ISSN 0005-1055, Automatic Documentation and Mathematical Linguistics, 2023, Vol. 57, No. 3, pp. 135–139. © Allerton Press, Inc., 2023.
Russian Text © The Author(s), 2023, published in Nauchno-Tekhnicheskaya Informatsiya, Seriya 2: Informatsionnye Protsessy i Sistemy, 2023, No. 5, pp. 7–12.

Optimization of Special-Purpose Information Systems Operation

V. I. Sumin[a], Yu. Yu Gromov[b], and V. M. Tyutyunnik[b, c, *]

[a] *Voronezh Institute of the Federal Correctional Service, Voronezh, Russia*
[b] *Tambov State Technical University, Tambov, Russia*
[c] *International Nobel Information Centre (INIC), Tambov, Russia*
**e-mail: vmtyutyunnik@gmail.com*

Received March 13, 2023

Abstract—An approach to assessing and predicting the state of complex special-purpose information systems is considered, taking into account the impact of external factors. The functioning of such systems was studied using multivariate statistical analysis, for which it was proposed to use a method based on the use of matrices of paired and partial correlation coefficients.

Keywords: complex special-purpose information system, dynamic system, state transition matrix, matrix of pairwise correlation coefficients, matrix of partial correlation coefficients, causal analysis method

DOI: 10.3103/S0005105523030044

INTRODUCTION

The possibility of using a multilevel hierarchical approach for modeling complex special-purpose information systems (CSPIS) has been substantiated in [1–3]. It has been shown in [1, 4–6] that it is advisable to consider the functioning of such systems in a finite-dimensional space of states, which makes it possible to use algebraic research methods based on the application of the theory of finite semi-groups. The results of analyzing the process of operation of the CSPIS allow for the adoption of effective management decisions; however, it is necessary to answer the following questions:

(1) How can the process of operation of the CSPIS be evaluated by comparing the real values of the parameters of this system with their predicted values?

(2) What must the algorithm be for involving multivariate statistical analysis to make effective management decisions in the CSPIS?

The answers to these questions will make it possible both to reveal the usually hidden mechanisms of the formation of links in the CSPIS and to reasonably use statistical methods in analyzing the process of operation of the CSPIS. Moreover, considering the process of operation of the CSPIS in a finite-dimensional state space will improve the efficiency of decisions made [5, 7, 8], which ensure the achievement of the goal set for this system.

THEORETICAL PART

To construct a mathematical description of the process of operation of a complex information system S with a finite number of states, we introduce the following sets and functions that we will use:

- The set of admissible inputs U,
- The set of valid outputs Y,
- The set of states G,
- The transition function λ: $G \times U \to G$, (1)
- The output function γ: $G \times U \to Y$. (2)

The dynamics of the system is described by two families of mappings:

$$P = \{\lambda : G \times U \to G\}, \quad (3)$$

$$R = \{\gamma : G \times U \to Y\}, \quad (4)$$

where: P is the state transition display (“input – state 1 – state 2”);

R is the reaction of the system (“input – state – output”).

It is assumed that the sets U, Y, and G are finite. This allows us to represent the model of a complex information system in the form:

$$S = \{U, Y, G, R, P\}. \quad (5)$$

To study the algebraic structure of model (5), we propose to use the methods of the theory of finite semi-groups. In this case, special attention is paid to the analysis of the connectivity of functional structures on a finite-dimensional space of states, which are considered as elements of the set of vertices of a simplicial complex. In [1], the possibility of partitioning and combining subsystems of the CSPIS in analyzing the functioning of this system is substantiated on the basis of the following theorems.

Theorem 1. The number of steps K' required to obtain a structure with limit connections is less than the dimension of the state transition matrix ($K' \leq n$).

Theorem 2. If there are N dynamical systems S_i, ($i = 1, ..., N$), the description of which has the form (1)–(4) and which operate on time intervals $[T_{U_i}, T_{y_i}]$, then the system S described by expressions (1)–(4) is formed from the set of systems $\{S_i\}$ in accordance with the arrangement of time intervals $[T_{U_i}, T_{y_i}]$ if and only if the following conditions are satisfied:

(a) Connectivity of the time interval of operation

$$T_{yi} = T_{Ui+1}, \quad i = 1, \ldots, N; \tag{6}$$

(b) Continuity of state

$$\lim_{t \to T_{yi-0}} g(t) = \lim_{t \to T_{Ui+1+0}} g(t), \quad i = 1, \ldots, N-1. \tag{7}$$

Theorem 3. The dynamic description of systems S_i that simultaneously function and are interconnected has the form:

$$S = U \times Y \times G, \tag{8}$$

$$P = \{\lambda : G \times U \to G\}, \tag{9}$$

$$R = \{\gamma : G \times U \to Y\}, \tag{10}$$

$$S = \{U, Y, G, P, R\}, \tag{11}$$

where the components of the system S are obtained from the components of the dynamic systems $S_i \in S$ in accordance with the expressions:

$$U = U \backslash (U \cap Y); \tag{12}$$

$$Y = Y \backslash (Y \cap U); \tag{13}$$

$$g = \mathop{\Xi}_{i=1}^{N} g_i; \tag{14}$$

$$P = \mathop{\Xi}_{i=1}^{\mathrm{N}} P_i; \tag{15}$$

$$R = \mathop{\Xi}_{i=1}^{N} R_i, \tag{16}$$

where g_i, P_i, and R_i are the components of the description of the system S_i, which form a single description of the system S in the form (8)–(11).

The procedure for the formation of the systems $S_i \in S$ are called dynamic composition and denoted by the symbol ⊗. Then we can write $S = \otimes_{i=1}^{N} S_i, \quad S_i \in S$.

The use of Theorems 1–3 makes it possible to evaluate the state of complex special-purpose information systems. It follows from the systemic representation of the operation of the CSPIS that the level of achievement of the set goal by such a system is estimated by the values of model parameters (5). We assume that the predicted point is such a point in the finite-dimensional space of system states, in the coordinates of which the CSPIS takes the maximum value according to (5). The trajectories of the characteristics of this system that pass through the predicted points will be called predictable trajectories [4].

We consider the predicted CSPIS to be the system given by the predicted values:

$$S_n = \{U_n, Y_n, G_n, R_n, P_n\}, \tag{17}$$

$$t_n = U_n \times Y_n \times G_n. \tag{18}$$

The assessment of the operation of the CSPIS will be determined by the measure of the closeness of the real values of its parameters to the predicted ones. We will consider the system with its real values as a real CSPIS:

$$S_{\mathrm{r}} = \{U_{\mathrm{r}}, Y_{\mathrm{r}}, G_{\mathrm{r}}, R_{\mathrm{r}}, P_{\mathrm{r}}\}, \tag{19}$$

$$t_{\mathrm{r}} = U_{\mathrm{r}} \times Y_{\mathrm{r}} \times G_{\mathrm{r}}. \tag{20}$$

The input values of the CSPIS parameters are defined as the Cartesian product of pairs of sets of the predicted and real system. In the event that the process of operation of the CSPIS is defined, it is possible to identify the predicted values of the CSPIS parameters as a result of solving the optimization problem under given criteria. The operation of the predicted CSPIS is described over the entire time interval $[T_U, T_Y]$ by the expressions:

$$S_{\mathrm{p}} = \{U_{\mathrm{p}}, Y_{\mathrm{p}}, G_{\mathrm{p}}, R_{\mathrm{p}}, P_{\mathrm{p}}\}, \tag{21}$$

$$P_{\mathrm{p}} = \{l_{\mathrm{p}} : G_{\mathrm{p}} \times U_{\mathrm{p}} \to G_{\mathrm{p}}\}, \tag{22}$$

$$R_{\mathrm{p}} = \{g_{\mathrm{p}} : G_{\mathrm{p}} \times U_{\mathrm{p}} \to Y_{\mathrm{p}}\}. \tag{23}$$

If the conditions of theorems 1–3 are satisfied, the composition of the predicted (24) and real (25) CSPIS is possible in the form:

$$S_n = \bigcup_{i=1}^{N} S_{ni}, \tag{24}$$

$$S_{\mathrm{r}} = \bigcup_{i=1}^{N} S_{\mathrm{r}i}, \tag{25}$$

where: $S_{\mathrm{p}i}$, $S_{\mathrm{r}i}$ are the corresponding CSPIS.

If the predicted and real parameters of the CSPIS $\mathrm{S_p}$ and $\mathrm{S_r}$ differ slightly, then we will assume that the identity is true:

$$S_{\mathrm{p}} = S_{\mathrm{r}}. \tag{26}$$

If the predicted and real parameters of CSPIS S_{p} and S_{r} do not match, the following expression is valid:

$$\delta S_{\mathrm{pr}} = S_{\mathrm{p}} - S_{\mathrm{r}}. \tag{27}$$

Taking into account expressions (24), (25) and (26), we get:

$$\begin{aligned} S_{\mathrm{r}} = \{&U_{\mathrm{p}} + \delta U_{\mathrm{pr}}, Y_{\mathrm{p}} + \delta Y_{\mathrm{pr}}, G_{\mathrm{p}} + \delta G_{\mathrm{pr}}, \\ &R_{\mathrm{p}} + \delta R_{\mathrm{pr}}, P_{\mathrm{p}} + \delta P_{\mathrm{pr}}\}. \end{aligned} \tag{28}$$

In expression (28) δU_{pr}, δY_{pr}, δG_{pr}, δR_{pr}, δP_{pr} characterize deviations in the processes of functioning of the predicted and real values of the CSPIS parameters.

Taking into account expressions (24) and (28), the following will be true:

$$S_r = \bigcup_{i=1}^{N} \{U_{pi} + \delta U_{pri}, Y_{pi} + \delta Y_{pri}, G_{pi} + \delta G_{pri}, R_{pi} + \delta R_{pri}, P_{pi} + \delta P_{pri}\}. \quad (29)$$

Employees using the CSPIS are required to monitor the following to ensure its effective operation: installation, operation and reliability of security equipment (SE) at protected objects (POs); actions of mobile groups (MGs); determination of the reasons for the receipt of false signals from SE with POs, etc.

To study the process of operation of the CSPIS, the determination of the predicted values of its parameters (factors) requires the use of multivariate statistical analysis in the following areas: preliminary processing of statistical data on the operation of security equipment, the actions of mobile groups and the establishment of links between them; building the regression models that determine the relationship between the reasons for the receipt of false alarms by security equipment with integrated performance indicators of the CSPIS; and assessment of the stability of the operation of the CSPIS for each parameter or combination of parameters [9–11].

The operation of the CSPIS is determined by a variety of procedures: synthesizing, analyzing, predicting, and evaluating procedures, which provide the ability to determine, analyze and predict the consequences of the influence of destructive factors on this system. Destructive factors (including the criminal actions of intruders) affect security equipment at protected objects [7, 12, 13]. Regression models are used to predict the results of their impact, about which information is accumulated in databases.

For the CSPIS we are considering, it is characteristic that a large number of false signals from protected objects coming from security equipment arise due to the influence of destructive influences on these objects, including:

- Poor quality maintenance (F_1);
- The owner's improperly turning over the object for protection (F_2);
- Malfunction or interference in telecommunication systems (F_3);
- Failures of security equipment (F_4);
- Power failures (F_5);
- Wrong actions of mobile groups (F_6);
- Incorrect design of security equipment at protected objects (F_7);
- Influence of external factors on the work of security equipment (F_8);
- Penetration of an intruder into a protected object (F_9);
- Lack of security of information in the CSPIS (F_{10}).

It should be noted that the consequences of the influence of the first two destructive impacts can be reduced due to organizational measures at protected objects, i.e. it is necessary to conduct better technical maintenance of security equipment and fine owners for improperly turning over objects for protection.

It is possible to reduce the consequences of other destructive impacts on the protected object by:

- Modernization of telecommunication systems;
- Replacement of unstably working security equipment with more reliable one;
- Installation of security equipment with backup power supply;
- Training employees of mobile groups in actions in case of receipt of alarm signals from the protected object; and
- Correct design of security equipment at the protected object.

The first priority for employees of units using the CSPIS is to reduce the number of such signals. Measures to reduce the number of false signals that are aimed at improving the reliability of the operation of security equipment at protected objects must be comprehensive and systematic.

Let us dwell in more detail on the solution of the issue that was raised at the beginning of this article: the use of methods of multivariate statistical analysis to assess the process of operation of the CSPIS. At the first stage, it is necessary to build linear dependences that connect the values of the output parameters $\bar{Y}$ with the values of destructive influences $X_1, \ldots X_m$ that affect $\bar{Y}$. This is achieved by constructing a dependence that takes into account errors in determining the values of destructive effects and has the form:

$$\bar{Y} = f(X_1, X_2, \ldots, X_m) + \varepsilon, \quad (30)$$

where: ε is a normally distributed random error with mathematical expectation $M[\varepsilon] = 0$ and variance $D[\varepsilon] = \sigma^2$.

Let's represent expression (30) in a vector-matrix form:

$$\bar{Y} = X\beta + \varepsilon, \quad (31)$$

where: $\bar{Y} = (y_1, y_2, \ldots, y_n)^T$ is the column vector of height n of the observed values of the dependent parameter (variable) Y; $X = (x_{ji})$, $j = 1', n$, $i = 0', m$ is the matrix of size $n(m+1)$ for observations of independent variables $X_0, X_1, \ldots X_m$ (regression matrix); X_0 is a dummy variable whose observation column vector has the form $(x_{10}, x_{20}, \ldots x_{n0})^T$ with the equality $x_{10} = x_{20} = \ldots = x_{n0} = 1$ being satisfied; $\beta = (\beta_0, \beta_1, \ldots, \beta_m)^T$ is the column vector of the height $(m+1)$ of the coefficients of equation (31); $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T$ is the column vector of the height n of errors (residuals).

The significance of the regression coefficients β_i is checked by Student's t-test:

$$t_i = |\beta_i / \sigma\{\beta_i\}| \quad (32)$$

with the number of degrees of freedom $\nu = n - m - 1$, where $\sigma\{\beta_i\} = \sigma_{\text{res}}\{Y\}\sqrt{c_{ii}}$ is the estimate of the standard deviation of the regression coefficient β_i; c_{ii} is the diagonal element of the matrix $(X^T X)^{-1}$; $\sigma_{\text{res}}\{Y\} = \sqrt{\sigma^2_{\text{res}}\{Y\}}$ is the sample residual standard deviation; $\sigma^2_{\text{res}}\{Y\} = \nu^{-1}\sum_{j=1}^{n}\varepsilon_j^2$ is the sample residual variance; $\varepsilon_j = y_j - \hat{y}_j$ is the j-th element of the residual vector ε, which is equal to the deviation of the observed value y_j of the dependent parameter (variable) Y from the estimated regression line $\hat{y}_j$ formed by the column vector $\hat{Y} = X\beta = (\hat{y}_1, \hat{y}_2, \ldots \hat{y}_n)^T$ of the height n of the predicted values of the output parameter Y according to the regression equation.

Then, if $t_i < t_{\text{cr}}(\nu, \alpha)$, where $t_{\text{cr}}(\nu, \alpha)$ is the critical value of Student's test (t-criterion) determined from the tables for the two-sided Student's distribution at the selected confidence probability α (as a rule, for engineering applications it is enough to set $\alpha = 0.95$) and the number of degrees of freedomν, then the corresponding coefficient β_i is recognized as insignificant ($\beta_i = 0$) with the confidence probability α, and the estimation of the parameters is repeated for a new type of equation.

In the event that the first stage ended unsuccessfully, i.e., we failed to build a linear model that meets the specified requirements, then it is necessary to proceed to the next stage, which is to search for more complex nonlinear regression models. In this case, both variables and parameters can be nonlinear. There are two options for constructing an estimate of the parameters of a nonlinear regression model.

1. In some cases, a non-linear regression equation can be brought to a linear form using a change of variables. To do this, it is necessary to find transformations $g(Y)$ and $h_i(X_i)$, $i = 1', m$, which will provide

$$g(Y) = \sum_{i=1}^{m} L_i h_i(X_i),$$

where: L_i is equal to β_i or a known function of β_i. In this case, the theory of linear regression can be applied to our equation. A wide class of functions that allow such transformations is given in [9, 10].

2. If the non-linear regression equation cannot be reduced to a linear form, then in this case the non-linear least squares method can be used [14].

A detailed analysis of the operation of the CSPIS requires identifying the mechanisms for the formation of links, as noted earlier. Meanwhile, we note that the pairwise correlation coefficient r_{ij} of two factors X_i and X_j is a reliable estimate of their linear statistical relationship only in the case when X_i and X_j are not affected by other factors; otherwise, with a false correlation, the relationship between the factors X_i and X_j will be "induced."

If there are differences between the matrices of pairwise and partial correlation coefficients of factors affecting the operation of the CSPIS, it is necessary to use the causal analysis method, which will provide additional information for the analysis of this system, in contrast to standard statistical methods. The use of causal analysis methods [8, 15] makes it possible to take into account two issues related to problem solving: determining the structural causal relationships between destructive influences and quantifying them. The solution of the first problem is possible by removing the false correlation, and the remaining relationships are considered causal. The second task is solved through path analysis (P-analysis), which is based on the use of linear functional relationships between destructive influences, where the regression coefficients are interpreted in terms of causal relationships.

The path coefficient determines the intensity of the influence of X_i on X_j and is calculated as follows:

$$P_{ij} = b_{ij}\frac{\sigma_i}{\sigma_j},$$

where: σ_i, σ_j are standard deviations of X_i and X_j; b_{ij} is the corresponding regression coefficient.

An assessment of the causal relationship between two factors X_i and X_j, if they are not neighboring $X_i \rightarrow X_{i+1} \rightarrow \ldots \rightarrow X_{i+1} = X_j$, is carried out using the expression:

$$C_{ij} = \prod_{k=1}^{l} P_{i+k-1,i+k}\sigma_j^2.$$

Determining the predicted values of parameters for complex special-purpose information systems requires a more thorough analysis, as shown by the experience of their use [16–18].

CONCLUSIONS

At present, it is important to improve the efficiency of operation of complex special-purpose information systems through the use of automated static analysis systems, taking into account the specific features of the work of protected objects and processing the parameters of theifqr functioning. The specifics of the 1
work of many protected objects involves:

(1) Collecting and processing the functioning parameters of the entire diverse set of system objects;

(2) Determining and analyzing the causes of alarm signals coming from security equipment and the reliability of its functioning;

(3) Generating statistical information about the causes of alarms coming from security equipment

(based on information in databases) and identifying patterns of their occurrence using multivariate statistical analysis in the form of a linear analytical dependence of regression analysis;

(4) Analyzing the potential of building system management models that have the ability to form management decisions for employees of departments of complex special-purpose information systems based on computer use of statistical and causal analysis methods.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interests. The authors declare that they have no conflict of interest.

REFERENCES

1. Sumin, V.I., Smolentseva, T.E., Gromov, Yu.Yu., and Tyutyunnik, V.M., Analysis of the functioning and structural decomposing of special-purpose information systems, *Autom. Doc. Math. Linguist.*, 2021, vol. 55, no. 4, pp. 169–177. https://doi.org/10.3103/s0005105521040087
2. Tyutyunnik, V.M., Data analysis and information process model for application information systems, *Promyshlennye ASU Kontrollery*, 2019, no. 4, pp. 19–29. https://doi.org/10.25791/asu.04.2019.576
3. Gromov, Yu.Yu. and Tyutyunnik, V.M., Materials to construction of the theory of information. I: Measuring the amount and quality of information, *Fundam. Issled.*, 2011, no. 8, pp. 347–355.
4. Rytov, M.Yu., Musienko, N.O., Gubskov, Yu.A., and Minin, Yu.V., Audit and monitoring of the state of informatization objects in the process of designing complex information protection systems for significant objects of critical information infrastructure, *Prib. Sist. Upr., Kontrol', Diagn.*, 2022, no. 10, pp. 10–18. https://doi.org/10.25791/pribor.10.2022.1364
5. Sumin, V.I. and Durdenko, V.A., *Teoreticheskie osnovy avtomatizatsii proektirovaniya sistem upravleniya podrazdelenii vnevedomstvennoi okhrany sub''ekta federatsii* (Theoretical Foundations of Automation of Design of Control Systems of Extra-Deparmental Guards in a Subject of Federation), Voronezh: Voronezhsk. Gos. Univ., 1997.
6. Sumin, V.I. and Durdenko, V.A., *Osnovy proektirovaniya sistem upravleniya okhrannoi deyatel'nost'yu sub''ekta federatsii* (Foundations of Design of Control Systems of Guarding Activity in a Subject of Federation), Voronezh: Voronezhskaya Vysshaya Shkola MVD Rossii, 1998.
7. Sumin, V.I., Durakov, S.G., Chulyukov, V.A., and Yaroshenko, M.V., Building a probabilistic algorithm for finding arbitrary in a given route digraph models the structure of information systemsand the developments on protected objects, *Prib. Sist. Upr., Kontrol', Diagn.*, 2012, no. 5, pp. 47–49.
8. Gaganov, I.N. and Turumbaev, A., *Prichinnyi analiz slozhnykh sistem* (Cause-and-Effect Analysis of Complex Systems), Alma-Ata: Mektep, 1984.
9. Seber, G.A.F., *Linear Regression Analysis*, New York: Wiley, 1977.
10. Draper, N.R. and Smith, H., *Applied Regression Analysis*, New York: Wiley, 1966.
11. Bard, Y., *Nonlinear Parameter Estimation*, New York: Academic, 1974.
12. Gromov, Yu.Yu. and Tyutyunnik, V.M., Models for assessing the security of network information systems against negative external influences, *Autom. Doc. Math. Linguist.*, 2021, vol. 55, no. 5, pp. 196–202. https://doi.org/10.3103/S0005105521050071
13. Tyutyunnik, V.M., Cybernetic model of the control system, *Inf. Protsessy, Sist. Tekhnol.*, 2022, vol. 3, no. 3, pp. 5–7. https://doi.org/10.52529/27821617_2022_3_3_05
14. Saraev, P.V., Nonlinear least squares method and block recurrent and iterative procedures in neural networks teaching, *Upr. Bol'shimi Sist.*, 2010, no. 30, pp. 24–34.
15. Heise, D.R., *Causal Analysis*, John Wiley & Sons, 1975.
16. Abd Alkhadi, Kh., Minin, Yu.V., and Tyutyunnik, V.M., Analytical models of decision-making support in an enterprise, *Inzh. Fiz.*, 2021, no. 8, pp. 22–33. https://doi.org/10.25791/infizik.8.2021.1221
17. Sumin, V.I., Churakov, D.Yu., and Tsarkova, Ye.G., Development of models and algorithms information structures processes and objects of special importance,, 2019, pp. 30–39.
18. Kravchenko, A.S. and Sumin, V.I., Algorithm of informational process of calculation of the shortest way of penetration through projected system of protection, *Vestn. Voronezh. Gos. Tekh. Univ.*, 2011, vol. 7, no. 8, pp. 177–179.

Translated by L. A. Solovyova

SPELL: 1. theifqr