	9.683E-003	1	0.12	-0.38	-0.40

Final Equation in Terms of Coded Factors:

$$k0 = 11.11 - 0.72 \times A - 0.28 \times B - 0.23 \times A \times B - 0.80 \times A^2 + 9.683E\text{-}003 \times B^2 \quad (10.22)$$

Final Equation in Terms of Actual Factors:

$$k0 = 10.09787 + 1.44670 \times \text{Alginate} - 3.24583E\text{-}003 \times \text{Acrycoat E30D} - 0.078241 \times \text{Alginate} \times \text{Acrycoat E30D} - 0.35441 \times \text{Alginate}^2 + 2.42083E\text{-}003 \times \text{Acrycoat E30D}^2 \quad (10.23)$$

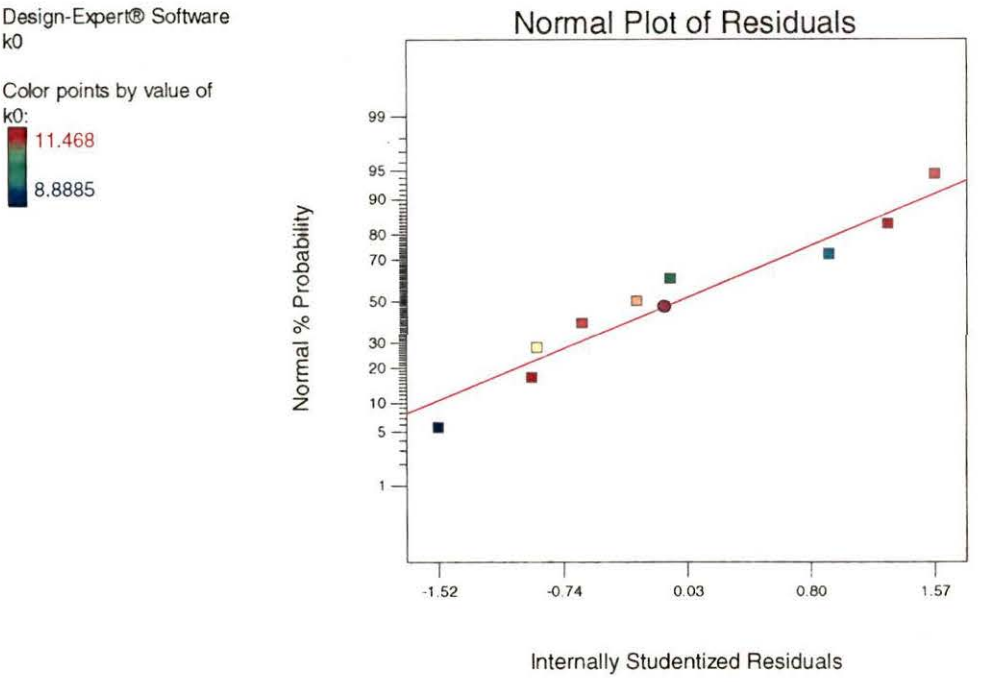

Figure 10.1 Internally Studentized Residuals vs Normal % Probability

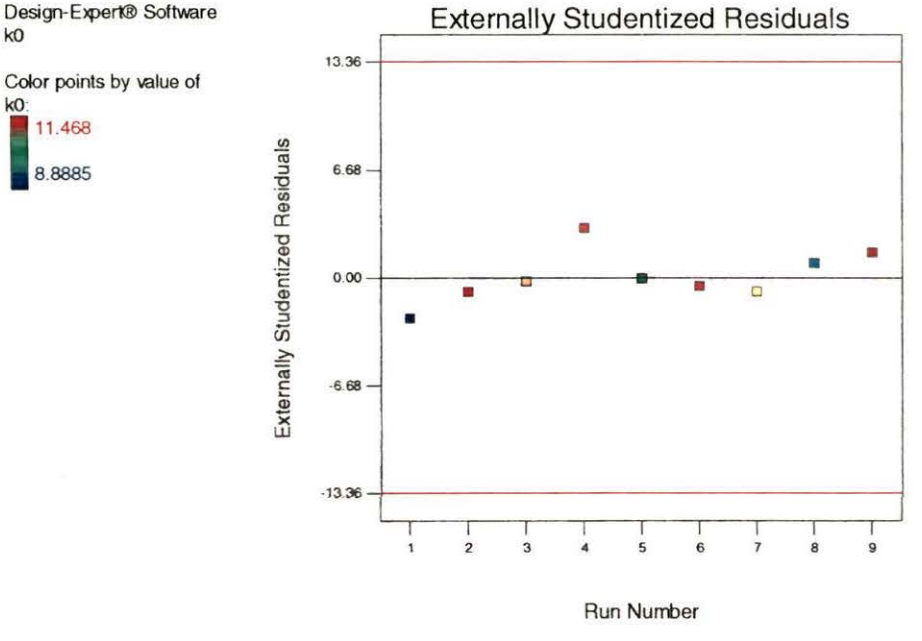

Figure 10.2 Externally Studentized Residuals vs Run Number

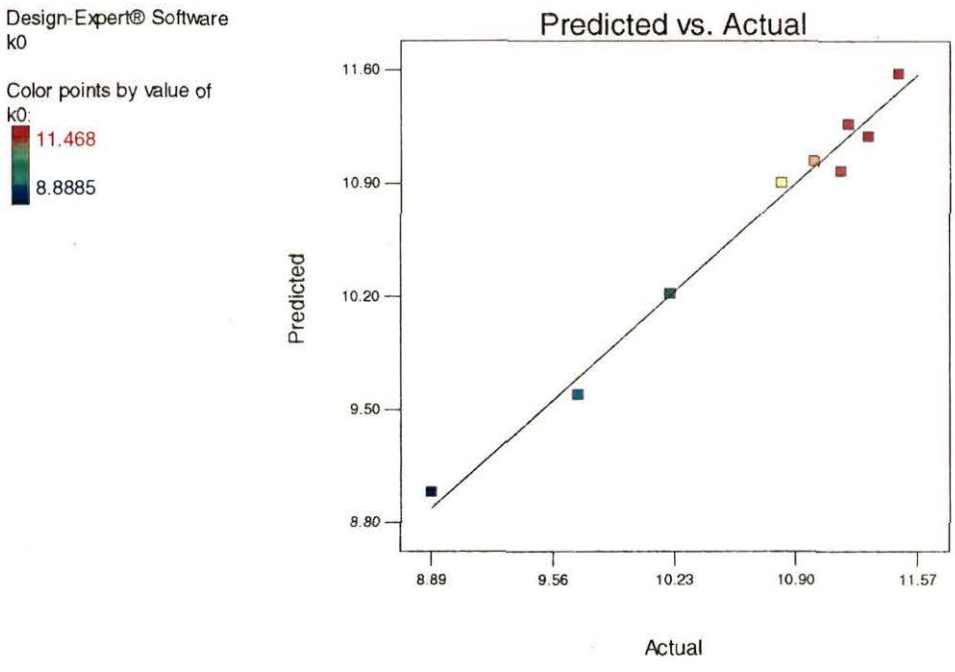

Figure 10.3 Predicted vs Actual

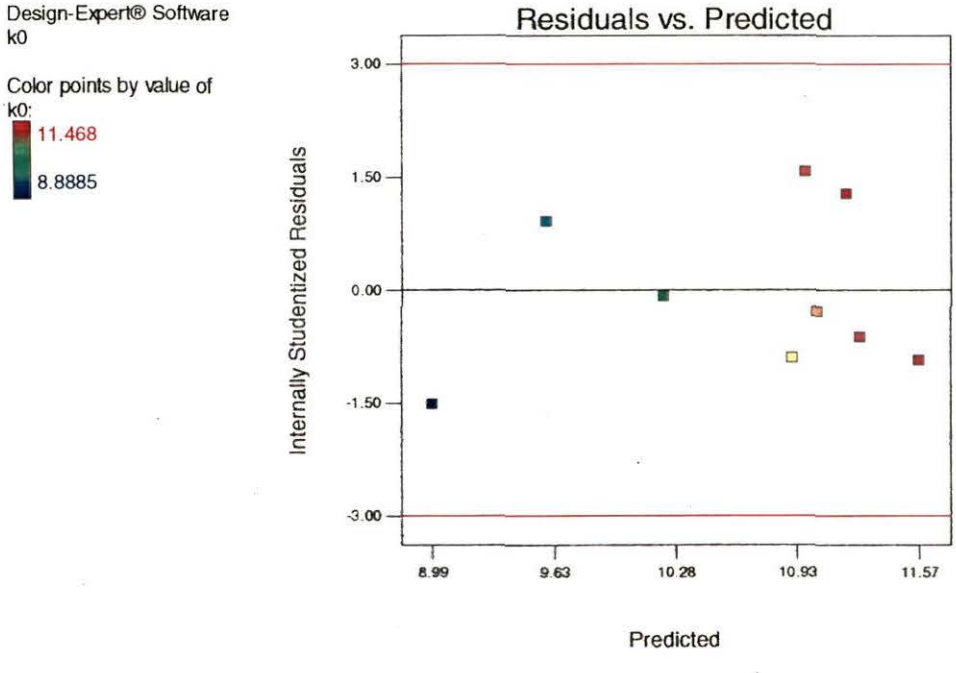

Figure 10.4 Internally Studentized Residuals vs Predicted

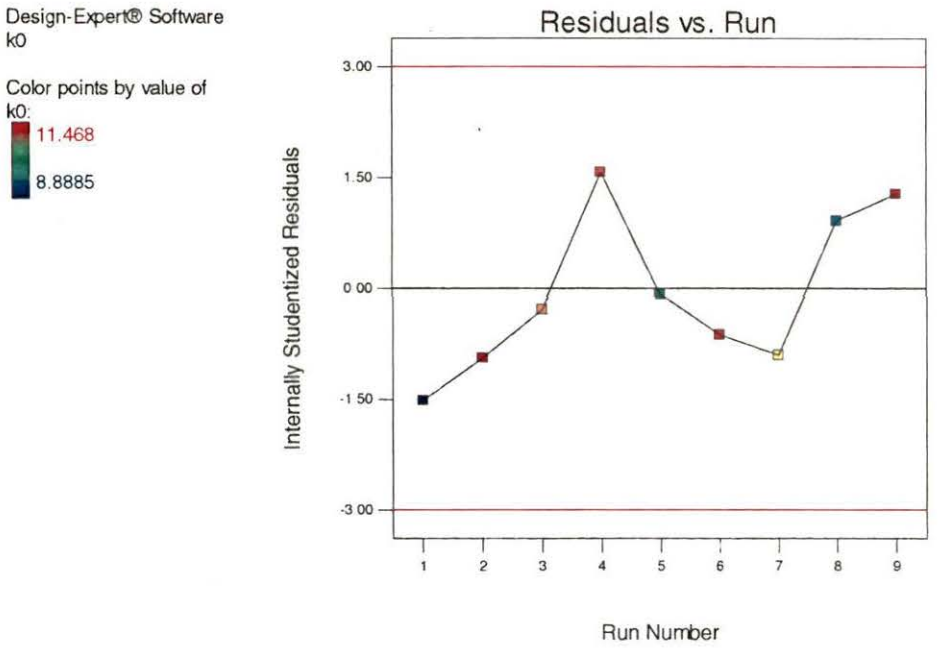

Figure 10.5 Internally Studentized Residuals vs Run Number

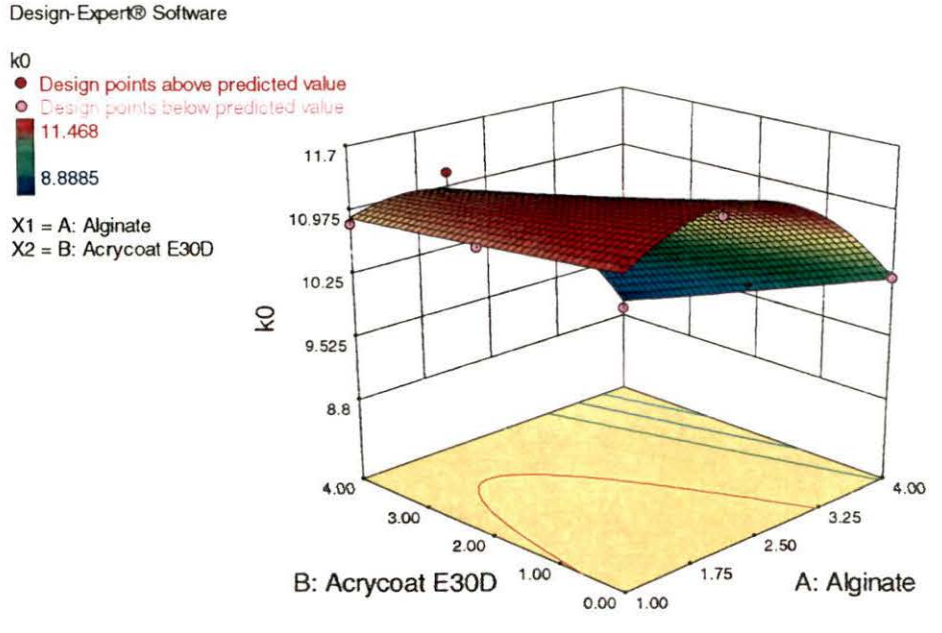

Figure 10.6 Effect of interaction between Alginate and Acrycoat E30D on Zero order rate constant (Ko) represented in 3D Response curve

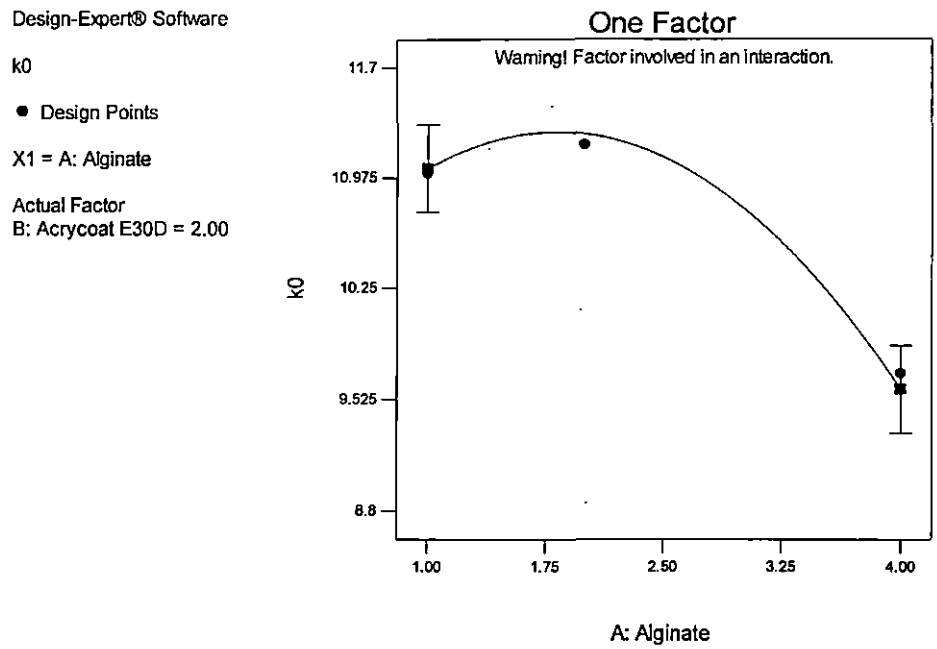

Figure 10.7 Sole effect of Alginate Concentration on Ko

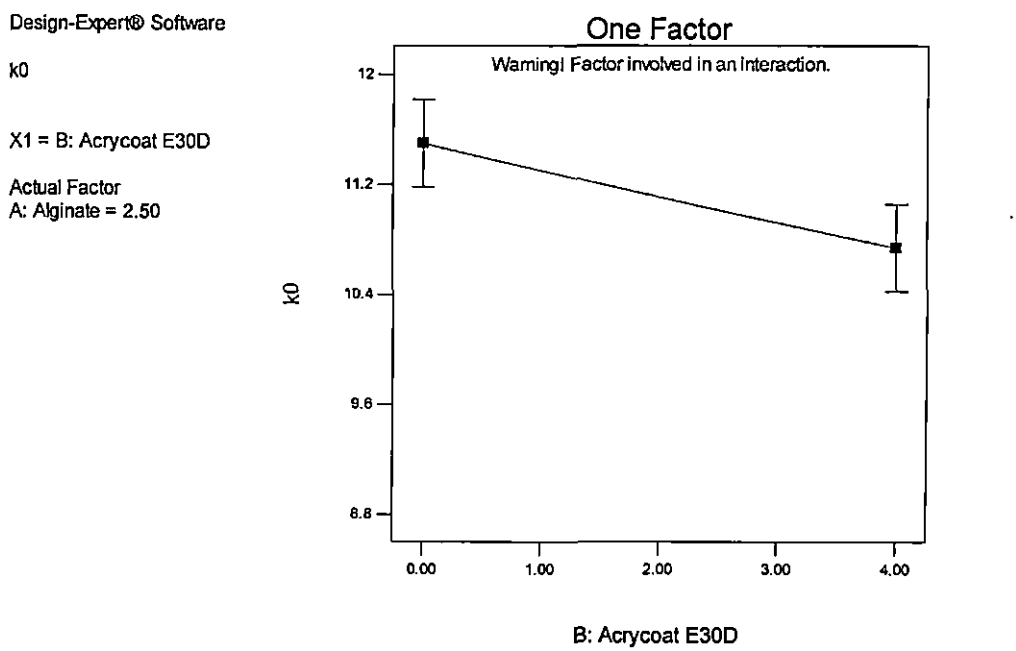

Figure 10.8 Sole effect of Acrycoat E30D Concentration on Ko

10.8.3 Use of Response Surface Methodology (RSM) and Experimental Design to study the Effect of Polymers on the Burst effect of the micropellets (X 120)

Table10.10 Analysis of Variance (ANOVA) table for Response Surface Quadratic Model

[Partial sum of Squares – Type - III]

Source	Sum of Squares	Degree of Freedom dF	Mean Square	F value	p-value prob> F	Effect
Model	1012.2	5	202.44	33.26	0.0079	**significant**
A-Alginate %w/w	558.39	1	558.39	91.74	0.0024	
B-Acrycoat E30D %w/w	70.4	1	70.4	11.57	0.0424	
AB	43.53	1	43.53	7.15	0.0754	
A^2	517.97	1	517.97	85.10	0.0027	
B^2	5.74	1	5.74	0.94	0.4032	
Residual	18.26	3	6.09			
Cor total	1030.46	8				

Std. Dev.	2.47	R-Squared	0.9823
Mean	22.51	Adj R-Squared	0.9527
C.V. %	10.96	Pred R-Squared	0.8068
PRESS	199.13	Adeq Precision	13.743

Result:

The Model F-value of 33.26 implies that the model is significant. There is only a 0.79% chance that a large "Model F-Value" could occur due to noise. Values of "Probability > F" less than 0.0500 indicate model terms are significant. In this case A, B and A^2 are

significant model terms. Values greater than 0.1000 indicate the model terms are not significant. If there are many insignificant model terms (not counting those required to support hierarchy), model reduction may improve the model.

The "Pred R-Squared" of 0.8068 is in reasonable agreement with the "Adj R-Squared" of 0.9527. "Adeq Precision" measures the signal to noise ratio. A ratio greater than 4 is desirable. The ratio of 13.743 indicates an adequate signal. This model can be used to navigate the design space.

Table 10.11 Coefficient estimation of X_{120}

Factor	Coefficient Estimate	dF	Standard error	95 % CI Low	95 % CI High
Intercept	9.59	1	2.02	3.16	16.02
A- Alginate %w/w	-9.65	1	1.01	-12.85	-6.44
B-Acrycoat E30D %w/w	-3.46	1	1.02	-6.69	-0.22
AB	-3.24	1	1.21	-7.09	0.62
A^2	18.44	1	2.00	12.08	24.80
B^2	-1.69	1	1.74	-7.25	3.86

Final Equation in Terms of Coded Factors:

$$X_{120} = 9.59 - 9.65 \times A - 3.46 \times B - 3.24 \times A \times B + 18.44 \times A^2 - 1.69 \times B^2 \quad (10.24)$$

Final Equation in Terms of Actual Factors:

$$X_{120} = 73.24356 - 45.24245 \times \text{Alginate} + 2.66558 \times \text{Acrycoat E30D} - 1.07986 \times \text{Alginate} \times \text{Acrycoat E30D} + 8.19417 \times \text{Alginate}^2 - 0.4234 \times \text{Acrycoat E30D}^2 \quad (10.25)$$

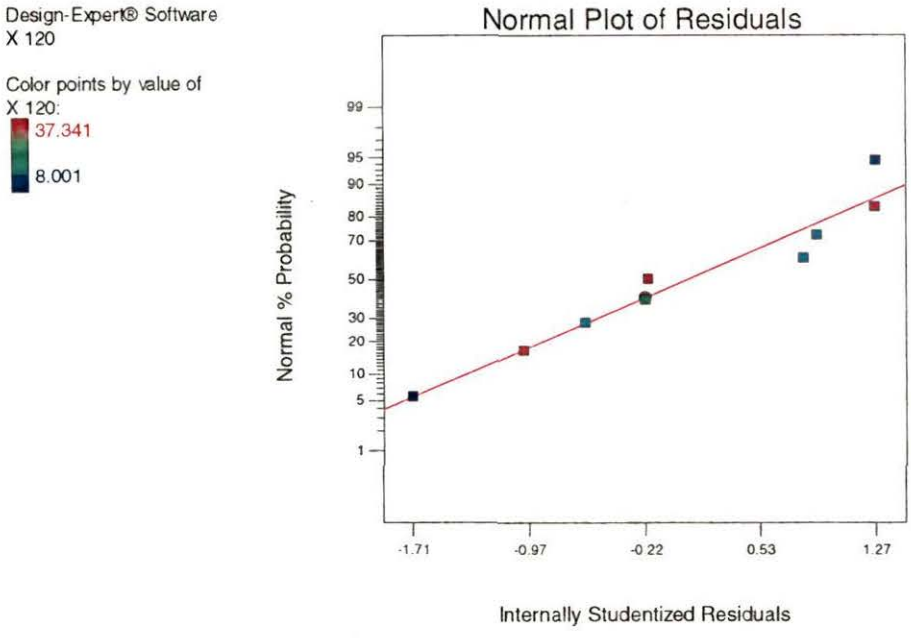

Figure 10.9 Internally studentized residuals vs normal % probability

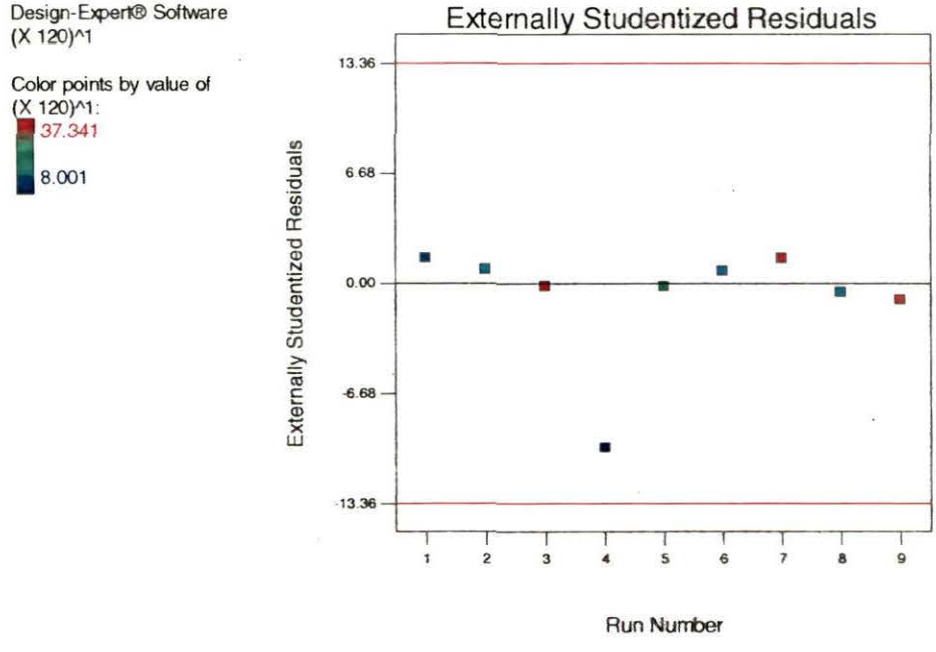

Figure 10.10 Externally Studentized Residuals vs Run Number

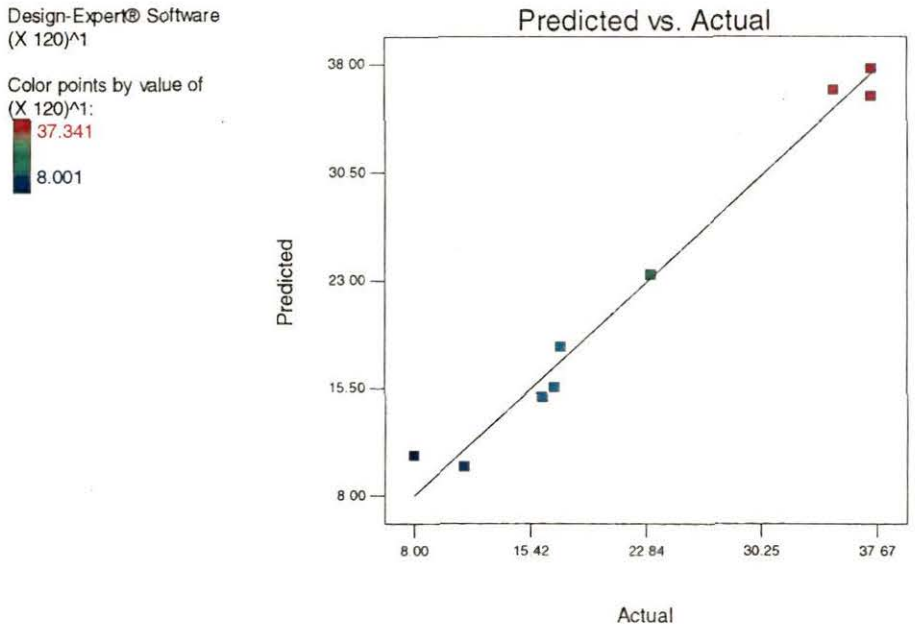

Figure 10.11 Predicted vs Actual

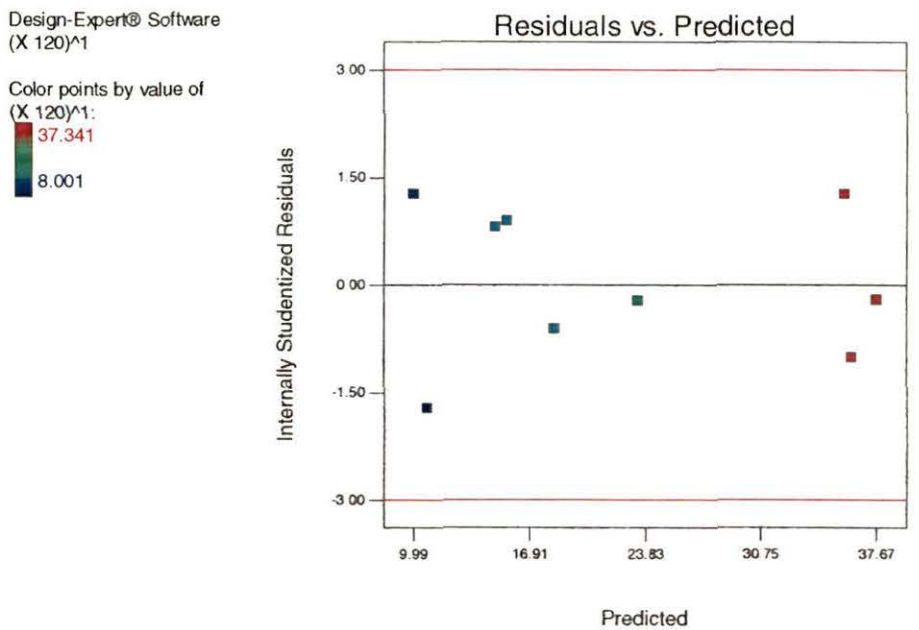

Figure 10.12 Internally Studentized Residuals vs Predicted

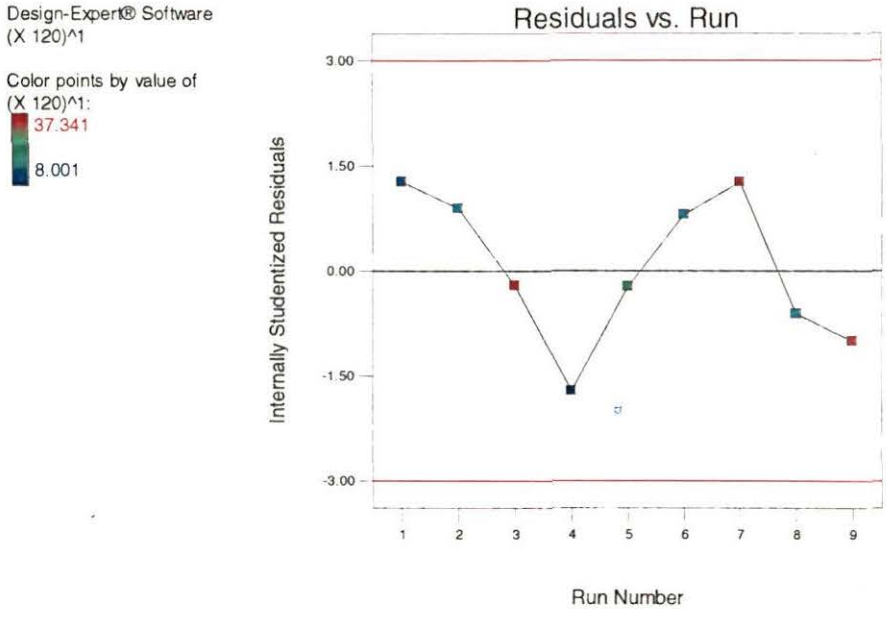

Figure 10.13 Internally Studentized Residuals vs Run Number

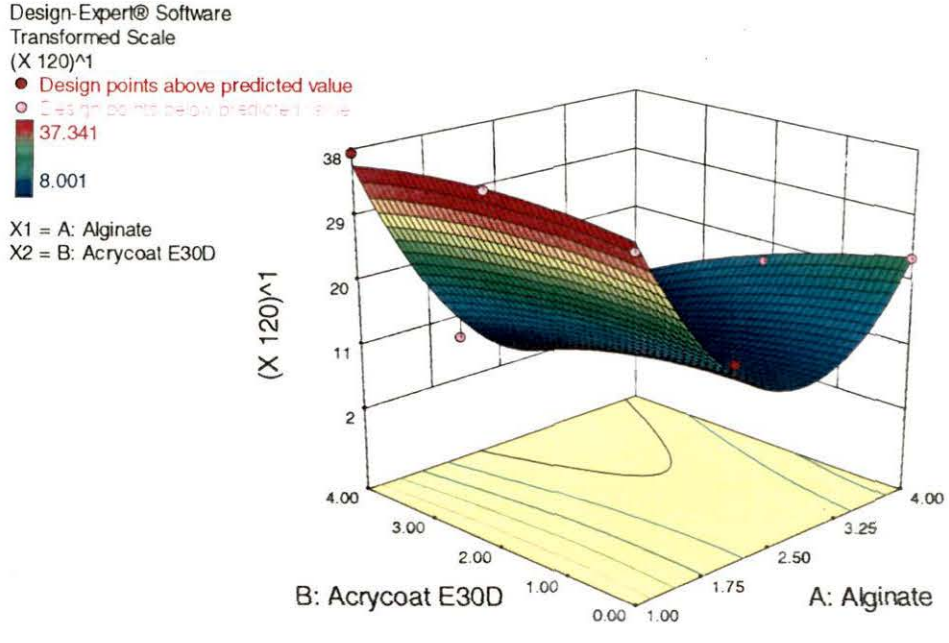

Figure 10.14 Effect of interaction between Alginate and Acrycoat E30D on the amount of drug released in 120 minutes (2 hr) X_{120}, represented in 3D response curve

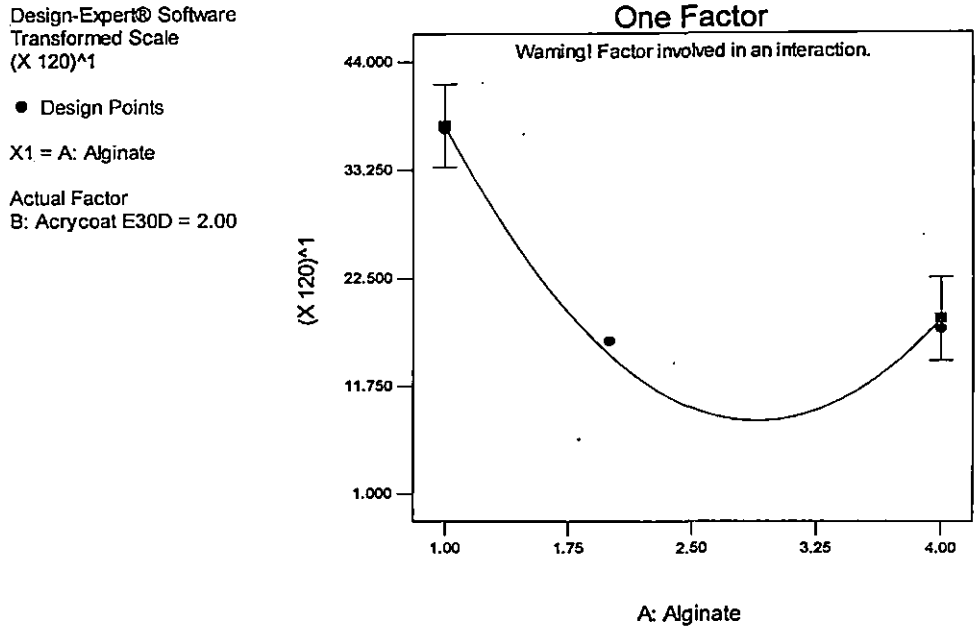

Figure 10.15 Sole effect of Alginate Concentration on X_{120}

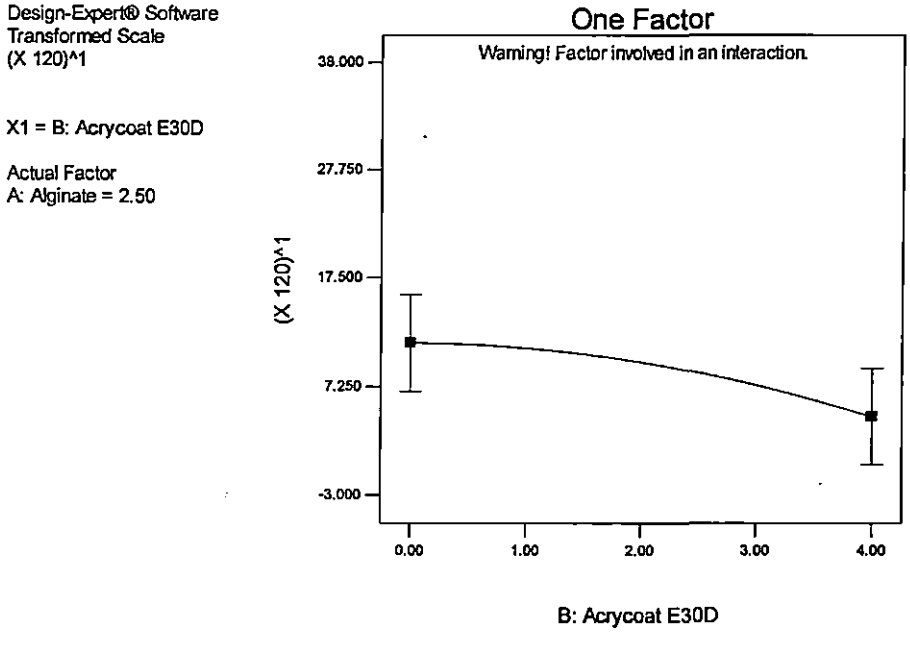

Figure 10.16 Sole effect of Acrycoat E30D Concentration on X_{120}

10.8.4 Use of Response Surface Methodology (RSM) and Experimental Design to study the Effect of Polymers on the time required for 80 % release (t_{80})

Table 10.12 Analysis of Variance (ANOVA) table for Response Surface Quadratic Model

[Partial sum of Squares – Type - III]

Source	Sum of Squares	Degree of Freedom dF	Mean Square	F value	p-value prob> F	Effect
Model	19.15	5	3.83	38.12	0.0065	**significant**
A-Alginate %w/w	16.34	1	16.34	162.55	0.0010	
B-Acrycoat E30D %w/w	1.99	1	1.99	19.76	0.0212	
AB	2.176E-003	1	2.176E-003	0.03	0.8743	
A^2	2.63	1	2.63	26.12	0.0145	
B^2	0.014	1	0.014	0.14	0.7348	
Residual	0.30	3	0.10			
Cor total	19.46	8				

Std. Dev.	0.32	R-Squared	0.9845
Mean	7.12	Adj R-Squared	0.9587
C.V. %	4.45	Pred R-Squared	0.8264
PRESS	3.38	Adeq Precision	17.234

Result:

The Model F-value of 38.12 implies that the model is significant. There is only a 0.65% chance that this large "Model F-Value" could occur due to noise. Values of "Probability > F" less than 0.0500 indicate model terms are significant. In this case A, B and A^2 are

significant model terms. Values greater than 0.1000 indicate the model terms are not significant. If there are many insignificant model terms (not counting those required to support hierarchy), insignificant model terms (not counting those required to support hierarchy), model reduction may improve the model.

The "Pred R-Squared" of 0.8264 is in reasonable agreement with the "Adj R-Squared" of 0.9587."Adeq Precision" measures the signal to noise ratio. A ratio greater than 4 is desirable. The ratio of 17.234 indicates an adequate signal. This model can be used to navigate the design space.

Table10.13 Coefficient estimation of t_{80}

Factor	Coefficient Estimate	dF	Standard error	95 % CI Low	95 % CI High
Intercept	8.17	1	0.26	7.35	9.00
A- Alginate %w/w	1.65	1	0.13	1.24	2.06
B-Acrycoat E30D %w/w	0.58	1	0.13	0.16	1.00
AB	-0.027	1	0.16	-0.52	0.47
A^2	-.1.31	1	0.26	-2.13	-0.50
B^2	0.083	1	0.22	-0.63	0.08

Final Equation in Terms of Coded Factors:

$$t_{80} = 8.17 + 1.65 \times A + 0.58 \times B - 0.027 \times A \times B - 1.31 \times A^2 + 0.083 \times B^2 \quad (10.26)$$

Final Equation in Terms of Actual Factors:

$$t_{80} = 1.23611 + 4.03452 \times \text{Alginate} + 0.22917 \times \text{Acrycoat E30D} - 8.92857E\text{-}003 \times \text{Alginate} \times \text{Acrycoat E30D} - 0.58333 \times \text{Alginate}^2 + 0.020833 \times \text{Acrycoat E30D}^2 \quad (10.27)$$

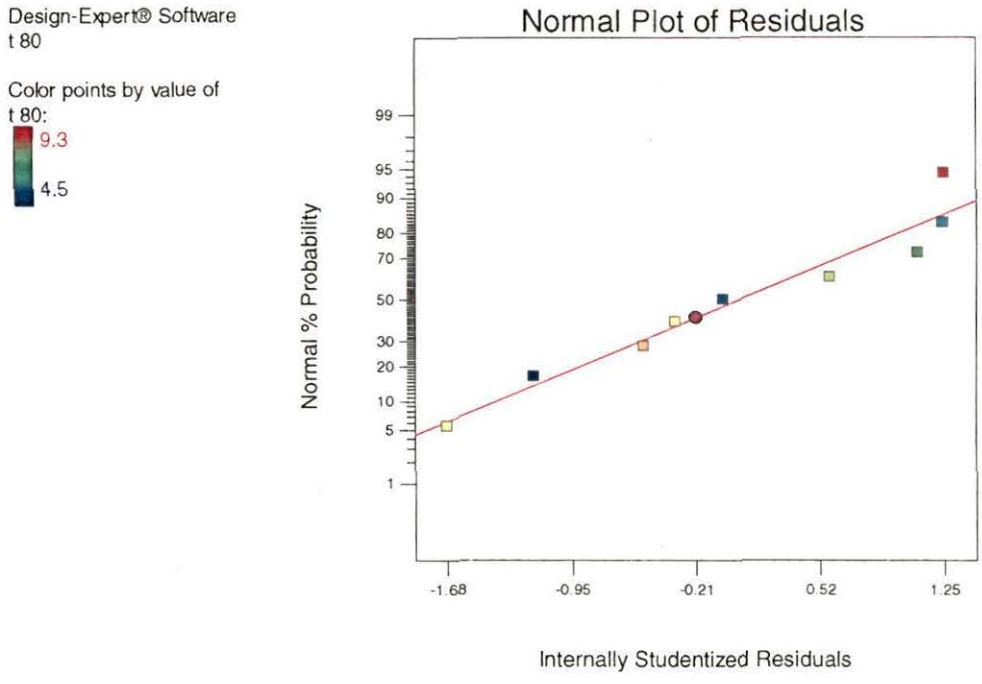

Figure 10.17 Internally Studentized Residuals vs Normal % Probability

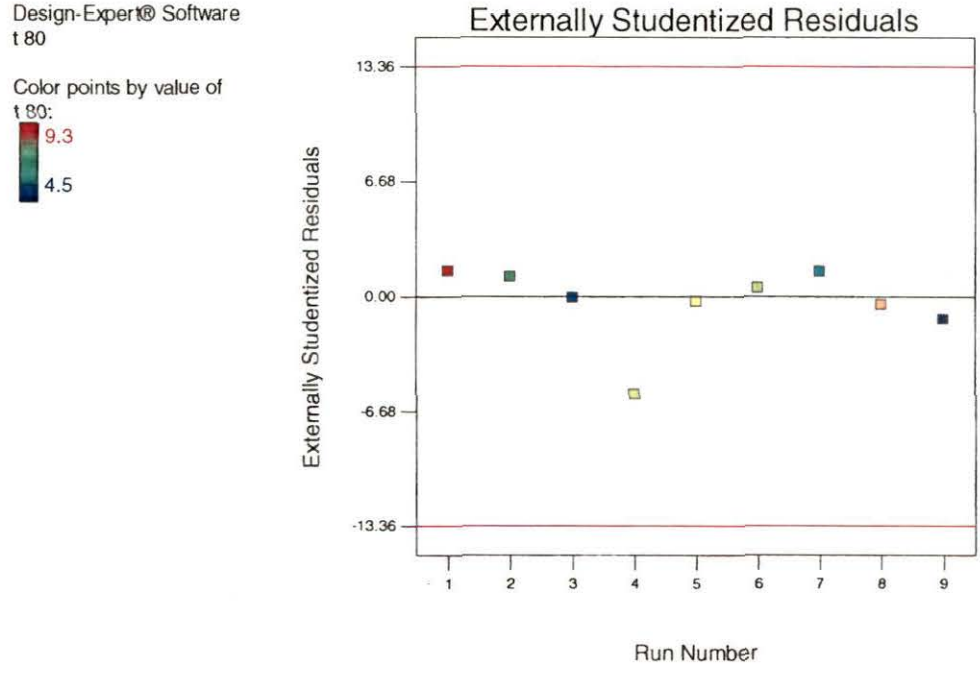

Figure 10.18 Externally Studentized Residuals vs Run Number

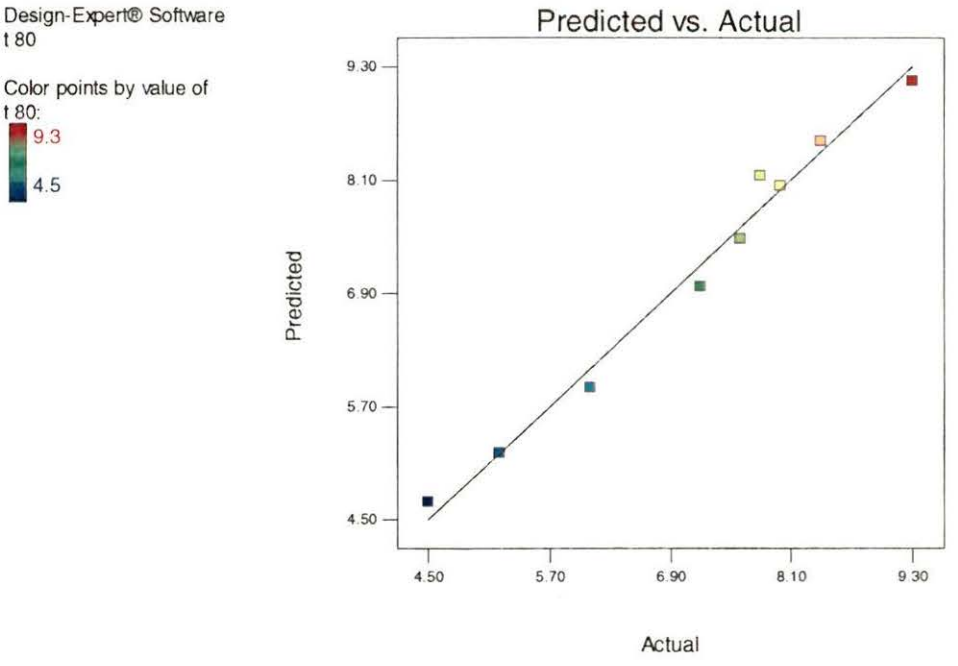

Figure 10.19 Predicted vs Actual

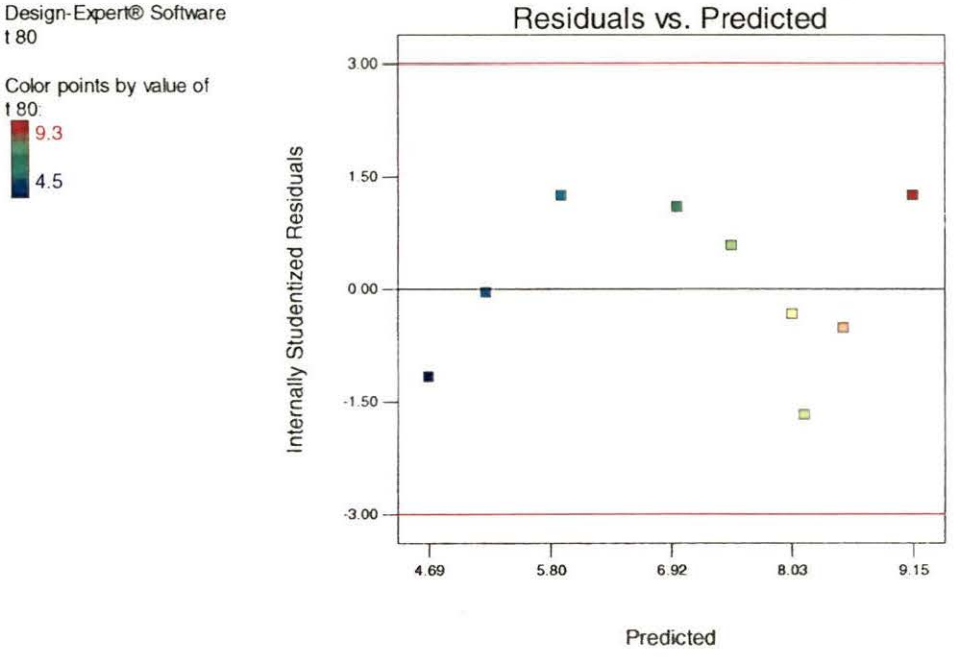

Figure 10.20 Internally Studentized Residuals vs Predicted

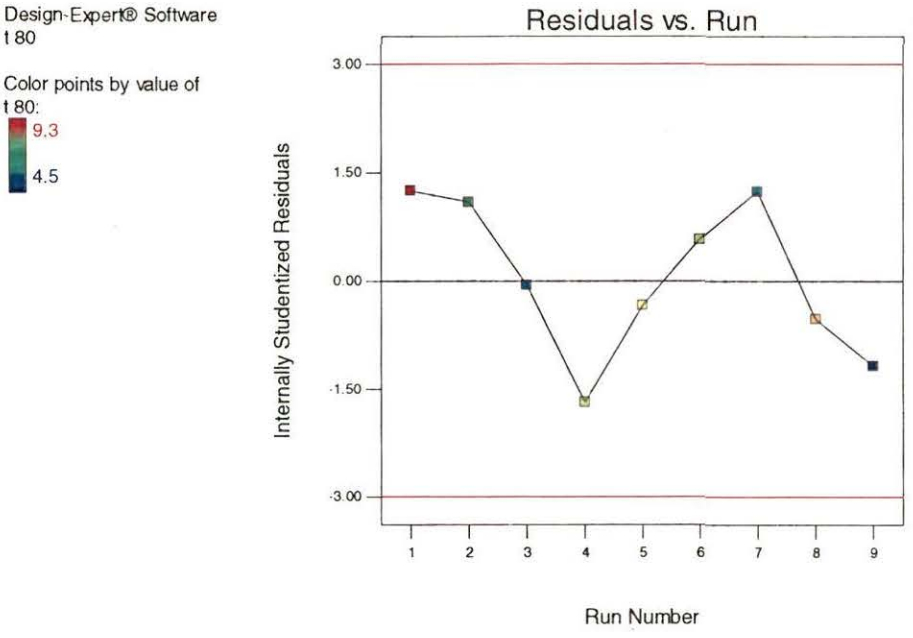

Figure 10.21 Internally Studentized Residuals vs Run Number

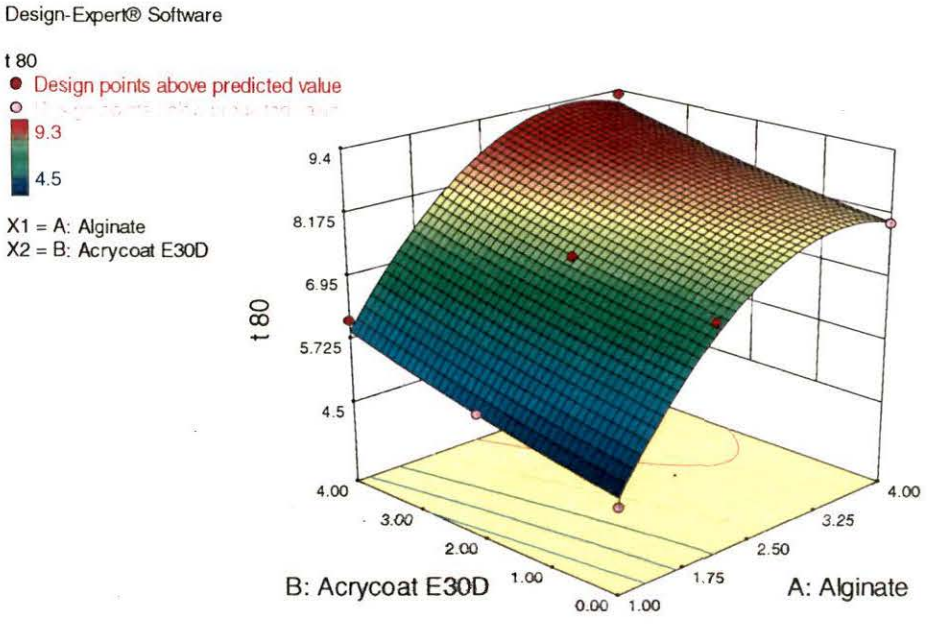

Figure 10.22 Effect of interaction between Alginate and Acrycoat E30D on time of 80% drug release (t_{80}) represented in 3D Response curve

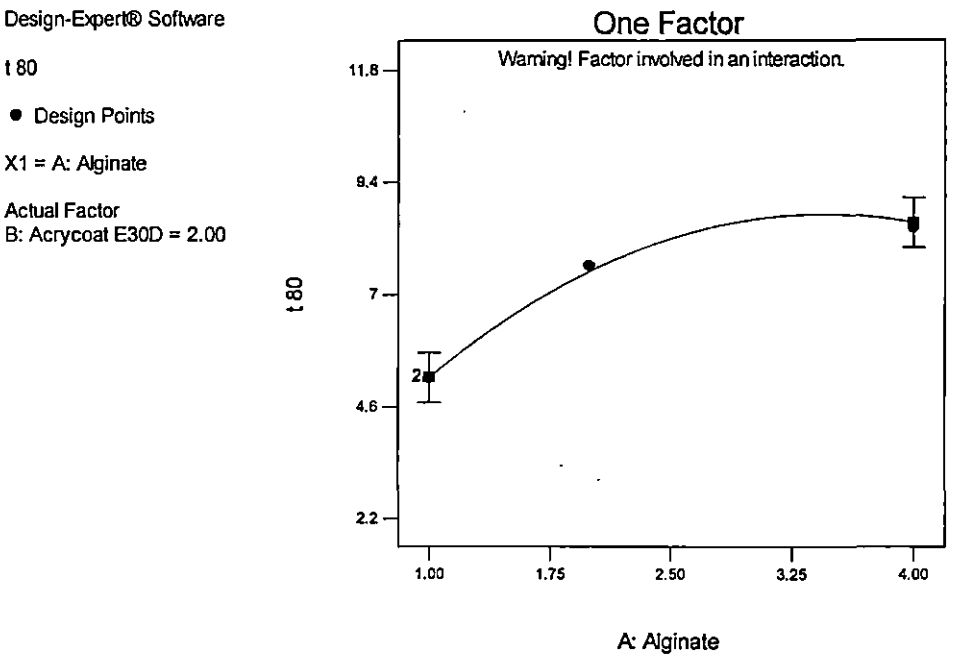

Figure 10.23 Sole effect of Alginate Concentration on t_{80}

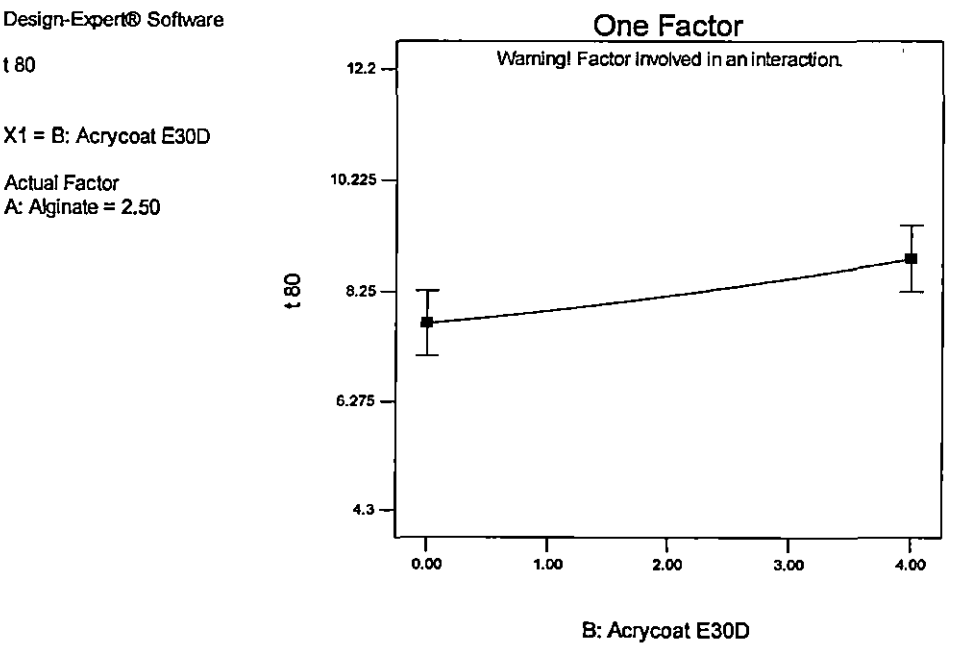

Figure 10.24 Sole effect of Acrycoat E30D Concentration on t_{80}

10.8.5 Use of Response Surface Methodology (RSM) and Experimental Design to study the Effect of Polymers on the diffusion coefficient of Peppas equation (n)

Table10.14 Analysis of Variance (ANOVA) table for Response Surface Quadratic Model

[Partial sum of Squares – Type - III]

Source	Sum of Squares	Degree of Freedom dF	Mean Square	F value	p-value prob> F	Effect
Model	1.11	5	0.22	5.56	0.0943	**Not significant**
A-Alginate %(w/w)	0.33	1	0.33	8.38	0.0627	
B-Acrycoat E30D %w/w	0.17	1	0.17	4.15	0.1345	
AB	0.013	1	0.013	0.33	0.6043	
A^2	0.73	1	0.73	18.43	0.0232	
B^2	0.025	1	0.025	0.63	0.4864	
Residual	0.12	3	0.04			
Cor total	1.23	8				

Std. Dev.	0.20	R-Squared	0.9026
Mean	1.1	Adj R-Squared	0.7403
C.V. %	19.67	Pred R-Squared	0.0269
PRESS	1.19	Adeq Precision	6.349

Result:

The Model F-value of 5.56 implies that there is only 9.43% chance that this large "Model F-Value" could occur due to noise. Values of "Prob > F" less than 0.0500 indicate model terms are significant. In this case A^2 are significant model terms. Values greater than 0.1000 indicate the model terms are not significant. If there are many insignificant model terms (not counting those required to support hierarchy), model reduction may improve the model.

The "Pred R-Squared" of 0.0269 is not as close to the "Adj R-Squared" of 0.7403 as one might normally expect. This may indicate a large block effect or a possible problem with the model and/or data. Things to consider are model reduction, response tranformation, outliers,etc. "Adeq Precision" measures the signal to noise ratio.A ratio greater than 4 is desirable.The ratio of 6.349 indicates an adequate signal. This model can be used to navigate the design space.

Table 10.15 Coefficient estimation

Factor	Coefficient Estimate	dF	Standard error	95 % CI Low	95 % CI High
Intercept	1.45	1	0.16	0.93	1.97
A- Alginate %w/w	0.24	1	0.081	-0.023	0.50
B-Acrycoat E30D %w/w	0.17	1	0.082	-0.094	0.43
AB	0.057	1	0.098	-0.26	0.37
A^2	-0.69	1	0.16	-1.21	-0.18
B^2	0.11	1	0.14	-0.34	0.56

Final Equation in Terms of Coded Factors:

$$n = 1.45 + 0.24 \times A + 0.17 \times B + 0.057 \times A \times B - 0.69 \times A^2 + 0.11 \times B^2 \quad (10.28)$$

Final Equation in Terms of Actual Factors:

$$n = -0.82788 + 1.66177 \times \text{Alginate} - 0.075108 \times \text{Acrycoat E30D} + 0.018850 \times \text{Alginate} \times \text{Acrycoat E30D} - 0.30844 \times \text{Alginate}^2 + 0.027921 \times \text{Acrycoat E30D}^2 \quad (10.29)$$

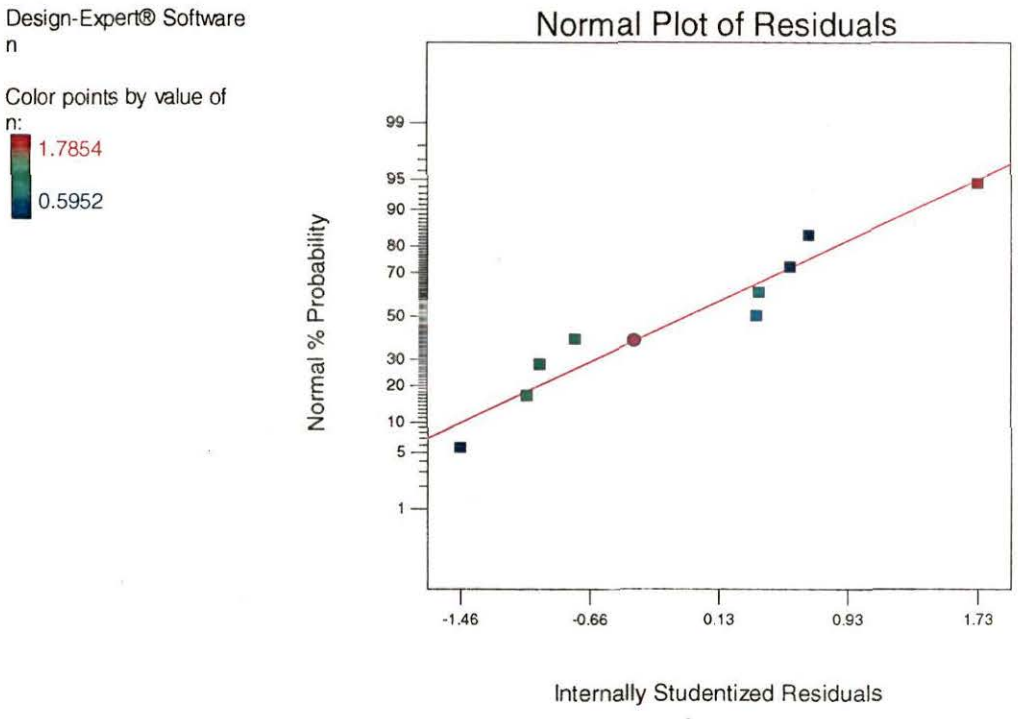

Figure 10.25 Internally Studentized Residuals vs Normal % Probability

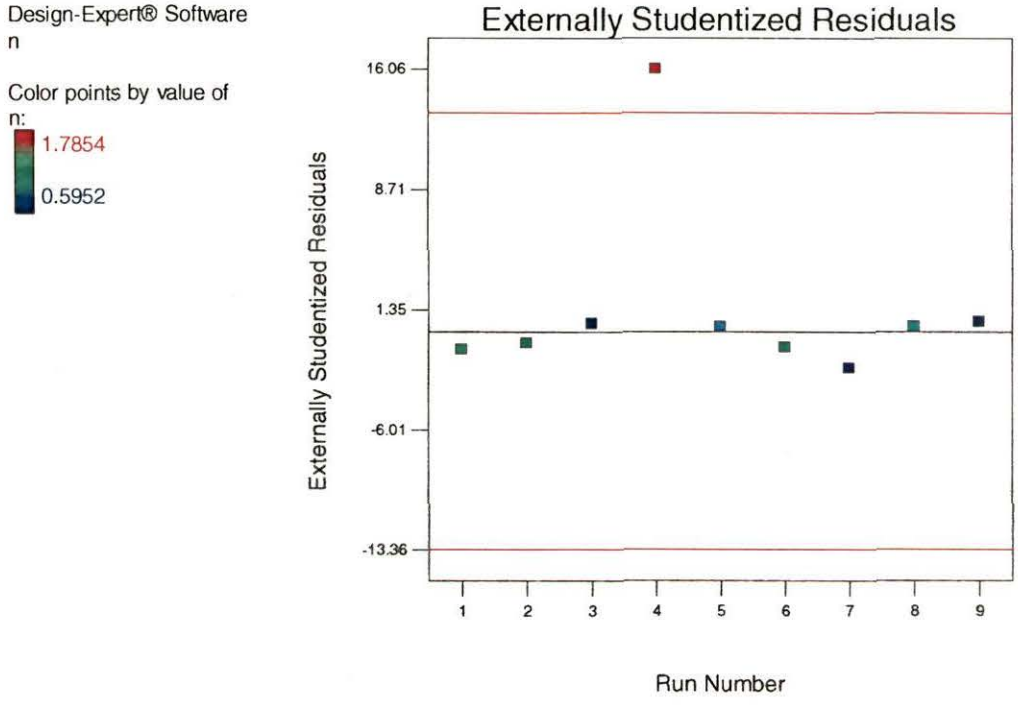

Figure 10.26 Externally Studentized Residuals vs Normal % Probability

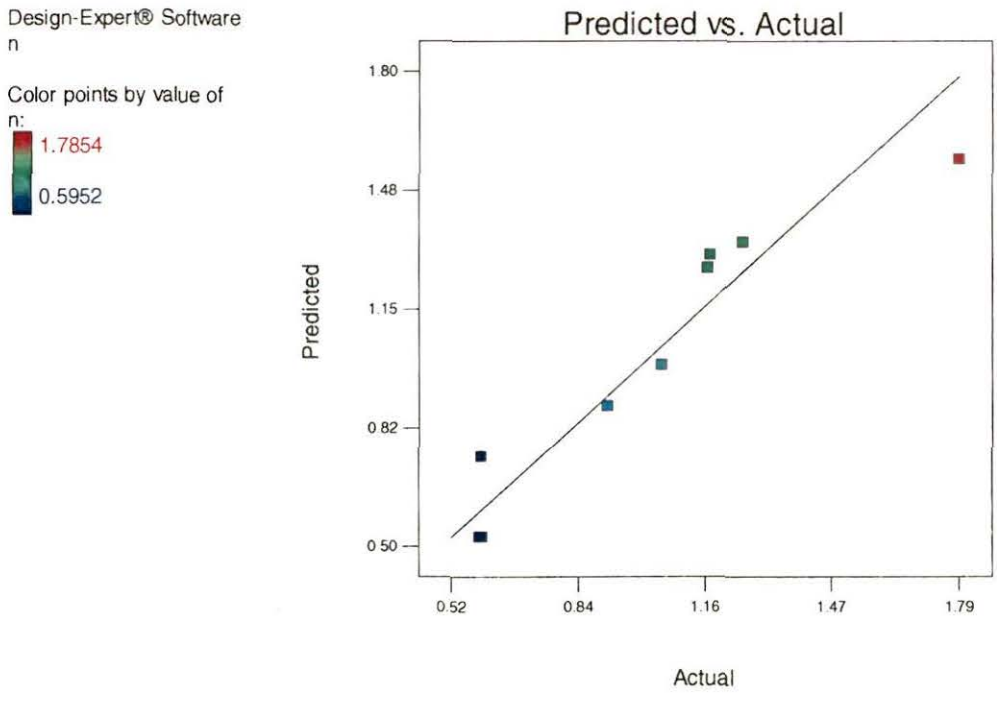

Figure 10.27 Predicted vs Actuals

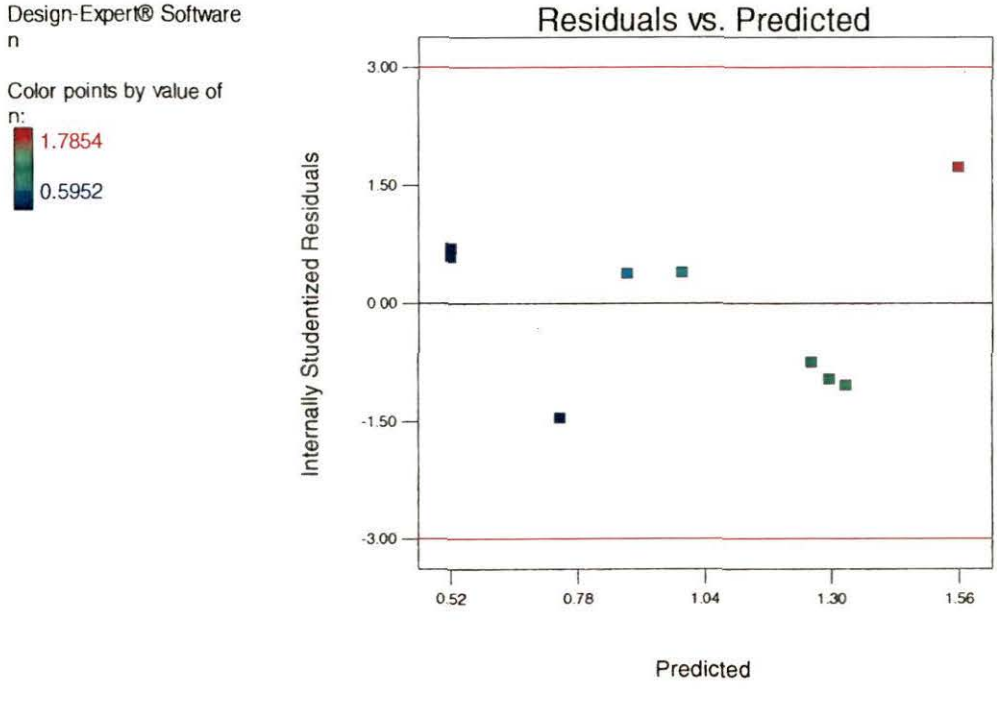

Figure 10.28 Internally Studentized Residuals vs Predicted

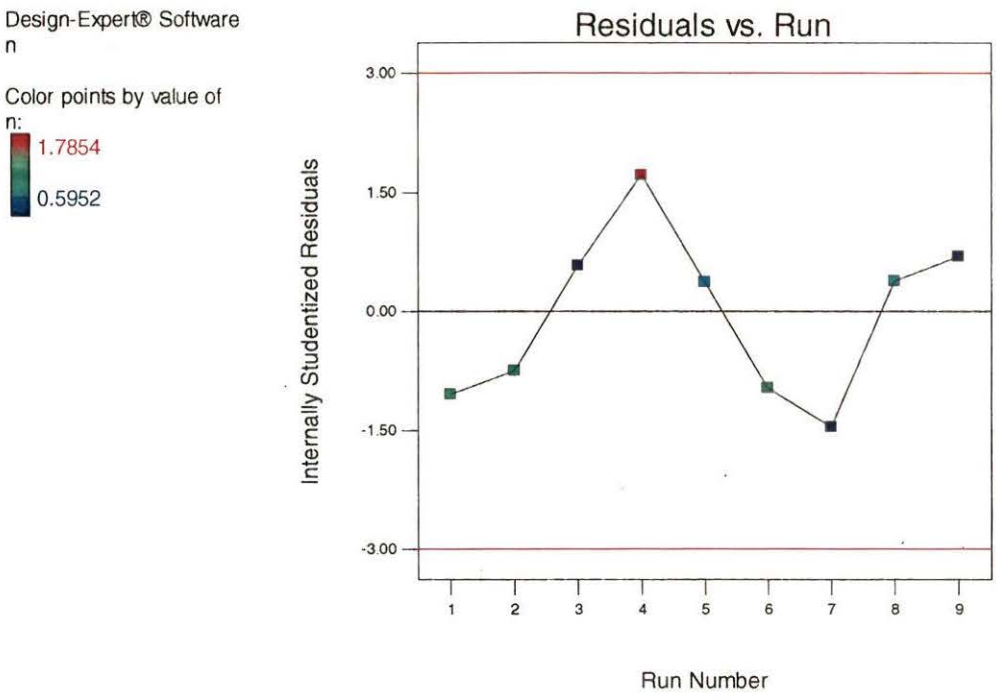

Figure 10.29 Internally Studentized Residuals vs Run Number

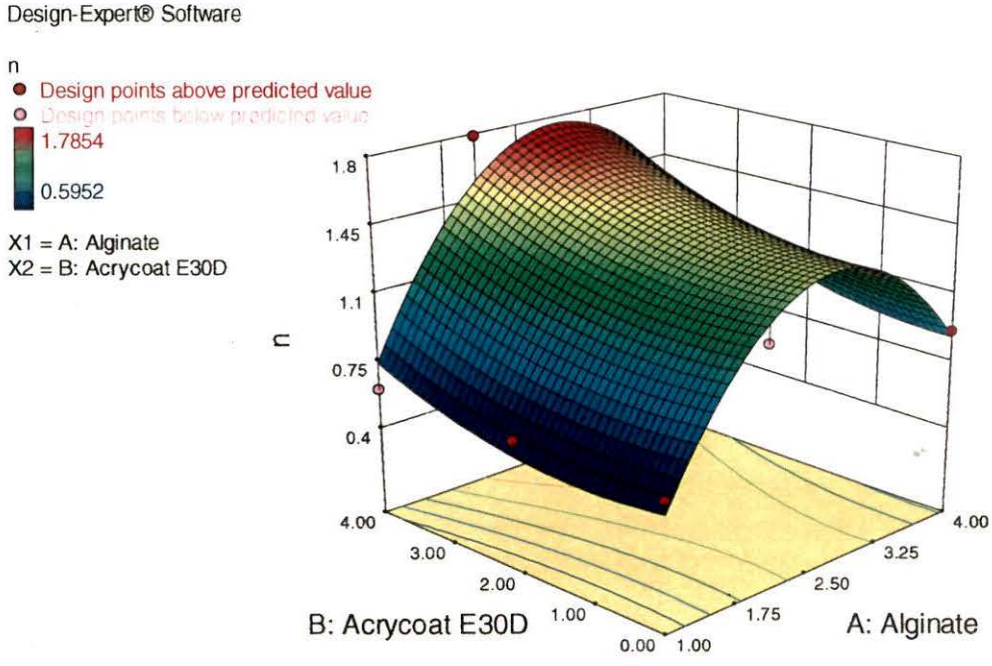

Figure 10.30 Effect of interaction between Alginate and Acrycoat E30D on the diffusion coefficient (n) represented in 3D Response curve

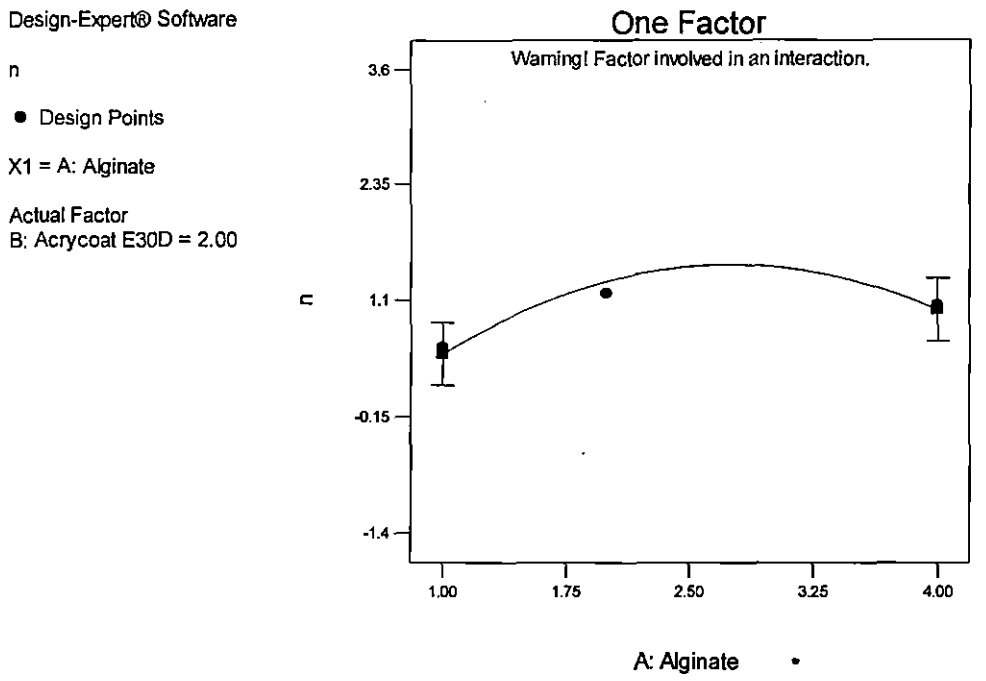

Figure 10.31 Sole effect of Alginate Concentration on 'n'

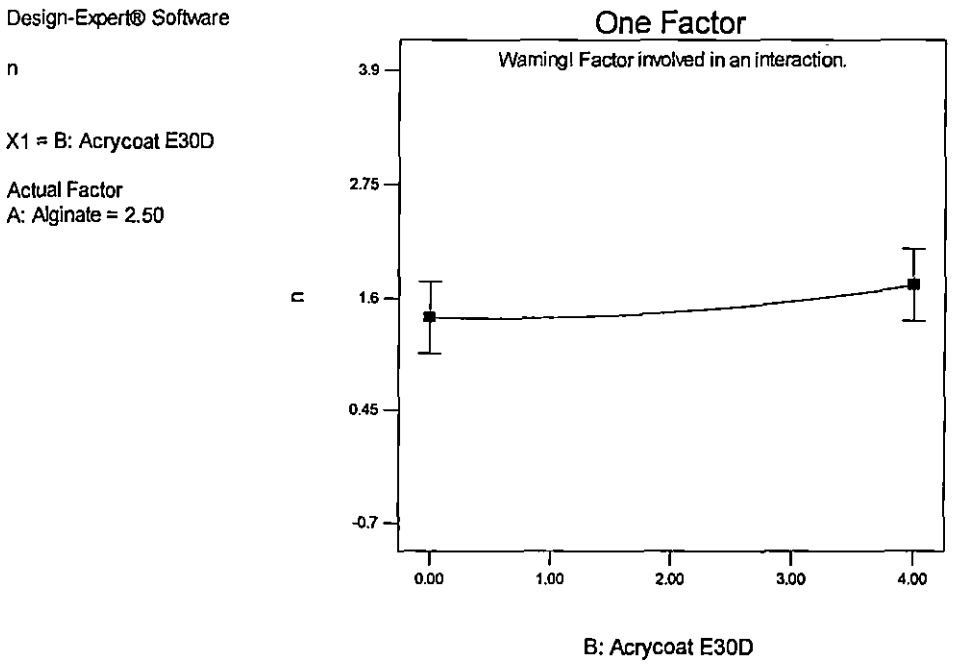

Figure 10.32 Sole effect of Acrycoat E30D Concentration on 'n'

10.9 Discussion

10.9.1 Effect on Ko (mg/hr) - Zero order release rate constant

The values of Ko obtained under the different experimental conditions for all the final nine formulations are summarized in Table 10.4. The variability associated with test samples was small as indicated by coefficient of variation values given in Table 10.6.

The application of RSM offers, on the basis of parameter estimate, an empirical relationship between the response variable Ko and the test variables under considerations. Quadratic model (Partial sum of squares – Type – III) was selected for all the RSM studies. By applying multiple regression analysis on the experimental data, the response variable Ko and the test variables A (concentration of Sodium alginate % w/w) and B (concentration of Acrycoat E30D % w/w) are related by the following second order polynomial equation:

$$Ko = 11.11 - 0.72 \times A - 0.28\ B - 0.23\ \times A \times B - 0.80 \times A^2 + 9.683E\text{-}003 \times B^2 \qquad (10.22)$$

A summary of the analysis of variance (ANOVA) for the selected quadratic predictive model is shown in Table 10.8. Statistical testing of the model was done in the form of analysis of variance (ANOVA) which is required to test the significance and adequacy of the model. Here the ANOVA of regression model demonstrates that the model is highly significant, as evident from the calculated F-value (40.21) and a very low probability value ($p < 0.006$). The model was found to be adequate for prediction within the range of the variables employed.

The coefficient values of the equation (10.22) were calculated and tested for their significance and are listed in the Table 10.8 and 10.9. The p-values are used as a tool to check the significance of each of the coefficients, which in turn may indicate the pattern of interactions between the variables. The smaller the value of p, the more significant is the corresponding coefficient. It is evident from the Table 10.8 that the significant coefficients were A ($p = 0.0019$), B ($p = 0.0127$) and A^2 ($p = 0.0105$) the p-values of all of them being small at 5% confidence interval ($p < 0.0500$). The coefficients A, B and A^2 had a negative effect on the zero order release rates (Ko) of the drug from the prepared micropellets.

The graphical representation of the regression equations (10.22 and 10.23) called the response surfaces and 3-D curves were obtained using the software Design Expert 7.1.2 and are presented in Figures 10.1 to 10.8. Figure 10.1 and 10.2 displaying normal probability curve of the studentized residuals which checks for the normality of residuals. Studentized

residuals versus predicted values were plotted (Fig.10.4 and 10.5) to check for constant error and it was found that error on statistical calculation was within very short and acceptable range. To measure the predictive power of regression equation obtained by RSM, the Pearson's correlation coefficient (r) was calculated between experimentally obtained actual Ko values and predicted Ko values. The result of the plot (Fig.10.3) gave close Pearson's correlation value (r = 0.9853) making the formulation optimized statistically. The 3D-curve in figure 10.6 shows a downward trend of the wire mesh depicting at higher level (+1) of sodium alginate concentration, the release rate (Ko) was definitely got affected and was in the line of the objective of the research work. When individual effect of the polymers were investigated in figure 10.7 and 10.8 it was clearly found that effect of sodium alginate was profound. With the increase in its concentration, values of Ko came down significantly. Whereas Acrycoat E30D showed flat effect and comply with the ANOVA result. So it is clearly evident from r values that the RSM can be used for formulation studies with high prediction power for Ko.

10.9.2 Effect on X_{120} (mg) - Amount of drug released in 2 hours (Burst effect)

The values of X_{120} obtained under the different experimental conditions for all the final nine formulations are summarized in Table 10.4. The application of RSM offers, on the basis of parameter estimate, an empirical relationship between the response variable X_{120} and the test variables under considerations. Quadratic model (Partial sum of squares – Type – III) was selected for all the RSM studies. By applying multiple regression analysis on the experimental data, the response variable X_{120} and the test variables A (concentration of Sodium alginate % w/w) and B (concentration of Acrycoat E30D % w/w) are related by the following second order polynomial equation:

$$X_{120} = 9.59 - 9.65 \times A - 3.46 \times B - 3.24 \times A \times B + 18.44 \times A^2 - 1.69 \times B^2 \qquad (10.24)$$

A summary of the analysis of variance (ANOVA) for the selected quadratic predictive model is shown in Table 10.10. Statistical testing of the model was done in the form of analysis of variance (ANOVA) which is required to test the significance and adequacy of the model. Here the ANOVA of regression model demonstrates that the model is highly significant, as evident from the calculated F-value (33.26) and a very low probability value

($p = 0.0079$). The model was found to be adequate for prediction within the range of the variables employed.

The coefficient values of the equation (10.24) were calculated and tested for their significance and are listed in the Table 10.10 and 10.11. The p-values are used as a tool to check the significance of each of the coefficients, which in turn may indicate the pattern of interactions between the variables. The smaller the value of p, the more significant is the corresponding coefficient. It is clear from the Table 10.10 that the significant coefficients were A ($p = 0.0024$), B ($p = 0.0424$) and A^2 ($p = 0.0027$) the p-values of all of them being small at 5% confidence interval ($p < 0.0500$). The coefficients A, B and A^2 had a negative effect on the burst release (X_{120}) of the drug from the prepared micropellets.

The graphical representation of the regression equations (10.24 and 10.25) called the response surfaces and 3-D curves were obtained using the software Design Expert 7.1.2 and are presented in Figures 10.9 to 10.16. Figure 10.9 and 10.10 displaying normal probability curve of the studentized residuals which checks for the normality of residuals. Studentized residuals versus predicted values were plotted (Fig.10.12 and 10.13) to check for constant error and it was found that error on statistical calculation was within very short and acceptable range. To measure the predictive power of regression equation obtained by RSM, the Pearson's correlation coefficient (r) was calculated between experimentally obtained actual X_{120} values and predicted X_{120} values. The result of the plot (Fig.10.11) gave close Pearson's correlation value ($r = 0.9823$) making the formulation optimized statistically. So it can be concluded from r values that the RSM can be used for studying X_{120} values of the formulation with high prediction power.

The 3D-curve in figure 10.14 shows a steep downward trend of the wire mesh depicting that by increasing the level of sodium alginate concentration from lower(-1) to medium(0) and then higher level (+1), the burst release amount of drug (X_{120}) got significantly reduced and was in accordance of the objective of the research work. When individual effect of the polymers were investigated in figure 10.15 and 10.16 it was clearly found from the u-shaped curve that effect of sodium alginate was pronounced above 2% (w/w) level. With the increase in its concentration, values of X_{120} came down significantly. Whereas Acrycoat E30D showed flat effect and comply with the ANOVA result.

10.9.3 Effect on t_{80} (hr) - Time required for 80% drug release

The values of t_{80} obtained under the different experimental conditions for all the final nine formulations are summarized in Table 10.4. The application of RSM offers, on the basis of parameter estimate, an empirical relationship between the response variable t_{80} and the test variables under considerations. Quadratic model (Partial sum of squares – Type – III) was selected for all the RSM studies. By applying multiple regression analysis on the experimental data, the response variable t_{80} and the test variables A (concentration of Sodium alginate % w/w) and B (concentration of Acrycoat E30D % w/w) are related by the following second order polynomial equation:

$$t_{80} = 8.17 + 1.65 \times A + 0.58 \times B - 0.027 \times A \times B - 1.31 \times A^2 + 0.083 \times B^2 \quad (10.26)$$

A summary of the analysis of variance (ANOVA) for the selected quadratic predictive model is shown in Table 10.12. Statistical testing of the model was done in the form of analysis of variance (ANOVA) which is required to test the significance and adequacy of the model. Here the ANOVA of regression model demonstrates that the model is highly significant, as evident from the calculated F-value (38.12) and a very low probability value ($p = 0.0065$). The model was found to be adequate for prediction within the range of the variables employed.

The coefficient values of the equation (10.24) were calculated and tested for their significance and are listed in the Table 10.12 and 10.13. The p-values are used as a tool to check the significance of each of the coefficients, which in turn may indicate the pattern of interactions between the variables. The smaller the value of p, the more significant is the corresponding coefficient. It can be seen from the Table 10.12 that the significant coefficients were A ($p = 0.0010$), B ($p = 0.0212$) and A^2 ($p = 0.0145$) the p-values of all of them being small at 5% confidence interval ($p < 0.0500$). The coefficients A, B and A^2 had a negative effect on the time taken to release 80% (t_{80}) of the drug from the prepared micropellets.

The graphical representation of the regression equations (10.26 and 10.27) called the response surfaces and 3-D curves were obtained using the software Design Expert 7.1.2 and are presented in Figures 10.17 to 10.24. Figure 10.17 and 10.18 displaying normal probability curve of the studentized residuals which checks for the normality of residuals. Studentized residuals versus predicted values were plotted (Fig.10.20 and 10.21) to check

for constant error and it was found that error on statistical calculation was within very short and acceptable range. To measure the predictive power of regression equation obtained by RSM, the Pearson's correlation coefficient (r) was calculated between experimentally obtained actual t_{80} values and predicted t_{80} values. The result of the plot (Fig.10.19) gave close Pearson's correlation value (r = 0.9845) making the formulation optimized statistically. So it is clearly evident from r values that the RSM can be used for formulation studies with high prediction power for t_{80}.

The 3D-curve in figure 10.22 shows a upward trend of the wire mesh depicting at medium level (0) onwards of sodium alginate concentration, the time required for the formulations to release 80% of the drug content (t_{80}) got retarded and was also in accordance with the objective of the research work. When individual effect of the polymers were investigated in figure 10.23 and 10.24 it was clearly found that effect of sodium alginate was profound and jumped from 5 hr to 7 hr with increasing its concentration from 1% to 2% (w/w). Further increment in the concentration does not show any pronounced retardation. In contrast, Acrycoat E30D showed flat effect and comply with the ANOVA result.

10.9.4 Effect on n – Diffusion coefficient of Peppas equation

The values of n obtained under the different experimental conditions for all the final nine formulations are summarized in Table 10.4. The application of RSM offers, on the basis of parameter estimate, an empirical relationship between the response variable n and the test variables under considerations. Quadratic model (Partial sum of squares – Type – III) was selected for all the RSM studies. By applying multiple regression analysis on the experimental data, the response variable n and the test variables A (concentration of Sodium alginate % w/w) and B (concentration of Acrycoat E30D % w/w) are related by the following second order polynomial equation:

$$n = 1.45 + 0.24 \times A + 0.17 \times B + 0.057 \times A \times B - 0.69 \times A^2 + 0.11 \times B^2 \qquad (10.28)$$

A summary of the analysis of variance (ANOVA) for the selected quadratic predictive model is shown in Table 10.14. Statistical testing of the model was done in the form of analysis of variance (ANOVA) which is required to test the significance and adequacy of the model. Here the ANOVA of regression model demonstrates that the model is

insignificant at a confidence interval of 5%, as evident from the calculated F-value (5.56) and a high probability value ($p = 0.0943$). The model was found to be inadequate for prediction within the range of the variables employed.

The coefficient values of the equation (10.28) were calculated and tested for their significance and are listed in the Table 10.14 and 10.15. The p-values are used as a tool to check the significance of each of the coefficients, which in turn may indicate the pattern of interactions between the variables. The smaller the value of p, the more significant is the corresponding coefficient. It is evident from the Table 10.14 that the significant coefficient is A^2 ($p = 0.0232$) the p-value being small at 5% confidence interval ($p < 0.0500$). The coefficient A, B had no significant effect and only the term A^2 had a statistically significant effect on the diffusion coefficient (n) of the drug from the prepared micropellets.

The graphical representation of the regression equations (10.28 and 10.29) called the response surfaces and 3-D curves were obtained using the software Design Expert 7.1.2 and are presented in Figures 10.25 to 10.32. Figure 10.25 and 10.26 displaying normal probability curve of the studentized residuals which checks for the normality of residuals. Studentized residuals versus predicted values were plotted (Fig.10.28 and 10.29) to check for constant error and it was found that error on statistical calculation was within very short and acceptable range. To measure the predictive power of regression equation obtained by RSM, the Pearson's correlation coefficient (r) was calculated between experimentally obtained actual n values and predicted n values. The result of the plot (Fig.10.27) gave low Pearson's correlation value ($r = 0.9026$). It thus signifies the fact the actual n-value does not closely correlate with the statistically predicted value. The 3D-curve in figure 10.30 shows an upside-down U-curve of the wire mesh depicting at medium(0) level highest value of n was observed with slow dipping at higher level (+1) of sodium alginate concentration. The n-values were near 1 at both the levels but at lower level it was near to 0.5 following purely Higuchi diffusion. When individual effect of the polymers were investigated in figure 10.31 and 10.32 it was found that effect of both sodium alginate and Acrycoat E30D were less pronounced. Overall, it can be inferred that the combination of polymers does not have any significant effect on the n-values. Concentrations of both the polymers above 2% (w/w) fail to affect the dissolution behavior of the drug from the formulated micropellets.

10.9.5 Inference

One of the primary objectives of the research work was to achieve controlled release microparticulate drug delivery systems in the form of micropellets, through the use of a combination of water-soluble polymer such as Sodium alginate and a water-insoluble polymer Acrycoat E30D. To control the release of the drug the polymer system must quickly hydrate to form a gelatinous outer layer. A rapid formation of a gelatinous layer is critical in preventing wetting of the interior and disintegration of the core of the micropellets. Once the original protective gel layer is formed, it controls the penetration of additional water into the micropellets. As the outer gel layer fully hydrates and dissolves, a new inner layer must replace it and be cohesive and continuous enough to retard the influx of water and control drug diffusion. Although gel strength is controlled by polymer concentration, polymer chemistry also plays a significant role. Research evidence[11] suggest that the chemistry of Sodium alginate and Acrycoat E30D encourage a strong, tight gel formation compared to other cellulosics. The formation of protective gel around the micropellets while in contact with dissolution medium depends on polymer concentration. There must be sufficient polymer content in a matrix system to form a uniform barrier. This barrier protects the drug from immediately releasing into the dissolution medium. If the polymer level is too low, a complete gel layer may not form. In the research work, increased polymer level in the formulations resulted in decreased drug-release rates (Ko) and burst effect (X_{120}) and increased time (t_{80}) for the release of 80% of the drug from the formulated micropellets. The effect of the primary polymer, sodium alginate was much more significant statistically ($p< 0.05$) than the copolymer Acrycoat E30D. The combined effect was proportional to the sole effect of sodium alginate. The justification of using a copolymer thus can be attributed only in retarding the release rate. All other pharmacokinetic parameters remained fairly unchanged with its incorporation. It can also be concluded that the concentration range within 2 – 4% (w/w) of sodium alginate can give a predictive and reproductive formulation. Though, concentrations above 4% was not investigated, it can be assumed from the statistical result that no robust change in the dissolution behavior will be achieved. On the contrary, higher concentration of polymers shall increase the particle size and may also pose mechanical problems during extrusion technique due to high polymer viscosity.

10.10 REFERENCES

1. L. Lachman, H.A.Lieberman and J.L. Kanig, *The Theory and Practices of Industrial Pharmacy*, 3rd Edition, Varghese Publishing House, Delhi, 1991, 243, 269.
2. D.C. Montgomery, *Design & Analysis of Experiments*, 4th Edition, John Wiley & Sons, New York, 1996, 243
3. R.A. Fisher, *The Design of Experiments*, 8th Edition, Hafner Publishing Co., New York, 1996,
4. S. Bolton, *Factorial Designs* in *Pharmaceutical Statistics: Practical and Clinical Applications*, 3rd Edition, Marcel Dekker Inc., New York, 1997.
5. W.G. Cochran & G.W. Snedecor, *Factorial Experiments in Statistical Methods*, 6th Edition, The Iowa State University Press, Iowa, 1967.
6. O.I. Davies, *Design & Analysis of Industrial Experiments*, Hafner Publishing Co., New York, 1956.
7. W.G. Cochran, *Statistical Methods*, 6th edition, Oxford & IBH Publishing Co., India, 1967, 339-377.
8. H. R. Lindman, *Analysis of Variance in Complex Experimental Designs*, W. H. Freeman & Co., San Francisco, 1974.
9. D.V. Lindley, *Regression and Correlation Analysis* in *New Palgrave: A Dictionary of Economics*, **Vol. 4**, 1987, 120-23.
10. N.A. Peppas, *Pharma Acta Helv.*, **9**, 1985, 298-302.
11. N.A.Peppas, R.Langer, *Review of Macromolecule Chem. Phys.*, **Vol. C23**, 1983, 61-126