



RESEARCH OF DIAGNOSTIC PARAMETERS OF COMPOSITE MATERIALS USING JOHNSON DISTRIBUTION

Vitaliy Babak ¹⁾, Volodymyr Eremenko ²⁾, Artur Zaporozhets ³⁾

¹⁾ Institute of Engineering Thermophysics of NAS of Ukraine, Ukraine, vdoe@ukr.net

²⁾ National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, nau_307@ukr.net

³⁾ Institute of Engineering Thermophysics of NAS of Ukraine, Ukraine, a.o.zaporozhets@nas.gov.ua

Paper history:

Received 05 February 2019

Received in revised form 26 October 2019

Accepted 29 November 2019

Available online 31 December 2019

Keywords:

Johnson distribution;
Gaussian distribution;
uniform distribution;
uncertainty;
approximation;
quantiles;
statistical diagnostics.

Abstract: In this paper, it was proposed to carry out a preliminary normalization of diagnostic parameters using the Johnson distribution, which with three basic distribution groups (S_L , S_B , S_U), covers a wide class of empirical distributions. The mathematical description of the family allows us to find the approximating probability density function in an explicit form, to determine the distribution parameters for obtaining the corresponding function (curve), as well as the inverse function for finding the quantiles of the specified levels. To assess the accuracy of the obtained normalized data, they were compared with the data obtained by replacing the resulting law with a Gaussian one. Percentages of values were compared in the implementation under study, which concentrated in the limits of estimated quantiles. Implementations were obtained using the simulation method. By the same method, the correctness (relative systematic error) of determining the quantile values of the specified levels was evaluated. The error value δ was estimated between the conditionally true quantile value calculated from the generated pseudo-general complex and the value estimated using the methods considered in the paper. Obtained data show that the relative error in the calculation of quantiles using the Johnson distribution does not exceed 0.07% and decreases in two orders of magnitude than the currently accepted procedure for replacing sample laws with Gaussian.

Copyright © Research Institute for Intelligent Computer Systems, 2019.

All rights reserved.

1. INTRODUCTION

Products made of composite materials, in contrast to products made of metals, are formed from primary raw materials simultaneously with the formation of the materials [1, 2, 3]. Due to the complexity of their manufacturing technology, it becomes impossible to build a priori models describing the definitions of informative parameters of controlled objects [4, 5], and ignorance of the laws of probability distribution of changes does not allow to form the corresponding decision rule [6].

Among the existing methods for selecting informative parameters, not everything can be used to describe a large number of information signals and diagnostic parameters of composite materials selected on their basis [7, 8, 9]. Most of the existing criteria are based on the use of the normal law of distribution of the studied data [10]. But this creates

certain difficulties [11, 12]. First, it is quite difficult to verify the laws of the distribution of the entire set of diagnostic parameters, and, therefore, it cannot be argued that they are all normal [13]. Secondly, a change in the mechanical properties of the studied zones leads to a change in the distribution laws of the studied informative parameters [14, 15]. Therefore, it is necessary to apply criteria that are not sensitive to changes in the distribution laws [16]. The quality of selection of diagnostic parameters on the selection criteria can only be determined after the construction of a decision rule and the assessment of the reliability of control over the combination of the selection method and the method for constructing a decision rule [17, 18, 19].

Normalizing transformations are highlighted in [7], however, all transformations are associated with the problem of determining the general transformation for all classes (in the case of

parametric methods, determining the general transformation parameters), after which the distributions of all classes approach the normal one [20, 21].

There are parametric and non-parametric distribution methods [16]. In [1], it was shown that on small sample volumes, parametric methods give more accurate results. They are divided into the following groups:

- 1) power and logarithmic transformations from the Tukey series using normalization and shift [22];
- 2) transformations that use a power series decomposition (Cornish-Fisher transformation) [23];
- 3) Blom's functional transformation followed by the Cornish-Fisher transformation [24];
- 4) transformations from the Box-Cox-Tukey family (generalization of power and logarithmic transformations);
- 5) transformations based on the use of a priori information about the distribution of a random variable (approximation by χ^2 -distribution, the Fisher distribution, etc., lead to a normal distribution with transformation) [25];

6) Johnson transform (approximation from a family of distributions followed by transformation into a Gaussian distribution) [26, 27].

The first five transformations do not allow us to fully cover all classes of possible distributions of diagnostic features, since they are mainly focused on the group of distributions, special cases of which are Gaussian [28, 29]. The use of Pearson and Fisher type distributions for data normalization is possible only for cases when an appropriate distribution is established with a sufficient level of significance, which is possible only with large amounts of empirical data and the application of additional statistical procedures. Therefore, the Johnson transform is used for a wider class of distributions [30].

The article describes the approach for reasonable choice of the "rejection level" (threshold value) in non-destructive testing and diagnostics using the approximation of the unknown distribution of the informative parameter by the Johnson distribution. Based on experimental data, estimates of the corresponding statistical characteristics of the parameter are calculated (for example, the first four moments of the distribution), according to which the corresponding Johnson distribution is adjusted. Then, according to the quantiles calculated by the fitted distribution, the "rejection level" is determined taking into account the errors (risks) of the first and second kind.

2. JOHNSON NORMALIZATION TRANSFORM

Among possible normalizing transformations, the Johnson family of distributions, with three groups of distributions, covers a wide class of empirical distributions. A sufficient mathematical description of the family allows one to find the approximating probability density function (PDF) in an explicit form, the distribution parameters for obtaining the equation of the corresponding curve, and also the inverse function for finding quantiles [31, 32]. This makes it possible both to normalize diagnostic parameters for their further statistical processing, and to apply PDF approximations to select and substantiate threshold levels and calculate errors of the first and second kind that determine the probability of diagnostics. Therefore, the work proposed and investigated the use of the Johnson transform to normalize diagnostic parameters.

In general, the Johnson normalization transform is

$$z = \gamma + \eta \tau(x, \xi, \lambda); \quad \eta > 0; \quad -\infty < \gamma < \infty; \quad (1)$$

$$\lambda > 0; \quad -\infty < \xi < \infty$$

where $\tau(x, \xi, \lambda)$ is one of the three special functions; $\gamma, \eta, \xi, \lambda$ are distribution parameters; z is the normalized random variable distributed according to the Gaussian law.

The Johnson distribution system is described by the following three transformation equations:

– Johnson family of distributions S_L (log-normal distribution with three parameters)

$$f_1(x) = \frac{\eta}{\sqrt{2\pi}(x-\xi)} \times \exp\left\{-\frac{1}{2}\eta^2\left[\frac{\gamma^*}{\eta} + \ln(x-\xi)\right]^2\right\}; \quad (2)$$

$$\gamma^* = \gamma - \eta \ln \lambda; \quad x \geq \xi; \quad \eta > 0; \quad -\infty < \xi < \infty;$$

$$\lambda > 0; \quad -\infty < \xi < \infty,$$

– Johnson family of distributions S_B

$$f_2(x) = \frac{\eta}{\sqrt{2\pi}} \cdot \frac{\lambda}{(x-\xi)(\lambda-x+\xi)} \times \exp\left\{-\frac{1}{2}\left[\gamma + \eta \ln\left(\frac{x-\xi}{\lambda-x+\xi}\right)\right]^2\right\}; \quad (3)$$

$$\xi \leq x \leq \xi + \lambda; \quad \eta > 0; \quad -\infty < \gamma < \infty; \quad \lambda > 0;$$

$$-\infty < \xi < \infty,$$

– Johnson family of distributions S_U

$$f_3(x) = \frac{\eta}{\sqrt{2\pi}} \cdot \frac{1}{(x-\xi)^2 + \lambda^2} \cdot \exp\left(-\frac{1}{2}\gamma\right) \times \exp\left[\eta \ln\left\{\left(\frac{x-\xi}{\lambda}\right) + \left[\left(\frac{x-\xi}{\lambda}\right)^2 + 1\right]^{1/2}\right\}^2\right]. \quad (4)$$

For families of distributions S_B and S_U , the parameters γ, η are responsible for the shape of the distribution, ξ is the parameter characterizing the distribution center, and λ is the scale parameter.

The equations for estimating the quantiles of the desired level of significance based on the inverse of the normalizing transformation of the function for each type of Johnson distribution are given in Table 1.

Table 1. Equations for calculating α -quantile

Johnson distribution type	Quantile estimates x_α
S_L	$x(\alpha) = \exp(\sigma^2 \cdot z(\alpha) + m)$
S_B	$x(\alpha) = \frac{\xi + (\xi + \lambda) \cdot \exp\left(\frac{z(\alpha) - \gamma}{\eta}\right)}{1 + \exp\left(\frac{z(\alpha) - \gamma}{\eta}\right)}$
S_U	$x(\alpha) = \frac{\lambda \cdot \exp\left(\frac{z(\alpha) - \gamma}{\eta}\right) - 1}{2 \exp\left(\frac{z(\alpha) - \gamma}{\eta}\right)} + \xi$

To find estimates of the parameters of the normalizing Johnson transform, two methods are used – the quantile method and the method of moments.

2.1 QUANTILE METHOD

This method is based on comparing estimates of quantiles of an empirical distribution of $x(\alpha)$ with the values of the corresponding quantiles of Gaussian distributions $z(\alpha)$. Estimating the parameters of the Johnson distribution from empirical data to obtain the corresponding PDF is covered in [33]. Summary information on the estimation of parameters for each type of distribution is given in [34].

2.2 METHOD OF MOMENTS

This method is based on an estimate of the sample moments and uses the functional relationship between the moments and the parameters of the Johnson distribution.

The Johnson distribution system occupies large areas in the plane of moments, which makes it

possible to describe various laws of the distribution of the studied data. The article [35] shows the possible values of the coefficients of asymmetry and excess for the corresponding types of Johnson distribution.

The essence of the estimation of the distribution parameters by the method of moments consists in equating the expressions of the distribution moments as a function of their parameters to the sample values of the corresponding moments. The solution of the system of equations thus obtained will be the moment estimates of the desired values of the distribution parameters. The method of moments uses the following numerical characteristics of a random variable in magnitude PDF: moment of the first order (expectation) m , moment of the second order (dispersion) σ^2 , asymmetry sk (expressed through the moment of the third order), excess ex (expressed through the moment of the fourth order).

Mathematically the essence of the method of moments is described by the system of equations (5), which is obtained on the basis of the definition of moments.

The solution of this system of equations is the desired values of the parameters $\gamma, \eta, \xi, \lambda$. Numerical methods for solving the system involve the use of iterative solution methods, are rather difficult to implement in practice and provide an individual approach depending on different initial conditions [36].

To estimate the characteristics of the accuracy of the normalization method using the Johnson transform, a symmetric type distribution S_B^S is investigated, which allows us to approximate a wide class of truncated symmetric distributions of diagnostic parameters.

$$\begin{cases} \int_{\xi}^{\xi+\lambda} x \cdot f(x; \xi, \lambda, \eta, \gamma) dx = m, \\ \int_{\xi}^{\xi+\lambda} (x - m_x)^2 \cdot f(x; \xi, \lambda, \eta, \gamma) dx = \sigma^2, \\ \int_{\xi}^{\xi+\lambda} (x - m_x)^3 \cdot f(x; \xi, \lambda, \eta, \gamma) dx = sk, \\ \int_{\xi}^{\xi+\lambda} (x - m_x)^4 \cdot f(x; \xi, \lambda, \eta, \gamma) dx = ex. \end{cases} \quad (5)$$

In [37], functional dependencies between the parameters of distributions and sampling moments are presented that implement the solutions of the reduced system with the indicated restrictions based on them. A method is developed for estimating the

parameters of a symmetric truncated Johnson type S_B^S distribution using the method of moments:

1) according to the sample values of the indicators of asymmetry and excess the shape parameters γ and η are determined: for a symmetric distribution $\gamma=0$, the functional dependence $\eta(\beta_2)$ was obtained:

$$\eta(\beta_2) \approx \begin{cases} \frac{(0,62\beta_2 - 0,408)}{(3 - \beta_2)^{0,749}} & \text{if } \beta_2 \geq 1,8, \\ 0,8(\beta_2 - 1) & \text{if } \beta_2 < 1,8. \end{cases} \quad (6)$$

2) using the formulas [38], we calculate the corresponding values of the expected value My and the dispersion Dy as a function of γ and η :

$$My = \mu_1(\gamma, \eta) = \exp\left(\frac{1}{2\eta^2} - \frac{\gamma}{\eta}\right) \{1, k\},$$

$$Dy = \mu_2 = \exp\left(\frac{2}{\eta^2} - \frac{2\gamma}{\eta}\right) \times \left(\{2, k\} - \omega^{-1} \{1, k\}^2\right). \quad (7)$$

3) for sample values of mean $My = \bar{x}$ and dispersion $Dy = \sigma^2$, the parameters of the shift ξ and scale λ are determined:

$$\lambda = \sigma / \sqrt{Dy}, \quad \xi = \bar{x} - \lambda / 2, \quad (8)$$

4) according to the parameter estimates, the required quantile of the given level is calculated:

$$x(\alpha) = \frac{\xi + (\xi + \lambda) \cdot \exp\left(\frac{z(\alpha) - \gamma}{\eta}\right)}{1 + \exp\left(\frac{z(\alpha) - \gamma}{\eta}\right)}. \quad (9)$$

Thus, the obtained equations make it possible to calculate the value of a quantile of a given level for an arbitrary distribution that corresponds to the above conditions.

3. MATHEMATICAL PROCESSING OF EXPERIMENTAL RESULTS

3.1 RESEARCH OF THE ACCURACY OF ESTIMATING PARAMETERS BY THE QUANTILE METHOD

The choice of specific percentiles according to the method [39] is arbitrary. However, if using

different percentiles for the same initial data, this will lead to a change in the parameter estimates. According to the theoretical justification [40], in order to obtain a satisfactory approximation in the region of large deviations (at the ends of the distribution), the percentiles lying in this region should be chosen. Using too extreme percentiles can lead to a loss of overall accuracy in the selection of an approximating distribution due to the large variability of their estimates [41].

In Figs. 1, 2 estimates of the empirical distribution laws and their approximations by the Johnson distribution are given depending on the selected quantile levels:

(a) percentiles of the level of 0.025 0.05 0.95 and 0.975 (approximation at the ends of the distribution);

(b) percentiles of the level of 0.025 0.15 0.85 and 0.975 (approximation in the vicinity of the level of 5% and 95%);

(c) percentiles of level 0.3 0.4 0.6 0.6 and 0.7 (approximation of the middle of the distribution).

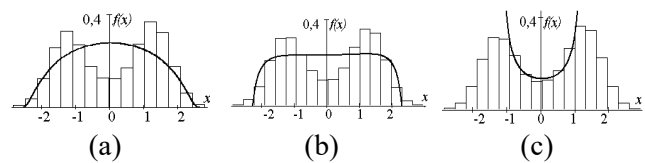


Figure 1 – Approximation of the sum of the arc sinusoidal, triangular, and uniform distribution of PDFs by the Johnson type distribution

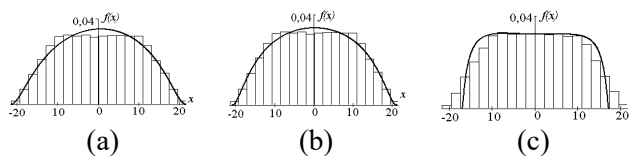


Figure 2 – Approximation of the sum of three uniform distribution laws (one uniform dominant) of PDF by the distribution of the S_B Johnson type

In Figs. 1, 2 ?? the choice of the values of empirical quantiles, for which the approximation is carried out, is shown, the values of distribution parameters are determined that affects the form of the calculated PDF. The choice of quantile level of 0.025, 0.05, 0.95 and 0.975 provides the best approximation at the ends of the distribution, the choice of quantile level of 0.025, 0.15, 0.85 and 0.975 provides approximation between the quantile level of 0.2–0.4 and 0.6–0.8. The best approximation of the middle of the distribution is provided by the choice of empirical quantiles of level 0.3, 0.4, 0.6 and 0.7.

Approximation of PDF from Johnson distributions should be used when it is necessary to achieve an accuracy of approximation on a certain set of values. Approximation should be used with

caution when it comes to finding the best approximation function in terms of the minimum mean-square error and additionally used statistical agreement criteria. In this case, the best results are obtained using smoothing Pearson curves.

To assess the accuracy of finding quantiles depending on the Percentile partition, conditionally true quantiles of the required level of significance were calculated in aggregate, approaching the general (volume $N = 10^6$ values). Quantile values were estimated with a uniform split of percentile (level 0.1, 0.4, 0.6 and 0.9), as well as for the case when they are concentrated in the ambit of the quantiles sought, for example, to find the expanded uncertainty at $P=0.95$ (quantile's levels 0.025 and 0.975) during solving a system of four equations, percentiles of the level of 0.025, 0.05, 0.95 and 0.975 were selected. In the study of the accuracy of quantile estimation, the considered method [42] was repeated $N=10000$ times, which made it possible to estimate the variation in the estimates of the quantiles sought. The results of the study are given in Table. 2.

It can be seen in the Table 2 that a uniform splitting of quantiles leads to a systematic error in the estimate. The displacement relative to the conventionally true value is the greater than the more closely sought quantiles approach the ends of the distribution. At the same time, the choice of quantiles in the ambit does not practically contribute to a systematic component. This is explained by the fact that the choice of quantiles determines the most accurate approximation. The standard deviation (SD) of quantile estimates for different partitions is not significantly different, since the studies were conducted on the same volumes of data.

3.2 RESEARCH OF THE ACCURACY OF ESTIMATING PARAMETERS BY METHOD OF MOMENTS

Estimation of the parameters of the Johnson's S_B -distribution by the method of moments was carried out according to the method described in [40]. However, they were derived under certain assumptions, so the question of investigating the accuracy of estimating the parameters of the Johnson distribution using these formulas arises.

The calculation of the moments for symmetric distributions according to the available information in accordance with the distribution law is made according to the formulas given in article [1].

For an experimental study, data was generated for the general population with the Johnson's S_B -distribution.

The data was generated with the transformation (Table 3) which makes it possible to obtain a

random variable y_{S_B} with the Johnson's S_B -distribution and arbitrarily specified distribution parameters using the normalized Gaussian distribution law z_N .

The accuracy of quantile estimation was investigated by approximation of the Johnson distribution, constructed by the method of moments compared to the accuracy of quantile estimation by replacing PDF with the Gaussian distribution as the most common method.

Fig. 3 shows the histograms of the distribution of quantile level estimates of 0.025 and 0.05, respectively, for the case of the sum of Gaussian, triangular and uniform distribution laws, $f1(x)$, $f3(x)$ are the distributions of estimates obtained by replacing the resulting law with Gaussian, $f2(x)$, $f4(x)$ are the distributions of estimates obtained using the normalization procedure.

The dotted line denotes conditionally true quantile values of the corresponding level, calculated as the quantile of the corresponding significance level.

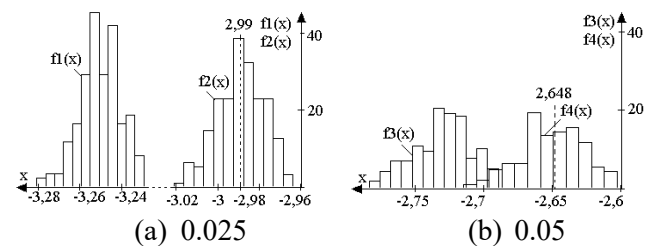


Figure 3 – Distributions of quantile estimates of different levels

Analysis of Fig. 3 shows that the variation of the calculated estimates depends on the level of quantile, but does not depend on the method of assessment. The distribution of the quantile estimate (average), calculated by replacing the resulting Gaussian distribution, has a systematic error, thereby overestimating the real limits of the diagnostic attribute values, and the average value of the quantile estimate, calculated using the normalization procedure, coincides with the conventionally true value.

From above results, it can be concluded that the convergence of the obtained quantile estimates is the same (SD levels), but the correctness of the estimate is different. It is much better for the method based on the procedure of approximation by the Johnson distribution.

Another approach to comparing the procedure for normalizing and replacing the resultant Gaussian law is to compare the percentage of values in the resulting implementation which locates within the estimated quantiles. Using the calculated values of

quantiles, the actual percentage of values was calculated between them, it was compared with real values: $S_{95}=95\%$, $S_{90}=90\%$ (between the quantile level of 0.025 and 0.975 should be 95% of all values, between the quantile of 0.05 and 0.95 – 90%). Relative error was also considered.

As a percentage of the values between the estimated quintiles, it was taken an average of N values for calculating the percentage of values. The results are presented in Table 4.

Table 4 shows that the value of the relative error using approximations by Johnson distribution is much less than with the replacement by the Gaussian law. The relative error in calculating quantiles, as well as the percentage of real values that are within 90%-95% does not exceed 0.01%, which suggests that the proposed method for estimating quantiles of the distribution of diagnostic parameters is possible, because its accuracy increases in two orders of magnitude.

3.3 RESEARCH OF THE ACCURACY OF THE METHOD DEPENDING ON THE VOLUME OF THE STUDIED DATA

The quantile data normalization method is based on the use of percentiles, which are obtained for an ordered sample. In the case when the volume of the studied data is 100 values, the index of each value in the ordered sample corresponds to one percentile. If the sample size exceeds 100 values, one or several values in the sample will correspond to the same percentile level, which will not distort the result. However, in the case when the volume of the studied data is less than 100 values, the same percentile corresponds to several quantiles, which will introduce significant uncertainty in the calculations of the parameters of the approximating distribution.

The study was conducted by statistical Monte-Carlo simulation with different sample sizes, ranging from 20 to 500 in 10-step increments for quantiles starting from the level of 0.025 to 0.975 with a step of changing 0.025. The obtained plane of the calculated values of standard errors is shown in Fig. 4. Fig. 5 shows the dependence of standard errors of estimation of moments on the amount of sample values.

From the data obtained, it can be concluded that the standard errors of the estimates of the extreme quantiles of the level of 0.025 and 0.975 take values of 0.6 for small data volumes, decrease when approaching the distribution center (the standard error is 0.33 for the quantile of level 0.5) and approach to 0 with increasing data.

The error value increases for each next moment and asymptotically approaches to 0 with increasing volume. The standard errors of determining the first

three moments do not exceed the standard errors of quantiles (0.6), but the error of the 4th moment increases in 3 times.

The study of the correctness of quantile evaluation was carried out by the method of simulation. The relative error modulus between the conditionally valid quantile value (calculated from the general subset) and the quantiles estimated by the proposed method for a sample size from 20 to 100 values was calculated. The experiment was conducted $N = 10000$ times, the value of the relative error for each sample size was calculated as the average value. The results obtained for different volumes (from 20 to 100 values) differed slightly, therefore, in Table 5 the average value of the relative error is shown.

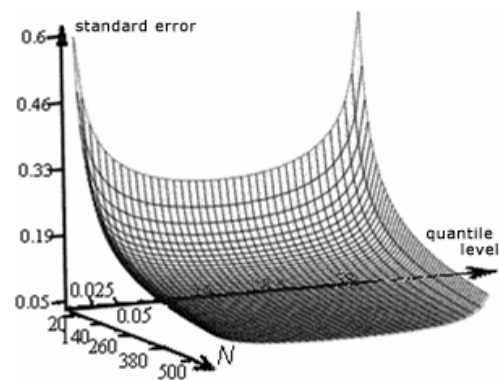


Figure 4 – Standard errors of quantiles

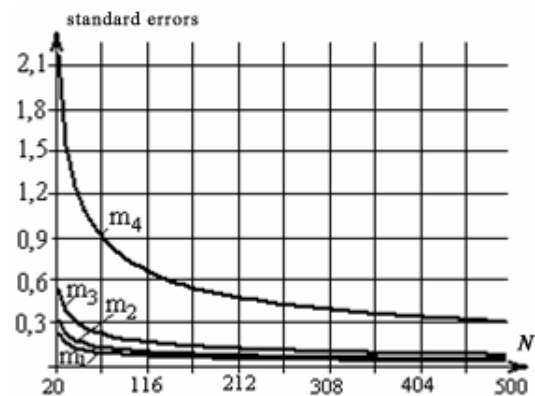


Figure 5 – Standard errors of moments

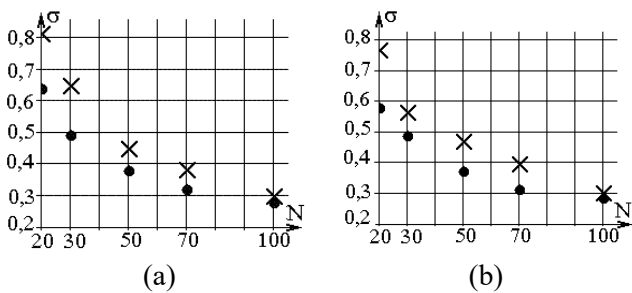
Since the experimental study of the method has been carried out on samples of limited volume, the characteristic is convergence, which is calculated as SD for repeated tests and allow us to show how much the obtained estimates characterize the general population.

The dependences of the SD of quantile estimation in the construction of approximating distributions by the method of moments with a confidence probability of 0.9 and 0.95 of the sample size are shown in Fig. 6 (for (a): X is SD of quantile

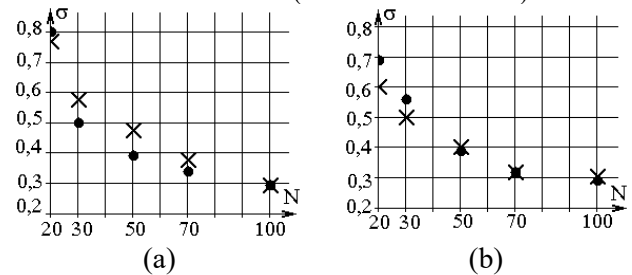
estimates 0.025, \cdot is SD of quantile estimates 0.05; for (b): X is SD of quantile estimates 0.975, \cdot is SD of quantile estimates 0.95).

The analysis of the obtained results allowed us to draw the following conclusions.

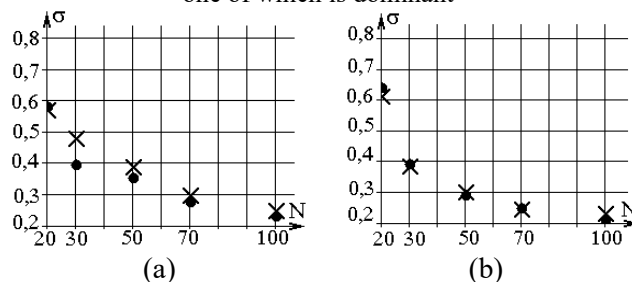
The method of estimating the parameters of the distribution of diagnostic parameters using the approximation of the Johnson distribution allows us to cover a wide class of distributions, including dual-modal distributions and does not require a priori information about the type of law, its truncation, has a relatively simple implementation, the only mathematical complexity of which is to use numerical methods for solving a system of nonlinear equations, which in turn have a wide representation in various software packages. The accuracy of the procedure of approximation and normalization depends on the accuracy of the calculation of distribution parameters, the proximity of the quantiles to empirical and Gaussian law quantiles during solving a system of equations and practically does not depend on the distribution law of the combined uncertainty and the level of the quantile.



(a) The combination of Gaussian, uniform and triangular distribution laws (uniform is dominant)



(a) The combination of three uniform laws of distribution, one of which is dominant



(a) Combination of one arcsinusoidal and two uniform distribution laws (arcsinusoidal is dominant)

Figure 6 – Dependence of SD of quantile estimates on the volume of data under study in the quantile method of parameter estimation

In cases of small sample sizes (from 20 to 100 values), the moment method of parameter estimation should be used, because it provides higher accuracy and provides satisfactory convergence. With large amounts of data (from 100 or more values), it's possible to use both the moment and quantile method, the accuracy of which significantly depends on the type of distribution law of diagnostic parameters.

The convergence of estimates of quantiles calculated by the method of approximation by Johnson distribution and the method of replacing the resulting law with Gaussian is not significantly different. However, the correctness of the proposed method is much higher: the relative error of the method of replacing the resulting law with Gaussian for the sum of Gaussian, uniform and triangular laws of distribution and the sum of three uniform laws of distribution takes values from 2.6% to 8.7% for a dual-modulated resultant law (the sum of the arcsinusoidal and two uniform laws) of the distribution of total uncertainty – from 8% to 23%. For the method using the procedure of normalization, the value of the relative error does not exceed 2% for critical case of dual mode.

3.4 RESEARCH OF THE ACCURACY OF THE APPROXIMATION

During the experimental study of the proposed method, situations were considered in which the distribution law of the regenerated samples had an importance of excess and asymmetry, consistent with the data obtained during the study of samples of composite panels [43].

An important task during carrying out a simulation based on the Monte Carlo method is the characteristics of a random number generator. In the course of the study, random samples were generated with given values of excess and asymmetry. The deviation of the obtained values from the values specified in the modeling process may affect the results of the study. Thus, during studying the characteristics of a random number generator, the following tasks were set:

1. Estimation of confidence intervals for the values of excess and asymmetry coefficients of the obtained samples.
2. Establishing the degree of repeatability of samples.

Table 6 shows the calculated values of the confidence intervals for the estimates of asymmetry and excess coefficients obtained in the study of a random number generator. In the process of modeling, samples were generated with a volume of 20000 with asymmetry and excess coefficients, which varied from -0.4 to +0.4 for a fixed value of

one of the parameters. The confidence probability for all cases was 0.95.

As it can be seen from the results, the width of the confidence interval for the asymmetry coefficient remains stable throughout the entire simulation range, while for the excess it tends to

expand in accordance with the shift of the planned value to positive values.

Fig. 7 shows the dependence of the change in the quantile estimate on the value of the asymmetry coefficient of the distribution law of the original sample.

Table 2. Correctness and convergence of quantile estimates depending on quantile partitioning

Quantile level	Quantile value			SD of quantile estimates with uniform location	SLE quantile estimates with located in the ambit of the desired quantiles
	Conditionally true value	Uniform location of percentiles	Location in the ambit of the sought-for quantiles		
The sum of arcsinusoidal, triangular and uniform distributions					
0.025	-2.039	-1.711	-2.039	0.009	0.007
0.05	-1.857	-1.689	-1.855	0.006	0.006
0.95	1.887	1.714	1.890	0.007	0.005
0.975	2.066	1.728	2.071	0.007	0.006
The sum of a Gaussian, triangular and uniform distributions					
0.025	-2.993	-2.792	-2.989	0.199	0.010
0.05	-2.648	-2.575	-2.648	0.011	0.008
0.95	2.648	2.575	2.648	0.010	0.011
0.975	2.988	2.791	2.989	0.199	0.008
The sum of three uniform distributions laws					
0.025	-16.828	-15.787	-16.826	0.066	0.041
0.05	-15.138	-14.658	-15.136	0.048	0.040
0.95	15.139	14.651	15.137	0.057	0.037
0.975	16.824	15.775	16.819	0.078	0.042

Table 3. Estimations of moments for different distribution laws

Distribution parameters	Moments			
	1 st	2 nd	3 rd	4 th
Gaussian				
m, σ^2	m	σ^2	0	3
uniform				
a, b	$(a+b)/2$	$(b-a)/\sqrt{12}$	0	$Ex \cdot \sigma^4$ ($ex=1.8$)
triangular				
a, b	$(a+b)/2$	$(b-a)/\sqrt{24}$	0	$Ex \cdot \sigma^4$ ($ex=2.4$)
arcsinusoidal				
m, σ^2	m	σ^2	0	$Ex \cdot \sigma^4$ ($ex=1.5$)

Table 4. The percentage of values in the sample located in the confidence interval with a given level of probability P

	$P, \%$	Normalization	Gaussian Replacement	$\delta_{norm}, \%$	$\delta_{gaus.rep}, \%$
	The sum of the Gaussian, uniform and triangular distributions laws				
$S, \%$	90	89.99	91.37	0.01	1.52
	95	95.01	97.37	0.01	2.49
The sum of three uniform distributions laws					
$S, \%$	90	90.01	91.20	0.01	1.33
	95	95.01	97.30	0.01	2.45
The sum of two uniform and arcsinusoidal distribution laws					
$S, \%$	90	89.95	93.96	0.01	4.40
	95	95.01	99.38	0.01	4.61

Table 5. Relative systematic error (in percent) of quantile estimates N = 20..100

The combination of the laws of distribution	Gaussian, uniform and triangular		Three uniform		Two uniform and arcsinusoidal	
	0,025&0,975	0,05&0,95	0,025&0,975	0,05&0,95	0,025&0,975	0,05&0,95
Method of Gaussian replacement	9.8	2.9	8.9	2.8	27.6	15.7
Method of moments	2.1	0.9	1.2	0.4	2.1	1.2

Table 6. Confidence intervals of sample values of excesses and asymmetries

Setpoints S_k and E_x	Asymmetry coefficient			Excess coefficient		
	Average S_k	Upper limit of interval	Lower bound of interval	Average E_x	Upper limit of interval	Lower bound of interval
-0.4	-0.40	-0.44	-0.36	-0.40	-0.49	-0.31
-0.2	-0.20	-0.23	-0.16	-0.20	-0.21	-0.08
0.0	-0.01	-0.04	0.03	-0.00	-0.14	0.14
0.2	0.20	0.17	0.24	0.20	0.03	0.38
0.4	0.40	0.36	0.44	0.40	0.19	0.62

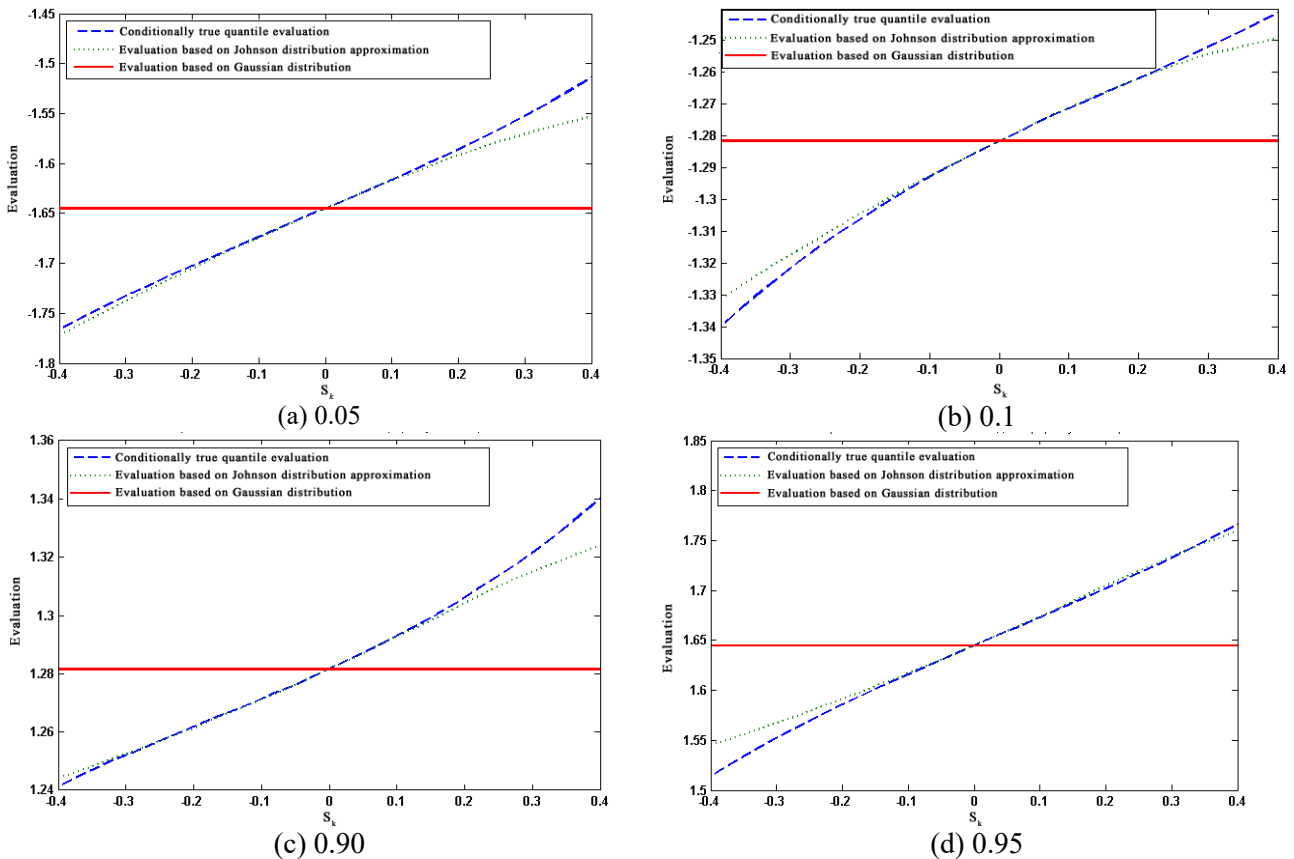


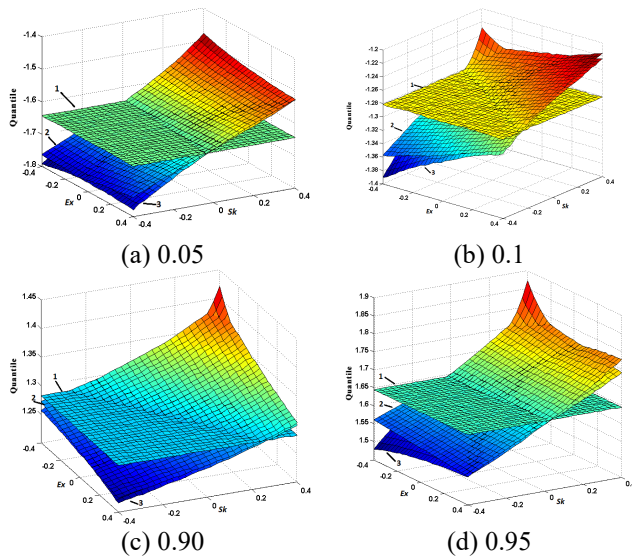
Figure 7 – Dependencies of the assessment of the value of quantile levels on the value of the asymmetry coefficient

As it can be seen in Fig. 7, the value of the quantile estimate obtained by replacing the empirical distribution law of the sample under study with Gaussian remains unchanged. An increase in the absolute value of the asymmetry coefficient leads to an increase in the error in determining quantiles as a result of a shift in the absolute value of the estimate. At extreme points, with maximum values of sample asymmetry coefficients, for quantiles of 0.05 level, the determination error with the replacement by the

Gaussian law can reach 20%, while the error of determination of the quantile with the Johnson distribution does not exceed 3%, which is essential in the process of forming the threshold regulations.

To test the effectiveness of using the Johnson distribution, as well as assessing the possible errors arising from replacing the distribution law of the original sample with Gaussian, it simulated a simultaneous change in the coefficients of asymmetry S_k and excess E_x . The procedure of the

model experiment was implemented similarly to the one that was performed during assessing the effect of a change in the asymmetry coefficient on the quantile estimate. The results of the study are shown in Fig. 8.



1 – estimation based on replacement by Gaussian law; 2 – estimation based on Johnson distribution approximation; 3 – conditionally true evaluation

Figure 8 – Dependencies of quantile estimates on asymmetry and excess coefficients

The above dependencies show that approximation using the Johnson distribution causes significantly less error than approximation by the Gaussian law, therefore, for large absolute values of excess and asymmetry of the laws of the distribution of diagnostic parameters, to increase the reliability of diagnosis, it is necessary to apply approximations based on the Johnson distribution.

4. CONCLUSION

A method has been developed to study the statistical characteristics of diagnostic parameters, taking into account the type of their distribution laws, which made it possible to develop a method for normalizing the probability distribution of diagnostic parameters using the Johnson transform, as well as obtaining the density equation for the probability distribution of the parameters, which significantly expands the scope of solving diagnostic problems and improves the accuracy of determining the threshold and reliability of the assessment of the state of the product.

The accuracy of determining the quantile of given levels of empirical laws of distributions using the Johnson transform approximation, which allowed us to justify the method of constructing approximations, is investigated.

5. REFERENCES

- [1] S. Gholizadeh, "A review of non-destructive testing methods of composite materials," *Procedia Structural Integrity*, vol. 1, pp. 50-57, 2016
- [2] B. Salski, W. Gwarek, P. Korpas, S. Reszewicz, A.Y.B. Chong, P. Theodorakeas, I. Hatzioannidis, V. Kappatos, C. Selcuk, T.-H. Gan, M. Kouli, M. Iwanowski, B. Zielinski, "Non-destructive testing of carbon-fibre-reinforced polymer materials with a radio-frequency inductive sensor," *Composite Structures*, vol. 122, pp. 104-112, 2015
- [3] J. Dong, B. Kim, A. Locquet, P. McKeon, N. Declercq, D.S. Citrin, "Nondestructive evaluation of forced delamination in glass fiber-reinforced composites by terahertz and ultrasonic waves," *Composites Part B: Engineering*, vol. 79, pp. 667-675, 2015
- [4] A. Zaporozhets, V. Eremenko, R. Serhienko, S. Ivanov, "Methods and hardware for diagnosing thermal power equipment based on Smart Grid Technology", in: N. Shakhovska, M.O. Medykovskyy (Eds.), *Advances in Intelligent Systems and Computing III*, Springer, Cham, 2019, pp. 476-492.
- [5] A.A. Zaporozhets, V.S. Eremenko, R.V. Serhienko and S.A. Ivanov, "Development of an intelligent system for diagnosing the technical condition of the heat power equipment," *Proceedings of the 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, Lviv, Ukraine, 11-14 September, 2018, pp. 48-51.
- [6] M.-Y. Moon, K.K. Choi, N. Gaul, D. Lamb, "Treating Epistemic Uncertainty Using Bootstrapping Selection of Input Distribution Model for Confidence-Based Reliability Assessment," *Journal of Mechanical Design*, vol. 141, issue 3, 031402, 2019
- [7] Z.Q. John Lu, "The elements of statistical learning: data mining, inference, and prediction," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 173, issue 3, pp. 693-694, 2010.
- [8] S. Nunes, M. Pimentel, F. Riberio, P. Milheiro-Oliveira, A. Carvalho, "Estimation of the tensile strength of UHPFRC layers based on non-destructive assessment of the fibre content and orientation," *Cement and Concrete Composites*, vol. 83, pp. 222-238, 2017
- [9] M.R. Jolly, A. Prabhakar, B. Sturzu, K. Hollstein, R. Singh, S. Thomas, P. Foote, A. Shaw, "Review of Non-destructive Testing (NDT) Techniques and their Applicability to

- Thick Walled Composites,” *Procedia CIRP*, vol. 38, pp. 129-136, 2015
- [10] Y. Ameer, H. Hedenmalm, N. Makarov, “Random normal matrices and Ward identities,” *The Annals of Probability*, vol. 43, issue 3, pp. 1157-1201, 2015.
- [11] W. He, G. Li, P. Hao, Y. Zeng, “Maximum entropy method-based reliability analysis with correlated input variables via hybrid dimension-reduction method,” *Journal of Mechanical Design*, vol. 141, issue 10, 101405, 2019.
- [12] A.A. Hassen, H. Taheri, U.K. Vaidya, “Non-destructive investigation of thermoplastic reinforced composites,” *Composites Part B: Engineering*, vol. 97, pp. 244-254, 2016.
- [13] K. Marhadi, S. Venkataraman, S.S. Pai, “Quantifying uncertainty in statistical distribution of small sample data using Bayesian inference of unbounded Johnson distribution,” *International Journal of Reliability and Safety*, vol. 6, issue 4, pp. 311-337, 2012.
- [14] T. Nishizu, A. Takezawa, M. Kitamura “Eigenfrequency-based damage identification method for non-destructive testing based on topology optimization,” *Engineering Optimization*, vol. 49, issue 3, pp. 417-433, 2015.
- [15] V. Eremenko, A. Zaporozhets, V. Isaenko, K. Babikova, “Application of Wavelet Transform for Determining Diagnostic Signs,” *CEUR Workshop Proceedings*, vol. 2387, pp. 202-214, 2019.
- [16] G. Barbato, G. Genta, A. Germak, R. Levi, G. Vicario, “Treatment of experimental data with discordant observations: issues in empirical identification of distribution,” *Measurement Science Review*, vol. 12, issue 4, pp. 133-140, 2012.
- [17] N.I.E. Farhana, M.S. Abdul Majid, M.P. Paulraj, E. Ahmadhilmi, M.N. Fakhzan, A.G. Gibson, “A novel vibration based non-destructive testing for predicting glass fibre/matrix volume fraction in composites using a neural network model,” *Composite Structures*, vol. 144, pp. 96-107, 2016.
- [18] A. Katunin, K. Dragan, M. Dziendzikowski, “Damage identification in aircraft composite structures: A case study using various non-destructive testing techniques,” *Composite Structures*, vol. 127, pp. 1-19, 2015.
- [19] M. Pavlovic, M. Doycinovic, S. Martinovic, M. Vlahovic, Z. Stevic, T. Volkov Husovic, “Non destructive monitoring of cavitation erosion of cordierite based coatings,” *Composites Part B: Engineering*, vol. 97, pp. 84-91, 2016.
- [20] M. Alatefi, S. Ahmad, M. Alkahtani, “Performance evaluation using multivariate non-normal process capability,” *Processes*, vol. 7, 833, 2019.
- [21] C. Mineo, S.G. Pierce, P.I. Nicholson, I. Cooper, “Robotic path planning for non-destructive testing – A custom MATLAB toolbox approach,” *Robotics and Computer-Integrated Manufacturing*, vol. 37, pp. 1-12, 2016.
- [22] X. Zhang, S. Du, “A linear Dirichlet mixture model for decomposing scenes: Application to analyzing urban functional zonings,” *Remote Sensing of Environment*, vol. 169, pp. 37-49, 2015.
- [23] C.-O. Amedee-Manesme, F. Barthelemy, D. Maillard, “Computation of the corrected Cornish–Fisher expansion using the response surface methodology: application to VaR and CVaR,” *Annals of Operations Research*, vol. 281, issue 1-2, pp. 423-453, 2019.
- [24] I. Holmes, M.T. Lacey, B.D. Wick, “Bloom’s inequality: commutators in a two-weight setting,” *Archiv der Mathematik*, vol. 106, issue 1, pp. 53-63, 2016.
- [25] M. Lints, S. Dos Santos, A. Salupere, “Solitary waves for non-destructive testing applications: Delayed nonlinear time reversal signal processing optimization,” *Wave Motion*, vol. 71, pp. 101-112, 2017.
- [26] P.J.A. Cayton and D.S. Mapa, *Time-varying conditional Johnson SU density in value-at-risk (VaR) methodology*, 2012, [Online]. Available: <https://mpra.ub.uni-muenchen.de/id/eprint/36206>
- [27] W.C. Torrez, J.T. Durham and R.D. Trueblood, “Performance measures for neural nets using Johnson distributions,” *Proceedings of the IEEE International Conference on Neural Networks*, San Francisco, CA, USA, 28 March-1 April, 1993, pp. 506-510.
- [28] P. Duchene, S. Chaki, A. Ayadi. P. Kraqczak, “A review of non-destructive techniques used for mechanical damage assessment in polymer composites,” *Journal of Material Science*, vol. 53, pp. 7915-7938, 2018.
- [29] M.C. Porcu, L. Pieczonka, A. Frau, W.J. Staszewski, F. Aymerich, “Assessing the scaling subtraction method for impact damage detection in composite plates,” *Journal of Nondestructive Evaluation*, vol. 36, 33, 2017.
- [30] H. Hayashi, Y. Kurita and T. Tsuji, “A non-Gaussian approach for biosignal classification based on the Johnson SU translation,” *Proceedings of the 2015 IEEE 8th International Workshop on Computational Intelligence and Applications (IWCIA)*, Hiroshima, Japan, 6-7 November, 2015, pp. 115-120.

- [31] L. Lin, Y. Saad, C. Yang, "Approximating Spectral Densities of Large Matrices," *SIAM Review*, vol. 58, issue 1, pp. 34-65, 2016.
- [32] A.P. Bartok, G. Csanyi, "Gaussian approximation potentials: A brief tutorial introduction," *Quantum Chemistry*, vol. 115, issue 16, pp. 1051-1057, 2015.
- [33] S. Anitha and B.M. Ramesh, "Network reconfiguration for loss minimization by using Johnson's algorithm," *Proceedings of the 2018 4th International Conference on Electrical Energy Systems (ICEES)*, Chennai, India, 7-9 February, 2018, pp. 680-684.
- [34] R.H. Storer, J.J. Swain, S. Venkatraman and J.R. Wilson, "Comparison of methods for fitting data using Johnson translation distributions," *Proceedings of the 1998 Winter Simulation Conference*, San Diego, CA, USA, 12-14 December, 1998, pp. 476-481.
- [35] S. Nakagawa, P.C.D. Johnson, H. Schielzeth, "The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded," *J. R. Soc. Interface*, vol. 14, issue 134, 2017.
- [36] G. Letac, L. Mattner, M. Piccioni, "The median of an exponential family and the normal law," *Statistics & Probability Letters*, vol. 133, pp. 38-41, 2018.
- [37] M. Baak, S. Gadatsch, R. Harrington, W. Verkerke, "Interpolation between multi-dimensional histograms using a new non-linear moment morphing method," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 771, pp. 39-48, 2015.
- [38] O.S. Sonbul, A.N. Kalashnikov, "Determining the operating distance of air ultrasound range finders: calculations and experiments," *International Journal of Computing*, vol. 13, issue 2, pp. 125-131, 2014.
- [39] A.O. Zaporozhets, O.O. Redko, V.P. Babak, V.S. Eremenko, V.M. Mokiychuk, "Method of indirect measurement of oxygen concentration in the air," *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, issue 5, pp. 105-114, 2018.
- [40] Y. Zhang, H. Liu, Y. Huang, Z. Deng, "A nonlinear ARMA-GARCH model with Johnson S_u innovations and its application to sea clutter modeling," *IEEE Access*, vol. 6, pp. 11888-118969, 2018.
- [41] K.K. Phoon and J. Ching, "Multivariate model for soil parameters based on Johnson distributions," *Proceedings of the 2018 Foundation Engineering in the Face of Uncertainty: Honoring Fred H. Kulhawy*, San Diego, USA, 3-7 March, 2013, pp. 337-353.
- [42] V.P. Babak, V.M. Mokiychuk, A.A. Zaporozhets, A.A. Redko, "Improving the efficiency of fuel combustion with regard to the uncertainty of measuring oxygen concentration," *Eastern-European Journal of Enterprise Technologies*, vol. 6, no. 8(84), pp. 54-59, 2016.
- [43] L. Zhang, H. Cheng, H. He, Q. Zhou and P. Zeng, "Johnson system based reliability evaluation of composite power system with wind farms," *Proceedings of the International Conference on Renewable Power Generation (RPG 2015)*, Beijing, China, 17-18 October 2015, p. 1-6.

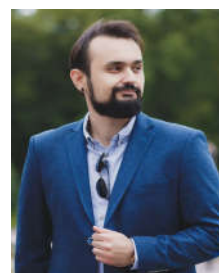


Vitaliy Babak is a Corresponding Member of the National Academy of Sciences of Ukraine, Doctor of Technical Sciences, Professor, Honored Scientist of Ukraine, laureate of the State Prize of Ukraine in science and technology, Deputy Director for Science in Institute of Engineering Thermophysics of NAS of Ukraine. He graduated from the National Technical University of Ukraine "Kiev Polytechnic Institute". Research interests: technical diagnostics, signal processing, information-measuring systems.



Volodymyr Eremenko is a Doctor of Technical Sciences, Professor of the National Technical University of Ukraine "Kiev Polytechnic Institute", Honored Scientist of Ukraine. He graduated from the National Technical University of Ukraine "Kiev Polytechnic Institute".

Research interests: technical diagnostics, signal processing, information-measuring systems, non-destructive testing of composite structure.



Artur Zaporozhets is Candidate of Technical Sciences (Ph.D.), Senior Researcher of the Department for Monitoring and Optimization of Thermal Power Processes of the Institute of Engineering Thermophysics of the National Academy of Sciences of Ukraine. He graduated from the National Aviation University with a degree in Applied Physics. Research interests: data analysis, Big Data, environmental monitoring, optimization of energy processes.