Geometric modeling and optimization of multidimensional data in Radischev integrated drawing

E V Konopatskiy¹, A A Bezditnyi²

¹Donbas National Academy of Civil Engineering and Architecture, Ukraine
²Sevastopol branch of «Plekhanov Russian University of Economics», Russia

Abstract. The article deals with the issues of systematization and analysis of multidimensional data by presenting them in Radishchev’s integrated drawing. It is depicted in the form of arcs of algebraic curves passing through predetermined points. The principal approach to optimization of multidimensional data is that on the curve corresponding to the response function, the optimal value is selected and the values of the influence factors are fixed. The combined interaction of these factors provides the optimal value of the response function.

1. Introduction

An important task is to process the collected experimental and statistical information for systematization and analysis of multifactorial processes (or phenomena). Usually, this problem is solved by modeling the initial experimental and statistical data, which allows obtaining an analytical description of the influence of factors dependence on the response function. In the future, this will optimize the process under study. Any multi-factor process can be represented as a multi-parameter geometric object of multidimensional space. However, it should be taken into account the complexity of the visual perception of multidimensional space, which leads to the need to use not visual, but logical clarity, which is based on methods of generalization and analogy. One of the possible ways to reduce the dimension of space is the application of projection algorithms, the model of which is represented by a set of projections on the projection planes of the global coordinate system [1]. The problems of geometric modeling of processes and phenomena using time series are solved in a similar way. The time series are a sequence of observations, which is ordered in time, although it is possible to order through some other parameters. The main feature that distinguishes time series analysis from other types of statistical analysis is the materiality of the order in which observations are made [2]. For example, in the research [3] the methods of visual representation and interpretation of the received results are used for choosing rational technological modes of functioning, for assessing the degree of implementation of the set requirements to the created products, for clarifying regularities of functioning, and analyzing the influence of factors on systems quality indicators, etc. The research [4] uses the similar approach for solving economic problems of the analysis and forecasting.

It should be noted that for the analytical description and computer representation of such models it is more effective to use not a set of projections on projection planes, but a set of projections on the axis of the global coordinate system. This is so because each axis of the global coordinate system corresponds to a specific influence factor or response function. In this case, the modeling result will be a system of parametric equations. In case of using BN-calculation [5-7], these equations will be of the same type where only the coordinates of the initial points will change. The process of transition from a
2. Application of Radischev integrated drawing for systematization of multidimensional data
To systematize multidimensional data with one of the parameters it is convenient to use the integrated drawing of V.P. Radischev [8]. It can be represented as a set of orthogonal planar projections. Hence, the process model can be represented as a set of curves being consistent with the same parameter. The relationship between the individual curves is carried out by means of inter-projection lines. The analytical description of such curves directly depends on the amount of initial experimental and statistical information. In the case of a small amount of initial data, it is convenient to use interpolation methods (for example, parameterized arcs of algebraic curves passing through predetermined points [9-10]). The parametrized curves allow us to define easily the fixed value of the response function and the corresponding values of the factors influencing the investigated process. With a large number of initial data, it is advisable to apply methods of approximation or interpolation through piecewise functions (for example, convex contours of the first order of smoothness [11-12]).

The principal approach to the optimization of multidimensional data presented in the Radishchev integrated drawing is that the optimal value of the response function is selected on the indicator curve. This value most often coincides with one of the extrema of the function, and the value of the parameter at which it was achieved is fixed. Further, the scientific hypothesis is put forward that the joint interaction of the factors fixed by the obtained specific value of the parameter provides the optimal value of the response function. Substituting the value of the optimal parameter into the equations of each factor, we obtain the optimal values of the factors affecting the response function. The implementation of this approach is possible only when the movement of the current points of all simulated curves is consistent with the same parameter. It should be noted that the proposed scientific hypothesis is fully justified provided that all possible factors affecting the behavior of the response function are taken into account.

The proposed approach is especially effective when the initial experimental and statistical data are not enough to build a full-fledged geometric model.

3. Systematization of multidimensional data using the example of modeling and optimizing the technological process of obtaining tetraethoxysilane
Let’s consider the proposed approach to the systematization of multidimensional data using the example of modeling and optimizing the technological process of obtaining tetraethoxysilane, described in the research [13]. The initial experimental data for the simulation are presented in Table 1. The task is to optimize the values of the input parameters to achieve the maximum content of tetraethoxysilane in the product (without ethyl alcohol).

| Item | The content of the main substance in the four-chloride silicon, % | The content of water in ethyl alcohol, % | The molar ratio of the reactants | Reaction temperature, °C | The content of tetraethoxysilane in a product (without ethyl alcohol) |
|------|---------------------------------------------------------------|----------------------------------------|---------------------------------|--------------------------|---------------------------------------------------------------|
| 1    | 99,0                                                          | 0,5                                    | 1:4,6                           | −5                       | 75                                                            |
| 2    | 99,0                                                          | 0,2                                    | 1:4,4                           | +20                      | 85                                                            |
| 3    | 97,2                                                          | 0,2                                    | 1:4,4                           | +20                      | 84                                                            |
| 4    | 98,0                                                          | 0,5                                    | 1:4,7                           | 0                        | 73                                                            |
| 5    | 98,0                                                          | 0,5                                    | 1:4,7                           | −2                       | 72                                                            |
| 6    | 98,0                                                          | 0,8                                    | 1:4,6                           | −5                       | 75                                                            |
| 7    | 97,2                                                          | 0,2                                    | 1:4,7                           | +25                      | 80                                                            |
| 8    | 98,0                                                          | 0,3                                    | 1:4,7                           | +25                      | 80                                                            |

To work with the given Table 1 we introduce the notation:
\[ C_1 \] – content of the main substance in silicon tetrachloride, %;
\[ C_2 \] – content of water in ethyl alcohol, %;
\[ C_3 \] – the molar ratio of the reacting substances, %;
\[ T \] – reaction temperature, ° C;
\[ P \] – content of tetraethoxysilane in a product, %.

A geometrically four-factor process is a four-parameter hypersurface belonging to a five-dimensional space. It is obvious that the source data presented in Table 1 are not enough to build a full-fledged geometric model and get an accurate picture of the behavior of the model. Therefore, the desired geometric model is represented in the form of a set of five projection planes. In this case, the horizontal axis corresponds to the experiment number, and the vertical axis combines the values of the input parameters, which play the role of influence factors, and the response functions – the content of tetraethoxysilane in the product (without ethyl alcohol).

If the number of experiments equals 8, we will use for modeling the point equation of the arc of the 7th order curve passing through 8 predetermined points:
\[
M = M_1 p_1 + M_2 p_2 + M_3 p_3 + M_4 p_4 + M_5 p_5 + M_6 p_6 + M_7 p_7 + M_8 p_8,
\]
where \( M_i \) – starting 8 points, through which the arc of the 7th order curve passes, corresponding to the data from Table 1;

\[ p_i \] – polynomial functions of the current parameter \( t \in [0,1] \);

\[
\bar{T} = 1 - t - \text{addition of a parameter } t \text{ to } 1;
\]

\[
p_1 = \frac{1}{20} 19^{10} - \frac{222}{180} 19^{9} + \frac{7159}{720} 19^{8} - \frac{42881}{180} 19^{7} + \frac{7159}{20} 19^{6} - \frac{223}{6} 19^{5} + \overline{n}^{6};
\]

\[
p_2 = \frac{49}{10} 19^{5} - \frac{5047}{180} 19^{4} + \frac{78253}{720} 19^{3} - \frac{223097}{360} 19^{2} + \frac{32291}{180} 19^{1} - \frac{49}{6} 19^{0} + \overline{n}^{0};
\]

\[
p_3 = \frac{147}{2} 19^{4} - \frac{25431}{180} 19^{3} + \frac{320117}{180} 19^{2} - \frac{62573}{60} 19^{1} + \frac{6321}{20} 19^{0} + \overline{n}^{0};
\]

\[
p_4 = \frac{147}{3} 19^{3} - \frac{52861}{36} 19^{2} + \frac{78449}{144} 19^{1} - \frac{281995}{36} 19^{0} + \overline{n}^{0};
\]

\[
p_5 = \frac{147}{4} 19^{2} - \frac{63778}{36} 19^{1} + \frac{78253}{60} 19^{0} + \overline{n}^{0};
\]

\[
p_6 = \frac{147}{5} 19^{1} - \frac{63217}{20} 19^{0} + \overline{n}^{0};
\]

\[
p_7 = \frac{147}{6} 19^{0} + \overline{n}^{0};
\]

After completing the each coordinate calculation, we obtain the following system of parametric equations:

\[
C_1 = 4640, 6t^6 - 16830, 34t^5 + 24486, 86t^4 - 18170, 91t^3 + 7167, 65t^2 - 1391, 06t + 96, 18t + 99;
\]

\[
C_2 = 2026, 18t^7 - 6862, 86t^6 + 9169, 15t^5 - 6122, 55t^4 + 2116, 5t^3 - 345, 04t^2 + 18, 42t + 0, 5;
\]

\[
C_3 = -18, 07t^8 + 71, 76t^7 - 112, 31t^6 + 87, 05t^5 - 34, 14t^4 + 6, 01t^3 - 0, 3t + 0, 22;
\]

\[
T = -75164, 64t^7 + 258174, 19t^6 - 351079, 56t^5 + 239066, 24t^4 - 83825, 39t^3 + 13462, 07t^2 - 602, 92t + 0, 5;
\]

\[
P = -18627, 76t^8 + 68138, 38t^7 - 99091, 27t^6 + 72080, 02t^5 - 26712, 55t^4 + 4405, 1t^3 - 186, 92t + 75.
\]
The graphic visualization of the obtained geometric model has the following form (Figure 1).

Figure 1. Graphical visualization of the model of tetraethoxysilane production

It should be noted that for the illustration purposes the vertical scale of 1:100 is used for the influence factors \( C_2 \) and \( C_3 \) (Figure 1).

Figure 1 shows that the maximum content of tetraethoxysilane in the product lies in the interval between the 1st and 2nd experiment. The authors suggest using the methods of mathematical analysis to determine this extreme point, which corresponds to the maximum function. Thus, one need to solve the following equation:

\[
\frac{\partial P}{\partial t} = 0.
\]  

(3)

The solution of equation (3) should be sought when changing the parameter on the interval from 0 to 1. In the specified interval, the desired equation has 4 roots. As a result, the maximum content of tetraethoxysilane in the product \( P = 87.873\% \) is achieved when the value of the parameter makes up \( t = 0.204 \). The combination of values of the input parameters includes the content of the main substance in silicon tetrachloride \( C_1 = 97.618\% \), the content of water in ethanol \( C_2 = 0.038\% \), the molar ratio of the reactants \( C_3 = 1:4.317 \), and the reaction temperature \( T = 28.095^\circ C \).

4. Conclusion
The authors propose an approach to the systematization of multidimensional data, which are represented by a system of projections on the axis of the global coordinate system. For visualizing the results of modeling multidimensional data, Radischev's integrated drawing was used. This approach makes it possible to effectively systematize multidimensional data for subsequent optimization of multifactor processes (or phenomena) with an insufficient amount of initial experimental and statistical information. This was shown using the example of the technological process optimization for obtaining tetraethoxysilane.

In addition, it should be highlighted that the proposed approach can be effectively applied for an objective assessment in solving the problems of determining the degree of influence of each factor on the response function as a whole. This is especially important in the modeling of socio-economic
processes and phenomena and is absolutely necessary for making decisions to improve the performance of the researched system.

References

[1] Chizhik M A, Moskovtsev M N and Monastyrenko D P 2013 Geometric modeling of multifactor processes based on projection algorithms Omsk scientific Bulletin (Omsk OmGTU) No1(117) pp 14-17

[2] Makashina E V 2013 Geometric modeling of time series in multidimensional space Geometry and graphics (Moscow: Infra-M) vol 1 No1 pp 20-21

[3] Yakovenko K S, Volkov V Ya and Prokopets V S 2012 The method of geometric analysis of multifactorial models processes Vestnik SibADI (Omsk: SibADI) Vol 3(25) pp 87-91

[4] Okhotnikova M L 2004 Geometric modeling of problems of analysis and forecasting in economics and algorithms for solving them (M.: MAI) 192 p

[5] Baluba I G 1995 Constructive geometry of manifolds in point calculus (Makeevka: MakISI) 227 p

[6] Baluba I G (2015) Dot calculus (Melitopol: MGPU by B Khmelnitsky) 236 p

[7] Naidysh V M, Baluba I G and Vereshchaga V M 2012 Algebra of BN-calculation Applied geometry and engineering graphics (Kiev: KNUBA) Vol 90 pp 210-215

[8] Radishchev V P 1947 On the application of the geometry of four dimensions to the construction of disequilibrium physicochemical diagrams News of the physico-chemical analysis Sector Vol 15 pp 129–134

[9] Konopatsky E V 2018 Geometric modeling and optimization of multifactor processes and phenomena by the method of multidimensional interpolation Proceedings Of the international scientific conference on physical and technical Informatics CPT2018 (Moscow-Protvino) pp 299-306

[10] Konopatsky E V 2018 Principles of building computer models of multifactor Playback processes and phenomena multi-dimensional interpolation method The collection of materials of II International scientific-practical conference "Software engineering: methods and technology development of informational-computational systems (PIIS-2018)" (Donetsk: DonNTU) pp 277-287

[11] Konopatsky E V, Krysko A A and Bumaga A I 2018 Computational Algorithms for Simulating Single - Dimensional Contours through k Predefined Points Geometry and graphics (Moscow: Infra-M) No4(172) pp 20-32

[12] Konopatsky E V 2018 Principles of modeling of multifactorial Playback processes and phenomenas with a large number of initial data Information technologies in design and production (Moscow: NTTS «Kompas») No4(172) pp 20-25

[13] Vertinskaya N D 2009 Mathematical modeling of multifactor and multiparameter processes in multicomponent systems based on constructive geometry (Irkutsk: IrGTU) 230 p