ISSN 2413-5577, Ecological Safety of Coastal and Shelf Zones of Sea, 2024, No. 1, pp. 20–30.

Original article

Distribution of Sea Surface Elevations in the Form of a Two-Component Gaussian Mixture

A. S. Zapevalov *, A. S. Knyazkov

Marine Hydrophysical Institute of RAS, Sevastopol, Russia
** e-mail: sevzepter@mail.ru*

Abstract
The approximation of the probability density function of sea surface elevations by a two-component Gaussian mixture has been verified. For verification, the data of direct wave measurements obtained on a stationary oceanographic platform, installed in the Black Sea, were used. The approximation correctness criterion is the relative error ε of deviation of the model of probability densities function from the experimental function calculated from the measurement data. The average error $\langle\varepsilon\rangle$ over the ensemble of situations is small if $|\xi| < 3$. The standard deviation δ is minimal if $|\xi| \approx 0$ and is equal to 0.12, if $|\xi| = 3$ then $\delta \approx 0.5$. It is shown that the error $\langle\varepsilon\rangle$ has a systematic component, which depends on the deviations of the third and fourth statistical moments from the values corresponding to the Gaussian distribution. A semi-empirical relationship has been constructed to take this component into account. It is noted that the approximation accuracy can be increased by 2–3 times by eliminating the systematic component.

Keywords: Gaussian mixture, sea waves, surface elevation, nonlinear waves, statistical moment, Black Sea

Acknowledgements: The work was performed under state assignment of Marine Hydrophysical Institute of RAS on topic FNNN-2021-0004 “Fundamental studies of oceanological processes which determine the state and evolution of the marine environment influenced by natural and anthropogenic factors, based on observation and modeling methods”.

For citation: Zapevalov, A.S. and Knyazkov, A.S., 2024. Distribution of Sea Surface Elevations in the Form of a Two-Component Gaussian Mixture. *Ecological Safety of Coastal and Shelf Zones of Sea*, (1), pp. 20–30.

Распределение возвышений морской поверхности в форме двухкомпонентной гауссовой смеси

А. С. Запевалов *, А. С. Князьков

Морской гидрофизический институт РАН, Севастополь, Россия
* *e-mail: sevzepter@mail.ru*

Аннотация
Верифицирована аппроксимация функции плотности вероятностей возвышений морской поверхности двухкомпонентной гауссовой смесью. Для верификации использованы данные прямых волновых измерений, полученные на стационарной океанографической платформе, установленной в Черном море. Критерием корректности аппроксимации выбрана относительная ошибка ε отклонения модельной функции плотности вероятности от экспериментальной функции, рассчитанной по данным волновых измерений. Средняя по ансамблю ситуаций относительная ошибка $\langle\varepsilon\rangle$ мала, если значения нормированного на среднеквадратическую величину возвышения поверхности лежат в области $|\xi| < 3$. Среднеквадратическое отклонение относительной ошибки минимально при $|\xi| \approx 0$ и равняется 0.12, при $|\xi| = 3$ возрастает до ~ 0.5. Показано, что ошибка $\langle\varepsilon\rangle$ имеет систематическую составляющую, которая зависит от отклонений третьего и четвертого статистических моментов от значений, соответствующих распределению Гаусса. Построена полуэмпирическая зависимость, позволяющая учесть эту составляющую. Отмечено, что точность аппроксимации можно повысить в 2–3 раза, исключив систематическую составляющую.

Ключевые слова: гауссова смесь, морская поверхность, нелинейные волны, статистический момент, Черное море

Благодарности: работа выполнена в рамках госзадания ФГБУН ФИЦ МГИ по теме FNNN-2021-0004 «Фундаментальные исследования океанологических процессов, определяющих состояние и эволюцию морской среды под влиянием естественных и антропогенных факторов, на основе методов наблюдения и моделирования».

Для цитирования: *Запевалов А. С., Князьков А. С.* Распределение возвышений морской поверхности в форме двухкомпонентной гауссовой смеси // Экологическая безопасность прибрежной и шельфовой зон моря. 2024. № 1. С. 20–30. EDN EHKUET.

Introduction

Sea surface waves are a weakly nonlinear process, and the statistical distributions of sea surface elevations and slopes are close to the Gaussian distribution [1]. Although deviations from the Gaussian distribution are small, they play an important role in applications related to ocean remote sensing [2, 3], as well as when forecasting the occurrence of anomalous waves [4].

As a rule, distributions based on truncated Gram–Charlier or Edgeworth series are used for the sea surface statistical description [5, 6]. The distributions are the expansion of the desired probability density function in Chebyshev–Hermite orthogonal polynomials. The use of truncated series leads to distortions in the desired probability density function due to the appearance of negative values in it, as well as several local maxima [7–9].

The relevance of the search for new approaches to the statistical description of the sea surface is determined by the fact that existing models do not make it possible to construct a probability density function of sea surface elevations over the entire range of their changes. One possible solution to this problem is to approximate the distribution of a quasi-Gaussian process by a two-component Gaussian mixture. Distributions of this type have not yet found wide application in oceanology, which may be due to the complex procedure for calculating their parameters [10]. For the first time, the use of such a model to describe the sea surface was independently proposed in [11, 12], in which probability density functions were constructed for the sea surface slopes. Recently, a two-component Gaussian mixture has been proposed to describe the distributions of sea surface elevations [13]. Unknown parameters for the desired Gaussian mixture are calculated based on the known statistical moments as in the construction of the Gram–Charlier and Edgeworth distributions.

This work aims at analyzing the possibility and limits of a two-component Gaussian mixture in order to describe the distribution of sea surface elevations. The analysis is based on direct measurements of sea waves carried out in the Black Sea.

Two-component Gaussian mixture

Finite Gaussian mixtures are widely used in various fields to approximate unknown probability density functions [9, 14]. The two-component Gaussian mixture of random variable ξ is as follows [15]

$$P_S(\xi) = \sum_i \frac{\alpha_i}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(\xi - m_i)^2}{2\sigma_i^2} \right), \quad (1)$$

where α_i – weight of the i-th component (i = 1, 2), $\alpha_i \in (1, 2)$; m_i – expected value; σ_i^2 – variance. Weighting coefficients satisfy the condition

$$\alpha_1 + \alpha_2 = 1. \quad (2)$$

Taking into account condition (2), it is necessary to find five parameters: m_1, m_2, σ_1, σ_2 and α_1 to construct $P_S(\xi)$. In [13], it was proposed to calculate them based on the first five statistical moments of sea surface elevations. The disadvantage of this approach is that according to wave measurements under marine conditions, as a rule, statistical moments are determined only up to the fourth order inclusive [16–18]. Therefore, we will use the first four statistical moments to calculate the model parameters (m_1, m_2, σ_1, σ_2) leaving the fifth parameter (α_1) free [11]. Parameter α_1 will be varied to satisfy the condition of distribution unimodality.

The procedure for calculating model parameters (1) is described in [10]. It amounts to solving the system of equations

$$\alpha_1 m_1 + (1 - \alpha_1) m_2 = \mu_1, \quad (3)$$

$$\alpha_1 \left(m_1^2 + \sigma_1^2 \right) + (1 - \alpha_1) \left(m_2^2 + \sigma_2^2 \right) = \mu_2, \quad (4)$$

$$\alpha_1 \left(m_1^3 + 3 m_1 \sigma_1^2 \right) + (1 - \alpha_1) \left(m_2^3 + 3 m_2 \sigma_2^2 \right) = \mu_3, \quad (5)$$

$$\alpha_1 \left(m_1^4 + 6 m_1^2 \sigma_1^2 + 3\sigma_1^4 \right) + (1 - \alpha_1) \left(m_2^4 + 6 m_2^2 \sigma_2^2 + 3\sigma_2^4 \right) = \mu_4, \quad (6)$$

where μ_i – statistical moment of order i :

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

Let us assume that the average level of the surface is zero ($\mu_1 = 0$), and the variance of the analyzed random variable is equal to 1 ($\mu_2 = 1$). Parameters μ_3 and $\mu_4 - 3$ are the skewness and excess kurtosis, respectively. System of equations (3)–(6) is symmetric with respect to triples of parameters (m_1, σ_1^2, α_1) and (m_2, σ_2^2, α_2).

Verification

To verify the model probability density function of sea surface elevations (1), the data of wave measurements obtained on a stationary oceanographic platform of Marine Hydrophysical Institute of RAS were used [19]. The measurements were carried out during December 2018. The platform was located in the Black Sea 600 m from the coast at a depth of about 30 m. The waves were measured with a string wave recorder [20].

The measurements were carried out under wind conditions that varied from calm to wind speed of 25 m/s. Significant wave heights (the average height of 1/3 of the highest waves) varied from 0.23 m to 2.26 m, the maximum wave height reached 4.9 m. The wavelengths corresponding to the peak of the wave spectrum ranged from 10 to 120 m.

The verification took place as follows. Continuous wave measurements were divided into wave records lasting 20 min. The total volume of data for analysis was more than 2200 wave records. Each wave record was centered and normalized so that its variance was equal to one, then experimental probability density function $P_E(\xi)$ was calculated for each wave record. Statistical moments $\mu_3 = \langle \xi^3 \rangle$ and $\mu_4 = \langle \xi^4 \rangle$ were also determined so that to calculate the parameters of two-component Gaussian mixture $P_S(\xi)$. Here and below, symbol $\langle \; \rangle$ means averaging.

According to wave measurements previously carried out in the Black Sea, the values of statistical moments μ_3 and μ_4 can be found mainly in the following ranges [19]

$$-0.2 < \mu_3 < 0.3 \quad \text{и} \quad 2.6 < \mu_4 < 3.4 . \tag{7}$$

The same ranges were determined from measurements in the North Sea [18]. As a rule, exceeding the specified ranges occurs in situations where abnormally high waves (rogue waves) are observed. [17]. In this work, we will limit ourselves to the analysis of situations when μ_3 and μ_4 satisfy condition (7).

The experimental probability density function is calculated based on the analysis of the histogram of sea surface elevations. Width of intervals $\Delta\xi$ was taken equal to 0.45. Function $P_E(\xi)$ was obtained from the histogram by normalizing it to the total number of points in the wave record and to the width of the interval.

The verification procedure for a two-component Gaussian mixture model consists of comparing functions $P_E(\xi)$ and $P_S(\xi)$. The criterion for the correspondence of model (1) to wave measurement data is relative error

$$\varepsilon(\xi) = \frac{P_S(\xi) - P_E(\xi)}{P_E(\xi)},$$

for which average value $\langle\varepsilon(\xi)\rangle$ and standard deviation $\delta(\xi) = \langle(\varepsilon(\xi) - \langle\varepsilon(\xi)\rangle)^2\rangle^{0.5}$ are calculated.

To calculate Gaussian mixture $P_S(\xi)$, the procedure described in [10] was chosen. Taking into account condition (2), system of equations (3)–(6) was reduced to one sixth degree polynomial equation in m_1

$$2\alpha_1^2(\alpha_1 - \alpha_1^2 - 1)m_1^6 - 4\mu_3\alpha_1(2\alpha_1 - 1)(\alpha_1 - 1)^2 m_1^3 + \\ +3(\mu_4 - 3)\alpha_1(\alpha_1 - 1)^3 m_1^2 + \mu_3^2(\alpha_1 - 1)^4 = 0, \quad (8)$$

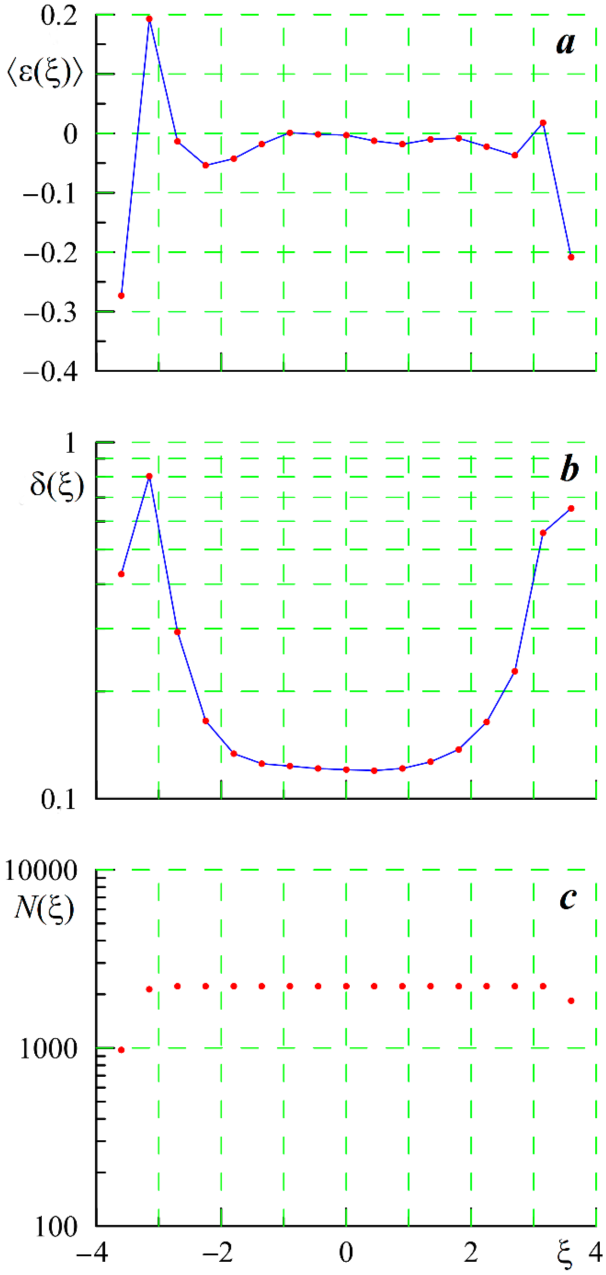
the solutions of which for given values μ_3 and μ_4 were found numerically by Newton's method by varying α_1 The coefficients included in equation (8) were analyzed in [10], where it was shown that, except for the rare case when $\mu_3 = 0$ and $\mu_4 > 3$, it could always be solved and the construction of a probability density function was possible. From several solutions obtained for various possible α_1, the one was chosen that corresponded to the physical condition of unimodality of the resulting distribution and the positivity of values σ_1^2 и σ_2^2, which were recalculated, like m_2, from value m_1 according to the method discussed in [3]. Values μ_3 and μ_4 calculated for model Gaussian mixture $P_S(\xi)$ obtained as a result of solving equation (8) were compared with the values calculated from the wave record and used in original equations (3)–(6). The accuracy of agreement between values μ_3 and μ_4 calculated from the Gaussian mixture and from the wave record is achieved no worse than 10^{-3}.

Figure 1 shows functions $\langle\varepsilon(\xi)\rangle$ and $\delta(\xi)$. Here, $N(\xi)$ is number of points from which statistical characteristics were calculated in given interval $\Delta\xi$. Functions $\langle\varepsilon(\xi)\rangle$ and $\delta(\xi)$ are average over the ensemble of situations in which measurements were carried out, with μ_3 and μ_4 satisfying condition (7). Parameters $\mu_3 = 0$ and $\mu_4 > 3$ in 11 wave records led to their exclusion from consideration for the reason stated above.

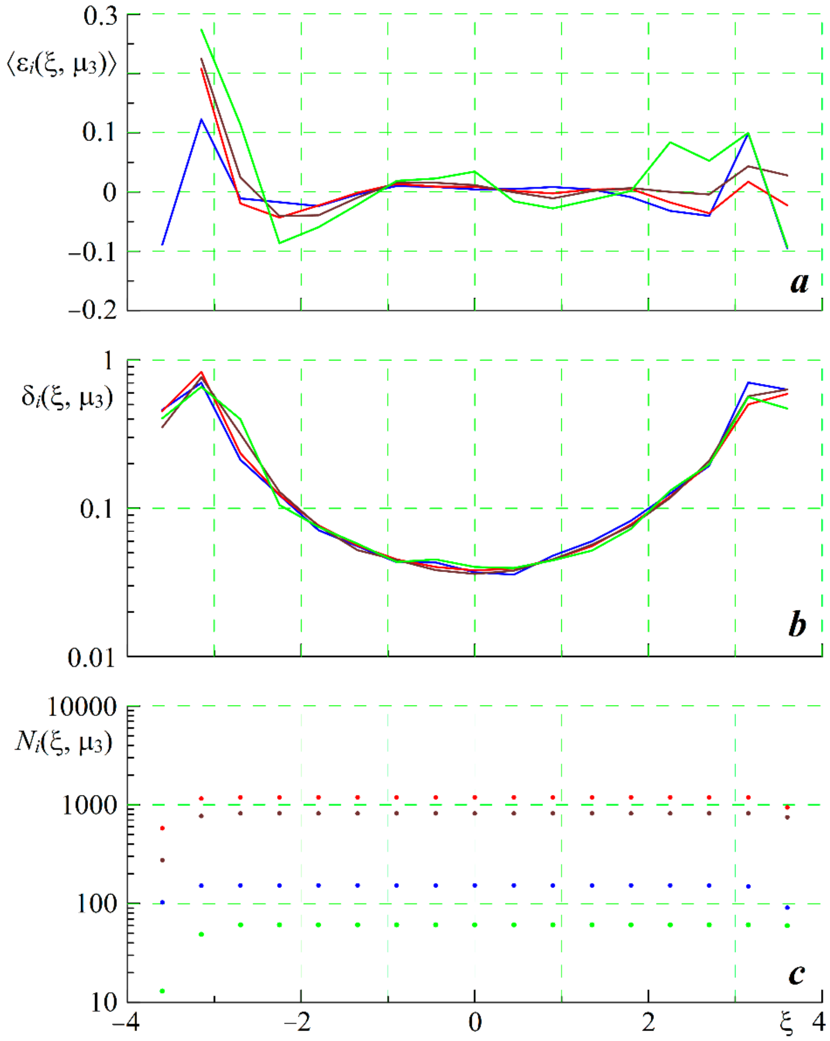
Analysis of deviations of model function $P_S(\xi)$ from experimental one $P_E(\xi)$ given in Fig. 1 indicates their smallness in the vicinity of point $\xi = 0$ and increase with $|\xi|$. As for range $|\xi| < 3$, the parameters characterizing this deviation satisfy conditions

$$|\langle\varepsilon(\xi)\rangle| < 0.05, \quad \delta(\xi) < 0.3.$$

For further analysis, all data were divided into groups corresponding to four ranges of the third statistical moment: group 1 – $-0.2 < \mu_3 \le 0$, group 2 – $0 < \mu_3 \le 0.1$, group 3 – $0.1 < \mu_3 \le 0.2$, group 4 – $0.2 < \mu_3 \le 0.3$. Figure 2 shows variables $\langle\varepsilon(\xi, \mu_3)\rangle$ and $\delta_i(\xi, \mu_3)$ calculated for each group. Here, index i taking values from one to four corresponds to the group number. Parameter $N_i(\xi, \mu_3)$ shows the number of



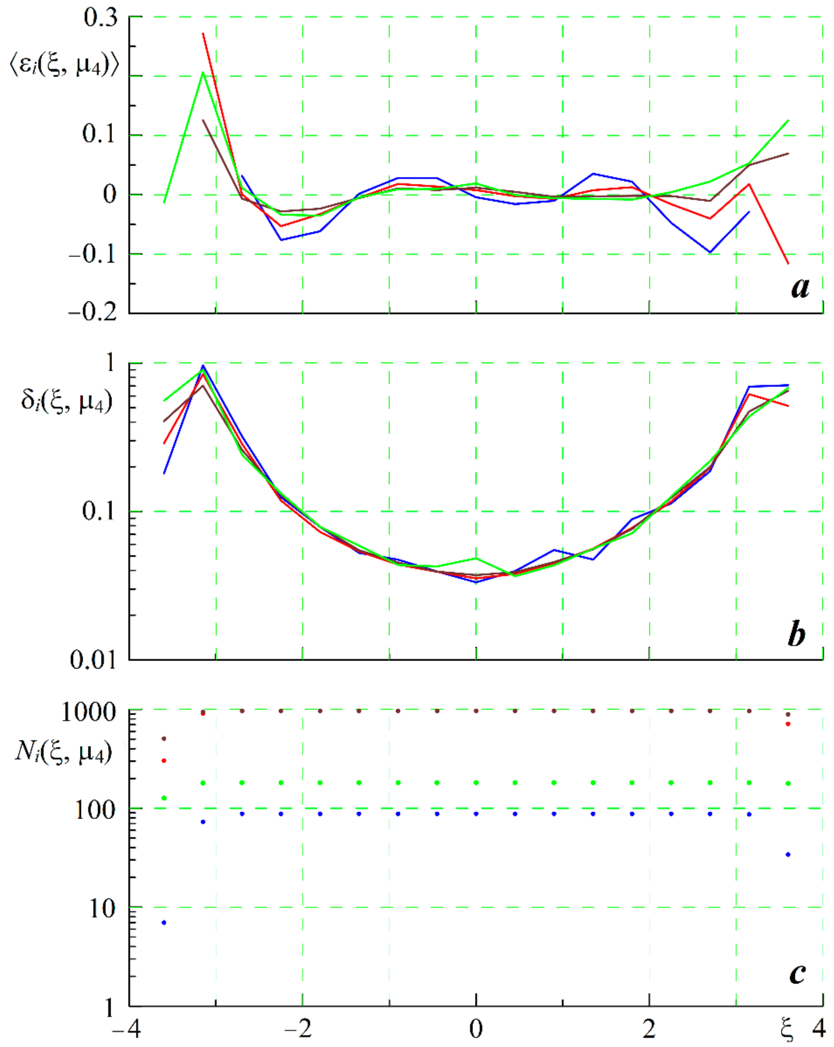
F i g . 1 . Relative error ε(ξ) (*a*) and standard deviation δ(ξ) (*b*) calculated for an ensemble of situations, the number of points *N*(ξ) from which statistical characteristics were calculated in a given interval Δξ (*c*)



F i g . 2 . Variables $\varepsilon(\xi)$ (*a*), $\delta(\xi)$ (*b*), $N(\xi)$ (*c*) calculated for four ranges μ_3: $-0.2 < \mu_3 \le 0$ (blue), $0 < \mu_3 \le 0.1$ (red), $0.1 < \mu_3 \le 0.2$ (brown), $0.2 < \mu_3 \le 0.3$ (green)

points from which the values $\langle \varepsilon_i(\xi, \mu_3) \rangle$ and $\delta_i(\xi, \mu_3)$ were calculated. Average value of relative error $\langle \varepsilon_i(\xi, \mu_3) \rangle$ depends significantly on the group for which it was calculated. At the same time, standard deviation $\delta_i(\xi, \mu_3)$ is almost the same for all groups. The discrepancy between $P_S(\xi)$ and $P_E(\xi)$ depends on how much statistical moment μ_3 deviates from the zero value corresponding to the Gaussian distribution. The greatest discrepancies are observed for group 4.

We use a similar approach to analyze the approximation of the probability density of sea surface elevations for different values of the fourth statistical moment. Let us divide the data into groups corresponding to four ranges μ_4: group 1 – $2.6 < \mu_4 \le 2.8$, group 2 – $2.8 < \mu_4 \le 3.0$, group 3 – $3.0 < \mu_4 \le 3.2$, group 4 – $3.2 < \mu_4 \le 3.4$. Figure 3 shows variables $\langle\varepsilon_i(\xi, \mu_4)\rangle$ and $\delta_i(\xi, \mu_4)$ calculated for the specified groups.



F i g . 3 . Variables $\varepsilon(\xi)$ (*a*), $\delta(\xi)$ (*b*), $N(\xi)$ (*c*) calculated for four ranges μ_4: $2.6 < \mu_4 \le 2.8$ (blue), $2.8 < \mu_4 \le 3.0$ (red), $3.0 < \mu_4 \le 3.2$ (brown), $3.2 < \mu_4 \le 3.4$ (green)

Division into groups according to the range of changes in statistical moments μ_3 and μ_4 results in a significant change in the relative error in the approximation of the probability density of sea surface elevations. In range $|\xi| < 2$, values $\delta_i(\xi, \mu_3)$ and $\delta_i(\xi, \mu_4)$ are 2–3 times lower than values $\delta(\xi)$ calculated for the entire ensemble of situations. This makes it possible to describe the probability density function by the semi-empirical relationship

$$P_\xi(\xi) = P_S(\xi)\left(1 + \langle \varepsilon_E(\xi) \rangle\right),$$

where $\langle \varepsilon_E(\xi) \rangle$ is average relative error calculated for corresponding ranges μ_3 and μ_4.

Conclusion

The approximation of the probability density function of sea surface elevations by a two-component Gaussian mixture was verified for the values of the third and fourth statistical moments which vary within $-0.2 < \mu_3 < 0.3$ and $2.6 < \mu_4 < 3.4$ and are characteristic of the Black Sea coastal zone. The criterion for the correctness of the approximation is the deviation of the model probability density function from that one calculated from wave measurement data, which is characterized by relative error.

In range $|\xi| < 3$, values of average relative error $\langle \varepsilon(\xi) \rangle$ and its standard deviation $\delta(\xi)$ are small and satisfy condition $|\langle \varepsilon(\xi) \rangle| < 0.05$, $\delta(\xi) < 0.3$. Approximation error $\langle \varepsilon(\xi) \rangle$ has a systematic component which depends on the deviations of the third and fourth statistical moments from the values corresponding to the Gaussian distribution. A semi-empirical relationship has been constructed to take this component into account. The elimination of the systematic component will reduce $\delta(\xi)$, and the approximation accuracy can accordingly be increased by 2–3 times.

REFERENCES

1. Longuet-Higgins, M.S., 1963. The Effect of Non-Linearities on Statistical Distributions in the Theory of Sea Waves. *Journal of Fluid Mechanic*s, 17(3), pp. 459–480. https://doi.org/10.1017/S0022112063001452
2. Hayne, G.S., 1980. Radar Altimeter Mean Return Waveforms from Near-Normal-Incidence Ocean Surface Scattering. *IEEE Transactions on Antennas and Propagation*, 28(5), pp. 687–692. https://doi.org/10.1109/TAP.1980.1142398
3. Kay, S., Hedley, J.D. and Lavender, S., 2009. Sun Glint Correction of High and Low Spatial Resolution Images of Aquatic Scenes: A Review of Methods for Visible and Near-Infrared Wavelengths. *Remote Sensing*, 1(4), pp. 697–730. https://doi.org/10.3390/rs1040697
4. Annenkov, S.Y. and Shrira, V.I., 2014. Evaluation of skewness and kurtosis of wind waves parameterized by JONSWAP spectra. *Journal of Physical Oceanography*, 44(6), pp. 1582–1594. https://doi.org/10.1175/JPO-D-13-0218.1
5. 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. https://doi.org/10.1029/2005JC003343
6. Callahan, P.S. and Rodriguez, E., 2004. Retracking of Jason-1 Data. *Marine Geodesy*, 27(3–4), pp. 391–407. https://doi.org/10.1080/01490410490902098

7. 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. https://doi.org/10.1080/03610926.2020.1833219
8. Lin, W. and Zhang, J.E., 2022. The Valid Regions of Gram–Charlier Densities with High-Order Cumulants. *Journal of Computational and Applied Mathematics*, 407, 113945. https://doi.org/10.1016/j.cam.2021.113945
9. Blinnikov, S. and Moessner, R., 1998. Expansions for nearly Gaussian Distributions. *Astronomy and Astrophysics Supplement Series*, 130(1), pp. 193–205. https://doi.org/10.1051/aas:1998221
10. 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.
11. Zapevalov, A.S. and Ratner, Y.B., 2003. Analytic Model of the Probability Density of Slopes of the Sea Surface. *Physical Oceanography*, 13(1), pp. 1–13. https://doi.org/10.1023/A:1022444703787
12. 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. https://doi.org/10.1175/1520-0426(2003)020<1697:MROTCD>2.0.CO;2
13. Gao, Z. and Sun Z. Liang, S., 2020. Probability Density Function for Wave Elevation Based on Gaussian Mixture Models. *Ocean Engineering*, 213, 107815. https://doi.org/10.1016/j.oceaneng.2020.107815
14. Carreira-Perpinan, M.A., 2000. Mode-Finding for Mixtures of Gaussian Distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), pp. 1318–1323. https://doi.org/10.1109/34.888716
15. 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. https://doi.org/10.1134/S1054661813030024
16. Babanin, A.V. and Polnikov, V.G., 1995. On the Non-Gaussian Nature of Wind Waves. *Physical Oceanography*, 6(3), pp. 241–245. https://doi.org/10.1007/BF02197522
17. Guedes Soares, C., Cherneva, Z. and Antão, E.M., 2003. Characteristics of Abnormal Waves in North Sea Storm Sea States. *Applied Ocean Research*., 25(6), pp. 337–344. https://doi.org/10.1016/j.apor.2004.02.005
18. Jha, A.K. and Winterstein, S.R., 2000. Nonlinear Random Ocean Waves: Prediction and Comparison with Data. In: ASME, 2000. *Proceedings of the 19th International Offshore Mechanics and Arctic Engineering Symposium*. New Orleans, USA. Paper No. 00-6125.
19. 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. https://doi.org/10.22449/1573-160X-2021-4-414-425
20. Toloknov, Yu.N. and Korovushkin, A.I., 2010. The System of Collecting Hydrometeorological Information. In: MHI, 2010. *Monitoring Systems of Environment*. Sevastopol: ECOSI-Gidrofizika. Iss. 13, pp. 50–53 (in Russian).

Submitted 06.07.2023; accepted after review 03.09.2023;
revised 27.12.2023; published 25.03.2024

About the authors:

Aleksandr S. Zapevalov, Chief Research Associate, Marine Hydrophysical Institute of RAS (2 Kapitanskaya St., Sevastopol, 299011, Russian Federation), Dr.Sci. (Phys.-Math.), **Scopus Author ID: 7004433476**, **ResearcherID: V-7880-2017**, **ORCID ID: 0000-0001-9942-2796**, *sevzepter@mail.ru*

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

Contribution of the authors:

Aleksandr S. Zapevalov – task statement, review of literature on the study topic, article preparation

Aleksandr S. Knyazkov – algorithm development and calculation performance, critical analysis of the calculations

The author has read and approved the final manuscript.