The Object According State Prediction to Diagnostic Data

L A Baranov¹, E P Balakina², A I Godyaev³

¹Head of the department Control and protection of the Information, RUSSIAN UNIVERSITY OF TRANSPORT (MIIT), 9b9 st. Obrazcova, Moscow 127994, Russia
²Associate professor, Control and protection of the Information, RUSSIAN UNIVERSITY OF TRANSPORT (MIIT), 9b9 st. Obrazcova, Moscow 127994, Russia
³Head of the department, Automation and telemechanics on Railway, Far Eastern State Transport University, Khabarovsk, st. Serysheva 47, Russia

E-mail: Baranov.miit@gmail.com, Balakina_e@list.ru, a.i.godyaev@mail.ru

Abstract. The predicting methodology the state of the object based on diagnostic data is considered. With the selected parameter that determines the state of the object, it is measured in real time at a fixed sampling step. According to the measurement data, the value of this parameter is predicted in the future. This operation is implemented by an extrapolator of the l order - a l degree polynomial, built using the least squares method based on the previous measurements results. The changing process model of the diagnosed parameter is a random time function described by the stationary centered random component sum and a mathematical expectation deterministic change. The estimating prediction error method and the extrapolator parameters influence on its value are presented.

1. Introduction

Prediction based on diagnostic data is used in many fields of science, including significant results obtained in the field of economics, medicine and technology [1] - [13].

The modern technical means presence, software, algorithms and accurate sensors allow diagnosing the state of systems and technological processes in real time. Decisions are made on the diagnostic object functioning termination based on the information received, or on the need for technical measures, including repairs for the further operation possibility. The decision is made after the measured parameter (or group of parameters) exceeds the permissible value, or goes out of the permissible range.

With the aging of the diagnostic object, we can talk about the possibility of predicting its state in real time based on the results of measuring its parameters. The presence of the forecast allows you to make an appropriate decision on the further facility operation, the technological process until the parameter reaches the "critical value", which ultimately increases the functional reliability and survivability of the system. With this approach, the predicted parameter value is compared with the value of a threshold or several thresholds, which makes it possible to make the necessary decisions in advance. The decision-making process in modern intelligent systems can be fully automated.

In this paper, we consider the forecasting process based on l-order extrapolators, built on the use of polynomials orthogonal on equally spaced points.
2. Extrapolator model

Let \( x[nT] \) be a parameter that is a random function of time \( t \), measured at the moments \( t = nT \), \( n = 0,1,2,\ldots, T \) is the time sampling increment. The extrapolator determines at the moment of time \( nT \) predicted value of the function \( z[(n + \varepsilon)T] \), \( 0 < \varepsilon \leq 1 \), according to the known measurement results at the time moments \( (n - M + i)T \), \( i = 0,1,2,\ldots, M \), i.e. by the \( (M + 1) \) th previous measurement result. This approach allows real-time prediction. Consider the resulting the function value \( z[(n + \varepsilon)T] \) as a polynomial of \( l \)-th degree. For \( l < M \) coefficients of the polynomial are calculated using the least squares procedure from the known values \( x[(n - M + i)T], i = 0,1,2,\ldots, M \) [14],[15].

\[
z[(n + \varepsilon)T] = \sum_{l=0}^{M} x[(n - M + l)T] L_{l}^{e_i,ls,l}[0, \varepsilon],
\]

where

\[
L_{l}^{e_i,ls,l} = \sum_{j=0}^{l} \frac{\varphi_{j,M}(\varepsilon)\varphi_{j,M}(M+\varepsilon)}{\|\varphi_{j,M}\|},
\]

\[
\|\varphi_{j,M}^{2}\| = \sum_{l=0}^{M} \varphi_{l,M}(i);
\]

\[
\varphi_{0,M}(\bar{t}), \varphi_{1,M}(\bar{t}), \ldots, \varphi_{j,M}(\bar{t}), \ldots, \varphi_{l,M}(\bar{t}) - \text{Chebyshev polynomials of degree } 0,1,\ldots,l, \text{ orthogonal}
\]

on the equidistant points set \( \bar{t} = \frac{t_{0} - t_{1}}{T} \) - distance interval, \( t_{0}, t_{1}, \ldots, t_{M} \) - correspond to points \( i = 0,1,2,\ldots, M \).

Chebyshev polynomials are calculated from the recurrence relation

\[
\bar{t}\varphi_{j,M}(\bar{t}) = -\frac{(j+1)(M-j)}{2(2j+1)} \varphi_{j+1,M}(\bar{t}) + \frac{M}{2} \varphi_{j,M}(\bar{t}) - \frac{j(M+j+1)}{2(2j+1)} \varphi_{j-1,M}(\bar{t}),
\]

for \( \varphi_{0,M} = 1, \varphi_{1,M} = 1 - \frac{2\bar{t}}{M} \).

The extrapolation order \( l \) is the degree of polynomial approximation. For the extrapolation order increased from \( l \) to \( (l + 1) \)

\[
L_{l}^{e_i,ls,l+1}[0, \varepsilon] = L_{l}^{e_i,ls,l}[0, \varepsilon] + \frac{\varphi_{l+1,M}(\varepsilon)\varphi_{l+1,M}(M+\varepsilon)}{\|\varphi_{l+1,M}\|}.
\]

Proven that [14]:

\[
\sum_{l=0}^{M} L_{l}^{e_i,ls,l}[0, \varepsilon] = 1.
\]

The function expressions \( L_{l}^{e_i,ls,l}[0, \varepsilon] \) are given in the table 1 for \( l = 1, l = 2 \).

| Table 1. The expressions \( L_{l}^{e_i,ls,l}[0, \varepsilon] \) for extrapolators of the first and second orders. |
| --- |
| \( l \) | \( L_{l}^{e_i,ls,l}[0, \varepsilon] \) |
| --- | --- |
| 1 | \( 2(M^2 - M + 3M\varepsilon - 3Mi - 6\varepsilon) \) |
| 2 | \( 2(M^2 - M + 3M\varepsilon - 3Mi - 6\varepsilon) \) |
| 3 | \( 5(6\varepsilon(M + \varepsilon) + M(M - 1))(6\varepsilon(M - i) - M(M - 1)) \) |
| 4 | \( M(M + 1)(M + 2)(M + 3) \) |

Thus, in accordance with (1), using the data of Table 1, it is possible to calculate the predicted value of the random variable \( z[(n + \varepsilon)T] \) for any fixed value \( \varepsilon, 0 < \varepsilon \leq 1 \), which is compared with the threshold value and a decision is made on the object functioning possibility.

3. The prediction error estimation

Various estimates of the forecast error are used to analyze the forecast, and their classification is also given in the literature [16] [17] [18] [19].
The absolute value of the prediction error at \( t = (n + \varepsilon)T \) defined as
\[
\theta[n, \varepsilon] = x[(n + \varepsilon)T] - z[(n + \varepsilon)T].
\] (7)

Measured parameter \( x[(n + \varepsilon)T] \) in time is a random process, is not stationary, because the object degradation is associated with its aging. At the same time, the forecast is effective only when the measurement results are dependent random variables.

Consider a random process model, which is the sum of a centered stationary process and a deterministic function that describes the change in the mathematical expectation of a parameter characterizing the object degradation. Such model should be consistent with a priori knowledge of the object. A priori information is stored using Big Data technology. Taking into the linearity of operators (1) and (2), the resulting forecast error can be obtained after estimating the forecast errors of stationary and deterministic random processes.

Let be
\[
x[(n + \varepsilon)T] = \hat{x}[(n + \varepsilon)T] + x_d[(n + \varepsilon)T],
\]
\[
z[(n + \varepsilon)T] = \hat{z}[(n + \varepsilon)T] + z_d[(n + \varepsilon)T] = 
\]
\[
= \sum_{i=0}^{M} \hat{x}[(n - M + i)T]L^{e,ls,l}_{0,0} [0, \varepsilon] + \sum_{i=0}^{M} x_d[(n - M + i)T]L^{e,ls,l}_{i,0}[0, \varepsilon],
\] (8)

where \( \hat{x} \) and \( \hat{z} \) are the random component corresponding values of measured and predicted value, \( x_d \) and \( z_d \) are deterministic components corresponding values.

Let us separately estimate the forecast error of the deterministic and random stationary components.

If the model of deterministic component is a time function, given by a polynomial whose degree does not exceed \( l \), then it is obvious that using \( l \)-degree extrapolator the extrapolation error is zero. For parabolic model of deterministic component, the error is zero using an extrapolator, the which order is 2 and more.

To analyze the forecast error of the stationary component, we use the root-mean-square estimate. The random process \( \hat{x}[(n + \varepsilon)T] \) is given by its autocorrelation function.
\[
R_{xx}(\tau) = \sum_{m=0}^{\infty} \frac{R^{(2m)}_{xx}(0)}{(2m)!} \tau^{2m},
\] (9)

where \( R^{(2m)}_{xx}(0) \) - \((2m)\)-order derivative of the function \( R_{xx}(\tau) \) for \( \tau = 0 \).

Such expansion is valid for the class of differentiable random processes [14], [20].

We introduce the root-mean-square estimate as
\[
\sqrt{\hat{\theta}^2} = \sqrt{[\hat{x}[(n + \varepsilon)T] - \hat{z}[(n + \varepsilon)T]]^2},
\] (10)

where the bar above is the mean value of the expression square enclosed in brackets.

Using the results [14], [20], we can show
\[
\hat{\theta}^2 = R_{xx}(0) - 2 \sum_{i=0}^{M} L^{e,ls,l}_{i,0} [0, \varepsilon] R_{xx}[(M - i + \varepsilon)T] + \sum_{i=0}^{M} \sum_{r=0}^{M} L^{e,ls,l}_{i,0} [0, \varepsilon] L^{e,ls,l}_{r,0}[0, \varepsilon] R_{xx}[(i - r)T],
\]

where \( \hat{\theta}^2 \) is the forecast error variance, \( R_{xx}(\tau) \) - autocorrelation function of a centered random process \( \hat{x}[(n + \varepsilon)T] \).

Let us pay attention to the dependence of the error variance from \( \varepsilon \), indicated the random function non-stationarity of the forecast error.

We substitute in (11) the expression (9) of the autocorrelation function expansion in a power series and restrict ourselves to the series main term. Following [14], [15], we obtain expressions for the upper estimate of the forecast error variance shown in Table 2 for extrapolators of the first and second orders.
Table 2. Expressions of the error variance random component of the forecast error.

| Expression                                                                                         |
|----------------------------------------------------------------------------------------------------|
| The series main term $\overline{\theta^2}[0, \varepsilon]$                                       |
| $a_m = \frac{\delta^2 \theta \left( \theta^2 + 6 \theta \varepsilon - M + 6 \varepsilon^2 \right)^2}{20}$ |
| $= \frac{a_m}{\theta^2} \left( \theta^2 + 12 \theta \varepsilon - 3 M + 10 \varepsilon^2 + 2 \right)^2$  |

In table 2 coefficients are

$$a_m = \frac{R_{xx}^{(2m)}(0)}{(2m)!},$$

where $m = 2$ in case $l = 1$, $m = 3$ in case $l = 2$.

As an example, we present the results for random processes with a constant spectral power density in the band from $\omega = 0$ to $\omega = \omega_{\text{max}}$ and equal to zero outside this band and a random process with a decreasing spectral power density [15]. In the first case

$$R_{xx}(\tau) = R_{xx}(0) \sin(\omega_{\text{max}} \tau) and a_m = \frac{(-1)^m(\omega_{\text{max}} \tau)^{2m}}{(2m+1)!} R_{xx}(0), \quad (11)$$

in the second case

$$R_{xx}(\tau) = R_{xx}(0) e^{-a_2 \tau^2} and a_m = \frac{(-1)^m(a \tau)^{2m}}{(2m)!} R_{xx}(0). \quad (12)$$

To analyze the forecast of the centered random component, we introduce the relative estimate

$$\sqrt{\delta^2[0, \varepsilon]} = \sqrt{\frac{a^2[0, \varepsilon]}{R_{xx}(0)}}$$

In this case $a_m$ is divided by $R_{xx}(0)$ - the variance of the input signal, that is, in the first case $\frac{a_m}{R_{xx}(0)} = \frac{(-1)^m(\omega_{\text{max}} \tau)^{2m}}{(2m+1)!}$, in the second case $\frac{a_m}{R_{xx}(0)} = \frac{(-1)^m(a \tau)^{2m}}{(2m)!}$.

The considered estimate is the relative root-mean-square estimate (RMS) of the forecast error (with respect to the RMS of the input signal).

Let us bring to a different form the expression of the estimates $\sqrt{\delta^2[0, \varepsilon]}$ for the examples under consideration. With autocorrelation functions (11) and (12), the results presented in Table 3.

Table 3. Expressions for estimating the relative standard deviation forecast for extrapolators of the first and second orders.

| Expression                                                                                         |
|----------------------------------------------------------------------------------------------------|
| $R_{xx}(\tau) = R_{xx}(0) \frac{\sin(\omega_{\text{max}} \tau)}{\omega_{\text{max}} \tau}$        |
| $= \frac{(\omega_{\text{max}} \tau)^{2m}}{12 \sqrt{5}} \left( 1 + \frac{6}{M} \varepsilon - \frac{1}{M} + \frac{6}{M^2} \varepsilon^2 \right)$ |
| $\frac{(\omega_{\text{max}} \tau)^{2m}}{120 \sqrt{7}} \left( 1 + \frac{2}{M} \varepsilon \right) \left( 1 + \frac{10}{M} \varepsilon - \frac{3}{M} + \frac{10}{M^2} \varepsilon^2 \right)$ |
| $+ \frac{2}{M^2}$                                                                                 |

| Expression                                                                                         |
|----------------------------------------------------------------------------------------------------|
| $R_{xx}(\tau) = R_{xx}(0) e^{-a_2 \tau^2}$                                                         |
| $= \frac{(a \tau)^{2m}}{12 \sqrt{5}} \left( 1 + \frac{6}{M} \varepsilon - \frac{1}{M} + \frac{6}{M^2} \varepsilon^2 \right)$ |
| $\frac{(a \tau)^{2m}}{120 \sqrt{7}} \left( 1 + \frac{2}{M} \varepsilon \right) \left( 1 + \frac{10}{M} \varepsilon - \frac{3}{M} + \frac{10}{M^2} \varepsilon^2 \right)$ |
| $+ \frac{2}{M^2}$                                                                                 |

$\tau = \tau$ is the time interval between the zero and $M$-th measurement points of the predicted process. This value determines the forecast error. An increase $\tau$ corresponds to decrease the correlation.
between the measurement results at the zero and $M$-th points. Hence, we can conclude that an accurate forecast takes place only if there is a correlation between the measured values of the predicted process.

Consider the functions $F_1(M) = \left(1 + \frac{6}{M}\epsilon - \frac{1}{M} + \frac{6}{M^2}\epsilon^2\right)$ and $F_2(M) = \left(1 + \frac{2}{M}\epsilon\left(1 + \frac{10}{M}\epsilon - \frac{3}{M} + \frac{10}{M^2}\epsilon^2 + \frac{2}{M^2}\right)\right)$. The graphs of these functions for $\epsilon = 1$, which corresponds to the forecast time $T$, are shown in Fig. 1. We can see from the graphs that an increase in the number of predicted process known values, the factors under consideration in the estimate of the forecast relative standard deviation $\sqrt{\delta^2[0,\epsilon]}$ for the extrapolators of the first and second orders tend to unity.

![Figure 1. Dependence of functions F1 and F2 on M.](image1)

Figures 2a and 3a show the dependences $R_{xx}(\omega_{max}TM)$ and estimates of the forecast relative standard deviation $\sqrt{\delta^2[0,\epsilon]}$ of the first-order extrapolator for the autocorrelation functions given by expressions (11) and (12) respectively.

Figures 2b and 3b show the dependences $R_{xx}(\omega_{max}TM)$ and estimates of the forecast relative standard deviation $\sqrt{\delta^2[0,\epsilon]}$ of the second-order extrapolator for the autocorrelation functions given by expressions (11) and (12) respectively.

![Figure 2. The dependences $R_{xx}(\omega_{max}TM)$ and estimates of the forecast relative standard deviation $\sqrt{\delta^2[0,\epsilon]}$ of the first-order extrapolator for the autocorrelation functions given by expressions:](image2)
a) \( R_{xx}(\tau) = R_{xx}(0) \frac{\sin(\omega_{\text{max}}\tau)}{\omega_{\text{max}}\tau} \)

b) \( R_{xx}(\tau) = R_{xx}(0) e^{-\alpha\tau^2} \)

\[ \text{Figure 3.} \] The dependences \( R_{xx}(\omega_{\text{max}} TM) \) and estimates of the forecast relative standard deviation \( \sqrt{\delta^2 [0, \epsilon]} \) of the second-order extrapolator for the autocorrelation functions given by expressions:

\[ \text{a)} \ R_{xx}(\tau) = R_{xx}(0) \frac{\sin(\omega_{\text{max}}\tau)}{\omega_{\text{max}}\tau} \]

\[ \text{b)} \ R_{xx}(\tau) = R_{xx}(0) e^{-\alpha\tau^2} \]

This implies:

to ensure the forecast error using the first order extrapolator with the autocorrelation function \( R_{xx}(\tau) = R_{xx}(0) \frac{\sin(\omega_{\text{max}}\tau)}{\omega_{\text{max}}\tau} \) up to 10% we should choose \( \omega_{\text{max}} TM \) from 0 to 1,6. To ensure the forecast error up to 5% \( \omega_{\text{max}} TM \) we should choose \( \omega_{\text{max}} TM \) from 0 to 1,15. To ensure the forecast error up to 2% \( \omega_{\text{max}} TM \) we should choose \( \omega_{\text{max}} TM \) from 0 to 0,73.

to ensure the forecast error using the first order extrapolator with the autocorrelation function \( R_{xx}(\tau) = R_{xx}(0) e^{-\alpha\tau^2} \) up to 10% we should choose \( \omega_{\text{max}} TM \) from 0 to 1,09. To ensure the forecast error up to 5% \( \omega_{\text{max}} TM \) we should choose \( \omega_{\text{max}} TM \) from 0 to 0,77. To ensure the forecast error up to 2% \( \omega_{\text{max}} TM \) we should choose \( \omega_{\text{max}} TM \) from 0 to 0,49.

to ensure the forecast error using the second order extrapolator with the autocorrelation function \( R_{xx}(\tau) = R_{xx}(0) \frac{\sin(\omega_{\text{max}}\tau)}{\omega_{\text{max}}\tau} \) up to 10% we should choose \( \omega_{\text{max}} TM \) from 0 to 3,16. To ensure the forecast error up to 5% \( \omega_{\text{max}} TM \) we should choose \( \omega_{\text{max}} TM \) from 0 to 2,51. To ensure the forecast error up to 2% \( \omega_{\text{max}} TM \) we should choose \( \omega_{\text{max}} TM \) from 0 to 1,85.

to ensure the forecast error using the second order extrapolator with the autocorrelation function \( R_{xx}(\tau) = R_{xx}(0) e^{-\alpha\tau^2} \) up to 10% we should choose \( \omega_{\text{max}} TM \) from 0 to 2,28. To ensure the forecast error up to 5% \( \omega_{\text{max}} TM \) we should choose \( \omega_{\text{max}} TM \) from 0 to 1,81. To ensure the forecast error up to 2% \( \omega_{\text{max}} TM \) we should choose \( \omega_{\text{max}} TM \) from 0 to 1,33.

4. Conclusion
The obtained forecast error analysis estimates of the stationary component of a random process shows:

at the constant values of \( \omega_{\text{max}} = \alpha \) and \( T = \text{const} \), the value \( \sqrt{\delta^2 [0, \epsilon]} \) grows with an increase the extrapolation order and forecast time \( \epsilon T \);

with an increase in the value of \( \omega_{\text{max}} \) and \( \alpha \), indicated a decrease in the statistical relationship between the measuring results the diagnostic parameter, the estimate of the forecast error grows;
changing the values of $T$ and $M$, it is possible to choose the random component permissible forecast error with the extrapolation order $l$, which ensured that the forecast error of the deterministic component is equal to zero.

5. References

[1] Klyachkin V N, Zhukov D A 2019 Forecasting the state of a technical object using machine learning methods *International journal Software products and systems* **22** 244-250 DOI:10.15827/0236-235X.126.244-250

[2] Kuvayskova Yu E 2017 The prediction algorithm of the technical state of an object by means of fuzzy logic inference models *Procedia Engineering* **201** 767-772 DOI:10.1016/j.proeng.2017.09.624

[3] Makhtutov N A, Permyakov V N, Akhmetkhanov R S etc 2017 Diagnostics and monitoring of the state of complex technical systems: a tutorial (Tyumen: TIU) 632 ISBN 978-5-9961-1433-7

[4] Chi-Phi Do, Quang-Huy Le, Duy-Phuoc Pham, Dinh-Kha Le Forecast of Energy Consumption of Drying System According to The Environmental Temperature and Humidity on IoT by Arima Algorithm https://www.researchgate.net/publication/351546709_Forecast_of_Energy_Consumption_of_Drying_System_According_to_The_Environmental_Temperature_and_Humidity_on_IoT_by_Arima_Algorithm

[5] Cheng Zhang, Hedvig Kjellström, Bo C Bertilson Diagnostic Prediction Using Discomfort Drawings 1612.01356.pdf (arxiv.org)

[6] Adalat Kerimov, Jeyran Khalilova, Sadagat Abdulzade Recognition And Evolutionary Computation IJCSI-9-2-3-247-251.pdf

[7] Zenkin A I, Ryazanov V V 1977 Algorithms of forecasting of conditions of control objects *The journal of calculus mathematics and the mathematical physics* vol 17 **6** pp 1564-1573

[8] Zhukov D A, Klyachkin V N, Krasheninnikov V R and Kuvayskova Yu E Selection of aggregated classifiers for the prediction of the state of technical objects Open Access proceedings *Journal of Physics: Conference series (ceur-ws.org)*

[9] Gaskarov D V, Golinkevich T A and Mozgalevskij A V 1974 Technical condition and reliability prediction of electronic equipment (Moscow: Soviet radio) p 224

[10] Krasheninnikov V R, Klyachkin V N and Kuvayskova Yu E 2018 Models updating for technical objects state forecasting 3rd Russian-Pacific Conf. on Computer Technology and Applications (RPC) IEEE Xplore 1-4

[11] Merkov A B 2011 Pattern recognition Introduction to statistical learning methods (Moscow: Editorial URSS) p 256

[12] Krasheninnikov V R, Krasheninnikov V R, Klyachkin V N, Kuvayskova Yu E Models Updating for Technical Objects State Forecasting https://www.researchgate.net/publication/328157525_Models_Updating_for_Technical_Objects_State_Forecasting

[13] Baranov L A, Balakina E P 2020 The random processes prediction based on orthogonal polynomials on the set of equally spaced points *Electrical engineering* **9** pp 39-45

[14] Baranov L A 1990 Level quantization and temporal discretization in digital control systems (M.: Energoatomizdat) 304

[15] Chao Chen, Jamie Twycross, Jonathan M Garibaldi A new accuracy measure based on bounded relative error for time series forecasting https://doi.org/10.1371/journal.pone.0174202

[16] Shcherbakov M V, Brebels A, Shcherbakova N L, Tyukov A P, Janovsky T A, Kamaev V A A survey of forecast error measures https://www.researchgate.net/publication/281718517_A_survey_of_forecast_error_measures

[17] Ummul Khair, Hasanul Fahmi, Sarudin Al Hakim, Robbi Rahim Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error
https://www.researchgate.net/publication/321801771_Forecasting_Error_Calculation_with_Mean.Absolute.Deviation_and_Mean.Absolute.Percentage_Error

[18] Boylan John & Syntetos Aris 2021 Forecasting the Variance of Demand and Forecast Error 10.1002/9781119135289.ch7.

[19] Baranov L A, Balakina E P, Godyaev A I 2020 The Disturbances Prediction in the Systems of Automatic Train Traffic Control of Urban Transport Systems International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)

**Acknowledgments**

The reported study was funded by RFBR, Sirius University of Science and Technology, JSC Russian Railways and Educational Fund “Talent and success”, project number 20-37-51001.