Applicability of statistical modelling in problems of cardiopulmonary resuscitation device’s state control

K Osipov\textsuperscript{1,*}, A Krugovoy\textsuperscript{1} and V Novikov\textsuperscript{2}

\textsuperscript{1}Sevastopol state university, Sevastopol, Russia
\textsuperscript{2}Black Sea Higher Naval School n.a. P.S. Nakhimov

*assistanttmm@mail.ru

Abstract. The methods of experimental study of complex machines cardiopulmonary resuscitation are developed. They involve the statistical modeling of the tested objects on the basis of technical parameter values and adaptive filtering, utilizing Kullback-Leibler information measure as a criterion of the state estimation. The measuring information on technical parameters obtained during their tests is analyzed as an example.

Experimental study of complex systems based on statistical analysis of measurement data is a known tool for their modeling, especially at incomplete information on their properties and behavior [1]. However, the universal methodology of experimental and statistical approach to the study of technical objects has not been established so far, although it would be useful for many applications, such as quality control of complex machines and mechanisms after assembly.

The aim is to develop the methods of experimental study of complex machines the functioning of which is accompanied by various physical processes, during tests to make decision on preparation for exploitation. These can be cardiopulmonary resuscitation device, compressors, gas turbines, internal combustion engines (ICE). The method involves the statistical modeling of the tested objects on the basis of technical parameter values and adaptive filtering, utilizing Kullback-Leibler information measure as a criterion of the state estimation [2]. The measuring information on engine diagnostic parameters obtained during their acceptance tests is analyzed as an example.

The tests are one of the most important stages of the complex machines (cardiopulmonary resuscitation device, engine) quality control. During the tests compliance of quality characteristics (reliability, security, etc.) with the technical requirements is inspected, the values technical parameters, which are representing the state of the complex machines (cardiopulmonary resuscitation device, engine), are controlled, and decisions on engines’ availability for exploitation are made. For example, for gasoline automobile engines such parameters are power, fuel consumption, oil pressure, etc.

For some machines, for example, engines the costs of the tests and control are 25 - 30% of their value, thus a development of new methods of testing and analysis of measurement information focused on cost reducing at improving the accuracy and reliability of test results, is still relevant scientific and industrial problem.

For example, the most probable manufacturing defects for engines are: depressurization of the space over pistons, lubrication and cooling systems, increasing or decreasing the mounting gap, the gap rubber seals, the curvature of the longitudinal axis of the revolution body, connecting rods, etc. They cause a decrease in gas pressure on the piston blow-by into the crankcase on the steps of...
compression, combustion and expansion, an increase in the mechanical friction losses, reduced efficiency, and deterioration of the process of filling the combustion chamber with fresh charge. As a consequence, engine power is reduced, fuel consumption is increased, and the concentration of toxic compounds in the flue gas grows.

To estimate the technical state of new engines at the tests accurately and to make right decisions on their availability for exploitation is necessary to measure fuel consumption $g_e$ (g / kWh), torque $M_k$, fuel flow rate $G_t$ (kg / h), and concentration of CO and CH$_4$ in the exhaust gases (Fig. 1) [3].

![Figure 1. Measured parameters of engine.](image)

Monitoring of the parameters should be performed at changing modes of engine through changing rotation velocity of cranked shaft from 150 to 350 s$^{-1}$ and at variable (0.2–0.7$M_k$) load.

The measurement of fuel consumption and fuel rate permits to determine the mechanical losses and to estimate the state of bearings’ interfaces and ICE adjustment accuracy. The torque values are generalized index, reflecting engine availability to functioning. Concentration of CH$_4$ in the exhaust gas characterizes the technical state of the cylinder group, and the concentration of CO characterizes a state of the power supply system. During the tests the analyzed information is sequences of random values of the measured parameters, or discrete random processes. The main problem of their analysis is to identify dynamic trends. In some cases they can be clearly traced, but more often they are not visible due to different noises. Under conditions of modes’ changing it is necessary to smooth
influence of periodic components of the processes. It can be achieved by spectral analysis through decomposition of autocorrelation function into Fourier series.

As it is known, spectral density is a continuous non-negative function and is associated with the theoretical autocorrelation function \( \rho(k) \) by the relation:

\[
I(\omega) = D[1 + 2\rho_1 \cos(\pi\omega) + 2\rho_2 \cos(2\pi\omega) + ...],
\]

where \( D \) is a process variance. Autocorrelation function and the spectral density are mathematically equivalent and they are mutual transformants [3]. Here the smoothed periodogram is used to estimate the spectrum. For non-stationary processes with a smooth trend periodogram contains a sharp at the low frequencies and reflects the deterministic frequency with a long period. The peaks of the periodogram values reflect the distribution of the variance between different harmonic components (Fig. 2). Spectral analysis of the processes shows a maximal peak at the periodogram corresponding to a period of 6 (12 minutes taking into account time interval between measurements).

\[ x(t) = TC(t) + S(t) + E(t) \]  \hspace{1cm} (1)

where \( TC(t) \) is a deterministic component, \( S(t) \) is a periodic component, and \( E(t) \) is the (a) random component.

Deterministic components of the model (1) are determined by the classic method of decomposition [4] used a moving average with the period 6. Initial random processes are converted into the new processes by subtracting the value of periodic components. Their mean values are not changed in comparison with the average of the initial processes, but the variances reduce through (at the expense of) reducing the scatter in the values of the parameters.
A multivariate random process of changing of the parameter vector \( x_t = (X_{t1}, \ldots, X_{tn})^T \), where \( i = 1 \div n \) is parameter number, is considered. The vector autoregressive model (\( VAR \)) describes the vector dynamics:

\[
x_t = A_0 + \sum_{j=1}^{p} A_j x_{t-j} + \varepsilon_t, t = 1, \ldots, T
\]

where \( A_0 \) is the vector of constants; \( A_j = (a_{ik}(j)); i, k = 1, \ldots, n, j = 1, \ldots, p \) is the coefficient matrix; \( \varepsilon_t = (\varepsilon_{t1}, \ldots, \varepsilon_{tn})^T \) is the vector of estimation errors (residuals); \( p \) is the model order. The coefficients of the model (2) are calculated by the least squares method. Model order is determined by \( Schw"{a}rz "{c} Bayesian Information Criterion:\n
\[
SBC = T \log|\Sigma| + N \log(T)
\]

where \(|\Sigma|\) is the determinant of the covariance matrix of residuals, \( N \) is the total number of parameters. Thus, for each equation of \( VAR \) has the order \( p \) and the constant, then \( N = n^2 p + n \); each of \( n \) the equations has \( np \) members and constant. The number of regression members is reduced \( \log|\Sigma| \) due to the increase \( N \). Thus the optimal order of the model (2) corresponds to the lowest value of the information criterion. For the investigated process optimal order equal to 1, and the model (2) is:

\[
x_t = A_0 + A_{1t} x_{t-1} + \varepsilon_t
\]

where

\[
x_t = \begin{pmatrix} M_{kt} \\ G_t \\ ge_t \\ CO_t \\ CH_4_t \end{pmatrix}, \quad A_0 = \begin{pmatrix} 33,54 \\ -7,62 \\ 0,16 \\ -0,40 \\ 78,37 \end{pmatrix}, \quad A_{1t} = \begin{pmatrix} 0,21 \\ 0,01 \\ 0,07 \\ 0,01 \\ 0,01 \end{pmatrix}, \quad \varepsilon_t = \begin{pmatrix} 1,34 \\ 0,67 \\ 21,70 \\ -1,99 \\ 207,86 \end{pmatrix}, \quad \omega_t = \begin{pmatrix} -9,49 \\ 21,70 \\ -0,19 \\ 0,60 \\ -24,53 \end{pmatrix}, \quad \varepsilon_t = \begin{pmatrix} 2,87 \\ -0,19 \\ -0,01 \\ -0,04 \\ 0,68 \end{pmatrix}
\]

For correct functioning engines there are stable linear relationships for nonstationary processes of the different parameters’ (elements of vector \( x_t \)) changing. For example:

\[
M_{kt} = -41 G_t + 464,67 CO_t - 0,83 CH_4_t - 349,17 ge_t
\]

Algorithm for finding of such dependencies among engine diagnostic parameters is given in [1,2]. Model (4) is rewritten in the state space:

\[
x_t = A_0 + A_{1t} x_{t-1} + \omega_t
\]

\[
y_t = B x_t + \nu_t
\]

where \( x_t \) is the \( n \times 1 \) state vector, \( y_t \) is the \( m \times 1 \) vector of observations, the initial time vector \( x_0 \) is normally distributed with covariance \( \Sigma_0, \omega_t \), and \( \nu_t \) relatively uncorrelated, \( Q_t \) and \( R_t \) are noises covariances. The matrix \( A_{1t} \) varies.

Based on the well-known approach [6], the model (6), (7) can be simplified by replacing the system of equations with constant coefficients:

\[
x_{t+1} = A_0 + A x_t + w_t
\]

\[
\hat{y}_t = B \cdot x_t + \nu_t, \quad t = 1,2, \ldots, N
\]

\( \hat{y}_t \) denotes a new vector of observations over the system (7), (8). The original and the transformed covariance matrices of the observations are
\[ T = A_{t-1} \Sigma_{t-1} A_{t-1}^T + Q_{t-1}, \hat{T} = A \hat{\Sigma}_{t-1} A^T + Q_{t-1} \]  \hspace{1cm} (10)

At this step the problem is reduced to the determination of the matrix \( A \) for the system that generates a sequence of observations with probabilistic characteristics, as better as possible approximating the probability characteristics of the original vector \( y \). We introduce the information Kullback-Leibler divergence [7], which has the sense of "closeness" of the two probability distributions. For the probability distributions with densities \( g(y) \) and \( f(y) \) the information divergence is

\[ J(g, f) = \int \left[ g(y) - f(y) \right] \log \frac{g(y)}{f(y)} dy \]  \hspace{1cm} (11)

Optimal approximating model (8), (9) for (6), (7) by an information Kullback-Leibler divergence is a system that generates a vector of observations \( \hat{y} \), the parameters of the probability distribution of the closest to the parameters of the distribution of the initial vector \( y \).

The information divergence between the parameters of probability distributions of vectors \( y \) and \( \hat{y} \) is determined by the expression [7]:

\[ J_i = J(g, f) = \frac{1}{2} tr \left\{ \left[ T_i - \hat{T}_i \right] \cdot \left[ T_i^{-1} - \hat{T}_i^{-1} \right] \right\}, \]  \hspace{1cm} (12)

Here \( tr \) is a trace of a matrix \([•]\). Consider the Kullback-Leibler information divergence (13) as a functional of a matrix argument \( A \), showing matrix space \( R_{mxn} \) in \( R \). To determine \( A \) which provides the minimum value for (12), we use the method of steepest descent [9]. At initial estimate \( A^{(0)} \) of the matrix \( A \),

\[ A^{(i+1)} = A^{(i)} - \gamma^{(i)} \frac{\partial J_i}{\partial A} A = A(i), \]  \hspace{1cm} (13)

The gradient term \( \frac{\partial J_i}{\partial A} \) of (14) is the derivative of the matrix functional (13) with the matrix argument \( A \). It can be defined through the rules of matrix differentiation [9]:

\[ \frac{\partial J_i}{\partial A} = \left[ T_i^{-1} - \hat{T}_i^{-1} \right] A \Sigma_{t-1} \downarrow \]  \hspace{1cm} (14)

vector \([•]\downarrow\) is a result of extraction (rewriting) of matrix \([•]\) in column row-by-row.

Thus, at the given initial estimates of matrices \( A \) and \( \hat{A} \) coefficient \( \gamma \) filtering algorithm that minimizes the difference between the parameters of information distribution of the vector of observations and the title of the "new" vector is defined by (14) and (15). At each step of the calculations of the matrix \( A \) is converted, the corresponding covariance matrix of the vector is defined, and the value of the information divergence between the distribution parameter vectors \( y \) and \( \hat{y} \) is tested. By the methods of functional analysis the convexity of the functional \( J_i \) is verified. If the functional to be minimized is convex and bounded below, the sequence defined by (13) is minimizing.

The proposed algorithm is realized by Maple V. The simplified model of the system (6), (7) is determined:

\[ x_{t+1} = A_0 + A x_t + \omega_t \]  \hspace{1cm} (15)
\[ \hat{y}_t = x_t, \ t = 1,2...N. \]  \hspace{1cm} (16)
where

\[
A_i = \begin{pmatrix}
0,210 & 1,39 & -9,30 & 2,80 & 0,000598 \\
-0,00139 & 0,670 & 21,702 & -0,19 & 0,0114 \\
0,0037 & 0,000917 & 0,605 & -0,01 & -0,004 \\
0,00027 & 0,07 & 2,16 & -0,04 & 0,00178 \\
0,01 & 1,75 & 208 & -24,50 & 0,670 \\
\end{pmatrix}
\]

with eigenvalues \( \lambda = [0.809 \ 0.859 \ 0.368 \ 0.0518 \ 0.0518] \).

To calculate the revised matrix values \( A_i \), the estimate of the noise covariance matrix \( Q_t \), resulting in the model (4), was used.

To calculate the covariance matrix of the observations sample estimate \( T_i = \text{cov}(y_i) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})(y_i - \bar{y})' \) was used.

After 40 observations the elements of the covariance matrix of observations are established. When steady-state values of the matrix \( T_i \) and \( \hat{T}_i \) are used, the matrix \( A_i \) is established at 3 - 6 step of the algorithm, depending on the length of the step \( \gamma \). The initial value of the matrix \( A \) was obtained through multivariate OLS. Length of the step is assumed to be constant and equal to the lowest order of the numbers of \( A \).

Model (16), (17) was used to calculate the predicted values of the engine parameters in real time. This allows to reduce labor-intensity and to improve the quality of adjustment work, as well as to increase the test result reliability without increasing their terms. Predicted values of the observed parameters are plotted at the Fig. 3. The estimation errors \( \varepsilon_i = 100\% \frac{|x_i - \hat{x}_i|}{x_i} \) are shown in Fig. 4.
The proposed experimental and statistical approach unlike to existing methods for analysis of measuring sequences, obtained in the tests, allows to implement the improved methods for tests of complex machines (cardiopulmonary resuscitation device, engine). These methods are based on the replacement of the established modes of the test machines by nonstationary ones by means of change of work regimes of the machines and change of the values of the speed and load law that simulates machine in operation.

Acknowledgements
Project "Development and creation of an experimental portable automatic device for cardiopulmonary resuscitation based on devices for indirect heart massage and artificial lung ventilation", ID 32/06-31

References
[1] Pugachev V.S. (2002) The Theory of Probability and Mathematical Statistics: Training, allowance. - 2nd ed., Rev. and add. 496., M. FIZMATLIT.
[2] Golikova V., Pervukhina E., Emmenegger J.-F. (2008), Cointegration of Control Parameter Time Series Measured on Motors at the Stage of Assemblage / / International Journal of Pure and Applied Mathematics, Vol. 42. - N 1. - P. 139-151.
[3] Konstantin Osipov (2011) Identification of modes acceptance tests of ICE. Advanced technologies and systems engineering: Intern. Sat scientific. mp. - Donetsk, Donetsk National Technical University, Vol. 41. - S. 261 - 266.
[4] Halavyan AA Statistic 6 (2007). Statistical analysis of the data. 3rd ed. Textbook. 512s. Moscow: OOO "Bean-Press".
[5] Pervukhina E.L. (2009) Evaluation of engineering products during manufacturing test after assembly. Assembly in the mechanical engineering, instrument. Number 10. - S. 3-9.
[6] Box George E. P., Jenkins Gwilym M., Reinsel Gregory C. (1994) Time Series Analysis, Forecasting and Control. 598 p. - New Jersey: Prentice Hall, Englewood Cliffs.
[7] Makridakis S. G., Wheelwright S. C. (1989) Forecasting methods for management. 497 p. New York: Wiley. Enders W. (1995) Applied Econometric Time Series / 433 p. - New York: John Wiley & Sons.