VOLUME 2, ISSUE 2

Journal of ongoing Chemical Research

EXPERIMENTAL DESIGN METHODS WITH PRACTICAL APPLICATIONS

Oğuz Akpolat

December 19, 2016

Experimental design term usually refers to a two-stage modeling. The first of all named as working strategy is the establishment of the experiment execution model in which the operating range of the experimental factors called as working parameters shown the simultaneous changes including their interactions with each other on the reaction. Second step is to determine a linear or nonlinear mathematical model for response surface function to define the relationship between the factors and the experimental result. Predetermination of the experimental factor levels for the experimental works in the laboratory is the most important stage for the creation of experimental strategies. The linear or non-linear mathematical models chosen for the response surface function to define the relationship between the factors and the experimental results may also be as similar complex structure as the models planned for the experiment execution in the working strategy determined.

EXPERIMENTAL DESIGN METHODS WITH PRACTICAL APPLICATIONS

Oğuz Akpolat

Muğla Sıtkı Koçman University, Faculty of Science, Chemistry Department, Muğla-48000 Turkey

Corresponding author's name: Oğuz Akpolat

Email: oakpolat@mu.edu.tr

Phone: +90-252-2111589

ABSTRACT

Experimental design term usually refers to a two-stage modeling. The first of all named as working strategy is the establishment of the experiment execution model in which the operating range of the experimental factors called as working parameters shown the simultaneous changes including their interactions with each other on the reaction. Second step is to determine a linear or nonlinear mathematical model for response surface function to define the relationship between the factors and the experimental result. Predetermination of the experimental factor levels for the experimental works in the laboratory is the most important stage for the creation of experimental strategies. The linear or non-linear mathematical models chosen for the response surface function to define the relationship between the factors and the experimental results may also be as similar complex structure as the models planned for the experiment execution in the working strategy determined. In this case, it is clear that advanced optimization techniques for the mathematical analysis of the response surface function would need for optimization of the working factors. Finally it should be noted that the understanding ability of the experimental design could be improved only by examining laboratory works performed by a chemical transformation as one of its practical applications. In this study prepared in that context experimental design methods would be explained both theoretically in general and practically with examples using alternative soft-wares for some experimental design applications,

Keywords: Experimental Design, Enzyme, Full Factorial Design, Immobilization, Optimization, Reaction, Response Surface Function.

INTRODUCTION

Experiment is very important to explain, to understand and to identify a system or a process. At the same time experiment is also to observe the outputs of a system or a process when its inputs are changed and to analyze its behavior. No matter is in which area of experimental studies were done, experimental works could be carried out more efficiently not only randomly but also designing of them in advance and evaluating of their results statistically, thereby experimental resources and time were not consumed in vain. An experimental design is generally to determine which parameters affect experimental results previously and to fictionalize an execution process of the experimental study. As to statistically evaluation basic approach for the comparison of two factors to determine their effects on the results is to compare the averages of the results. For the comparison of two different methods or more is it used comparing of the deviations, which means analysis of variances (ANOVA). However, if there would be numerous factors and interactions among these factors evaluation of test results might become much more complex or even impossible statistically from time to time. In this case, a design strategy to be selected according to the aim of the experiment should be planned formerly and various methods were developed for these type of strategies in the literature. As scan designs could be determine important factors affecting the results of the experimental studies the best ones of these factors could be find out by optimization. Finally experimental results are simulated for different working conditions by modeling of the relation between the factors and the experimental result. From the literature studies it is understood that there would be some design options such as full factorial design, fractional factorial design and central composite design used for experimental design strategies [1].

As called design it comes to mind first that it is imagined an object such as a building or a machine how to be formed or that it is planned a journey or a fiction step by step in details. Expressing of all planned details for a fiction in a mathematical equation is defined as modeling of that but a prototype is an animation of the fiction with all details in a hard body, if that animation is taking shape on a computer screen it is a simulation. A design of an experiment is generally to fictionalize an execution of an experimental process and the term of experimental design strategy usually refers to a two-stage modeling. The first of all named as working strategy is the determination of the experimental execution model in which the operating range of the experimental factors called as working parameters was shown their simultaneous interactions with each other. The other one is that a linear or a non-linear mathematical model for response surface function to define the relationship between the factors and the experimental result was chosen. Pre-determination of the experimental factor levels for an experimental work in the laboratory is the most important step for the creation of experimental strategy and experimental strategies might have different complexities because of the fact that many factors would affect the experimental results simultaneously more or less. The primary objective for designing of chemical or biochemical processes or of analysis processes of reaction products or raw materials is to optimize the process factors in another words process parameters or to minimize the cost of the process, so the processes could be modeled assuming that they take place in the optimum parameters. Today, studies in many fields such as chemistry, biochemistry, earth and environmental sciences are done experimentally and all analysis of semi-finished and finished products or raw materials for quality control are carried out in the laboratories. To succeed in these experimental studies and to reach the minimum cost are needed the well-designed test strategies [2].

Considering the work done about experimental design models it is understood that some specific rules were developed for different design strategies, and they were named according to using of the factors to effect the experimental result. In the following parts it will be emphasized these experimental strategies and experimental design methods in detailed and in this context the importance of experimental design were explained in four items [3]:

-Scan (Screening): This type of design is used to identify important factors that affect the results in experimental studies. For example the factors affecting the efficiency of a chemical reaction are concentration of reactants, catalyst concentration, temperature, pH, reaction time, stirring speed, etc. The first question is which factors affect the reaction result and which of them are important and which of them are eliminated, and which ones should be examined in details. The answers of these questions could be found by screening tests and the models used for screening design could be given as "factorial design " and "Plackett-Burman design."

-Optimization: The important factors found out by scan designs, for example like reaction efficiency or chromatographic separation of reaction products are improved by optimization. The most commonly used optimization methods are "central composite design" and "simplex optimization".

-Time Saving: By traditional or classic methods the effects of the factors are determined only one factor changing and the others keeping constant but when much factors are examined both traditional method takes longer time and costs more, and the effects of the factors cannot be followed simultaneously.

A simple example to show the advantages offered by the statistically experimental design and the optimization approach was given here. In the study done by classic method two factors, pH and catalyst concentration, affecting the yield of the reaction were investigated. In this work firstly the pH was changed keeping constant the catalyst concentration of 2.0 mM and it was found that the reaction yield would be optimal for the pH of 3.4 then the catalyst concentration changed keeping constant the pH of 3.4 as the reaction yield was found optimal for the catalyst concentration of 1.6. Maximum factor values affecting the reaction yield were shown on Figure 1 and Figure 2.

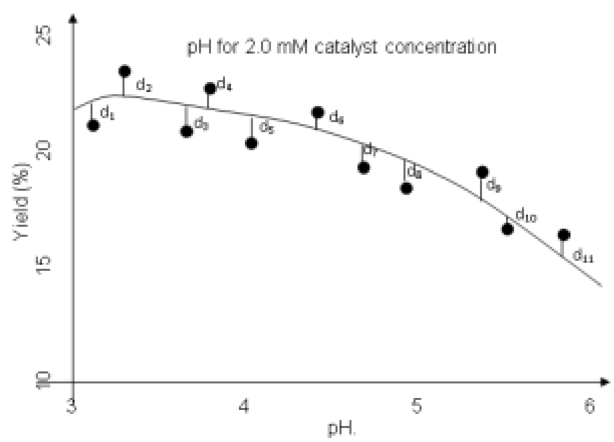


Figure 1: Yield versus pH for constant catalyst concentration of 2.0 mM (optimum)

After that the same work was performed by statistical experimental design method and the results obtained by statistical experimental design was given in Figure 3. Optimum reaction factors obtained for maximum reaction yield by this method are pH=4.4 and catalyst concentration=1.0 mM and it is seen that the factors determined by the experimental design would be quite different than those of the classical method. The reason of this difference is the interactions between the factors simultaneously.

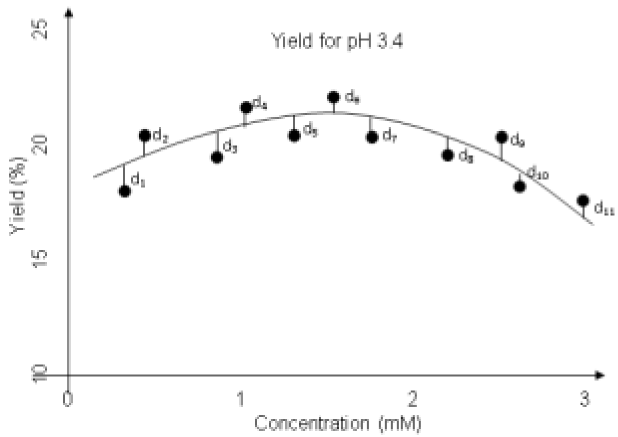
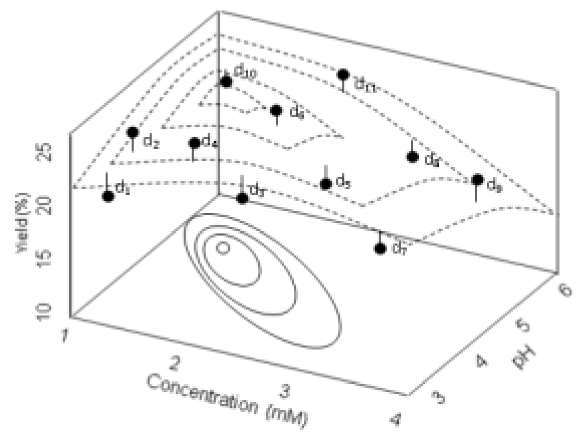


Figure 2: Yield versus catalyst concentration for constant pH of 3.4 (optimum)



Fgure 3: Optimization of the reaction yield depending on the pH and the catalyst concentration

The simplest one of the scan designs is full factorial design and this method has two levels for the factors and it is used to determine which factors are effected on the result of the experiment. For example if a chemical reaction is affected by pH and temperature a full factorial design for the reaction could be done for two-factors in two-levels. Experiment number is calculated by formula 2^k. Here k is the number of factors and the number of levels is 2 coded as -1 and +1. The number of experiments is 4 for 2 factors. The number of experiments is 8 for 3 factors. To set the table of the full factorial design the values of the factors are determined for each factor level. In this sample temperature are chosen as 30ºC and 60ºC for the levels of -1 and +1 and pH as 4 and 6 for the levels of -1 and +1 [4].

Table 1: Full factorial design with two factors and two levels

Experiment number	**1**	**2**	**3**	**4**
x_1 (Factor 1)	-1	+1	-1	+1
x_2 (Factor 2)	-1	-1	+1	+1

Table 1 gives a full factorial design with two-factors in two-levels and the design fiction of Table 1 was pictured on Figure 4 by analogy of a cube where each experiment locates on the corners.

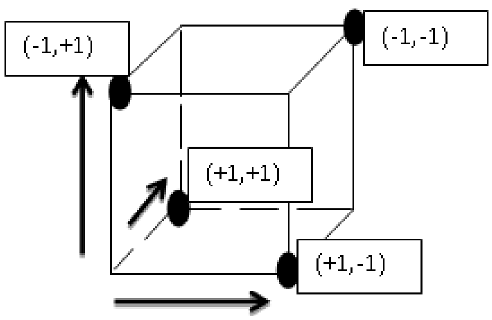


Figure 4: Full factorial design with two-factors in two-levels

It could be understood easily to image a design with more experiment numbers or more dimensions if a new points would be located on the cube surface or a new dimension would be added to the cube. The second step of the experimental design is to model a response surface function to define the relationship between the factors and the experimental result. The response surface function of this reaction depending on the factors of pH and catalyst concentration was plotted to show the reaction yield affected by the factors in Figure 3. What done firstly here was that a mathematical model of this surface would be proposed and a simulation graphic of the surface model would be plotted. For the response surface function can be offered some linear or non-linear models. The next chapter will be on the creation of these models and focuses on the mathematical analysis of these models in detail. Here a linear model not consisting of the parabolic effects of the factors was proposed only as an example.

$$y = b_0x_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 \qquad (1)$$

Calculating of the coefficients in the mathematical model based on an optimization process covers minimization of the sum of the distances between factors calculated from model and experimental values and this optimization are named as least square method. Written optimization function for the response surface is as follows.

$$J = \sqrt{\sum_{i=1}^{n} {d_i}^2} = \sqrt{\sum_{i=1}^{n} (y_{i-\text{experimental}} - y_{i-calculated})2} \rightarrow Minimum \qquad (2)$$

Analysis of this optimization function could be carried out by some numerical techniques.

The term of experimental design is referred generally as determination of the factors or in another words the parameters which affects the experiment results, as finding of the effects of these factors and minimizing of the effects of uncontrollable factors. The factor is the name given to any experimental condition affecting experiment results. If a factor which can be changed or controlled by the researcher it is called as controllable factor, if no changed or selected randomly called as uncontrollable factor. If a factor can be stated as numerically is named as quantitative factor, if no as qualitative factor. The values which a factor presents as dimensionless in a determined intervals are named as its level. Due to many factors can affect the experimental results quite complex experimental designs are encountered. The first stage of experimental design strategy is that the experimental work to plan is classified according to its results expected, determining of the factors to affect the experimental results and modeling of the experimental process. After completion of the experimental work the statistical evaluation of the experimental data should be done, therefore a mathematical model should be chosen for the response surface function and optimized the factors by solution of the response surface function. Many execution process models to use in various work areas like chemistry, biochemistry, agriculture and environment were proposed for some design strategies. In this study some basic experimental design models are focused on but mathematically analysis of these models by using computer programs was left to next chapters. Here the experimental design process models selected for screening designs are full factorial design cases and it was focused on the software used in need of coding like MATLAB and in need of without coding like Design Expert [5, 6].

RESULT & DISCUSSION

In this study some basic experimental design models are focused on but mathematically analysis of these models by using computer programs was left to next chapters. Here the experimental design process models selected for screening designs are full factorial design cases and it was focused on the software MATLAB. As the practical application a full factorial design with three factors and two-levels was conducted on the reaction set given in the following reaction for the optimization of the reaction parameters. Considering of three factors which are the amount of catalyst (%w), the reaction temperature (°C) and the reaction time (min) would be affected on the yield of the reaction (%), the working limits of the factors and the experimental design table were prepared according to a three-factors and two-levels full factorial design strategy. As proposing of a response surface function with a linear model optimization of the reaction calculate the factors. The coefficients of the response surface function are solved by the linear equation set and thus the reaction factors would be described for all working conditions and optimized. Selected response surface function here reflects only the linear relationship between the factors. Examined the calculated coefficients of the response surface function given in Table 4 and the simulations shown on the Figure 6, Figure 7 and Figure 8, it could be seen the effects of the factors on the reaction yield clearly. The biggest effect of the reaction factors on the reaction yield is amount of catalyst, and the effect of reaction temperature and the effect of reaction time on the reaction yield follow it in order.

EXPERIMENTAL

In this section a full factorial design with three factors and two-levels was conducted on the reaction set given in the following reaction for the optimization of the reaction parameters. Considering of three factors which are the amount of catalyst (%w), the reaction temperature (°C) and the reaction time (min) would be affected on the yield of the reaction (%), the working limits of the factors and the experimental design table prepared according to a three-factors and two-levels full factorial design strategy were shown on Table 2 and Table 3. As proposing of a response surface function with a linear model optimization of the reaction calculate the factors. For the solution it is written first the linear model of the response surface function as given in equation 3. Then using of the data of Table 2 and Table 3 the linear equation set with eight parameters of the function is formed. The coefficients of the response surface function are solved by the linear equation set and thus the reaction factors would be described for all working conditions and optimized. Selected response surface function here reflects only the linear relationship between the factors.

$$y = b_0x_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 \quad \textbf{(3)}$$

Calculating of the coefficients in the mathematical model based on an optimization process covers minimization of the sum of the distances between factors calculated from the model and those of experimental. This optimization are called as least square method. Written optimization function to minimize the distance is as follows.

$$J = \sqrt{\sum_{i=1}^{n}(y_{i-\text{experimental}} - y_{i-calculated})2} \rightarrow Minimum \quad \textbf{(4)}$$

Table 2: The boundaries of experimental study by full factorial design (hab_04_chapt_04_algorx _prg_04_04_exp_are.xls: File used for MATLAB program)

Factors	**(-1) level**	**(+1) Level**
Amount of catalyst (%w)	0.1	0.3
Reaction temperature (°C)	60	80
Reaction time (min)	20	40

Table 3: The interaction of factors and responses in the full factorial design model (hab_04_chapt_04_algorx_prg_04_04_exp_des.xls: File used for MATLAB program)

Experiment	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	Y (Yield)
1	+1	-1	-1	-1	+1	+1	+1	-1	73
2	+1	+1	-1	-1	-1	-1	+1	+1	71
3	+1	-1	+1	-1	-1	+1	-1	+1	79
4	+1	+1	+1	-1	+1	-1	-1	-1	82
5	+1	-1	-1	+1	+1	-1	-1	+1	78
6	+1	+1	-1	+1	-1	+1	-1	-1	89
7	+1	-1	+1	+1	-1	-1	+1	-1	83
8	+1	+1	+1	+1	+1	+1	+1	+1	93

For the solution taking of the partial derivatives of the coefficients in the optimization function, that means the slops of the tangents on related points of the function, are equaled to zero.

$$\sum_{i=1}^{n} \frac{\partial J}{\partial b_i} |_{b_{..i-1,i+1,...}} = 0 \qquad \textbf{(5)}$$

Here the cluster of b constants defines the response surface function. The linearly dependent equation set is written as follows,

$$y = bX \qquad \textbf{(6)}$$

and mathematically solution of this set was shown on equation 7.

$$b = (X'X)^{-1} X'y \qquad \textbf{(7)}$$

The MATLAB programs prepared for the optimization of the factors by full factorial design and simulation of the response surface function were coded below as the experimental data of the reaction were taken instantly excel files for the calculations.

```
% Full Factorial Design

% -----------------------------------

clc, clear all, clf, format long, '% Screening Design

BACCompM_Data_01_01_Den_Bol

=xlsread('Hab_04_Chapt_04_Algorx_Prg_04_04_Exp_Are.xls');

BACCompM_Data_01_01

=xlsread('Hab_04_Chapt_04_Algorx_Prg_04_04_Exp_Des.xls');

XXD=BACCompM_Data_01_01(:,1:1);XX0=BACCompM_Data_01_01(:,2:2);

XX1 =BACCompM_Data_01_01(:,3:3);XX2=BACCompM_Data_01_01(:,4:4);

XX3=BACCompM_Data_01_01(:,5:5); XXB=BACCompM_Data_01_01(:,2:9);

R   =BACCompM_Data_01_01(:,10:10);            nK=length(XXD);

% Fur response surface function

% Y_X1_X2_X3_X4=  +b0*x0+b1*X1+b2*X2+b3*X3                   ...

%                 +b12*X1*X2+b13*X1*X3+b23*X2*X3           …

%                 +b123*X1*X2*X3

y=R; A=XXB; R_R_X1_X2_X3(:,1)=y;  % y=bX; B=(X'*X)^-1*X'*y

%   Calculating of the coefficients of response surface function

'y', R, 'X', A, bb=(((transpose (A)*(A)))^-1)*transpose(A)*R; b=bb, K=b;

% Calculated function values

for r=1:nK
```

```
   X1=A(r,2);X2=A(r,3);X3=A(r,4);

R_R_X1_X2_X3(r,1)   =  K(1)+K(2)*X1+K(3)*X2+K(4)*X3+              ...

                       K(5)*X1*X2+K(6)*X1.*X3+K(7)*X2*X3+     ...

                       K(8)*X1*X2*X3;

end

% R_01_00_X2_X3 For simulation of the experimental data

  X1=0; X2=XX2; X3=XX3;                          (X1 Constant)

figure(1) % Plotting of the experimental data

plot3(X2, X3, R, 'k*'); grid

title('BACCompMGraphic 01 00 X2 X3')

xlabel('X2'); ylabel('X3'); zlabel('R');grid ;  hold on

xx=2;yy=xx; DelX=(xx-(-xx))/(nK-1); % Simulation

DelY=DelX; XX=-xx:DelX:xx; YY=-yy:DelY:yy;

[X2 X3]=meshgrid(XX,YY);

R_X1_X2_X3   =  K(1)+K(2)*X1+K(3)*X2+K(4)*X3+                   ...

                       K(5)*X1.*X2+K(6)*X1.*X3+K(7)*X2.*X3+    ...

                        K(8)*X1.*X2.*X3;

mesh(X2,X3,R_X1_X2_X3); grid

% R_01_X1_00_X3 For simulation of the experimental data

  X1=XX1; X2=0; X3=XX3; ;                          (X2 Constant)
```

```
figure(2) % Plotting of the experimental data

plot3(X1, X3, R, 'k*'); grid

title('BACCompMGraphic 01 X1 00 X3')

xlabel('X1'); ylabel('X3'); zlabel('R');grid ;  hold on % !

xx=2;yy=xx;   DelX=(xx-(-xx))/(nK-1);   DelY=DelX;   XX=-xx:DelX:xx;   YY=-yy:DelY:yy;   % Simulation

[X1 X3]=meshgrid(XX,YY);

 R_X1_X2_X3   =  K(1)+K(2)*X1+K(3)*X2+K(4)*X3+                        ...

                  K(5)*X1.*X2+K(6)*X1.*X3+K(7)*X2.*X3+    ...

                   K(8)*X1.*X2.*X3;

mesh(X1,X3,R_X1_X2_X3); grid

% R_01_00_X2_X3 For simulation of the experimental data

  X1=XX1; X2=XX2; X3=0; ;                              (X3 Constant)

figure(3) % Plotting of the experimental data

plot3(X1, X2, R, 'k*'); grid

title('BACCompMGraphic 01 X1 X2 00')

xlabel('X1'); ylabel('X2'); zlabel('R');grid ;  hold on % !

xx=2;yy=xx;   DelX=(xx-(-xx))/(nK-1);   DelY=DelX;   XX=-xx:DelX:xx;   YY=-yy:DelY:yy;   % Simulation

[X1 X2]=meshgrid(XX,YY);
```

```
R_X1_X2_X3  =  K(1)+K(2)*X1+K(3)*X2+K(4)*X3+          ...
               K(5)*X1.*X2+K(6)*X1.*X3+K(7)*X2.*X3+   ...
               K(8)*X1.*X2.*X3;          mesh(X1,X2,R_X1_X2_X3); grid

figure(4) % Correlation of response surface function

plot(R, R_R_X1_X2_X3, 'k*');

title('BACCompMGraphic 01 X1 X2 X3')

xlabel('Experimental data'); ylabel('Calculated value');grid ;  hold on % !

% ------------------------------------------------------------
```

The steps of the MATLAB program written for the optimization algorithm of full factorial experiment scan design consist of calculation of the response surface function coefficients, simulation of the response surface function by using these coefficients and statistically correlation of the optimized response surface function to the experimental data. Calculated coefficients of the response surface function were given in Table 4.

Table 4: Calculated coefficients of the response surface function

b0	**b**1	**b**2	**b**3	**b**12	**b**13	**b**23	**b**123
81.0000	2.7500	3.2500	4.7500	0.5000	2.5000	-1.0000	-0.7500

Simulations of the response surface were plotted on Figure 5, Figure 6 and Figure 7 and the correlation of the response surface function was given on Figure 8

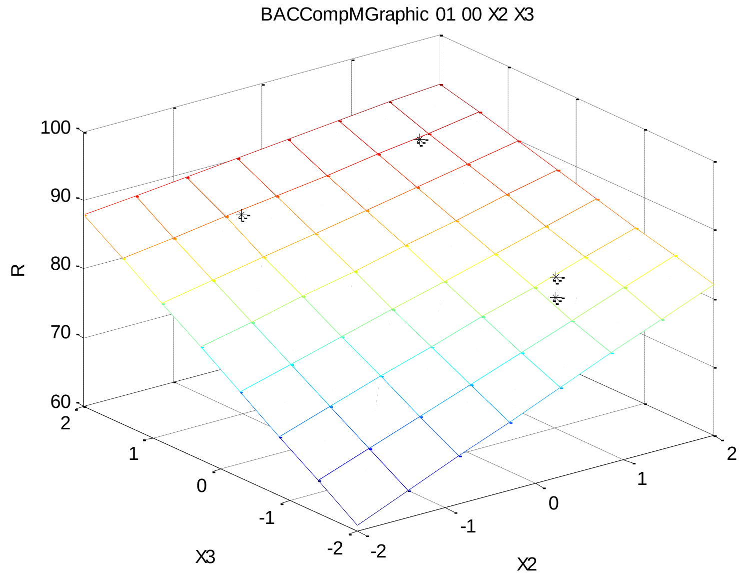


Figure 5: The simulation of the response surface function plotted keeping constant of one parameter (x1: amount of catalyst) for the full factorial design

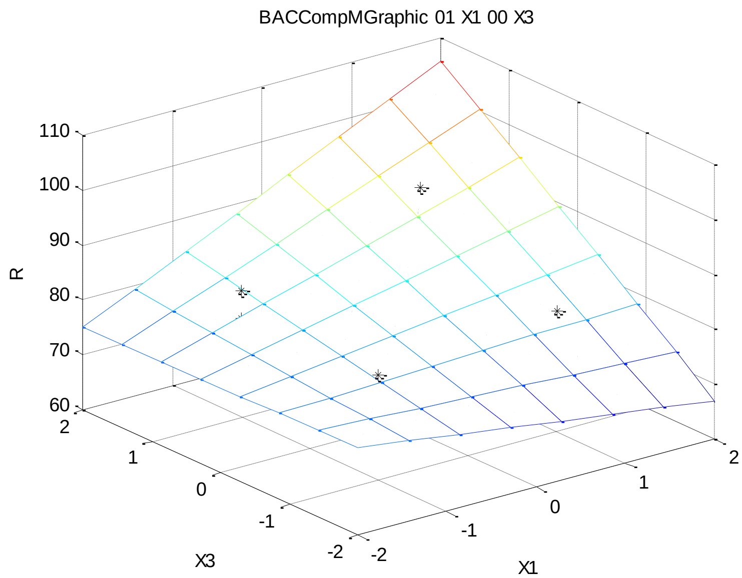


Figure 6: The simulation of the response surface function plotted keeping constant of one parameter (x2: reaction temperature) for the full factorial design

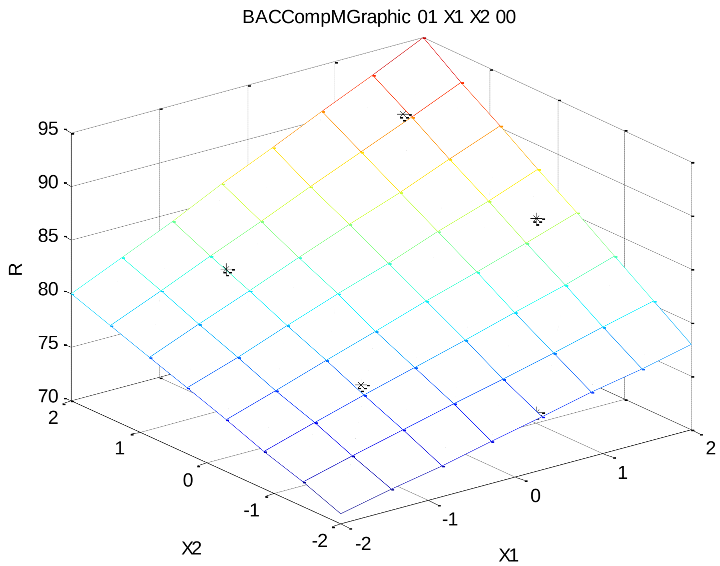


Figure 7: The simulation of the response surface function plotted keeping constant of one parameter (x3: reaction time) for the full factorial design

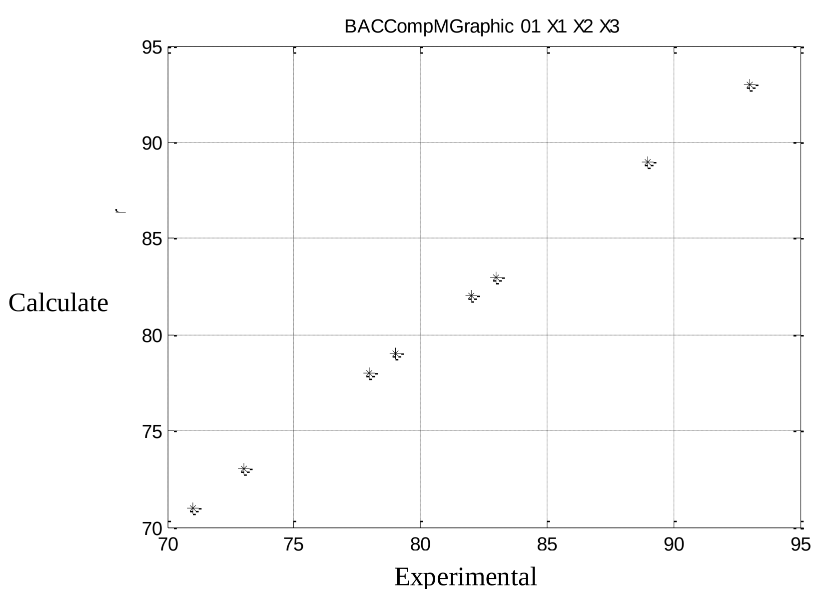


Figure 8: Correlation of the yields as response

ACKNOWLEDGEMENT:

We are thankful to the Chemistry department of Muğla Sıtkı Koçman University, Muğla, Turkey.

REFERENCES

[1] Özevci, G., İpteş, N., Akpolat, O., (2013), Statistical design of experiments, Central Anatolia Region 1 Agriculture and Food Congress, TARGID2013, 2-4 Oct.

[2] Brereton, R. G., (2003), Chemometrics: Data Analysis for thje Laboratory and Chemical Plant, John Wiley & Sons Ltd.

[3] Özdemir, D., (2012), Kemometri Temel Eğitimi Ders Notları, Kimya Akademisi, Antalya.

[4] Ertaş, H., Ertaş, F. N., (2001), Analitik Kimyada Veri İşleme ve Değerlendirme, Ders Notları, Ege Üniv. Fen Fak. Kim. Böl. Analitik Kim. A. B. D.

[5] http://www.mathworks.com/products/matlab/, (2012)

[6] http://www.designexpert.com, (2012), Design Expert Tutorial, General One Factor Tutorial-Part 1. 15 April.