Informatics and Control Problems 39 Issue 2 (2019)
journal homepage: www.icp.az

Algorithms for determining the confidence interval of the estimates of distribution density of the noise of a noisy signal

T.A. Aliev[a,b], N.F. Musaeva[a*], M.T. Suleymanova[b]

[a] *Azerbaijan University of Architecture & Construction, Baku, Azerbaijan*
[b] *Institute of Control Systems of ANAS, Baku, Azerbaijan*

ARTICLE INFO

Article history:
Received 11.10.2019
Received in revised form 24.11.2019
Accepted 23.12.2019
Available online 30.12.2019

Keywords:
Noise
Noisy signal
Noise distribution density
Confidence interval
General population
Sample
Control
Diagnostics
Fault monitoring
Defect development dynamics

ABSTRACT

The authors propose algorithms and technologies for determining the confidence interval of the estimates of distribution density of the noise in the general population and by a sample of the stochastic process. It is noted that these algorithms can be used to indicate the early period of the initiation of a fault and the dynamics of its development in monitoring, control and diagnostic systems in various industries.

1. Introduction

It is known that noisy signals received from sensors in systems of monitoring, control, management, diagnostics, etc. are random processes. To date, methods are known for determining the estimates of mathematical expectation, variance and correlation function of an ergodic random process [1, 2]. In addition, for an ergodic random process, probability density estimates are obtained. To this end, the method of the relative residence time of the implementation of the signal above a given level and the method of discrete samples are used. The obtained measurement results are graphically presented in the form of a histogram [1, 2].

In systems of monitoring, control, management, diagnostics, etc., these methods allow estimating the probability density of a signal coming from a sensor installed in the most informative points of the object under investigation. At the same time, it was shown in [3–13] that at the early stage of the appearance of various faults of equipment, devices, structures, engine, mechanism, motor, etc., as a result of, for instance, wear, corrosion, cracks, breakages, etc., the additive noise $\mathrm{E}(t)$ is superimposed on the useful signal $X(t)$: $G(t)=X(t)+\mathrm{E}(t)$. In this case, using the existing methods for analyzing measurement information, one can estimate the mathematical expectation, variance,

[*]Corresponding author.
E-mail addresses: director@cyber.az (T.A. Aliev), musanaila@gmail.com (N.F. Musaeva), metanet_suli@yahoo.com (M.T. Suleymanova)

www.icp.az/2019/2-01.pdf

correlation function, probability density of the noisy signal $G(t)$.

It was noted in [14-17] that for oil and gas production facilities, deep-sea offshore platforms, drilling rigs, etc., detecting the early stage of fault formation and determining the dynamics of its development requires calculating the above characteristics for the noise $\mathrm{E}(t)$ of the noisy signal $G(t)$. However, the characteristics of the noise $\mathrm{E}(t)$ are not directly measurable, since the samples of the noise cannot be extracted from the samples of the noisy signal $G(t)$. Considering that only the parameters of the noisy signal $G(t)$ can be measured, methods were developed in [3–13] for calculating estimates of the noise characteristics using realizations of $G(t)$.

For instance, in [3], algorithms are proposed for determining the robust estimates of the correlation functions of random noisy signals as a result of calculating the estimates of the noise characteristics and eliminating their effects on the estimates of the useful signal. In [4], algorithms are proposed for determining the estimates of the spectral characteristics of the noise of noisy signals and technologies for replacing the samples of the noise that are not measurable with their approximate equivalent values. In [5-13], algorithms are proposed for calculating the estimates of the distribution density of the noise of noisy signals.

In this paper, we propose algorithms and technologies for determining the confidence interval of the estimates of the distribution density of the noise $\mathrm{E}(t)$ in the general population and by a sample of the noisy signal $G(t)$, also demonstrating the possibility of using these algorithms to identify a fault and the dynamics of its development at an early stage of initiation in specific technical objects.

2. Problem statement

Let us analyze a noisy digital signal $G(t)$ consisting of the useful signal $X(t)$ and the additive noise $\mathrm{E}(t)$: $G(t)=X(t)+\mathrm{E}(t)$, which comes from a sensor located in the coverage area of factors affecting the object. In this case, the appearance of the noise $\mathrm{E}(t)$ is due to the presence of a technical fault, defects, malfunctions. The signals $X(t)$, $\mathrm{E}(t)$, $G(t)$ are random stationary ergodic processes, and the noise $\mathrm{E}(t)$ is impossible to extract from $G(t)$. For the noisy signal $G(t)$, we can determine the estimates of the mathematical expectation m_G, the variance D_G, the mean square deviation σ_G, the correlation function $R_{GG}(\tau)$ [18, 19]:

$$m_G = \frac{1}{N}\sum_{i=1}^{N} G(i\Delta t),$$

$$D_G = \frac{1}{N}\sum_{i=1}^{N}[G(\Delta t) - m_G]^2,\ \sigma_G = \sqrt{D_G},$$

$$R_{GG}(\mu) = \frac{1}{N}\sum_{i=1}^{N}\overset{\circ}{G}(i\Delta t)\overset{\circ}{G}((i+\mu)\Delta t), \tag{1}$$

where Δt is the sampling interval, $\mu=0, \Delta t, 2\Delta t, \ldots$ is the time shift.

In this case, the useful signal $X(t)$ reflects the current state of the process under investigation. A priori it is known that in the system of monitoring, control, diagnostics, forecasting, management, identification, etc. the noise emerged as a result of, for instance, wear, corrosion, cracks, breakdowns, damage, failure of units, engines, mechanisms, destruction of building structures, etc., has the normal distribution $N(\varepsilon; m_{\mathrm{E}}, \sigma_{\mathrm{E}})$ and zero mean $m_{\mathrm{E}} = 0$ in the general population [14-17]. In addition, at the time of the defect initiation, the correlation time of the noise $\mathrm{E}(t)$ is much shorter than the correlation time of the useful signal $X(t)$.

Since the stationary random noise $\mathrm{E}(t)$ is ergodic, its mathematical expectation m_{E} and the mean square deviation σ_{E} have the same value for any of the random signals in the population. Therefore, we represent the density function of the normal distribution of the noise $N(\varepsilon; m_{\mathrm{E}}, \sigma_{\mathrm{E}}) = N(\varepsilon)$ as follows:

$$N(\varepsilon) = \frac{1}{\sigma_{\mathrm{E}}\sqrt{2\pi}} e^{-\frac{(\varepsilon - m_{\mathrm{E}})^2}{2\sigma_{\mathrm{E}}^2}}. \tag{2}$$

From formula (2), it is obvious that by the type and dynamics of changes in the estimates of the distribution density $N(\varepsilon)$ of the noise $\mathrm{E}(t)$, we can assess the changes in the technical condition of the object under investigation. This is due to the fact that different values of the mean square deviation of the noise correspond to different degrees of malfunction. Besides, despite the fact that in the general population the noise has zero mean $m_{\mathrm{E}} = 0$, in the sample this value fluctuates within a certain confidence interval depending on the nature and degree of development of the malfunction. Therefore, different types of the $N(\varepsilon)$ curve with different maximum values and the coordinates of the inflection points will correspond to different degrees of malfunction. This property of the noise distribution curve allows us to use it to assess the technical condition of an object

For this reason, in the following paragraphs, we propose algorithms for determining the confidence interval for the mathematical expectation and the estimates of the distribution density $N(\varepsilon)$ of the noise $\mathrm{E}(t)$ in the general population and by a sample of the noisy signal $G(t)$.

3. Algorithms for determining the confidence interval.

3.1. Algorithms for determining the lower limit of the confidence interval.

It is known that the normal distribution $N(\varepsilon)$ of the noise $\mathrm{E}(t)$ of the noisy signal $G(t)$ is characterized by two parameters: the mathematical expectation m_{E} and the mean square deviation $\sigma_{\mathrm{E}} = \sqrt{D_{\mathrm{E}}}$. Since in the general population the noise $\mathrm{E}(t)$ is distributed according to the normal law with the zero mean $m_{\mathrm{E}} = 0$, the problem comes down to determining only the estimate of the parameter σ_{E}. For this purpose, we first calculate the variance D_{E} of the noise. We shall use expression (1):

$$R_{GG}(\mu) = \frac{1}{N}\sum_{i=1}^{N}\left(\overset{\circ}{X}(i\Delta t) + \mathrm{E}(i\Delta t)\right)\left(\overset{\circ}{X}((i+\mu)\Delta t) + \mathrm{E}((i+\mu)\Delta t)\right),$$

$\overset{\circ}{X}(i\Delta t) = X(i\Delta t) - m_X$, $m_X = \frac{1}{N}\sum_{i=1}^{N} X(i\Delta t)$ is the mathematical expectation of $X(t)$.

Considering that the signal $X(t)$ and the noise $\mathrm{E}(t)$ are not correlated, i.e.

$$\frac{1}{N}\sum_{i=1}^{N}\overset{\circ}{X}(i\Delta t)\mathrm{E}((i+\mu)\Delta t) = 0, \quad \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}(i\Delta t)\overset{\circ}{X}((i+\mu)\Delta t) = 0,$$

we can write the following [3-13]:

$$R_{GG}(\mu) = R_{XX}(\tau) + R_{\mathrm{EE}}(\mu), \tag{3}$$

where $R_{XX}(\mu) = \frac{1}{N}\sum_{i=1}^{N}\overset{\circ}{X}(i\Delta t)\overset{\circ}{X}((i+\mu)\Delta t)$ is the correlation function of $X(t)$,

$R_{\mathrm{EE}}(\mu) = \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}(i\Delta t)\mathrm{E}((i+\mu)\Delta t)$ is the correlation function of the noise $\mathrm{E}(t)$.

In practice, in such low-frequency slow-flowing technological processes as oil refining, petrochemistry, when $\mu = \Delta t$ is much smaller than the observation time T, the noise $\mathrm{E}(t)$ forms from high-frequency spectra as a result of such malfunctions as wear, corrosion, carbon deposition, etc. and has a higher spectrum than the useful component $X(t)$. The value of the useful component does not have enough time to change during Δt, and $X(t+\Delta t)$ coincides with the value of $X(t)$, i.e.

$$X(t+\Delta t) = X(t). \quad (4)$$

Then, for the indicated objects, if condition (4) is fulfilled, the relation $\frac{R_{XX}(\Delta t)}{R_{XX}(0)}$ equals unit, i.e. [19]:

$$R_{XX}(\Delta t) = R_{XX}(0). \quad (5)$$

At the same time, due to the fact that for a Gaussian random noise, the sampling interval Δt is selected based on the end correlation time μ_{cor} of the noise, then $R_{\mathrm{EE}}(\mu)$ can be represented as follows [20]:

$$R_{EE}(\mu) = \begin{cases} R_{EE}(0) & when\ \mu = 0 \\ 0 & when\ \mu \geq \Delta t \end{cases}. \quad (6)$$

Therefore, if, using formula (1), we calculate the estimates of the correlation function $R_{GG}(\mu)$ of the noisy signal at μ=0 and a sufficiently small time interval $\mu_{cor} = \Delta t$ compared to the observation time T and find the difference between these estimates, we get:

$$R_{GG}(0) - R_{GG}(\Delta t) = R_{XX}(0) + R_{\mathrm{EE}}(0) - R_{XX}(\Delta t) - R_{\mathrm{EE}}(\Delta t). \quad (7)$$

Taking into account conditions (5)-(6), formulas (3), (7), and the face that the estimates of the autocorrelation functions of the useful signal $X(t)$ and the noise $\mathrm{E}(t)$, respectively, at the zero time shift μ=0 are the variances of the useful signal and the noise, respectively: $R_{XX}(0) = D_X$, $R_{\mathrm{EE}}(0) = D_{\mathrm{E}}$, we get:

$$D^*_{\mathrm{E}} = R_{GG}(\mu = 0) - R_{GG}(\mu = \Delta t). \quad (8)$$

In earlier works, a formula was developed for calculating the noise variance for a more general case, when the sampling interval Δt is selected based on the frequency band of the spectrum of the noise $\mathrm{E}(t)$, i.e. $\Delta t = 1/(2 f_\varepsilon)$, where f_ε is the cutoff frequency, Hz [5-7]:

$$D^*_{\mathrm{E}} = R_{GG}(\mu = 0) - 2 R_{GG}(\mu = \Delta t) + R_{GG}(\mu = 2\Delta t). \quad (9)$$

Taking into account (2), (8), (9), digital algorithms are proposed for determining the distribution density $N^*(\varepsilon)$ of the normal noise $m_g = \frac{1}{N}\sum_{i=1}^{N} G(i\Delta t)$, the maximum $N^*_{\max}(0)$ and the inflection points with coordinates $\left(m_{\mathrm{E}} - \sigma^*_{\mathrm{E}}; \frac{1}{\sigma^*_{\mathrm{E}}\sqrt{2\pi e}}\right)$ and $\left(m_{\mathrm{E}} + \sigma^*_{\mathrm{E}}; \frac{1}{\sigma^*_{\mathrm{E}}\sqrt{2\pi e}}\right)$ or given that in the general population m_{E}=0, $\left(-\sigma^*_{\mathrm{E}}; \frac{1}{\sigma^*_{\mathrm{E}}\sqrt{2\pi e}}\right)$ and $\left(\sigma^*_{\mathrm{E}}; \frac{1}{\sigma^*_{\mathrm{E}}\sqrt{2\pi e}}\right)$:

1) The mean-square deviation of the noise $\mathrm{E}(t)$ of the noisy signal $G(t)$ is calculated from the expression:

$$\sigma^*_{\mathrm{E}} = \sqrt{D^*_{\mathrm{E}}} = \begin{cases} \sqrt{R_{GG}(\mu = 0) - R_{GG}(\mu = \Delta t)} & for\ the\ special\ case я \\ \sqrt{R_{GG}(\mu = 0) - 2R_{GG}(\mu = \Delta t) + R_{GG}(\mu = 2\Delta t)} & for\ the\ general\ case \end{cases}. \quad (10)$$

2) Given that in the general population $m_{\mathrm{E}} = 0$ and for a normally distributed random signal, the deviation from the mathematical expectation in absolute value does not exceed the triple mean square deviation, discrete values of the distribution density $N^*(\varepsilon)$ of the noise $\mathrm{E}(t)$ are calculated in

the interval $\pm 3\sigma_{\mathrm{E}}^{*}$, i.e. at $-3\sigma_{\mathrm{E}}^{*} \le \mathrm{E}(t) \le +3\sigma_{\mathrm{E}}^{*}$:

– the values of $\mathrm{E}(t)$ are calculated: $\varepsilon_{\min} = -3\sigma_{\mathrm{E}}^{*}$; $\varepsilon_{\max} = +3\sigma_{\mathrm{E}}^{*}$;

– with a certain step $\Delta\varepsilon$ the sequence of possible values of $\mathrm{E}(t)$ is set: $\varepsilon(1) = \varepsilon_{\min}$, $\varepsilon(i+1) = \varepsilon(i) + \Delta\varepsilon$, …, $\varepsilon_{\max}$, for which the condition $\varepsilon(i-1) < \varepsilon(i)$ is fulfilled;

– the distribution density is calculated in the points $\varepsilon_{\min} = \varepsilon(1),\ \varepsilon(2),\ \ldots,\ \varepsilon_{\max}$:

$$N^{*}(\varepsilon) = \frac{1}{\sigma_{\mathrm{E}}^{*}\sqrt{2\pi}} e^{-\frac{(\varepsilon - m_{\mathrm{E}})^2}{2(\sigma_{\mathrm{E}}^{*})^2}} \tag{11}$$

or taking into account the condition m_{E} =0, we get:

$$N^{*}(\varepsilon) = \frac{1}{\sigma_{\mathrm{E}}^{*}\sqrt{2\pi}} e^{-\frac{\varepsilon^2}{2(\sigma_{\mathrm{E}}^{*})^2}}. \tag{12}$$

3) The maximum of the density of the normal distribution $N_{\max}^{*}(\varepsilon)$ of the noise is determined, which is located in the point m_{E} =0:

$$N_{\max}^{*}(0) = \frac{1}{\sigma_{\mathrm{E}}^{*}\sqrt{2\pi}} =$$

$$= \begin{cases} \dfrac{1}{\sqrt{2\pi\cdot\left(R_{GG}(0) - R_{GG}(\Delta t)\right)}} & \mathit{for\ the\ special\ case} \\ \dfrac{1}{\sqrt{2\pi\cdot\left(R_{GG}(\mu=0) - 2R_{GG}(\mu=\Delta t) + R_{GG}(\mu=2\Delta t)\right)}} & \mathit{for\ the\ general\ case} \end{cases} \tag{13}$$

4) The inflection points $\left(-\sigma_{\mathrm{E}}^{*}; \dfrac{1}{\sigma_{\mathrm{E}}^{*}\sqrt{2\pi e}}\right)$ and $\left(\sigma_{\mathrm{E}}^{*}; \dfrac{1}{\sigma_{\mathrm{E}}^{*}\sqrt{2\pi e}}\right)$ are determined:

– for the first and second points along the abscissa:

$$AE1^{*} = \begin{cases} -\sqrt{\left(R_{GG}(0) - R_{GG}(\Delta t)\right)} & \mathit{for\ the\ special\ case} \\ -\sqrt{\left(R_{GG}(\mu=0) - 2R_{GG}(\mu=\Delta t) + R_{GG}(\mu=2\Delta t)\right)} & \mathit{for\ the\ general\ case} \end{cases} \tag{14}$$

$$AE2^{*} = \begin{cases} \sqrt{\left(R_{GG}(0) - R_{GG}(\Delta t)\right)} & \mathit{for\ the\ special\ case} \\ \sqrt{\left(R_{GG}(\mu=0) - 2R_{GG}(\mu=\Delta t) + R_{GG}(\mu=2\Delta t)\right)} & \mathit{for\ the\ general\ case} \end{cases} \tag{15}$$

– for the first and second points along the ordinate:

$$OE^{*} = \begin{cases} \dfrac{1}{\sqrt{2\left(R_{GG}(0) - R_{GG}(\Delta t)\right)\pi e}} & \mathit{for\ the\ special\ case} \\ \dfrac{1}{\sqrt{2\left(R_{GG}(\mu=0) - 2R_{GG}(\mu=\Delta t) + R_{GG}(\mu=2\Delta t)\right)\pi e}} & \mathit{for\ the\ general\ case} \end{cases} \tag{16}$$

However, it is known that the confidence interval of the estimates of the mathematical expectation of the noise in a sample with the known mean squarre deviation σ_{ε} is calculated from the expression [21]: $\left(m_{\mathrm{E}} - z_p \cdot \frac{\sigma_{\mathrm{E}}}{\sqrt{N}}; m_{\mathrm{E}} + z_p \cdot \frac{\sigma_{\mathrm{E}}}{\sqrt{N}}\right)$, where z_p is the the critical value of the distribution, which can be found by setting a certain confidence probability $p = 1 - \alpha = \Phi(z)$; $\Phi(z)$ is the Laplace function. To build an interval with a 95% confidence level, цу must select α = 0,05. For instance, for the probability p=0.95 we have $z_{0.95}$=1.65; N is the sample size.

Then, taking into account expressions (8), (9), we can calculate the confidence interval for mathematical expectation m_{E}^* of the noise in the sample:

$$m_{\mathrm{E}} - z_p \cdot \frac{\sigma_{\mathrm{E}}^*}{\sqrt{N}} \le m_{\mathrm{E}}^* \le m_{\mathrm{E}} + z_p \cdot \frac{\sigma_{\mathrm{E}}^*}{\sqrt{N}}.$$

Since $m_{\mathrm{E}} = 0$ and the mathematical expectation of the sample cannot be a negative number, we get $0 \le m_{\mathrm{E}}^* \le z_p \cdot \frac{\sigma_{\mathrm{E}}^*}{\sqrt{N}}$.

Then estimates of the distribution density at the lower limit of the confidence interval with the mathematical expectation $m_{\mathrm{E}} = 0$ will be calculated from expressions (12)-(16).

3.2. Algorithms for determining the upper limit of the confidence interval.

At the upper limit of the confidence interval with the mathematical expectation $m_{\mathrm{E}v}^* = z_p \cdot \frac{\sigma_{\mathrm{E}}^*}{\sqrt{N}}$, formula for calculating the noise distribution density (11) takes the form:

$$N_v^*(\varepsilon) = \frac{1}{\sigma_{\mathrm{E}}^* \sqrt{2\pi}} e^{-\frac{\left(\varepsilon - z_p \cdot \frac{\sigma_{\mathrm{E}}^*}{\sqrt{N}}\right)^2}{2\left(\sigma_{\mathrm{E}}^*\right)^2}}, \quad (17)$$

where the estimates of the distribution density of the noise should be calculated in the interval: $\varepsilon_{\min} = z_p \cdot \frac{\sigma_{\mathrm{E}}^*}{\sqrt{N}} - 3\sigma_{\mathrm{E}}^*$; $\varepsilon_{\max} = z_p \cdot \frac{\sigma_{\mathrm{E}}^*}{\sqrt{N}} + 3\sigma_{\mathrm{E}}^*$.

In this case, the maximum of the density of the normal distribution $N_{\max}^*(\varepsilon)$ of the noise is determined, which is located in the point $m_{\mathrm{E}}^* = z_p \cdot \frac{\sigma_{\mathrm{E}}^*}{\sqrt{N}}$ will also be calculated from expression (13).

However, the inflection points $\left(m_{\mathrm{E}}^* - \sigma_{\mathrm{E}}^*; \frac{1}{\sigma_{\mathrm{E}}^* \sqrt{2\pi e}}\right)$ and $\left(m_{\mathrm{E}}^* + \sigma_{\mathrm{E}}^*; \frac{1}{\sigma_{\mathrm{E}}^* \sqrt{2\pi e}}\right)$ are determined from the expression:

$$\left(z_p \cdot \frac{\sigma_{\mathrm{E}}^*}{\sqrt{N}} - \sigma_{\mathrm{E}}^*; \frac{1}{\sigma_{\mathrm{E}}^* \sqrt{2\pi e}}\right) \text{ and } \left(z_p \cdot \frac{\sigma_{\mathrm{E}}^*}{\sqrt{N}} + \sigma_{\mathrm{E}}^*; \frac{1}{\sigma_{\mathrm{E}}^* \sqrt{2\pi e}}\right).$$

It follows from this expression that the coordinates for the first and second inflection points along the abscissa vary depending on the nature and degree of the defect, while the coordinates of the first and second inflection points along the ordinate remain unchanged.

4. Algorithms for using the confidence interval of the estimates of the distribution density of the noise for early detection of malfunctions of industrial facilities and monitoring the dynamics of their development

To detect the appearance of defects at an early stage, it is necessary to create a databank consisting of the estimates of the characteristics of the noise obtained at different times t_i, $i = 1,2,3,\dots$. The following algorithm is proposed for processing this data..

1. At the instant t_1, the values of the noise characteristics $D^*_{E_{t1}}$, $\sigma^*_{E_{t1}}$, $N^*_{t1}(\varepsilon)$, $N^*_{t1,\max}(m^*_E)$, $A1^*_{t1}$, $A2^*_{t1}$, O^*_{t1} are calculated from expressions (8)-(17). If all these values are equal to some steady-state value corresponding to an operable value $D^*_{E_{t1}} \approx D_{const}$, $\sigma^*_{E_{t1}} \approx \sigma_{const}$, $N^*_{t1}(\varepsilon) \approx N_{const}(\varepsilon)$, $N^*_{t1,\max}(m^*_E) \approx N_{const}(m^*_E)$, $A1^*_{t1} \approx A1_{const}$, $A2^*_{t1} \approx A2_{const}$, $O^*_{t1} \approx O_{const}$, this means that there are no malfunctions in the technical condition of the facility. These values are entered into the databank of informative operable attributes S_{t1}.

2. Then at the instant t_2, the values of the similar noise characteristics $D^*_{E_{t2}}$, $\sigma^*_{E_{t2}}$, $N^*_{t2}(\varepsilon)$, $N^*_{t2,\max}(m^*_E)$, $AE1^*_{t2}$, $AE2^*_{t2}$, OE^*_{t2} are calculated and entered into the databank of informative attributes S_{t2}.

The analysis of the noise estimates obtained at the instants t_1 and t_2 is carried out and conclusions are drawn. If $D^*_{E_{t1}} = D^*_{E_{t2}}$, $\sigma^*_{E_{t1}} = \sigma^*_{E_{t2}}$, $N^*_{t1}(\varepsilon) = N^*_{t2}(\varepsilon)$, $N^*_{t1,\max}(m^*_E) = N^*_{t2,\max}(m^*_E)$, $AE1^*_{t1} = AE1^*_{t2}$, $AE2^*_{t1} = AE2^*_{t2}$, $OE^*_{t1} = OE^*_{t2}$, then no malfunction is observed. Then the state of the production facility is considered stably operable. If $D^*_{E_{t1}} \neq D^*_{E_{t2}}$, $\sigma^*_{E_{t1}} \neq \sigma^*_{E_{t2}}$, $N^*_{t1}(\varepsilon) \neq N^*_{t2}(\varepsilon)$, $N^*_{t1,\max}(m^*_E) \neq N^*_{t2,\max}(m^*_E)$, $AE1^*_{t1} \neq AE1^*_{t2}$, $AE2^*_{t1} \neq AE2^*_{t2}$, $OE^*_{t1} \neq OE^*_{t2}$, there is a malfunction. Then a correspondence is established between the values of the noise characteristics and the type of defect for the vector of informative attributes S_{t2}.

3. Then the noise characteristics $D^*_{E_{t3}}$, $\sigma^*_{E_{t3}}$, $N^*_{t3}(\varepsilon)$, $N^*_{t3,\max}(m^*_E)$, $AE1^*_{t3}$, $AE2^*_{t3}$, OE^*_{t3} are calculated at the instant t_3; the obtained values are also entered into the databank of informative attributes S_{t3}. If $D^*_{E_{t3}} > D^*_{E_{t2}}$, $\sigma^*_{E_{t3}} > \sigma^*_{E_{t2}}$, it means that the malfunction is in the process of development. Then preventive maintenance should be carried out in the normal operation mode.

If $D^*_{E_{t1}} >> D^*_{E_{t2}}$, $\sigma^*_{E_{t1}} >> \sigma^*_{E_{t2}}$, it means that damage is developing rapidly. Then, in the vector of informative attributes, these values of the noise characteristics are associated with the state of "rapidly developing degree of defect". In this case, it is necessary to stop the operation of the equipment in order to avoid an accident.

4. Then the noise characteristics are calculated at the instants t_4, t_5, …, t_n, the databank of informative attributes S_{t4}, S_{t5}, …, S_{tn} is compiled for each instant; for each of these vectors, a correspondence is established between the values of the noise characteristics and the degree of development of this type of defect. A matrix of informative attributes of the dynamics of the development of this type of malfunction of a technical object is built:

$$S_t = \begin{pmatrix} D^*_{E_{t1}} & \sigma^*_{E_{t1}} & N^*_{t1}(\varepsilon) & N^*_{t1,\max}(m^*_E) & AE1^*_{t1} & AE2^*_{t1} & OE^*_{t1} \\ D^*_{E_{t2}} & \sigma^*_{E_{t2}} & N^*_{t2}(\varepsilon) & N^*_{t2,\max}(m^*_E) & AE1^*_{t2} & AE2^*_{t2} & OE^*_{t2} \\ D^*_{E_{t3}} & \sigma^*_{E_{t3}} & N^*_{t3}(\varepsilon) & N^*_{t3,\max}(m^*_E) & AE1^*_{t3} & AE2^*_{t3} & OE^*_{t3} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ D^*_{En} & \sigma^*_{E_{tn}} & N^*_{tn}(\varepsilon) & N^*_{tn,\max}(m^*_E) & AE1^*_{tn} & AE2^*_{tn} & OE^*_{tn} \end{pmatrix} \quad (18)$$

5. The state vector is built:

$$S = \{S_{t1} \quad S_{t2} \quad \dots \quad S_{tn}\}. \quad (19)$$

6. To each of the informative attributes, the degree of damage development is assigned.:

$$S = \begin{cases} S_{t1} & \text{no damage} \\ S_{t2} & \text{minor damage appeared} \\ S_{t3} & \text{major damage appeare} \\ S_{t4} & \text{damage is developing with minor intensity} \\ S_{t4} & \text{damage is developing with great intensity} \\ \dots & \dots \\ S_{tn} & \text{damage is severe, developing rapidly} \end{cases} \quad (20)$$

7. A table of actions to be carried out in each individual case is compiled:

Table 1

Technical condition of the facility and relevant actions

State	Degree of damage	Required action
S_{t1}	No damage	Continue operation
S_{t2}	Minor damage appeared	Conduct preventive maintenance
S_{t3}	Major damage appeared	Conduct repairs
S_{t4}	Damage is developing with minor intensity	Conduct repairs in the process of normal operation
S_{t5}	Damage is developing with great intensity	Stop operation and conduct repairs
…		
S_{tn}	Damage is severe, developing rapidly	Stop operation, call the emergency crew and conduct repairs

8. Similar matrices of informative attributes of the dynamics of the development of malfunctions of technical objects are built for other types of defects. This allows detecting the moment of emergence of the defect and conduct early monitoring of accidents.

5. Using the confidence interval of the distribution density of the noise for early monitoring of compressor station malfunction.

In the following paragraphs, we shall demonstrate the possibility of using the confidence interval of the distribution density of the noise to determine the early stage of the initiation of a compressor station malfunction and the intensity of its development.

At present, at oilfields, oil and gas pipelines, and oil refineries, compressor stations are widely used to provide the transfer of oil and gas. After a certain period of normal operation of compressor stations, the period of latent transition to an emergency state begings.

The majority of failures of oil and gas pumping units occur in the mechanical part and the lubrication system and are caused by malfunctions of mounting groups, gearboxes, couplings, pumps. For instant, a typical bearing damage is the appearance of scratches, rubbings, cracks, melting, etc. on the surface. Failures of gears occur as a result of surface layer fatigue, which leads to local surface damage or peeling of individual particles of material. Damage to gear wheels is seizing, plastic deformation, peeling. Damage of gear couplings is abrasive wear, dents and ridges on working surfaces, tooth breakage.

Therefore, the task of monitoring the technical condition of a compressor station comes down to providing a reliable indication of the beginning of the time of occurrence of one of these defects.

Currently, monitoring and control systems cannot ensure registration of the beginning of the latent period of the transition of the facility into an emergency state. This is one of the drawbacks of compressor station control systems. In the following paragraphs, we consider the possibility of determining the early latent period of the onset of a fault and the degree of its development at a compressor station as a result of calculating the confidence intervals of the mathematical expectation and the distribution density of the noise of noisy signals coming from vibration sensors mounted on the electric motor, on the first and second gearboxes, on low and high pressure cylinders, mounting groups.

The following algorithm of early monitoring is proposed. The noise matrix S_t of informative attributes is built (18) and the state vector S is created (19), (20). If at some instant in time the value of one of the estimates D^*_{E}, σ^*_{E}, $N^*(\varepsilon)$, $N^*_{\max}(m^*_{\mathrm{E}})$, $A1^*$, $A2^*$, O^* differs from the reference values D_{const}, σ_{const}, $N_{const}(\varepsilon)$, $N_{const}(m^*_{\mathrm{E}})$, $A1_{const}$, $A2_{const}$, O_{const}, it means that a defect has occurred and a correspondence is established between the values of the noise characteristics and the type of defect. At the same time, the corresponding element is selected in the state vector S and the type and severity of the damage is noted. Types of defect are recorded in coded form:

1 – scratches on the bearing surface;
2 – rubbing on the bearing surface;
3 – cracks on the bearing surface, etc.;
4 – damage on the surface of gears;
5 – seizing of gear wheels;
6 – plastic deformation of the gears;
7 – gears peeling;
8 – abrasive wear of gear couplings;
9 – dents and ridges on the working surfaces of gear couplings;
10 – breakage of the teeth of gear couplings, etc.

After determining the type and degree of compressor station malfunction, the monitoring system generates information on the actions that need to be taken in each specific case: 1 – repair is not required; 2 – regular maintenance is required; 3 – major repairs are required; 4 – the damaged part must be replaced; 5 – the operation of the unit must stop, etc.

6. Conclusions

The application of the developed algorithms for determining the confidence interval of the estimates the distribution density of the noise in monitoring and control systems allows calculating the minimum and maximum threshold values, beyond which we can draw a conclusion about the occurrence of a fault and determine the intensity of its development at an early stage. When the system operates in noise control mode during operation of a unit, a state vector, the elements of which correspond to various faults of the object under investigation in the latent period of fault initiation, is built using combinations of confidence intervals of the estimates the distribution density of the noise of all analyzed signals. This allows assessing the technical condition of the facility as a whole and conduct early monitoring of accidents [14-22].

The possibility of practical application of the developed algorithms is considered on the example of identifying the latent period of defect initiation and determining the dynamics of its development at a compressor station.

References

[1] А.И. Заико, Измерение плотности вероятности эргодического случайного процесса с линейной корреляционной функцией, Вестник УГАТУ. 21 No.4(78) (2017) 121–128. [In Russian: A.I. Zaiko, Measurement of the probability density of an ergodic random process with a linear correlation function, Vestnik UGATU]. http://journal.ugatu.ac.ru/index.php/Vestnik/article/view/113

[2] A.I. Zaiko, Random signal with uniform distribution, Measurement Techniques. 42 (1999) 11–13. https://doi.org/10.1007/BF02504193

[3] T.A. Aliev, Digital Noise Monitoring of Defect Origin, Springer, New York, 2007. https://doi.org/10.1007/978-0-387-71754-8
[4] T.A. Aliev, Noise Control of the Beginning and Development Dynamics of Accidents, Springer, 2019. https://doi.org/10.1007/978-3-030-12512-7
[5] T.A. Aliyev, N.F. Musaeva, An algorithm for Eliminating Microerrors of Noise in the Solution of Statistical Dynamics Problems, Automation and Remote Control. 59 No.5 (1998) 679-688. http://www.mathnet.ru/links/b526a48db645dc7b27175c25b11b1610/at2603.pdf
[6] N.F. Musaeva, Robust correlation coefficients as initial data for solving a problem of confluent analysis, Automatic Control and Computer Sciences. 41 No.2 (2007) 76-87. https://doi.org/10.3103/S0146411607020034
[7] N.F. Musaeva, Robust method of estimation with “contaminated” coarse errors, Automatic Control and Computer Sciences. No.6. (2003) 50-63. https://www.researchgate.net/publication/293583902_Robust_method_of_estimation_with_contaminated_coarse_errors
[8] T.A. Aliev, N.F. Musaeva, M.T. Suleymanova, B.I. Gazizade, Analytic representation of the density function of normal distribution of noise, Journal of Automation and Information Sciences. 47(8) No.4 (2015) 24-40. https://doi.org/10.1615/JAutomatInfScien.v47.i8.30
[9] T.A. Aliev, N.F. Musaeva, M.T. Suleymanova, B.I. Gazizade, Technology for calculating the parameters of the density function of normal distribution of the useful component in a noisy process, Journal of Automation and Information Sciences. 48 No.4 (2016) 35-55. https://doi.org/10.1615/JAutomatInfScien.v50.i11.30
[10] T.A. Aliev, N.F. Musaeva, B.I. Gazizade, Algorithms of building a model of the noisy process by correction of the law of its distribution, Journal of Automation and Information Sciences. 49 No.9 (2017) 61-75. https://doi.org/10.1615/ JAutomatInfScien.v49.i9.50
[11] T.A. Aliev, N.F. Musaeva, M.T. Suleymanova, Density Function of Noise Distribution as an Indicator for Identifying the Degree of Fault Growth in Sucker Rod Pumping Unit (SRPU), Journal of Automation and Information Sciences. 49 No.4 (2017) 1-11. https://doi.org/10.1615/JAutomatInfScien.v49.i4.10
[12] T.A. Aliev, N.F. Musaeva, M.T. Suleymanova, Algorithms for Indicating the Beginning of Accidents Based on the Estimate of the Density Distribution Function of the Noise of Technological Parameters, Automatic Control and Computer Sciences. 52 No.3 (2018) 231-242. https://doi.org/10.3103/S0146411618030021
[13] T.A. Aliev, N.F. Musaeva, B.I. Gazizade, Calculation Algorithms of the High Order Moments of Interference of Noisy Signals, Journal of Automation and Information Sciences. 50 No.6 (2018) 1-13. https://doi.org/10.1615/JAutomatInfScien.v50.i6.10
[14] Т.А. Алиев, Н.Ф. Мусаева, М.Т. Сулейманова, Б.И. Газызаде, Чувствительные алгоритмы выявления степени развития неисправности штанговой глубинной насосной установки, Мехатроника, автоматизация, управление, Москва. 18 No.2. (2017) 94-102. [In Russian: T.A. Aliev, N.F. Musaeva, M.T. Suleymanova, Sensitive Algorithms for Identifying the Degree of Fault Growth in Sucker Rod Pumping Units]. https://doi.org/10.17587/mau.18.91-102
[15] T.A. Aliev, G.A. Guluyev, F.H. Pashayev, A.B. Sadygov, Noise monitoring technology for objects in transition to the emergency state, Mechanical Systems and Signal Processing. No.27 (2012) 755–762.
[16] T.A. Aliev, T.A. Alizada, N.E. Rzayeva, Noise technologies and systems for monitoring the beginning of the latent period of accidents on fixed platforms, Mechanical Systems and Signal Processing. 87 Part.A (2017) 111–123. https://doi.org/10.1016/j.ymssp.2016.10.014
[17] T.A. Aliev, O.G. Nusratov, G.A. Guluev, As.G. Rzaev, F.G. Pashaev, M.G. Rezvan, Position-Binary Technology for Processing the Signals of a Load for Identification of the Technical condition of Deep-Well Sucker-Rod Pumps, Measurement Techniques. 61 No.9 (2018) 885-890. https://doi.org/10.1007/s11018-018-1519-x
[18] Е.С. Вентцель, Л.А. Овчаров, Теория случайных процессов и ее инженерные приложения, 5-е изд., Москва, КНОРУС, (2013). [In Russian: E.S. Wentzel, L.A. Ovcharov, The Theory of Random Processes and Its Engineering Applications, 5th ed., Moscow, KNORUS]. https://www.studmed.ru/ventcel-es-ovcharov-la-teoriya-sluchaynyh-processov-i-ee-inzhenernye-prilozheniya_32e9b501bdd.html https://www.twirpx.com/file/36748/
[19] Техническая кибернетика, Книга 2, Под ред. Солодовникова В.В., Москва, Машиностроение, (1967) 682 с. [In Russian: Technical Cybernetics, Book 2, Edited by Solodovnikov V.V., Moscow, Mashinostroenie]. https://www.libex.ru/detail/book29233.html

[20] В.И. Тихонов, Статистическая радиотехника, Москва, Радио и связь, (1982) 624 с. [In Russian: V.I. Tikhonov, Statistical Radio Engineering, Radio i Svyaz]. https://www.twirpx.com/file/1641910/
[21] В.М. Иванова, В.Н. Калинина, Л.А. Нешумова, И.О. Решетникова, Математическая статистика, Москва, Высшая школа, (1975) 398 с. [In Russian: V.M. Ivanova, V.N. Kalinina, L.A. Neshumova, I.O. Reshetnikova, Mathematical Statistics, Moscow, Vysshaya Shkola]. https://edu-lib.com/matematika-2/dlya-studentov/matematicheskaya-statistika-uchebnik-dlya-tehnikumov-dlina
[22] A.M. Abbasov, M.H. Mamedova, G.H. Orujov, H.B. Aliyev, Synthesis of the methods of subjective knowledge representations in problems of fuzzy pattern recognition, Mechatronics. No.11 (2001) 439-449.

Reviewer: Associate Professor A.K. Kazim-zadeh, PhD (Engineering)