The use of first-order plans to determine the degree of influence of external factors on the course of a chemical reaction

M O Fabrichkina¹, A K Shatrov², V V Baranov² and V I Litovchenko³

¹ Siberian Federal University, Krasnoyarsk, Russia
² JSC Academician M.F. Reshetnev «ISS» Zheleznogorsk, Krasnoyarsk region, Russia
³ Reshetnev Siberian University of Science and Technology, Krasnoyarsk, Russia

E-mail: fabrichkina@yandex.ru

Abstract. To estimate the significance of the influence of external factors on the course of a chemical reaction, experimental design methods can be used. This article discusses the possibility of applying first-order plans for this task. A linear model and a first-order plan were constructed and the significance of factors for a specific task was evaluated.

1. Introduction
When observing the course of chemical reactions, the question about significance of certain factors arises. Scientists use a great deal of different methods to solve a similar problem. This can be both traditionally applied methods used in the chemical industry, and modern methods of data analysis. To solve problems of this kind, you can use experimental design. The theory of experimental design is used to study and mathematical description of processes and phenomena by constructing their mathematical models, based on the description of this methodology, we can talk about the possibility of mathematical description of the chemical process by experimental design methods.

Experimental design is a procedure for selecting the number and conditions of experiments necessary and sufficient to solve the problem with the required accuracy [1]. The tasks for which experimental design can be used are extremely diverse. They include: the search for optimal conditions, the construction of interpolation formulas, the selection of essential factors, the estimation and refinement of the constants of theoretical models, the selection of the most acceptable hypotheses about the mechanism of phenomena, the study of composition - property diagrams, etc. [2].

The aim of the work is to demonstrate the possibility of using experimental design to determine the influence of external factors on the course of a chemical reaction.

2. Construction and research of first-order plans
Let the task to study the influence of certain factors and their interactions on a certain object be set. It is required to conduct experiments according to a preconceived plan, which allows you to implement all possible combinations of factor levels. In the theory of experimental design, an experiment is understood as the total of all possible operations that can be performed on the object of study in order to obtain information about its properties.
In the current article, the construction of a full factorial experiment will be considered. Full factorial experiment is an experiment in which a combination of all possible levels of factors is considered. Full factorial experiment is of type $N = p^m$, this type is obtained if the number of factors is $m$, and the number of levels of each factor is $p$. When constructing a linear model of the object, Full factorial experiment of the type $N = 2^m$ is used.

When planning an experiment, it is impossible to use factors if their original scales are in different units of measurement, all factors should be reduced to a single scale into dimensionless coordinates that take values $±1$. The table contains the values of the experimental conditions. In each table, the rows are different experiments, and the columns are the values of the factors. Such tables are called experiment planning matrices. For the convenience of mathematical calculations, a column of a dummy variable is often introduced into the planning matrix, which in all experiments takes the value $+1$ [3].

To demonstrate the possibility of using first-order plans, a chemical reaction was simulated under the influence of various factors, such as catalyst charge, temperature, pressure, and substance concentration. Data on the objects were taken from the site Statistical Engineering Devesion [7].

**Table 1.** Variable factors and their variation levels for the first object.

| Factor | Dimension | Low level of variation | Ground level of variation | Upper level of variation |
|--------|-----------|-------------------------|---------------------------|--------------------------|
| $x_1$  | Pounds    | 10                      | 12.5                      | 15                       |
| $x_2$  | С$^0$     | 220                     | 230                       | 240                      |
| $x_3$  | PSI       | 50                      | 65                        | 80                       |
| $x_4$  | %         | 10                      | 11                        | 12                       |
| $y$    | Not measured | -                        | -                        | -                        |

After the experiment, it is necessary to process its results and build a model, these actions are carried out in the following steps:
1. coding of factors - the transition from natural values to dimensionless;
2. expanding the planning matrix by taking into account the pair interactions of factors;
3. verification of experiment reproducibility;
4. the calculation of the coefficients of the equation of the model;
5. checking the calculated model coefficients for significance;
6. checking the obtained model equation for significance;
7. interpretation of the obtained model;
8. checking of model performance.

To simulate the course of the reaction, we carry out the first-order full factorial experiment. The linear model equation has the form:

$$\hat{y} = \beta_0 + \sum_i \beta_i x_i + \sum_{i<j} \beta_{ij} x_i x_j,$$

where $\beta_i$ – regression coefficient; $x_i$ – plan factors; $x_i x_j$ – interaction factors.

The authors developed a program in the high-level programming language C #, which allows, based on the results of the experiment, to construct a first-order plan for the selected model. Using the program developed by the authors, we will construct a full factorial experiment plan for the selected object. The result is shown in Figure 1.
Next, you need to calculate the necessary values. For the calculation, you need to rely on the algorithms for constructing and checking the adequacy of the model.

When implementing an orthogonal plan, there are situations when the fact remains unknown whether the dispersions of the outputs are the same at each point of the plans, in this situation, it is necessary for each point of the plan to make several additional measurements of the output. Based on the results of these additional studies, an estimate of the dispersion at each point is found and the hypothesis of homogeneity of dispersions is tested.

The estimation of dispersion at each selected point is found by the following formula

$$\hat{\sigma}_j^2 = \frac{1}{k-1} \sum_{i=1}^k (y_{jk} - \bar{y}_j)^2,$$

where $y_j$ – value of object output in $j$ point of plan; $\bar{y}_j$ – the average value of object output in $j$ point of plan; $k$ – the number of measurements of object output at each point of the plan.

There are many different ways to verify the uniformity of dispersions. The most commonly used statistics are Fisher and Kochren statistics. In the current task, to test the main hypothesis, we will use Kochen's criteria.

The hypothesis of homogeneity of a number of dispersions is accepted if the experimental value of the Kochen statistics $G_{max}$ does not exceed the threshold value for a given significance level $\alpha = 0.05$ and the number of degrees of freedom $v = k - 1 = 2$. At this object, we obtain the following values: $G_{max} = 0.212$, $G_{threshold} = 0.32$. Since $G_{max} < G_{threshold}$ the hypothesis of homogeneity of the dispersions series is accepted.

The next step is to determine the coefficients of the model. Coefficients assessments are found using the least — square technique [4]. Due to the orthogonality of the full factorial experiment plan, the following formulas can be applied to calculate the parameters $\beta$:

$$\beta_0 = \frac{\sum \bar{y}_i}{2^n}, \quad \beta_i = \frac{\sum x_i \bar{y}_i}{2^n}, \quad \beta_{ij} = \frac{\sum x_{ij} \bar{y}_i}{2^n},$$

where $\sum \bar{y}_i$ – sum of average values of measurements; $N$ – number of plan points.

The results of calculating the coefficients $\beta$ are shown in table 2.

There are several main problems when planning an experiment, one of the most significant is the choice of factor values during the experiment. This problem is of fundamental importance in the theory of experimental design, since the accuracy of assessing the degree of influence of individual factors in a multifactor experiment on the target parameter, as well as the duration of the experiment, depends on the method of varying the values of the factors.
Table 2. The results of the calculation of the coefficients $\beta$.

|   | 1   | 2   | 3   | 4   | 12  | 13  | 14  |
|---|-----|-----|-----|-----|-----|-----|-----|
| 0 | 72.3| 3.83| -11.732| 1.128| 2.822| 0.12| 0.367| 0.179|
| 1 | 23  | 24  | 34  | 123 | 124 | 134 | 234 | 1234|
| 2 | -0.48| 2.348| 0.001| 0.54| 0.706| 0.486| 0.569| -0.068|

By mathematical processing of the results of the experiment, it is necessary to extract information from the mixed information on the influence of individual factors, and this, as follows from the theory of the experiment, is not possible with any method of compiling a planning matrix [5].

In order to evaluate which of the factors and of the interactions are significant, it is necessary to check the significance of the model coefficients.

Checking the significance of the coefficients is to test the hypothesis of equality of the coefficients to zero. To test the hypothesis, we use Student’s statistics.

The dispersions of the model coefficients are calculated as follows

$$\sigma^2_{\beta_0} = \frac{\sigma^2_Y}{k \times 2^n}, \sigma^2_{\beta_i} = \frac{\sigma^2_Y}{k \times 2^n}, \sigma^2_{\beta_{ij}} = \frac{\sigma^2_Y}{k \times 2^n}$$

(4)

The results of the calculation of Student’s statistics for each coefficient of the model and the decision on its significance are given in table 3. Number of degrees of freedom $v = N(k - 1) = 32$, significance level $\alpha = 0.05$. Threshold value for this object $t_{\text{threshold}} = 2.038$.

Table 3. The results of the calculation of Student’s statistic.

| Index | 1   | 2   | 3   | 4   | 12  | 13  | 14  |
|-------|-----|-----|-----|-----|-----|-----|-----|
| t     | 14.73| 45.12| 4.39| 10.85| 0.46| 1.41| 0.69|
| Result | Signif.| Signif.| Signif.| Signif.| Not signif.| Not signif.| Not signif.|
| 23  | 24  | 34  | 123 | 124 | 134 | 234 | 1234|
| 1.85| 9.03| 0.005| 2.08| 2.72| 1.87| 2.19| 0.26|
| Not signif. | Signif. | Not signif. | Signif. | Signif. | Not signif. | Signif. | Not signif.|

The obtained equation for the output of the object model is as follows:

$$\hat{y} = 72.3 + 3.83x_1 - 11.732x_2 + 1.128x_3 + 2.822x_4 + 2.348x_2x_4 + 0.54x_1x_2x_3 + 0.706x_1x_2x_4 + 0.569x_2x_3x_4$$

(5)

Having received a decision on the significance or non-significance of the coefficients, we need to check the adequacy of the model we have chosen, for this we need to first calculate the model outputs at each point in the plan, and then calculate the dispersion of the model’s adequacy.

The values of the outputs of the model investigated in the plan are presented in table 4.

Tables 4. Model Outputs.

|   | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   |
|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 72.51| 62.358| 87.65| 82.482| 68.037| 60.045| 87.61| 80.283|
| 2 | 59.622| 52.295| 89.25| 81.258| 57.42| 52.26| 86.94| 76.785|
| 3 | 9   | 10  | 11  | 12  | 13  | 14  | 15  | 16  |
The dispersion of adequacy has the form
\[ \hat{\sigma}_{ad}^2 = \frac{1}{N-m-1} \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2. \]  \hspace{1cm} (6)

A check on the adequacy of the obtained model equation is carried out using Fisher statistics. The threshold value of the Fisher criterion for a given level of significance \( \alpha = 0.05 \) and the corresponding values of degrees of freedom \( \nu_1 = N - m - 1 = 16 - 8 - 1 = 7, \nu_2 = N(k - 1) = 32 \) for this model is \( F_{table} = 2.31 \), criterion value \( F = 1.38 \). Since \( F < F_{table} \), the obtained model equation can be considered adequate to the described object.

From the obtained equation of the model it follows that the factor \( x_2 \) - temperature, has the greatest influence on the output value, because it has the largest coefficient in absolute value. After it, according to the force of influence on the course of the chemical reaction, there is \( x_1 \) - the charge of the catalyst, \( x_4 \) - the concentration of the substance, \( x_2 \) - temperature and concentration level, \( x_3 \) - pressure level.

Since coefficients of \( x_1, x_3, x_4, x_2 x_4, x_1 x_2 x_3, x_1 x_2 x_4, x_2 x_3 x_4 \) are positive, then with an increase of these factors, the output value increases. Coefficient of factor \( x_2 \) is negative. This means that with an increase of this factor, the output value will decrease.

In natural variables, the equation of the model has the following form:
\[ \hat{y} = -904.904 + 94.5072u_1 + 4.009u_2 + 13.81u_3 + 86.72u_4 - 0.404u_1 u_2 - 0.33u_1 u_3 - 6.5u_1 u_4 - 0.06u_2 u_3 - 0.365u_2 u_4 - 0.873u_3 u_4 + 0.001u_1 u_2 u_3 + 0.028u_1 u_2 u_4 + 0.004u_2 u_3 u_4 \]  \hspace{1cm} (7)

The adequacy of the model does not guarantee its suitability for practical use in forecasting tasks. The main indicator of model quality is the coefficient of determination \( R^2 \). The obtained model can be considered functional (suitable for practical use for prediction purposes) if this model has a determination coefficient of \( R^2 \geq 0.75 \). The coefficient of determination is calculated by the following formula [6]:
\[ R^2 = \frac{\sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}. \]  \hspace{1cm} (8)

where \( \bar{y} = \sum_{i=1}^{N} y_i \).

In this case, to calculate the coefficient of determination, the following formula can be used:
\[ R^2 = 1 - \frac{k(N-m-1)\hat{\sigma}_{ad}^2 + N(k-1)\hat{\sigma}_y^2}{k \sum_{i=1}^{N} (y_i - \bar{y})^2 + N(k-1)\hat{\sigma}_y^2}. \]  \hspace{1cm} (9)

The obtained value of the coefficient of determination, equal to 0.99, indicates that the obtained model is functional and can be used for predicting the values of the object output.

3. Conclusion
In the course of studying and conducting an experiment on modeling the course of a chemical reaction under the influence of external factors, we were able to estimate the degree of influence of external factors using first-order plans.

Using this methodology for constructing an experimental design, you can evaluate any real object, evaluate the significance of factors, evaluate the model chosen to describe a particular object, and, most importantly, evaluate the values of the inputs and outputs of the model.
References
[1] Adler Yu P, Markova Ye V and Granovsky Yu V 1976 Planning an experiment to find optimal conditions 280
[2] Khamkhanov K M 2001 Experiment planning basics 50
[3] Ruban A I 2004 Data analysis methods 319
[4] Linnik Yu V 1958 Least squares method and the basis of the mathematical and statistical theory of observation processing 336
[5] Makarichev Yu A, Ivannikov Yu N 2016 Methods of experiment planning and data processing 131
[6] Fyorster E and Ryone B 1983 Methods of correlation and regression analysis 304