

Data analysis for variational assimilation of the surface temperature of the Black and Azov Seas

N. B. Zakharova¹ and E. I. Parmuzin¹

Received 5 October 2021; accepted 28 October 2021; published 29 November 2021.

The paper considers sensing data on the state of the Black and Azov Seas, provided by the shared use center “IKI-Monitoring” and their application in the data assimilation problems. Observation data on the sea surface temperature are obtained from various satellites (Aqua, Terra, SNPP) and instruments (MODIS, VIIRS), measured at different moments and received irregularly and often cover only a part of the investigated water area due to weather conditions and the characteristics of measuring instruments. The paper describes the specifics of the data obtained, the difficulties that had to be faced in the data processing for their correct use in the problems of mathematical modeling of the sea dynamics. An analysis of observation data was carried out which showed the presence of errors in the data. An algorithm is proposed based on the determination of weight coefficients characterizing the proximity of observation data to known verified “reference” values. The weight coefficients are constructed taking into account the received fields of observation data and an additional set of average daily data from the European Copernicus Marine Service. The calculated matrix of weight coefficients is used in the algorithm of variational assimilation of observation data for the numerical model of thermodynamics of the Black and Azov Seas. The results of numerical experiments on variational assimilation of observational data using the constructed weight coefficients are presented. The results of assimilation are compared with near-real time *in situ* quality control observations. **KEYWORDS:** Data analysis; observation data; data errors; variational assimilation; sea surface temperature; optimality system.

Citation: Zakharova, N. B. and E. I. Parmuzin (2021), Data analysis for variational assimilation of the surface temperature of the Black and Azov Seas, *Russ. J. Earth. Sci.*, 21, ES6002, doi:10.2205/2021ES000779.

Introduction

Mathematical models in geophysics are used both for modeling natural processes and for analyzing the state of the studied environment. The relevance of environmental monitoring tasks is also due to the identified process of global warming, attracting the interest of the international community and researchers from around the world to monitoring the state of the Earth. For example, the Ocean State

¹Marchuk Institute of Numerical Mathematics RAS, Moscow, Russia

Report [*von Schuckmann et al., 2021*] highlights the trend of global ocean warming. Ocean heat content is an important quantitative variable in the Earth’s climate system, and provides a unique measure of the current status and prospects of global warming.

In order to more accurately reproduce actual physical processes, the models are corrected or verified with observation data. There are many systems that provide a large amount of observation data on the state of the environment, but in practical implementation there are still areas not covered with observation data or are known with some error [*Zakharova, 2016*]. Therefore, in problems of mathematical modeling this leads to formulation

and solving the problems with incomplete information. And at the same time the question arises of how to correctly use data from various sources. The problems of data processing and mathematical modeling of the environment state are combined in the problems of data assimilation in numerical models. There are different methods for combining mathematical models, observation data and a priori information. In the paper a variational data assimilation is used to solve this problems, based on the theory of optimal control and the theory of adjoint equations. Other methods of data assimilation can be used, such as Kalman ensemble filtering, but in any case, research shows [Stepanov *et al.*, 2021] that degree of improving model fields in the process of data assimilation is highly dependent on the structure of data at the input of the assimilation procedure.

In this paper the analysis of observation data from various sources (instantaneous data from satellites, average daily data from the Copernicus Marine Service, station data) are carried out. Based on the analysis of the data, matrices of weight coefficients are constructed, taking into account the reliability and accuracy of the incoming information. These matrices are used in the algorithm for variational data assimilation on sea surface temperature (SST), which is included in the INMOM 3D hydrothermodynamics model developed at the Marchuk Institute of Numerical Mathematics of the Russian Academy of Sciences (INM RAS) [Zalesny *et al.*, 2012]. To illustrate how the model works with the algorithm of variational data assimilation, the results of numerical calculations for the water area of the Black and Azov Seas are presented.

Data analysis

The data from “IKI-Monitoring” shared use center developed at Space Research Institute of the Russian Academy of Sciences [Proshin *et al.*, 2020] are used in the research. The study is carried out on sea surface temperature of the Black and Azov Seas data from various satellites: Aqua with the Moderate Resolution Imaging Spectroradiometer (MODIS), Terra (MODIS), SNPP with the Visible and Infrared Imager/Radiometer Suite (VIIRS) instrument. A detailed description of the data ob-

tained can be found on the NASA website www.oceandata.sci.gsfc.nasa.gov. Images from different satellites are received at different moments; on average, 6–8 surface temperature fields are received per day, partially or completely covering the water area, depending on the season and cloud cover.

However, when processing the data, it turned out that the values of data from different sources arriving at close points in time can differ significantly from each other. Thus, a comparison of SST data from the instruments of the Aqua (MODIS) and SNPP (VIIRS) satellites for September 2019 (the SST during this period of the year on average varies from 15°C to 22°C) showed that in most of the water area the difference in values, measured by different instruments, is kept within 0.5°C, however, in some areas, the difference in values can reach 1.5°C, which is a rather large indicator. The relative error reaches 6–10%. It can be explained by the fact that algorithms for recalculating the brightness temperatures into sea surface temperature for each device (MODIS and VIIRS) are different. For each satellite image the value of SST depends on the angle of incidence of the sun’s rays and the height of the sun above the horizon.

Assimilation in a numerical model of data from various sources that are not coordinated with each other can lead to jumps in the numerical solution, energy and flux when assimilating the data in its pure form. The presence of additional algorithms to account for the dispersion in data from different sources can significantly improve the accuracy of the calculations. So, one of the options for solving the problem of reconciling data from different sources with each other may be the assimilation of data with the addition of some weight coefficients determined by the level of confidence in the data. However, in order to calculate such weights, some of the data must be taken as the true or most reliable data.

Since the data from the Aqua and Terra satellites are equipped with identical MODIS instruments for measuring SST the further study will be carried out in relation to only two datasets: Aqua (MODIS) and SNPP (VIIRS).

The data of the European Copernicus Marine Service www.marine.copernicus.eu were selected as data for assessing the accuracy of the obtained data. The Service received the average daily observation data on the SST of the Black

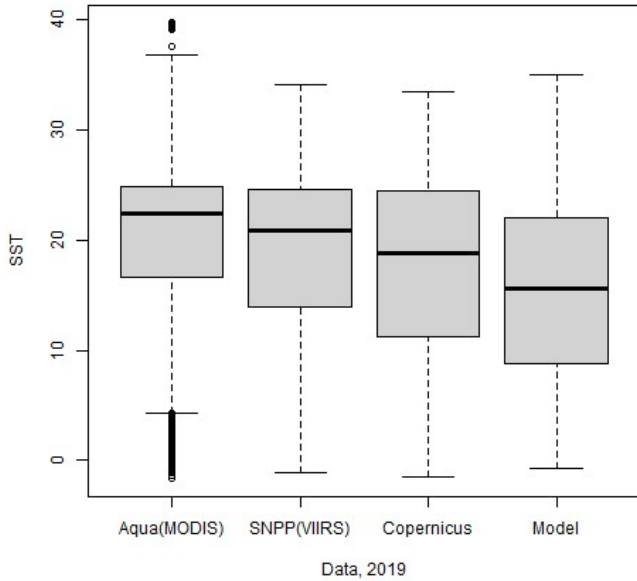


Figure 1. Distribution of SST values for whole year 2019, the bold line is the median, °C.

and Azov Seas (product ID `sst_eur_sst_l3s_nrt_observations_010_009_a`), the L3 multi-sensor (supercolated) product is built from bias-corrected L3 mono-sensor (collated) products at the resolution 0.02 degrees. A synthesis of the bias-corrected L3 mono-sensor (collated) files remapped is done through a selection of data based on the following hierarchy: AVHRR_METOP_B, VIIRS_NPP, SLSTRA, SEVIRI, AVHRR_L19, MODIS_A, MODIS_T, AMSR2. Further, for brevity, the indicated data are called Copernicus data in the paper.

To make sure that Copernicus data are the closest to the real, an analysis of the available data was carried out and a number of statistical indicators were calculated. The Figure 1 shows the distribution of SST values for 2019 according to the estimated data, Copernicus data, and calculations of the hydrodynamics numerical model. The figure Figure 1 also shows the median of the values.

Comparison of the data shows that the SST values of the numerical model are underestimated, the values of the data obtained from the MODIS and VIIRS are close to each other, the Copernicus data are between the model and estimated observation data. Additional analysis showed that the data obtained from the MODIS and VIIRS are the closest in summer and differ in winter periods of time. It could be explained by the fact that during the win-

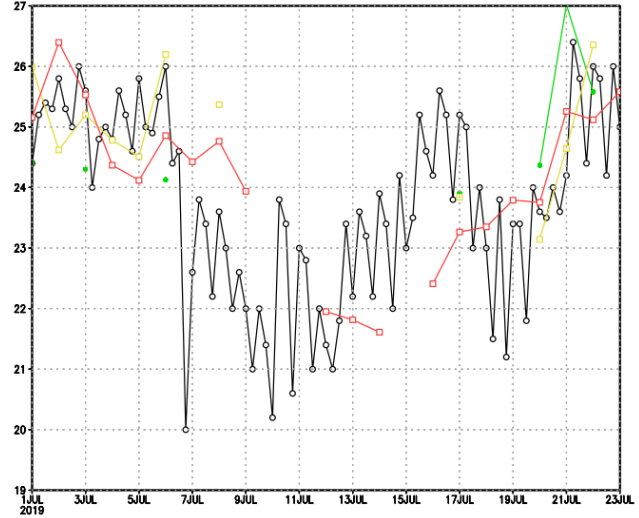


Figure 2. Time-series of SST. Station point 29.67°E, 45.15°N, July, 2019. Black line – station data, red – Copernicus data, green – SNPP (VIIRS), yellow – Aqua (MODIS), °C.

ter, due to the peculiarities of weather conditions, the amount of observation data from satellites is much less than in summer, less cloudy months.

The next step, in order to assess the proximity of Copernicus data to real data, was obtained *in situ* data – point data on sea surface temperature, from moored buoys and coastal stations. Such data are insufficient (on average, 10 points in the coastal zones of the water area with measurements during a month) for assimilation of data in the numerical model, however, they can be used to verify the temperature values, both observation data and the results of calculations of the numerical model. Thus, Black Sea – near real-time (NRT) *in situ* quality controlled observations (product id `in-situ_bs_nrt_observations_013_034`) were chosen from the Copernicus Marine Service.

The distribution of SST values according to Copernicus data, data from Aqua(MODIS) and SNPP(VIIRS) satellites and point data from coastal stations are demonstrated in Figure 2 with SST values at a point 29.67°E, 45.15°N. From a comparison of the presented data, it can be concluded that Copernicus data are close to real ones.

Thus, to further assess the accuracy of the estimated data of the MODIS and VIIRS instruments to reality in order to determine the weighting coefficients, the Copernicus data were selected. The next step was to calculate the deviations of the

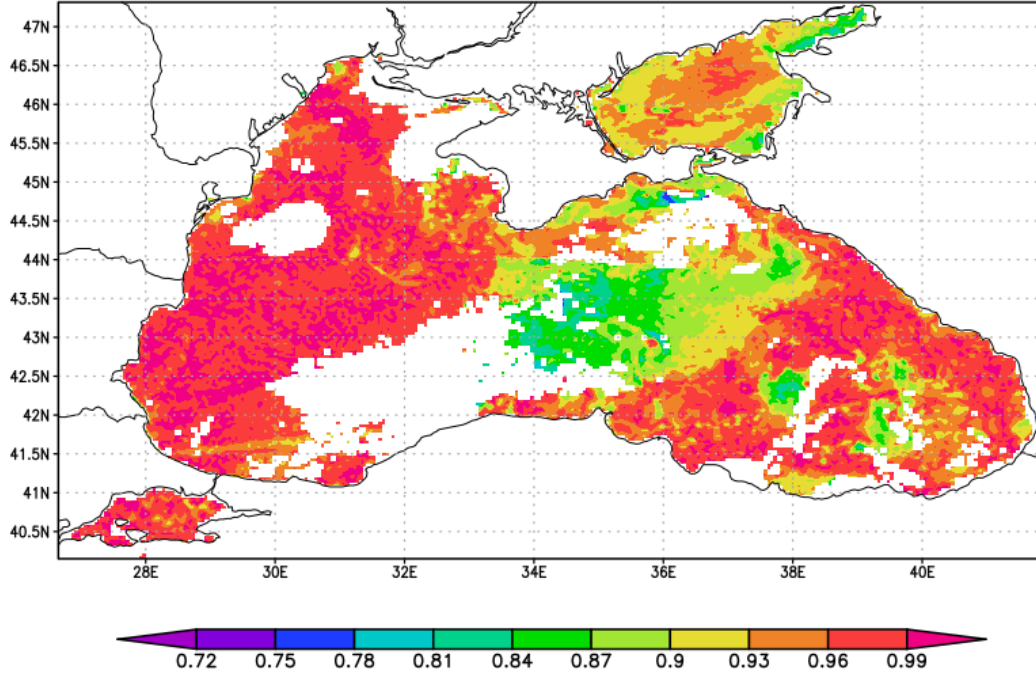


Figure 3. The field of weight coefficients calculated on 29.08.2019 based on all Aqua(MODIS) data observed on this date (09:20, 09:25, 11:00, 22:55, 23:00).

temperature field values according to the MODIS and VIIRS data from the Copernicus data in the part where the data of the studied satellites and Copernicus data are available. Weight coefficients for each day of the year, at each point of the water area, were calculated using the formula

$$\beta(x, y, t) = 10 \frac{|T_{\text{obs}}(x, y, t) - T_C(x, y, t)|}{|T_C(x, y, t)|},$$

where T_C – Copernicus data, T_{obs} – estimated observation data. This formula allows to take a weighting factor close to 1, if the data is close to the “reference”, and reduce the factor as the relative error between the data increases.

In the calculations the daily average values of the observation data instead of the instantaneous ones are used, since the comparison is with the daily average data of Copernicus.

Thus, weights were calculated for each day of the year for each mesh point in the water area (in which data was obtained on that day of the year) based on data for 2019 year. An example of weighting coefficients for Aqua(MODIS) and SNPP(VIIRS) data in the Black and Azov Seas is presented in Figure 3 and Figure 4.

It can be seen from the presented figures that the weight coefficients calculated from the first and second observation data are similar in some regions of the water area, which indicates that the data obtained with both MODIS and VIIRS differ from the Copernicus data in these areas and have a general character of deviations. But if the Copernicus data are considered to be the most reliable, then the obtained weighting coefficient field in the assimilation algorithm will allow the model to be closer to the data and weaker, where the MODIS and/or VIIRS data values are relatively different from the Copernicus data.

In the next section an algorithm for variational data assimilation with the matrices of weight coefficients β is formulated and iterative method for solving the problem is proposed.

Statement of Variational Data Assimilation Problem

Let us describe the algorithm of variational data assimilation using the example of the problem of

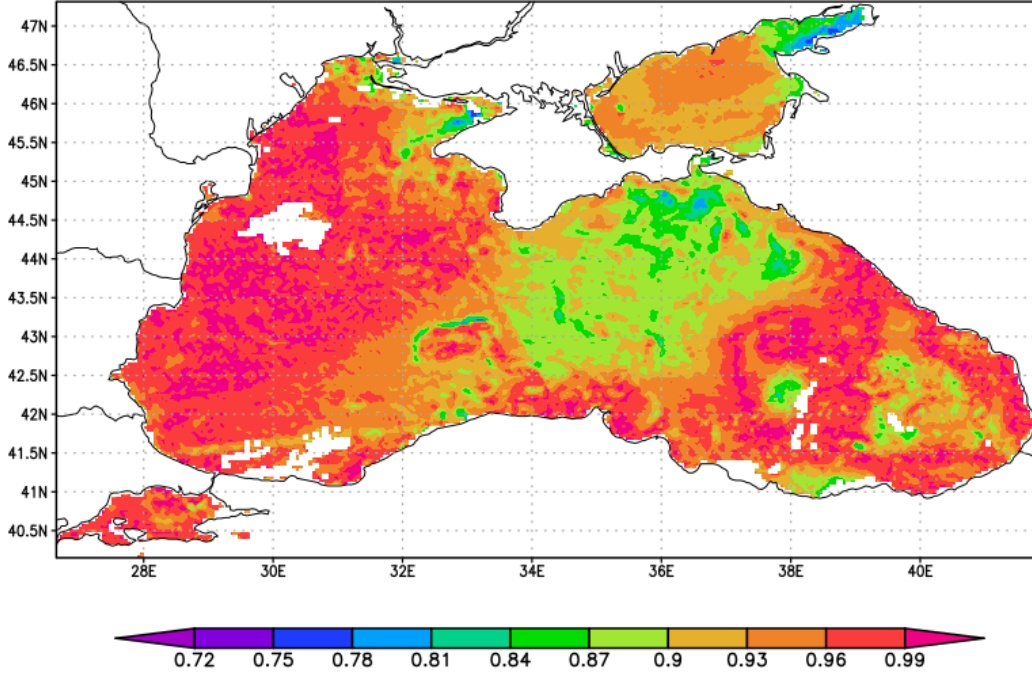


Figure 4. The field of weight coefficients calculated on 29.08.2019 based on all SNPP(VIIRS) data observed on this date (09:24, 09:30, 11:06, 23:24).

heat transfer in the sea. Consider the thermodynamic system of equations in the form:

$$\begin{aligned}
 & T_t + \frac{1}{2} \left(w_1 \frac{\partial T}{\partial z} + \frac{1}{r^2} \frac{\partial (r^2 w_1 T)}{\partial z} \right) - \\
 & \frac{1}{r^2} \frac{\partial}{\partial z} r^2 \nu_T \frac{\partial T}{\partial z} = \\
 & f_T \text{ in } D \text{ for } t \in (t_{j-1}, t_j), \\
 & T = T_1(t_j) \text{ for } t = t_{j-1}, \\
 & -\nu_T \frac{\partial T}{\partial z} = Q \text{ at } z = 0, \\
 & \nu_T \frac{\partial T}{\partial z} = 0 \text{ at } z = H,
 \end{aligned} \tag{1}$$

here T is a sea surface temperature, w_1 is a vertical velocity, r is a radius of the Earth, ν_T is a turbulent exchange coefficient, f_T is a known function of external sources, z is a coordinate in depth. The function:

$$Q \equiv Q_T - \gamma_T(T - T_a) - \bar{U}_n^{(-)}T + \bar{U}_n^{(-)}d_T$$

can be considered as a turbulent heat flux.

Suppose that in the considered part of the water area the function of observations T_{obs} on the

SST is given. The area where the data is specified is determined by the characteristic function of the observations m_0 , which is equal to 1 at the grid points where the observation data exists and 0 at the other points. Note that the values of m_0 can change over time. Let us consider the case when the unknowns are the solution of the system (1) and the value of the heat flux at the surface Q , which is commonly called “control”.

Let us introduce a functional characterizing the deviation of the temperature calculated by the model from the observation data:

$$\begin{aligned}
 J_\alpha(Q) = & \frac{\alpha}{2} \int_0^{\bar{t}} \int_\Omega |Q - Q^{(0)}|^2 d\Omega dt + \\
 & \sum_{j=1}^J \int_{t_{j-1}}^{t_j} \int_\Omega \beta |(m_0 T|_{z=0} - T_{\text{obs}})|^2 d\Omega dt,
 \end{aligned} \tag{2}$$

where $Q^{(0)} = Q^{(0)}(x, y, t)$ is a given function, $\alpha = \text{const} > 0$ is a regularization parameter. Here $\beta = \beta(x, y, t)$ are matrices of weight coefficients characterizing the “reliability” of the observation data, which is obtained from the statistical analy-

sis of various data sources, described in previous section.

Thus, the problem of variational assimilation in this case can be formulated as follows: it is required to find a solution to the equation (1) and the function Q , such that the functional (2) takes the least value.

One of the methods for solving this problem is the use of the theory of adjoint equations. So, using the results of [Shutyaev and Parmuzin, 2021], one can write out an optimality system, which, according to the necessary condition $\text{grad}J_\alpha(Q) = 0$, determines the solution of the problem of variational data assimilation and reduces to the sequential solution of the main equation, adjoint equation and the additional conditions on the control Q . In our case, the optimality system takes the following form:

$$\left\{ \begin{array}{l} T_t + \frac{1}{2} \left(w_1 \frac{\partial T}{\partial z} + \frac{1}{r^2} \frac{\partial(r^2 w_1 T)}{\partial z} \right) - \\ \frac{1}{r^2} \frac{\partial}{\partial z} r^2 \nu_T \frac{\partial T}{\partial z} = f_T, \quad T_{t=0} = T_0, \\ -\nu_T \frac{\partial T}{\partial z} \Big|_{z=0} = Q, \quad \nu_T \frac{\partial T}{\partial z} \Big|_{z=H} = 0, \end{array} \right. \quad (3)$$

$$\left\{ \begin{array}{l} T_t^* - \frac{1}{2} \left(w_1 \frac{\partial T^*}{\partial z} + \frac{1}{r^2} \frac{\partial(r^2 w_1 T^*)}{\partial z} \right) - \\ \frac{1}{r^2} \frac{\partial}{\partial z} \left(r^2 \nu_T \frac{\partial T^*}{\partial z} \right) = 0, \quad T_{t=T}^* = 0, \\ \nu_T \frac{\partial T^*}{\partial z} \Big|_{z=H} = 0, \\ \left(-w_1 T^* - \nu_T \frac{\partial T^*}{\partial z} \right) \Big|_{z=0} = \\ 2\beta m_0 (T - T_{\text{obs}}), \end{array} \right. \quad (4)$$

$$\alpha_0 (Q - Q^{(0)}) + T^* = 0. \quad (5)$$

The approximate solution of the numerical model with variational data assimilation in this case is described as follows: the system (3) is solved with the initial approximation Q^k , the system (4) is solved with known solution of the direct problem (3), and then the new approximation Q^{k+1} is calculated according to (5) as follows:

$$Q^{(k+1)} = Q^{(k)} - \gamma_k (\alpha(Q^{(k)} - Q^{(0)}) + T^*)$$

with the parameter γ_k , which ensures the convergence of this iterative process.

Other variants of using the considered algorithm in models of hydrothermodynamics of seas and oceans can be found for example in [Agoshkov et al., 2019; Sheloput, 2018].

Numerical Experiments and Discussion

In order to verify the proposed method the numerical experiments are performed in the water area of Marmarian, Black and Azov seas. The numerical model, developed at the INM RAS, is based on the splitting method [Zalesny et al., 2012] and σ -grid on Z direction. The spatial resolution of the numerical model is 4 km and location of the first point of the computational mesh is 26.65°E and 40.15°N. Numerical experiments are carried out with a time step equal to 2.5 minutes. The data assimilation procedure [Shutyaev and Parmuzin, 2021] is embedded in the model with the aim of restoring the heat flux Q on the surface using sea surface observation data T_{obs} from Aqua(MODIS), Terra(MODIS) and SNPP(VIIRS) satellites, described above.

The numerical experiment is a calculation for one year with the assimilation of observation data T_{obs} during the whole year 2019. The data assimilation scheme can be described as follows: at the moment of time when there is a data from the satellite, a 10-minute assimilation window is selected (5 model time steps) and at this time the data assimilation procedure works in the model. When there is no observed data, a model calculation is performed without assimilation.

The climatic flux obtained from the NCEP (National Centers for Environmental Prediction) reanalysis is used as $Q^{(0)}$. The results of the calculation by the model without assimilation procedure, averaged over one day, are compared with the results of calculation with assimilation procedure and matrices of weight coefficients $\beta(x, y, t)$, averaged over one day. The results of this numerical experiments are shown in Figure 5 and Figure 6. Figure 5 shows the time series from 1 of July 2019 till 23d of July 2019 at station point 27.95°E, 43.2°N in Black Sea water area, which is located in the western part of the sea. The green line presents the result obtained by the model without assimilation procedure, blue line – the calculation with assimilation and matrices of weight coefficients $\beta(x, y, t)$, red

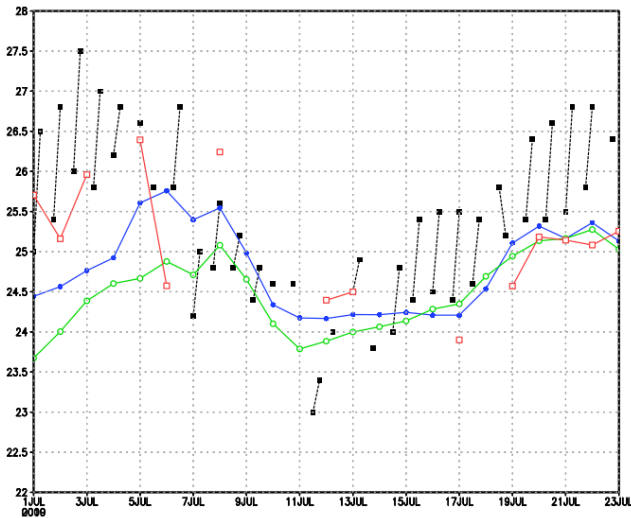


Figure 5. Time-series of temperature, point 27.95°E, 43.2°N, July 2019, °C. Black line – station point, red – daily-mean from Copernicus, green – model without assimilation, blue – model with assimilation.

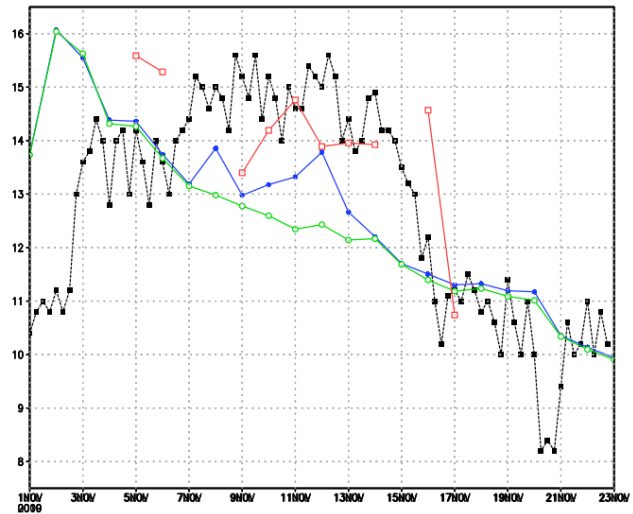


Figure 6. Time-series of temperature, point 29.67°E, 45.15°N, November 2019, °C. Black line – station data, red – daily-mean from Copernicus, green – model without assimilation, blue – model with assimilation.

line – daily-mean data from Copernicus and black dash line – near real-time the *in situ* data from station. The results of the experiment for the period from 1st of November till 23d of November 2019 at station point 29.67°E, 45.15°N are presented at Figure 6.

As can be seen from the figures the average daily sea surface temperature becomes 0.5–1.0°C closer to the *in situ* data from the coastal station when calculated with the assimilation procedure. This is especially noticeable in the summer period of time, since at this time, due to weather conditions and the characteristics of measuring instruments on satellites, the bigger amount of data are observed. In November (see Figure 6) the influence of data assimilation is less due to the small number of observations. It should be noted that sometimes the difference between the satellites observation data (both assimilated data and “reference” data) and *in situ* data reaches 2–3°C, which does not allow the model, even with an assimilation procedure, to obtain a result close to the data from the station. However at those points in time when *in situ* data and Copernicus data have good correlation, the model with data assimilation shows good calculation accuracy. In the period from November 7 to 13, the daily mean of sea surface temperature calculated by the model with assimilation is on av-

erage 0.5–1.0°C closer to the *in situ* data than the daily mean of sea surface temperature calculated by the model without assimilation. Thus, preliminary processing and analysis of observational data, which are used later in the procedures of variational data assimilation, is an important part of modeling and forecasting the state of the studied water area. It worth to mentioned that *in situ* data from coastal stations are located at the very boundary of the computational domain of the model; therefore, special attention should be paid to setting boundary conditions, including conditions for river runoff [Lebedev *et al.*, 2020].

Conclusions

In the paper the data analysis is performed and algorithms to take into account information obtained from the satellites to further use in the data assimilation problem are considered.

Based on the studies carried out, it can be concluded that the embedding the assimilation procedure with matrices of weight coefficients into the hydrothermodynamic model makes it possible to obtain the results of numerical calculations that are closer to the observation data than with-

out this procedure. In some parts of the considered water area, the difference between the values obtained with and without data assimilation can reach 2.0°C. Also, the analysis of data and the construction of matrices of weight coefficients on its basis makes the process of data assimilation more realistic, allowing us to reduce the impact of insufficiently accurate or non-faithful data.

It is worth noting that the weight coefficient β depends not only on the accuracy of the observational data, but also on the choice of data for evaluating the accuracy of the data – “reference”. In this paper, the Copernicus data are used, but, depending on the task being solved, to assess the accuracy of operational data, one can choose other data fields for evaluating accuracy – climatic, averaged operational, model, observational data from other sources and combinations of available data.

Numerical experiments have shown that in the months when there are more observation data on the state of the studied environment, the solution of the numerical model with data assimilation procedures is closer to the observation data. From this can be concluded that by filling in the gaps in the data and thereby increasing the amount of incoming information, it is possible to make the numerical solution more accurate. In this case, it makes sense to introduce additional procedures for interpolation and extrapolation of observation data.

Acknowledgments. The work was supported by the Russian Science Foundation (project 19-71-20035, Informational Computational System for Variational Data Assimilation “INM RAS – Black Sea” and its integration with the hardware-software complex of the CKP “IKI-Monitoring”).

References

Agoshkov, V. I., N. R. Lezina, T. O. Sheloput (2019), Domain Decomposition Method for the Variational Assimilation of the Sea Level in a Model of Open Water Areas Hydrodynamics, *Journal of Marine Science and Engineering*, 7, 195, [Crossref](#)

- Lebedev, S. A., A. G. Kostianoy, D. M. Soloviev, et al. (2020), On a relationship between the river runoff and the river plume area in the northeastern Black Sea, *International Journal of Remote Sensing*, 41, No. 15, 5806–5818, [Crossref](#)
- Proshin, A. A., M. A. Burtsev, I. V. Balashov, et al. (2020), “IKI-Monitoring” shared use center support and development – possible solutions, *Problems in Remote Sensing of the Earth from Space*, 17, No. 6, 51–55, [Crossref](#)
- von Schuckmann, Karina, Pierre-Yves Le Traon, et al. (2021), Copernicus Marine Service Ocean State Report, *Journal of Operational Oceanography*, 14, No. 5, 1–185, [Crossref](#)
- Sheloput, T. O. (2018), Numerical solution of the problem of variational assimilation of the sea level on the liquid (open) boundary in the Baltic Sea hydrothermodynamics model, *Problems in Remote Sensing of the Earth from Space*, 15, No. 7, 15–23, (in Russian) [Crossref](#)
- Shutyaev, V. P., E. I. Parmuzin (2021), Numerical solution of the problem of variational data assimilation to restore heat fluxes and initial state for the ocean thermodynamics model, *Russ. J. Numer. Anal. Math. Modelling*, 36, No. 1, 43–53, [Crossref](#)
- Stepanov, V. N., Yu. D. Resnyanskii, et al. (2021), Evaluating Effects of Observational Data Assimilation in General Ocean Circulation Model by Ensemble Kalman Filtering: Numerical Experiments with Synthetic Observations, *Russian Meteorology and Hydrology*, 46, No. 2, 94–105, [Crossref](#)
- Zakharova, N. B. (2016), Verification of the sea surface temperature observation data, *Problems in Remote Sensing of the Earth from Space*, 13, No. 3, 106–113, [Crossref](#)
- Zalesny, V. B., N. A. Diansky, V. V. Fomin, et al. (2012), Numerical model of the circulation of the Black Sea and the Sea of Azov, *Russ. J. Numer. Anal. Math. Modelling*, 27, No. 1, 95–111, [Crossref](#)

Corresponding author:

N. B. Zakharova, Marchuk Institute of Numerical Mathematics Russian Academy of Sciences, Moscow, Russia. (n.zakharova@inm.ras.ru)