UDC 519.2 **doi:** https://doi.org/10.20998/2522-9052.2023.3.05

Lev Raskin, Viacheslav Karpenko, Yuriy Ivanchykhin, Dmitro Sokolov

National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine

DIAGNOSIS OF SYSTEMS UNDER CONDITIONS OF SMALL INITIAL DATA SAMPLING

Annotation. Object of the study is to assess systems state in conditions of a small sample of initial data. **Relevance** of the problem is as follows. The functioning of a significant number of real objects takes place under conditions of poorly predicted changes in the values of environmental factors affecting system efficiency. The resulting heterogeneity of the results of objects experimental study and the environment of their functioning leads to reduction in sample size. At the same time, the standard requirements regarding the correspondence of the number of experiments and the number of coefficients of regression equation determining system state are not met. **Purpose of the study** is to develop methods for assessing systems state operating in a changing environment, in conditions of small sample of initial data. Tasks to be solved to achieve the goal: the first is the equivalent transformation of the set of observed initial data forming a passive experiment in aggregate into an active experiment, which corresponds to an orthogonal plan; the second is the construction of a truncated orthogonal representative sub-plan of the general orthogonal plan obtained as a result of solving the first problem**. Research methods**: statistical methods of experimental data processing, regression analysis, method for solving a triaxial boolean assignment problem. **The results obtained**: orthogonal representative subplan of the complete factorial experiment being formed makes it possible to calculate a truncated regression equation containing all the influencing factors and their interactions. Analysis of the coefficients of this equation by known methods makes it possible to cut off its insignificant elements.

Keywords: objects state assessment; small sample of initial data; truncated representative regression equation.

Introduction

The traditional approach to solving real problems of analysis, evaluation of the effectiveness of complex systems and their management is based on finding specific mathematical models of these systems [1, 2]. The purpose of such models is to establish a relationship between the values of the characteristics of these systems and the environment of their functioning, on the one hand, and the value of some selected indicator that uniquely determines the state and efficiency of the system. The mathematical description of the model being searched for in each case is determined by the type of system, the nature and features of the operating environment and is selected individually [3, 4]. At the same time, the relevance and demand of numerous practical tasks have led to the development and widespread use of some special types of models, the mathematical apparatus of which provides an adequate description of the processes of systems functioning, solving problems of system analysis and management [5]. Such models include regression analysis models [6, 7]. These models are formed as follows.

Suppose a set of

$$F_{(t)} = \left(F_1(t), F_2(t), \dots, F_m(t)\right)$$

system and environment parameters is given, presumably affecting the numerical value of the selected $y(F,t)$ system indicator. To describe the relationship between these variables, a relation is introduced

$$\begin{gathered} y(F_{(t)}) = a_0 + a_1F_1(t) + a_2F_2(t) + \dots + a_mF_m(t) + \\ + a_{12}F_1(t)F_2(t) + a_{13}F_1(t)F_2(t) + \dots + \\ + a_{m-1,m}F_{m-1}(t)F_m(t) + \\ + \dots + a_{d,d+1}F_d(t)F_{d+1}(t) + \\ + \dots + a_{1,2,d+m}F_1(t)F_2(t)\dots F_{d+m}(t). \end{gathered} \quad (1)$$

The ratio (1) includes all influencing factors and their interactions of the order not higher $d+m$. It is assumed that all these factors depend on time (for example, the system is slowly aging and the values of all indicators are deteriorating). Simplification of the model occurs if the process of changing the values of factors is such that, without large losses in accuracy, the interval of the system functioning can be divided into subintervals, within which this change can be neglected. Thus, a piecewise constant approximation of the real process of changing the system parameters is introduced [8, 9]. Let's write down the ratio obtained in this case for the special case when the maximum order of the interactions of factors taken into account is equal to three:

$$\begin{gathered} y(F) = a_0 + a_1F_1 + \dots + a_mF_m + a_{12}F_1F_2 + \\ + \dots + a_{m-1,m}F_{m-1}F_m + a_{123}F_1F_2F_3 + \dots + \\ + a_{124}F_1F_2F_4 + a_{m-2,m-1,m}F_{m-2}F_{m-1}F_m. \end{gathered} \quad (2)$$

At the same time, the total number of coefficients to be evaluated is equal to $M = 1 + m + +(m-1)m + +(m-2)(m-1)m$.

It is clear that with an increase in the number of factors, the value of M grows rapidly. For example, for an absolutely real number of factors we $m = 6$ have $M = 15$. Then the required number of experiments N should be on the order of 1000, which is hardly feasible taking into account the requirements for the uniformity of the processed data array. Thus, when solving the mentioned practical problems, there is a clear discrepancy between the number of available experiments and the number of coefficients of the regression equation to be evaluated [10]. This discrepancy practically cannot be leveled by increasing the number of experiments. The only real way to solve the problem is to reduce the number of estimated parameters [10–12].

Thus, the purpose of the study is to develop a method for estimating the coefficients of the response function containing all the influencing factors and their interactions, in conditions of insufficient initial data.

Materials and Methods

Let's consider possible ways to achieve the formulated goal.

A. Evaluation of correlations between the values of the controlled parameters and the values of the indicator that determines the effectiveness of the system. At the same time, the corresponding correlation matrix is calculated and the parameters that have little effect on the main indicator are eliminated.

B. Assessment of the Level of Informativeness of the Monitored Indicators. At the same time, the effectiveness of the approach is determined by the level of differences between the main probabilistic characteristics of these indicators (for example, distribution densities) for different states of the system. The numerical value of this level for a specific indicator determines its informational value, which makes it possible to eliminate insignificant indicators. The main disadvantage of these approaches is the difficulty of choosing and justifying the value of the acceptable threshold.

C. Approximation of the desired function (1) by a simpler function determined, in addition to factors, by interactions whose order does not exceed the specified one. Such a function, for example, for three factors, taking into account only paired interactions, will have the form:

$$y(F_1) = a_0 + a_1F_1 + a_2F_2 + a_3F_3 + \\ +a_{12}F_1F_2 + a_{13}F_1F_3 + a_{23}F_2F_3. \qquad (3)$$

Disadvantage is the difficulty of a priori justification of the “threshold” of truncation in the formation of the desired regression polynomial. A promising direction for solving the formulated problem of reducing the dimensionality of the problem is as follows. Let the F_1, ... , F_m results of *N* experiments be given for a set of factors, that is, the matrix *M* and the vector *Y* are determined:

$$M = \begin{pmatrix} \hat{F}_{11} & \hat{F}_{12} & \dots & \hat{F}_{1N} \\ \hat{F}_{21} & \hat{F}_{22} & \dots & \hat{F}_{2N} \\ - & - & - & - \\ \hat{F}_{m1} & \hat{F}_{m2} & \dots & \hat{F}_{mN} \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{pmatrix},$$

where $\hat{F}_{ij}$ – the value of factor *i* in the *j*-m experiment, $i = 1,2,\cdots,m,\ j = 1,2,\cdots,N,\ y_j$ – values of the main indicator of the system efficiency in the *j*-m experiment. The measured values of the monitored indicators are displayed in the interval [-1,1] according to the formula:

$$F_{ij} = \left(2\hat{F}_{ij} - \hat{F}_{ij\max} - \hat{F}_{ij\min}\right) \Big/ \left(\hat{F}_{ij\max} - \hat{F}_{ij\min}\right); \quad (4)$$

$$\hat{F}_{ij\min} = \min_j \hat{F}_{ij}; \quad \hat{F}_{ij\max} = \max_j \hat{F}_{ij}.$$

As a result of transformation (4), the values of the monitored indicators will be in the range [-1,1], and all

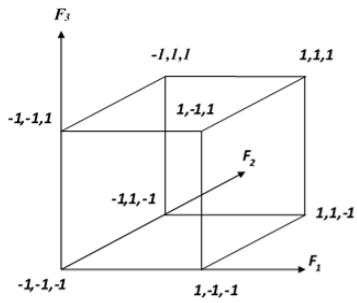

Fig. 1. A three-dimensional hypercube in space (F_1, F_2, F_3)

measurements of these indicators will be displayed by a set of points in the *m*--dimensional factor space lying inside an *m*--dimensional hypercube centered at the origin, whose edge is 2. This hypercube consists of 2^m cubes with a side equal to 1. For example, for *m*=3, the corresponding hypercube in space F_1, F_2, F_3 has the form shown in Fig. 1. This hypercube consists of $2^3 = 8$ cubes (Fig. 2). The cubes (Fig. 2) that make up the original hypercube (Fig. 1) are marked with the values of the factors at the vertices in order to unambiguously determine their position in the coordinate system F_1, F_2, F_3. Now, for each of the cubes, we independently renumber the experiments that got into it. After that, the problem of finding a hyperplane describing the state of experiments inside this cube and determining the dependence of the result of each experiment $y(F)$on the values of factors in this experiment is solved for each cube. This hyperplane is found by the least squares method [13]. Let $M_s$be a matrix of factor values on the set *S* of experiments in the selected cube and $(y_1, y_2, \dots, y_s)$be a set of measured parameter values$y(t)$:

$$M_s = \begin{pmatrix} F_{11} & F_{12} & \dots & F_{1S} \\ F_{21} & F_{22} & \dots & F_{2S} \\ - & - & - & - \\ F_{m1} & F_{m2} & \dots & F_{mS} \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_S \end{pmatrix}.$$

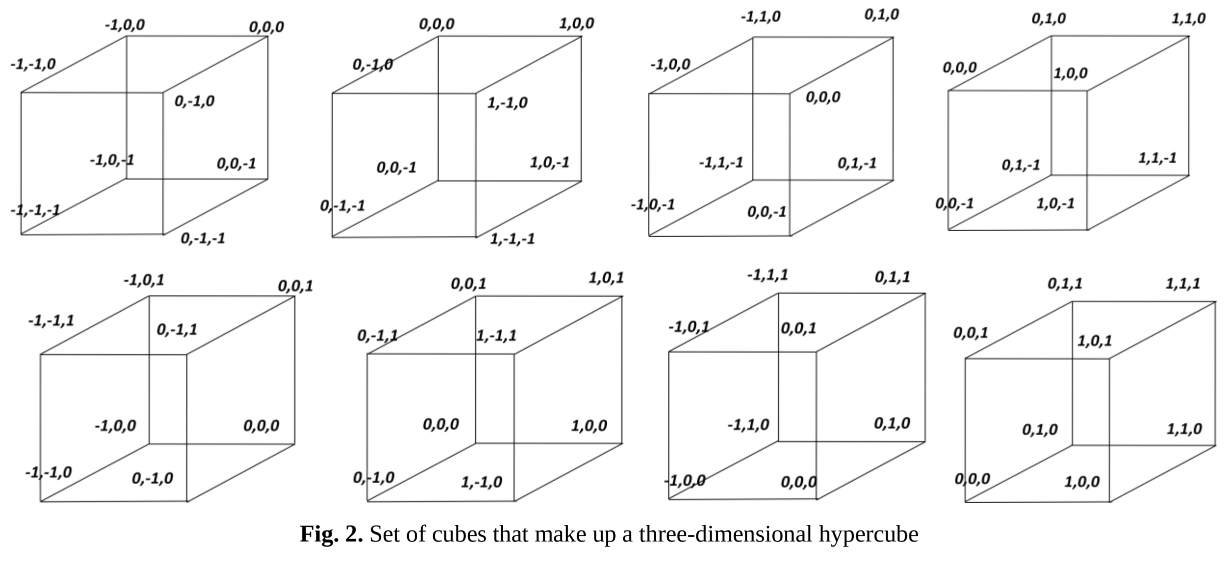

Fig. 2. Set of cubes that make up a three-dimensional hypercube

Introduce

$$y(F) = a_0 + a_1 F_1 + \dots + a_m F_m. \qquad (5)$$

Then the problem boils down to finding a vector that $A^T = (a_0, a_1, \dots, a_m)$ minimizes the least squares functional

$$\varphi(A) = (M_s A - Y)^T (M_s A - Y).$$

The desired vector A is determined by the ratio

$$A = (M_s^T M_s)^{-1} M_s^T Y. \qquad (6)$$

Now, using (6), we calculate the values of the approximating polynomial at the vertex of this cube by the formula (5) and the variance of this value by the formula

$$\vartheta_K(F_K) = (M_s^T M_s)^{-1} \sigma^2, \qquad (7)$$

where σ^2 - variance of response functions values evaluation at the points belonging to k-th cube. The operation of calculating the values will be $y(F)$ repeated for the vertices of all the cubes of the hypercube. As a result of these operations, a plan for a complete factorial experiment (PEF) will be obtained.

Assuming a piecewise constant dependence of the response function on time, we write the complete regression equation (1) as follows:

$$y(F) = a_0 + a_1 F_1 + \dots + a_m F_m + a_{12} F_1 F_2 + \dots + a_{m-1,m} F_{m-1} F_m + \dots + a_{12\dots m} F_1 F_2 \dots F_m. \qquad (8)$$

If all interactions together with factors are renumbered in order, then by introducing new variables, we get an expression for response functions that looks like

$$y(x) = \sum_{j=0}^{N} b_j x_j, \; N = 2^m. \qquad (9)$$

The corresponding matrix H of the complete factorial experiment is presented in Table 1.

Table 1 – **Matrix *H* of a Complete Factorial Experiment**

n/n	x_0	x_1	x_2	…	x_n	y
0	+	-	-	…	-	y_0
1	+	-	-	…	+	y_1
…	…	…	…	…	…	…
n	+	+	+	…	+	y_n

The plan of a complete factorial experiment has the following properties.

1. Symmetry with respect to the center of the experiment

$$\sum_{i=0}^{N} x_{ij} = 0, \quad j = \overline{1,n}.$$

2. Fulfillment of the normalization condition

$$\sum_{i=0}^{N} x_{ij}^2 = 2^m, \quad j = \overline{1,n}.$$

3. Orthogonality of the planning matrix columns

$$\sum_{i=0}^{N} x_{ij1} \cdot x_{ij2} = \begin{cases} 2^m, & j_1 = i_2; \\ 0, & j_1 \neq i_2. \end{cases}$$

Using the matrix H of the full factorial experiment, we define the vector B of the coefficients of the polynomial (9) as follows.

Let's write the relation (9) in matrix form:

$$Y = XB.$$

The components of the unknown vector B will be found by the least squares method, minimizing

$$J(B) = (XB - Y)^T (XB - Y). \qquad (10)$$

To this end, we differentiate the given function by vector B, after which we equate the resulting derivative to zero and solve the corresponding equation.

To implement the vector differentiation operation, we introduce an auxiliary function generated by (10):

$$\begin{gathered} J(B + ht) = [X(B + ht) - Y]^T [X(B + ht) - Y] = \\ = [XB + Xht - Y]^T [XB + Xht - Y] = \\ = [B^T X^T + H^T X^T t - Y^T][XB + Xht - Y] = \\ = B^T X^T XB + B^T X^T Xht - B^T X^T Y + \\ + h^T X^T XBt + h^T X^T Xht^2 - h^T X^T Yt - \\ - Y^T XB - Y^T Xht + Y^T Y. \end{gathered}$$

Next, we will find

$$\left. \frac{dJ(B+ht)}{dt} \right|_{t=0} = B^T X^T Xh + h^T X^T XB + 2h^T X^T Xht -$$
$$-h^T X^T Y - Y^T Xh \,|_{t=0} = B^T X^T Xh + h^T X^T XB - h^T X^T Y - Y^T Xh.$$

Since $h^T X^T XB$ and $h^T X^T Y$ are scalars, then

$$(h^T X^T XB)^T = B^T X^T Xh \text{ and } (h^T X^T Y)^T = Y^T Xh.$$

Then $\left. \frac{dJ(B+ht)}{dt} \right|_{t=0} = 2(B^T X^T X - Y^T X)h.$

Since $\left. \frac{dJ(B+ht)}{dt} \right|_{t=0} =$

$$= \left. \frac{dJ(B+ht)}{d(B+ht)} \cdot \frac{d(B+ht)}{dt} \right|_{t=0} = \frac{dJ(B)}{dB} \cdot h, \qquad (11)$$

then the desired derivative is $\frac{dJ(B)}{dB}$ equal to the matrix operator acting on h on the left in relation (11). At the same time

$$\frac{dJ(B)}{dB} = 2(B^T X^T X - Y^T X) = 0,$$

where from $B^T X^T X = Y^T X$ or $X^T XB = X^T Y$

and $$B = (X^T X)^{-1} X^T Y. \qquad (12)$$

Now let's take into account that the columns of the planning matrix are orthogonal and normalized. Then, since

$$X = (X_0 X_1 \dots X_n), \; X^T X = 2^n \begin{pmatrix} 1 & \dots & 0 \\ \dots & 1 & \dots \\ 0 & \dots & 1 \end{pmatrix},$$

$$(X^T X)^{-1} = \frac{1}{2^n} \begin{pmatrix} 1 & \dots & 0 \\ \dots & 1 & \dots \\ 0 & \dots & 1 \end{pmatrix}, \qquad B = \frac{1}{2^n} \begin{pmatrix} X_0^T & Y \\ X_1^T & Y \\ \dots & \dots \\ X_n^T & Y \end{pmatrix},$$

then $$b_j = \frac{1}{2^n} \sum_{i=0}^{n} X_{ij} y_i, \; i = 0, 1, \dots, n. \qquad (13)$$

Thus, for each vertex of the hypercube, using (13), the corresponding value of the response function can be calculated, as well as the variance of this value.

The resulting PFE plan can be used to calculate all coefficients of the complete regression polynomial [14, 15]. However, the quality level of the resulting description of the response functions is $y(F)$ unpredictable. The reason for this is the uncontrolled heterogeneity of the results of response functions measurements in a real experimental study, resulting from an unpredictable number of experiments in each of the cubes. Moreover, some cubes may turn out to be empty at all (or contain an insufficient number of experiments). The real possibility of obtaining acceptable results in this promising direction consists in choosing an orthogonal sub-plan from the available complete plan that has the required representation. In accordance with this, we set the task of finding an orthogonal symmetric subplane of PFE general plan satisfying the following requirements:

- total number of experiments used is the maximum,
- there are no empty cubes in the resulting subplane, as well as cubes with an unacceptably small number of experiments.

Let us proceed to the consideration of the method of forming an orthogonal representative replica [14] of the PFE plan. For clarity, we will present the technology of solving the problem for the special case when the number of factors $m = 6$. In this case, the PFE plan will have $2^6 = 64$ lines. The resulting set of rows will be placed in a three-dimensional cubic matrix (*i*,*j*,*k*) forming squares of dimension 4x4 in each section of the matrix (horizontal, vertical and frontal planes). In this case, we will put a triple (*i*,*j*,*k*) in accordance with each element of the matrix, where *i* is the row number, *j* is the column number, *k* is the column number. Now, for each specific experiment, it is easy to determine which of the cubes the point corresponding to this measurement falls into. At the same time, the total number of points included in the set of cubes included in this plan can be chosen as a criterion for the representativeness of a replication-like plan. The analytical relation for the criterion is obtained as follows. Let's introduce the indicator $\eta_{kj} = 1$, if the result for *j*-th dimension will fall into the *k*-th cube, and 0, otherwise. Now, for a specific replication-like plan *L*, we define

$$M(L) = \sum_{k=1}^{2^m} \sum_{j\in L}^{N} \eta_{kj}. \quad (14)$$

Next, the one that maximizes (14) is selected from the set of plans. The natural cut has the form

$$\sum_{j\in L} \eta_{kj} > m, k \in L, \quad (15)$$

that is, in the selected plan *L* there should be no “empty” cubes that do not contain a single dimension, as well as cubes with an unacceptably small number of experiments.

It is also interesting to solve this problem by another criterion. Significant differences in the number of experimental points that fall into each of the cubes of the hypercube lead to corresponding differences in the variances of the values of the response functions at the vertices of these cubes. The natural criterion for plan quality, taking into account this circumstance, has the form

$$D(L) = \sum_{k=1}^{2^m} \sum_{j\in L}^{m} D_{kj}\eta_{kj}. \quad (16)$$

At the same time, a plan minimizing (16) and satisfying (15) is selected from the set of orthogonal replication-like plans of the PFE. The direction of further research is the development of technology for the formation of a representative orthogonal subplan of the general plan of a complete factor experiment. We obtain a formal model of the problem of choosing an orthogonal representative replication-like plan as follows.

Let's introduce a set of indicators $X_{ijk} = 1$, if (*i*,*j*,*k*)-th element of the PFE plan is included in the sub-plan sought, and 0, otherwise. The system of restrictions for selecting the desired sub - plan has the form:

$$\sum_{i=1}^{m} X_{ijk} = 1; \quad \sum_{j=1}^{n} X_{ijk} = 1; \quad \sum_{k=1}^{p} X_{ijk} = 1; \quad i = \overline{1,m}; \quad j = \overline{1,n}; \quad k = \overline{1,p}. \quad (17)$$

Any solution of the system of equations (17) defines a plan in which the selected matrix elements are $\{C_{ijk}\}$located one at a time in each of the one-dimensional sections of the PFE plan. In this case, a set $\{X_{ijk}\}$ satisfying the constraints (15) and maximizing

$$L(X) = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{p} X_{ijk}, \quad (18)$$

defines a representative orthogonal subplan of the general plan of the factor experiment. The resulting problem (17), (18) has a special name - the triaxial assignment problem and is solved by the method developed in [16].

The orthogonality of the resulting subplan solution allows using the relations (12), (13) to calculate the values of all coefficients in the complete regression equation (8) and to filter out weakly influencing factors and interactions. The truncated regression polynomial formed in this case makes it possible to assess the state of the system in conditions of a small sample of initial data.

Conclusions

1. The analysis of methods for solving the problem of assessing the state of objects operating in a changing environment. It is established that a characteristic feature of these problems is the phenomenon of a small sample of initial data.

2. Possible models and methods for solving problems of assessing the state of objects in these conditions are considered. The most promising approach is reasonably chosen - the construction of a truncated regression equation linking the values of influencing factors and their interactions, on the one hand, and the value of the indicator of the object functioning effectiveness, on the other.

3. Method for constructing a truncated regression equation based on artificial orthogonalization of a passive experiment with the subsequent solution of a triaxial boolean assignment problem has been developed.

REFERENCES

1. Kwon, Baik (eds.) (2004), *Systems modeling and simulation: theory and applications*, third Asian Simulation Conf., AsiaSim 2004, Jeju Island, Korea, October 4–6, Revised Selected Papers, ISBN 3-540-24477-8, doi: https://doi.org/10.1007/b10561.
2. Derek W., Bunn and Erik R., Larsen (1997), *Systems modelling for energy policy*, Wiley, 250 p., ISBN 0-471-95794-1, available at: https://www.wiley.com/en-us/Systems+Modelling+for+Energy+Policy-p-9780471957942.
3. Hans-Jörg, Kreowski, Ugo, Montanari, Fernando, Orejas, Grzegorz, Rozenberg, and Gabriele, Taentzer (eds.) (2005), *Formal methods in software and systems modeling*, Springer, ISBN 3-540-24936-2, doi: https://doi.org/10.1007/b106390.

4. Jiming, Liu, Xiaolong, Jin and Kwok Ching, Tsui (2005), *Autonomy oriented computing: from problem solving to complex systems modeling*, Springer, doi: https://doi.org/10.1007/b101185.
5. Michael, Pidd (2004), *Systems Modelling: Theory and Practice*, John Wiley & Sons, 240 p., ISBN 0-470-86732-9, available at: https://www.wiley.com/en-us/Systems+Modelling:+Theory+and+Practice-p-9780470867310.
6. Václav, Pinkava (1988), *Introduction to Logic for Systems Modelling*, Taylor & Francis, 1988, 132 p., ISBN 0-85626-431-8, available at: https://www.biblio.com/book/introduction-logic-systems-modeling-pinkava-vaclav/d/1094804495.
7. Draper, N. R. and Smith, H. (1998), *Applied Regression Analysis*, ISBN: 978-0-471-17082-2, 736 p., available at: http://www.pzs.dstu.dp.ua/DataMining/mls/bibl/Applied_Regression_Analysis.pdf.
8. Strutz, T. (2013), *Data Fitting and Uncertainty*, A practical introduction to weighted least squares and beyond, Vieweg+Teubner, ISBN 978-3-8348-1022-9, available at: https://link.springer.com/book/9783658114558.
9. Stulp, Freek, and Olivier, Sigaud (2015), "Many Regression Algorithms, One Unified Model: A Review", *Neural Networks*, vol. 69, Sept. 2015, pp. 60–79, doi: https://doi.org/10.1016/j.neunet.2015.05.005.
10. Malakooti, B. (2013), *Operations and Production Systems with Multiple Objectives*, John Wiley & Sons, 1114 p., ISBN: 978-0-470-03732-4, available at: https://www.amazon.com/Operations-Production-Systems-Multiple-Objectives/dp/0470037326.
11. Chicco, Davide, Warrens, Matthijs J. and Jurman, Giuseppe (2021), "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation", *PeerJ Computer Science*, **7** (e623): e623, doi*:* https://doi.org/10.7717/peerj-cs.623.
12. Avriel, Mordecai (2013), *Nonlinear Programming: Analysis and Methods*, Dover Publishing, 512 p., ISBN 0-486-43227-0, available at: https://books.google.com.ua/books/about/Nonlinear_Programming.html?id=Gq1InQAACAAJ&redir_esc=y.
13. Bonnans, J. Frédéric, Gilbert, J. Charles, Lemaréchal, Claude and Sagastizábal, Claudia A. (2016), *Numerical optimization: Theoretical and practical aspects*, Berlin: Springer-Verlag, doi*:* https://doi.org/10.1007/978-3-540-35447-5.
14. Luenberger, David G. and Ye, Yinyu (2008), *Linear and nonlinear programming*, International Series in Operations Research & Management Science, Vol. 228 (Fourth Edition), New York: Springer, ISBN 978-0-387-74502-2, doi: https://doi.org/10.1007/978-3-319-18842-3.

Надійшла (received) 26.03.2023
Прийнята до друку (accepted for publication) 16.08.2023

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Раскін Лев Григорович– доктор технічних наук, професор, професор кафедри мультимедійних та інтернет технологій і систем, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна;
Lev Raskin – Doctor of Technical Sciences, Professor, Professor in the Department of Multimedia and Internet Technologies and Systems, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;
e-mail: topology@ukr.net; ORCID: http://orcid.org/0000-0002-9015-4016.

Карпенко Вячеслав Васильович – кандидат технічних наук, старший викладач кафедри мультимедійних та інтернет технологій і систем, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна;
Viacheslav Karpenko – Candidate of Technical Sciences, Senior Lecturer in the Department of Multimedia and Internet Technologies and Systems, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;
e-mail: kvvkml@gmail.com; ORCID: https://orcid.org/0000-0002-8378-129X.

Іванчихін Юрій Володимирович – кандидат технічних наук, доцент кафедри мультимедійних та інтернет технологій і систем, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна;
Yuriy Ivanchykhin – Candidate of Technical Sciences, Associate Professor in the Department of Multimedia and Internet Technologies and Systems, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;
e-mail:iwanywed@gmail.com; ORCID: http://orcid.org/0000-0002-3999-6541.

Соколов Дмитро Дмитрович – аспірант кафедри мультимедійних та інтернет технологій і систем, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна;
Dmytro Sokolov – postgraduate student in the Department of Multimedia and Internet Technologies and Systems, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;
e-mail: sokolovddd@gmail.com; ORCID: https://orcid.org/0000-0002-4558-9598.

Діагностика стану систем в умовах малої вибірки вихідних даних

Л. Г. Раскін, В. В. Карпенко, Ю. В. Іванчихін**,** Д. Д. Соколов

Анотація. Об'єкт дослідження - оцінка стану систем в умовах малої вибірки вихідних даних. **Актуальність** проблеми полягає в такому. Функціонування значної кількості реальних об'єктів відбувається в умовах погано прогнозованої зміни значень факторів зовнішнього середовища, що впливають на ефективність системи. Неоднорідність результатів експериментального дослідження об'єктів і середовища їхнього функціонування, що виникає при цьому, призводить до скорочення обсягу вибірки. При цьому стандартні вимоги щодо відповідності числа експериментів і числа коефіцієнтів рівняння регресії, що визначає стан системи, не виконуються. **Мета дослідження** – розробити методи оцінювання стану систем, що функціонують у середовищі, що змінюється, в умовах малої вибірки вихідних даних. **Завдання**, розв'язувані для досягнення мети: перше – еквівалентне перетворення множини спостережуваних вихідних даних, що формують у сукупності пасивний експеримент, на активний експеримент, якому відповідає ортогональний план; друге - побудова усіченого ортогонального показного підплану загального ортогонального плану, отриманого внаслідок розв'язання першого завдання. **Методи дослідження**: статистичні методи опрацювання експериментальних даних, регресійний аналіз, метод розв'язання триаксіальної булевої задачі призначення. **Отримані результати**: ортогональний представницький підплан повного факторного експерименту, який формують, дає можливість розрахунку усіченого рівняння регресії, що містить усі впливові фактори та їхні взаємодії. Аналіз коефіцієнтів цього рівняння відомими методами дає змогу відсікти малозначущі його елементи.

Ключові слова: оцінка стану об'єктів; мала вибірка вихідних даних; усічене представницьке рівняння регресії.