Моделирование и анализ данных
2020. Том 10. № 1. С. 7–34
DOI: 10.17759/mda.2020100101
ISSN: 2219-3758 (печатный)
ISSN: 2311-9454 (online)

Modelling and Data Analysis
2020. Vol 10, no 1, pp. 7–34
DOI: 10.17759/mda.2020100101
ISSN: 2219-3758 (print)
ISSN: 2311-9454 (online)

УДК 351.8, 629.7

Evaluating the Contribution of Human Factor to Performance Characteristics of Complex Technical Systems

Kuravsky L.S.*
MSUPE, Moscow, Russia,
ORCID: http://orcid.org/0000-0002-3375-8446,
e-mail: l.s.kuravsky@gmail.com

Yuryev G.A.**
MSUPE, Moscow, Russia,
ORCID: https://orcid.org/0000-0002-2960-6562,
e-mail: g.a.yuryev@gmail.com

Zlatomrezhev V.I.***
GosNIIAS, Moscow, Russia,
ORCID https://orcid.org/0000-0003-1776-6881
e-mail: vizlatomr@2100.gosniias.ru

Yuryeva N.E.****
MSUPE, Moscow, Russia,
ORCID: https://orcid.org/0000-0003-1419-876X,
e-mail: yurieva.ne@gmail.com

Mikhaylov A.Y.*****
GosNIIAS, Moscow, Russia,
ORCID: https://orcid.org/0000-0003-0278-1819
e-mail: aymihaylov@2100.gosniias.ru

For citation:
Kuravsky L.S., Yuryev G.A., Zlatomrezhev V.I., Yuryeva N.E., Mikhaylov A.Y. Evaluating the Contribution of Human Factor to Performance Characteristics of Complex Technical Systems. *Modelirovanie i analiz dannykh = Modelling and Data Analysis*, 2020. Vol. 10, no. 1, pp. 7–34. DOI: 10.17759/mda.2020100101

Kuravsky Lev Semenovich, DSc (Engineering), Professor, Dean of Computer Science Faculty, Moscow State University of Psychology and Education, Moscow, Russia, ORCID: http://orcid.org/0000-0002-3375-8446, e-mail: l.s.kuravsky@gmail.com

****Yuryev Grigory Aleksandrovich,*** PhD (Physics and Mathematics), Associate Professor, Head of Scientific Laboratory, Moscow State University of Psychology and Education, Moscow, Russia, ORCID: https://orcid.org/0000-0002-2960-6562, e-mail: g.a.yuryev@gmail.com

*****Zlatomrezhev Valentin Igorevich,*** Head of Laboratory, State Research Institute of Aviation Systems (GosNIIAS), Moscow, Russia, ORCID https://orcid.org/0000-0003-1776-6881, e-mail: vizlatomr@2100.gosniias.ru

Estimating the influence of human factor on the activity of operators of complex technical systems is an important problem for condition monitoring, personnel training and diagnostics. Presented are both an overview and mutual comparisons of the approaches which are useful to reveal the effect of human factor and have already shown their performances in practical applications. Under consideration are: the structural equation modeling, the Bayesian estimations for probabilistic models represented by Markov random processes, the multivariate statistical techniques including the discriminant and cluster analysis as well as wavelet transforms.

Keywords: Operators of complex technical systems, human factor, condition monitoring, factor analysis, structural equation modeling, Markov random processes, wavelet transform, multivariate statistical techniques, principal components analysis, multidimensional scaling, cluster analysis.

*****Yuryeva Nataliya Evgenievna,*** PhD (Engineering), Research fellow, Moscow State University of Psychology and Education, Research Associate, Moscow, Russia, ORCID: https://orcid.org/0000-0003-1419-876X, e-mail: yurieva.ne@gmail.com

******* Mikhaylov Artem Yurievich,*** Engineer, State Research Institute of Aviation Systems (GosNIIAS), Moscow, Russia, ORCID: https://orcid.org/0000-0003-0278-1819, e-mail: aymihaylov@2100.gosniias.ru

1. INTRODUCTION

Estimating the influence of human factor on the activity of operators of complex technical systems is an important problem for condition monitoring personnel training and diagnostics. Presented are the approaches which are useful to reveal the effect of human factor and have already shown their performances in practical applications. Under consideration are both the structural equation modeling presented in Section 2 and the multivariate statistical techniques as well as Bayesian estimations for probabilistic models based on Markov random processes, which are presented in Section 3. The first approach is intended for revealing influences of human factor on variances of observed characteristics whereas the second one recognizes influences of training level and psychophysiological state of operators of complex technical systems on their activity performance.

As a rule, available parameters measured for condition monitoring do not represent characteristics of a system under study in the mode that is suitable directly for understanding system status and formulating reliable conclusions sufficient for proper diagnostics. For multivariate measurements, which condition monitoring usually deals with, it is important to reveal some latent factors responsible for joint variability of observed measurable parameters, determine their nature and scope of influences, and use the obtained information to identify system condition.

It is desirable to replace the parameters those are easy to measure by the parameters those are easy to interpret and understand the system behavior, with minimal information losses being expected during this data mining. Functional relationships between revealed factors and observed parameters are also to be determined for further analysis. As a result of this study, a researcher should get the structure of causal connections between revealed factors and observed variables as well as immediate factor values to differentiate system status, if necessary.

To meet all the indicated requirements, empirical mathematical models and corresponding methods of multivariate statistical analysis were developed [3-4, 13, 33, 37].

The most appropriate in the discussed situation are exploratory and confirmatory factor models and methods of their analysis. Both approaches are based on the analysis of sample covariance or correlation matrices of the observed parameters under study. The exploratory analysis assumes unknown number of uncorrelated factors with a priori undetermined interpretation[1], whereas the confirmatory one assumes the factors, their interpretation, causal connections with observed variables and correlation connections between latent factors to be known beforehand. Confirmatory models also admit a convenient technique for estimating statistical significance of each their component.

Since substantial hypotheses of the reasons of possible influences on the observed variables are usually available in practice, the latter approach is preferable.

Nevertheless, condition monitoring usually needs to take into account time dynamics of observed parameters, with their magnitudes for different time points being formally interpreted as different quantities to be analyzed. To comply with this demand, the simplex method of the structural equation modelling was developed [16]. However, it has serious inherent limitations, which frequently make its practical applications questionable, viz.: capacity of studying factor interaction for adjacent checkpoints only, impossibility of associating factors with time periods, acceptability for analysis of covariance/correlation matrices with simplex structure merely, etc. Besides, the traditional structural equation modelling has its own intrinsic defect. It needs solution of the laborious local multivariate optimization problem to estimate the values of free model parameters that brings about impossibility of the global minimum estimation and ambiguous solution.

To overcome these problems, an approach combining capabilities of both wavelet transforms and trained confirmatory factor structures was developed. Its features and advantages, including the possibility of finding the values of free model parameters by direct (non-iterative) methods ensuring an unambiguous optimal solution, flexible capacity of studying factor interaction, applicability for the analysis of arbitrary covariance/correlation matrices et al., are presented in Section 2.

Presented in Section 3 is the approach for supporting the outcome grading for activities of operators of complex technical systems. It is based on comparisons of current exercises with the activity database patterns in the wavelet representation metric associated with observed parameters as well as on probabilistic assessments of skill class recognition using sample distribution functions of exercise distances to cluster centers in a scaling space and Bayesian likelihood estimations with the aid of probabilistic profile of staying in activity parameter ranges. These techniques have demonstrated the capabilities of recognizing sets of abnormal exercises in the scaling spaces with the wavelet coefficient metric and detection of parameters characterizing operator mistakes to reveal the causes of abnormality. The techniques presented overcome limitations of existing methods and provide advantages over manual data analysis since they greatly reduce the combinatorial enumeration of the options considered.

An objective assessment of activity performance is essential for training process of operators of complex technical systems. One of the critical aspects here is development of the training evaluation criteria. The objective data based on trainee activity character-

1 Factors are usually interpreted using variables, which they are connected with: to identify a factor it is necessary to assign it a name generalizing the meanings of relevant variables.

istics may be an effective indicator for objective judging of the level of training effectiveness and skills obtained after training. So, in demand are computer-aided diagnostics techniques that can be employed for selection of operator candidates to estimate the level of building-up of knowledge, ability and skills. These techniques can raise objectivity, informational content and accuracy of estimations together with standardization and automation of measurements. Of special importance are development and analysis of new approaches which are used to estimate level of training and psychophysiological state of operators. The principal focus area in this regard is diagnostics as a result of work on contemporary *simulators* where special conditions are worked out.

By now, a certain amount of results related to selection of abnormal exercise implementations has been accumulated [2, 7, 8-12, 14, 15, 18, 34-36, 41, 43-44].

Presented are the techniques for skill assessments associated with relevant activity data obtained with the aid of both experiments and mathematical analysis of their results. As a result of its application, classes of skills should be determined using activity parameters and operator actions revealed during exercises.

The presented approach is developed to support assessment process. To evaluate the data under study an activity record database is required, in which patterns of training data representing exercise implementations by different operator crews are collected. A *pattern* in question is a representation of a certain activity fragment to be analyzed, which is referred to as an *exercise*, via the set of parameters describing this exercise. These patterns are to be related to one of the recognizable skill classes of trainees.

Collected training data should include exercise parameters as well as relevant expert instructor’s assessment/supervision comments from different sources including various types of activity simulators (mobile, motionless, etc.), virtual reality systems and real operator work. The expert comments in use should reveal weak points of operator performance using information about typical mistakes in terms of activity parameters and advices to a local instructor how to correct these weaknesses. Certain attributes related to recognizable classes can be stored in the database both for cluster centers of patterns belonging to the classes under consideration and for each exercise pattern to solve application problems of interest.

General assumption of the approach in use is that the activities implemented in different styles and quality as well as exercises of different types can be discriminated in the multidimensional space formed by wavelet coefficient representation. This statement is proved by computer experiments based on the exercises implemented on a simulator. The general method that results from this conclusion is pattern selection.

After an operator completes a certain test exercise, then, in order to collect a meaningful data set representing his results accumulated by some time check point, the following quantities are to be calculated with the aid of the given database:

– Distances to cluster centers of patterns belonging to recognizable classes, which are determined in the metric associated with the wavelet representations of observed parameters, together with the sample distribution functions of distances to centers of such clusters;
– The nearest pattern in the same metric, following to which the relevant attributes are ascribed to the operator under assessment.

The fragment of exercise results must be comparable with patterns in the database in terms of the number of test exercises under consideration and the time of their implementation.

Presented techniques make it possible to carry out quantitative and qualitative assessments of skill class, which are based on three ways, viz.:

- direct comparison of current exercises with the activity database patterns in the wavelet representation metric associated with observed parameters (selecting activity database patterns in the wavelet representation metric), which is considered as a basic technique as well as on
- probabilistic assessment of skill class recognition using both sample distribution functions of exercise distances to cluster centers in a scaling space (selecting a skill class using exercise distributions in a scaling space), which is considered as a supportive technique, and
- Bayesian likelihood estimation (selecting a skill class with the aid of probabilistic profile of staying in activity parameter ranges), which is also considered as a supportive technique.

The techniques presented provide tremendous advantages over manual data analysis since they greatly reduce the combinatorial enumeration of the options considered.

The proposed techniques differ significantly from the probabilistic methods applied for system control, predictive diagnostics of technical failures, condition monitoring and operator activity support [17].

2. STRUCTURAL EQUATION MODELING

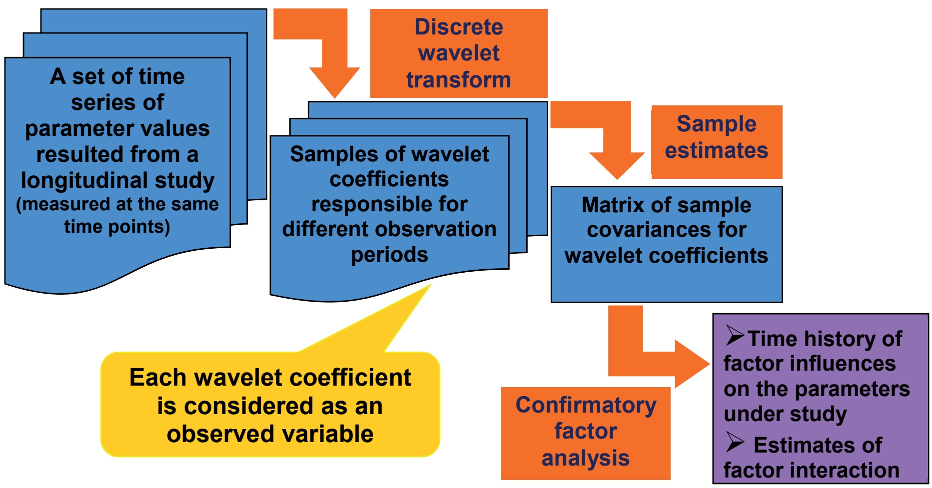

Figure 1. Principal stages of the analysis

Principal stages of the structural equation modeling are presented in Figure 1. This technology combines capabilities of wavelet transforms and trained factor structures. According to the proposed approach, the samples of coefficients resulted from discrete wavelet transform of initial parameter time series under study and responsible for different observation periods are considered as values of observed variables in the subsequent structural equation modelling to reveal time history of factor influences and estimates of factor interaction. Data representation created with the aid of wavelet transforms makes it possible to reveal differences in process characteristics for diverse scales. Identification of free factor model parameters (usually factor variances and covariances) is carried out by a new direct (noniterative) procedure based on the maximum likelihood method, which is an alternative to traditional local iterative solution of optimization problems.

2.1. Principal Components of the Technology: Wavelet Transforms

Monitoring process representation to be analyzed is created with the aid of wavelet transforms. These transforms make it possible to reveal differences in process characteristics for diverse scales, with the process features being available for analysis in different time points of some interval under study. If the dependence under test is a usual one-variable function, resulting wavelet-spectrum is the function of two arguments, viz.: scale parameter characterizes oscillation time cycles whereas shift parameter – time displacements. Wavelet-spectra are calculated using wavelets, which are special functions in the form of short waves with both zero integral value and localization along the axis of the independent variable, which are able to shift along this axis as well as to scaling (stretching/contraction.)

Wavelet-analysis has clear superiority over the traditional spectral analysis since it yields correct representation in case of transition (non-stationary) processes and keeps more useful information about the object behavior under study. These and other advantages have made the approach in question very popular among researchers of different specialties now. Its discrete variant is used here to represent initial data patterns in the form of points of a certain metric functional space with a wavelet-basis. This representation is necessary for a recognition procedure described below.

2.2. Principal Components of the Technology: Alternative Variant of the Structural Equation Modelling

Principal components of the structural equation modelling are presented in Figure 2.

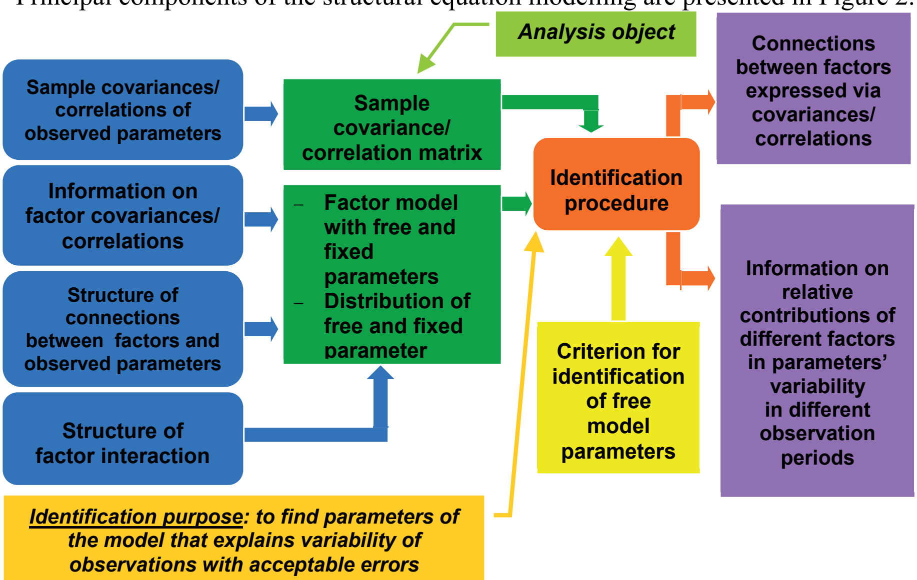

Figure 2. Principal components of the structural equation modelling

Note. In case of several types of observed random processes a mixed covariance matrix for wavelet coefficients is formed.

2.2.1. Traditional Structural Equation Modelling

Strictly determined factor model of the phenomenon under study is assumed in the traditional structural equation modelling. A factor model that connects latent and observed variables is formed using knowledge about the application domain. The hypotheses concerning the model structure have to be based on the analysis of the investigated factors nature (that is to say, both theory and observations are taken into account) [37]. It is admissible to formulate quantitative assumptions concerning correlations between latent variables as well as factor loadings. Free model parameters are calculated to get the best approximation of correlation (covariance) matrices for observed variables (from the viewpoint of a given criterion).

Objects of the traditional structural equation modelling are correlation or covariance matrices for observed variables. Analysis purpose is to find model parameters that explain variability of observations with acceptable errors.

Structural equation modelling has a following peculiarities:

- Nonzero (free) factorial loadings in the equations of model and number of investigated factors are defined in advance
- Correlation between errors of measurements are possible
- Factor loadings and covariances between latent variables can be free model parameters or be equal to the given constants
- The analysis of several model groups is supposed
- It is possible to check whether the given restrictions meet observation results
- Possibility of checking hypothesises about model properties by selecting optimum values of free parameters
- Estimations of free parameters are determined by the maximum likelihood method.

Using the maximum likelihood method the following statistic is to be minimized as a criterion for selection of free parameters:

$$\boldsymbol{F} = [\boldsymbol{ln}\,|\boldsymbol{\Sigma}| - \boldsymbol{ln}\,|\mathbf{S}| + \boldsymbol{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \boldsymbol{m}]\,(\boldsymbol{N\text{-}1}),$$

where $\mathbf{S}$ – sample covariance matrix for observed variables, $\boldsymbol{\Sigma}$ – expected covariance matrix for observed variables, $|\boldsymbol{\Sigma}|$ and $|\mathbf{S}|$ – determinants of matrices $\boldsymbol{\Sigma}$ and $\mathbf{S}$, $\boldsymbol{tr}\,(\mathbf{S}\boldsymbol{\Sigma}^{-1})$ – trace of matrix $(\mathbf{S}\boldsymbol{\Sigma}^{-1})$, $\boldsymbol{N}$ – size of the sample used to calculate matrix $\mathbf{S}$, $\boldsymbol{m}$ – number of observed variables [33].

Elements of the expected covariance matrix are analytical expressions consisting of free model parameters. In case of multivariate normalcy of observed variables values of the statistic $\boldsymbol{F}$ are distributed as $\boldsymbol{\chi}^2$.

Thus, to estimate the values of free model parameters it is necessary to solve numerically a sufficiently laborious local multivariate optimization problem by the iteration methods. In general case, this way results in impossibility of the global minimum estimation, since one of the possible local minima depending on its initial approximation is usually found. Consequently, the solution is ambiguous.

2.2.2. Alternative Variant of the Structural Equation Modelling: General Principles of the Approach

Since the traditional structural equation modelling expects decision of the laborious local multivariate optimization problem to estimate the values of free model parameters that results in impossibility of the global minimum estimation and solution ambiguities. Proposed alternative variant of the structural equation modelling allows to find the values of free model parameters by direct (noniterative) methods ensuring an unambiguous optimal solution.

Expressing observed variances and covariances via free factor variances and covariances with the aid of a factor model, in the alternative variant of the structural equation modelling it is proposed:

- To compose overdetermined sets of equations;
- To solve them by a direct (noniterative) method using a certain form of the maximum likelihood approach, which is different from the one used in the structural equation modelling [20];
- To examine for the adequacy of the obtained equation sets to observations with the aid of statistical goodness-of-fit tests.

To avoid solving nonlinear equation sets as respects correlation coefficients and factor loadings the variance component's path model in which path coefficients (factor loadings) equal to unity is used.

Hereinafter, each observed variance and covariance is associated with an equation that expresses analytically their expected value via free variances and covariances of latent variables and equates it with the corresponding sample estimation [20]. In particular, tracing rules of the path diagram[2] analysis may be used for that. Detour begins against a causal relationship, then change of a direction on covariance communications, and then movement along a causal relationship. It is necessary to remember also, that covariance communication cannot be bypassed twice. As a result the set of the equations is obtained, in which number of the equations equals to the number of observed variances and covariances. If this number of equations exceeds the number of free model parameters, the overdetermined set of equations is the case that is necessary for the further decision. The method under consideration needs also multivariate normalcy of observed variables.

Let us represent the obtained overdetermined set of equations in matrix notation:

$$\mathbf{Ax}=\mathbf{b},$$

where **A** – system $\boldsymbol{n}$Н$\boldsymbol{m}$ matrix, which coefficients are determined using the factor model (path diagram) under consideration; **b** – column $\boldsymbol{n}$Н1 vector of variance and covariance sample estimates, which are determined using observation results; **x** – column $\boldsymbol{m}$Н1 vector of unknown free model parameters of interest (viz.: variances and covariances for latent variables).

Now, let us consider the vector $\varepsilon=\mathbf{Ax}_*-\mathbf{b}$ that represents the residual of pseudosolution $\mathbf{x}_*$ of the overdetermined system obtained by the least- squares method. Assuming in the general case that components of the residual vector are correlated let us express their nonsingular covariance matrix as $\boldsymbol{\sigma}^2\mathbf{V}$.

[2] In path diagrams, ovals (circles) correspond to latent factors, rectangles correspond to observed variables, unidirectional arrows correspond to causal relationships, double-headed arrows correspond to covariances, variances or correlations.

After substitutions vector **b** and matrix **A** can be expressed the following way:

$$\mathbf{b}=\mathbf{V}^{½}\mathbf{b}_0 \text{ and } \mathbf{A}=\mathbf{V}^{½}\mathbf{A}_0 ,$$

where $\mathbf{V}=\mathbf{V}^{½}\mathbf{V}^{½}$. The only symmetric nonnegatively defined matrix $\mathbf{V}^{1/2}$, which is called the square root of **V**, exists for every symmetric nonnegatively defined covariance matrix **V**, so that $(\mathbf{V}^{1/2})^2=\mathbf{V}$.

Thus, let us turn to the set:

$$\mathbf{A}_0\mathbf{x}=\mathbf{b}_0,$$

for which the covariance matrix of the residual vector $\varepsilon_0= \mathbf{V}^{-½}\varepsilon$ looks like $\boldsymbol{\sigma}^2\mathbf{E}$ where **E** is identity matrix.

Whether now it is necessary to define is the given pseudosolution a maximal likelihood estimation. It will occur, if following conditions will be satisfied. First, the system should be nonsingular, that is *rank* **A**=***m***, where ***m*** is the number of free model parameters. Secondly, a vector are residual $\mathbf{e}_0$ should have multivariate normal distribution.

If these conditions are satisfied, the pseudosolution is a maximum likelihood estimation. The pseudosolution is calculated by the least squares method using the formula:

$$\mathbf{x}_* = (\mathbf{A}_0^T\mathbf{A}_0)^{-1}\mathbf{A}_0^T\mathbf{b}_0 = (\mathbf{A}^T\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{V}^{-1}\mathbf{b}.$$

At that the statistics

$$X^2=(\mathbf{b}_0-\mathbf{A}_0\,\mathbf{x}_*)^T(\mathbf{b}_0-\mathbf{A}_0\,\mathbf{x}_*)/s^2=(\mathbf{b}-\mathbf{A}\,\mathbf{x}_*)^T\mathbf{V}^{-1}(\mathbf{b}-\mathbf{A}\,\mathbf{x}_*)/s^2$$

has χ^2-distribution with ***n-m*** degrees of freedom.

Last statistics makes it possible to check the model validity. Under the assumptions indicated above, the presented statistics $\mathbf{X}^2$ makes it possible to test the hypothesis of representability of sample variances and covariances constituting the vector **b** with the aid of variances and covariances of latent variables contained in the model under study. Acceptance region is $\mathbf{X}^2\leq\chi^2_{n-m;\alpha}$ where α is criterion significance level.

At realization of the given approach the necessary simplifications shown in features of the decision are made:

1. Components of the residual vector e are assumed to be uncorrelated;
2. Mean-square deviation values of different components of the vector e are set equal to the same fixed proportion (percentage) of the corresponding components of the vector **b** (the hypothesis of proportionality);
3. The mentioned proportion (percentage) is selected to realize the equality $\mathbf{X}^2\leq\chi^2_{n-m;\alpha}$ at the significance level α=**0.05**, after that the admissibility of this quantity is evaluated. (It is convenient to evaluate the level of this characteristics having determined its reasonable critical value, for example 0.1. Thus, a new criterion (critical percentage) appears instead of the significance level.

Advantages of the suggested technique are:

- The solution of a problem is not reduced to the local multivariate optimization
- The new way of a choice of adequate model where the percent of mistakes is estimated via the estimation of a residual vector
- Since this method is direct there is no multiplicity of solutions
- No need in search of global minima.

Taking into account that the direct method of solution allows studying laws of interrelations easily, connections between free parameters of the factor model have been investigated, namely: numerical estimations of certain parameters for the given combinations of other ones with the aid of a matrix formula were carried out to reveal the dependencies of interest.

Since no special contingencies are provided for positiveness of calculated variance values and correct dependencies between connected covariances and variances, these quantities may possess incorrect values for some models. In this case vector $\mathbf{x}_c$ that meets the given conditions and is closest to pseudosolution $\mathbf{x}_*$ in Euclidian metrics is accepted as the result to be found. To calculate vector $\mathbf{x}_c$ simple linear optimization problem is solved. The hypothesis of representability of sample variances and covariances constituting the vector **b** with the aid of variances and covariances of latent variables constituting the vector $\mathbf{x}_c$ is verified, as before, by means of statistics

$$X^2=(\mathbf{b}-\mathbf{A}\ \mathbf{x}_r)^T\mathbf{V}^{-1}(\mathbf{b}-\mathbf{A}\ \mathbf{x}_r)/s^2$$

that has χ^2-distribution with ***n-m*** degrees of freedom.

As in the traditional structural equation modelling, the considered method also allows to make conclusions on statistical significance of different model components and judge about the importance of the model components under study using goodness-of-fit tests.

To do this one should compare X^2 statistics for two models: saturated model containing the component of interest and simplified model where this component is absent (equals to zero.) Let's denote hypothesis that the saturated model coincides with observation results as H_f. Significance level of the component of interest is revealed if there is no grounds to discard hypothesis H_f. At first one should estimate free parameters of the simplified model. The obtained value for X^2 statistics is compared with similar characteristics for the saturated model.

Since the difference in these statistics is asymptotically distributed as $\mathbf{c}^2$ with the number of degrees of freedom equal to the difference in degrees of freedom of saturated and simplified models, this difference is used to verify zero hypothesis H_r that the simplified model coincides with the observation results against alternative hypothesis H_f.

If H_r hypothesis is not discarded at the given significance level then the component under study is treated as statistically insignificant and the conclusion is made that the available data do not evidence the influence of the studied model part on the observed characteristic under consideration. If H_r hypothesis is discarded (and H_f hypothesis is accepted), then one can talk about the influence of the studied component on the given characteristic.

2.3. Alternative Factor Models for the Wavelet-Based Structural Equation Modelling

The ways of constructing new factor models for longitudinal studies instead of traditional simplex ones relying on the wavelet-based structural equation modelling include development of path coefficients factor models and variance components factor models as well as and their modifications. Typical examples of these two model variants are shown in Figures 3 and 4. Composition of wavelet coefficients to be analyzed in the capacity of observed variables depends on an application problem under consideration and may be varied. It is usually supposed that the number of time points under study equals to some power of 2.

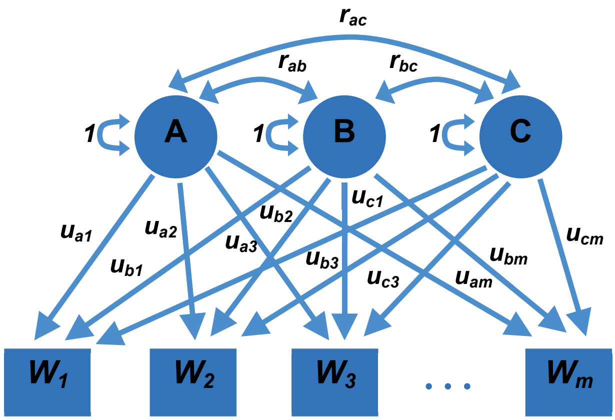

Figure 3. Path coefficients factor model represented by a path diagram

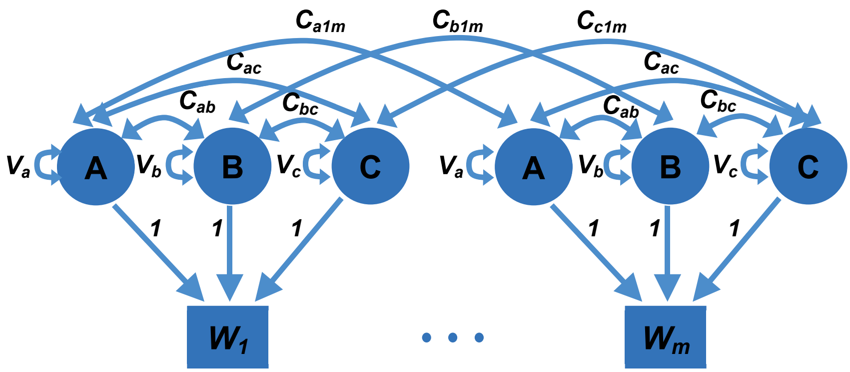

Figure 4. Variance components factor model represented by a path diagram

In case of path coefficients factor models, expressions for covariances and variances of wavelet coefficients W_i are non-linear:

$$Cov(W_i,W_j)=\sum_k \sum_l r_{kl} u_{ki} u_{lj} \ ;$$

$$Var(W_i)=\sum_k \sum_l r_{kl} u_{ki} u_{li} \ ,$$

where k and l are factor numbers, u_{**} – path coefficients, r_{**} – correlations between factors. These non-linearities make it impossible to get simple direct unambiguous estimations of free model parameters of interest. As contrasted to this, in case of variance components factor models, similar expressions are linear:

$$Cov(W_i,W_j)=\sum_k C_{kij} \ ;$$

$$Var(W_i) = \sum_k V_k + \sum_k \sum_l C_{kl} \; ,$$

where k and l are factor numbers, V_* – variances, C_{**} and C_{***} – covariances between factors. This fact makes it possible to obtain direct estimations of free model parameters using the alternative variant of the structural equation modelling described hereinbefore. Thus, it is the model type that may be used for solution of application problems in reality.

In practical situations, the basic variance components factor model generates a set of particular modifications representing problem peculiarities that are important for solution. For example, simultaneous analysis of different model groups can be useful for studying factor influences in case of several variants of observation conditions (see Figure 5).

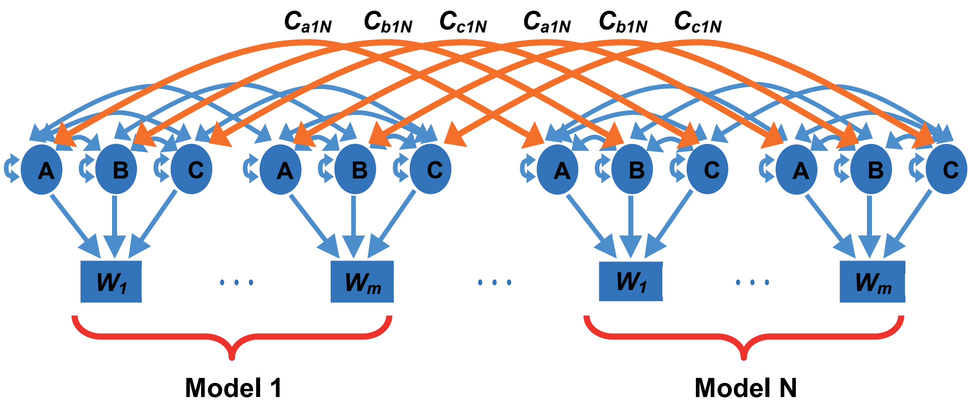

*Figure 5. Studying factor influences in case of several variants of observation conditions: simultaneous analysis of model groups. C*** are covariances between factors*

Typical representation of the wavelet-based structural equation modelling results destined for further interpretation includes:

- Factor variances and covariances estimated as free model parameters;
- Estimated correlations between different factors relevant to the same time points;
- Estimated correlations between the same factors relevant to different time points;
- Statistical significance estimations for different model components.

Corresponding examples can be found in papers [1, 20].

2.4. Model Singularity

If some model derived from an application domain yields system matrix $\mathbf{A}$ which rank is less than the number of free model parameters, pseudosolution $\mathbf{x}_* = (\mathbf{A}^T\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{V}^{-1}\mathbf{b}$ cannot be calculated properly because of singularity of system matrix $\mathbf{A}^T\mathbf{V}^{-1}\mathbf{A}$. In this case, one should reduce the number of free model parameters eliminating certain dependent variables to transform the given matrix into nonsingular form. Number of eliminated variables equals to the defect of matrix $\mathbf{A}^T\mathbf{V}^{-1}\mathbf{A}$. The following technique can be used to determine redundant quantities subjected to this operation:

– Solution of the eigenvalue problem for the matrix $\mathbf{A}^T\mathbf{V}^{-1}\mathbf{A}$, which is symmetric and nonnegatively definite, to obtain the proper subspace defined by the eigenvectors corresponding to nonzero eigenvalues of the given matrix
– Rotation of the obtained proper subspace basis keeping it within this subspace to attain maximal correspondence between directions of coordinate axes of the proper subspace and ones of the initial basis that formally results in transformation of coordinates of the proper subspace axes into either substantial or negligible values[3].

The axes of the initial basis, which are represented in the expressions of all the rotated basis directions by negligible coordinate values only, can be considered as lines that are approximately orthogonal with respect to the calculated nonzero proper subspace. Therefore, these lines approximately determine a subspace corresponding to the zero eigenvalues and, accordingly, define variables to be eliminated from the model to reproduce nonsingulary of the matrix in question. Since these quantities cause matrix singularity, they may be considered as dependent (redundant) ones. Their elimination turns into either expressing these variables via independent ones or assigning constant values to them and usually results in the transformed matrix nonsingularity.

If these transformations result in an obviously unacceptable model, one can keep the initial model representation and calculate an approximation of the pseudosolution using the Gauss-Seidel iteration method[4] (as well as other relevant approaches for singular system matrices).

2.5. Examples of models for practical applications

The approach under consideration was successfully applied to solution of the following problems:
– Assessing influence of maneuvering loads occurrences and climatic conditions of basing on aircraft damage accumulation rate [1]
– Estimating the balance of flight proficiency and psychophysiological crew status influences on the variances of the oculomotor activity characteristics under study [25].

Corresponding factor models represented in the form of path diagrams are shown in Figures 6-7.

First model is destined for 8-point dynamic series realizations of parameters under study. In the aforementioned examples, they represented rates of damage accumulation and results of psychological testing, correspondingly. Only four last wavelet coefficients (of eight) were used for analysis: this detail was conditioned by the application features. Influences of national features of pilotage technique (factors **R** and **F**) and influences of national environment exploitation (factors **D** and **A**) are under investigation in the first model, whereas influences of different sorts of original phonological awareness (factors **R**, **P** and **C**) and their associated distortions conditioned by test imperfection (factors $\mathbf{E_R}$, $\mathbf{E_P}$ and $\mathbf{E_C}$) are

[3] This standard procedure called Quartimax is usually available in the widespread statistical software packages.

[4] This method always converges in case of the symmetric nonnegatively definite system matrices under consideration.

under investigation in the second model. The analysis that has been carried out revealed causal and temporal relationships between the latent factors in question as well as statistical significance of different model components.

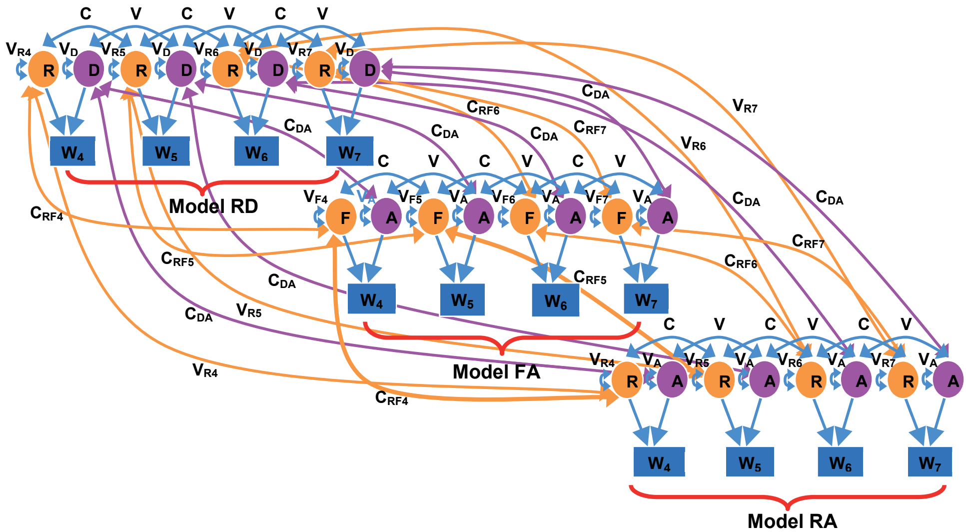

*Figure 6. Model to study influence of maneuvering load occurrences and climatic conditions of basing on aircraft damage accumulation rate: influences of national features of pilotage technique are represented by factors **R** and **F**, influences of national environment exploitation – by factors **D** and **A**, **V**$_*$ – variances, **C**$_{**}$ and **C**$_{***}$ – covariances between factors*

The factor model, which is represented in Figure 7, takes into account the effects of aircraft crew proficiency, its psychophysiological condition and measurement errors on oculomotor activity. Used as observed variables are jointly the following gaze movement primary indexes estimated during experiments at the Universal Crew Cockpit Prototyping Bench (Figure 8) with the aid of an eye tracker (Figure 9) for comparable maneuvers performed by the four pilots under study: duration of gaze fixations, rate of saccades and gaze entropy. This model takes into account: effects of pilot proficiency (N_1, N_2, N_3 and N_4) as well as combined effects of both pilot psychophysiological condition and measurement errors (S_1, S_2, S_3 and S_4) – on the observed estimates of the given gaze movement primary indexes, which are calculated for each of the four pilots under study during the time periods of comparable maneuver implementations (P_1, P_2, P_3 and P_4). Free model parameters to be identified are: factor loadings (n, s) and correlation coefficients (, ,). Values of the observed parameters were standardized using mean values and standard deviations. The factors S_1, S_2, S_3 and S_4 are assumed to be independent of each other.

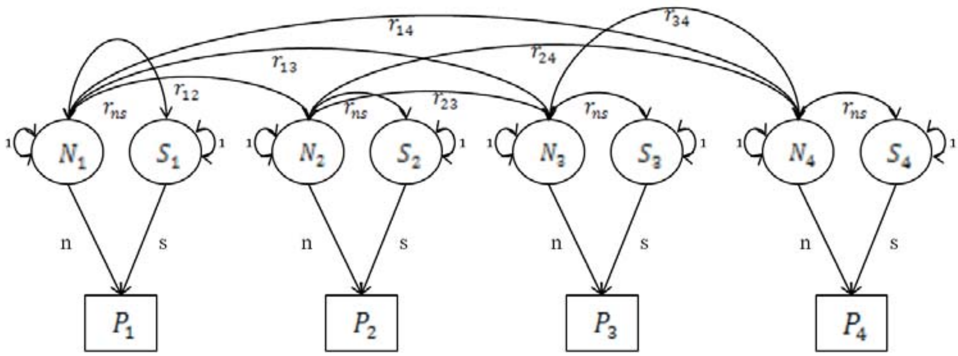

*Figure 7. Factor model components: effects of pilot proficiency (**N_1**, **N_2**, **N_3** and **N_4**) and combined effects of both pilot psychophysiological condition and measurement errors (**S_1**, **S_2**, **S_3** and **S_4**) – on the observed estimates of the given gaze movement primary indexes, which are calculated for each of the four pilots under study during the time periods of comparable maneuver implementations (**P_1**, **P_2**, **P_3** and **P_4**). Free model parameters to be identified are: factor loadings (**n**, **s**) and correlation coefficients (, ,)*

Figure 8. The Universal Crew Cockpit Prototyping Bench

Figure 9. Eye tracker to register oculomotor activity

To eliminate statistically insignificant components of the saturated model presented in Figure 7 and reveal the structure that is optimal from the viewpoint of its matching the observations, the following models were compared:

- Saturated model;
- Reduced model without correlations between corresponding factors $\boldsymbol{N_i}$ and $\boldsymbol{S_i}$ ($i=1,2,3,4$);
- Reduced model without correlations between corresponding factors $\boldsymbol{N_i}$ and $\boldsymbol{S_i}$ ($i=1,2,3,4$) and corresponding factors $\boldsymbol{N_i}$ and $\boldsymbol{N_j}$ ($i,j=1,2,3,4;\ i\neq j$);
- Reduced model without factors $\boldsymbol{N_i}$ ($i=1,2,3,4$).

Reduced model without factors $\boldsymbol{S_i}$ ($i=1,2,3,4$) were not under consideration since the application problem under study implies their obligatory presence.

Results of model fitting are shown in Table 1. For each model the hypothesis on its fit to observation data was tested, with statistics $\boldsymbol{F}$ being used as a goodness-of-fit measure.

Table 1.

Results of model fitting

Model	**Statistics *F***	**Degrees of freedom**	***p*-value**	**Model fit. Statistical significance of model reduction**
Saturated model	*8.857*	*5*	*0.115*	Model fit.
Reduced model without correlations between corresponding factors $\boldsymbol{N_i}$ and $\boldsymbol{S_i}$ ($i=1,2,3,4$) – optimal model	*8.857*	*6*	*0.182*	Model fit. Difference in ***F***-statistic between the saturated model and reduced model is not statistically significant

Model	Statistics *F*	Degrees of freedom	*p*-value	Model fit. Statistical significance of model reduction
Reduced model without correlations between corresponding factors $\boldsymbol{N}_i$ and $\boldsymbol{S}_i$ (*i=1,2,3,4*) and corresponding factors $\boldsymbol{N}_i$ and $\boldsymbol{N}_j$ (*i,j=1,2,3,4; i≠j*)	*17.390*	*8*	*0.026*	No model fit. Difference in ***F***-statistic between the optimal model and this reduced model is statistically significant
Reduced model without factors $\boldsymbol{N}_i$ (*i=1,2,3,4*)	*17.390*	*9*	*0.043*	No model fit. Difference in ***F***-statistic between the optimal model and this reduced model is statistically significant

Table 1 shows that both the saturated model and the reduced model without correlations between corresponding factors $\boldsymbol{N}_i$ and $\boldsymbol{S}_i$ (*i=1,2,3,4*) fit observation data ($p>0.05$). At the same time, the hypothesis on model fit should be rejected for the reduced model without correlations between corresponding factors $\boldsymbol{N}_i$ and $\boldsymbol{S}_i$ (*i=1,2,3,4*) and corresponding factors $\boldsymbol{N}_i$ and $\boldsymbol{N}_j$ (*i,j=1,2,3,4; i≠j*) as well as for the reduced model without factors $\boldsymbol{N}_i$ (*i=1,2,3,4*) (p<0.05).

Difference in *F*-statistic for the saturated and reduced models is asymptotically distributed as χ^2, with number of degrees of freedom being equal to the difference in their numbers of degrees of freedom. Thus, the reduced model without correlations between corresponding factors $\boldsymbol{N}_i$ and $\boldsymbol{S}_i$ (*i=1,2,3,4*) should be recognized as *optimal* one since there are no significant changes in the given statistic.

In case of the models for which the hypothesis on fit is rejected, the corresponding differences in *F*-statistic are statistically significant. Therefore, the indicated changes in factors and correlations degrade the model fit.

Since correlations between corresponding factors $\boldsymbol{N}_i$ and $\boldsymbol{S}_i$ (*i=1,2,3,4*) equal zero, the variances of observed variables consist of two components () representing, respectively, the factor of pilot proficiency and the combined factor of both pilot psychophysiological condition and measurement errors. The *n* and *s* values identified for the optimal model allow to conclude that the variance representing variability of the observed parameters is conditioned by 38% influence of pilot proficiency and 62% combined influence of both psychophysiological condition of pilots and measurement errors. This fact suggests the comparable influences of the factors under study on the analyzed characteristics.

3. THE EMPIRICAL APPROACH BASED ON WAVELET TRANSFORMS AND MULTIVARIATE STATISTICAL TECHNIQUES

Algorithmic aspects of this approach are represented by the following steps. It is assumed that the operator's activity is represented by a set of time series describing the dynamics of technical system parameters as well as operator's state, if possible.

Step 1. Preliminary processing: selection of time intervals for exercise comparison and data normalization. Selected are time series subsets corresponding to common time intervals, which are suitable for comparison of the same type exercises to be analyzed. Prior to the following computations, time series representing the history of performance of exercise completion are brought to a single scale, where the maximum is one and the minimum is zero.

Step 2. Elimination of redundant information. Where the data are broken down by sub exercises or represented by several measured parameters, the redundant information contained in the given time series is removed using *the Principal Components Analysis* [40, 45]. To do so, matrices of mutual correlations of time series values are computed, the algebraic problem of eigenvalues is solved and it is figured out to what extent it is possible to decrease the dimension of eigen subspace of the researched parameters so that this would contain a sufficiently representative (in practice, from 70% or higher) part of variability of the observed parameters. For each of the selected eigendirections of this subspace (principal components), for one of the highest component loads *a representative is elected from sub exercises or registered parameters (transition to the basis of principal components is impractical due to uncertain substantial interpretation of principal components and, in a number of applied tasks, due to no precise synchronization of researched processes for different exercises in time)*. The purpose of this stage is to leave only relatively independent characteristics replacing groups of significantly dependent representatives with only one characteristic parameter to avoid distortions stipulating to combined effect of strongly dependent characteristics in the subsequent phases.

Step 3. Transition to integral characteristics for time intervals by means of discrete wavelet transform. Time series representing the training processes under study are replaced with series of wavelet coefficients obtained as a result of *the Multiresolution Analysis* [39]. In this case, the original processes as functions of time are replaced with the integral characteristics of the time intervals, which are associated with these functions domain. In addition, significant saving (for about an order of magnitude) in the number of coefficients necessary for a correct representation of these process becomes available. *Due to the rules for assigning wavelet coefficients to time series fragments, which are in use in multiresolution analysis, problems associated with the need to precisely synchronize processes relating to various same-type training exercises in time have been cancelled since the most significant coefficients relating to relatively long time intervals are almost insensitive to moderate time shifts.* Length of the wavelet representations used during subsequent analysis stages can be significantly (approximately an order of magnitude) smaller than the length of the corresponding original time series, with no loss in the estimations accuracy.

Step 4. Computing mutual distance matrices. For each researched sub exercise or parameter one needs to compute the matrix of mutual distances between wavelet representations of source processes for different operator exercises, which are obtained in stage 3. Dimensions of such matrices are equal to the volume of the sample of analyzed exercises. Such matrices of mutual distances for all considered sub exercises are added forming the total matrix of mutual distances between the exercises under study.

Step 5. Multidimensional scaling to analyze the mutual allocation of operator exercises within a space with acceptable dimensions. The computed allocation of exercises under study in the resulting space of *the Multidimensional Scaling* [5, 42] is further used to define distances between the exercises to make diagnostic decisions. *The dimensions of the scaling space are defined based on the condition of sufficient differentiation of exercise samples relating to different recognizable classes.*

Step 6. Cluster analysis of patters in the obtained scaling space to reveal clusters representing various types of exercises and operator skill classes. Steps 5 and 6 are necessary to reasonably select clusters of abnormal exercises.

Step 7. Using the identification procedure, creation of probabilistic models represented by Markov random processes with discrete states and continuous time [6, 19, 21-27, 38] for each pattern cluster to represent probabilistic dynamics for each operator skill class to forecast probabilistic class behavior in two ways: by distribution of probabilities of being in model states and dynamics of mathematical expectations for each independent parameter determined in Step 2 (parameters are considered approximately as independent ones since they are calculated with the aid of the Principal Component Analysis).

Step 8. Computing distances to pattern cluster centers or to the pattern nearest to the operator based on results of completion of a sequence of test exercises. Where the volume of pattern samples is fairly large, distances are defined to centers of cluster patterns. These are computed based on multidimensional scaling data obtained at Step 5. Where pattern samples are small, the nearest pattern is defined, which definition may be done in two ways: either immediately through the computing of the pattern being nearest in the Euclidian metric of wavelet representation or through the identifying of the pattern in the resulting space of multidimensional scaling, such pattern being nearest in the Euclidian metric.

Step 9. Probabilistic assessments of operator skill class recognition using sample distribution functions of distances to cluster centers. Probabilistic assessments of recognition are defined using sample distribution functions of Euclidean distances to cluster centers of patterns belonging to the relevant recognizable operator skill class $i \in \{0, ..., z\}$ in a multidimensional scaling space. Computed values of , where is the Euclidean distance of i^{th} cluster in the space of multidimensional scaling, are interpreted as probabilistic assessments of belonging to the given classes. Their distribution among classes characterizes the reliability of the obtained classification. In fact, *the given approach implements the idea of the linear discriminant analysis in an generalized form*. But unlike the latter there are no restrictions on observed data distributions.

Step 10. Probabilistic assessments of operator skill class attribution using Bayesian likelihood estimations. These estimations are calculated basing on identified probabilistic models.

Step 11. Forecasting behavior of the parameters selected at Step 2 for the operator skill class cluster recognized at Steps 9 and 10 with the aid of probabilistic models created at Step 7 by both distribution of probabilities of being in model states and dynamics of mathematical expectations for each independent parameter to estimate probability of hitting into dangerous situations associated with the certain intervals of the parameters under consideration.

The techniques, which implementation results are presented below, have been developed by the authors of this paper and presented in works [29-32]. The *Intelligent System for Flight Analysis (ISFA)* implementing the developed approache with the aid of the LabVIEW graphical programming environment provided the required calculations. ISFA has been officially registered at the Russian Patent Agency ("ROSPATENT") [28].

Partial comparisons to estimate parameter's contributions to the mutual distances are available in case of small exercise sample sizes. In case of greater sample sizes, plural comparisons to get Bayesian likelihood estimations for operator skill clusters attribution, which contain parameter's contributions, are available.

The actions, which an expert is responsible for, are:

- Assessment to select clusters of abnormal exercises;
- Analysis of parameter's balance to select relevant parameters to be under study;
- Identification and interpretation of abnormalities and mistakes.

4. PRINCIPAL RESULTS AND CONCLUSIONS

1. The wavelet-based structural equation modelling, which combines capabilities of wavelet transforms and trained factor structures, can be used to reveal the influence of human factor on the activity of operators of complex technical systems. According to the proposed approach, the samples of coefficients resulted from discrete wavelet transform of initial parameter time series under study and responsible for different observation periods are considered as values of observed variables in the subsequent structural equation modelling to reveal time history of factor influences and estimates of factor interaction.
2. Identification of free factor model parameters (usually factor variances and covariances) can be carried out by a direct (noniterative) procedure based on the maximum likelihood method that is an alternative to traditional ambiguous local iterative solutions of multivariate optimization problems, which depend on the initial approximations. Main features of this procedure are:
   - Composing overdetermined sets of equations as respects free factor model parameters and their following solution by the maximum likelihood method;
   - Using variance components path model;
   - Using new criterion instead of traditional significance level for testing model representability of observation results, viz.: the critical percentage of mean-square devia-

tion values of residual vector components for the corresponding observed variances and covariances;
- Possibility of significance tests for model components using statistical goodness-of-fit measures.

3. With the operator's activity being represented by a set of time series describing the dynamics of technical system parameters as well as operator's state, if possible, the empirical approach which is based on wavelet transforms and multivariate statistical techniques, makes it possible to:
   - Support the outcome grading for current exercises by means of comparing their parameters with the exercise patterns collected beforehand in the activity record database;
   - Carry out quantitative and qualitative assessments of operator skill class, which are based on comparisons of current exercises with the activity database patterns in the wavelet representation metric associated with observed parameters (basic technique) as well as on probabilistic assessments of skill class recognition using sample distribution functions of exercise distances to cluster centers in a scaling space (supportive technique) and Bayesian likelihood estimations with the aid of probabilistic profile of staying in activity parameter ranges (also supportive technique).
4. The multidimensional scaling and cluster analysis of patters in an obtained scaling space can be used to reasonably select abnormal exercises. The techniques in use provide both discrimination in a scaling space between the trial types and correct skill assessments.
5. The techniques in use provide possibility for creating certain classification rules to separate different scale levels of trial quality assessment in a scaling space. Wherein cluster differences for an exercise type can be explained by the exercise implementation resulted from individual skills.
6. The pattern selection technique in use makes it possible to ensure assessments of acceptable correctness.
7. Both the technique for probabilistic recognition of skill classes via Bayesian likelihood estimations (when a skill class is selected with the aid of probabilistic profile of staying in flight parameter ranges) and the technique of calculating parameter's contributions to the mutual distances in the wavelet coefficient metric in pairwise exercises comparison, have demonstrated the capabilities of:
   - Recognizing sets of abnormal exercises in the scaling spaces with the wavelet coefficient metric,
   - Detection of parameters characterizing operator mistakes to reveal the causes of abnormality.
8. The given empirical approach can be applied even if the exercise sample sizes are small since both the pattern selection and calculation of the parameter's contributions to the mutual distances in the wavelet coefficient metric in pairwise exercises comparison are available in this case.
9. The factor analysis shows comparable influences of the factor of pilot proficiency and the combined factor of both pilot psychophysiological condition and measurement errors on the analyzed primary indexes of oculomotor activity.

Acknowledgement

This work was performed as a part of the “SAFEMODE” Project (Grant Agreement No 814961) with the financial support of the Ministry of Science and Higher Education of the Russian Federation (Project UID RFMEFI62819X0014).

References

1. Baranov S.N., Kuravsky L.S., Baranov N.I. Studying influence of maneuvering loads occurrences and climatic conditions of basing on aircraft damage accumulation rate with the aid of trained structures. In: Proc. 5th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, Edinburgh, Scotland, United Kingdom, July 2008.
2. Bastani V., Marcenaro L., Regazzoni C. Unsupervised trajectory pattern classification using hierarchical Dirichlet Process Mixture hidden Markov model. 2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP) / IEEE. 2014. Pp. 1–6.
3. Bishop Y.M.M., Fienberg S.E., Holland P.W. Discrete multivariate analysis: Theory and practice. (Cambridge, MA, M I T Press, 1975.)
4. Bollen K.A. Structural equations with latent variables. (New York, John Wiley, 1989.)
5. Borg, I., Groenen, P. J. F. Modern Multidimensional Scaling Theory and Applications. – Springer, 2005. – P.140.
6. Cramer H. Mathematical Methods of Statistics. Princeton: Princeton University Press. 1999. – 575 pp.
7. Eerland W.J., Box S. Trajectory Clustering, Modelling and Selection with the focus on Airspace Protection. AIAA Infotech@ Aerospace. _ AIAA, 2016. Pp. 1–14.
8. Enriquez M. Identifying temporally persistent flows in the terminal airspace via spectral clustering. Tenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2013). Federal Aviation Administration (FAA) and EUROCONTROL. Chicago, IL, USA: 2013. June 10-13.
9. Enriquez M., Kurcz C. A Simple and Robust Flow Detection Algorithm Based on Spectral Clustering. International Conference on Research in Air Transportation (ICRAT). Federal Aviation Administration (FAA) and EUROCONTROL. – Berkeley, CA, USA: 2012. May 22-25.
10. Gaffney S., Smyth P. Joint probabilistic curve clustering and alignment. In Advances in Neural Information Processing Systems. Vol. 17. Cambridge, MA: MIT Press, 2005. Pp. 473–480.
11. Gaffney S., Smyth P. Trajectory clustering with mixtures of regression models. Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining. 1999. Pp. 63–72.
12. Gariel M., Srivastava A., Feron E. Trajectory clustering and an application to airspace monitoring. IEEE Transactions on Intelligent Transportation Systems. 2011. Vol. 12, no. 4. Pp. 1511–1524.
13. Goldstein H. Multilevel statistical models. (3rd ed., London, Arnold, 2003.)
14. Grevtsov N. Synthesis of control algorithms for aircraft trajectories in time optimal climb and descent. Journal of Computer and Systems Sciences International. 2008. Vol. 47, no. 1. Pp. 129–138.
15. Hung C., Peng W., Lee W. Clustering and aggregating clues of trajectories for mining trajectory patterns and routes. The VLDB Journal – The International Journal on Very Large Data Bases. 2015. Vol. 24, no. 2. Pp. 169–192.
16. Jöreskog K.G. Estimation and testing of simplex models. British Journal of Mathematical and Statistical Psychology, Vol. 23, pp.121-145, 1970.
17. Krasilshchikov M.N., Evdokimenkov V.N., Bazlev D.A. Individually adapted airborne systems for monitoring the aircraft technical condition and supporting the pilot control actions. – Moscow, MAI Publishing House, 2011. – 440 pp (in Russian).
18. Kuchar J. K., Yang L. C. A review of conflict detection and resolution modeling methods. IEEE Transactions on Intelligent Transportation Systems. 2000. Vol. 1, no. 4. Pp. 179–189.
19. Kuravsky L.S., Artemenkov S.L., Yuriev G.A., Grigorenko E.L. New approach to computer-based

adaptive testing. – Experimental Psychology. 2017. Vol. 10. No. 3. Pp. 33–45. doi:10.17759/exppsy.2017100303.

20. Kuravsky L.S., Baranov S.N. Development of the wavelet-based confirmatory factor analysis for monitoring of system factors. – In: Proc. 5th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies, Edinburgh, United Kingdom, July 2008, pp.818-834.
21. Kuravsky L.S., Margolis A.A., Marmalyuk P.A., Panfilova A.S. , Yuriev G.A. Mathematical aspects of the adaptive simulator concept – Psychological Science and Education. 2016. Vol. 21. No. 2. Pp. 84–95. doi: 10.17759/pse.2016210210 (in Russian).
22. Kuravsky L.S., Margolis A.A., Marmalyuk P.A., Panfilova A.S., Yuryev G.A., Dumin P.N. A Probabilistic Model of Adaptive Training. – Applied Mathematical Sciences, Vol. 10, 2016, no. 48, 2369 – 2380, http://dx.doi.org/10.12988/ams.2016.65168.
23. Kuravsky L.S., Marmalyuk P.A., Yurev G.A. Diagnostics of Professional Skills Based on Probability Distributions of Oculomotor Activity. – RFBR Journal, No. 3 (91), 2016, pp.72-82 (Supplement to "Information Bulletin of RFBR" No. 24, in Russian).
24. Kuravsky L.S., Marmalyuk P.A., Yuryev G.A. and Dumin P.N. A Numerical Technique for the Identification of Discrete-State Continuous-Time Markov Models – Applied Mathematical Sciences. Vol. 9, 2015, No. 8, pp. 379–391. URL: http://dx.doi.org/10.12988/ams. 2015.410882.
25. Kuravsky L.S., Marmalyuk P.A., Yuryev G.A., Belyaeva O.B. and Prokopieva O.Yu. Mathematical Foundations of Flight Crew Diagnostics Based on Videooculography Data. – Applied Mathematical Sciences, Vol. 10, 2016, no. 30, 1449–1466, http://dx.doi.org/10.12988/ams.2016.6122.
26. Kuravsky L.S., Marmalyuk P.A., Yuryev G.A., Dumin P.N., Panfilova A.S. Probabilistic modeling of CM operator activity on the base of the Rasch model. – In: Proc. 12th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies, Oxford, UK, June 2015.
27. Kuravsky L.S., Yuriev G.A. Probabilistic method of filtering artefacts in adaptive testing. – Experimental Psychology, Vol.5, No. 1, 2012, pp. 119-131 (in Russian).
28. Kuravsky L.S., Yuryev G.A. Certificate of state registration of the computer program №2018660358 Intelligent System for Flight Analysis v1.0 (ISFA#1.0). – Application №2018617617; declared 18 July 2018; registered 22 August 2018. – (ROSPATENT).
29. Kuravsky L.S., Yuryev G.A. Detecting abnormal activities of operators of complex technical systems and their causes basing on wavelet representations. – International Journal of Civil Engineering and Technology (IJCIET). 10(2), 2019, pp. 724–742. http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=10&IType=2.
30. Kuravsky L.S., Yuryev G.A. On the approaches to assessing the skills of operators of complex technical systems. – In: Proc. 15th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies, Nottingham, UK, September 2018. – 25 pp.
31. Kuravsky L.S., Yuriev G.A., Dumin P.N.. Estimating the influence of human factor on the activity of operators of complex technical systems in civil engineering with the aid of adaptive diagnostics, International Journal of Civil Engineering and Technology (IJCIET). 10(2), 2019, pp. 1930-1941. http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=10&IType=02.
32. Kuravsky L.S., Yuryev G.A., Zlatomrezhev V.I. New approaches for assessing the activities of operators of complex technical systems. Eksperimental'naya psikhologiya = Experimental psychology (Russia), 2019, vol. 12, no. 4, pp. 27–49. doi:10.17759/exppsy.2019120403.
33. Lawley D.N., Maxwell A.E. Factor analysis as a statistical method. (London, Butterworths, 1971, 2nd ed., 153 pp).
34. Laxhammar R., Falkman G. Online learning and sequential anomaly detection in trajectories. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2014. Vol. 36, no. 6. Pp. 1158–1173.

35. Li Z., et al. Incremental clustering for trajectories. Database Systems for Advanced Applications. Lecture Notes in Computer Science. 2010. Vol. 5982. Pp. 32–46.
36. Liu H., Li J. Unsupervised multi-target trajectory detection, learning and analysis in complicated environments. 2012 21st International Conference on Pattern Recognition (ICPR) / IEEE. 2012. Pp. 3716–3720.
37. Loehlin J.C. Latent variable models: An introduction to factor, path, and structural analysis. (Hillsdale, NJ, Erlbaum, 1987.)
38. Markov models in the diagnostics and prediction problems: Textbook. /Edited by L.S. Kuravsky. – 2nd Edition, Enlarged. – Moscow: MSUPE Edition, 2017. – 203 pp. (in Russian).
39. Neal P.G. Multiresolution Analysis for Adaptive Refinement of Multiphase Flow Computations. – University of Iowa, 2010. –116 pp.
40. René Vidal, Yi Ma, Shankar Sastry. Generalized Principal Component Analysis. – Springer-Verlag: New York, 2016.URL: http://www.springer.com/ us/book/9780387878102
41. Rintoul M., Wilson A. Trajectory analysis via a geometric feature space approach. Statistical Analysis and Data Mining: The ASA Data Science Journal. 2015.
42. Trevor F. Cox, M.A.A. Cox. Multidimensional Scaling, Second Edition. – Chapman & Hall/CRC, 2001. – P. 299.
43. Wei J., et al. Design and Evaluation of a Dynamic Sectorization Algorithm for Terminal Airspace. Journal of Guidance, Control, and Dynamics. 2014. Vol. 37, no. 5. Pp. 1539–1555.
44. Wilson A., Rintoul M., Valicka C. Exploratory Trajectory Clustering with Distance Geometry. International Conference on Augmented Cognition. – Springer. 2016. Pp. 263–274.
45. Xiangyu Kong, Changhua Hu, Zhansheng Duan. Principal Component Analysis Networks and Algorithms. – Springer, 2017. URL: http://www.springer.com/us/book/9789811029134.

Оценка вклада человеческого фактора в эксплуатационные характеристики сложных технических систем

Куравский Л.С.*
МГППУ, Москва, Россия,
ORCID: http://orcid.org/0000-0002-3375-8446,
e-mail: l.s.kuravsky@gmail.com

Юрьев Г.А.**
МГППУ, Москва, Россия,
ORCID: https://orcid.org/0000-0002-2960-6562,
e-mail: g.a.yuryev@gmail.com

Златомрежев В.И. **
ГосНИИАС, Москва, Россия,
ORCID https://orcid.org/0000-0003-1776-6881
e-mail: vizlatomr@2100.gosniias.ru

Юрьева Н.Е.****
МГППУ, Москва, Россия,
ORCID: https://orcid.org/0000-0003-1419-876X,
e-mail: yurieva.ne@gmail.com

Михайлов А.Ю.*****
ГосНИИАС, Москва, Россия,
ORCID: https://orcid.org/0000-0003-0278-1819
e-mail: aymihaylov@2100.gosniias.ru

Для цитаты:
Куравский Л.С., Юрьев Г.А., Златомрежев В.И., Юрьева Н.Е., Михайлов А.Ю. Оценка вклада человеческого фактора в эксплуатационные характеристики сложных технических систем // Моделирование и анализ данных. 2020. Том 10. № 1. С. 7–34. DOI: 10.17759/mda.2020100101 (In Engl., abstr. In Russ.)

Куравский Лев Семёнович, доктор технических наук, профессор, декан факультета информационных технологий, Московский государственный психолого-педагогический университет, Москва, Россия, ORCID: http://orcid.org/0000-0002-3375-8446, e-mail: l.s.kuravsky@gmail.com

****Юрьев Григорий Александрович,*** кандидат физико-математических наук, доцент, заведующий научной лабораторией, Московский государственный психолого-педагогический университет, Москва, Россия, ORCID: https://orcid.org/0000-0002-2960-6562, e-mail: g.a.yuryev@gmail.com

*****Златомрежев Валентин Игоревич,*** Заведующий лабораторией, Государственный научно-исследовательский институт авиационных систем (ГосНИИАС), Москва, Россия, ORCID https://orcid.org/0000-0003-1776-6881, E-mail: vizlatomr@2100.gosniias.ru

******Юрьева Наталия Евгеньевна,*** кандидат технических наук, научный сотрудник, Московский государственный психолого-педагогический университет, Москва, Россия, ORCID: https://orcid.org/0000-0003-1419-876X, e-mail: yurieva.ne@gmail.com

Оценка влияния человеческого фактора на деятельность операторов сложных технических систем является важной задачей для мониторинга состояния, подготовки и диагностики персонала. Представлены обзор и взаимные сравнения подходов, которые используются для оценки влияния человеческого фактора и уже показали свою эффективность в практическом применении. Рассматриваются: моделирование структурными уравнениями (конфирматорный факторный анализ), байесовские оценки вероятностных моделей, представленные марковскими случайными процессами, многомерные статистические методы, включающие дискриминантный и кластерный анализ, а также вейвлет-преобразования.

Ключевые слова: Операторы сложных технических систем, человеческий фактор, мониторинг состояния, факторный анализ, конфирматорный факторный анализ, марковские случайные процессы, вейвлет-преобразование, многомерные статистические методы, анализ главных компонент, многомерное шкалирование, кластерный анализ.

*****_**Михайлов Артём Юрьевич,**_ инженер 1 кат, Государственный научно-исследовательский институт авиационных систем (ГосНИИАС), Москва, Россия, ORCID: https://orcid.org/0000-0003-0278-1819, E-mail: aymihaylov@2100.gosniias.ru

Благодарность

Работа выполнена как часть проекта «SAFEMODE» (грант № 814961) при финансовой поддержке Министерства науки и высшего образования Российской Федерации (проект UID RFMEFI62819X0014).

Литература

1. Baranov S.N., Kuravsky L.S., Baranov N.I. Studying influence of maneuvering loads occurrences and climatic conditions of basing on aircraft damage accumulation rate with the aid of trained structures. In: Proc. 5th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, Edinburgh, Scotland, United Kingdom, July 2008.
2. Bastani V., Marcenaro L., Regazzoni C. Unsupervised trajectory pattern classification using hierarchical Dirichlet Process Mixture hidden Markov model//2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)/IEEE. 2014. Pp. 1–6.
3. Bishop Y.M.M., Fienberg S.E., Holland P.W. Discrete multivariate analysis: Theory and practice. (Cambridge, MA, M I T Press, 1975.)
4. Bollen K.A. Structural equations with latent variables. (New York, John Wiley, 1989.)
5. Borg, I., Groenen, P. J. F. Modern Multidimensional Scaling Theory and Applications. – Springer, 2005. – P.140.
6. Cramer H. Mathematical Methods of Statistics. Princeton: Princeton University Press. 1999. – 575 pp.
7. Eerland W.J., Box S. Trajectory Clustering, Modelling and Selection with the focus on Airspace Protection //AIAA Infotech@ Aerospace. _ AIAA, 2016. Pp. 1–14.
8. Enriquez M. Identifying temporally persistent flows in the terminal airspace via spectral clustering // Tenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2013)/Federal Aviation Administration (FAA) and EUROCONTROL. Chicago, IL, USA: 2013. June 10–13.
9. Enriquez M., Kurcz C. A Simple and Robust Flow Detection Algorithm Based on Spectral Clustering // International Conference on Research in Air Transportation (ICRAT)/Federal Aviation Administration (FAA) and EUROCONTROL. – Berkeley, CA, USA: 2012. May 22–25.

10. Gaffney S., Smyth P. Joint probabilistic curve clustering and alignment//In Advances in Neural Information Processing Systems. Vol. 17. Cambridge, MA: MIT Press, 2005. Pp. 473–480.
11. Gaffney S., Smyth P. Trajectory clustering with mixtures of regression models//Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining. 1999. Pp. 63–72.
12. Gariel M., Srivastava A., Feron E. Trajectory clustering and an application to airspace monitoring// IEEE Transactions on Intelligent Transportation Systems. 2011. Vol. 12, no. 4. Pp. 1511–1524.
13. Goldstein H. Multilevel statistical models. (3rd ed., London, Arnold, 2003.)
14. Grevtsov N. Synthesis of control algorithms for aircraft trajectories in time optimal climb and descent // Journal of Computer and Systems Sciences International. 2008. Vol. 47, no. 1. Pp. 129–138.
15. Hung C., Peng W., Lee W. Clustering and aggregating clues of trajectories for mining trajectory patterns and routes//The VLDB Journal – The International Journal on Very Large Data Bases. 2015. Vol. 24, no. 2. Pp. 169–192.
16. Jöreskog K.G. Estimation and testing of simplex models. British Journal of Mathematical and Statistical Psychology, Vol. 23, pp.121–145, 1970.
17. Krasilshchikov M.N., Evdokimenkov V.N., Bazlev D.A. Individually adapted airborne systems for monitoring the aircraft technical condition and supporting the pilot control actions. – Moscow, MAI Publishing House, 2011. – 440 pp. (in Russian).
18. Kuchar J. K., Yang L. C. A review of conflict detection and resolution modeling methods//IEEE Transactions on Intelligent Transportation Systems. 2000. Vol. 1, no. 4. Pp. 179–189.
19. Kuravsky L.S., Artemenkov S.L., Yuriev G.A., Grigorenko E.L. New approach to computer-based adaptive testing. – Experimental Psychology. 2017. Vol. 10. No. 3. Pp. 33–45. doi:10.17759/exppsy.2017100303.
20. Kuravsky L.S., Baranov S.N. Development of the wavelet-based confirmatory factor analysis for monitoring of system factors. – In: Proc. 5th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies, Edinburgh, United Kingdom, July 2008, pp. 818–834.
21. Kuravsky L.S., Margolis A.A., Marmalyuk P.A., Panfilova A.S. , Yuriev G.A. Mathematical aspects of the adaptive simulator concept – Psychological Science and Education. 2016. Vol. 21. No. 2. Pp. 84–95. Doi: 10.17759/pse.2016210210 (in Russian).
22. Kuravsky L.S., Margolis A.A., Marmalyuk P.A., Panfilova A.S., Yuryev G.A., Dumin P.N. A Probabilistic Model of Adaptive Training. – Applied Mathematical Sciences, Vol. 10, 2016, no. 48, 2369–2380, http://dx.doi.org/10.12988/ams.2016.65168.
23. Kuravsky L.S., Marmalyuk P.A., Yurev G.A. Diagnostics of Professional Skills Based on Probability Distributions of Oculomotor Activity. – RFBR Journal, No. 3 (91), 2016, pp. 72–82 (Supplement to "Information Bulletin of RFBR" No. 24, in Russian).
24. Kuravsky L.S., Marmalyuk P.A., Yuryev G.A. and Dumin P.N. A Numerical Technique for the Identification of Discrete-State Continuous-Time Markov Models – Applied Mathematical Sciences. Vol. 9, 2015, No. 8, pp. 379–391. URL: http://dx.doi.org/10.12988/ams. 2015.410882.
25. Kuravsky L.S., Marmalyuk P.A., Yuryev G.A., Belyaeva O.B. and Prokopieva O.Yu. Mathematical Foundations of Flight Crew Diagnostics Based on Videooculography Data. – Applied Mathematical Sciences, Vol. 10, 2016, no. 30, 1449–1466, http://dx.doi.org/10.12988/ams.2016.6122.
26. Kuravsky L.S., Marmalyuk P.A., Yuryev G.A., Dumin P.N., Panfilova A.S. Probabilistic modeling of CM operator activity on the base of the Rasch model. – In: Proc. 12th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies, Oxford, UK, June 2015.
27. Kuravsky L.S., Yuriev G.A. Probabilistic method of filtering artefacts in adaptive testing. – Experimental Psychology, Vol.5, No. 1, 2012, pp. 119–131 (in Russian).
28. Kuravsky L.S., Yuryev G.A. Certificate of state registration of the computer program №2018660358 Intelligent System for Flight Analysis v1.0 (ISFA#1.0). – Application №2018617617; declared 18 July 2018; registered 22 August 2018. – (ROSPATENT).

29. Kuravsky L.S., Yuryev G.A. Detecting abnormal activities of operators of complex technical systems and their causes basing on wavelet representations. – International Journal of Civil Engineering and Technology (IJCIET). 10(2), 2019, pp. 724–742. http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=10&IType=2.
30. Kuravsky L.S., Yuryev G.A. On the approaches to assessing the skills of operators of complex technical systems. – In: Proc. 15th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies, Nottingham, UK, September 2018. – 25 pp.
31. Kuravsky L.S., Yuriev G.A., Dumin P.N., Estimating the influence of human factor on the activity of operators of complex technical systems in civil engineering with the aid of adaptive diagnostics, International Journal of Civil Engineering and Technology (IJCIET). 10(2), 2019, pp. 1930–1941. http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=10&IType=02.
32. Kuravsky L.S., Yuryev G.A., Zlatomrezhev V.I. New approaches for assessing the activities of operators of complex technical systems. Eksperimental'naya psikhologiya = Experimental psychology (Russia), 2019, vol. 12, no. 4, pp. 27–49. doi:10.17759/exppsy.2019120403.
33. Lawley D.N., Maxwell A.E. Factor analysis as a statistical method. (London, Butterworths, 1971, 2nd ed., 153 pp).
34. Laxhammar R., Falkman G. Online learning and sequential anomaly detection in trajectories // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2014. Vol. 36, no. 6. Pp. 1158–1173.
35. Li Z., et al. Incremental clustering for trajectories//Database Systems for Advanced Applications. Lecture Notes in Computer Science. 2010. Vol. 5982. Pp. 32–46.
36. Liu H., Li J. Unsupervised multi-target trajectory detection, learning and analysis in complicated environments//2012 21st International Conference on Pattern Recognition (ICPR)/IEEE. 2012. Pp. 3716–3720.
37. Loehlin J.C. Latent variable models: An introduction to factor, path, and structural analysis. (Hillsdale, NJ, Erlbaum, 1987.)
38. Markov models in the diagnostics and prediction problems: Textbook./Edited by L.S. Kuravsky. – 2nd Edition, Enlarged. – Moscow: MSUPE Edition, 2017. – 203 pp. (in Russian).
39. Neal P.G. Multiresolution Analysis for Adaptive Refinement of Multiphase Flow Computations. – University of Iowa, 2010. – 116 pp.
40. René Vidal, Yi Ma, Shankar Sastry. Generalized Principal Component Analysis. – Springer-Verlag: New York, 2016.URL: http://www.springer.com/ us/book/9780387878102
41. Rintoul M., Wilson A. Trajectory analysis via a geometric feature space approach//Statistical Analysis and Data Mining: The ASA Data Science Journal. 2015.
42. Trevor F. Cox, M.A.A. Cox. Multidimensional Scaling, Second Edition. – Chapman & Hall/CRC, 2001. – P. 299.
43. Wei J., et al. Design and Evaluation of a Dynamic Sectorization Algorithm for Terminal Airspace // Journal of Guidance, Control, and Dynamics. 2014. Vol. 37, no. 5. Pp. 1539–1555.
44. Wilson A., Rintoul M., Valicka C. Exploratory Trajectory Clustering with Distance Geometry // International Conference on Augmented Cognition/Springer. 2016. Pp. 263–274.
45. Xiangyu Kong, Changhua Hu, Zhansheng Duan. Principal Component Analysis Networks and Algorithms. – Springer, 2017. URL: http://www.springer.com/us/book/9789811029134.