Design Of Experiments

Plackett-Burman & Composite Design

Submitted by

Fatima Zgheib
Saja Al Ghazi
Ali Karbala
Hadi Joudieh

Submitted to

Dr. Zaher Khraibani

Contents

Abstract

In recent years, the combination of Response Surface Methodology models with experimental designs has gained popularity. The benefits of analyzing interactive effects among different factors, coupled with considerations of efficiency, have motivated a growing number of researchers to apply these models in natural product research. This review delves into the fundamental principles of factorial designs, central composite designs and plackett-Burman when utilized in conjunction with RSM, providing insights into their applications in the realm of natural product research. The goal of this review is to document experimental design techniques and their practical implementations within the field of natural product research. Additionally, the distinctive characteristics of each design are outlined, and the evolving trends of RSM applications in natural product research are thoroughly examined in this paper.

Keywords: Central composite design, RSM, Plackett-Burman

1 Approach And Definition:

Statistics is the scientific discipline that involves the collection, analysis, and interpretation of data. Various methods are employed for data collection, including sampling surveys, observational studies, and experiments [17]. Experiments and observational studies differ in their approach to variable manipulation, where experiments involve intentional changes made by researchers to observe the effects on response variables, while observational studies focus on naturally occurring changes in variables. An experiment is defined as a controlled test or series of tests where purposeful changes are made to input variables, allowing researchers to observe and understand the reasons for changes in the output response. In experiments, researchers manipulate or design variables to study their impact on the response variable. This concept aligns with definitions provided by [21], which emphasize the intentional and controlled nature of variable manipulation in experimental design.

2 History

In the 1920s and 1930s, *Ronald A.* Fisher pioneered the concept of experimental design at Rothamsted Experimental Station near London. Initially employed in agriculture, Fisher showcased the efficient derivation of valid conclusions from experiments conducted amidst natural variations and interfering factors such as temperature and soil conditions. This approach, addressing both known and unknown variables, gained popularity in the military and industry from the 1940s onward. Besse Day utilized it during World War II to resolve welding issues at a naval shipyard, while George Box optimized chemical processes at Imperial Chemical Industries. W. Edwards Dem-

ing introduced experimental design, applying it in Japan for quality improvement, and Genichi Taguchi, a pivotal figure in this movement, implemented his methods at Toyota. In the late 1970s, the U.S. industry embraced quality improvement initiatives like "Total Quality" and "Six Sigma," wherein the design of experiments played a crucial role as an advanced method in Six Sigma programs initiated by*Motorola* and *GE.*[7, 10, 12]

3 Aim

The objective for experimental design is demonstrated as follows:

1. Determining the variables with maximum influence on the response.
2. Determining where to set the influential variables so that the response is almost always near the desired outcome.
3. Determining where to set the influential variables so that the variability in y is small.
4. Determining where to set the influential variables so that the effect of uncontrollable variables are minimized [21].

4 Fundemantals

The core principles of experimental design aim to address issues related to nuisance and uncontrolled variables, thereby enhancing the efficiency of experiments. These fundamental principles include:

Randomization: This principle safeguards against unknown biases that may distort results.

Replication: The repetition of experiments is employed to increase precision.

Blocking: Precision is heightened by blocking,which eliminates the influence of known nuisance factors.

Orthogonality: This principle contributes to improved interpretation of results.

Factorial Experimentation: This approach estimates the effects attributable to individual factors as well as combinations of factors [7, 10, 12].

5 Different Designs

There are numerous experimental design types employed for various purposes,improved from trial and error to modern complicated and robust designs. SAS team in their book Design of Experiments Guide (SAS Institute,2015) adresses thirteen different types of experimental designs: Full Factorial Designs, Screening Designs, Definitive Screening Designs, Mixture Designs, Taguchi Array Designs, Evaluate Designs,Custom Designs, Response surface Designs, Augmented Designs, Non linear Designs, Accelerated Life Test Designs, Discrete Choice Designs, Space-Filling Designs. To maintain focus, we will briefly review several designs, such that full factorial design, response surface designs, and screening design. We will then delve deeper into **Central Composite (CCD) and Plackett-Burman Designs** for a more detailed exploration.

5.1 Full factorial design

Full factorial design contains all possible combinations of a set of factors. It supports both continuous and categorical factors with up to nine levels. The limitation in this design is the exponential growth in the number of runs, which is too expensive to run for most practical purpose.

5.2 Response surface designs

Response surface designs are useful for modeling a curved quadratic surface to continuous factors. A response surface model can approach a minimum or maximum response (the optimum value), if one exists inside the factor region. Three distinct values for each factor are necessary to fit a quadratic function, so the standard two-level designs cannot fit curved surfaces.The most popular RSM designs are the CCD and Box-Benkhen (BB) design.

5.3 Screening Designs

A screening design is a strategic experiment used when dealing with numerous potential factors. Its purpose is to identify the most influential ones affecting key responses, streamlining further experimentation by reducing the number of factors under investigation. In addition, it generally requires fewer experimental runs than other designs. This cost-effective process acts as a vital filter, eliminating unimportant factors before investing in more experiments that are extensive. For instance, imagine a manufacturing process with various parameters such as temperature, pressure, and material composition. Instead of conducting exhaustive experiments on all possible combinations of these factors, a screening design would help identify the most impactful ones. Notable features of a screening design include its ability to enhance quality control by establishing upper and lower control limits for specific variables, refine processes economically by identifying influential factors [22], and achieve the overarching goal of maximizing information while minimizing the number of experiments [33]. This structured approach not only improves product quality but also presents information in an understandable and readable format, allowing for efficient and reliable analysis of results expressed through mathematical expressions. Information gained can be used to optimize a process, and the repeatability of a process can be maintained [25], [31].In a screening design, each continuous factor is usually set at two levels to economize on the number of runs needed.The design consists of only a fraction of the possible combinations of factor levels.(SAS Institute,2015)

6 Plackett-Burman Design

A popular class of screening designs is the Plackett-Burman design (PBD), developed by R.L. Plackett and J.P. Burman in 1946.Plackett and Burman (PB) devised orthogonal arrays are useful for screening, which yield unbiased estimates of all main effects in the smallest design possible, Such designs are known as saturated designs. The main advantage of saturated designs is the minimum number of observations needed to calculate an effect for a certain factor. A selection of two-level Plackett-Burman designs is equal to the saturated fractional factorial designs. PBD require fewer experiments than the highly fractionated factorial designs that include the same number of factors. This technique is forming various combinations (which are called assemblies) of the components with varying amounts. With the help of this design, up to K=N-1 factors can be studied in N assemblies(runs) , where N must be a multiple of 4 , for more than seven factors and especially for n $\times$ 4 experiments,

i.e., 8, 12, 16, 20, etc., that are suitable for studying up to 7, 11, 15, 19, etc., factors respectively. After finding the critical factors, the next step was to optimize the concentrations of these components. Then we can use a response surface methodology (RSM) using a central composite design [25], [36], [27],[30], and [37]. The disadvantage of PB design is that the aliasing pattern is much more complex, each main effect is aliased with every two-way interaction not involving that effect. Lack of fit is difficult to assess, and first-order effects may be confounded with interaction effects.PB designs are Resolution III designs with the attribute of requiring the lowest number of runs, but do not allow the estimation of interactions between factors; it can identify the significant main factors that make up the possible significant interactions.Further analysis of the important main factors would allow the analyst to identify and estimate the significant interaction terms.Therefore, the use of a Plackett-Burman design is appropriate for screening [33].

The steps involved in a screening design, as outlined by [18], can be summarized as follows:

- Select the factors.
- Define the levels for the factors.
- Select the responses to be measured.
- Generate a design matrix of PBD.
- Randomize (block) and perform the experiments described in the experimental setup.
- Replicate the design.
- Develop a model.
- Conduct statistical or graphical analysis of the effects.
- Interpret and conclude from the statistical analysis.
- Recommend possible improvements, and if necessary, apply higher resolution designs.
- Build verification products.

6.1 Selection of Factors (Factor Levels and Responses):

The first critical step in a screening design is selecting the factors, defining their levels and responses, which have to be measured. Selection of factors is based on the experience of the researcher, who chooses one or several responses and discovers where in the experimental space, derivatives of the response in respect of each factor can be obtained [27].
If the factor range is too small, the variation of the response is too small and influence of the experimental error on the response is more; on the other hand if the range is too large, the first degree model used to interpret the results of the experimental design would not be valid. Therefore, in a screening PBD, a range between low and high levels of each factor is generally small in comparison with the ranges used during the optimization phase [27]
From the experiments, a number of responses can be determined depending on the experimenter's requirement and area of research. For instance, an experimenter would be interested in responses describing quantitative responses such as drug content, drug release in the area of formulation development or peak areas, and peak heights in chromatographic methods. Qualitative response factors such as resolution and relative retention can also be considered [34].

6.2 Design Matrix

After selection of factors and their levels, a design matrix is generated (for example, see Table 1). PBSD is used to indicate two level fractional factorials, although more levels are

possible. It allows efficient estimation of main effects of all factors being explored. An example of PBD with 12 runs and 11 factors is presented in Table 1. The columns represent factors with degrees of freedom equal to the number of levels in the column. The elements in the columns specify the levels high level (+1) and low level (−1) to be set for factors for the given experiment. PBD is particularly cyclical and the table of 11 factor 12 experimental design is obtained from a first line which describes the first experimental run and [+ − + − − − + + + − +]. The second line is obtained by moving the minus sign at the far right to the beginning of the next line and sliding the rest of the signs from the previous row one place along. The rest of the matrix is filled appropriately in the same way. Experiments 2–12 are obtained by noting all the cyclical permutations of this line. It is possible to verify that each factor is examined at six high and six low levels. Modified PBDs, such as centered and reflected designs, allow examination of factors at three levels [35]. Saturated PBDs can screen 'n' factors in n+1 runs, offering an economical approach. However, to accommodate non-linear responses, central points are added, transforming them into "Augmented Plackett and Burman" designs. This addition aids in estimating method repeatability. In essence, PBD provides researchers with a robust method for process control, balancing efficiency and accuracy in experimentation [14].

Trial	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
1	+	−	+	−	−	−	+	+	+	−	+
2	+	+	−	+	−	−	−	+	+	+	−
3	−	+	+	−	+	−	−	−	+	+	+
4	+	−	+	+	−	+	−	−	−	+	+
5	+	+	−	+	+	−	+	−	−	−	+
6	+	+	+	−	+	+	−	+	−	−	−
7	−	+	+	+	−	+	+	−	+	−	−
8	−	−	+	+	+	−	+	+	−	+	−
9	−	−	−	+	+	+	−	+	+	−	+
10	+	−	−	−	+	+	+	−	+	+	−
11	−	+	−	−	−	+	+	+	−	+	+
12	−	−	−	−	−	−	−	−	−	−	−

Figure 1: An example of 12 runs PB

6.3 Regression Coefficients and Analysis of Variance

Following the completion of all runs and the calculation of responses, the computation of regression coefficients is initiated in Plackett-Burman Designs (PBDs), which are constructed based on Hadamard matrices [14]. The results are interpreted using a first-degree polynomial model:

$$y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_{11}x_{11}$$

Where $'y'$ represents the predicted response, $x_1 - x_{11}$ denote the factor settings, $a_1 - a_{11}$ represent the regression coefficients and the intercept of mean.

As PBD is a saturated design, the main effect estimates do not show standard errors and all the degrees of freedom are used to estimate the factor main effects. Subsequently, after estimating the factor regression coefficients, the identification of significant factors influencing dependent variables undergoes an analysis of variance (ANOVA). ANOVA determines the factors significantly affecting the dependent variable through sum of squares (SS) calculations, F-ratios (F) representing the ratio of mean-square effect (MS) to mean-square error, and 'P'-probability values indicating the significance of factors impacting the response.

6.4 Graphical Representation of Results :

6.4.1 Diagnostic Plots of Residuals:

To start with, before accepting a particular "model" that includes a particular number of effects, the distribution of the residual values is examined. These are computed as the difference between the predicted values and the observed values. In this plot the actual residual

values are plotted along the horizontal X-axis; the vertical Y-axis shows the expected normal values for the respective values, after they are rank-ordered. If all values fall onto a straight line, then one can be satisfied that the residuals follow normal distribution. A graphical representation is shown in Figure 1 as generated by the software Design Expert V. 6.0.5 © Stat ease Inc., U.S.A

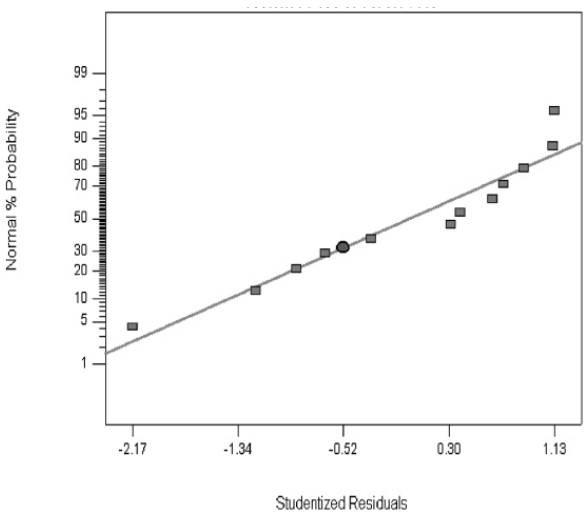

Figure 2: Graphical representation of diagnostic plots of residuals

6.4.2 Pareto Chart of Effects:

The Pareto chart of effects is often an effective tool for communicating the results of an experiment, in particular to laymen. In this graph, the ANOVA effect estimates are sorted from the largest absolute value to the smallest absolute value. The magnitude of each effect is represented by a column, and often, a line going across the columns indicates how large an effect has to be (i.e., how long a column must be) to be statistically significant (Figure 3).

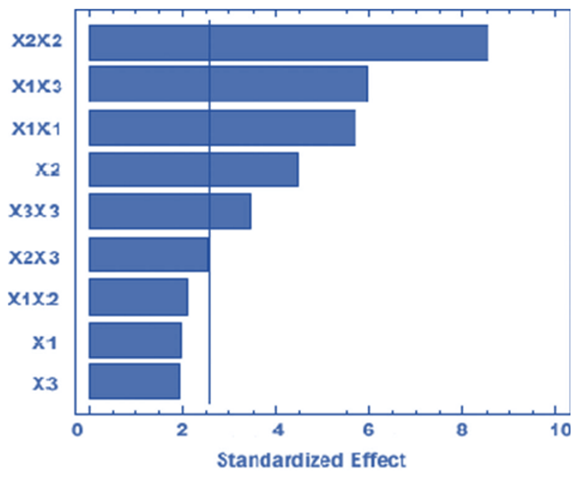

Figure 3: Representation of pareto chart

6.4.3 Detection of Significant Sources

Another useful technical summary graph, is the normal probability plot of the estimates. Because the number of degrees of freedom for the error term is small in saturated designs, the power of classical ANOVA will be too low. For this reason a graphical tool, the half-normal plot, can be used to which the algorithms of Length and Dong are applied to identify possible significant effects and to estimate standard deviation of the effects. Significant effects in half-normal plots are detected through visual inspection. A graphical representation is shown in Figure 3 as generated by the software Design Expert V. 6.0.5 © Stat ease Inc., U.S.A. The slope of the line through the effects assumed non-significant gives an estimate of the standard deviation (σ) of the error [27]

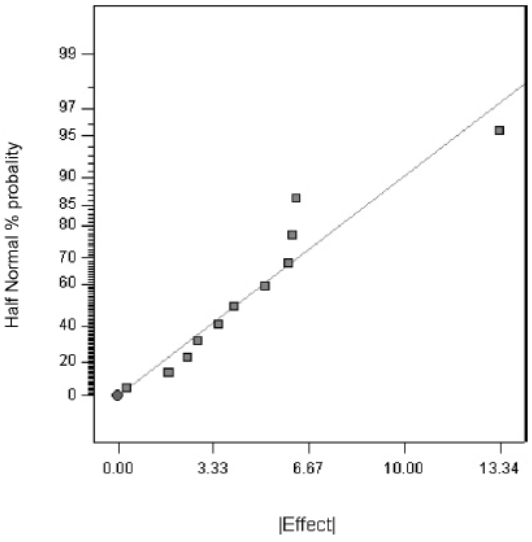

Figure 4: A graphical representation of half normal plot

6.4.4 Choice of Experimental Dummies:

Another important part of the PB screening design was the choice of dummies. A dummy is a component whose level does not change in the design. Factors known to have no effect can be chosen as dummies, or any factor not chosen as a variable can be included as a dummy. The dummies are used to obtain the estimate of error and normally three dummy variables will provide an adequate estimate of error [32]

6.4.5 Data Analysis for the Plackett-Burman Design:

Analysis for the Plackett-Burman experiment, referred to as the Plackett-Burman analysis henceforth, was carried out as follows. First, for all the components, including dummies, their effect on the response was calculated, which was the difference between the average response for the assemblies having a higher level of a given component and that for the ones having its lower level. Thus, the effect of a given component on the response parameter R could be written as:

$$Effect = \frac{2[\sum R(H) - \sum R(L)]}{N}$$

Where $R(H)$ = response parameter of an assembly in the screening design that contains the higher quantity of a given component, h^{-1} , $R(L)$ = response parameter of an assembly in the screening design that contains the lower quantity of a given component, h^{-1} , and N = number of assemblies. As indicated, the response parameter used in this study was μ. If there were no interactions among the variables, the effects shown by dummies would be zero. Similarly, the factor that had no effect would give a value of zero if no interactions existed. A positive sign for the effect meant that the component would increase the response if added at a higher level, and vice versa. After determining the effect of each component, a corresponding factor mean square (FMS), which is similar to the variance for the measurements of a single quantity, was calculated. This was the average squared effect of a component on the response, i.e.

$$\text{Factor mean square} = \frac{(effect)^2}{N} = \frac{4[\sum R(H) - \sum R(L)]^2}{N^3}$$

The error mean square (EMS) was calculated by averaging the FMS for all the dummies. Finally, an F-test on the ratios of the FMS to the EMS was performed to identify the factors that showed large effects.

6.4.6 Reverse Plackett-Burman Design:

To avoid misinterpretation of the results due to interactions among components, the Reverse Plackett-Burman design was also carried out experimentally, as suggested by *Nelson* (1982)[26].This design was constructed by reversing the signs (–1 to +1 and +1 to – 1) for all the assemblies that were used in the Plackett-Burman design . The analysis performed for all the assemblies used in the Plackett-Burman and Reverse Plackett-Burman designs, which will be referred to as the combined analysis henceforth, gave results that were not affected by two-factor interactions.

6.4.7 F-Test for the Null Hypothesis of Equal Means:

The F-test was performed to test the null hypothesis that there was no significant difference between the ratios of the FMS to the EMS for various components, which was equivalent to stating that none of the components had a significant effect on the specific growth rate. If the null hypothesis were true, and there were no sampling errors and interactions among the components, one would expect the value of F statistic to equal 1.0. However, since the sampling errors and interactions among components are usually present, as reflected by a nonzero value of the EMS, a sampling distribution can be used to make a probability statement. The sampling distribution of F, for a given ratio of the FMS to the EMS, determines the probability that one can expect the FMS to be greater than the EMS by a factor that equals the given ratio. If this probability meets

a preset criterion (P = 0.05), the null hypothesis can be rejected, leading to the conclusion that the component has a significant effect on the specific growth rate.

7 CCD

RSM is a powerful multivariate statistical tool that involves fitting empirical models to experimental data obtained through carefully designed experiments [4]. Employing lower-order polynomials [38], RSM has proven to be a reliable method for chemical process applications [28, 3]. To overcome the limitations of two-level designs in screening studies, CCD emerges as a compelling solution. While two-level designs excel in simplicity, they struggle to capture non-linear relationships and maxima, restricting analyses to linear models,[15]. Attempting a full factorial design with more than two levels intensifies the workload due to the increased number of experiments. Acknowledging these challenges, Box and Wilson [8] introduced CCD to address the need for greater level flexibility without the burden of exhaustive experiments, making it well-suited for fitting a quadratic surface—a model that typically proves effective in process optimization [29, 19, 24]. CCD comprises three integral components: a full or fractional factorial design, axial points, and the central composite design itself—a prevalent choice for fractional factorial designs within the response surface model.

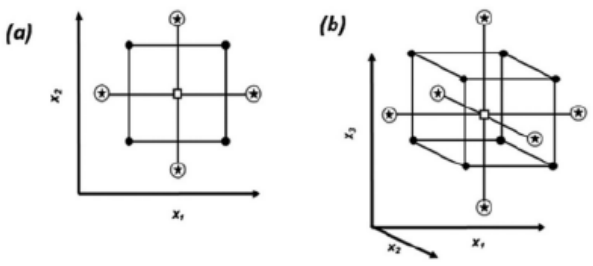

Figure 5: Illustration of CCD for (a) two factors and (b) 3 factors optimization.Every design consists of factorial points(•),star points(⋆),and central points(□)

Within the realm of Response Surface Methodology (RSM), CCD is a versatile tool frequently explored for optimizing diverse research problems. At the core of CCD's effectiveness lies its unique composition, organized into three key groups of design points [20].

Firstly, the foundation is established with a set of two-level factorial or fractional factorial design points (2k), meticulously constructed to encompass the entire spectrum of factor combinations at +1 and -1 levels. This foundational layer establishes a comprehensive groundwork for exploring linear relationships in the experimental space.

Building upon this, the second group introduces 2k axial points, often referred to as star points, strategically positioned at a specified distance (alpha) from the center. These axial points play a crucial role in generating quadratic terms, enabling CCD to capture and analyze intricate second-order polynomial equations. This dynamic addition extends the design's capability to discern non-linear relationships and optimize complex systems.

Lastly, the inclusion of center points completes the triad of design components. These center points serve a dual purpose: not only do they provide replicate terms essential for enhancing statistical reliability, but they also furnish independent estimates of experimental error. This meticulous arrangement empowers researchers with a robust framework for fitting precise second-order polynomial equations, making CCD a preferred choice in the RSM toolkit. In essence, CCD's synergy of factorial points, axial precision, and center point reliability renders it a potent instrument for navigating the intricacies of optimization in diverse research domains.

It can be seen in Figure 5 that the factorial

and star points lie equidistant from the central point. It allows the design to cover the factor space near the central point with more points than at the peripheral area of the factorial design. Considering these points, the number of experiments designed by CCD will be: $N = 2^k + 2 * k + n$

- N is the total number of experiments
- k the number of factors studied
- n is the number of replicates

7.1 Determination of α:

In CCD, value of alpha is important to calculate as it could determine the location of axial points in experimental domain. α Value can be defined as the calculated distance of each individual axial point (star point) from the center in the center composite design [2]. Depending on alpha value, design is spherical, orthogonal, rotatable, or face centered. If Alpha (α) is less than 1, it indicates the axial point must be a cube, and if it is greater than 1, it indicates it is outside the cube. In central composite design, each factor has five levels, i.e., Extreme high or otherwise called a star point, higher point, center point, low point, and finally, extreme low star point $(-\alpha, -1, 0, +1, +\alpha)$. Coming to the Alpha ($\alpha$) determination; which can be determined by the following equation:

$$\begin{aligned} \alpha &= (\text{Number of factorial runs})^{1/4} \\ &= (2^k \text{ or } 2^{k-r})^{1/4} \end{aligned}$$

Experimental results obtained are analyzed using response surface regression procedure of statistical analysis system. Correlation between responses and independent variables is obtained by fitting them into second order polynomial equation [16]

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_{ii}^2 + \sum_{i=1}^{k} \sum_{j=1, j \neq i}^{k} \beta_{ij} x_i x_j + \epsilon$$

Here,

- y represents the responses,
- k is the total number independent factors,
- β_0 is an intercept,
- i, ii, and ij with β represent the coefficient values for linear, quadratic, and interaction effects, respectively,
- x_i and x_j in the above equation show the coded levels for independent variables [13].

The quality of the polynomial model was expressed by the coefficient of determination, namely, R^2 and $Adj - R^2$. The statistical significance was verified with adequate precision ratio and by the F test. For statistical calculations, the variables X_i (the real value of an independent variable) were coded as x_i (dimensionless value of independent variable) according to the following equation:

$$x_i = \frac{X_i - X_0}{\delta X}$$

Where X_0 is the value of X_i at the center point, and δX represents the step change [23]

8 Application on Plakett-Burman and Central Composite Design:

There are numerous articles that cover applications that use Plackett-Burman and Central Composite Designs, here is a list of some of these articles:

1- Process optimization of intermediate-wave infrared drying: Screening by Plackett–Burman; comparison of Box-Behnken and central composite design and evaluation: A case study [6].

2- Optimization of ß-carotene production from agro-industrial by-products by Serratia marcescens ATCC 27117 using Plackett–Burman design and central composite design [1].

3- Optimization of Extraction or Purification Process of Multiple Components from Natural Products: Entropy Weight Method Combined with Plackett–Burman Design and Central Composite Design [11].

4- Application of Plackett-Burman and central composite designs for screening and optimization of factor influencing the chromatographic conditions of HPTLCmethod for quantification of efonidipine hydrochloride [9].

Without loss of generality and to dive into practice, after discussing theoretical part of Placket-Burman and Central Composite Designs, only the last article (Chaudhari-Shirkhedkar,2020) will be discussed. This article is from pharmaceutical field, aims to find the factor that influence the quantity of a drug called efonidipine at chromatographic conditions,and to find the optimum level of several factors for maximum quantification. We will focus on the screening and the optimization part and the rest details is out of our concern.

8.1 Statistical tool

R software (version 4.2.1) was used for developing PBD for the screening of independent variables and for designing a CCD and RSM for optimization of independent variables.

8.2 Plackett-Burman design screening of independent variables

The PBD serves as an effective screening design aimed at pinpointing the pivotal variables from a multitude of factors that ultimately influence the outcomes. It was applied to identify the significant independent variable impacting both peak area and R_f. The chosen factors for the PBD are wavelength, saturation time, dichloromethane (DCM) volume, triethylamine volume, development distance, and plate activation. These factors were systematically investigated to ascertain the crucial factor for peak area and R_f, selected as responses with a 95% confidence level. The specific levels of independent variables utilized in the PBD are detailed in table 1. The proposed study involved 12 experimental runs conducted using R to analyze the effects of six factors on responses. Each factor was examined at two levels: low level (-1) and high level (+1).The experimental runs for the PBD are outlined in Table 2. The PBD employed a first-order polynomial model, expressed in equation(1), for mathematical modeling.

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i \tag{1}$$

where Y is the responses (peak area and R_f), β_0 is the model intercept, β_i is a linear coefficient, and X_i is the level of independent variables (Naveena et al.2005; Vanaja and Shobha Rani 2007). [5]

Sr. No.	Independent variables	Unit	Experimental value	
			Low:(−1)	High:(+1)
1	Wavelength	nm	248	254
2	Saturation time	min	15	25
3	Volume of dichloromethane	mL	1.5	2.5
4	Volume of triethylamine	mL	0.3	0.7
5	Development distance	cm	7.5	8.5
6	Activation of plate	min	8.0	12

Table 1: Experimental Design Matrix for the Factors

Run order	Wave length	Saturation time	Volume of DCM	Volume of triethylamine	Development distance	Activation of plate	Peak area	Retention factor
1	1	1	-1	1	-1	-1	4985.45	0.77
2	-1	-1	-1	-1	-1	-1	6736.95	0.58
3	1	-1	1	1	-1	1	5383.96	0.52
4	-1	1	1	-1	1	-1	5565.23	0.5
5	1	-1	1	-1	-1	-1	5954.12	0.46
6	-1	-1	-1	1	1	1	9045.56	0.43
7	-1	1	1	1	-1	1	4005.6	0.49
8	1	-1	-1	-1	1	1	8195.95	0.48
9	1	1	1	-1	1	1	7902.53	0.47
10	-1	-1	1	1	1	-1	9265.45	0.55
11	-1	1	-1	-1	-1	1	4601.98	0.45
12	1	1	-1	1	1	-1	8315.76	0.46

Table 2 : Plackett-Burman design experimental runs for screening the significant independent variables affecting on the responses

In order to evaluate the suitability of our regression model, a comprehensive analysis was conducted using residual plots. This technique allowed to assess the adequacy of the model's fit to the data. The careful examination of residuals provided valuable insights into the model's performance, confirming its reliability for the dataset. Residuals should be scattered randomly around the horizontal axis such that residuals should have a mean value close to zero. A mean significantly different from zero might indicate a bias in the model.

In our investigation, we focused our attention on the first response variable, namely the peak area. By drawing the residual plots specific to this variable (figure 6) , we aimed to measure the effectiveness of our regression model in capturing the details of peak area variations. The results from the residual analysis were encouraging, providing a clear indication of the model's aptitude for explaining the observed patterns in peak area.

Conversely, when extending our analysis to the residual plots of the second response variable, namely the rtnfactor (figure 7) . The examination of residuals for this aspect showed that our model predictions didn't match the actual data very well. Unusual findings and noticeable patterns in the residuals hinted that our current model might not be the best fit for understanding the complexities in how it changes.

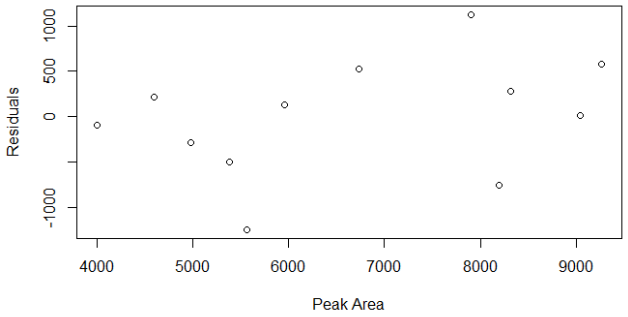

Figure 6: Residual plot of peak area

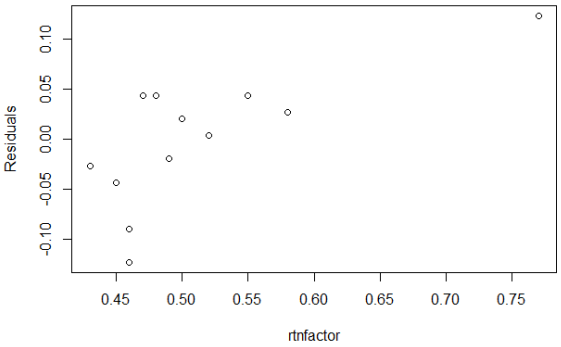

Figure 7: Residual plot of Retention factor

The statistical results of screening for significant factors are shown in Table 3. The first-order polynomial model equation for peak area ($Y1$) and retention factor ($Y2$) can be written as:

$Y1$ = 6663.2 + 126.4 (Wavelength)+ -767.1 (Saturation time) + -317.1 (volume of DCM) + 170.4 (Volume of triethylamine) + 1385.2 (Development distance) + -140.6 (Activation of plate)

$Y2$ = 0.51333 + 0.01333 (Wavelength) + 0.01 (Saturation Time) - 0.015 (Volume of DCM) + 0.02333 (Volume of triethylamine)- 0.03167 (Development distance)- 0.04 (Actiation of plate)

The coefficient of determination R^2 value of 87.77% and 47.57%, respectively, for peak area and retention factor indicated that up to 87.77% and 47.57% variability in peak area and retention factor could be estimated. This shows that the fitted model to Peak Area are significant but for retention factor is not significant.

The result of the study revealed that the impact of saturation time and development distance have a positive influence (p value < 0.05) on peak area, whereas the effect of wavelength, saturation time, volume of DCM, volume of triethylamine, development distance, and activation of plate has negative influence (p value > 0.05) on the peak area. While all factors for Retention Factor shows a high p-value (>0.05) which proves the claim of non significance. Consequently, the saturation time and development distance were selected for the further optimization stage and their influence on peak area.

The statistical analysis of experimental data was conducted using an F test for ANOVA, and the findings are presented in Table 4.

The model's F value of 5.98 in relation to peak area indicates a significant relationship between the response and independent variables, suggesting that independent variables in the proposed model improve the fit. Conversely, for retention factor, the F value of 0.76 suggests no significant relationship between the response and independent variables. The significance of individual independent variables was assessed using probability values, and a (p-value < 0.05) indicates the significance of each factor.

In addition, pareto charts are drawn for both models to have an idea about the effects of the factors (linear coefficients), see figure 8 and 9. It is clear that the development distance and Saturation time are the highest effects in the peak area experiment.

Factors	Effect	Coefficient	Standard error coefficient	T value	P value
Peak area					
Constant		6663.2	272.9	24.415	2.15e-06 ***
Wavelength		126.4	272.9	0.463	0.66268
Saturation Time		-767.1	272.9	-2.811	0.03751 *
Volume of DCM		-317.1	272.9	-1.162	0.29778
Volume of triethylamine		170.4	272.9	0.624	0.55973
Development distance		1385.2	272.9	5.075	0.00385 **
Activation of plate		-140.6	272.9	-0.515	0.62836
Retention factor					
Constant		0.51333	0.02836	18.099	9.46e-06 ***
Wavelength		0.01333	0.02836	0.470	0.658
Saturation Time		0.01000	0.02836	0.353	0.739
Volume of DCM		-0.01500	0.02836	-0.529	0.620
Volume of triethylamine		0.02333	0.02836	0.823	0.448
Development distance		-0.03167	0.02836	-1.116	0.315
Activation of plate		-0.04000	0.02836	-1.410	0.218

Table 3: PBD for statistical analysis and effect of independent variables on response variables

Source	DF	Adj. SS	Adj. MS	F value	p value
Peak area					
Model					
Linear					
Wavelength	1	191774	191774	0.2146	0.662683
Saturation Time	1	7061677	7061677	7.9005	0.037513 *
Volume of DCM	1	1206350	1206350	1.3497	0.297777
Volume of triethylamine	1	348509	348509	0.3899	0.559731
Development distance	1	23025404	23025404	25.7606	0.003849 **
Activation of plate	1	237271	237271	0.2655	0.628358
Error	5	4469113	893823		
Total	11	36540098			
Retention factor					
Model					
Linear					
Wavelength	1	0.002133	0.0021333	0.2210	0.6581
Saturation Time	1	0.001200	0.0012000	0.1243	0.7388
Volume of DCM	1	0.002700	0.0027000	0.2797	0.6195
Volume of triethylamine	1	0.006533	0.0065333	0.6768	0.4481
Development distance	1	0.012033	0.0120333	1.2465	0.3150
Activation of plate	1	0.019200	0.0192000	1.9890	0.2175
Error	5	0.048267	0.0096533		
Total	11	0.092066			

Table 4: Analysis of variance for peak area and retention factor using Plackett-Burman design

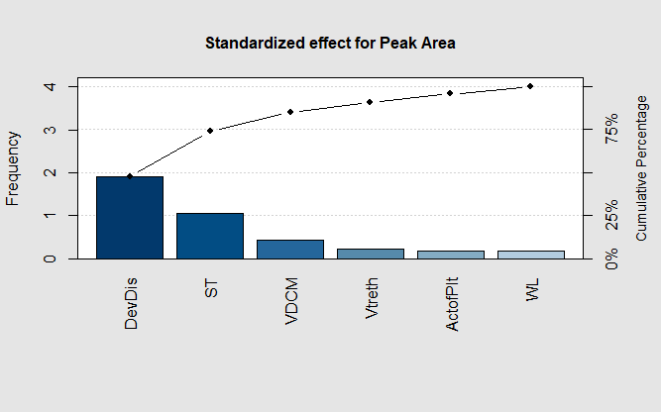

Figure 8 : Standardized effect of Peak Area

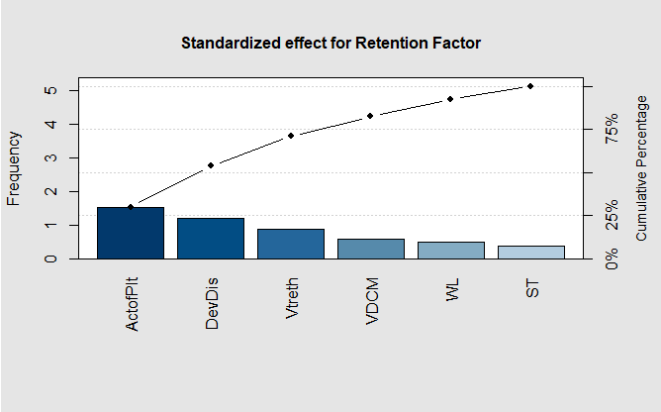

Figure 9 : Standardized effect of Retention Factor

8.3 Optimization of HPTLC conditions using face- centered CCD and RSM

Following the results obtained from the PBD in screening independent variables, the two most influential and significant factors were identified for further optimization. The aim was to determine the optimal values of these variables to achieve the maximum peak area for EFH. Consequently, in this study, Response Surface Methodology (RSM) was applied using a 2-factor, 3-level face-centered Central Composite Design (CCD) to optimize and analyze the NP-HPTLC chromatographic separation conditions. The chosen significant independent variables from PBD, namely development distance (X1) and saturation time (X2), were investigated at three different coded levels (7.5, 8.0, and 8.5 for X1 and 15, 20, and 25 for X2). The dependent response variable chosen for optimization was the peak area (Y), with the objective of obtaining the maximum peak area. A total of thirteen experimental runs were conducted based on the face-centered CCD matrix comprising eight non-center points and five center points, as outlined in Table 5.

Moreover, the statistical model was validated using Analysis of Variance (ANOVA), employing Fisher's statistical analysis, p-value, and the coefficient of determination (R^2) to assess the goodness of fit of the regression model (Priyadharshini and Bakthavatsalam 2016). By exploring multiple regression analysis, a second-order polynomial equation was generated to describe the relationship between the peak area (Y) to the elected independent variables X1 and X2 being represented in the following expression:

$$Y = 6677.541 + 210.323X_1 - 693.397X_2 - 729.18X_1X_2 - 215.492X_1^2 - 1191.182X_2^2$$

where Y is the peak area, X1 development distance (cm), and X2 saturation time (min).

By considering the ANOVA and the statistical analysis, results are shown in Tables 6 and 7.

The variable 'development distance' (X1) lacks statistical significance in both first and second-order terms (p-value > 0.05). Conversely, the variable 'saturation time' (X2) is statistically significant in both first and second-order terms. The interaction between these variables is significant, indicating that the effect of development distance is primarily manifested through interaction. Additionally, the high R-squared (0.967) and adjusted R-squared (0.944) values suggest the significance of the quadratic model.

The lack of fit value, 0.0173545, signifies there is no difference between the model and observed data. A lack of fit value close to zero suggests a good model fit. However, it's crucial to consider the significance level for interpretation. In this case, the lack of fit value being 0.0173545 should be compared to p-value at 5% significance level. The lack of fit is below the significance level , so it suggests that the model adequately fits the data.
All the fitted linear models in the experiment are supported by the residual plots that show random distribution around the factor axes, see Figure 10.

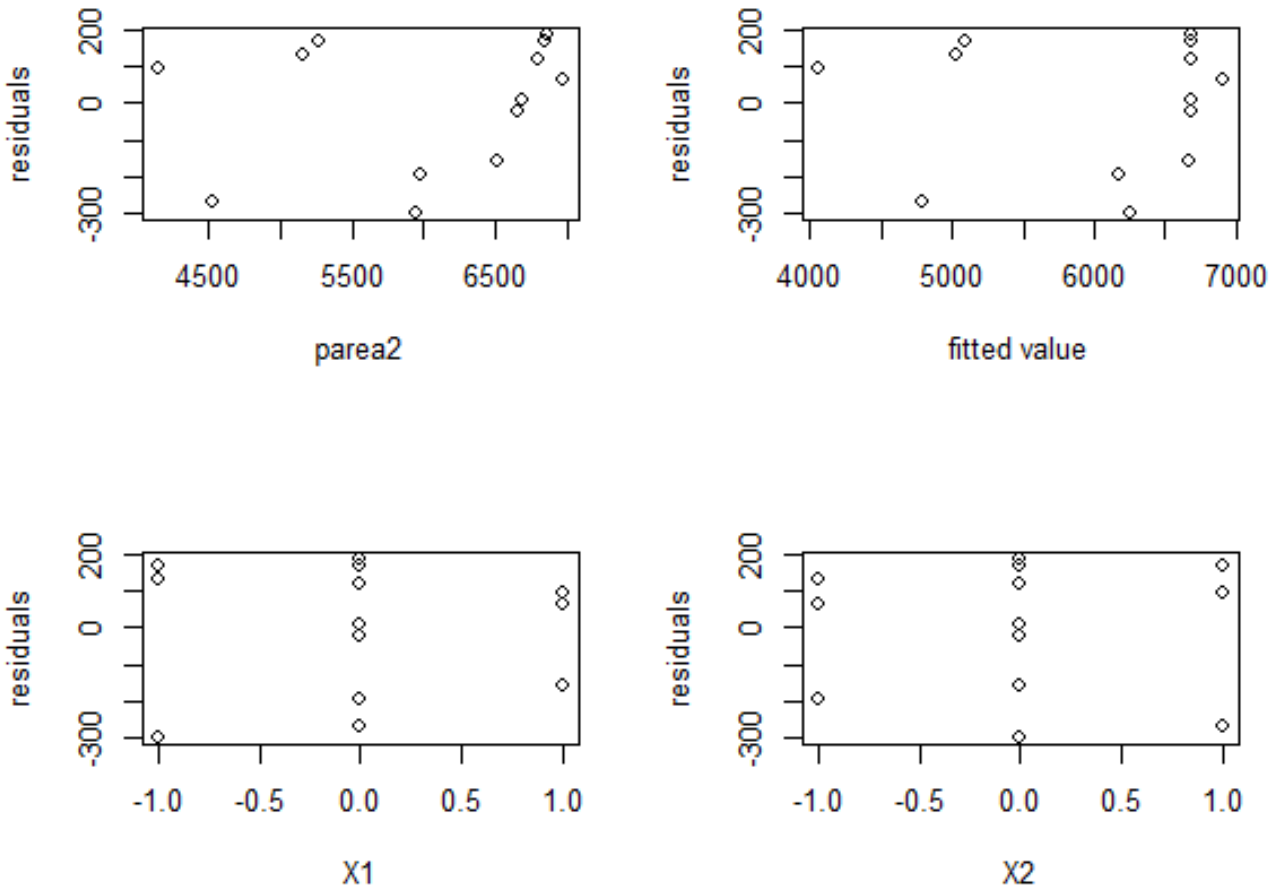

Figure 10 : Residual plots

The optimum value is visualized by the 3D plot and the contour plot (Figure 11), the maximum of the plot at a certain value of peak area, which theoretically is the solution of the nullity of the derivative of the model equation (saturation point). The optimum value tends to be at high level of Development Distance and the low level of Saturation time (8.5 , 15.432) which yields the Peak Area 6977.8.

Runs	Development distance(mm)	Saturation time (min)	Peak area
1	7.5	15	5156.98
2	8.5	15	6965.74
3	7.5	25	5264.65
4	8.5	25	4156.69
5	8.0	20	6865.48
6	8.0	20	6795.53
7	8.0	20	6656.69
8	8.0	20	6845.12
9	7.5	20	5951.18
10	8.5	20	6512.32
11	8.0	15	5985.56
12	8.0	25	4526.56
13	8.0	20	6685.48

Table 5: Face centered CCD matrix with two independent variables studied accompanied by results marked

	Estimate	Standard error	T value	Pr($>\|t\|$)
Intercept	6677.541	94.390	70.7442	2.965e-11 ***
x_1	210.323	92.803	2.2663	0.0577892
x_2	-693.397	92.803	-7.4717	0.0001406 ***
x_1: x_2	-729.180	113.661	-6.4154	0.0003619 ***
${x_1}^2$	-215.492	136.784	-1.5754	0.1591620
${x_2}^2$	-1191.182	136.784	-8.7085	5.282e-05 ***

Table 6: CCD for statistical analysis and effect of independent variables on response variables

	D_f	Sum sq	Mean sq	F value	pr(>F)
FO(x1, x2)	2	3150209	1575105	30.481	0.0003507
TWI(x1, x2)	1	2126814	2126814	41.158	0.0003619
PQ(x1, x2)	2	5366070	2683035	51.922	6.329e-05
Residuals	7	361724	51675		
Lack of fit	3	326330	108777	12.293	0.0173545
Pure error	4	35393	8848		

Table 7:Face-centered CCD ANOVA results for peak area

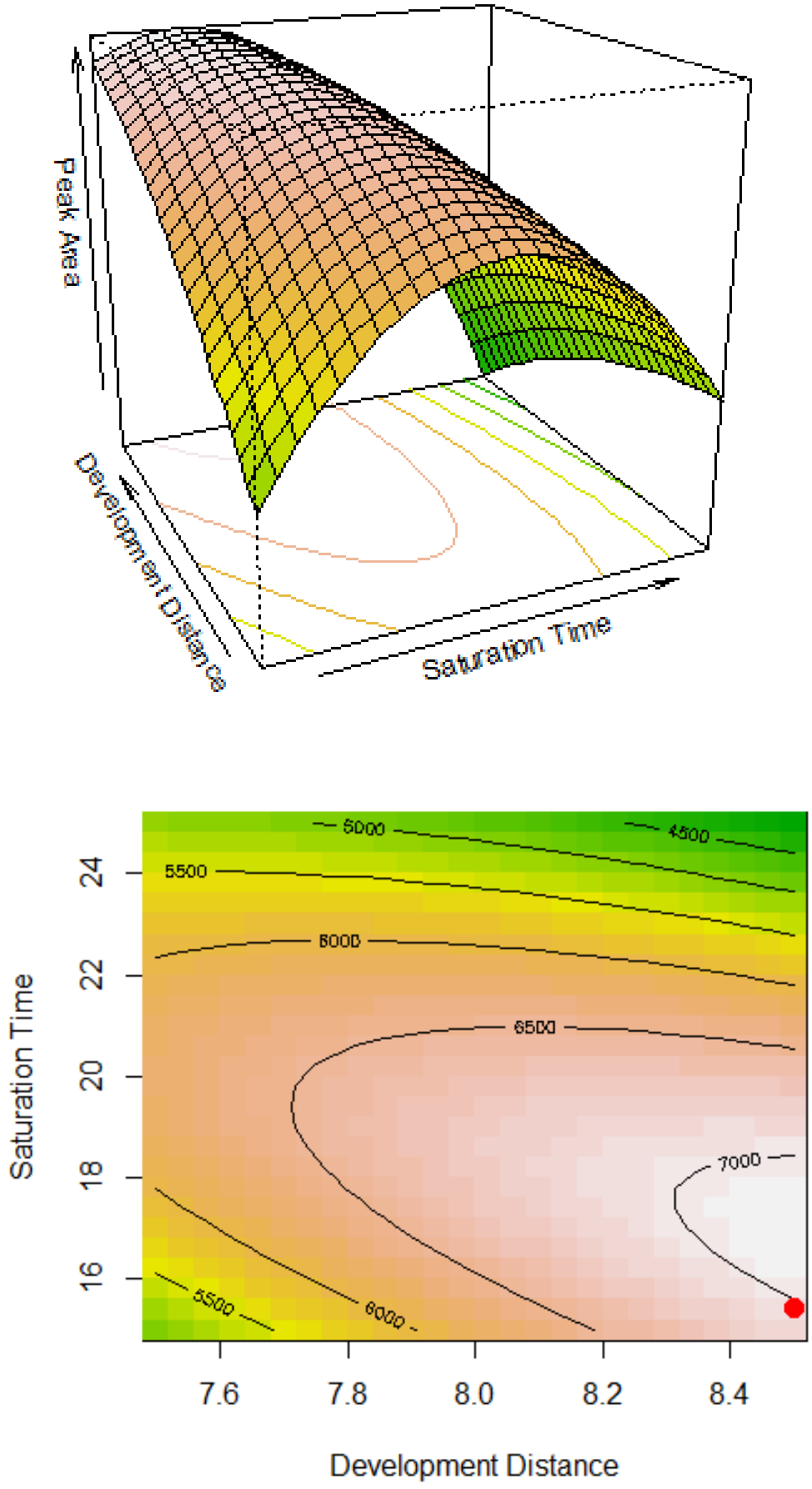

Figure 11 :3D response surface plots and counter plots demonstrating the relationship between the development distance and saturation time

9 Conclusion:

The CCD model is useful for modeling and analyzing programs in which the response of interest influences several variables. The CCD model can be considered as a robust statistical tool for process optimization. The best part of CCD is, as compared to Plackett–Burman design, a limited number of experiments are required with less computational experience. The biggest challenge of the CCD model is finding the critical factor. Central composite designs are beneficial in sequential experiments because you can often build on previous factorial experiments by adding axial and center points.

The proposed study outlines the utilization of both the Plackett-Burman design and central composite design to screen and optimize independent variables in an HPTLC method. The primary aim is to enhance the separation and detection of EFH in its pure form and within commercial pharmaceutical preparations. The key objective is to assess the significant independent variables influencing EFH peak area and retention factor, with the goal of achieving maximum peak area and an appropriate retention factor.

Out of the six independent variables considered, the study identifies development distance and saturation time as the crucial factors affecting EFH peak area. Central composite design and response surface methodology are then employed to optimize these variables. The results from experimental designs reveal that even small adjustments in development distance and saturation time directly impact the EFH peak area. Therefore, careful control of these variables is imperative during the chromatographic conditions optimization process.

In conclusion, the established method is characterized as simple, sensitive, and specific, making it suitable for quantifying EFH in both bulk and pharmaceutical formulations.

10 Abbreviations

CCD : Central Composite Design
RSM : Response Surfce Methodology
PBD : Pluckett - Burman Design
PB : Pluckett - Burman
F : F-ratios
MS: Mean-square effect
SS : Sum of squares
EMS: Error mean square
FMS : Factor mean square
Rf : Retention factor
EFH : Efonidipine hydrochloride

References

[1] Ahmed A. Abdelhafez et al. "Optimization of -carotene production from agro-industrial by-products by *Serratia marcescens* ATCC 27117 using Plackett–Burman design and central composite design". In: *Annals of Agricultural Sciences* 61.1 (2016), pp. 87–96. ISSN: 0570-1783. DOI: 10.1016/j.aoas.2016.01.005. URL: https://www.sciencedirect.com/science/article/pii/S0570178316000178.

[2] Gourav Barman, Atul Kumar, and Puja Khare. "Removal of Congo Red by Carbonized Low-Cost Adsorbents: Process Parameter Optimization Using a Taguchi Experimental Design". In: *Journal of Chemical and Engineering Data* 56.11 (2011), pp. 4102–4108.

[3] S Baroutian et al. "A packed bed membrane reactor for production of biodiesel using activated carbon supported catalyst". In: *Bioresource Technology* 102.2 (2011), pp. 1095–1102.

[4] MA Bezerra et al. "Response surface methodology (RSM) as a tool for optimization in analytical chemistry". In: *Talanta* 76.5 (2008), pp. 965–977.

[5] Naveena BJ et al. "Selection of medium components by Plackett–Burman design for production of L(+) lactic acid by *Lactobacillus amylophilus* GV6 in SSF using wheat bran". In: *Bioresource Technology* 96 (2005), pp. 485–490. DOI: 10.1016/j.biortech.2004.05.020. URL: https://doi.org/10.1016/j.biortech.2004.05.020.

[6] Isaac Duah Boateng and Xiao-Ming Yang. "Composite design and evaluation: A case study". In: *Industrial Crops and Products* 162 (2021), p. 113287. ISSN: 0926-6690. DOI: 10.1016/j.indcrop.2021.113287. URL: https://www.sciencedirect.com/science/article/pii/S0926669021000510.

[7] G. E. P. Box and N. R. Draper. *Empirical Model Building and Response Surfaces*. Hoboken, NJ: Wiley, 1987.

[8] George EP Box and Kenneth B. Wilson. "On the Experimental Attainment of Optimum Conditions". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 13.1 (1951), pp. 1–45.

[9] Suraj R. Chaudhari and Atul A. Shirkhedkar. "Application of Plackett-Burman and central composite designs for screening and optimization of factor influencing the chromatographic conditions of HPTLC method for quantification of efonidipine hydrochloride". In: *Journal of Analytical Science and Technology* 11.1 (Nov. 5, 2020), p. 48. ISSN: 2093-3371. DOI: 10.1186/s40543-020-00246-2. URL: https://doi.org/10.1186/s40543-020-00246-2.

[10] W. E. Deming. *Out of the Crisis*. Cambridge, MA: MIT Center for Advanced Engineering Study, 1982.

[11] Yu Du et al. "Optimization of Extraction or Purification Process of Multiple Components from Natural Products: Entropy Weight Method Combined with Plackett–Burman Design and Central Composite Design". In: *Molecules* 26.18 (2021). ISSN: 1420-3049. DOI: 10.3390/molecules26185572. URL: https://www.mdpi.com/1420-3049/26/18/5572.

[12] R. A. Fisher. *The Design of Experiments*. Edinburgh, Scotland: Oliver , Boyd, 1935.

[13] Abdellatif El-Ghenymy et al. “Optimization of the electro-Fenton and solar photoelectro-Fenton treatments of sulfanilic acid solutions using a pre-pilot flow plant by response surface methodology”. In: *Journal of Hazardous Materials* 221-222 (2012), pp. 288–297.

[14] Jacques Goupy. “What kind of experimental design for finding and checking robustness of analytical methods?” In: *Analytica Chimica Acta* 544 (2005), pp. 184–190.

[15] Donald B. Hibbert. “Experimental Design in Chromatography: A Tutorial Review”. In: *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* 910 (2012), pp. 2–13.

[16] Maryam Bagheri Kasiri, Hamidreza Aleboyeh, and Ali Aleboyeh. “Modeling and Optimization of Heterogeneous Photo-Fenton Process with Response Surface Methodology and Artificial Neural Networks”. In: *Environmental Science & Technology* 42.21 (2008), pp. 7970–7975.

[17] John Lawson. *Design and Analysis of Experiments with R*. USA, 2005.

[18] J.A. van Leeuwen et al. “An expert system for the choice of factors for a ruggedness test in liquid chromatography”. In: *Analytica Chimica Acta* 228 (1990), pp. 145–153.

[19] Douglas C. Montgomery. *Design and Analysis of Experiments*. 4th. New York: Wiley, 1996.

[20] Douglas C. Montgomery. *Design and Analysis of Experiments*. New York, NY, USA: John Wiley & Sons, 2001.

[21] Douglas C. Montgomery. *Design of Experiment*. 5th. John Willy, 2001.

[22] RJ Murphy. “Screening design”. In: (2003), p. 920.

[23] Raymond H Myers. *Response Surface Methodology*. Boston: Allyn and Bacon, 1971.

[24] Raymond H. Myers and Douglas C. Montgomery. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. 2nd. New York: Wiley, 2002.

[25] National Institute of Standards and Technology (NIST). *NIST/SEMATECH e-Handbook of Statistical Methods*. Accessed June 2006. 2006. URL: http://www.itl.nist.gov/div898/handbook.

[26] LS Nelson. “Extreme screening designs”. In: *J Qual Technol* 14 (1982), pp. 99–100.

[27] A. Nijhuis et al. “Strategy for ruggedness tests in chromatographic method validation”. In: *Anal Chim Acta* 391 (1999), pp. 187–202.

[28] T Olmez. “The optimization of Cr(VI) reduction and removal by electrocoagulation using response surface methodology”. In: *Journal of Hazardous Materials* 162.2-3 (2009), pp. 1371–1378.

[29] A. Özer et al. “Biosorption of copper (II) ions on Enteromorpha prolifera: application of response surface methodology (RSM)”. In: *Chemical Engineering Journal* 146 (2009), pp. 377–387.

[30] R.L. Plackett and J.P. Burman. “The design of optimum multifactorial experiments”. In: *Biometrika* 33 (1946), pp. 305–325.

[31] T. Prvan and D.J. Street. “An annotated bibliography of application papers using certain classes of fractional factorial and related designs”. In: *Journal of Statistical Planning and Inference* (2002), pp. 245–269.

[32] PF Stanbury, A Whitaker, and SJ Hall. *Media for industrial fermentations*. Oxford: Pergamon, 1986, pp. 93–122.

[33] Lori Trocine and Charles L. Malone. “Finding Important Independent Variables through Screening Designs: A Comparison of Methods”. In: *Proceedings of the Winter Simulation Conference*. Ed. by Jeffrey A. Joines et al. 2000, pp. 749–754.

[34] Y. Van Heyden, F. Questier, and D.L. Massart. “A ruggedness test strategy for procedure related factors: experimental set-up and interpretation”. In: *Journal of Pharmaceutical and Biomedical Analysis* 17 (1998), pp. 153–168.

[35] Y. Van Heyden et al. “Ruggedness tests on the High Performance Liquid Chromatography assay of the United States Pharmacopeia XXII for tetracycline hydrochloride. A comparison of experimental designs and statistical interpretations”. In: *Analytica Chimica Acta* 312 (1995), pp. 245–262.

[36] J.A. Wass. “Formal experimental design and analysis for immunochemical product development. Part I”. In: *IVD Technology Magazine* (Sept. 1997).

[37] L.L. Yannis. “A Plackett-Burman screening design directs the efficient formulation of multicomponent DRV liposomes”. In: *Journal of Pharmaceutical and Biomedical Analysis* 26 (2001), pp. 255–263.

[38] L Ye et al. “Optimization of inductive angle sensor using response surface methodology and finite element method”. In: *Measurement* 48 (2014), pp. 252–262.

11 Appendix

```r
library(rsm)
library(qcc)
library(daewr)

pbdes126 <- data.frame(c(1,-1,1,-1,1,-1,-1,1,1,-1,-1,1),
                       c(1,-1,-1,1,-1,-1,1,-1,1,-1,1,1),
                       c(-1,-1,1,1,1,-1,1,-1,1,1,-1,-1),
                       c(1,-1,1,-1,-1,1,1,-1,-1,1,-1,1),
                       c(-1,-1,-1,1,-1,1,-1,1,1,1,-1,1),
                       c(-1,-1,1,-1,-1,1,1,1,1,-1,1,-1))
colnames(pbdes126) <-c('WL','ST','VDCM','Vtreth','DevDis','ActofPlt')

parea <-c(4985.45, 6736.95, 5383.96, 5565.23, 5954.12, 9045.56, 4005.6,
    8195.95, 7902.53, 9265.45, 4601.98, 8315.76)
#the response Peak Area of the runs in pb design

parea2 <- c(4526.56, 6512.32, 5264.65, 6865.48, 6795.53, 6656.69,
   6845.12, 5156.98, 5951.18, 4156.69, 6965.74, 6685.48, 5985.56)
#the response Peak Area of the runs in ccf design

order_parea2 <- c(8,11,3,10,4,5,6,7,9,2,13,1,12)

parea2 = parea2[order_parea2] #order the vector same as the runs in the
    experiment

rtnfactor <- c(0.77, 0.58, 0.52, 0.5, 0.46, 0.43, 0.49, 0.48, 0.47,
   0.55, 0.45, 0.46) #the 2nd response Retention Factor

DevDis2 <- c(8,8.5,7.5,8,8,8,8,7.5,7.5,8.5,8.5,8,8) #the variable
   development for ccf design

ST2 <- c(25,20,25,20,20,20,20,15,20,25,15,20,15) #the factor saturation
    time for ccf design

modpbpa <- lm(parea ~ (.), data = pbdes126)#fit linear model-> pb
   design ~ peak area

#Residual plots for Peak Area vs Placket Burman design
par(mfrow = c(3,3)) #creates a grid of 3 rows and 3 columns

plot(parea, modpbpa$residuals,xlab = "Peak Area",ylab = "Residuals")
plot(modpbpa$fitted.values,modpbpa$residuals,xlab = "Fitted Values",
   ylab = "Residuals")
plot(pbdes126$WL,modpbpa$residuals,xlab = "WL",ylab = "Residuals")
```

```r
plot(pbdes126$ST,modpbpa$residuals,xlab = "ST",ylab = "Residuals")
plot(pbdes126$VDCM, modpbpa$residuals,xlab = "VDCM",ylab = "Residuals")
plot(pbdes126$Vtreth,modpbpa$residuals,xlab = "Vtreth",ylab = "
   Residuals")
plot(pbdes126$DevDis,modpbpa$residuals,xlab = "DevDis",ylab = "
   Residuals")
plot(pbdes126$ActofPlt,modpbpa$residuals,xlab = "Actofplt",ylab = "
   Residuals")

#coefficients of the model (effects)
cfspa <- coef(modpbpa)[2:7]
cfspa_standardized <- abs(cfspa)/sd(cfspa)
modpbrf <- lm(rtnfactor ~ (.), data = pbdes126) #fit linear model -> pb
    design ~ retention factor

#Residual plots for Retention factor vs Placket Burman design
par(mfrow = c(3,3))

plot(rtnfactor,modpbrf$residuals,xlab ="rtnfactor",ylab ="Residuals")
plot(modpbrf$fitted.values,modpbrf$residuals,xlab = "Fitted Values",
   ylab = "Residuals")
plot(pbdes126$WL,modpbrf$residuals,xlab = "WL",ylab = "Residuals")
plot(pbdes126$ST,modpbrf$residuals,xlab = "ST",ylab = "Residuals")
plot(pbdes126$VDCM, modpbrf$residuals,xlab = "VDCM",ylab = "Residuals")
plot(pbdes126$Vtreth,modpbrf$residuals,xlab = "Vtreth",ylab = "
   Residuals")
plot(pbdes126$DevDis,modpbrf$residuals,xlab = "DevDis",ylab = "
   Residuals")
plot(pbdes126$ActofPlt,modpbrf$residuals,xlab="Actofplt",ylab ="
   Residuals")

cfsrf <- coef(modpbrf)[2:7]
cfsrf_standardized <- abs(cfsrf)/sd(cfsrf)

summary(modpbpa)
summary(modpbrf)
anova(modpbpa)
anova(modpbrf)

par(mfrow=c(1,2))
pareto.chart(cfspa_standardized,plot = TRUE, main = "Standardized
   effect for Peak Area")
pareto.chart(cfsrf_standardized,plot = TRUE, main = "Standardized
   effect for Retention Factor")

```

```r
ccfdes <- ccd(2,n0 = c(4,1), alpha = "faces",randomize = FALSE) #define composite design model..face centered used in the article (alpha = "faces")
ccfd <- as.coded.data(ccfdes, x1 ~ (DevDis2 - 8)/0.5, x2 ~ (ST2 - 20)/5)
modquad <- rsm(parea2 ~ SO(x1,x2), data = ccfd) #define quadratic model (SO->second order)

summary(modquad)

#residual plots
#Residuals are expected to be randomly distributed
#Randomness is an assumption for the linear model

par(mfrow = c(2,2)) #creates a grid of 2 rows and 2 columns
plot(parea2, modquad$residuals)
plot(modquad$fitted.values,modquad$residuals)
plot(ccfd$x1,modquad$residuals)
plot(ccfd$x2,modquad$residuals)

#contour plot and perspective plot

par(mfrow = c(1,1))
par(cex.lab = 1.3) #increase the size of the axis labels
par(mgp = c(3, 2, 1))

contour(modquad, ~ x1+x2,
        image = TRUE,
        xlabs = c("Development Distance","Saturation Time")) #define contour plot

points(ccfd$x1,ccfd$x2)
par(cex.lab = 0.7)
persp(modquad, x1~x2,
      ticktype = "simple",
      col = terrain.colors(50),
      contours = "colors",
      zlab = "Peak Area",
      xlabs = c("Development Distance","Saturation Time")) #define perspective plot (3D plot)

peak_point <- data.frame(x1 = 1,x2 = -0.9134395)
predict(modquad,peak_point) #prediction of the value of peak area at maximum point
```

Listing 1: R Code