ISSN 0146-4116, Automatic Control and Computer Sciences, 2018, Vol. 52, No. 3, pp. 231–242. © Allerton Press, Inc., 2018.
Published in Russian in Avtomatika i Vychislitel'naya Tekhnika, 2018, No. 3, pp. 55–69.

Algorithms for Indicating the Beginning of Accidents Based on the Estimate of the Density Distribution Function of the Noise of Technological Parameters[1]

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

[a] *Institute of Control Systems of the ANAS, Baku, AZ1141 Azerbaijan*

[b] *Azerbaijan University of Architecture and Construction, Baku, AZ1073 Azerbaijan*

**e-mail: musanaila@gmail.com*

Received April 28, 2017; in final form, December 30, 2017

Abstract—A technology has been developed, which allows for calculating the probability density function of noise, its maximum and inflection points, using the discrete values of a signal corrupted by an additive random noise. Computational experiments have been conducted. It has been demonstrated that knowledge of those characteristics of noise allows systems of monitoring, control, diagnostics, forecasting, identification, management, etc. to register not only the initial period of fault origin, but also the moment when preventive maintenance measures, routine or major overhaul works are required.

Keywords: stochastic process, noise, noisy signal, probability density function of noise, inflection points of probability density function of noise

DOI: 10.3103/S0146411618030021

1. INTRODUCTION

It is a known fact that in most cases signals transmitted and received in diagnostics, forecasting, control, identification and other systems are distorted by external disturbances, which are regarded as interferences. Causes and sources of additive interferences (or noises) $\mathrm{E}(t)$ that are superimposed on the useful signal $X(t)$ and lead to the appearance of the noisy signal $G(t) = X(t) + \mathrm{E}(t)$ can be a variety of factors. In this case, if the cause of the noise is known and it is regular, then there are many methods for its elimination [1–5]. However, noise is very diverse both in its origin and in its physical properties. Therefore, the methods of managing them differ somewhat from each other, for instance, methods of noise suppression; building receivers that are noise-insensitive; minimization of noise transmission through communication channels, filtration methods, etc. [2–5].

However, among many different types of noise, there are also those caused by the influence of various destabilizing factors, such as defects, wear, corrosion, cracks, breakdowns and other malfunctions of equipment, devices, structures, engine, mechanism, motor, etc. [6, 7]. To solve these problems, methods for calculating the noise characteristics of noisy signals are developed [6–11]. In such cases, the noise signals the onset of defects or malfunctions. Such random noise, as a rule, appears as early as at the initial stage of an incipient defect, when nothing is known about the defect itself in practice and it cannot be detected. For instance, the noise caused by the presence of contamination or water vapor at the joint of two metals, noise from the influence of microcavities in the walls of the well during logging, etc. [4]. Information about the existence of defects arrives much later, only when they assume a clearly expressed form and a repair is required. Therefore, if we detect such noise in time and calculate their characteristics, they can become informational, allowing us to find out the nature of defects at an early stage, and thus prevent possible breakdowns, failures, accidents, etc. In this case, the information about the noise is useful information about the abnormal situation that has occurred, and the noise itself is regarded as the carrier of this information.

It is known that the interference is of random nature and is a random function with random amplitude and phase and a higher frequency than the useful signal. In addition, it is traditionally assumed that inter-

[1] The article is published in the original.

ference is white noise and is described by a normal distribution law with a zero mathematical expectation [1–5]. However, in practice, white noise is regarded as a mathematical idealization, since all real processes always have a spectral density decreasing at very high frequencies, and therefore have a finite correlation time $\tau_{\text{fin}} \neq 0$ and a limited average power [5].

But this information alone is not sufficient to solve the aforementioned problems, since such basic characteristics of random noise as high-order moments, the distribution density function, probability of random noise getting into some interval remain unknown. At the same time, it is these characteristics that contain a sufficient amount of information about the properties of the noise and are exhaustive for solving the above problems.

There are appropriate formulas to calculate the characteristics of noise in the theory of analysis of random processes [4, 5]. However, for the practical application of these formulas, it is necessary to know the discrete values of the additive random noise $\mathrm{E}(t)$, which cannot be isolated from the noisy signal $G(t)$. Therefore, in practice, we confine ourselves to calculating the characteristics of the noisiest random process $G(t)$, forming and controlling observations [11–15], suppressing noise [3], determining the stopping time [2], etc.

At the same time, the extraction of useful components from noisy recorded signals or the determination of the characteristics of noise and interference to extract useful information from them is one of the principal objectives of primary signal processing [5, 12–15]. Therefore, the present paper proposes a detection technology based on discrete observations representing an additive mixture of unobserved useful signal and normally distributed random noise $\mathrm{E}(t)$ with a zero mathematical expectation, the density distribution function of the noise and the resulting characteristics.

2. PROBLEM STATEMENT

In the time interval $0 \le t \le T$, a continuous random stationary ergodic noisy technological process $G(t)$ is observed, consisting of the sum of the random useful component $X(t)$ and random noise $\mathrm{E}(t)$, which are also stationary ergodic and cannot be isolated from $G(t)$. The random process $G(t)$ contains information about one technological parameter being examined, e.g., temperature, pressure, flow, etc. and can comply with different distribution laws.

For the random process $G(t)$, it is possible to calculate selective estimates of such characteristics as mathematical expectation m_g, variance D_g, mean square deviation σ_g, correlation function $R_{GG}(\tau)$ from to the formulas:

$$m_g = \frac{1}{T}\int_0^T G(t)dt,$$

$$D_g = \frac{1}{T}\int_0^T (G(t) - m_g)^2 dt = \frac{1}{T}\int_0^T \overset{\circ}{G}{}^2(t)dt,$$

$$\sigma_g = \sqrt{D_g},$$

$$R_{gg}(\tau) = \frac{1}{T}\int_0^T \overset{\circ}{G}(t)\overset{\circ}{G}(t+\tau)dt, \tag{1}$$

where $\overset{\circ}{G}(t) = G(t) - m_g$, $\tau = 0, \Delta t, 2\Delta t, 3\Delta t, \ldots$ is the time shift.

The useful component $X(t)$ evaluates the current state of the process under study. It is known a priori that the noise in a system of monitoring, control, diagnostics, forecasting, management, identification, etc. is caused by defects, faults, malfunctions, etc. and has normal distribution $N(\varepsilon; m_\varepsilon, \sigma_\varepsilon)$ and zero mean $m_\varepsilon = 0$. Moreover, within the amplitude-frequency characteristic of the system under investigation at the time of the defect generation, the correlation time of the noise $\mathrm{E}(t)$ is much shorter than that of the useful signal $X(t)$.

Since the stationary random noise $\mathrm{E}(t)$ is ergodic, its mathematical expectation m_ε and the mean square deviation σ_ε have the same value for any of the random functions in the set. Therefore, the density function of the normal distribution $N(\varepsilon, m_\varepsilon, \sigma_\varepsilon) = N(\varepsilon)$ of the noise can be represented as:

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

It is obvious from formula (2) that to determine the density function of the distribution $N(\varepsilon)$ of the noise $\mathrm{E}(t)$ one must know the mean square deviation σ_ε, which is unknown, and its value cannot be isolated from the noisy process $G(t)$. Then, in the form and dynamics of the change in the density function of the distribution $N(\varepsilon)$ of the noise $\mathrm{E}(t)$, one can assess the changes occurring in the technical condition of the object under study, i.e., extract the necessary useful information. This is due to the fact that in practice, real technological processes are not strictly stationary. These processes are quasi-stationary, that is, they can be regarded as stationary at some time interval.

Therefore, in monitoring, control, diagnostics, forecasting, management, identification and other systems, at the initial stage of defects, faults, malfunctions, etc., the noise is subject to the normal distribution law with a certain form. After a certain period of time, as the degree of failure increases, the distribution density function changes its form. When a defect acquires a clearly expressed form, the law of distribution can change to a law different from normal. Therefore, the problem arises of determining the density function of the normal distribution $N(\varepsilon)$ of the noise $\mathrm{E}(t)$. This will allow:

(1) to reveal the dynamics of the change in the form of the normal noise curve $N(\varepsilon)$ in time, and, accordingly, the dynamics of the change in the most probable values (which fall in the interval $m_\varepsilon - \sigma_\varepsilon \le \mathrm{E}(t) \le m_\varepsilon + \sigma_\varepsilon$), the frequently occurring values (which fall into intervals $m_\varepsilon - 2\sigma_\varepsilon \le \mathrm{E}(t) \le m_\varepsilon - \sigma_\varepsilon$ and $m_\varepsilon + \sigma_\varepsilon \le \mathrm{E}(t) \le m_\varepsilon + 2\sigma_\varepsilon$) and the unlikely values (which fall into the intervals $m_\varepsilon - 3\sigma_\varepsilon \le \mathrm{E}(t) \le m_\varepsilon - 2\sigma_\varepsilon$ and $m_\varepsilon + 2\sigma_\varepsilon \le \mathrm{E}(t) \le m_\varepsilon + 3\sigma_\varepsilon$) of the noise of the investigated noisy parameter $G(t)$, and determine the values that are accepted with probabilities:

$$P(|\mathrm{E}(t) - m_\varepsilon| < \sigma_\varepsilon) = 0.6827;$$
$$P(|\mathrm{E}(t) - m_\varepsilon| < 2\sigma_\varepsilon) = 0.9545;$$
$$P(|\mathrm{E}(t) - m_\varepsilon| < 3\sigma_\varepsilon) = 0.9973;$$

(2) determine the maximum of the density function of the normal distribution $N_{\max}(\varepsilon)$ of the noise of the investigated noisy parameter $G(t)$ and establish the correspondence of its value to each faulty state of the object;

(3) determine the coordinates of the inflection points of the density function of the normal distribution of the noise of the investigated noisy parameter $G(t)$ and establish the correspondence of their values to each faulty state of the object.

If the matrix of informative features is formed, the elements of which are the mean square deviation, the density function of the normal distribution, then it is possible to determine not only the initial period of the defect generation, but also the moments when it is necessary to carry out preventive maintenance, current or major repairs. Therefore, an algorithm for determining the density function of the normal distribution $N(\varepsilon)$ of the noise is proposed below.

3. TECHNOLOGY FOR CALCULATING THE SECOND-ORDER MOMENT OF THE NOISE OF THE NOISY SIGNAL

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: mathematical expectation m_ε and mean square deviation $\sigma_\varepsilon = \sqrt{D_\varepsilon}$ (or square root of the variance). Since the noise is distributed $\mathrm{E}(t)$ according to the normal law with zero mean $m_\varepsilon = 0$, the problem comes down to calculating only the parameter σ_ε. We first calculate the variance D_ε. To this end, we use expression (1) for calculating the correlation function $R_{gg}(\tau)$ of the noisy signal $G(t)$.

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

$$\frac{1}{T}\int_0^T \overset{\circ}{X}(t)\,\mathrm{E}(t+\tau)\,dt = 0;$$
$$\frac{1}{T}\int_0^T \mathrm{E}(t)\,\overset{\circ}{X}(t+\tau)\,dt = 0,$$

the following can be written:

$$R_{gg}(\tau) = R_{xx}(\tau) + R_{\varepsilon\varepsilon}(\tau), \tag{3}$$

where $R_{xx}(\tau)$, $R_{\varepsilon\varepsilon}(\tau)$ are the correlation functions of the useful signal $X(t)$ and the noise $\mathrm{E}(t)$, respectively.

In practice, for such infra-low frequency slow-flowing technological processes as oil refining, petrochemistry, when $\tau = \Delta t$ is significantly (manifold) small compared to the observation time T, the noise $\mathrm{E}(t)$ is formed from high-frequency spectra as a result of such faults as wear, corrosion, carbon formation, etc. and has a higher spectrum than the useful component $X(t)$ itself. The value of the useful component within the time interval Δt does not have time to change, and $X(t+\Delta t)$matches the value of $X(t)$, i.e.

$$X(t+\Delta t) = X(t). \tag{4}$$

This equality holds when T is, for instance, 10–20 h, and Δt seconds or minutes (depending on the specifics of the process being studied). In this case, the sampling interval Δt is selected based on the finite correlation time τ_{fin} of the noise $\mathrm{E}(t)$ with useful signal.

Obviously, such a strict equality is not true for all real processes, but for such as refining, petrochemistry. For other technological processes, an approximate equality is permissible. Then for the above industrial facilities, when condition (4) is satisfied, the relation $\frac{R_{xx}(\Delta t)}{R_{xx}(0)}$ is equal to unity, i.e. [17]:

$$R_{xx}(\Delta t) = R_{xx}(0). \tag{5}$$

At the same time, since the sampling interval Δt for the Gaussian random noise is selected based on the finite correlation time τ_{fin} of the noise, then the correlation function $R_{\varepsilon\varepsilon}(\tau)$ can be represented as follows [5]:

$$R_{\varepsilon\varepsilon}(\tau) = \begin{cases} R_{\varepsilon\varepsilon}(\tau), \text{when } \tau = 0, \\ 0, \quad\quad \text{when } \tau \geq \Delta t. \end{cases} \tag{6}$$

Therefore if, using formula (1), we calculate the estimates of the correlation function $R_{gg}(\tau)$ of the noisy signal at $\tau = 0$ and at the time interval that is sufficiently small in comparison to the observation time $\tau_{\text{fin}} = \Delta t$ and find the difference between these estimates, we will get

$$R_{gg}(0) - R_{gg}(\Delta t) = R_{xx}(0) + R_{\varepsilon\varepsilon}(0) - R_{xx}(\Delta t) - R_{\varepsilon\varepsilon}(\Delta t). \tag{7}$$

Taking into account conditions (5), (6) and the fact that the estimates of the autocorrelation functions of the useful signal $X(t)$ and the noise $\mathrm{E}(t)$, respectively, at zero time shift $\tau = 0$ are the variances of the useful signal and the noise, respectively:

$$R_{xx}(0) = D_x; \quad R_{\varepsilon\varepsilon}(0) = D_\varepsilon,$$

we get the estimate of the variance D_ε^* of the noise $\mathrm{E}(t)$ of the noisy signal $G(t)$:

$$D_\varepsilon^* = R_{\varepsilon\varepsilon}(0) = R_{gg}(0) - R_{gg}(\Delta t). \tag{8}$$

Thus, the variance D_ε^* of the noise $\mathrm{E}(t)$ can be calculated by determining the difference in the estimates of the autocorrelation function $R_{gg}(\tau)$ of the noisy signal at a zero time shift, $\tau = 0$, and a sufficiently small time shift $\tau = \Delta t$ equal to the correlation time τ_{fin} of the noise.

However, this formula is suitable for the "ideal" case, when the sampling interval is selected based on the finite correlation time τ_{fin} of the noise $\mathrm{E}(t)$. In earlier works, the formula for calculating the noise variance was derived for the more general case, when the choice of the sampling interval Δt was based on the frequency band of the spectrum of the noise $\mathrm{E}(t)$ rather than that of the useful component $X(t)$, i.e. $\Delta t = 1/2f_\varepsilon$, where f_ε is the noise cutoff frequency, Hz [7–10]:

$$D_\varepsilon^* = R_{\varepsilon\varepsilon}^*(\tau = 0) = R_{gg}(0) + R_{gg}(2\Delta t) - 2R_{gg}(\Delta t). \tag{9}$$

4. ALGORITHM FOR CALCULATING THE DENSITY FUNCTION OF THE NORMAL DISTRIBUTION OF THE NOISE OF THE NOISY SIGNAL

It will be shown in the following paragraphs that, considering the condition $m_\varepsilon = 0$, expression (2) and formulas (8), (9) for calculating the noise variance D_ε^* of the noise $\mathrm{E}(t)$, it is possible to determine the following characteristics of the noise $\mathrm{E}(t)$ of the noisy signal $G(t)$ listed in the problem statement.

(1) The distribution density function $N^*(\varepsilon)$ of the normally distributed noise $\mathrm{E}(t)$ of the noisy signal $G(t)$:

$$N^*(\varepsilon) = \frac{1}{\sqrt{2\pi D_\varepsilon^*}} e^{-\frac{\varepsilon^2}{2D_\varepsilon^*}}. \tag{10}$$

(2) The maximum of the normal distribution density function:

$$N_{\max}^*(0) = \frac{1}{\sqrt{2\pi D_\varepsilon^*}}. \tag{11}$$

(3) The coordinates of the inflection points of the normal distribution density function $N(\varepsilon)$ of the noise $\mathrm{E}(t)$:

$$\left(-\sqrt{D_\varepsilon^*};\ \frac{1}{\sqrt{2D_\varepsilon^*\pi e}}\right) \text{ and } \left(\sqrt{D_\varepsilon^*};\ \frac{1}{\sqrt{2D_\varepsilon^*\pi e}}\right). \tag{12}$$

Thus, we have developed algorithms for calculating the distribution density function $N^*(\varepsilon)$, its maximum $N_{\max}^*(0)$, and the inflection points of the normally distributed noise $\mathrm{E}(t)$ of the noisy signal $G(t)$.

5. TECHNOLOGY FOR CALCULATING THE DISCRETE VALUES OF THE DENSITY FUNCTION OF THE NORMAL DISTRIBUTION OF THE NOISE OF THE NOISY SIGNAL AND CONSEQUENTIAL CHARACTERISTICS

We propose in the following paragraphs the discrete algorithms for determining the distribution density function $N^*(\varepsilon)$ of the normally distributed noise $\mathrm{E}(t)$ with the mathematical expectation $m_\varepsilon = 0$, maximum $N_{\max}^*(0)$, inflection points with the coordinates from (12).

Assume that the noisy digital signal $G(\Delta t)$ consisting of the useful signal $X(\Delta t)$ and the additive noise $\mathrm{E}(\Delta t)$ arrive from a sensor located within the range of factors influencing the facility and receiving digital information from this facility. The signal $G(\Delta t)$ is sampled at the interval Δt selected based on the finite correlation time τ_{fin} of the noise. Then the time interval T consists of N very small intervals Δt, i.e. $T = N\Delta t$, and the signal $X(t)$ varies only slightly over the interval $t + \Delta t$. If we give t and τ discrete values that are multiples of Δt, i.e. $t = \nu\Delta t$, $\nu = 1, 2, \ldots$; $\tau = \mu\Delta t$, $\mu = 0, 1, \ldots$ and introduce the notation $R_{gg}(\mu\Delta t) = R_{gg}(\mu)$; $R_{xx}(\mu\Delta t) = R_{xx}(\mu)$, $R_{\varepsilon\varepsilon}(\mu\Delta t) = R_{\varepsilon\varepsilon}(\mu)$ for the estimates of the correlation functions, then the algorithm for determining the distribution density function $N^*(\varepsilon)$ of the noise is represented as follows:

(1) The estimates of the autocorrelation function of the noisy signal at $\mu = 0$ and $\mu = \Delta t$ are calculated:

$$R_{\overset{\circ}{g}\overset{\circ}{g}}(0) = \frac{1}{N}\sum_{i=1}^{N}\overset{\circ}{G}(i\Delta t)\overset{\circ}{G}(i\Delta t), \quad (13)$$

$$R_{\overset{\circ}{g}\overset{\circ}{g}}(\Delta t) = \frac{1}{N}\sum_{i=1}^{N}\overset{\circ}{G}(i\Delta t)\overset{\circ}{G}((i+1)\Delta t), \quad (14)$$

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

(2) The variance and the mean square deviation of the noise $\mathrm{E}(t)$ of the noisy signal $G(t)$ are calculated:

$$D_{\varepsilon}^{*} = R_{\varepsilon\varepsilon}^{*}(0) = R_{\overset{\circ}{g}\overset{\circ}{g}}(0) - R_{\overset{\circ}{g}\overset{\circ}{g}}(\Delta t) = \frac{1}{N}\sum_{i=1}^{N}\overset{\circ}{G}(i\Delta t)\overset{\circ}{G}(i\Delta t) - \frac{1}{N}\sum_{i=1}^{N}\overset{\circ}{G}(i\Delta t)\overset{\circ}{G}((i+1)\Delta t); \quad (15)$$

$$\sigma_{\varepsilon}^{*} = \sqrt{D_{\varepsilon}^{*}}. \quad (16)$$

(3) Considering that $m_{\varepsilon} = 0$ and the deviation from the mathematical expectation in absolute value for a normally distributed random parameter does not exceed tripled standard deviation, the discrete values of the distribution density function $N^{*}(\varepsilon)$ of the noise $\mathrm{E}(t)$ are calculated in the interval $\pm 3\sigma_{\varepsilon}^{*}$, i.e. at $-3\sigma_{\varepsilon}^{*} \le \mathrm{E}(t) \le +3\sigma_{\varepsilon}^{*}$. To this end:

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

—the sequence of possible values of $\mathrm{E}(t)$ is given at a certain interval $\Delta\varepsilon$, in ascending order from $\varepsilon_{\min}$ to $\varepsilon_{\max}$: $\varepsilon(1) = \varepsilon_{\min}$, $\varepsilon(i+1) = \varepsilon(i) + \Delta\varepsilon, \ldots, \varepsilon_{\max}$;

—the sequence of possible values of the noise $\varepsilon(1)$, $\varepsilon(2)$, $\varepsilon(3)$, $\varepsilon(4)$,..., $\varepsilon_{\max}$ is formed, for which the condition $\varepsilon(i-1) < \varepsilon(i)$ holds;

—the density function of the normal distribution at the points $\varepsilon(1)$, $\varepsilon(2)$, $\varepsilon(3)$, $\varepsilon(4)$,..., $\varepsilon_{\max}$ is calculated.

(4) The maximum of the normal distribution density function $N_{\max}^{*}(\varepsilon(i))$ of the noise $\mathrm{E}(\Delta t)$ of the noisy signal $G(\Delta t)$ is determined from expression (11), which is at the point $m_{\varepsilon} = 0$, i.e. $\varepsilon_{\max}(i) = 0$.

(5) The coordinates of the inflection points of the normal distribution density function $N_{\max}^{*}(\varepsilon)$ of the noise $\mathrm{E}(\Delta t)$ are determined from expression (12).

6. TECHNOLOGY OF THE COMPUTATIONAL EXPERIMENTS

To verify the validity of the algorithm for calculating the normal distribution density function $N^{*}(\varepsilon)$ of the noise $\mathrm{E}(t)$ of the noisy signal $G(t)$, the maximum $N_{\max}^{*}(\varepsilon)$ of this function, and the inflection points $\left(-\sqrt{D_{\varepsilon}^{*}}; \frac{1}{\sqrt{2D_{\varepsilon}^{*}\pi e}}\right)$ and $\left(\sqrt{D_{\varepsilon}^{*}}; \frac{1}{\sqrt{2D_{\varepsilon}^{*}\pi e}}\right)$, computational experiments were conducted using MATLAB computing environment. The computational experiments were carried out as follows.

First, the useful signal $X(t)$ was formed. Then the probability distribution was determined for discrete values of the useful signal $X(t)$, i.e. for $X(i\Delta t)$, where $i = 0, 1, 2, \ldots$, With the use of the random number generator, the normally distributed noise $\mathrm{E}(i\Delta t)$ with different preset values of the distribution parameters $m_{\varepsilon} = 0$, $\sigma_{\varepsilon} = \sqrt{D_{\varepsilon}}$ was formed. It was supposed to be the real noise. The noisy signals $G(i\Delta t) = X(i\Delta t) + \mathrm{E}(i\Delta t)$ were formed. The essence of the experiments came down to calculating the distribution density function $N^{*}(\varepsilon)$ of the noise from developed algorithms (8)–(16) using the values of the generated noisy signal $G(i\Delta t)$. The resulting distribution density function $N^{*}(\varepsilon)$ of the noise was compared with the distribution density function $N(\varepsilon)$ of the noise, which was constructed using the generated discrete values of the noise $\mathrm{E}(i\Delta t)$.

To show that a useful signal can obey different distribution laws, it is possible to put forward a hypothesis about the distribution, and then verify it by Pearson's chi-squared test (χ^2) $\chi^2 = \sum_{i=1}^{n} \frac{(l_i - Np_i)^2}{Np_i}$.

To test the hypothesis, it is necessary to calculate the empirical frequencies l_i. To do this, we need to divide the entire variation range of the useful signal $X(i\Delta t)$, consisting of N samples, into n intervals: $\Delta_1, \Delta_2, \ldots, \Delta_n$, count the number of values l_i that fell into each of the intervals Δ_i, and build a histogram. To this end, we can use the MATLAB standard function ***[l,xout]=hist(x,n)*** that counts the number of hits of $X(i\Delta t)$ in the intervals with the ***xout*** middle. Next, we calculate the theoretical frequencies Np_i of the hits in the intervals Δ_i, where p_i is the probability of the useful signal $X(i\Delta t)$ hitting the interval Δ_i. The theoretical probability p_i in MATLAB is calculated using the standard functions ***betacdf*** (beta-distribution), ***expcdf*** (exponential distribution), ***logncdf*** (lognormal distribution), ***normcdf*** (normal distribution), etc. After this, we select the significance level of the criterion α and determine the table value of Pearson's chi-squared test $\chi^2_{k;\alpha}$, where the number of degrees of freedom is $k = (n - r - 1)$, r is the number of distribution parameters. If $\chi^2 > \chi^2_{k;\alpha}$, then the hypothesis H_0 is rejected, if $\chi^2 \le \chi^2_{k;\alpha}$, then the hypothesis of the corresponding distribution law of the useful signal is admitted.

Then, using expressions (2), we calculated the distribution density function $N(\varepsilon)$, the maximum $N_{\max}(m_\varepsilon)$ of the distribution density function and the inflection points of the generated noise $\mathrm{E}(t)$ set to obtain the noisy signal $G(t)$. After that, the distribution density function $N^*(\varepsilon)$ of the noise $\mathrm{E}(t)$, the maximum $N^*_{\max}(m_\varepsilon)$ of this function and the inflection points were calculated using algorithms (8)–(16) proposed in the paper, and a comparative analysis was carried out. For this purpose, the following were determined:

(1) the relative errors of the discrete values of the distribution density function $N(\varepsilon(i))$ of the noise $\mathrm{E}(t)$, $i = 1, 2, \ldots$, in the interval $-3\sqrt{D^*_\varepsilon} \le \mathrm{E}(i\Delta t) \le 3\sqrt{D^*_\varepsilon}$:

$$\Delta N(\varepsilon(i)) = |N(\varepsilon(i)) - N^*(\varepsilon(i))| / |N(\varepsilon(i))| \times 100\%;$$

(2) the relative error of the maximum of the distribution density function:

$$\Delta N_{\max}(m_\varepsilon) = \left| N_{\max}(m_\varepsilon) - N^*_{\max}(m_\varepsilon) \right| \Big/ \left| N_{\max}(m_\varepsilon) \times 100\% \right|;$$

(3) the relative errors of the first and second inflection points along the abscissa and ordinate axes from expressions:

—along the abscissa axis for the first point $\left(m_\varepsilon - \sqrt{D_\varepsilon}\right)$ and the second point $\left(m_\varepsilon + \sqrt{D_\varepsilon}\right)$:

$$\Delta a1 = \left| \left(m_\varepsilon - \sqrt{D_\varepsilon}\right) - \left(m^*_\varepsilon - \sqrt{D^*_\varepsilon}\right) \right| \Big/ \left| \left(m_\varepsilon - \sqrt{D_\varepsilon}\right) \right| \times 100\%,$$

$$\Delta a2 = \left| \left(m_\varepsilon + \sqrt{D_\varepsilon}\right) - \left(m^*_\varepsilon + \sqrt{D^*_\varepsilon}\right) \right| \Big/ \left(m_\varepsilon + \sqrt{D_\varepsilon}\right) \times 100\%,$$

—along the ordinate axis $\left(\frac{1}{\sqrt{2 D_\varepsilon \pi e}}\right)$:

$$\Delta o = \left| \left(\frac{1}{\sqrt{2 D_\varepsilon \pi e}}\right) - \left(\frac{1}{\sqrt{2 D^*_\varepsilon \pi e}}\right) \right| \Big/ \left(\frac{1}{\sqrt{2 D_\varepsilon \pi e}}\right) \times 100\%,$$

where $m_\varepsilon = \sum_{i=1}^{N} \varepsilon(i\Delta t) \Big/ N$, and the value of m_ε^* was assigned based on the given assumption that $m_\varepsilon^* = 0$.

7. RESULTS OF THE COMPUTATIONAL EXPERIMENTS

The following experiments were performed. It is known that any stationary random process $X(t)$ on the infinite interval T can be approximated arbitrarily accurately with a linear combination of harmonic oscillations with random amplitude and phase [18]. In the general form, the set of functions [18]

$$X_k(t) = \sum_{\nu=1}^{n} \left(a_{\nu k} \cos\left(\frac{2\pi\nu}{T} t + \varphi_{1\nu l}\right) + b_{\nu k} \sin\left(\frac{2\pi\nu}{T} t + \varphi_{2\nu l}\right)\right),$$

characterizes a random process if the probability distribution functions of the coefficients $a_{\nu k}$, $b_{\nu k}$ and the phases $\varphi_{\nu k}$, $\varphi_{1\nu k}$, $\varphi_{2\nu k}$ are known. Since the values of the real technological parameter are $X(t) > 0$, a constant component was added to $X_k(t)$.

For this reason, in performing the computational experiments, the useful signals $X(t)$ were formed in the form of a sum of harmonic oscillations with different distribution laws. It was assumed that a useful signal is a stationary ergodic process and $X(t)$ is one of its realizations.

First Type Experiments

The deterministic useful signal $X(t) = 40\sin(1.1t + 0.5) + 25\cos(0.5t) + 35\sin(1.2t + 1.5) + 55\cos(1.4t + 0.3) + 20\sin(2.5t + 0.7) - 80\cos(2.3t + 1.5) + 300$ is formed as a sum of harmonic oscillations.

The noise $\mathrm{E}(t)$ obeys the normal distribution law with the mathematical expectation $m_\varepsilon \approx 0.136$ and the mean square deviation $\sigma_\varepsilon \approx 18$.

Second Type Experiments

The random useful signal

$X(t) = 40\sin\left(2\pi\frac{(k0.2)^n}{T} + \varphi\right) + 100$ is simulated in the form of a perturbed harmonic discrete function with the initial phase φ that has a uniform probability distribution (or with a uniform probability density), where $k \in [0, K]$, $K = 2400$, exponent $n = 1.5$; signal period $T = 600$; the initial phase φ is given in the form rand(size(k))pi/3 [18], where the function rand (size(k)) forms a vector commensurate with the vector k whose elements are random variables distributed according to the uniform law in the interval (0, 1).

The noise $\mathrm{E}(t)$ obeys the normal distribution law with the mathematical expectation $m_\varepsilon \approx 0.3335$ and the mean square deviation $\sigma_\varepsilon \approx 25$.

The **third type experiments** and the **fourth type experiments** demonstrate the suitability of the proposed method for a wider class of stochastic processes.

Third Type Experiments

The random useful signal

$X(t) = 20\cos\left(2\pi\frac{(k0.5)^{n1}}{T} + \varphi_1\right) + 25\sin\left(2\pi\frac{(k1.5)^{n2}}{T} + \varphi_2\right) + 100$ is simulated in the form of a perturbed harmonic discrete function with the amplitude and the initial phases φ_1, φ_2, that have a uniform probability distribution (or with a uniform probability density), where $k \in [0, K]$, $K = 2400$, exponents $n = 1.5$, $n = 0.5$; signal period $T = 800$; the amplitudes are given in the form rand(size(k)), rand(size(k)); the initial phases φ_1, φ_2 are given in the form rand(size(k))pi/3, rand(size(k))pi/3 [18].

The noise $\mathrm{E}(t)$ obeys the normal distribution law with the mathematical expectation $m_\varepsilon \approx 0.3335$ and the mean square deviation $\sigma_\varepsilon \approx 25$.

Table 1. Characteristics of the noise

	Characteristics	Given value	Calculated value	Relative error, %
1	Value of Pearson's chi-squared test	230.75 (tabulated)	6687.5	6687.5 > 230.75
2	Second-order moment of the noise	627.44	617.02	1.66
3	Mean square deviation	25.049	24.84	0.84047
4	Mathematical expectation	0	0.33345	33,345
5	Max of the distribution density function	0.015958	0.016059	0.6354
6	First inflection point along the abscissa axis	−24.667	−24.84	0.69793
7	Second inflection point along the abscissa axis	25.333	24.84	1.9869
8	Inflection point along the ordinate axis	0.0096788	0.0097412	0.64034

Fourth Type Experiments

The random useful signal

$X(t) = 15\cos\left(2\pi\frac{(k0.5)^{n1}}{T} + \varphi_1\right) + 25\sin\left(2\pi\frac{(k0.3)^{n2}}{T} + \varphi_2\right) + 100$ is simulated in the form of a perturbed harmonic discrete function with the amplitudes and the initial phases φ_1, φ_2 that have a uniform probability distribution (or with a uniform probability density), where $k \in [0, K]$, $K = 2400$, exponents $n1 = 1.5$, $n2 = 2.7$; signal periods $T = 10000$; the initial phases φ_1, φ_2 are given in the form rand(size(k))pi/3, rand(size(k))pi/5. The amplitudes obey the normal distribution law with the mathematical expectations m1 = 0; m2 = 0 and the mean square deviations s1 = 0.4; s2 = 0.5 and are given in the form normrnd(m1, s1, 1, K), normrnd(m2, s2, 1, K), where the function normrnd() allows obtaining the matrix of allows to obtain a matrix of pseudo-random numbers with dimension of 1K elements distributed according to the normal law for parameters m1, m2 (mathematical expectations) and s1, s2 (mean square deviations) [18].

The noise $\mathrm{E}(t)$ obeys the normal distribution law with the mathematical expectation $m_\varepsilon \approx 0.3068$ and the mean square deviation $\sigma_\varepsilon \approx 23$.

The results of the calculations for *Experiment N2* are presented in Table 1. Similar results were obtained for *Experiment N1, N3, N4.*

8. COMPARATIVE ANALYSIS OF THE COMPUTATIONAL EXPERIMENTS

The following conclusions have been drawn after analyzing the obtained results.

(1) In all experiments, the calculated value of Pearson's chi-squared test χ^2 for testing if the normal distribution of the useful signal is greater than the tabulated value $\chi^2_{k;\alpha}$ with the number of degrees of freedom $k = n - r - 1 = 200 - 2 - 1 = 197$ and the significance level $\alpha = 0.95$. Therefore, useful signals do not obey the normal distribution law (Table 1, row 1). This means that the developed algorithms can be applied to a wide class of useful signals even when the classical conditions of the theory of stochastic processes are violated.

(2) In all experiments, the predetermined D_ε and calculated D_ε^* estimates of the variance and the mean square deviations of the noises practically match (Table 1, rows 2, 3): $D_\varepsilon \approx D_\varepsilon^*$, $\sqrt{D_\varepsilon} \approx \sqrt{D_\varepsilon^*}$, and the values of the relative errors ΔD_ε and $\sqrt{\Delta D_\varepsilon}$ are 1.66% and 0.84047%.

(3) In all experiments, the given estimate m_ε of the mathematical expectation of the noise and the estimate $m_\varepsilon^* = 0$ of the mathematical expectation of the noise accepted by the condition of the problem practically match (Table 1, row 4): $m_\varepsilon \approx m_\varepsilon^*$, and the magnitude of the relative error Δm_ε does not exceed 33.345%.

(4) In spite of the fact that the absolute value of the noise $\mathrm{E}(i\Delta t)$ is estimated at hundredths of unity in the range $-\sqrt{D_\varepsilon^*} \le \mathrm{E}(i\Delta t) \le \sqrt{D_\varepsilon^*}$ of the most probable values, the values of the relative errors $\Delta N(\varepsilon(i))$

Table 2. The normal distribution density function of the noise

i	$N(\varepsilon(i))$	$N^*(\varepsilon(i))$	$\Delta N(\varepsilon(i))$, %	i	$N(\varepsilon(i))$	$N^*(\varepsilon(i))$	$\Delta N(\varepsilon(i))$, %
1	0.00017727	0.00017528	1.126	91	0.015505	0.015547	0.27227
11	0.00048926	0.00048777	0.30553	101	0.013329	0.013274	0.40916
21	0.0011862	0.0011904	0.35341	111	0.010066	0.0099398	1.2516
31	0.0025263	0.0025477	0.84755	121	0.0066775	0.0065272	2.2508
41	0.0047263	0.0047818	1.1744	131	0.0038913	0.0037589	3.402
51	0.0077674	0.0078709	1.3324	141	0.001992	0.0018984	4.6994
61	0.011214	0.011362	1.3207	151	0.00089578	0.00084081	6.1368
71	0.014221	0.014383	1.1394	161	0.00035386	0.00032658	7.7074
81	0.015843	0.015968	0.7893				

vary only in the range of 0.27–1.25% (Table 2, rows 71–111). In the intervals $-2\sqrt{D_\varepsilon^*} \le \mathrm{E}(i\Delta t) \le -\sqrt{D_\varepsilon^*}$ and $\sqrt{D_\varepsilon^*} \le \mathrm{E}(i\Delta t) \le 2\sqrt{D_\varepsilon^*}$ of frequently encountered values of the noise $\mathrm{E}(i\Delta t)$, the absolute values of errors amount to thousandths of unity, and the relative errors in these intervals slightly exceed the previous values and vary within the range of 0.84–3.4% (Table 2, rows 31–61, 121–131). In the intervals $-3\sqrt{D_\varepsilon^*} \le \mathrm{E}(i\Delta t) \le -2\sqrt{D_\varepsilon^*}$ and $2\sqrt{D_\varepsilon^*} \le \mathrm{E}(i\Delta t) \le 3\sqrt{D_\varepsilon^*}$ of rarely encountered values of the noise $\mathrm{E}(i\Delta t)$, when absolute values of the errors amount to thousandths of unity, the relative errors do not exceed 0.3–7.7% (Table 2, rows 1–21, 141–161). This indicates that in all experiments the given estimate $N(\varepsilon)$ and the calculated estimate $N^*(\varepsilon)$ of the normal distribution density function of the noise practically match (Table 2): $N(\varepsilon(i)) \approx N^*(\varepsilon(i))$.

(5) In all experiments, the given and the calculated maximums of the normal distribution density function of the noise practically match (Table 1, row 5): $N_{\max}(m_\varepsilon) \approx N^*_{\max}(m_\varepsilon)$, and the value of the relative error $\Delta N_{\max}(m_\varepsilon)$ does not exceed 0.6354%.

(6) In all experiments, the given and the calculated inflection points of the normal distribution density function of the noise practically match (Table 1, rows 6–8): $\left(m_\varepsilon - \sqrt{D_\varepsilon}\right) \approx \left(m_\varepsilon^* - \sqrt{D_\varepsilon^*}\right)$, $\left(m_\varepsilon + \sqrt{D_\varepsilon}\right) \approx \left(m_\varepsilon^* + \sqrt{D_\varepsilon^*}\right)$, $\left(\frac{1}{\sqrt{2D_\varepsilon \pi e}}\right) \approx \left(\frac{1}{\sqrt{2D_\varepsilon^* \pi e}}\right)$, and the values of the relative error $\Delta a1$, $\Delta a2$ and Δo do not exceed 0.6973–1.9869%.

Thus, the computational experiments demonstrate that the distribution density function, its maximum value and the inflection point of the given noise and the distribution density function, its maximum value and inflection points calculated using the developed technology, practically match.

9. TECHNOLOGY FOR DETECTING INCIPIENT DEFECTS WITH THE USE OF THE DENSITY FUNCTION OF THE NORMALLY DISTRIBUTED NOISE

Let us consider one of the possible variants of solving the problem of indicating the onset of the defect generation process using the estimates of the variance and the mean square deviation of the noise calculated from expressions (8)–(15), the distribution density function, its maximum and inflection points. It will be shown in the following paragraphs that these estimates can be used as reliable indicators of the occurrence of faults.

Assume that in the normal state of the facility, before the start of the defect generation process, the useful signal does not contain a noise. Then the following equalities hold:

$$D_\varepsilon^* = 0,\quad \sigma_\varepsilon^* = 0,\quad N^*(\varepsilon) = 0,\quad N^*_{\max}(m_\varepsilon) = 0.$$

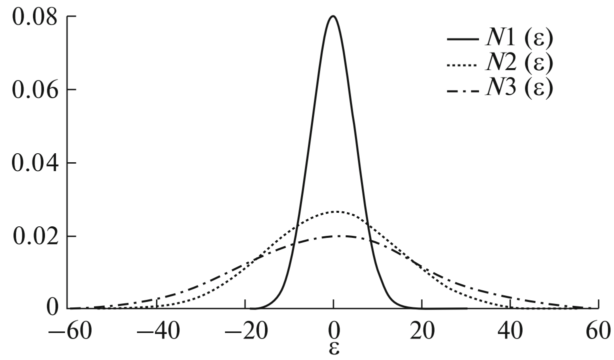

Fig. 1. The distribution density function of the noise at different states of the facility.

When the latent period of the defect inception begins and the facility goes into an emergency state, these conditions are violated, i.e.:

$$D_\varepsilon^* \neq 0, \quad \sigma_\varepsilon^* \neq 0, \quad N^*(\varepsilon) \neq 0.$$

Since the distribution density function $N^*(\varepsilon)$ requires that we specify a set of points, we can confine ourselves to the maximum value $N_{\max}^*(m_\varepsilon)$ and the inflection points.

For instance, let us consider an acoustic signal that comes from the output of an acoustic sensor installed on a gas pipeline (or oil pipeline). In the normal state of the gas pipeline (or oil pipeline), a useful signal appears in the output of the sensor in the form of a "hum" that does not contain high-frequency noise. When a microcrack appears, the "hum" turns into "whistling" under the influence of the high-frequency component that appears, which a noise is. The variance, the mean squared deviation, the maximum and the inflection points of the distribution density function of that noise have a certain value, e.g. $D_{\varepsilon 1}^*, \sigma_{\varepsilon 1}^*, N1_{\max}^*(m_\varepsilon)$.

As the microcrack grows, the "whistling" also increases and becomes "snoring." Then the characteristics of the noise change, taking on the values: $D_{\varepsilon 2}^*, \sigma_{\varepsilon 2}^*, N2_{\max}^*(m_\varepsilon)$.

Then the "snoring" turns into "gurgling," and the characteristics take on the values: $D_{\varepsilon 3}^*, \sigma_{\varepsilon 3}^*, N3_{\max}^*(m_\varepsilon)$.

Thus, depending on the fault degree, the mean square deviation, maximum and inflection points of the distribution density function of the noise change. This indicates that these characteristics determine the nature of the fault. Over time, unless proper repair work is performed, the distribution law itself changes, which already indicates an emergency state. Figure 1 shows the graphs of the distribution density functions of the noise for all three cases.

Thus, the estimates of the variance, mean square deviation, distribution density function, its maximum and inflection points can be used to indicate the onset of microchanges in the technical condition of the control object.

10. CONCLUSION

The technologies proposed in this paper have wide practical application and allow systems of monitoring, control, diagnostics, forecasting, management, identification, etc. to identify not only nascent changes, but also the moments when preventive maintenance, current and major repairs must be carried out [5, 6, 19]. This is because knowing the distribution density function of the noise, its maximum value and inflection points, as well as the dynamics of the curve shape change, we can make conclusions on the nature of the variation of the noise. Thus, with the use of the said characteristics, it is possible to obtain sufficiently exhaustive information about the state of the system, facility or device under study.

REFERENCES

1. Ott, H.W., *Noise Reduction Techniques in Electronic Systems,* John Wiley & Sons Inc, 1976.
2. Shiryaev, A.N., *Veroyatnostno-statisticheskie metody v teorii prinyatiya reshenii* (Probabilistic-Statistical Methods in Decision Theory), Moscow: MCCME, 2011.

3. Vishnyakov, A.N. and Tsypkin, Ya.Z., Detection of regularity violations by means of observed data in the presence of errors, *Autom. Remote Control,* 1991, vol. 52, no. 12, pp. 1744–1751.
4. Kharkevich, A.A., *Bor'ba s pomekhami* (Struggle with Noise), Moscow, 1965.
5. Tikhonov, V.I., *Statisticheskaya radiotekhnika* (Statistical Radiotechnics), Moscow: Radio i Svyaz', 1982.
6. Aliev, T., *Digital Noise Monitoring of Defect Origin,* London: Springer-Verlag, 2007.
7. Aliev, T.A., Guluyev, G.A., Pashayev, F.H., and Sadygov, A.B., Noise monitoring technology for objects in transition to the emergency state, *Mech. Syst. Signal Process.,* 2012, vol. 27, pp. 755–762.
8. Musaeva, N.F., Robust correlation coefficients as initial data for solving a problem of confluent analysis, *Autom. Control Comput. Sci.,* 2007, no. 2, pp. 76–87.
9. Aliyev, T.A. and Musaeva, N.F., An algorithm for eliminating microerrors of noise in the solution of statistical dynamics problems, *Autom. Remote Control,* 1998, vol. 59, no. 5, pp. 679–688.
10. Musaeva, N.F., Robust method of estimation with "contaminated" coarse errors, *Autom. Control Comput. Sci.,* 2003, vol. 37, no. 6, pp. 50–63.
11. Musaeva, N.F., Technology for determining the magnitude of robustness as an estimate of statistical characteristic of noisy signal, *Autom. Control Comput. Sci.,* 2005, vol. 39, no. 5, pp. 53–62.
12. Dubov, I.R., Formation of direct observations and approximation of probability density for rounding experimental data, *Autom. Remote Control,* 2000, vol. 61, no. 3, pp. 438–449.
13. Kolnogorov, A.V., On rational control of the mean level of random noise, *Autom. Remote Control,* 2000, vol. 61, no. 1, pp. 65–74.
14. Burkatovskaya, Yu.B. and Vorobeichikov, S.E., Detection of change-points in a noisy autoregression process, *Autom. Remote Control,* 2000, vol. 61, no. 3, pp. 425–437.
15. Markov, A.S., Estimation of the autoregression parameter with infinite dispersion of noise, *Autom. Remote Control,* 2009, vol. 70, no. 1, pp. 92–106.
16. Ventsel, E.S., *Teoriya veroyatnosti* (Probability Theory), Moscow: Nauka, 1969.
17. Solodovnikov, V.V., Matveyev, P.S., Valdenberg, Y.S., and Baburin, V., *Vychislitel'naya tekhnika v primenenii k statisticheskim issledovaniyam v avtomatike* (Computing Technology in Application for Statistical Research and Calculations of Automatic Control Systems), Moscow: Mashinostroenie, 1963.
18. Korolenko, P.V. and Ryzhikova, Y.V., *Modelirovanie i obrabotka stokhasticheskikh signalov i struktur* (Modeling and Processing of Random Signals and Structures), Moscow, 2012. http://optics.sinp.msu.ru.
19. Abbasov, A.M., Mammadova, M.H., Orujov, G.H., and Aliyev, H.B., Synthesis of the methods of subjective knowledge representations in problems of fuzzy pattern recognition, *Mechatronics,* 2001, no. 11, pp. 439–449.