Original Russian Text © A. S. Zapevalov, A. S. Knyazkov, 2022,
published in MORSKOY GIDROFIZICHESKIY ZHURNAL, Vol. 38, Iss. 4 (2022)

Statistical Description of the Sea Surface by Two-Component Gaussian Mixture

A. S. Zapevalov ✉, A. S. Knyazkov

Marine Hydrophysical Institute of RAS, Sevastopol, Russian Federation
✉ *sevzepter@mail.ru*

Abstract
Purpose. The aim of the study is to analyze the possibility of applying the two-component Gaussian mixture with unequal dispersions in order to approximate the probability density function (PDF) of the sea surface elevations.
Methods and Results. The Gaussian mixture is constructed in the form of a sum of the Gaussians with different weights. Construction of the two-component Gaussian mixture with the regard for the condition imposed on the weight coefficients requires presetting of five parameters. The first four statistical moments of the sea surface elevations are applied for their calculation. The fifth parameter is used to fulfill the condition of unimodal distribution. To assess the possibility of using the approximations in the form of the Gaussian mixture, they were compared with the approximation based on the Gram-Charlier distribution, which was previously tested with direct wave measurement data. It is shown that at positive values of the excess kurtosis, in the range of a random value variation with a unit dispersion $|\xi| < 3$, two types of approximations are close; whereas at negative values of the excess kurtosis, noticeable discrepancies are observed in the area $|\xi| < 1$ (here ξ is the surface elevation normalized to the RMS value). Besides, it is also demonstrated that at the zero skewness, the PDF approximation in the form of the Gaussian mixture can be obtained only at the negative excess kurtosis.
Conclusions. At present, the models based on the truncated Gram-Charlier series, are usually applied to approximate the PDF elevations and slopes of the sea surface. Their disadvantage consists in the limited range, in which the distribution of the simulated characteristic can be described. The Gaussian mixtures are free from this disadvantage. A procedure for calculating their parameters is developed. To clarify the conditions under which the Gaussian mixtures can be used, direct comparison with the wave measurement data is required.

Keywords: sea surface, probability density function, Gaussian mixture, Gram-Charlier distribution, skewness, kurtosis

Acknowledgements: the study was carried out within the framework of the state assignment on theme No. 0555-2021-0004.

For citation: Zapevalov, A.S. and Knyazkov, A.S., 2022. Statistical Description of the Sea Surface by Two-Component Gaussian Mixture. *Physical Oceanography*, 29(4), pp. 395-403. doi:10.22449/1573-160X-2022-4-395-403

DOI: 10.22449/1573-160X-2022-4-395-403

Introduction

Describing the probability density function (PDF) of elevations generated by sea surface waves, approximations based on the Gram-Charlier distribution [1] are most widely used. The fundamental problem of applying these approximations is related to the fact that in practice the Gram-Charlier distribution is used in a truncated form, which allows to describe the distribution only in a limited variation range of a random value [2]. The need to solve a wide range of applied problems, primarily related to remote sensing of the ocean from space [3–5], has led to the search for new approaches to constructing PDF approximations of sea surface elevations.

For simulation of sea surface elevations, it was recently proposed to use Gaussian mixtures [6], which have long been widely used in other areas in fundamental and applied research [7–9]. Previously, Gaussian mixtures were applied to approximate the PDF of sea surface slopes [10, 11].

In the general case, the problems of determining the number of unimodality area modes and boundaries for a Gaussian mixture have not been solved [12, 13], so it is necessary to check the correctness of its use for each physical problem. The present paper is aimed to analyze the possibility of using and limitations for a two-component Gaussian mixture with different variances in the PDF approximation of sea surface elevations. The analysis is carried out for the ranges of variation of the third and fourth statistical moments of the sea surface elevation, determined from the Black Sea measurement data [14].

Two-component Gaussian mixture

A two-component Gaussian mixture of a random variable ξ has the following form [13]:

$$P_S(\xi) = \frac{\alpha_1}{\sqrt{2\pi}\,\sigma_1}\exp\left(-\frac{(\xi - m_1)^2}{2\sigma_1^2}\right) + \frac{\alpha_2}{\sqrt{2\pi}\,\sigma_2}\exp\left(-\frac{(\xi - m_2)^2}{2\sigma_2^2}\right), \tag{1}$$

where α_i is the *i*-th component weight (i = 1, 2), $\alpha_i \in (0,1)$; m_i is the mean value; σ_i^2 is the variance. The weight coefficients satisfy the condition

$$\alpha_1 + \alpha_2 = 1\,. \tag{2}$$

To construct model (1), taking into account condition (2), it is necessary to find five parameters: m_1, m_2, σ_1^2, σ_2^2, α_1. In [6], it was proposed to calculate them from the first five statistical moments of sea surface elevations. The disadvantage of this approach is that statistical moments no older than the fourth order are determined from the data of field measurements. Therefore, following [11], the first four statistical moments will be used to calculate the model parameters, leaving the fifth parameter (α_1) free. The parameter α_1 will be varied to satisfy the unimodality condition in PDF.

The statistical moments of a random value ξ are defined as

$$\mu_j = \int \xi^j P(\xi)\, d\xi\,.$$

For a two-component Gaussian mixture

$$\mu_j = \alpha_1 \mu_{i,1} + \alpha_2 \mu_{i,2}\,, \tag{3}$$

where $\mu_{j,i} = \int \xi^j P_i(\xi)\, d\xi$, $P_i(\xi)$ are the first and second terms in the model (1).

A general system for calculating the parameters of Gaussian mixture was proposed in [15]. Further, it will be assumed that the analyzed random value variation ξ is equal to one. Taking the average surface level equal to zero, taking into account (2) and (3), a system of equations for calculating the parameters of the model (1) is obtained:

$$\alpha_1 m_1 + (1-\alpha_1) m_2 = 0\,, \tag{4}$$

$$\alpha_1 \left(m_1^2 + \sigma_1^2 - 1 \right) + (1-\alpha_1)\left(m_2^2 + \sigma_2^2 - 1 \right) = 0\,, \tag{5}$$

$$\alpha_1 \left(m_1^3 + 3m_1 \sigma_1^2 - \mu_3 \right) + (1-\alpha_1)\left(m_2^3 + 3m_2 \sigma_2^2 - \mu_3 \right) = 0\,, \tag{6}$$

$$\alpha_1 \left(m_1^4 + 6m_1^2 \sigma_1^2 + 3\sigma_1^4 - \mu_4 \right) + (1-\alpha_1)\left(m_2^4 + 6m_2^2 \sigma_2^2 + 3\sigma_2^4 - \mu_4 \right) = 0\,. \tag{7}$$

The parameters μ_3 and $\mu_4 - 3$ are the skewness and excess kurtosis, respectively. System (4) – (7) will be investigated for the values $-0.2 < \mu_3 < 0.4$, $-0.4 < \mu_4 - 3 < 0.4$, which for the Black Sea corresponds to the ranges of their change for wind waves and swell [14].

Following the approach [16], we reduce the system (4) – (7) to one equation, successively excluding the unknowns. From the equation (4) we have $m_2 = \alpha_1 m_1 / (\alpha_1 - 1)$, then from the equations (5) – (7), introducing an intermediate unknown

$$\beta = \frac{\left(m_1^2 + \sigma_1^2 - 1 \right)}{m_1} = \frac{\left(m_2^2 + \sigma_2^2 - 1 \right)}{m_2}\,, \tag{8}$$

we obtain

$$\frac{m_1^3 + 3m_1 \sigma_1^2 - \mu_3}{m_1} = \frac{m_2^3 + 3m_2 \sigma_2^2 - \mu_3}{m_2}\,, \tag{9}$$

$$\frac{m_1^4 + 6m_1^2 \sigma_1^2 + 3\sigma_1^4 - \mu_4}{m_1} = \frac{m_2^4 + 6m_2^2 \sigma_2^2 + 3\sigma_2^4 - \mu_4}{m_2}\,. \tag{10}$$

Using the equation (8), we express the dispersions:

$$\sigma_1^2 = m_1 \beta + 1 - m_1^2\,, \qquad \sigma_2^2 = m_2 \beta + 1 - m_2^2\,.$$

Substituting the expressions for σ_1^2 and σ_2^2 in (9) and (10), after the transformations we obtain

$$m_1 m_2 (3\beta - 2(m_1 + m_2)) = -\mu_3\,, \; m_1 m_2 (3\beta^2 - 2(m_1^2 + m_1 m_2 + m_2^2)) = -\mu_4 + 3\,.$$

After a symmetrical change of variables of $w = m_1 + m_2$ and $v = m_1 m_2$ form we have

$$3\beta v - 2vw = -\mu_3\,, \qquad 3\beta^2 v - 2v(w^2 - v) = -\mu_4 + 3\,.$$

Combining these two equations, we obtain

$$6v^3 - 2v^2 w^2 - 4vw\mu_3 + 3(\mu_4 - 3)v + \mu_3^2 = 0\,.$$

By reverse substitution, we express variables w and v in terms of m_1 and finally obtain

$$2\alpha_1^2(\alpha_1-\alpha_1^2-1)m_1^6-4\mu_3\alpha_1(2\alpha_1-1)(\alpha_1-1)^2m_1^3+$$
$$+3(\mu_4-3)\alpha_1(\alpha_1-1)^3m_1^2+\mu_3^2(\alpha_1-1)^4=0. \tag{11}$$

Thus, the original system of equations (8) – (11) is reduced to one sixth degree equation with respect to variable m_1 with known μ_3, μ_4 values and a free parameter α_1.

Let us consider the general properties of the sparse polynomial (11) as its coefficients change. In general terms

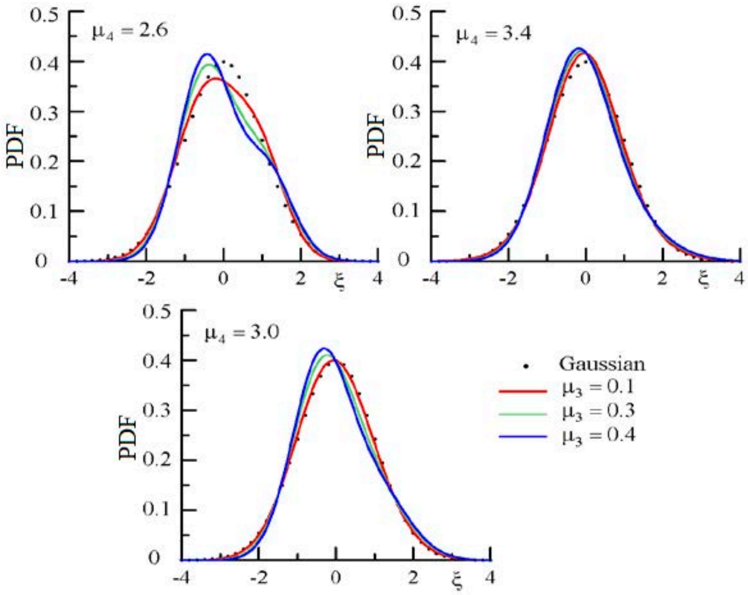
$$b_0m_1^6+b_3m_1^3+b_4m_1^2+b_6=0, \tag{12}$$

where $b_0=2\alpha_1^2(\alpha_1-\alpha_1^2-1)$; $b_3=-4\mu_3\alpha_1(2\alpha_1-1)(\alpha_1-1)^2$; $b_4=3(\mu_4-3)\alpha_1(\alpha_1-1)^3$; $b_6=\mu_3^2(\alpha_1-1)^4$. In the range $\alpha_1\in(0,1), b_0<0$, $b_6\geq 0$, while $b_6=0$ only when $\mu_3=0$. In case when $b_6>0$, since the polynomial degree is even, according to the Descartes rule, equation (12) has both positive and negative real roots, since there is a change of sign in the series of its coefficients. When $b_3>0$ and $b_4<0$, the number of sign changes is three, and in other cases – only one. The sign of the coefficient b_4 depends only on the sign of μ_4-3, b_3 can change sign when changing both μ_3 and α_1.

Let us consider separately the case when $\mu_3=0$. Equation (12) takes the form of $m_1^2(b_0m_1^4+b_4)=0$. Since $b_0<0$, it has nonzero real solutions only for positive values of b_4, which corresponds to the condition $\mu_4-3<0$. Consequently, when $\mu_3=0$ and $\mu_4-3>0$ the two-component mixture model has no real solutions and cannot be used. This limitation of two-component mixtures is obtained in general form for any process with the indicated values μ_3 and μ_4.

The values of m_1, satisfying (11), are found numerically by Newton's method for given μ_3 and μ_4, varying α_1. Some of the solutions obtained should be excluded based on the condition of σ_1^2 and σ_2^2 positivity. For the obtained values of m_1, satisfying (11), and corresponding to α_1 using the original system (4) – (7) m_2, σ_1^2 and σ_2^2 were calculated and Gaussian mixture PDF was built. In the general case, model (1) can be both unimodal and bimodal [13, 17]. Since the distribution of wind wave elevations is unimodal, the PDF derivative was additionally analyzed and only the parameter values were selected when $P_S(\xi)$ had a single extremum (this is equivalent to the unimodality condition).

Along with symmetry with respect to triples of numbers (m_1, σ_1^2, α_1) and (m_2, σ_2^2, α_2), the system of equations (8) – (10) has an additional symmetry property: replacing (m_1, m_2, μ_3) by (m_1, m_2, $-\mu_3$) gives identical solutions, so it is sufficient to analyze only for positive values of μ_3, i.e. for positive values of the asymmetry coefficient. PDF approximations in the form (1) are shown in Fig. 1.



F i g. 1. PDF approximations by the Gaussian mixture

Comparison with Gram-Charlier distribution

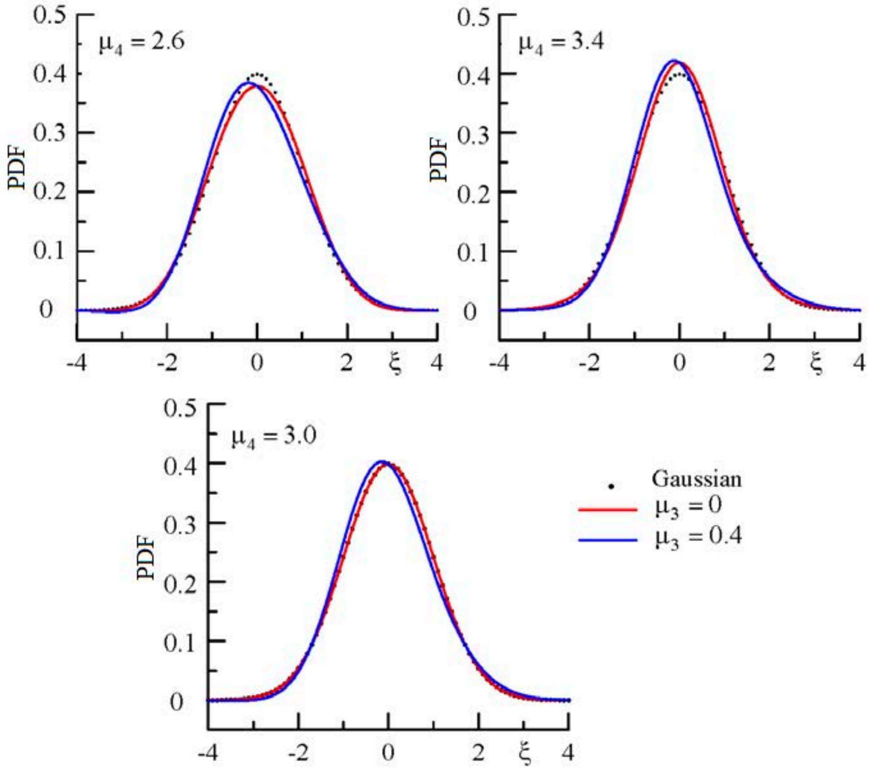
Sea surface waves are a quasi-Gaussian process [1, 18, 19]. The probability density function of such a process with unit variance can be represented as follows [2]:

$$P_{G-C}(\xi) = \sum_{i=0}^{\infty} C_i H_i(\xi) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\xi^2\right), \qquad (13)$$

where C_i are the series coefficients; H_i are the orthogonal Hermite *i*-order polynomials. C_i coefficients are calculated by statistical moments. Since the statistical moments of the sea surface elevations are known only up to the fourth order inclusive, instead of (13) we obtain

$$P_{G-C}(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\xi^2\right)\left\{1 + \frac{\mu_3}{6} H_3(\xi) + \frac{(\mu_4 - 3)}{24} H_4(\xi)\right\}. \qquad (14)$$

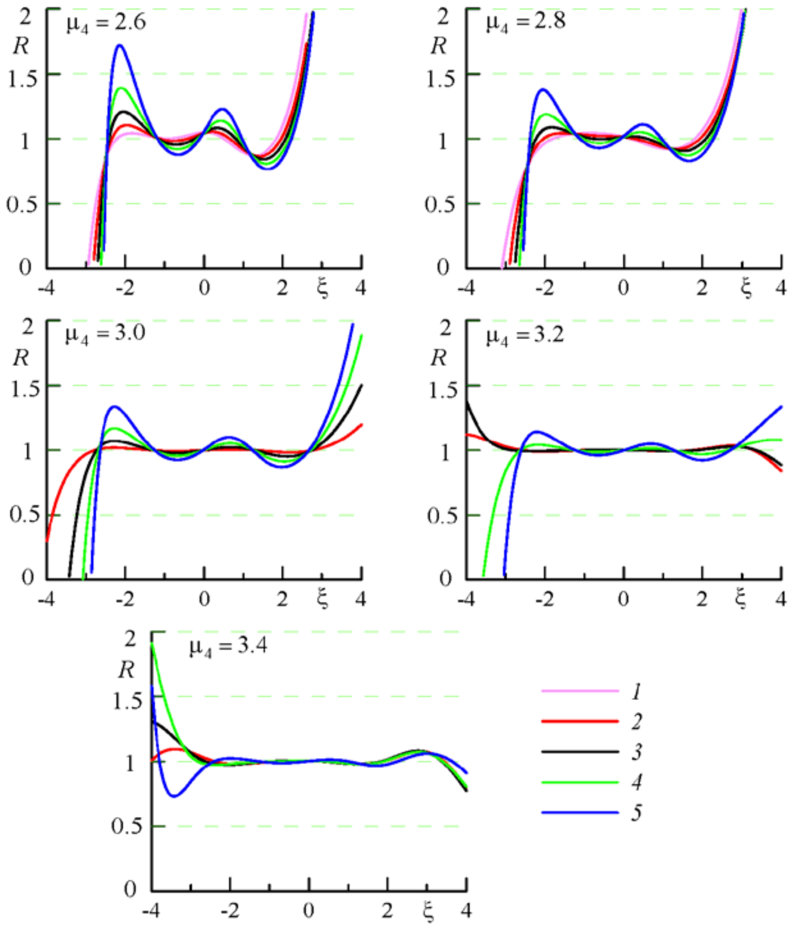
PDF approximations in the form (14) are shown in Fig. 2. Expansion of the function into a series that includes a relatively small number of terms leads to a narrowing of the region where this approximation is valid [2]. In particular, it can be seen that function $P_{G-C}(\xi)$ with μ_3 and μ_4 values determined in the field experiments can take negative values.



F i g. 2. PDF approximation by the Gram-Charlier distribution

Previously, the approximation (14) was compared with empirical PDFs of sea surface elevations obtained from measurements of sea waves carried out on a stationary oceanographic platform of Marine Hydrophysical Institute [20]. The relative error ε averaged over the ensemble of situations for the range $|\xi| < 3$ lies in the range of – 0.02 ... 0.07. The scatter of ε values in the domain $|\xi| < 1$ does not exceed the 0.08 level; outside the specified domain, the scatter grows rapidly.

Approximation (14) verified according to field measurements can be used for a preliminary assessment of the $P_S(\xi)$ correctness. The ratio $R(\xi) = P_{G-C}(\xi) / P_S(\xi)$ is shown in Fig. 3. It can be seen that in the case when the excess kurtosis is less than zero, the functions $P_S(\xi)$ and $P_{G-C}(\xi)$ differ noticeably. Moreover, differences are observed even in the domain $|\xi| < 1$ where, as noted above, there was a coincidence of $P_{G-C}(\xi)$ with the data of wave measurements. For positive values of excess kurtosis in the domain $|\xi| < 3$, the functions $P_S(\xi)$ and $P_{G-C}(\xi)$ are close.



F i g. 3. Dependences of ratio $R = P_{G-C}/P_S$ on the parameters μ_3 and μ_4. Curves *1–5* correspond to the μ_3 values from 0 to 0.4 with a step 0.1

The reliability evaluation of the PDF approximation of sea surface elevations by a Gaussian mixture comparing with a distribution constructed on the basis of a truncated Gram-Charlier distribution is preliminary. The next step should be a direct comparison of $P_S(\xi)$ with empirical PDFs of sea surface elevations.

Conclusion

The main results of the study carried out are as follows.

1. A technique for calculating the two-component Gaussian mixture parameters for approximation of PDF of sea surface elevations has been developed. The analysis was carried out for the ranges of changes in the third (μ_3) and fourth (μ_4) statistical moments of the sea surface elevations, determined from the data of direct wave measurements in the Black Sea.

2. Symmetry properties of the Gaussian mixture equations, which reduce the number of calculations, are distinguished. In general, it is shown that PDF

approximations in the form of a Gaussian mixture in a particular case $\mu_3 = 0$ can be obtained only under the condition $\mu_4 < 3$.

3. A comparison between the PDF approximation in the form of a two-component Gaussian mixture and the approximation based on the truncated Gram-Charlier distributions is carried out. When $\mu_4 > 3$ in $|\xi| < 3$ domain the approximations are close, at $\mu_4 < 3$ significant discrepancies are observed. To clarify the conditions under which Gaussian mixtures can be used, a direct comparison with wave measurement data is needed.

REFERENCES

1. Longuet-Higgins, M.S., 1963. The Effect of Non-Linearities on Statistical Distributions in the Theory of Sea Waves. *Journal of Fluid Mechanics*, 17(3), pp. 459-480. doi:10.1017/S0022112063001452
2. Kwon, O.K., 2022. Analytic Expressions for the Positive Definite and Unimodal Regions of Gram-Charlier Series. *Communications in Statistics – Theory and Methods,* 51(15), pp. 5064-5084. doi:10.1080/03610926.2020.1833219
3. Cox, C. and Munk, W., 1954. Measurements of the Roughness of the Sea Surface from Photographs of the Sun’s Glitter. *Journal of the Optical Society of America*, 44(11), pp. 838-850. doi:10.1364/JOSA.44.000838
4. Bréon, F.M. and Henriot, N., 2006. Spaceborne Observations of Ocean Glint Reflectance and Modeling of Wave Slope Distributions. *Journal of Geophysical Research: Oceans,* 111(C6), C06005. doi:10.1029/2005JC003343
5. Pokazeev, K.V., Zapevalov, A.S. and Pustovoytenko, V.V., 2013. The Simulation of a Radar Altimeter Return Waveform. *Moscow University Physics Bulletin*, 68(5), pp. 420-425. doi:10.3103/S0027134913050135
6. Gao, Z., Sun, Z. and Liang, S., 2020. Probability Density Function for Wave Elevation Based on Gaussian Mixture Models. *Ocean Engineering,* 213, 107815. doi:10.1016/j.oceaneng.2020.107815
7. Teicher, H., 1963. Identifiability of Finite Mixtures. *The Annals of Mathematical Statistics,* 34(4), pp. 1265-1269. doi:10.1214/aoms/1177703862
8. Ray, S. and Ren, D., 2012. On the Upper Bound of the Number of Modes of a Multivariate Normal Mixture. *Journal of Multivariate Analysis,* 108, pp. 41-52. doi:10.1016/j.jmva.2012.02.006
9. Améndola, C., Engström, A. and Haase, C., 2020. Maximum Number of Modes of Gaussian Mixtures. *Information and Reference: A Journal of the IMA*, 9(3), pp. 587-600. doi:10.1093/imaiai/iaz013
10. Tatarskii, V.I., 2003. Multi-Gaussian Representation of the Cox-Munk Distribution for Slopes of Wind-Driven Waves. *Journal of Atmospheric and Oceanic Technology*, 20(11), pp. 1697-1705. doi:10.1175/1520-0426(2003)020<1697:MROTCD>2.0.CO;2
11. Zapevalov, A.S. and Ratner, Yu.B., 2003. Analytic Model of the Probability Density of Slopes of the Sea Surface. *Physical Oceanography*, 13(1), pp. 1-13. doi:10.1023/A:1022444703787
12. Aprausheva, N.N. and Sorokin, S.V., 2015. [*Notes on Gaussian Mixture*]. Moscow: Dorodnicyn Computing Centre of the RAS, 144 p. doi:10.13140/RG.2.2.33609.34404 (in Russian).
13. Aprausheva, N.N. and Sorokin, S.V., 2013. Exact Equation of the Boundary of Unimodal and Bimodal Domains of a Two-Component Gaussian Mixture. *Pattern Recognition and Image Analysis,* 23(3), pp. 341-347. doi:10.1134/S1054661813030024
14. Zapevalov, A.S. and Garmashov, A.V., 2022. The Appearance of Negative Values of the Skewness of Sea-Surface Waves. *Izvestiya, Atmospheric and Oceanic Physics,* 58(3), pp. 263-269. doi:10.1134/S0001433822030136

15. Pearson, K., 1894. III. Contributions to the Mathematical Theory of Evolution. *Philosophical Transactions of the Royal Society A. Mathematical, Physical and Engineering Sciences*, 185, pp. 71-110. doi:10.1098/rsta.1894.0003
16. Cohen, A.C., 1967. Estimation in Mixtures of Two Normal Distributions. *Technometrics*, 9(1), pp. 15-28. doi:10.1080/00401706.1967.10490438
17. Carreira-Perpiñán, M.Á., 2000. Mode-Finding for Mixtures of Gaussian Distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), pp. 1318-1323. doi:10.1109/34.888716
18. Zapevalov, A.S. and Garmashov, A.V., 2021. Skewness and Kurtosis of the Surface Wave in the Coastal Zone of the Black Sea. *Physical Oceanography*, 28(4), pp. 414-425. doi:10.22449/1573-160X-2021-4-414-425
19. Babanin, A.V. and Polnikov, V.G., 1995. On the Non-Gaussian Nature of Wind Waves. *Physical Oceanography*, 6(3), pp. 241-245. doi:10.1007/BF02197522
20. Zapevalov, A.S., Bol'shakov, A.N. and Smolov, V.E., 2011. Simulating of the Probability Density of Sea Surface Elevations Using the Gram-Charlier Series. *Oceanology*, 51(3), pp. 407-414. doi:10.1134/S0001437011030222

About the authors:
Aleksandr S. Zapevalov, Chief Researcher, Marine Hydrophysical Institute of RAS (2 Kapitanskaya St., Sevastopol, Russian Federation, 299011), Dr. Sci. (Phis.-Math.), **Scopus Author ID: 7004433476**, **Web of Science ResearcherID: V-7880-2017**, sevzepter@mail.ru

Aleksandr S. Knyazkov, Leading Engineer, Marine Hydrophysical Institute of RAS (2 Kapitanskaya St., Sevastopol, Russian Federation, 299011), **ORCID ID: 0000-0003-1119-1757**, fizfak83@yandex.ru

Contribution of the authors:
Aleksandr S. Zapevalov – formulation of the aims and tasks of the study; development of requirements imposed on the model; the paper text preparation; editing and supplementing the paper text
Aleksandr S. Knyazkov – obtaining a mathematical model, testing the model; supplementing to the paper text

The authors have read and approved the final manuscript.
The authors declare that they have no conflict of interest.