ISSN 1573-160X, *PHYSICAL OCEANOGRAPHY*, Vol. 33, Iss. 2, pp. 263-280 (2026)
Original Russian Text © E. V. Zhuk, V. N. Belokopytov, 2026,
published in *Morskoy Gidrofizicheskiy Zhurnal*, 2026, Vol. 42, Iss. 2, pp. 216–234

Original article

Regional Methodology for Reconstructing Thermohaline Structure of the Black Sea Based on Remote Sensing Data Using Regression Relationships

E. V. Zhuk ✉, V. N. Belokopytov

Marine Hydrophysical Institute of RAS, Sevastopol, Russian Federation
✉ *elena.zhuk@mhi-ras.ru*

Abstract
Purpose. The purpose of the work is to develop a regional statistical methodology for deriving the Black Sea thermohaline structure based on satellite data on sea surface temperature and sea level.
Methods and Results. The study included *in-situ* temperature and salinity data, daily anomalies of altimetry sea level (0.125° grid) and sea surface temperature (0.05° grid) at the L4 processing level for 1993–2023, as well as a climatic array of the Black Sea thermohaline fields for 1950–2023. To reproduce the hydrological fields, the method of multiple linear regression was used. It made it possible to determine statistical relationships between the physical parameters of the sea surface layer and the values of temperature and salinity at depths up to 300 m. In order to take into account long-term changes, corrections to modern climatic values and linear trends were applied. Validation of the methodology through comparison of the calculated 30-year reanalysis and the *in-situ* measurement data showed that the total root-mean-square errors of temperature and salinity reconstruction were 0.4 °C and 0.2 psu, respectively, which corresponded to an average normalized error of 0.4, indicating a good result for a statistical method. The errors were highest in the seasonal thermocline and permanent halocline layers, where the amplitude of mesoscale noise increases sharply. The best values of calculation accuracy were observed in the cyclonic gyres in the central sea, while the worst ones were found in the sea periphery (Batumi anticyclonic eddy and Rim Current zone), and, as for salinity, on the north-western shelf. A comparison of the error values of several arrays of the Black Sea thermohaline field reanalyses showed that the accuracy of this method is close to that of the CMEMS arrays calculated using modern hydrodynamic models and complex data assimilation methods. Some indicators relating to salinity show that, despite the absence of an *in-situ* data assimilation algorithm in this version, the regression model even exceeds the accuracy values of the specified reanalyses.
Conclusions. A method for operational diagnosis and reanalysis of the Black Sea 3-dimensional thermohaline structure based on satellite data and using regression relationships has been developed. It provides a sufficiently good level of errors in temperature and salinity reconstruction while minimizing computational resources. The methodology implies the application of regular open-source data, requires no high-performance equipment (which is of practical importance), and can be regarded as a quick and effective instrument for assessing the state of the marine environment. The regression algorithm can also be used in operational data assimilation procedures in forecast models when *in-situ* oceanographic data are absent.

Keywords: Black Sea, thermohaline structure, linear regression, altimetry, reanalysis

Acknowledgments: The study was carried out within the framework of theme of state assignment of FSBSI FRC MHI FNNN-2024-0014.

For citation: Zhuk, E.V. and Belokopytov, V.N., 2026. Regional Methodology for Reconstructing Thermohaline Structure of the Black Sea Based on Remote Sensing Data Using Regression Relationships. *Physical Oceanography*, 33(2), pp. 263-280.

Introduction

Currently, ocean reanalysis arrays, calculated using hydrodynamic models with the assimilation of meteorological information, in situ and remote sensing data, are widely used in oceanology. Their degree of mutual disparity is quite significant [1, 2] and depends on the applied mathematical model, the source of atmospheric forcing information, and the data assimilation method. Traditional estimates of the water thermohaline structure, based solely on primary oceanographic measurements, are highly dependent on interpolation and statistical processing methods under conditions of significant data sparsity in time and space. Nevertheless, statistical arrays remain relevant as estimates that are most closely aligned with direct measurements. They are in demand both in climate research and for data assimilation and validation of results in hydrodynamic models.

Statistical arrays for the World Ocean, such as ISAS [3], EN3, EN4 [4], and CORA [5], are most often constructed using optimal interpolation [1] [6] or the successive correction method [7, 8]. With the development of Earth remote sensing technologies, it has become possible to use satellite data in their creation. Starting from the 1990s, studies [9, 10] emerged focusing on determining regression relationships between dynamic topography, altimetric sea level, and the vertical distribution of temperature and salinity. The objective of this approach is to generate "pseudo-measurements" of temperature and salinity calculated from remote sensing data on sea surface conditions, which allows for the estimation of the thermohaline structure of the water column in regions of the ocean with sparse observational data. The most well-known product in this area is the regularly produced ARMOR3D dataset [2] [11, 12], which combines calculations based on regression relationships and optimal interpolation of Argo profiling float data. According to the classification of the ORA-IP project [1], dedicated to comparing global reanalysis arrays, ARMOR3D is also considered a product based on measurements.

Analysis of the errors in reproducing ocean thermohaline fields in ARMOR3D and in similar methods developed by Marine Hydrophysical Institute (MHI) of the Russian Academy of Sciences has shown that for most regions of the World Ocean, the errors of regression models, given minimal input information and computational speed, are comparable to the accuracy achieved by far more complex and labor-intensive methods. However, in seas significantly isolated from the ocean, such as the Black Sea and the Mediterranean Sea, the errors of this method increase substantially, necessitating the development of regional methodologies.

---

[1] Gandin, L.S., 1963. *Objective Analysis of Meteorological Fields*. Leningrad: Gidrometeoizdat, 287 p. (in Russian).

[2] Godin, E.A., Belokopytov, V.N., Ingerov, A.V., Plastun, T.V., Galkovskaya, L.K., Kasyanenko, T.E., Zhuk, E.V. and Isaeva, E.A., 2019. *Black Sea: Hydrology – 2018:* Database. Moscow: Federal Research Center "Marine Hydrophysical Institute of the Russian Academy of Sciences". State Registration No. RU 2019621008 (in Russian).

For the Black Sea, a method using altimetric information was developed and applied [13, 14] for this purpose. It was based on the concept of the predominance of adiabatic processes in the basin, which determine the variability of thermohaline fields across a wide range of spatiotemporal scales. A "reference level" was identified, and the calculation of thermohaline fields was performed relative to this surface.

The aim of this work is to develop a regional methodology, based on an approach analogous to ARMOR3D, for calculating the thermohaline structure of the Black Sea using remote sensing data, taking into account the specific hydrological structure and temporal variability of the basin. As in [11, 12], the methodology can operate in conjunction with optimal interpolation of in situ observations; however, this study considers only the algorithm based on regression relationships.

Research data and methods

In situ measurements of temperature and salinity, satellite data on sea surface temperature and sea level, as well as climatological thermohaline fields, were used in the present study.

In situ measurements of temperature (T) and salinity (S) for 1993–2023, corresponding to the era of altimetric data availability, were collected from the Oceanographic Data Bank of Marine Hydrophysical Institute (ODB MHI) [2], the SeaDataNet [3] information resources, and the Argo [4] profiling float databases. A total of 26,693 vertical profiles that passed standard oceanographic quality tests were used in the study. In many grid cells of 20' × 30' resolution, fewer than 50 soundings were performed over the 30-year period (Fig. 1). The lowest observational data coverage occurs during the winter months, with measurements completely absent in many areas of the sea. Spatially, the data are predominantly from the southern coast of Crimea, the coast of Novorossiysk, and the coasts of Bulgaria. Such heterogeneity of measurements further indicates that in the Black Sea, as in other seas, the use of in situ data assimilation methods in hydrodynamic models is effective only in certain regions.

The remote sensing data are represented by arrays of sea level anomaly (*SLA*) [5] and sea surface temperature (*SST*) [6] at the L4 processing level. The spatial resolution of the *SLA* array is 0.125° (1/8°) with a temporal resolution of one day. The *SST* array is also provided as daily data, but on a finer grid with a spacing of 0.05° (1/20°).

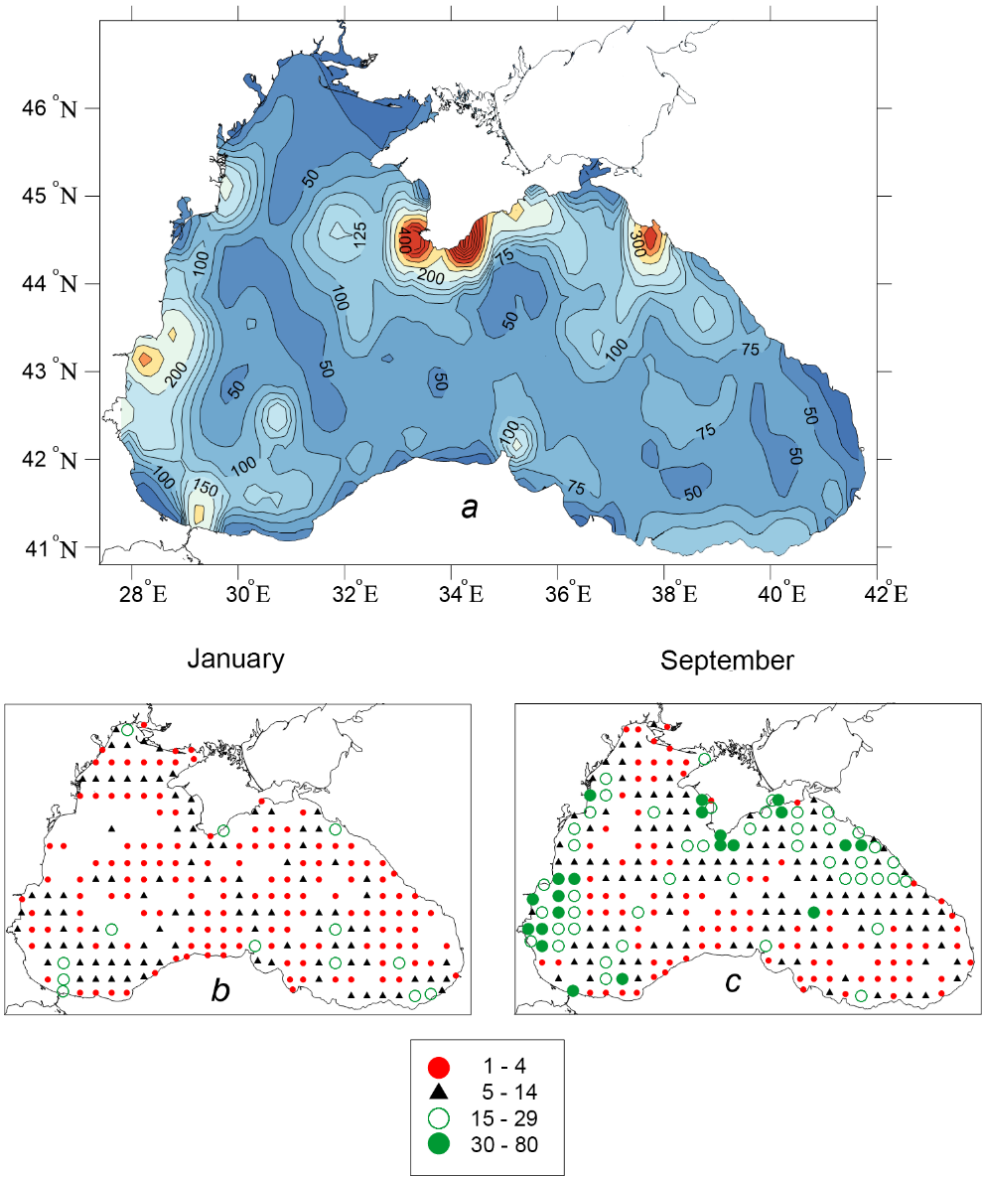
---

[3] SeaDataNet. *Marine Data Access*. [online] Available at: https://cdi.seadatanet.org/search [Accessed: 16 March 2023].

[4] Coriolis. *Data Selection*. [online] Available at: https://www.coriolis.eu.org/Data-Products/Data-selection [Accessed: 29 December 2023].

[5] European Union – Copernicus Marine Service, 2023. *Global Ocean Gridded L4 Sea Surface Heights and Derived Variables Reprocessed (1993–Ongoing)* (SEALEVEL_GLO_PHY_L4_REP_OBSERVATIONS_008_047): [dataset]. https://doi.org/10.48670/moi-00141

[6] European Union – Copernicus Marine Service, 2023. *Global Ocean Sea Surface Temperature (SST) L4 Reprocessed (1981–Ongoing)* (SST_GLO_SST_L4_REP_OBSERVATIONS_010_011): [dataset]. https://doi.org/10.48670/moi-00160



F i g. 1. Number of hydrological stations for 1993–2023: for the sea as a whole (*a*), and in the 20' × 30' grid cells (37 × 39 km) in January (*b*) and September (*c*)

For calculating the hydrological fields, the method of multiple linear regression was used, similarly to [11, 12], establishing statistical relationships between the state of the sea surface layer, described by *SST* and *SLA* data, and the values of *T* and *S* at depths of up to 300 m. These relationships are expressed in the following form:

$$T_{\mathrm{r}}(x,y,z,t) = \alpha(x,y,z)SLA'(x,y,t) + \beta(x,y,z)SST'(x,y,t) + \lambda(x,y,z) + \\ + T_{\mathrm{clm}}(x,y,z,t) + T_{\mathrm{cor}}(z,t) + T_{\mathrm{tr}}(z,t), \quad (1)$$

$$S_{\mathrm{r}}(x,y,z,t) = \gamma(x,y,z)SLA'(x,y,t) + \mu(x,y,z) + S_{\mathrm{clm}}(x,y,z,t) + \\ + S_{\mathrm{cor}}(z,t) + S_{\mathrm{tr}}(z,t), \quad (2)$$

where T_{r} and S_{r} are water temperature and salinity calculated from the regression relationships; t is time; x, y, z are spatial coordinates; $SLA' = SLA - SLA_{\mathrm{clm}} - SLA_{\mathrm{tr}}$ (SLA anomaly); $SST' = SST - SST_{\mathrm{clm}} - SST_{\mathrm{tr}}$ (SST anomaly); SLA_{clm}, SST_{clm} are the climatological daily mean values of SLA and SST; SST_{tr} and SLA_{tr} are corrections for linear trends of SST and SLA in 1993–2023; T_{clm}, S_{clm} are the climatological daily mean values of T and S; T_{cor} and S_{cor} are corrections to the climatological T and S values for the specified period; T_{tr} and S_{tr} are trend corrections for T and S over the same period; α, β, γ are regression coefficients; λ, μ are the intercept terms of the regression at the x and y grid nodes:

$$\alpha(z) = \frac{\mathrm{cov}(SST', SST')\mathrm{cov}\left(SLA', T'(z)\right) - \mathrm{cov}(SLA', SST')\mathrm{cov}\left(SST',\ T'(z)\right)}{\mathrm{cov}(SLA', SLA')\mathrm{cov}(SST',\ SST') - \mathrm{cov}(SLA', SST')^2}, \quad (3)$$

$$\beta(z) = \frac{\mathrm{cov}(SLA',\ SLA')\mathrm{cov}\left(SST', T'(z)\right) - \mathrm{cov}(SLA', SST')\mathrm{cov}\left(SLA', T'(z)\right)}{\mathrm{cov}(SLA',\ SLA')\mathrm{cov}(SST', SST') - \mathrm{cov}(SLA', SST')^2}, \quad (4)$$

$$\gamma(z) = \frac{\mathrm{cov}\left(SLA',\ S'(z)\right)}{\mathrm{cov}(SLA',\ SLA')}, \quad (5)$$

$$\lambda(z) = \overline{T'}(z) - (\alpha(z)\overline{SLA'} + \beta(z)\overline{SST'}), \quad (6)$$

$$\mu(z) = \overline{S'}(z) - \gamma(z)\overline{SLA'}, \quad (7)$$

where cov is covariance; $T'(z) = T(z) - T_{\mathrm{clm}}(z) - T_{\mathrm{cor}}(z) - T_{\mathrm{tr}}(z)$ (water temperature anomaly);

$S'(z) = S(z) - S_{\mathrm{clm}}(z) - S_{\mathrm{cor}}(z) - S_{\mathrm{tr}}(z)$ (salinity anomaly).

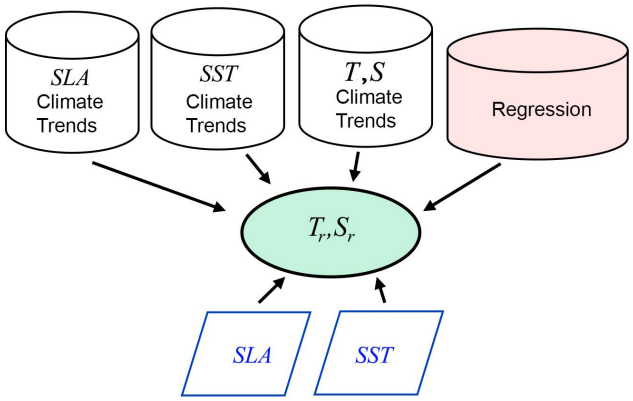
Unlike [11, 12], the correlations with ocean dynamic topography are not taken into account in this methodology; instead, the intercept terms λ and μ, as well as the corrections T_{cor}, T_{tr}, S_{cor}, and S_{tr} are added. The calculations are performed on a finer grid (0.125°), and a different climatological array for T_{clm} and S_{clm} is used. No additional physical assumptions, such as those in [13, 14], are considered in this method. The methodology is as close as possible to purely statistical relationships; in the oceanographic part, regional climatology is used and long-term trends are taken into account.

General block diagrams of the computational process, information flows, and calculation sequence are shown in Figs. 2 and 3.

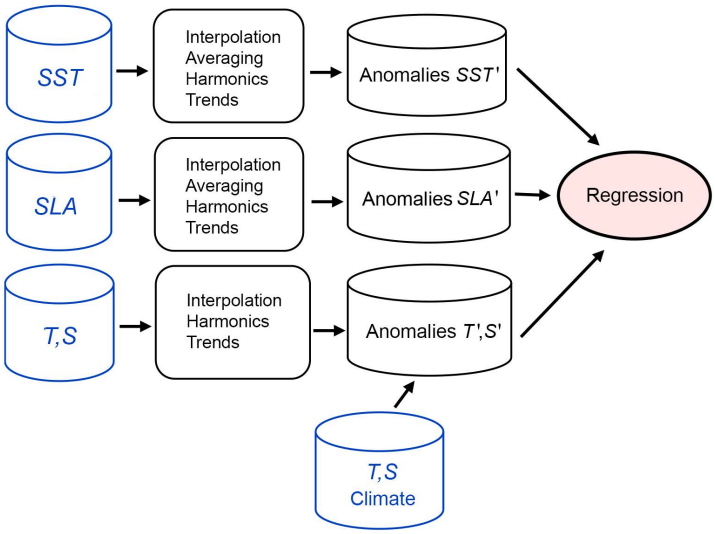
The *Tr* and *Sr* fields are calculated using equations (1) and (2) for a given date (Fig. 2) based on the *SLA* and *SST* fields, using pre-prepared arrays of regression parameters α, β, γ, λ, and μ according to equations (3)–(7), climatological daily fields T_{clm}, S_{clm}, SLA_{clm}, and SST_{clm}, and long-term trends T_{tr}, S_{tr}, SLA_{tr} and SST_{tr}.

The calculation of the regression parameter arrays α, β, γ, λ, and μ (Fig. 3) was carried out in several stages: preparation of the primary database of *T*, *S*, *SLA*, and

SST; calculation of the daily climatological seasonal cycle of T_{clm}, S_{clm}, SLA_{clm}, and SST_{clm}; interpolation of all variables onto a common grid; compilation of joint time series for the variables; and calculation of regression parameters.



F i g. 2. General block diagram for calculating three-dimensional T_r and S_r fields based on *SLA* and *SST* satellite measurements



F i g. 3. General block diagram for calculating regression relationships between the *SLA* and *SST* fields and the *T* and *S* fields

Monthly mean values from the climatological array for T_{clm} and S_{clm} over the 1950–2023 period [15] with corrections T_{cor} and S_{cor} for 1993–2023 were taken as the basis for climatological values; for SST_{clm} and SLA_{clm}, averaged daily mean values of *SST* and *SLA* over the period 1993–2023 were used. For each day of the year, the values were approximated by Fourier series: two harmonics for T_{clm}, S_{clm}, and SST_{clm}, and three harmonics for SLA_{clm}:

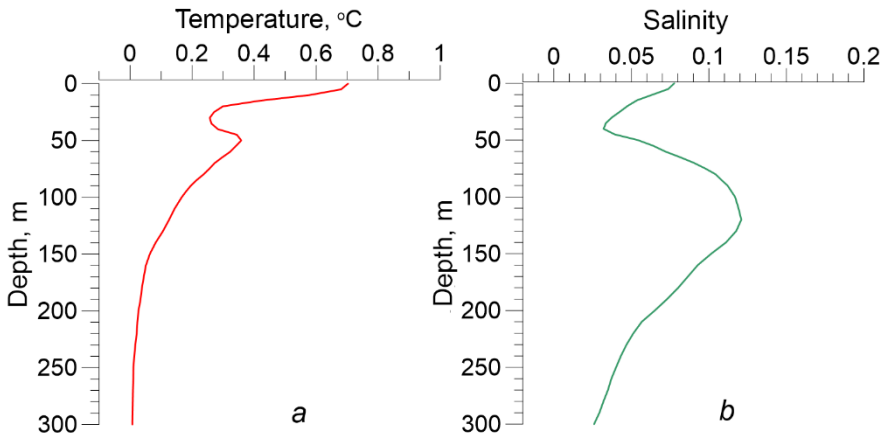
$$y_t = \bar{y} + \sum_{k=1}^{n/2} [A_k \cos\left(\frac{2\pi kt}{n}\right) + B_k \cos\left(\frac{2\pi kt}{n}\right)],$$

where y_t are the approximated values; $\bar{y}$ is the mean value of the series; n is the number of terms in the series; k is the harmonic number;

$$A_k = \frac{2}{n}\sum_{t=1}^{n} [y_t \cos\left(\frac{2\pi kt}{n}\right)];$$

$$B_k = \frac{2}{n}\sum_{t=1}^{n} [y_t \sin\left(\frac{2\pi kt}{n}\right)].$$

All original and climatological *SST* fields, as well as T_{clm} and S_{clm}, were interpolated onto a common grid corresponding to the *SLA* array (0.125°).



F i g. 4. Corrections of climatic temperature and salinity for 1993–2023: T_{cor} (*a*) and S_{cor} (*b*)

Since the regression relationships are calculated for anomalies relative to the mean seasonal cycle, both the coefficients α, β, γ calculated by equations (3)–(5) and the final results T_r and S_r calculated by equations (1) and (2) by summing the calculated anomalies and the climatological seasonal fields depend strongly on the quality of the climatological array. Using a thermohaline climatology for 1993–2023 would be optimal and would not require the introduction of special corrections. However, previous attempts to create such a climatological array were unsuccessful because the amount of hydrological data over the last 30 years was insufficient to reliably extract the seasonal cycle against the background of various types of variability. Therefore, the chosen approach was to adjust the better-observed climatological array for 1950–2023 [15]. The corrections to the 1950–2023 climatology T_{cor} and S_{cor} (Fig. 4) were calculated as the average of the annual mean *T* and *S* anomalies over the 1993–2023 period:

$$T_{\mathrm{cor}}(z) = \frac{\sum_{i=1}^{n} \overline{(T(z) - T_{\mathrm{clm}}(z))}_i}{n},$$

$$S_{\mathrm{cor}}(z) = \frac{\sum_{i=1}^{n} \overline{(S(z) - S_{\mathrm{clm}}(z))}_i}{n},$$

where n is the number of years.

It was also necessary to account for persistent trends in thermohaline characteristics, which in the Black Sea are quite high compared with those in other regions of the World Ocean [15]. The trend corrections T_{tr}, SST_{tr}, and SLA_{tr} for the 1993–2023 (Fig. 5, *a*, *c*, *d*, and the formulas for calculating the T_{tr}, S_{tr} and SST_{tr}, and SLA_{tr} trend corrections) were determined from the coefficients of linear trends calculated using the least squares method based on annual mean anomalies of *T*, *SST*, and *SLA*. The S_{tr} corrections for salinity were initially also approximated by linear trends, but to reduce errors they were replaced by third-degree polynomials (Fig. 5, *b* and the formulas for calculating the T_{tr} and S_{tr} trend corrections). This is due to the fact that, although the overall increase in Black Sea salinity over the last 30 years can be approximated by a linear relationship, in the upper layer of the sea at the beginning of the period a decrease occurred, followed by a sharp increase in salinity [15].

The formulas for calculating the T_{tr} and S_{tr} trend corrections for 1993–2023 are given below:

– at the 0 m horizon:

$T_{\mathrm{tr}} = 0.1 + 0.0466667x - 0.8,$

$S_{\mathrm{tr}} = 435342.9375 - 646.1815746x + 0.3196859612x^2 - 5.271547356 \cdot 10^{-5} x^3;$

– at the 50 m horizon:

$T_{\mathrm{tr}} = -0.55 + 0.045\,x - 0.125,$

$S_{\mathrm{tr}} = 53652.61413 - 76.7578727x + 0.03652614647x^2 - 5.780193564 \cdot 10^{-6} x^3;$

– at the 75 m horizon:

$T_{\mathrm{tr}} = -0.4 + 0.03333(x - 1993) - 0.1,$

$S_{\mathrm{tr}} = -120843.1024 + 182.4786739x - 0.09184225926x^2 + 1.54068499 \cdot 10^{-5} x^3;$

– at the 100 m horizon:

$T_{\mathrm{tr}} = -0.3 + 0.02333(x - 1993) - 0.05,$

$S_{\mathrm{tr}} = -18404.67283 + 28.63593108x - 0.01483085264x^2 + 2.557031364 \cdot 10^{-6} x^3;$

– at the 150 m horizon:

$T_{\mathrm{tr}} = -0.13 + 0.01033(x - 1993) - 0.025,$

$S_{\mathrm{tr}} = 15457.18 - 22.62038x + 0.01102795851x^2 - 1.79102376 \cdot 10^{-6} x^3;$

– at the 200 m horizon:

$T_{\mathrm{tr}} = -0.00433 + 0.0035(x - 1993) - 0.015,$

$S_{\mathrm{tr}} = 17552.31449 - 25.94658379x + 0.01278281352x^2 - 2.098795106 \cdot 10^{-6} x^3;$

– at the 250 m horizon:

$T_{\mathrm{tr}} = -0.03 + 0.003167(x - 1993) - 0.0175,$

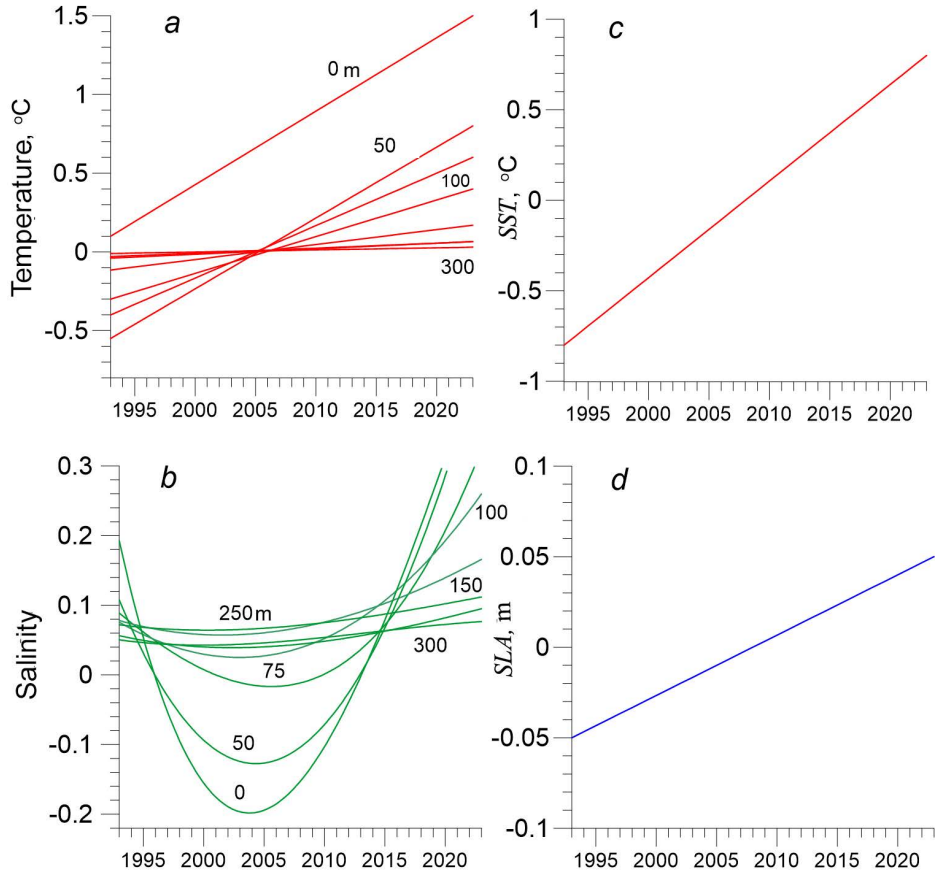
$S_{\mathrm{tr}} = 14678.00879 - 21.74886849x + 0.01074044889x^2 - 1.767750388 \cdot 10^{-6} x^3;$

– at the 300 m horizon:

$T_{\mathrm{tr}} = -0.015 + 0.001833(x - 1993) - 0.0125,$

$S_{tr} = 27686.98886 - 41.25085665x + 0.0204856215x^2 - 3.390964865 \cdot 10^{-6} x^3$,
where x is a year.

The SST_{tr}, SLA_{tr} trend corrections for 1993–2023 are as follows:
$SST_{tr} = -0.8 + 0.05333(y - 1993)$,
$SLA_{tr} = -0.05 + 0.00333(y - 1993)$,
where y is a year.



F i g. 5. Trend corrections for 1993–2023: T_{tr} (*a*), S_{tr} (b), SST_{tr} (*c*) and SLA_{tr} (*d*)

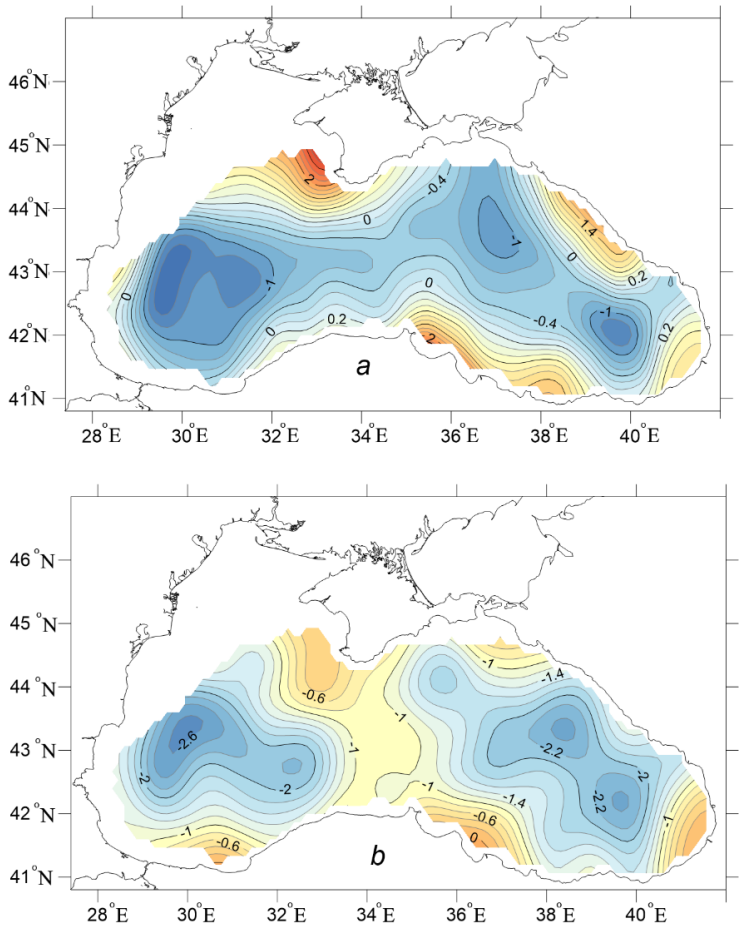
As a result of calculations using the combined 30-year time series of T, S, SST, and SLA anomalies, five three-dimensional arrays of regression coefficients were obtained on a 0.125° grid. Together with the climatological arrays T_{clm}, S_{clm}, SLA_{clm} and SST_{clm}, they form the basis of the computational algorithm.

Discussion of the results

It should be emphasized that after filtering out the seasonal cycle and trends, the characteristics of the correlation structure of thermohaline fields pertain to the range of synoptic and interannual variability. Correlations associated with

the seasonal cycle are not considered here, since the climatological fields are used directly in the calculations.

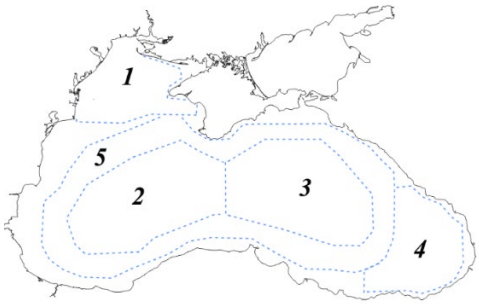
In the spatial distribution of the regression coefficients α, β, and γ, regional differences between the deep-sea part and the periphery of the sea are clearly visible. With depth, especially in the upper part of the pycnocline, the influence of the basin-scale circulation becomes more evident (Fig. 6). In the zone of the Black Sea Rim Current (RC), the correlation and regression coefficients cross zero and change sign.



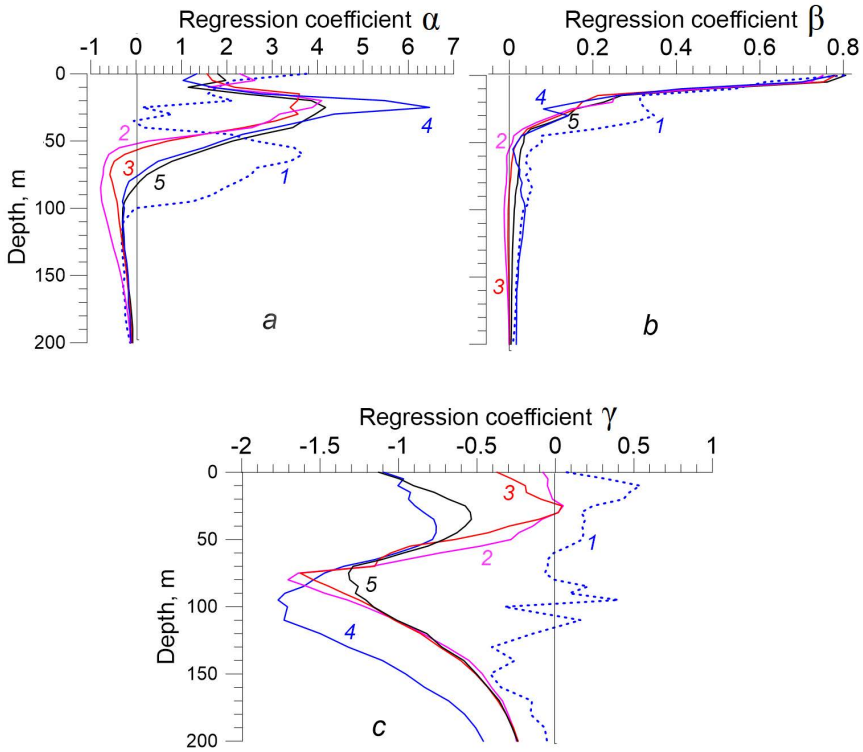
F i g. 6. Horizontal distribution of regression coefficients α (*a*) and γ (*b*) at the 75 m depth

The vertical distribution of the regression coefficients, which is largely related to the general water circulation (Figs. 7, 8), also exhibits distinct patterns. The β coefficients for water temperature are almost entirely positively correlated with *SST* throughout the layer and rapidly decrease with depth. The γ coefficients for salinity are negatively correlated with *SLA* throughout the water column, except on the shallow northwestern shelf. The α coefficients for water temperature show

a more complex distribution, characterized by sign changes in the main pycnocline. The γ coefficients for salinity have a similar curve shape, but without a sign change; instead, they exhibit minima within the pycnocline. These features of the curves reflect the influence of vertical motions associated with the intensity of the basin-scale circulation. The shape of the curves also depends on differences in the vertical position of the pycnocline in the central and peripheral parts of the sea.



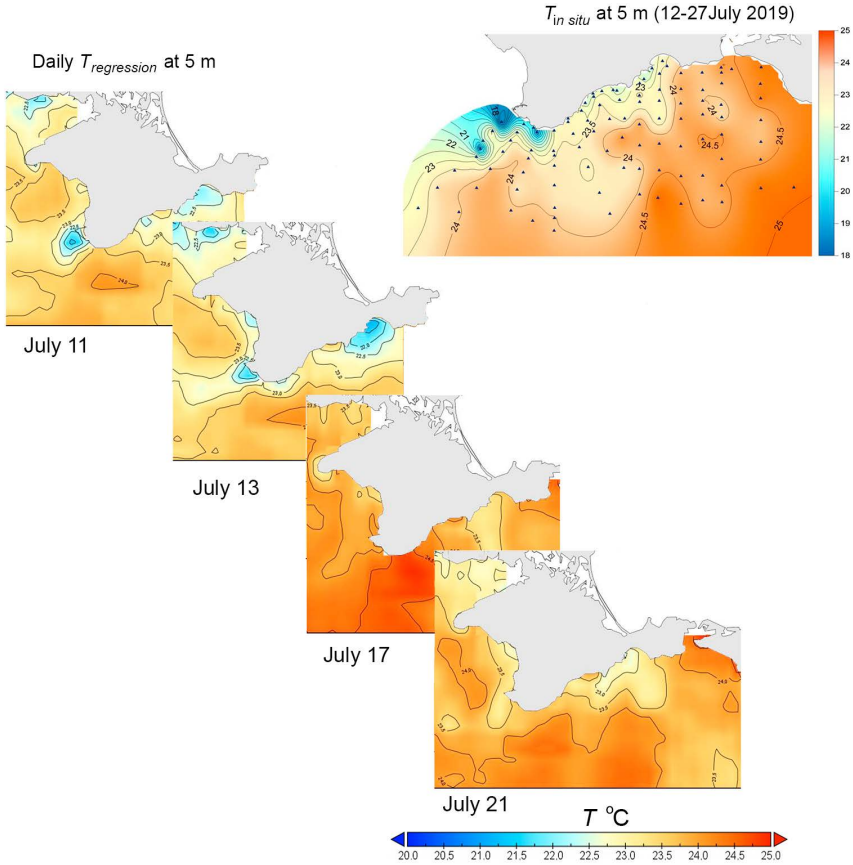
F i g. 7. Scheme of the Black Sea regions used for analyzing regression relationships and errors in reproducing the thermohaline structure: *1* – northwestern shelf; *2* – Western cyclonic gyre; *3* – Eastern cyclonic gyre; *4* – Batumi anticyclonic eddy; and *5* – Rim Current



F i g. 8. Vertical distribution of the regression coefficients α (*a*), β (*b*), and γ (*c*) by Black Sea region (the numerals on the curves correspond to those shown in Fig. 7)

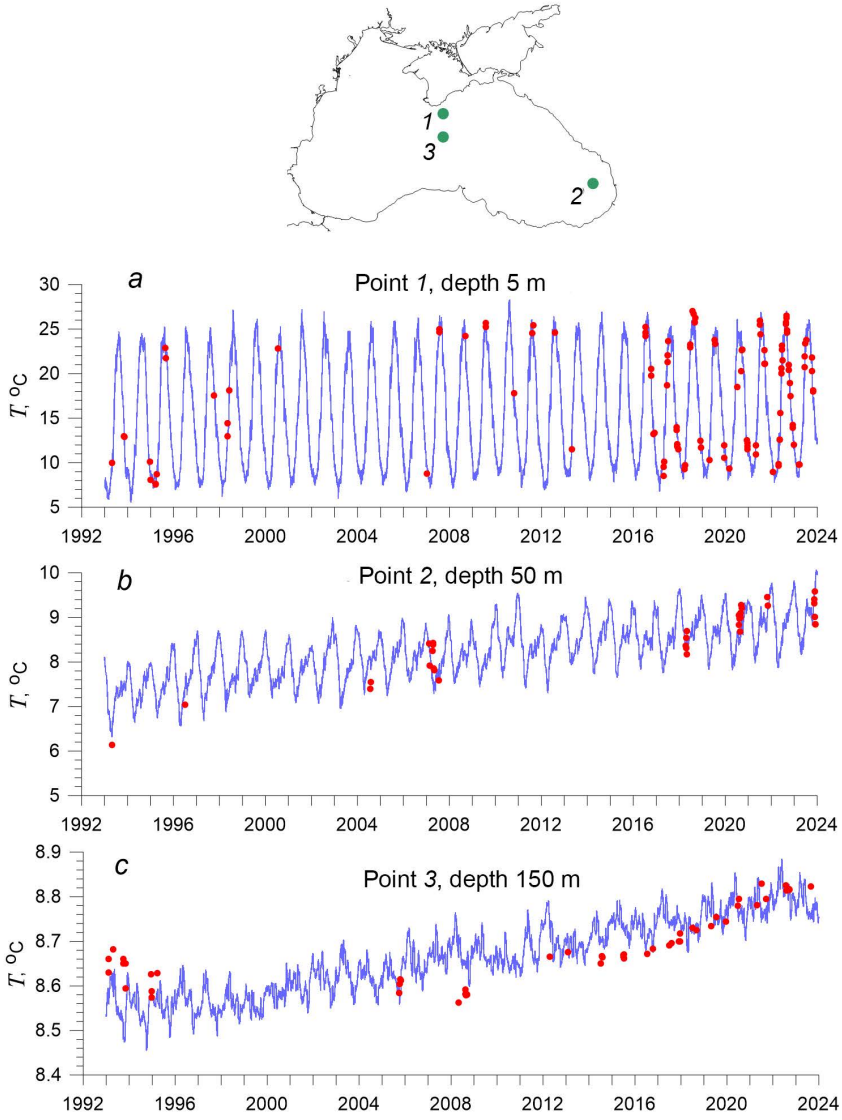
To assess the quality of the simulated thermohaline structure of the Black Sea, a reanalysis array for 1993–2023 was calculated. The following are examples of water temperature reconstruction.

Fig. 9 shows the simulation of coastal upwelling off the southern coast of Crimea during a shipboard hydrographic survey in July 2019. The vessel sailed along the coast on transects for 15 days, during which the upwelling developed rapidly and subsided quickly. Comparing the corresponding reanalysis maps with the measurement data, taking into account the dates and the vessel's position, it can be concluded that the simulation of this upwelling is quite satisfactory.



F i g. 9. Surface-layer water temperature (°C) during the upwelling period off the southern coast of Crimea in July 2019, based on measurement data obtained during the 108th cruise of the R/V *Professor Vodyanitsky* and on the reanalysis array of the regression model

Fig. 10 shows examples of 30-year water temperature time series obtained from the reanalysis array and from measurement data at three points characterizing the surface layer in the RC (point *1*), the cold intermediate layer in the Batumi anticyclonic eddy (point *2*), and the main pycnocline in the central part of the sea (point *3*). The qualitative simulation of seasonal and interdecadal variability is also satisfactory.



F i g. 10. Water temperature time series based on the regression model reanalysis (blue curves) at the 0.125° grid nodes with the following coordinates: point *1* – 44.17°N, 34.25°E (*a*); point *2* – 42.17°N, 40.75°E (*b*); and point *3* – 43.50°N, 34.25°E (*c*). Red symbols denote *in situ* measurement data in the vicinity of the selected grid nodes ±10′ (18 km)

To obtain quantitative characteristics of the errors in the water thermohaline structure simulation, the root-mean-square error S_q, the mean error S_m, and the normalized error S_n were calculated:

$$S_q = \sqrt{\frac{\sum_{i=1}^{n}(x_r - x_i)^2}{n-m}},$$

$$S_m = \frac{\sum_{i=1}^{n}(x_r - x_i)}{n-m},$$

$$S_{\mathrm{n}} = \frac{\sigma_x}{S_{\mathrm{q}}},$$

where x are the actual values; x_{r} are the values calculated from the regression relationships; n is the number of terms in the series; m is the number of degrees of freedom (3 for temperature, 2 for salinity); σ_x is the standard deviation of natural variability.

Fig. 11 shows the vertical distribution of the mean simulation errors for temperature and salinity across different regions of the Black Sea. The root-mean-square error S_{q} for water temperature in the 0–300 m layer is 0.4 °C; in the seasonal thermocline layer, the S_{q} errors are maximal and reach 2 °C (Fig. 11, *a*). The mean S_{q} for salinity in the 0–300 m layer is 0.2; the errors are maximal in the main halocline and reach 0.4, with their depth position differing between the central and peripheral parts of the sea (Fig. 11, *b*). The location of the error maximum in the pycnocline layers, where the amplitude of mesoscale noise increases sharply, is typical for all reanalysis arrays regardless of the calculation technology.

The practical significance of the error level can be assessed using the normalized error S_{n}. Theoretically, the calculation method is effective if the normalized error $S_{\mathrm{n}} < 1$, i.e., its skill is higher than that of a climatological forecast. According to [7], acceptable criteria in hydrometeorological practice range from $S_{\mathrm{n}} = 0.67$ for short-term forecasts to $S_{\mathrm{n}} = 0.8$ for long-term forecasts, although these estimates are rather conditional since the natural variability indicators σ_x themselves are determined with large uncertainties.

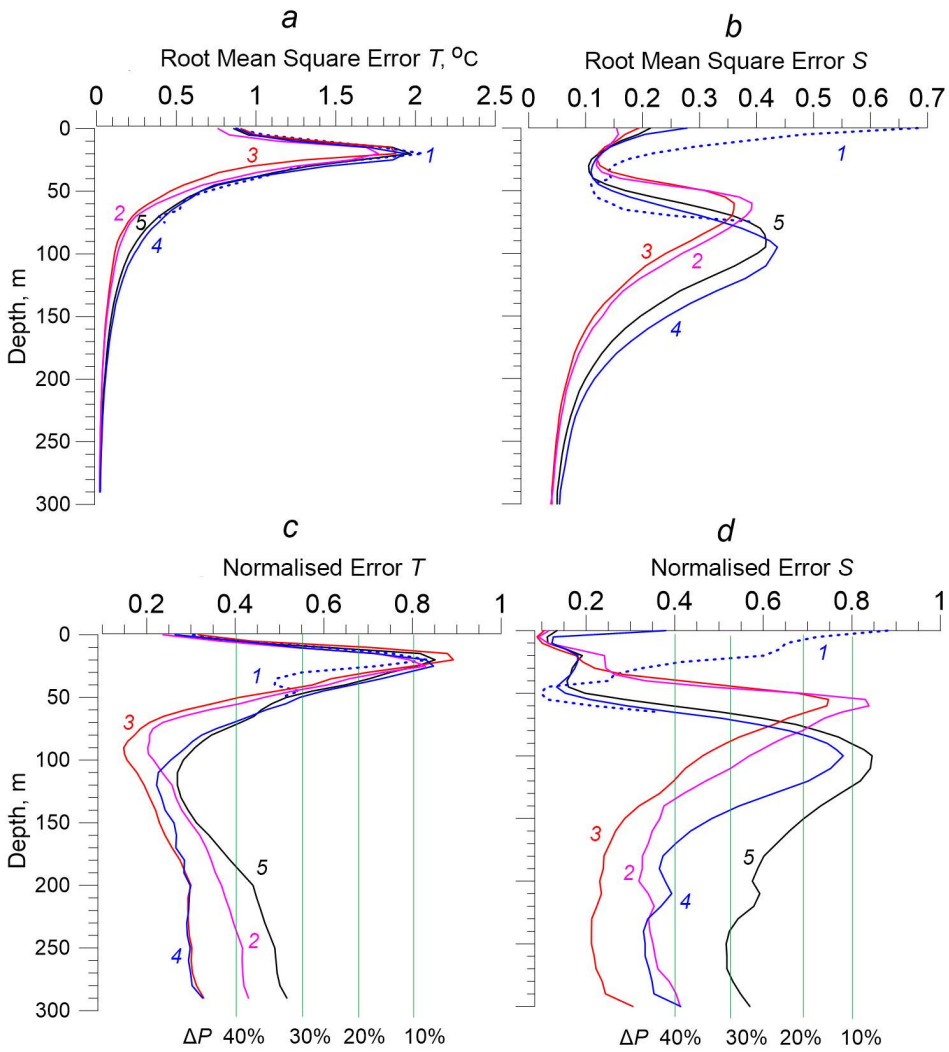
According to this criterion, water temperature calculated by this method is simulated well in the surface layer down to 10 m and in deeper layers below 50 m, with the method's skill improvement relative to the climatological forecast exceeding 40% ($S_{\mathrm{n}} < 0.4$) (Fig. 11, *c*). The mean normalized error S_{n} for water temperature in the 0–300 m layer is 0.4. For salinity, a skill improvement of more than 40% ($S_{\mathrm{n}} < 0.4$) is achieved in layers down to 50 m and below 150 m (Fig. 11, *d*). For both temperature and salinity, the best calculation accuracy indicators are found in the cyclonic gyres in the central part of the sea, while the worst are found on the periphery of the sea, in the Batumi anticyclone, and in the RC zone. Salinity is also poorly simulated on the shallow northwestern shelf. The mean normalized error S_{n} for salinity in the 0–300 m layer, like that for water temperature, is 0.4.

In further development of the regression model, the effectiveness of using satellite sea surface salinity data to create a two-parameter dependence, analogous to *SST* for water temperature, should be assessed.

The validation results presented here pertain to the entire dataset for 1993–2023. Independent test samples are not shown here in order to allow comparison with other studies where validation results are presented for the entire time series. Figs. 12 and 13 show the averaged errors of several reanalysis arrays of thermohaline fields

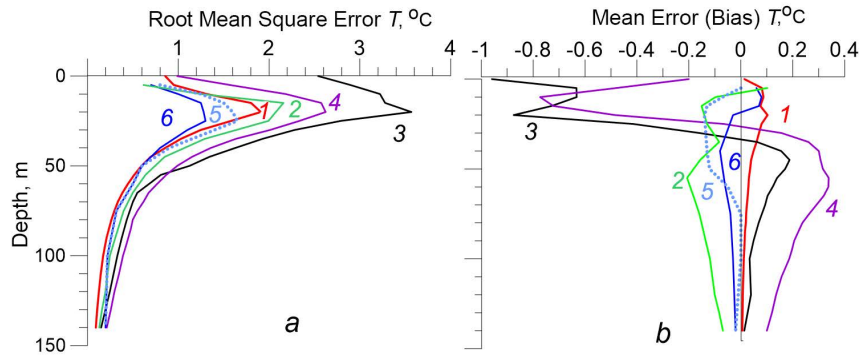
[7] Rosgidromet, 2011. [*Guidelines for Forecast Services. Section 3. Part III. Marine Hydrological Forecast Service: Guideline RD 52.27.759-2011*]. Moscow: TRIADA LTD, 189 p. (in Russian).

for the Black Sea: MHI, ARMOR3D, and CMEMS [8]. Overall, the accuracy of our method (red curves) is comparable to that of the CMEMS arrays, which are calculated using modern models with sophisticated assimilation techniques. For certain metrics, such as S_q for salinity in the 0–50 m layer and S_m for salinity throughout the entire layer, the regression model even outperforms them, despite the absence of an *in situ* data assimilation block in this version.

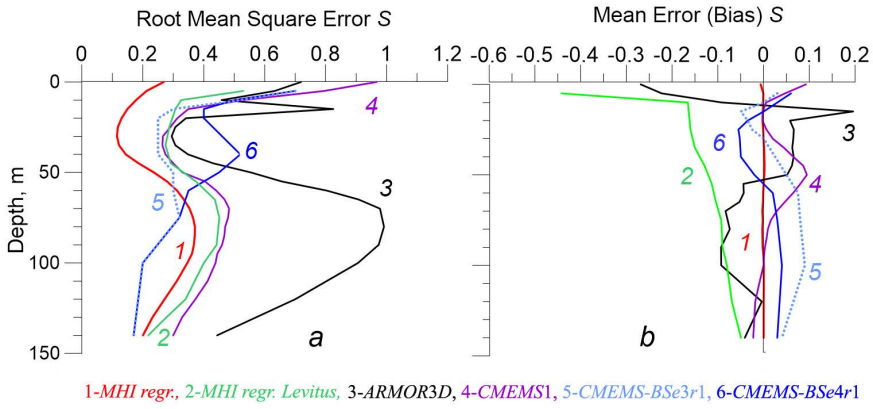


F i g. 11. Root-mean-square errors S_q of water temperature (*a*) and salinity (*b*), and normalized errors S_n of water temperature (*c*) and salinity (*d*) for different Black Sea regions (the numerals on the curves correspond to those shown in Fig. 7); ΔP is the increase in method reliability relative to a climatic forecast

[8] CMEMS, 2023. *Black Sea Physics Reanalysis (1993–2022)* (BLKSEA_MULTIYEAR_PHY_007_004): [dataset]. CMEMS Black Sea Production Centre; Copernicus Marine Service. https://doi.org/10.25423/CMEMS/BLKSEA_MULTIYEAR_PHY_007_004



F i g. 12. Root-mean-square S_q (*a*) and mean S_m (*b*) errors of water temperature for different reanalysis arrays of the Black Sea water thermohaline structure



F i g. 13. Root-mean-square error S_q (*a*) and mean error S_m (*b*) of water salinity for different reanalysis arrays of the Black Sea thermohaline structure

Conclusion

A methodology for the operational diagnosis and reanalysis of the three-dimensional thermohaline structure of the Black Sea waters using satellite data, based on mathematical statistics methods, has been developed. In situ measurements of temperature and salinity for 1993–2023 from the Oceanographic Data Bank of Marine Hydrophysical Institute of the Russian Academy of Sciences, the SeaDataNet information resources, and the Argo profiling float database were used. The remote sensing data are represented by *SLA* (0.125°) and *SST* (0.05°) arrays at the L4 processing level.

The methodology accounts for the specific hydrological structure of the Black Sea and its climatic variability, providing temperature and salinity reconstruction error levels comparable to those of the best hydrodynamic models, such as NEMO. At the same time, it does not require the application of complex data assimilation algorithms like 3DVar, 4DVar, or the Kalman filter; instead, the methodology uses

regularly available, widely accessible information from satellite altimeters and infrared radiometers. In the absence of hydrological measurement data, the regression algorithm can also be applied within assimilation blocks of operational forecast models.

The spatial distribution of the regression relationships between the sea surface height and the vertical thermohaline structure of the waters shows a significant difference between the deep-sea part and the periphery of the sea. With depth, especially in the upper part of the pycnocline, the influence of the basin-scale circulation becomes more evident. The vertical distribution of the regression coefficients reflects the influence of vertical motions associated with the intensity of the basin-scale circulation.

Accounting for long-term changes in the water thermohaline structure in the form of adjustments to climatological values for the contemporary period, as well as linear and polynomial trends, reduced the overall root-mean-square error level for temperature to 0.4 °C and for salinity to 0.2, corresponding to a mean normalized error of 0.4, which is a good indicator for a statistical method. The simulation errors of the thermohaline fields are maximal in layers with increased mesoscale noise – the seasonal thermocline and the permanent halocline. The best calculation accuracy indicators are found in the cyclonic gyres in the central part of the sea, while the worst are found on the periphery of the sea, in the Batumi anticyclone, in the Rim Current zone, and on the northwestern shelf.

The methodology minimizes the use of computational resources, which has practical significance. Its implementation as a ready-made software product that does not require high-performance hardware can be considered a fast and effective method for assessing the state of the marine environment.

To improve the accuracy of salinity simulation, future work should estimate the effectiveness of using satellite sea surface salinity data to develop a two-parameter dependence.

REFERENCES

1. Balmaseda, M.A., Hernandez, F., Storto, A., Palmer, M.D., Alves, O., Shi, L., Smith, G.C., Toyoda, T., Valdivieso, M. [et al.], 2015. The Ocean Reanalyses Intercomparison Project (ORA-IP). *Journal of Operational Oceanography*, 8(suppl. 1), pp. s80-s97. https://doi.org/10.1080/1755876X.2015.1022329
2. Storto, A., Alvera-Azcárate, A., Balmaseda, M.A., Barth, A., Chevallier, M., Counillon, F., Domingues, C.M., Drevillon, M., Drillet, Y. [et al.], 2019. Ocean Reanalyses: Recent Advances and Unsolved Challenges. *Frontiers in Marine Science*, 6, 418. https://doi.org/10.3389/fmars.2019.00418
3. Gaillard, F., Reynaud, T., Thierry, V., Kolodziejczyk, N. and Von Schuckmann, K., 2016. In Situ-Based Reanalysis of the Global Ocean Temperature and Salinity with ISAS: Variability of the Heat Content and Steric Height. *Journal of Climate*, 29(4), pp. 1305-1323. https://doi.org/10.1175/JCLI-D-15-0028.1
4. Good, S.A., Martin, M.J. and Rayner, N.A., 2013. EN4: Quality Controlled Ocean Temperature and Salinity Profiles and Monthly Objective Analyses with Uncertainty Estimates. *Journal of Geophysical Research: Oceans*, 118(12), pp. 6704-6716. https://doi.org.10.1002/2013JC009067
5. Cabanes, C., Grouazel, A., Von Schuckmann, K., Hamon, M., Turpin, V., Coatanoan, C., Paris, F., Guinehut, S., Boone, C. [et al.], 2012. The CORA Dataset: Validation and Diagnostics of

In-Situ Ocean Temperature and Salinity Measurements. *Ocean Science*, 9(1), pp. 1273-1312. https://doi.org/10.5194/os-9-1-2013

6. Bretherton, F.P., Davis, R.E. and Fandry, C.B., 1976. A Technique for Objective Analysis and Design of Oceanographic Experiments Applied to MODE-73. *Deep-Sea Research*, 23, pp. 559-582. https://doi.org/10.1016/0011-7471(76)90801-1
7. Locarnini, R.A., Mishonov, A.V., Baranova, O.K., Boyer, T.P., Zweng, M.M., Garcia, H.E., Reagan, J.R., Seidov, D., Weathers, K.W. [et al.], 2019. *NOAA Atlas NESDIS 81. World Ocean Atlas 2018. Volume 1: Temperature*. USA: NOAA, 52 p.
8. Zweng, M.M., Reagan, J.R., Seidov, D., Boyer, T.P., Locarnini, R.A., Garcia, H.E., Mishonov, A.V., Baranova, O.K., Weathers, K.W. [et al.], 2019. *NOAA Atlas NESDIS 82. World Ocean Atlas 2018. Volume 2: Salinity*. USA: NOAA, 50 p.
9. Carnes, M.R., Mitchell, J.L. and De Witt, P.W., 1990. Synthetic Temperature Profiles Derived from Geosat Altimetry: Comparison with Air-Dropped Expendable Bathythermograph Profiles. *Journal of Geophysical Research: Oceans*, 95(C10), pp. 17979-17992. https://doi.org/10.1029/JC095iC10p17979
10. Gilson, J., Roemmich, D., Cornuelle, B. and Fu, L.-L., 1998. Relationship of TOPEX/Poseidon Altimetric Height to Steric Height and Circulation in the North Pacific. *Journal of Geophysical Research: Oceans*, 103(C12), pp. 27947-27966. https://doi.org/ 10.1029/98JC01680
11. Guinehut, S., Le Traon, P.Y., Larnicol, G. and Philipps, S., 2004. Combining Argo and Remote-Sensing Data to Estimate the Ocean Three-Dimensional Temperature Fields – A First Approach Based on Simulated Observations. *Journal of Marine Systems*, 46(1–4), pp. 85-98. https://doi.org/10.1016/j.jmarsys.2003.11.022
12. Guinehut, S., Dhomps, A.-L., Larnicol, G. and Le Traon, P.Y., 2012. High Resolution 3-D Temperature and Salinity Fields Derived from in Situ and Satellite Observations. *Ocean Science*, 8(5), pp. 845-857. https://doi.org/10.5194/os-8-845-2012
13. Korotaev, G.K., Lishaev, P.N. and Knysh, V.V., 2015. Technique of the Black Sea Temperature and Salinity Measurement Data Analysis Using Dynamic Altimetry Level. *Physical Oceanography*, (2), pp. 24-38. https://doi.org/10.22449/1573-160X-2015-2-24-38
14. Korotaev, G.K., Knysh, V.V., Lishaev, P.N. and Demyshev, S.G., 2018. Application of the Adaptive Statistics Method for Reanalysis of the Black Sea Fields Including Assimilation of the Temperature and Salinity Pseudo-Measurements in the Model. *Physical Oceanography*, 25(1), pp. 36-51. https://doi.org/10.22449/1573-160X-2018-1-36-51
15. Belokopytov, V. and Zhuk, E., 2024. Climatic Variability of the Black Sea Thermohaline Characteristics (1950–2023). *Physical Oceanography*, 31(6), pp. 788-801.

Submitted 18.08.2025; approved after review 25.09.2025;
accepted for publication 28.01.2026.

About the authors:

Elena V. Zhuk, Junior Researcher, Marine Hydrophysical Institute of RAS (2 Kapitanskaya Str., Sevastopol, 299011, Russian Federation), **SPIN-code: 3814-6300**, **ORCID ID: 0000-0002-4263-7734**, **WoS ResearcherID: JCD-8660-2023**, **Scopus Author ID: 57191412660**, elena.zhuk@mhi-ras.ru

Vladimir N. Belokopytov, Leading Researcher, Head of Oceanography Department, Marine Hydrophysical Institute of RAS (2 Kapitanskaya Str., Sevastopol, 299011, Russian Federation), DSc. (Geogr.), **WoS ResearcherID: ABA-1230-2020**, **ORCID ID: 0000-0003-4699-9588**, **Scopus Author ID: 6602381894**, **SPIN-code: 5697-5700**, belokopytov.vn@mhi-ras.ru

Contribution of the co-authors:

Elena V. Zhuk – data collecting and processing; preparation and carrying out calculations; visualization and analysis of the results, discussion and conclusions

Vladimir N. Belokopytov – methods development; qualitative and quantitative analysis of research results, discussion and conclusions

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