Mathematical Modeling of Climate, Dynamics Atmosphere and Ocean: to the 95th Anniversary of G. I. Marchuk and the 40th Anniversary of the INM RAS

V. P. Dymnikov[a], *, E. E. Tyrtyshnikov[a], **, V. N. Lykossov[a], ***, and V. B. Zalesny[a], ****

[a]Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, 119333 Russia
*e-mail: dymnikov@inm.ras.ru
**e-mail: eugene.tyrtyshnikov@gmail.com
***e-mail: lykossov@yandex.ru
****e-mail: vzalesny@yandex.ru
Received January 20, 2020; revised January 27, 2020; accepted February 5, 2020

Abstract—Introductory article on the issue of the journal Izvestiya RAN, Atmospheric and Oceanic Physics, 2020, vol. 56, no. 3 dedicated to the 40th anniversary of the Marchuk Institute of Numerical Mathematics of the Russian Academy of Sciences (INM RAS) and the 95th anniversary of the birth of its founder—Academician G.I. Marchuk. The main features of the scientific-organizational and pedagogical activities of G.I. Marchuk, who had a significant impact on the development of modern geophysical fluid dynamics, physics of the atmosphere and ocean, and their reflection in the INM RAS research and development activities over a 40-year period.

> *Remember now your Creator in the days of your youth...*
> *While the sun, or the light, or the moon, or the stars,*
> *be not darkened, nor the clouds return after rain...*
> *Or ever the silver cord be loosed, or the golden bowl be broken, or*
> *the pitcher be broken at the fountain, or the wheel broken at the cistern...*
> *The words of the wise are as goads, and as nails fastened*
> *by the masters of assemblies, which are given from one shepherd.*
> *(Ecclesiastes, XII:1-6, 11)*

The issue of the journal is dedicated to the 40th anniversary of the Marchuk Institute of Numerical Mathematics of the Russian Academy of Sciences and the 95th anniversary of the birth of its founder—Academician Gury Ivanovich Marchuk.

The eminent scientist and organizer, teacher and public figure, Guri Ivanovich Marchuk left a bright mark in the development of several areas of modern science, the organization of scientific institutes and educational departments of Soviet and Russian Universities. We will single out the five keystones of his scientific, organizational, and pedagogical activity—two Institutes and three university Chairs. In the Novosibirsk Academgorodok, these are the Computing Center of the Siberian Branch of the USSR Academy of Sciences and the Chair of Mathematical Methods in Dynamic Meteorology at the Mechanics and Mathematics Department of Novosibirsk State University, with a major in Computational Mathematics, organized in 1964 (it was included in the Chair of Computational Mathematics in 1972); in Moscow—the Department (in 1991, the Institute) of Numerical Mathematics of the Academy of Sciences of the USSR (later the Russian Academy of Sciences) and the Chair of Mathematical Modeling of Physical Processes at the Department of Problems of Physics and Energetics of the Moscow Institute of Physics and Technology (MIPT), and also organized in 2004, the Chair of Computational Technologies and Modeling at the Faculty of Computational Mathematics and Cybernetics of Lomonosov Moscow State University (MSU).

Works by G.I. Marchuk in the field of computational mathematics [1−3] and mathematical modeling of climate, atmospheric and ocean dynamics [3, 4], developing of national scientific traditions [5−7], had a significant impact on modern science. First of all,

this refers to the central direction of atmospheric and ocean physics —mathematical modeling and numerical solution of weather prediction and climate change problems. The deep connection of numerical mathematics and geophysical fluid dynamics is the main feature of the work of his scientific school and the Institute of Numerical Mathematics (INM) of the Russian Academy of Sciences founded by him.

Among the scientific studies of G.I. Marchuk and his scientific school in the field of geophysical fluid dynamics, the atmospheric and oceanic physics can distinguish several main directions. This is a simulation of the Earth's climate system; mathematical problems of climate theory; adjoint equations and methods of the control theory in geophysical fluid dynamics problems; mathematical simulation and numerical algorithms of atmospheric and oceanic dynamics.

It should be noted an important feature of the organizational activities of G.I. Marchuk, to this day manifesting itself in the traditions of the scientific teams founded by him. It is expressed in the creation of a key science group, which is taken to solve extremely large, multidisciplinary problems [8−10]. The group, as a rule, small and mobile, brings together both experienced researchers, graduate students, and students. The solution of extremely large problems requires, in addition to the key group, the involvement of specialists of various scientific profiles from other research institutes and universities, and the creation of external scientific teams. On this basis, the scientific directions, structure and modern research staff of the Marchuk Institute of Numerical Mathematics are formed.

Guri Ivanovich said that in its activities, the INM RAS should follow the following principles: be aimed at solving important urgent problems; consist of three almost equal parts uniting researchers, graduate students and students; instead of traditional laboratories and departments, have a more flexible structure—creative research teams, the composition and subjects of which can change over time, following the challenges of modern science [11, 12].

The main research directions of the INM RAS are related to the scientific interests of its creator. In 1980 there were four of them: numerical mathematics; simulation of atmospheric and oceanic dynamics; mathematical immunology and medicine; parallel computing. Currently, the research areas of the Institute has changed somewhat. Now it is computational mathematics, tensors and optimization methods [13−19]; methods and technologies of numerical mathematics and problems of biology and medicine [20, 21]; modeling the Earth system dynamics and environmental problems [22−25]; mathematical modeling of the atmosphere and ocean dynamics; and the variational data assimilation problems [10, 12, 17, 26−29].

Celebrating its 40th anniversary in 2020, the Marchuk Institute of Numerical Mathematics RAS is currently in consortium with Lomonosov Moscow State University, Keldysh Institute of Applied Mathematics RAS is part of the Moscow Center for Fundamental and Applied Mathematics.

The research directions of the INM RAS correspond to the Program of Fundamental Studies of the State Academies of Sciences for 2013−2020. In addition to 16 basic research topics, the Institute over the past 5 years has completed more than 100 various projects under the Programs of the Presidium and Departments of the Russian Academy of Sciences, Federal Target Programs, Programs of Ministries and Departments, Russian Foundation for Basic Research, Russian Science Foundation, and International Programs (https://www.inm.ras.ru).

In the field of geophysical fluid dynamics, climate theory and environmental problems, the Institute actively cooperates with leading scientific teams of the Obukhov Institute of Atmospheric Physics RAS, Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Institute of Monitoring of Climate and Ecological Systems SB RAS, Shirshov Institute of Oceanology RAS, Moscow State University, Hydrometeorological Center of Russia, Marine Hydrophysical Institute RAS, Nuclear Safety Institute RAS, Zubov's State Oceanographic Institute and other organizations.

Along with scientific research, the Institute pays more attention to scientific and pedagogical activities in the framework of the graduate school of the INM RAS and its two basic Chairs: Computing Technologies and Modeling in Geophysics and Mathematics—at the Department of Control and Applied Mathematics of the Moscow Institute of Physics and Technology and Computing Technologies and Modeling—at the

Faculty of Computational Mathematics and Cybernetics of Moscow State University [16, 30]. Of the 55 scientific staff at present, 29 graduates of the Moscow Institute of Physics and Technology and 11 at Moscow State University work at the INM RAS.

The joint efforts of researchers, graduate students and students at the INM RAS develop ideas that were laid down by the founder of the Institute, Academician Gury Ivanovich Marchuk.

REFERENCES

1. G. I. Marchuk, *Methods of Computational Mathematics* (Nauka, Moscow, 1989) [in Russian].

2. G. I. Marchuk, *Selected Works,* vol. I. *Methods of Computational Mathematics* (RAN, Moscow, 2018) [in Russian].

3. G. I. Marchuk, *Selected Works,* vol. II. *Adjoint Equations and the Analysis of Complex Systems* (RAN, Moscow, 2018) [in Russian].

4. G. I. Marchuk, *Selected Works,* vol. III. *Models and Methods in Problems of Atmosphere and Ocean Physics* (RAN, Moscow, 2018) [in Russian].

5. A. M. Obukhov, *Turbulence and Atmosphere Dynamics* (Gidrometeoizdat, Leningrad, 1988) [in Russian].

6. A. S. Monin, *Introduction to the Climate Theory* (Gidrometeoizdat, Leningrad, 1982) [in Russian].

7. G. S. Golitsyn, *Statistics and Dynamics of Natural Processes and Phenomena* (KRASAND, Moscow, 2012) [in Russian].

8. G. I. Marchuk, V. P. Dymnikov, V. B. Zalesny, V. N. Lykosov, and V. Ya. Galin, *Mathematical Modeling of the Atmosphere-Ocean General Circulation* (Gidrometeoizdat, Leningrad, 1984) [in Russian].

9. G. I. Marchuk, V. P. Dymnikov, and V. B. Zalesny, *Mathematical Models in Geophysical Hydrodynamics and Numerical Methods for Their Implementation* (Gidrometeoizdat, Leningrad, 1987) [in Russian].

10. G. I. Marchuk and B. E. Paton, "The Black Sea as a simulation ocean model," Russ. J. Numer. Anal. Math. Modell. **27** (1), 1−4 (2012).

11. G. I. Marchuk, *A Life in Science* (Nauka, Moscow, 2000) [in Russian].

12. G. I. Marchuk, B. E. Paton, G. K. Korotaev, and V. B. Zalesny, "Data-computing technologies: A new stage in the development of operational oceanography," Izv., Atmos. Ocean. Phys. **49** (6), 579−591 (2013).

13. V. P. Shutyaev, *Control Operators and Iterative Algorithms in Variational Data Assimilation Problems* (Nauka, Moscow, 2001) [in Russian].

14. E. E. Tyrtyshnikov, "Tensor approximations of matrices generated by asymptotically smooth functions," Mat. Sb. **194** (5−6), 941−954 (2003).

15. I. Oseledets and E. E. Tyrtyshnikov, "TT-Cross approximation for multidimensional arrays," Linear Algebra Appl. **432** (1), 70−88 (2010).

16. Yu. V. Vasilevskii, A. A. Danilov, K. N. Lipnikov, and V. N. Chugunov, *Automatized Technologies of Design on Unstructured Calculation Grids* (Fizmatlit, Moscow, 2016) [in Russian].

17. V. I. Agoshkov, *Domain Decomposition Methods in Ocean and Sea Hydrothermodynamics Problems* (IVM RAN, Moscow, 2017) [in Russian].

18. A. B. Bogatyrev, "Real meromorphic differentials: a language for describing meron configurations in planar magnetic nanoelements," Theor. Math. Phys. **193,** 1547−1559 (2017).

19. Yu. M. Nechepurenko and M. Sadkane, "Computing humps of the matrix exponential," J. Comput. Appl. Math. **319,** 87−96 (2017).

20. G. I. Marchuk, V. Shutyaev, and G. Bocharov, "Adjoint equations and analysis of complex systems: application to virus infection modeling," J. Comput. Appl. Math. **184,** 177−204 (2005).

21. K. Nikitin, M. Olshanskii, K. Terekhov, and Y. Vassilevski, "A splitting method for numerical simulation of free surface flows of incompressible fluids with surface tension," Comput. Methods Applied Math. **15** (1), 59−77 (2015).

22. A. E. Aloyan, *Modeling Dynamics and Kinetics of Gas Admixtures and Aerosols in the Atmosphere* (Nauka, Moscow, 2008).

23. V. N. Lykosov, A. V. Glazunov, D. V. Kulyamin, E. V. Mortikov, and V. M. Stepanenko, *Supercomputer Simulation in Physics of Climate Systems* (Mosk. Gos. Univ., Moscow, 2012) [in Russian].

24. V. P. Dymnikov, V. N. Lykosov, and E. M. Volodin, "Mathematical simulation of Earth system dynamics," Izv., Atmos. Ocean. Phys. **51,** 227−240 (2015).

25. E. M. Volodin, V. Ya. Galin, A. S. Gritsun, A. V. Gusev, N. A. Diansky, V. P. Dymnikov, R. A. Ibrayev, V. V. Kalmykov, S. V. Kostrykin, D. V. Kulyamin, V. N. Lykossov, E. V. Mortikov, O. O. Rybak, M. A. Tolstykh, R. Yu. Fadeev, I. A. Chernov, V. V. Shashkin, and N. G. Yakovlev, *Mathematical Modeling of the Earth System,* Ed. by N. G. Yakovlev (MAKS Press, Moscow, 2016) [in Russian].

26. G. I. Marchuk and V. B. Zalesny, "Modeling of the World Ocean circulation with the four-dimensional assimilation of temperature and salinity fields," Izv., Atmos. Ocean. Phys. **48** (1), 15−29 (2012).

27. M. A. Tolstykh, R. Yu. Fadeev, V. V. Shashkin, G. S. Goyman, R. B. Zaripov, D. B. Kiktev, S. V. Makhnorylova, V. G. Mizyak, and V. S. Rogutov, "Multiscale global atmosphere model SL-AV: the results of medium-range weather forecasts," Russ. Meteorol. Hydrol. **43,** 773 (2018).

28. V. V. Kalmykov, R. A. Ibraev, M. N. Kaurkin, and K. V. Ushakov, "Compact Modeling Framework V3.0 for high-resolution global ocean-ice-atmosphere models," Geosci. Model Dev. **11** (10), 3983−3997 (2018).

29. V. P. Dymnikov and V. B. Zalesny, *Fundamentals of Computational Geophysical Fluid Dynamics* (GEOS, Moscow, 2019) [in Russian].

30. E. E. Tyrtyshnikov, *Basic Algebra,* (Fizmatlit, Moscow, 2017) [in Russian].

Simulation of Possible Future Climate Changes in the 21st Century in the INM-CM5 Climate Model

E. M. Volodin[a], * and A. S. Gritsun[a]

[a]*Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, 119333 Russian*
***e-mail: volodinev@gmail.com**

Abstract—Climate changes in 2015–2100 have been simulated with the use of the INM-CM5 climate model following four scenarios: SSP1-2.6, SSP2-4.5, and SSP5-8.5 (single model runs) and SSP3-7.0 (an ensemble of five model runs). Changes in the global mean temperature and spatial distribution of temperature and precipitation are analyzed. The global warming predicted by the INM-CM5 model in the scenarios considered is smaller than that in other CMIP6 models. It is shown that the temperature in the hottest summer month can rise more quickly than the seasonal mean temperature in Russia. An analysis of a change in Arctic sea ice shows no complete Arctic summer ice melting in the 21st century under any model scenario. Changes in the meridional streamfunction in atmosphere and ocean are studied.

INTRODUCTION

Forecasting future climate changes is one of the most important applications of climate simulation. This problem can be solved in various ways on different time scales. In this work, we consider probable climate changes in the 21st century associated with changes in the concentrations or emissions of greenhouse and other gases and aerosols associated with human activities.

There are several dozen climate models in the global scientific community, and they differ in the methods used to solve atmosphere and ocean dynamics equations and to parameterize physical processes, spatial resolution, and the presence or absence of certain blocks (for example, atmospheric chemistry, aerosol, and carbon cycle blocks). Climate models are compared every 7–8 years within the CMIP (Coupled Model Intercomparison Project) by means of coordinated experiments with the models. The data of such experiments enter a single database and are processed by different research teams. The processing results are published in scientific papers and in the Reports of the Intergovernmental Panel on Climate Change (IPCC).

Experiments within the 5th phase of CMIP project took place in 2009–2011, and the corresponding 5th IPCC report was published in 2013. The next, 6th phase of the experiments was carried out in 2017–2019. The 6th IPCC report, which should include, among other things, the results of these experiments, is expected to be released in 2022.

The CMIP6 structure was as follows: model groups should first perform the required numerical experiments: a preindustrial experiment for at least 500 years, in which all impacts on the climate system were fixed to 1850; experiments with instant and gradual quadrupling of CO_2 concentration, which allow estimating the equilibrium sensitivity of a model; an experiment with the atmospheric block of a model and a specified state of the ocean surface; and a historical experiment where the climate change for 1850–2014 was simulated and impacts on the climate system were set in accordance with available observations during that time period. After the required experiments, models could participate in different subprojects. One such subproject is the simulation of probable future climate changes in the 21st century according to different scenarios (ScenarioMIP). Numerical experiments, the results of which are discussed in this work, have been performed within this subproject.

The climate is understood as the totality of statistical characteristics of the instantaneous states of the atmosphere, ocean, and other climate system components averaged over a long time period. Therefore, we restrict ourselves to an analysis of some of the most important climate parameters, such as average temperature and precipitation. A more detailed analysis of individual aspects of climate change, such as changes in extreme weather and climate situations, will be the subject of another work. This study is not aimed at a full comparison with the results of other climate models, where calculations follow the same scenarios, since the results of other models have not yet been published in peer-reviewed journals by the time of this writing.

MODEL AND NUMERICAL EXPERIMENTS

The INM-CM5 climate model [1, 2] is used for the numerical experiments. It differs from the previous version, INMCM4, which was also used for experiments on reproducing climate change in the 21st century [3], in the following: an aerosol block has been added to the model, which allows inputting anthropogenic emissions of aerosols and their precursors; the concentrations and optical properties of aerosols are calculated, but not specified, like in the previous version; the parametrizations of cloud formation and condensation are changed in the atmospheric block; the upper boundary in the atmospheric block is raised from 30 to 60 km; the horizontal resolution in the ocean block is doubled along each coordinate; and the software related to adaptation to massively parallel computers is improved, which allows the effective use a larger number of compute cores. The model resolution in the atmospheric and aerosol blocks is $2° \times 1.5°$ in longitude and latitude and 73 levels and, in the ocean, $0.5° \times 0.25°$ and 40 levels. The calculations were performed at supercomputers of the Joint Supercomputer Center, Russian Academy of Sciences, and Moscow State University, with the use of 360 to 720 cores. The model calculated 6−10 years per 24 h in the above configuration.

Four scenarios were used to model the future climate: SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-5.8. The scenarios are described in [4]. The figure after the abbreviation SSP (Shared Socioeconomic Pathway) is the number of the mankind development path (see the values in [4]). The number after the dash means the radiation forcing (W m^{-2}) in 2100 compared to the preindustrial level. Thus, the SSP1-2.6 scenario is the most moderate and assumes rapid actions which sharply limit and then almost completely stop anthropogenic emissions. Within this scenario, greenhouse gas concentrations are maximal in the middle of the 21st century and then slightly decrease by the end of the century. The SSP5-8.5 scenario is the warmest and implies the fastest climate change. Each scenario includes the time series of carbon dioxide, methane, nitrous oxide, and ozone concentrations; emissions of anthropogenic aerosols and their precursors; the concentration of volcanic sulfate aerosol; and the solar constant. The scenarios are recommended for use in the project on comparing CMIP6 (Coupled Model Intercomparison Project, Phase 6, [5]) climate models.

One model experiment was carried out for each of the above scenarios. It began at the beginning of 2015 and ended at the end of 2100. The initial state was taken from the so-called historical experiment with the same model, where climate changes were simulated for 1850−2014, and all impacts on the climate system were set according to observations. The results of the ensemble of historical experiments with the model under consideration are given in [6, 7]. For the SSP3-7.0 scenario, five model runs was performed differing in the initial data taken from different historical experiments. The ensemble of numerical experiments is required to increase the statistical confidence of conclusions about climate changes.

RESULTS

Let us describe some simulation results of climate change in the 21st century. Figure 1 shows the change in the globally averaged surface air temperature with respect to the data of the corresponding historical experiment for 1850−1899. In the warmest SSP5-8.5 scenario, the temperature rises by more than 4° by the end of the 21st century. In the SSP3-7.0 scenario, different members of the ensemble show warming by 3.4°−3.6°. In the SSP2-4.5 scenario, the temperature increases by about 2.4°. According to the SSP1-2.6 scenario, the maximal warming by ~1.7° occurs in the middle of the 21st century, and the temperature exceeds the preindustrial temperature by 1.4° by the end of the century. The results for other CMIP6 models have not yet been published in peer-reviewed journals. However, according to the preliminary analysis (see, e.g., https://cmip6workshop19.sciencesconf.org/data/Session1_PosterSlides.pdf, p.29), the INM-CM5 model shows the lowest temperature increase among the CMIP6 models considered for all the scenarios due to the minimal equilibrium sensitivity to the CO_2 concentration doubling, which is ~2.1° for the current model version, like for the previous version, despite new condensation and cloud formation blocks.

The changes in the surface air temperature are similar for all scenarios; therefore, we analyze the difference between temperatures in 2071−2100 and 1981−2010 under the SSP5-8.5 and SSP1-2.6 scenarios (Fig. 2). The warming is maximal in the Arctic; it reaches 10° and 3°, respectively. Other features mainly correspond to CMIP5 data [8], including the INMCM4 model, which participates in the comparison. The warming on the continents of the Northern Hemisphere is about 2 times higher than the mean, and the warming in the Southern Hemisphere is noticeably less than in the Northern Hemisphere. The land surface is getting warmer than the ocean surface in all the scenarios except SSP1-2.6, because the greenhouse effect is expected to weaken in the second half of the 21st century in this scenario, and the higher heat capacity of the ocean prevents it from cooling as quickly as the land.

The changes in the temperature in individual seasons can noticeably differ from the annual average, and the changes in extreme temperatures can differ from the changes in the averages. The changes in extreme weather and climate events in scenario experiments will be considered in detail in another work. The quality of reproducing the extreme value indices by the INM-CM5 model when reproducing the modern climate is discussed in [9]. Here, we only show how the summer average temperature in 2071−2100 changes as compared to 1980−2010 in Eurasia in the ensemble of
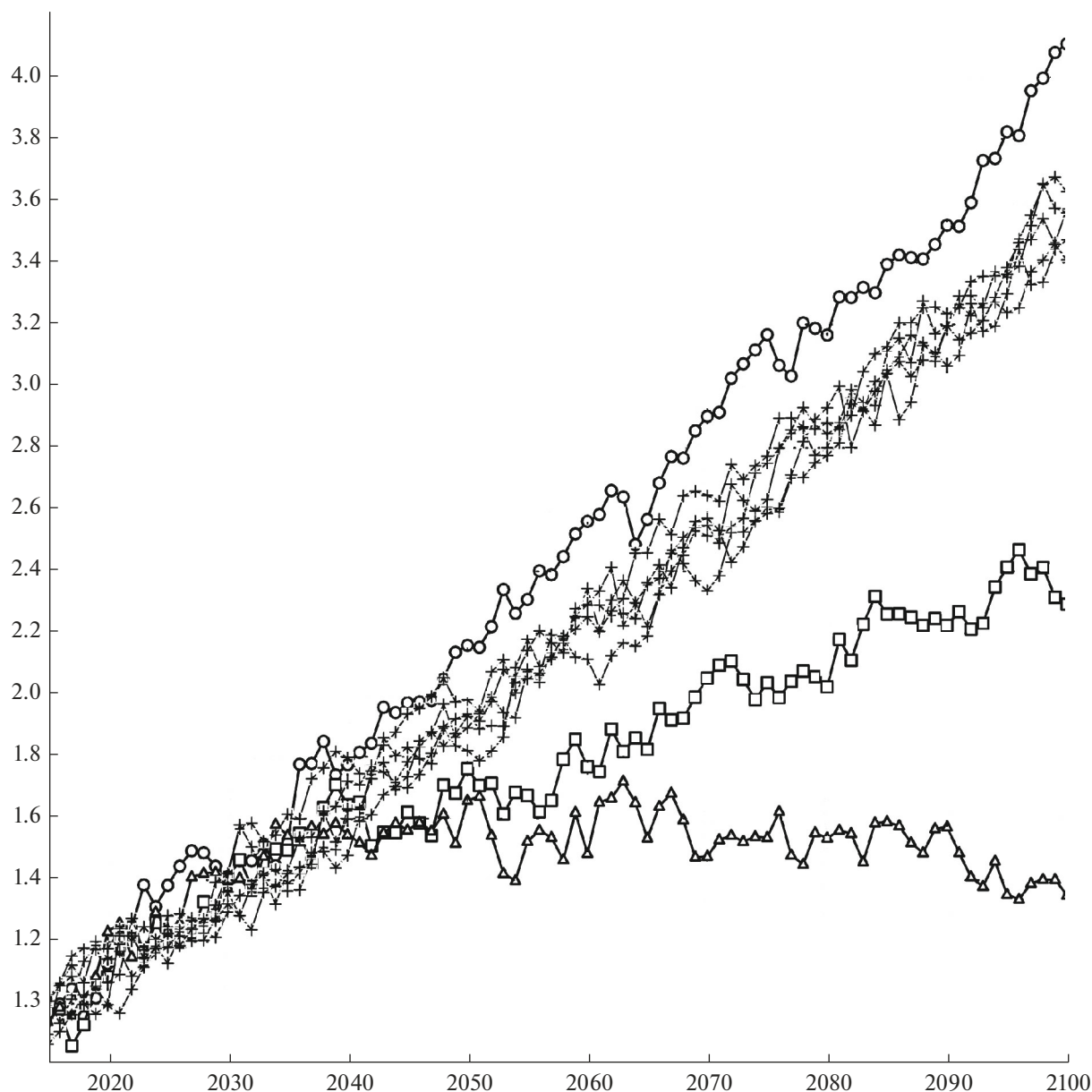
Fig. 1. Changes in the global average surface temperature (K) with respect to the pre-industrial level in experiments according to the SSP1-2.6 (triangles), SSP2-4.5 (squares), SSP3-7.0 (crosses), and SSP5-8.5 (circles) scenarios.

five experiments under the SSP3-7.0 scenario, as well as the temperature of the warmest month over these 30 years. The increase in the summer average temperature is the strongest in southern Europe due more intense soil drying and weaker evaporation. Thirty-year changes in the temperature in the warmest month correspond to an increase in the summer average temperature on the largest area of the globe. However, extremely high temperatures rise faster than the averages in most of Russia (Fig. 3). The probable reason discussed, e.g., in [10] is that the soil usually remains moist, but it can dry sometimes, which, in combination with suitable dynamic conditions, gives a particularly high temperature.

The changes in precipitation in December−February and June−August for the SSP3-7.0 scenario averaged over five members of the ensemble are shown in Fig. 4. All members of the ensemble show an increase in precipitation in the winter in a significant part of middle and high latitudes. In summer, the border between the increase and decrease in precipitation in Eurasia passes mainly around or to the north of 60°. In southern and central Europe, all members of the ensemble show a decrease in precipitation. Precipitation also increases in the region of the summer Asian monsoon, over the equatorial Pacific, due to a decrease in the upwelling and an increase in ocean surface temperature (OST). The distribution of changes in precipi-
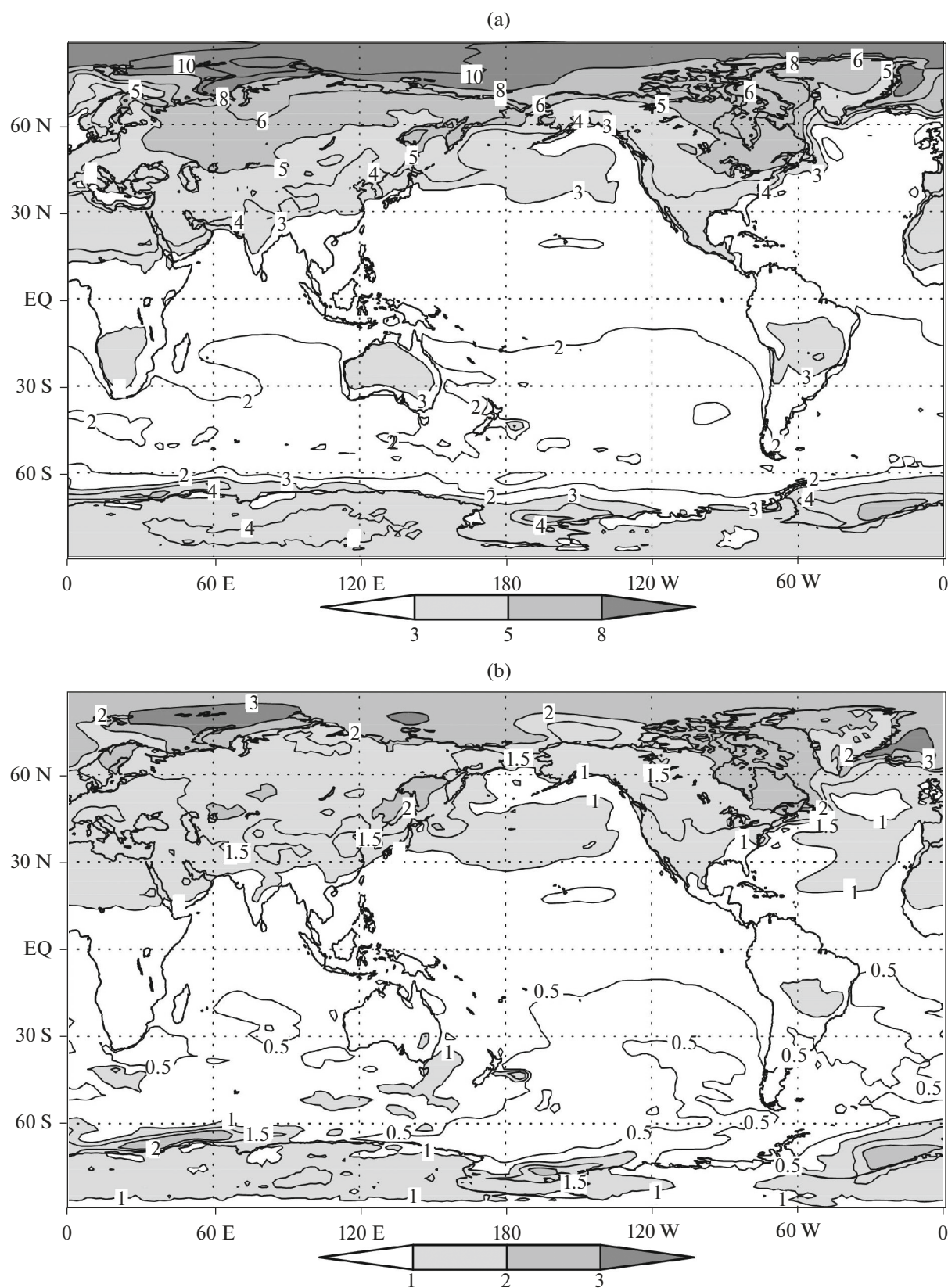
Fig. 2. Differences between the annual average surface air temperatures (K) in 2071−2100 and 1981−2010 for the (a) SSP5-8.5 and (b) SSP1-2.6 scenarios.
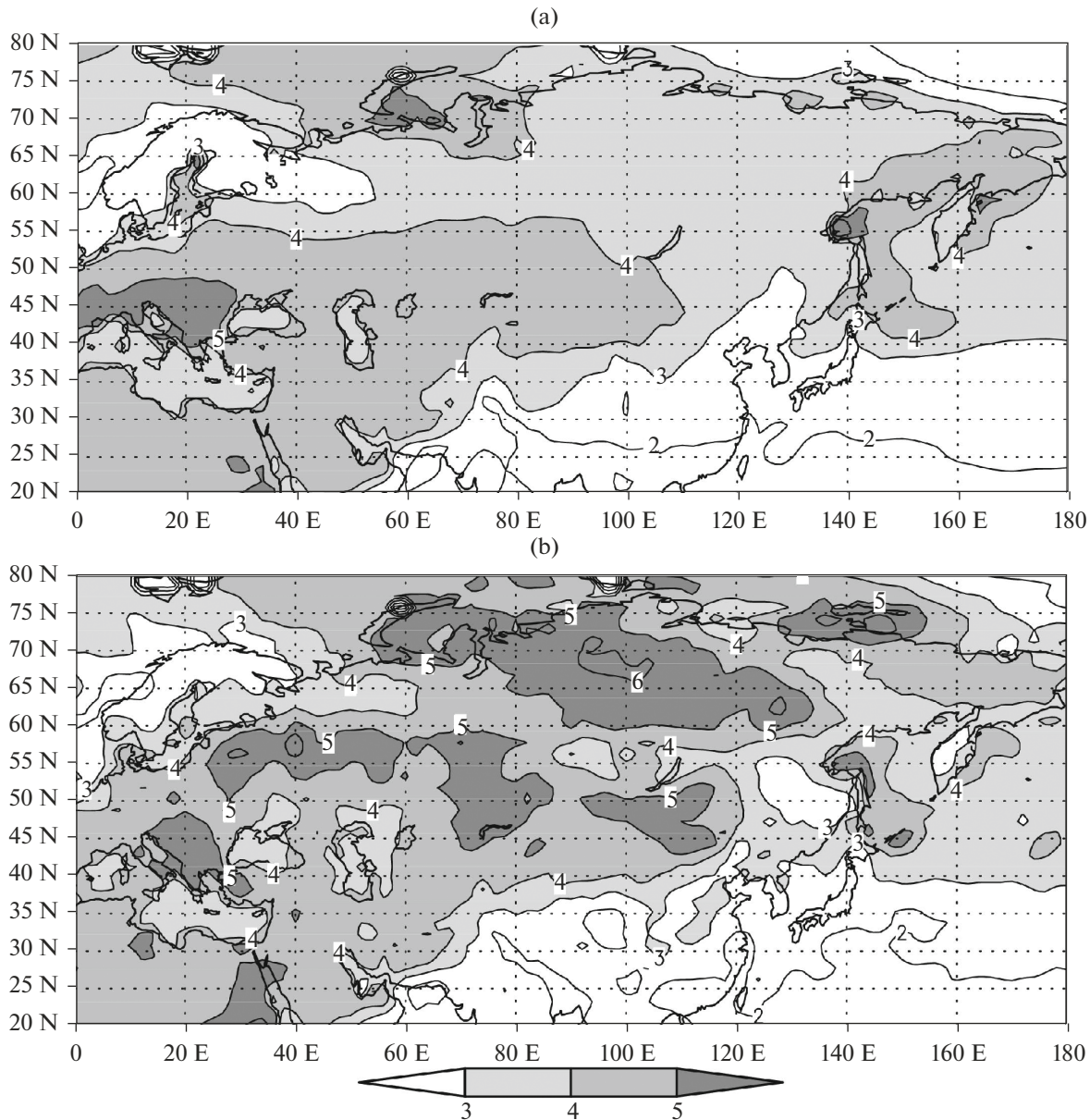
(a)

(b)

Fig. 3. Differences between the surface temperatures (K) in June−August of 2071−2100 and 1981−2010 for the (a) summer on the average and (b) the warmest months over 30 years.

tation mainly corresponds to that given in [6, Fig. 12.22] for all CMIP5 models.

The change in the Arctic sea ice area in September, when the ocean ice cover is minimal over the year, is of interest. Figure 5 shows the sea ice area in September 2015−2019 to be 4−6 million km² in all experiments, which corresponds to the estimate from observations in [11]. The Arctic sea ice does not completely melt in any of the experiments and under any scenario. However, according to [8, Figs. 12.28 and 12.31], many models participating in CMIP6, where the Arctic ice area is similar to that observed at the beginning of the 21st century, show the complete absence of ice by the end of the 21st century, especially under the RCP8.5

scenario, which is similar to SSP5-8.5. The reason for these differences is the lower equilibrium sensitivity of the INM-CM5 model. Note that the scatter of data between experiments under different scenarios in the first half of the 21st century is approximately the same as between different members of the ensemble under the SSP3-7.0 scenario and becomes larger only after 2070. The sea ice area values are sorted in accordance with the radiative forcing of the scenarios only after 2090. This indicates the large contribution of natural climate variability into the Arctic ice area. In the SSP1-2.6 experiment, the Arctic ice area at the end of the 21st century approximately corresponds to its area at the beginning of the experiment.
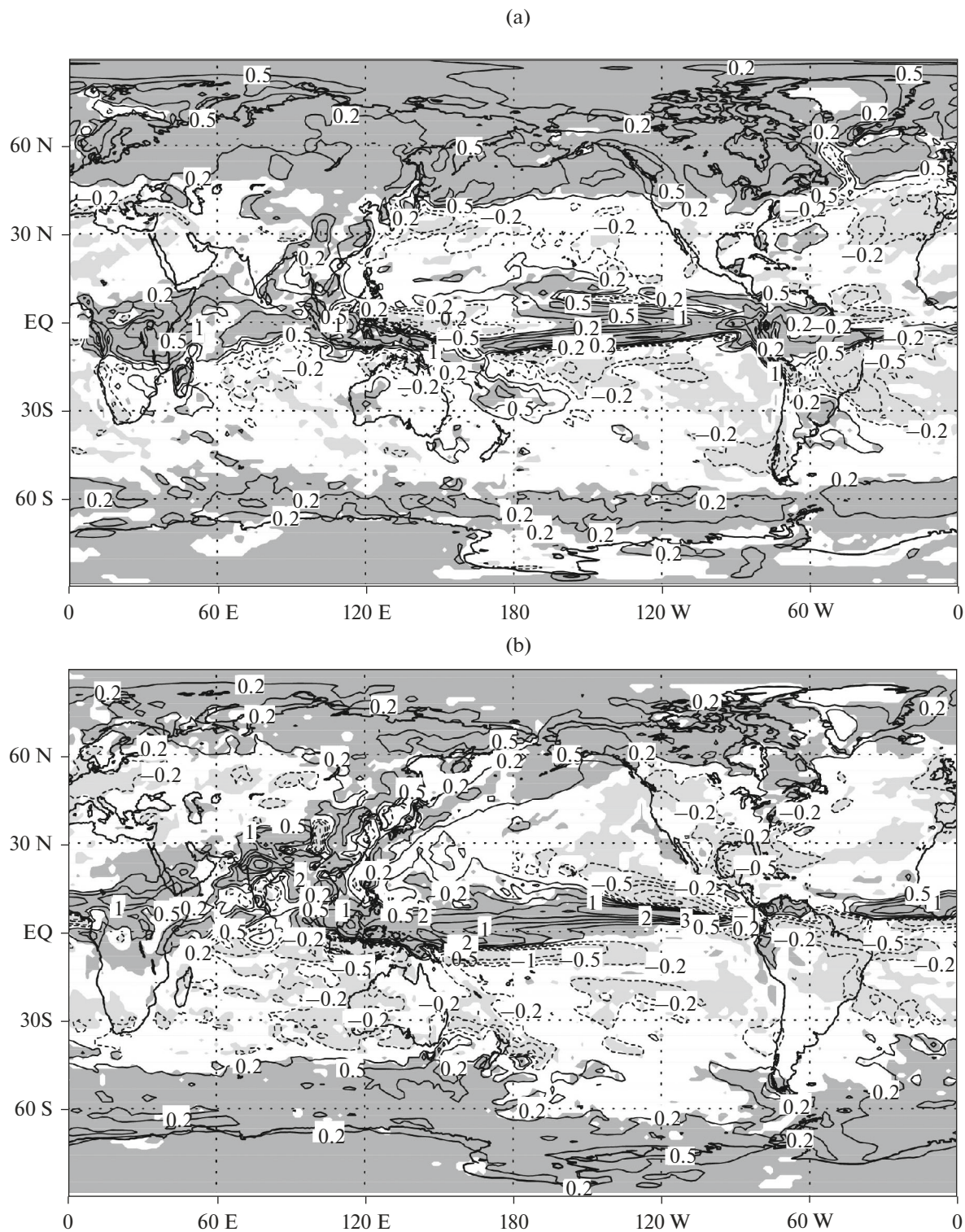
(a)

(b)

Fig. 4. Changes in precipitation (mm/day) in (a) December−February and (b) June−August 2071−2100 and 1981−2010 averaged over five SSP3-7.0 scenario experiments. Negative isolines are shown by the dashed curve; the regions where all five experiments show changes of the same sign are gray.

Climate changes can be also traced in the ocean circulation. Figure 6 shows the change in the 5-year averaged intensity of the Atlantic meridional circulation, defined as the maximum of the meridional streamfunction at 32° N. All experiments show a decrease in the intensity of meridional circulation in the 21st century
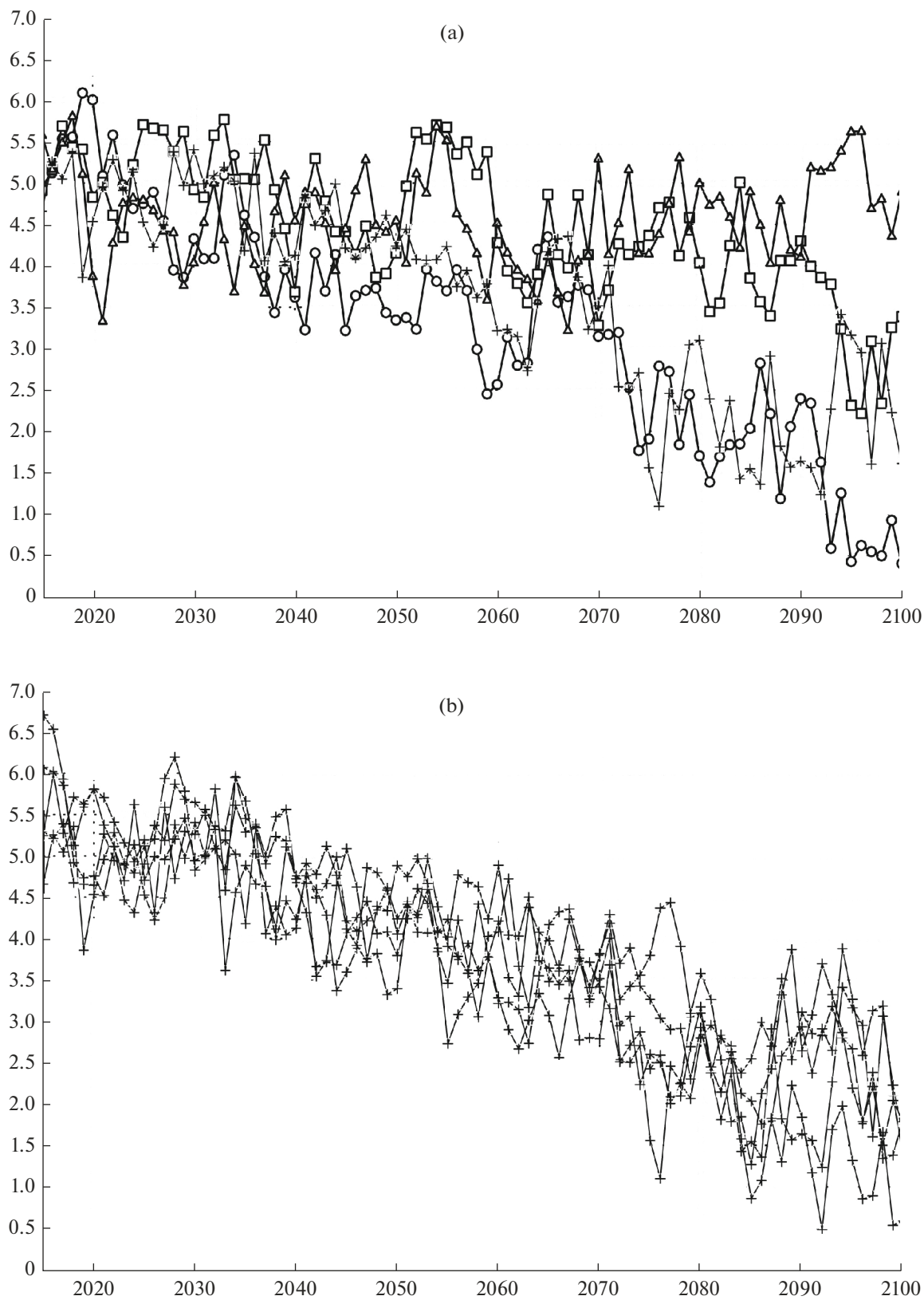
Fig. 5. Arctic ice area (millions km^2) in September: (a) data from experiments under the SSP1-2.6 (triangles), SSP2-4.5 (squares), and SSP5-8.5 (circles) scenarios and the first member of the SSP3-7.0 ensemble (crosses); (b) data of five SSP3-7.0 experiments.
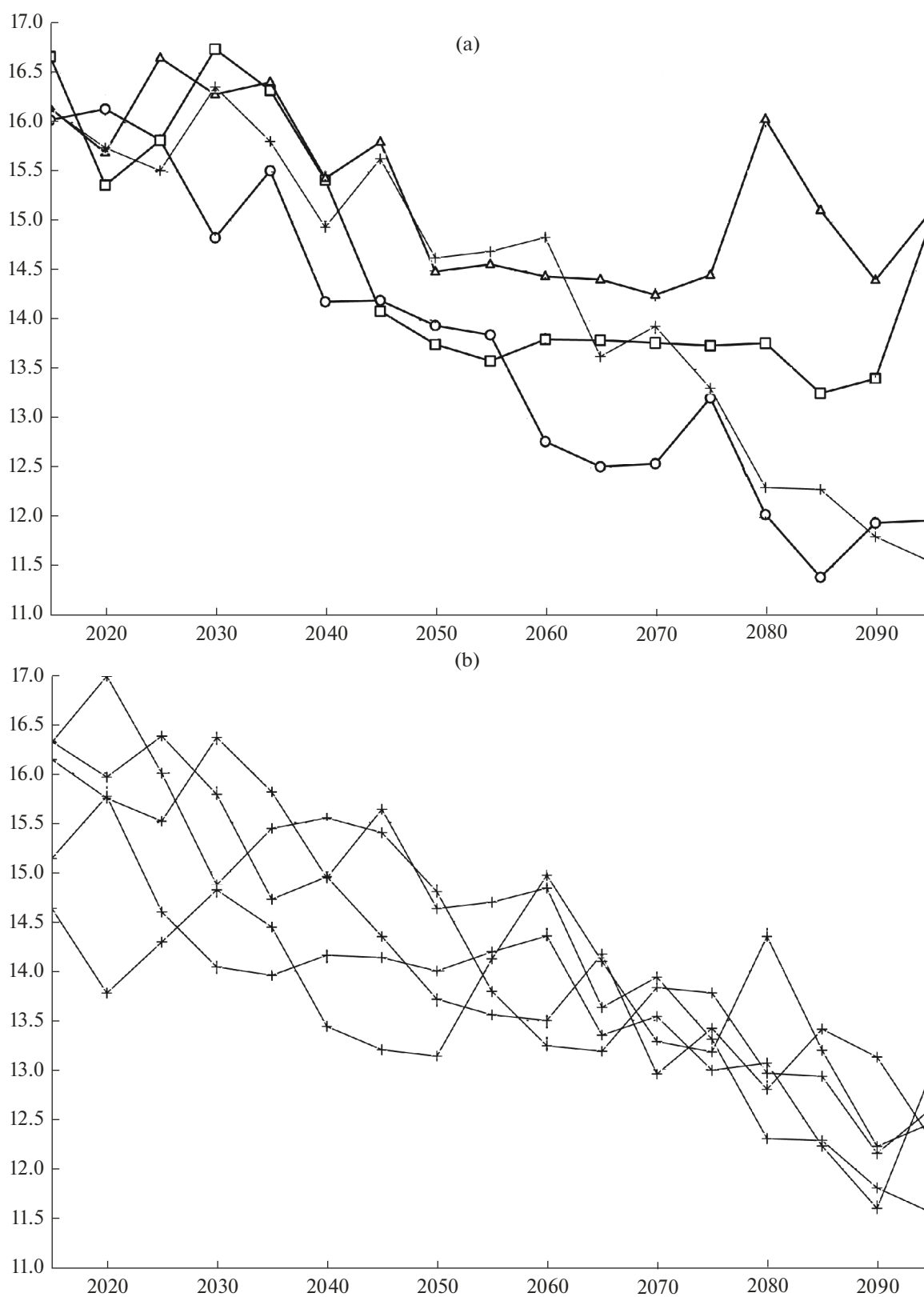
Fig. 6. Five-year averaged intensity of the Atlantic meridional circulation (Sv) defined as the maximum of the meridional stream-function at 32° N: (a) data from experiments under the SSP1-2.6 (triangles), SSP2-4.5 (squares), and SSP5-8.5 (circles) scenarios and the first member of the SSP3-7.0 ensemble (crosses); (b) data from five SSP3-7.0 experiments.
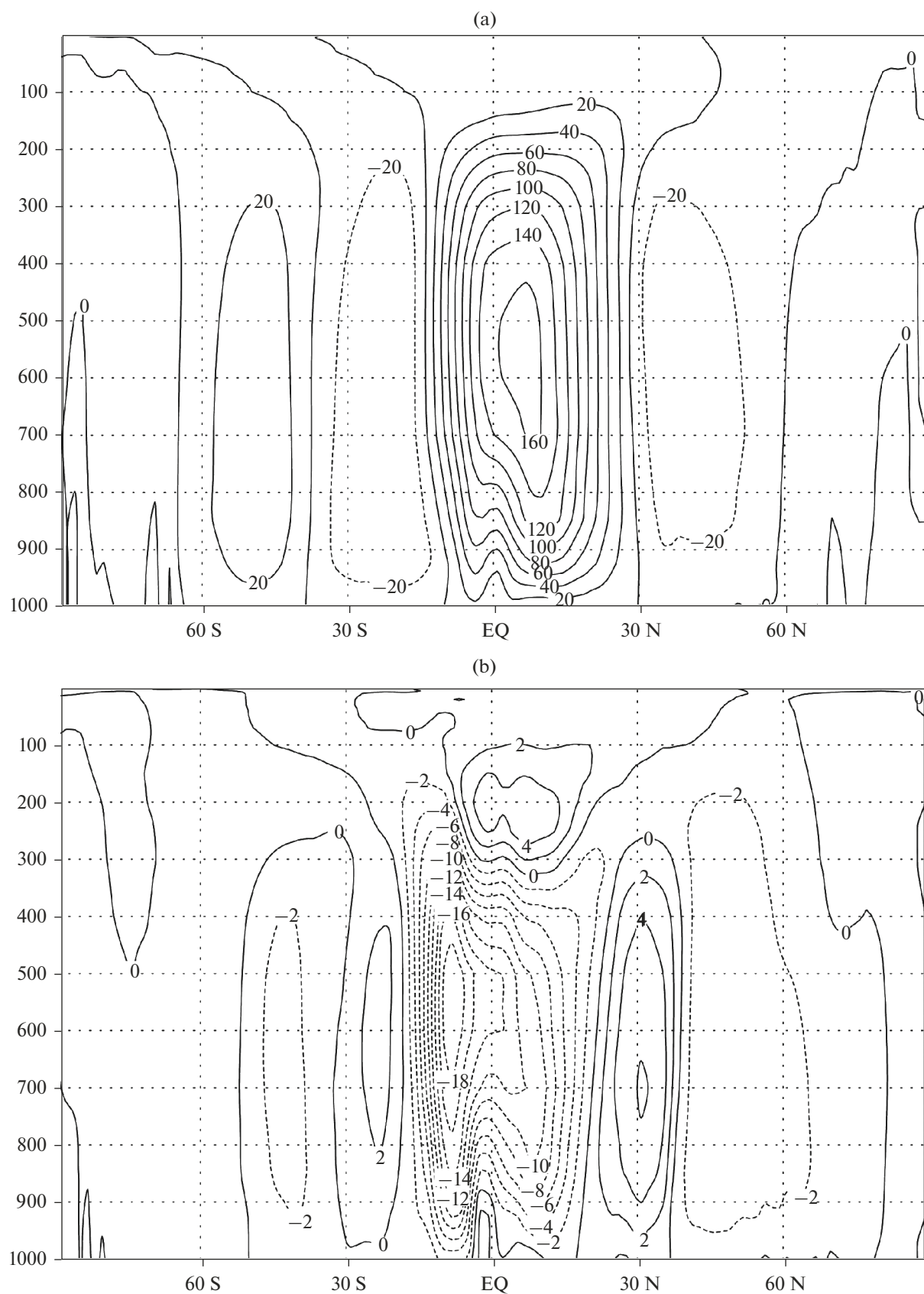
Fig. 7. (a) Meridional streamfunction (10^9 kg s^{-1}) in the atmosphere averaged over December–February 1981–2010; (b) its change in 2071–2100 as compared to 1981–2010.

and natural fluctuations against this decrease. The decrease is about 4.5−5 Sv for the SSP5-8.5 scenario, which is close to values obtained in the CMIP5 models [8, Fig. 12.35] under the RCP8.5 scenario. Under milder scenarios, the weakening of the meridional circulation is less pronounced. The reason for this weakening of the meridional circulation in the Atlantic, as far as we know, is not yet fully understood. In the first half of the 21st century, the scatter of the experimental data for different scenarios, which start with the same initial conditions, is less than the scatter of data of the SSP3-7.0 scenario experimental ensemble, different members of which start from different initial data. This may be evidence of the potential predictability of the meridional circulation intensity for several tens of years from the initial data. At the same time, the scatter of the SSP3-7.0 scenario ensemble significantly decreases by the end of the 21st century, which may mean a decrease in the natural variability in the warmer climate.

Let us mention several other changes in the ocean dynamics which occur by the end of the 21st century, as opposed to the end of the 20th century, in all scenario experiments except for SSP1-2.6.

First, we should note a decrease in the upwelling intensity at the equator in the Pacific, which might be due to a decrease in the trade wind velocity. This increases the OST more strongly than in latitudes farther from the equator and precipitation near the equator.

The Atlantic water inflow into the Arctic Ocean along the west coast of Europe also increases, as does the Arctic water flow into the North Atlantic along the east coast of Greenland. This can be due to both wind exposure (an increase in the North Atlantic Oscillation Index) and a change in the density gradients. In the Atlantic, the surface salinity significantly (by 0.5−1 PSU) increases from the equator to 40° N and similarly decreases in midlatitudes (40°−65° N). It again creases near the ice border in the north of the Barents Sea. These changes in salinity in the model are similar to those in CMIP5 models [8, Fig. 12.34].

The Gulf Stream intensity decreases in the Gulf of Mexico and near the coast of Florida. The Gulf Stream separation point shifts northward. The Kuroshio and the Antarctic circumpolar current intensify.

All these changes in the ocean dynamics during the development of the greenhouse effect, as far as the authors can tell, have no clear explanation in the modern scientific literature and need further research.

The dynamics of the atmosphere also slightly changes due to enhancement of the greenhouse effect. Here, we consider only changes in the meridional circulation during the winter of the Northern Hemisphere (December−February), when these changes are the most pronounced (Fig. 7). The meridional streamfunction shown in Fig. 7 is defined so there is a clockwise movement around the maximum, and counterclockwise around the minimum. In the upper fragment, which shows the intermediate state, the Hadley and Ferrel cells of the Northern and Southern hemispheres are clearly visible. The intensity of the Hadley circulation decreases in both hemispheres in the end of the 21st century, apparently due to an increase in the air moisture content at higher temperatures, which is faster than the increase in precipitation. Therefore, the vertical air stream, which accompanies precipitations, decreases. At the same time, as follows from Fig. 7, the Hadley cell extends upward. In addition, the Hadley circulation cell expands poleward, especially in the Northern Hemisphere. If the zero isoline of the streamfunction is considered its boundary, then the northern boundary shifts poleward by about 1° latitude. The probable cause of this is a change (northward shift) of the momentum fluxes generated by midlatitude Rossby waves, which support the Ferrel circulation. However, to prove this, additional studies are required. The Ferrel cell also shifts northward and slightly strengthens.

The dynamics of the stratosphere also changes. Like in other CMIP5 models ([8, Fig. 12.19]), west winds increase in midlatitudes due to warming of the troposphere and cooling of the stratosphere and a higher tropopause in the tropics than in midlatitudes. This intensifies the vertical wave propagation in midlatitudes and the Brewer−Dobson circulation.

CONCLUSIONS

Numerical experiments have been carried out to reproduce climate changes in the 21st century according to four scenarios of the CMIP6 program [4, 5], including an ensemble of five experiments under the SSP3-7.0 scenario. The changes in the global mean surface temperature are analyzed. It is shown that the global warming predicted by the INM-CM5 model is the lowest among the currently published CMIP6 model data. The geographical distribution of changes in the temperature and precipitation is considered. According to the model, the temperature in the warmest summer month will increase faster than the summer average temperature in Russia.

None of the experiments show the complete melting of the Arctic ice cover by the end of the 21st century. Some changes in the ocean dynamics, including the flow velocity and the meridional streamfunction, are analyzed. The changes in the Hadley and Ferrel circulation in the atmosphere are considered.

ACKNOWLEDGMENTS

REFERENCES

1. E. M. Volodin, E. V. Mortikov, S. V. Kostrykin, V. Ya. Galin, V. N. Lykosov, A. S. Gritsun, N. A. Diansky, A. V. Gusev, and N. G. Yakovlev, "Simulation of modern climate with the new version of the INM RAS climate model," Izv., Atmos. Oceanic Phys. **53,** 142−155 (2017).

2. E. M. Volodin, E. V. Mortikov, S. V. Kostrykin, V. Y. Galin, V. N. Lykosov, A. S. Gritsun, N. A. Diansky, A. V. Gusev, and N. G. Iakovlev, "Simulation of the present-day climate with the climate model INMCM5," Clim. Dyn. **49** (11−12), 3715−3734 (2017).

3. E. M. Volodin, N. A. Diansky, and A. V. Gusev, "Simulation and prediction of climate changes in the 19th to 21st centuries with the Institute of Numerical Mathematics, Russian Academy of Sciences, model of the Earth's climate system," Izv., Atmos. Oceanic Phys. **49** (4), 347−366 (2013).

4. B. C. O' Neill, C. Tebaldi, D. P. van Vuuren, V. Eyring, P. Friedlingstein, G. Hurtt, R. Knutti, E. Kriegler, J.-F. Lamarque, J. Lowe, G. A. Meehl, R. Moss, K. Riahi, and B. M. Sanderson, "The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6," Geosci. Model Dev. **9,** 3461−3482 (2016).

5. V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization," Geosci. Model Dev. **9,** 1937−1958 (2016).

6. E. Volodin and A. Gritsun, "Simulation of observed climate changes in 1850−2014 with climate model INM-CM5," Earth Syst. Dyn. **9,** 1235−1242 (2018).

7. E. M. Volodin and A. S. Gritsun, "Nature of the decrease in global warming at the beginning of the 21st century," Dokl. Earth Sci. **482,** 1221−1224 (2018).

8. M. Collins, R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W. J. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A. J. Weaver, and M. Wehner, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change,* Ed. by T. F. Stocker, D. Qin, G.- K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley (Cambridge University Press, Cambridge, UK and New York, NY, USA, 2013).

9. E. M. Volodin and M. A. Tarasevich, "Simulation of climate and weather extreme indices with the INM-CM5 climate model," Russ. Meteorol. Hydrol. **43,** 756−762 (2018).

10. E. M. Volodin and A. Y. Yurova, "Summer temperature standard deviation, skewness and strong positive temperature anomalies in the present-day climate and under global warming conditions," Clim. Dyn. **40** (5−6), 1387−1398 (2013).

11. J. C. Comiso and F. Nishio, "Trends in the sea ice cover using enhanced and compatible AMSR-E, SSM/I, and SMMR data," J. Geophys. Res.-Oceans **113** (C2), S07 (2008).
https://doi.org/10.1029/2007JC004257

Translated by O. Ponomareva

Variability of Extra Tropical Atmospheric Circulation and Periodic Trajectories in Simplified Models of Atmospheric Dynamics

A. S. Gritsun*

Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, 119333 Russia
e-mail: asgrit@mail.ru

Abstract—The possibility for approximate basic statistical characteristics of atmospheric models using periodic trajectories (closed solutions of equations of dynamics) is considered. The possibility of this approximation is based on ideas of the theory of dynamical systems which states that, in some important particular cases (e.g., for hyperbolic systems), the periodic trajectories define the invariant system measure related to a notion of the system climate. It is shown that this approximation also is possible in the case of the atmospheric models under consideration. Moreover, the principal modes of circulation variability are implemented as clusters of periodic orbits oriented along the leading eigenvectors of dynamical operators of models linearized with respect to their mean states. The analysis of observational data shows that probably the same conclusion could also be made for the real atmospheric circulation.

INTRODUCTION

Studying the nature of the variability in the midlatitude atmosphere has been a very popular scientific topic since the 1940s. The most popular mathematical approaches used to analyze the problem include the construction of reduced models of atmospheric circulation, which describe the atmospheric circulation on the intraseasonal time scales with a different degree of accuracy (barotropic models of atmospheric circulation, two- and three-layer quasi-geostrophic, and primitive equations). Using the linearization of the given equations with respect of the mean (modeled or really observed) state of the atmosphere, the eigen vectors of linearized operator can be related to the leading modes of variability (the so-called natural orthogonal functions (NOFs)) of fields of atmospheric characteristics (stream functions, surface pressure, geopotential, or velocity) [1–6]. Another approach is based on the empiric approach and reduced to using the data of observations to reconstruct the linear operator of evolution, which generates the observed atmosphere dynamics, with the subsequent analysis of the structure of its eigenvectors [7–11]. For lower frequency circulation components (on scales from a week to a month), one popular approach is the construction of the so-called circulation regimes—atmospheric states (e.g., "zonal" regime and the regime of "blocking")—the probability of finding the system in the vicinity of which is maximum [12–15]. The system dynamics in this case is transitions between the states. This transition is closely connected with the atmosphere dynamics approximation using the low-component Markov chains [16–20].

All these methods have their merits and demerits, leaving a gap for further studies. In this work, we consider one more, alternative, approach, which is, in a certain sense, a generalization of "the theory of circulation regimes": instead of stationary (quasi-stationary) states of the atmosphere for approximating its dynamics, more general sets are used: periodic trajectories. Behind the mathematical base of the method, the statement of theory of dynamical systems (proved for the so-called hyperbolic dynamical systems) is that any trajectory of the system (and any of its statistical characteristics) can be approximated using periodic solutions of the system: the system trajectories of a special form returning into the initial state for a finite time (period of orbit) [21–24]. In this case, statistical characteristics of the system are calculated using the weighted averages of the relevant orbit characteristics, while the weighted coefficients are determined through the orbit-instability characteristics used in averaging [24]. As a result of using this approach, both the individual statistical characteristics (mean state, variance, and variability modes) and the invariant measure itself of the system under study, which is related to the climate notion, can be approximated with a good accuracy [25]. In the case of models of atmosphere dynamics, the evidential ground of the method does not exist; however, in a number of works

(e.g., [26]), it is assumed that, with a calculation of macroscopic characteristics of the chaotic systems with a large number of degrees of freedom, it can be considered hyperbolic (the so-called "chaotic hypothesis"). In respect to models of atmosphere dynamics, a problem of searching for periodic trajectories is nontrivial, since it is reduced to solving the strongly nonlinear system of differential equations of high-dimensionality (equal to the dimensionality of a phase space of the system) with a poor initial condition. The issue of the existence of periodic solutions in the phase space of atmosphere models, the presence of links between the characteristic circulation regimes and periodic orbits, and the possibility of circulation approximation using periodic motions were studied through the example of models of the barotropic dynamics of atmosphere (ocean) in [27−32]. In this work we will consider in more detail the possibility of using periodic trajectories to approximate statistical characteristics of models of the atmosphere dynamics by also examining the quasi-geostrophic model of atmosphere (the so-called Marshall−Molteni model [33]) and using a fuller set of orbits for analysis. In particular, the conclusions on the structure of leading variability modes and their relation to regular period motions will be drawn.

This work is constructed as follows. In the second chapter, the models under study are formulated and their statistics are analyzed in comparison with observation data. In the third chapter, a technique for constructing periodic trajectories is considered and the results of approximating the model statistics using periodic orbits are given. In the fourth chapter, the specific features of the circulation in the plane of leading eigen modes of the models are considered. In the Conclusions, the work results are discussed.

MODELS OF ATMOSPHERE DYNAMICS

The barotropic atmospheric model is based on the Navier−Stokes two-dimensional equation in spherical geometry in terms of a stream function ψ (in approximation of antisymmetry with respect to the equator), taking into account the effects of rotation and friction on the underlying surface:

$$\frac{\partial \Delta \psi}{\partial t} + J\left(\psi, \Delta \psi + l + kh\right)$$
$$= -\alpha \Delta \psi + \mu \Delta^2 \psi + f_{\text{ext}}, \quad \psi(0) = \psi_0. \tag{1}$$

Here, ψ is the dimensionless stream function; the Coriolis parameter is denoted as l; the orography and its normalization are designated as h and k, respectively; external forcing is denoted f_{ext}; coefficients of turbulent viscosity and friction in the boundary layer are labelled μ and α, respectively; and the Laplace operator on the sphere is marked by the symbol Δ and the Jacobi operator by J.

The equation is solved using the Galerkin method (basis functions are the spherical harmonics, antisym-

metric with respect to the equator) over the space. Two spatial resolutions are used (T12 and T21), which lead to the system of ordinary differential equations with the dimensionality of 78 and 231. For time approximation, the "midpoint" scheme is used (X and τ are the right-hand side of the system and the time step):

$$\psi^{j+1} = \psi^j + \tau X\left(\psi^j + \frac{\tau X\left(\psi^j\right)}{2}\right).$$

The model parameters were specified as follows. The orography was obtained from real data of the Earth surface height in the Northern Hemisphere (antisymmetrically continued into the Southern Hemisphere) using the expansion in a Fourier series in terms of the basis functions of the Galerkin method. The normalizing factor k is 0.14, characteristic times corresponding to boundary layer friction and turbulent viscosity (for smallest scale) гармоники) are 25 and 5 days respectively. The external forcing was chosen such that a mean state of the system was close to a mean state of the real atmosphere in January−February for the field of the function calculated by the NCEP/NCAR reanalysis data.

The technique of external-forcing selection is substituting the real system trajectory (in our case, the NCEP/NCAR data) to the model equation and calculating the residual summand, which, after time averaging, gives the sought-after external forcing f_{ext}:

$$f_{\text{ext}} = P_N$$
$$\times \left\{ J\left(\psi_c, \Delta \psi_c + l + kH\right) + \alpha \Delta \psi_c - \mu \Delta^2 \psi_c \right\}.$$

Here, ψ_c is the stream function from the NCEP/NCAR data array (for a 250- or 300-millibar surface for the model of the coarse (T12) and "high" (T21) spatial resolutions); P_N is the operator of projection onto the system phase space. The description of this system in more detail can be found in [34, 35]. In the model construction, the archive data for December−January−February of 1979−2009 were used.

The linearization of the equation with respect to the mean state $\overline{\psi}$ has the following form:

$$\frac{\partial \psi'}{\partial t} = -\Delta^{-1}\left(J\left(\psi', \Delta \overline{\psi} + l + kh\right) + J\left(\overline{\psi}, \Delta \psi'\right)\right)$$
$$- \alpha \psi' + \mu \Delta \psi' \equiv A\psi', \psi'(0) = \psi_0'. \tag{2}$$

Quasi-geostrophic baroclinic model of the atmosphere [33] includes three equations for the evolution of the potential vorticity q_j (in approximation of antisymmetry with respect to equator) at the levels of 200, 500, and 800 mb (indices 1, 2, and 3, respectively). In this case, the effects of rotation and interaction

between levels and friction on the underlying surface are taken into consideration:

$$\frac{\partial q_j}{\partial t} + J\left(\psi_j, q_j\right) = -D_j + F_j,$$

$$\psi_j(0) = \psi_{j0}, \quad j = 1, 2, 3. \tag{3}$$

Expressions for the potential vorticity have the following form:

$$q_1 = \Delta\psi_1 - (\psi_1 - \psi_2)/R_1^2 + l,$$

$$q_2 = \Delta\psi_2 + (\psi_1 - \psi_2)/R_1^2 - (\psi_2 - \psi_3)/R_2^2 + l, \tag{4}$$

$$q_3 = \Delta\psi_3 + (\psi_2 - \psi_3)/R_2^2 + l(1 + H/H_0).$$

Here R_1 and R_2 are the Rossby deformation radii for the relevant levels with values 700 and 450 km respectively, ψ is the stream function, l is the Coriolis parameter, and H and H_0 are the orography and its maximum value. The dissipation terms D_j are defined by the expressions

$$-D_1 = (\psi_1 - \psi_2)/(\tau_R R_1^2) - R^8 \Delta^4 \dot{q}_1 / (\tau_H \lambda_{max}^4),$$

$$-D_2 = -(\psi_1 - \psi_2)/(\tau_R R_1^2) + (\psi_2 - \psi_3)/(\tau_R R_2^2)$$
$$- R^8 \Delta^4 \dot{q}_2 / (\tau_H \lambda_{max}^4), \tag{5}$$

$$-D_3 = -(\psi_2 - \psi_3)/(\tau_R R_2^2)$$
$$- EK_3 - R^8 \Delta^4 \dot{q}_3 / (\tau_H \lambda_{max}^4).$$

Here, τ_R denotes the relaxation time of vertical temperature profile (30 days), τ_H is the time scale of hyper diffusion (1.33 days), R is the Earth's radius, and λ_{max} is the absolute value of the largest eigenvalue of the Laplace operator in the model (for the T18 resolution, it is $\lambda_{max} = 18 \times 19$). Further, $EK_3 = \nabla(k_{surf}\nabla\psi_3)$ is the operator describing friction on the surface; $k_{surf} = (1 + 0.5LS + 0.5FH(h))/\tau_E$, where LS is the fraction of the Earth surface in the grid cell, $FH(h) = 1 - \exp(-h/1000m)$, and τ_E is the time scale of the Ekman friction (1.5 days). $\dot{q}_i = q_i - f$, $i = 1, 2$, and $\dot{q}_3 = q_3 - f(1 + h/H_0)$.

The procedure for selecting the right-hand side is similar to the procedure used for the barotropic model. For each equation, F_j is determined by the substitution of stream function data on the 200-, 500-, and 800-mb surfaces from the NCEP/NCAR reanalysis archive and by the corresponding averaging:

$$F_j = \overline{J\left(\psi_{Cj}, q_{Cj}\right) + D_{Cj}}.$$

Figure 1 presents statistics of the models under study in comparison with the observation data. The first row of panels presents the mean states of low-resolution barotropic model, the T21 barotropic model, the Marshal-Molteni model (level 1, corresponding to 200 mb) and stream functions on the 250-mb surface

from the NCEP/NCAR archive data. In the second row of panels, there are fields of standard deviations (in the NCEP/NCAR data, frequencies shorter than 3 days and longer than 90 days were filtered off preliminarily).

We note that, with the external-forcing construction for the T12 model, the 250-mb stream-function (for model T21, the 300-mb stream-function) data were used; therefore its mean state is the closest to the observations. Additionally, it is seen that all models understate a variance of the stream-function field; in this case the baroclinic model is most realistic. In the third and fourth rows in Fig. 1, the first and second leading natural orthogonal functions (NOF) of models and data are shown. During the NOF construction, filtration with a window of 3−90 days is conducted for the data again. It can be seen that the first NOF of the data, identified as the circular mode, is well reproduced by the barotropic models. The second NOF of the field of observations, which has a structure, similar to PNA, is close to the first NOF of the baroclinic model and has a certain similarity to the second NOFs of barotropic models. The second NOF of the baroclinic model is identified as wave three, also separated in the analysis of observation data. The absence of a circular mode in the leading NOF1 and NOF2 of the baroclinic model is associated with the fact that its data are not filtered and contain a significant fraction of the short-period baroclinic variability (this is necessary for unifying the analysis of models in the next chapters).

In the following, we consider a structure of linear modes of circulation on the level of 250 mb. As was shown in [36], the linearization of the barotropic vortex equation with respect to the mean state of the atmosphere on this level provides the highest values of correlation between leading eigenvectors and leading variability modes. Linear operator (2) has six pairs of unstable complex-valued eigenvectors whose periods are 12, 8, 5, 3, 90, and 30 days. In the first row of Fig. 2 (panels a1−c1), the imaginary parts of vectors, corresponding to periods of 8, 12, and 30 (eigenvectors 2, 1, and 6), are shown. The second and first vectors (a1 and a2) have a structure of waves six and five propagating from west to east; in the structure of the sixth vector (a3), a structure of the PNA type and (on the north) wave one are present. We note that the linearization of Eq. (2) with respect to the mean state of the real atmosphere (calculated from the NCEP/NCAR archive) gives similar results (all abovementioned modes are separated), with the only difference that, among the leading vectors, higher frequency modes with periods of 2 and 3 days and wavenumbers 7 and 8 also appear.

Regular oscillation processes are separated best of all using the construction of the complex-valued NOFs. The procedure for calculating the Hilbert EOFs for the system trajectory is as follows (see [37]). At first, a Fourier transform with respect to time is performed for each component of the physical field
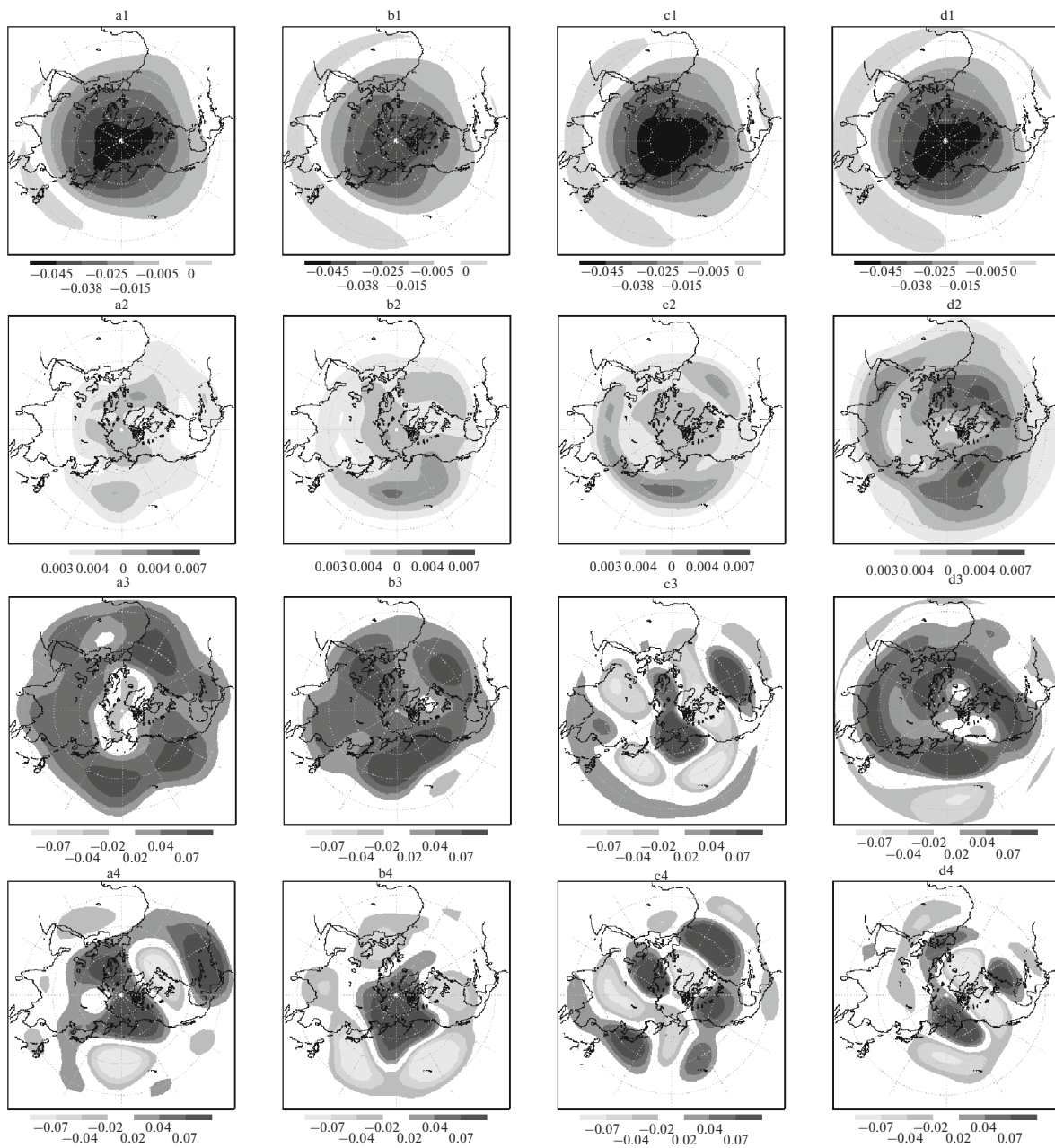
Fig. 1. Statistical characteristics of models and NCEP/NCAR reanalysis data: mean state (stream function), standard deviation (stream function), first and second NOFs of (a1–a4) T12 barotropic model, (b1–b4) T21 barotropic model, (c1–c4) Marshall–Molteni model, and (d1–d4) NCEP/NCAR reanalysis data. Stream function of the 250-mb surface, December–January–February of 1979–2009. Dimensional values (m^2/s) are obtained using multiplication by 2.97×10^9.

under study. Then, in the expansion, its shift by a quarter-period (90°) and the inverse Fourier transform are performed. As a result, we obtain the initial time series $\psi(j)$ and its Hilbert transformation $\tilde{\psi}(j)$. Further, it is necessary to compose a complex-valued variable $z(j) = \psi(j) + i\tilde{\psi}(j)$ and calculate the covariance matrix for it. The complex-valued eigenvectors of this matrix are the Hilbert (or complex) EOFs of the system, which determine the two-dimensional planes with a largest variability of the solution. Figure 2 shows

a result of a complex NOF analysis of the data of barotropic and baroclinic models, to which the filtration with a window of 7–9, 10–15, and 18–30 days is applied (only one component of the complex NOF, having the largest correlation with the corresponding linear mode, is given). The low-resolution barotropic model (the second row, panels a2–c2) poorly resolves the high-frequency modes; therefore, linear mode 2 is implemented only in the form of the 9th NOF in order (Fig. 2, panel a2), while the first mode is implemented in the form of the 2nd NOF. The 30-day structure of
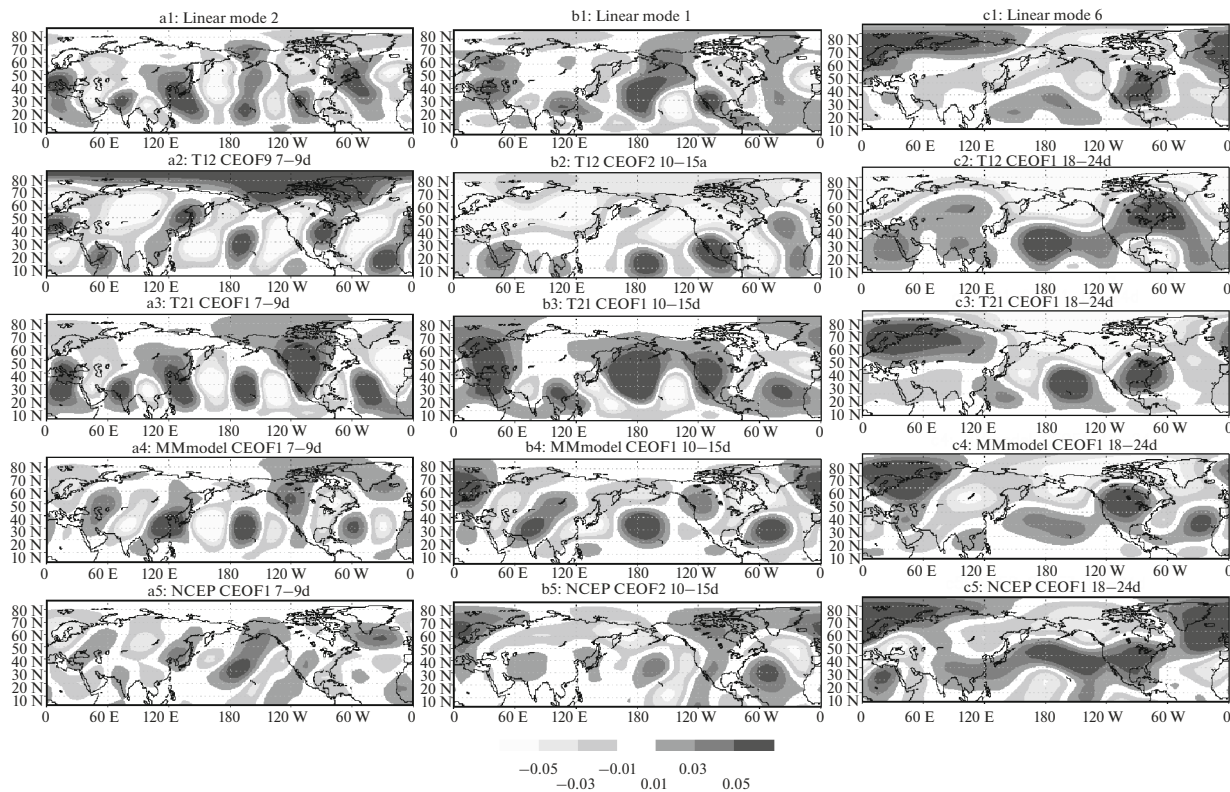
Fig. 2. Leading eigen modes of the (linearized with respect to the mean state) operator of the T21 barotropic model and the characteristic rotational modes of variability of models: (a1−c1) leading eigen modes (second, first, sixth) of the linear operator; (a2−c2) ninth, second, and first complex-valued NOFs of the T12 barotropic model for the filtration windows of 7−9, 10−15, and 18−24 days; (a3−c3) first complex-valued NOF of the T21 barotropic model; and (a4−c4) first complex-valued NOF for the Marshall−Molteni model. First, second, and third complex-valued NOFs by the NCEP data for filtration windows of 7−9, 10−15, and 18−24 days. Dimensionless units.

the 6th mode is separated in the first NOF. The structure of variability of high-resolution barotropic model is close to its eigen modes, all modes are separated in the form of leading complex NOFs (Fig. 2, panels a3−c3). The variability of the stream function of the upper level of the Marshall−Molteni (Fig. 2, panels a4−c4) is also close to eigen modes of the barotropic circulation, which are presented in the form of leading NOFs with a close structure. The real data have much richer structure of variability; however, the eigen modes of barotropic dynamics also are separated here (Fig. 2, panels a5−c5). Wave six is implemented in the form of a leading complex NOF; wave five is implemented in the form of the second NOF, while the structure wave one—PNA—is implemented in the form of the third NOF.

PERIODIC ORBITS OF THE SYSTEM AND ITS STATISTICS

By definition, a periodic orbit is a special trajectory of the system, which in the fixed time T (orbit period) returns to its initial position. It can be written symbolically as

$$\psi(T) \equiv S(T, \psi(0)) = \psi(0), \qquad (6)$$

where $S(T, \psi(0))$ is the nonlinear evolution operator, determining the state of the system with the initial state at the point $\psi(0)$ in the time T. The relationship defining the orbit is, in essence, a nonlinear system of N equations for $N + 1$ unknowns, since $\psi(T)$ nonlinearly depends on and is unambiguously determined by $\psi(0)$ and T. The system underdefiniteness is caused by the fact that the same orbit can be specified by different initial conditions, which are obtained from each other by a shift along the periodic trajectory. Different numerical methods can be used to solve the system of equations. In this work, we used the Newton method with step damping (see, e.g., [38]); in this case, it is necessary to extend the definition of the system, eliminating a freedom in choosing the initial condition. This can be done by adding to the equation system a phase condition, the meaning of which is that the search direction is selected to be orthogonal to the motion direction.

We illustrate in brief the numerical scheme of the method (details can be found in [29]). Let $\psi^{(k)}$ and $T^{(k)}$ k-th approximations be the approximations for $\psi(0)$ and T. We denote $\psi^{(k+1)} = \psi^{(k)} + h^{(k)}$ and

$T^{(k+1)} = T^{(k)} + \tau^{(k)}$. The standard iteration of the Newton method is based on expanding the equation

$$S\left(T^{(k+1)}, \psi^{(k+1)}\right) = \psi^{(k+1)}$$

to a Taylor series in terms of $h^{(k)}$ and $\tau^{(k)}$ with an accuracy of up to small quantities of first order. As a result, a linear system of equations is obtained for $h^{(k)}$ and $\tau^{(k)}$. That is,

$$0 = S\left(T^{(k+1)}, \psi^{(k+1)}\right) - \psi^{(k+1)} \approx S\left(T^{(k)}, \psi^{(k)}\right)$$
$$- \psi^{(k)} + \left(\frac{\partial S\left(T, \psi\right)}{\partial \psi}\bigg|_{(\psi, T) = \left(\psi^{(k)}, T^{(k)}\right)} - E\right) h^{(k)} \quad (7)$$
$$+ \frac{\partial S\left(T, \psi\right)}{\partial T}\bigg|_{(\psi, T) = \left(\psi^{(k)}, T^{(k)}\right)} \tau^{(k)} = 0.$$

Let us require the orthogonality of the search direction $h^{(k)}$ to the motion direction (the right-hand side of the system) in the current initial state:

$$\left(h^{(k)}, \mathrm{X}\left(\psi^{(k)}\right)\right) = 0. \quad (8)$$

The system of linear equations in the nondegenerate case unambiguously defines $h^{(k)}$ and $\tau^{(k)}$; in this case, the next approximation is calculated as

$$\psi^{(k+1)} = \psi^{(k)} + \eta h^{(k)}, \ T^{(k+1)} = T^{(k)} + \eta \tau^{(k)}.$$

The parameter η is unity in the standard Newton method. However, if the linear system is ill-conditioned, step $\left(h^{(k)}, \tau^{(k)}\right)$ may be large. In this case, the expansion to a Taylor series with an accuracy of up to terms of first order makes no sense, and the Newton method may diverge. As the results of the calculations show, for the problem under study, the step damping is a mandatory procedure. To choose the η value, the "line-search" technique, described in [38], was used in this work. We note that each iteration of the Newton method requires the integration of the linearized model in the full phase space.

To determine the average value of any statistical characteristic of the system $\Psi(\psi)$, the model is integrated for a long time; Ψ is estimated at each point of the obtained trajectory and is averaged over all time moments:

$$\langle\Psi\rangle = \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} \Psi(\psi(k)), \quad (9)$$

and, in this case, all $\psi(k)$ measurements are taken into account with the same weight equal to unity.

According to the theory of hyperbolic systems [24], by knowing a set of all periodic system trajectories with a period smaller than T, the average $\langle\Psi\rangle$ can be determined by calculating Ψ at points lying on periodic trajectories:

$$\langle\Psi\rangle = \lim_{T \to \infty} \frac{1}{W} \sum_{i=1}^{I(T)} w_i \Psi(\psi_i), \quad (10)$$

where ψ_i is the periodic point of the period $t_i < T$, $I(T)$ is the total number of such points, each orbit of the period t_i contains t_i/τ periodic points (τ is the time step in the model), w_i is the orbit weight (the same for all its periodic points), and W is the total weight of all periodic points. The orbit weight w_i is defined as

$$w_i = \exp\left(-t_i \sum_{m, \lambda(i,m) > 0} \lambda(i, m)\right), \quad (11)$$

where $\lambda(i, m)$ are the positive Lyapunov exponents of the orbit which corresponds to the ith periodic point. Thus, the formula makes the same sense as the standard definition of an average; only the value of the functional for a particular periodic point is taken into account with the weight proportional to characteristics of orbit instability (the phase-volume growth rate along its unstable directions), rather than with the same unit weight. The higher the growth rate is, the smaller the weight is. Weakly unstable orbits make a decisive contribution during the calculation of the average, which is justified on the grounds of physics, since the trajectory passes more time in the vicinity of these orbits.

Using the abovementioned techniques, a set of orbits was found for each model which will be used to calculate the statistics of the models together with the standard definition. For the T12 barotropic model, 2322 orbits were found; for its version of T21 resolution, 1851 orbits were found; and, for the Marshall—Molteni, 2711 orbits. We note that the validity of approximation is guaranteed by the theorem if the system under study is hyperbolic and if all orbits up to the orbits with sufficiently large periods are found. We certainly cannot guarantee these facts. Moreover, the number of orbits necessary for the correct reconstruction of the system trajectory exponentially depends on its complexity; therefore, it can be expected that, for the Marshall—Molteni model, having the largest dimension of the attractor and the highest rate of error growth, the tested approach will give the worst results.

The result of approximating the base statistical characteristics of the models under study from Fig. 1 (mean state, variability, first and second leading NOFs), using periodic trajectories and employing the presentations (10)−(11), is given in Fig. 3. It can be seen that characteristics of the T12 simplest barotropic model are reproduced exactly; in the case of the T21 model, the variance is understated and the second NOF is not reproduced as well. For the most elaborate Marshall—Molteni model, the variability is still more understated and the leading NOFs, having a characteristic time
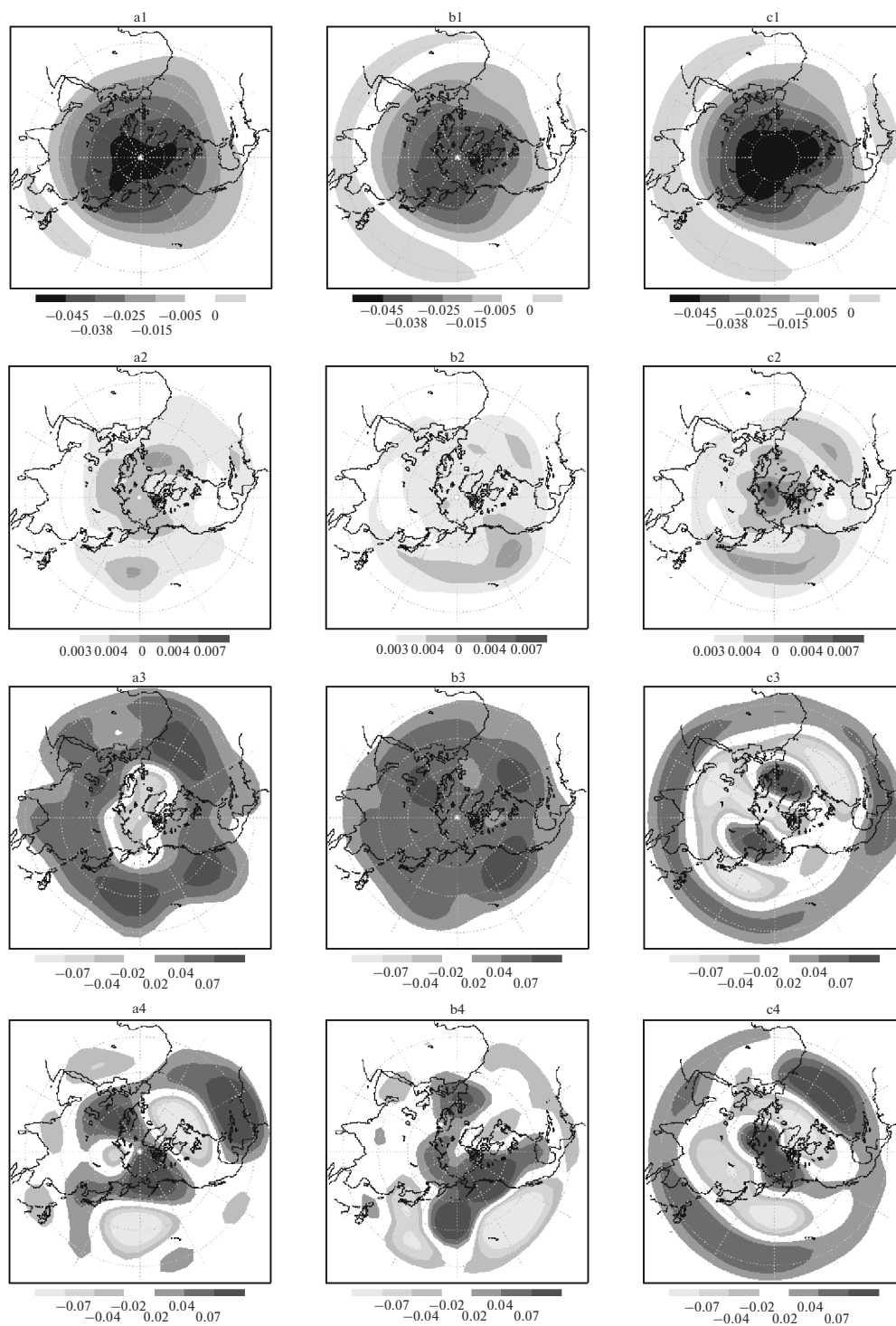
Fig. 3. Approximation of statistical characteristics of the models from Fig. 1 using periodic trajectories: mean state (stream function), standard deviation (stream function), and first and second NOFs for the (a1−a4) T12 barotropic model and (b1−b4) T21 barotropic model; (c1−c4) Marshall−Molteni model.

scale of about 20 days, are reproduced badly. This fact is explained by the situation that, for the Marshall−Molteni model, the number of orbits with periods of more than 10 days is small and the high-frequency modes dominate during NOF calculation.

CIRCULATION STRUCTURE IN THE PLANE OF LEADING NOFs

In the following, we consider in which way the circulation is arranged in the plane of eigen modes of the baro-
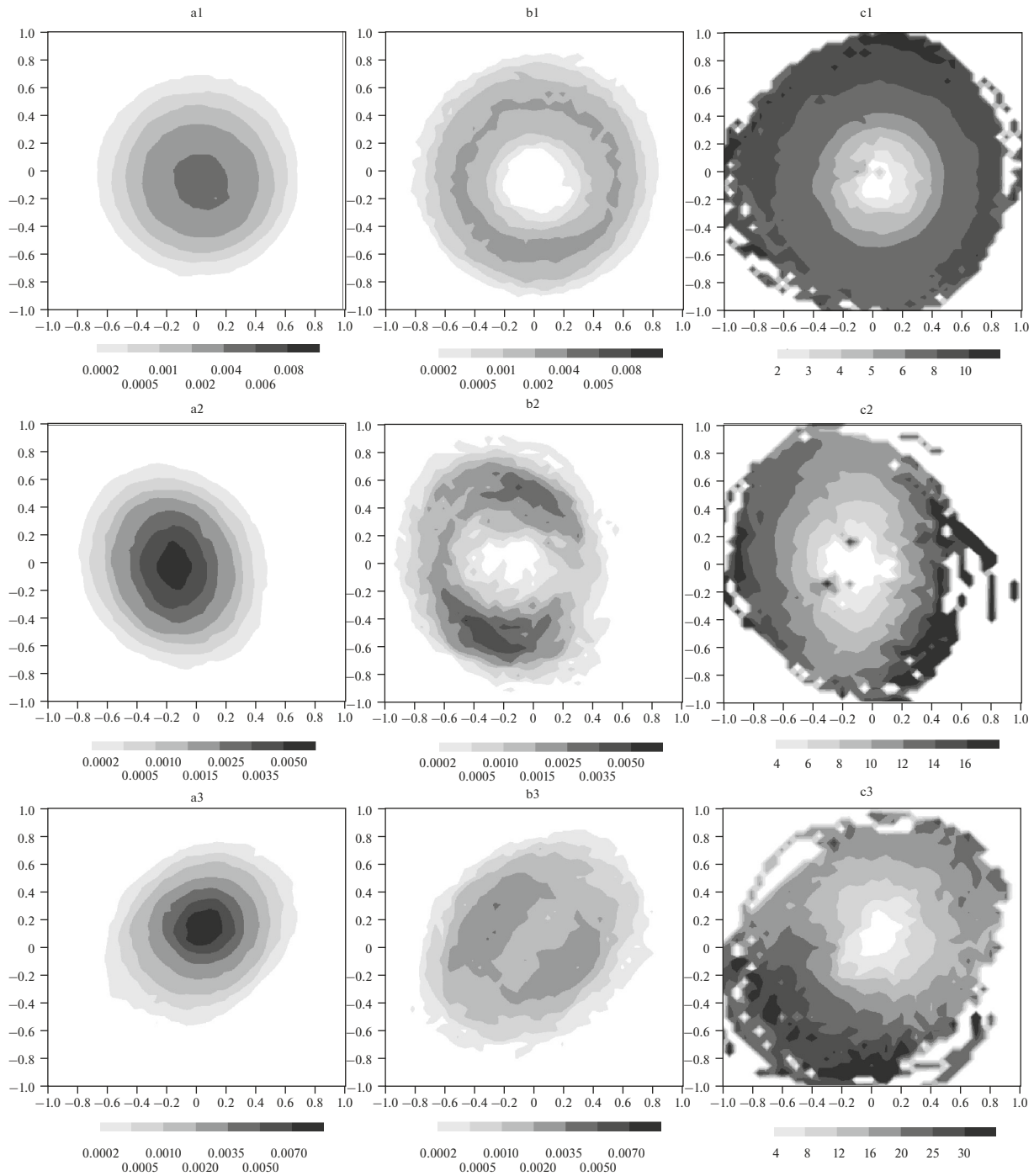
Fig. 4. Projections of the probability density for the plane of leading rotational modes of variability of (a1−a3) the T21 barotropic model, (b1−b3) the same for states whose velocities have a significant constituent along the projection plane, and (c1−c3) the characteristic period of revolution of the trajectory about the origin of coordinates.

tropic circulation. For this purpose, we project the trajectory of each model onto the relevant two-dimensional plane of the complex-valued NOF from Fig. 2. Thus, for each trajectory point, we obtain the two-component vector $(x_1(i), x_2(i)) \equiv r(i)$. A velocity of motion on the plane is $(x_1(i+1) - x_1(i), x_2(i+1) - x_2(i))/\tau \equiv u(i)$. The angle θ of rotation in the two-dimensional plane

for one step in time is (if the mean value is subtracted) $\theta = \arcsin[r(i+1) \times r(i)/|r(i+1)||r(i)|]$ and the corresponding period of revolution can be calculated as $T_\theta = 2\pi\tau/\theta$. Further, using the box method, we construct the probability density for the solution projection onto the chosen plane and the probability density for the points, the velocity of which along the
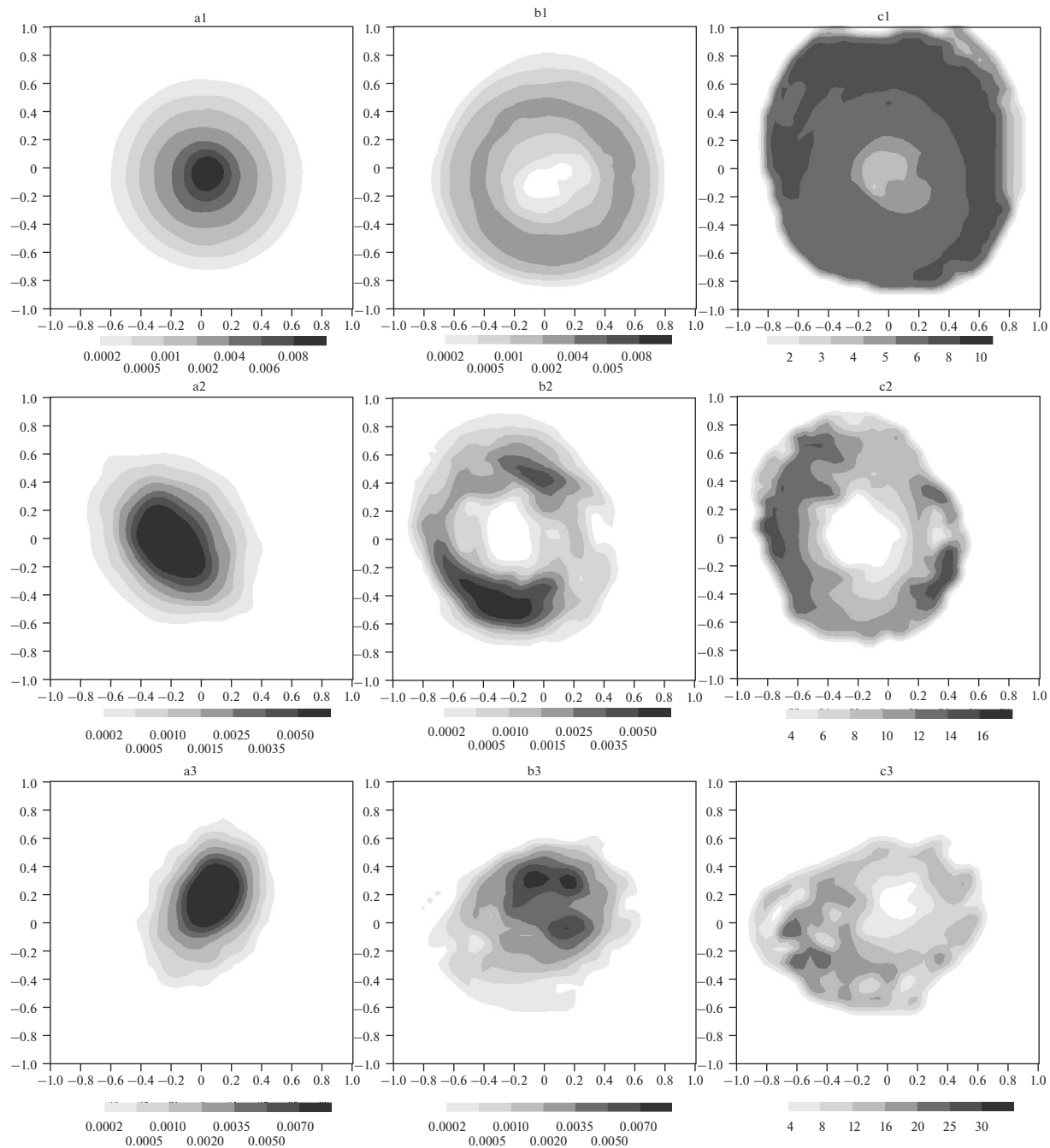
Fig. 5. Same as in Fig. 4, but obtained as a result of approximation by periodic orbits.

given plane $u(i)$ is in the norm, at minimum, 50% of the total velocity of the solution $(\psi(i+1) - \psi(i))/\tau$ (i.e., motion at the given moment of time occurs mainly along this plane).

Figure 4 (panels a1−a3) presents the probability densities of projections of the T21 barotropic model solution onto the plane of the complex-valued NOFs from Fig. 2 (panels a3, b3 and c3); in Fig. 4 (panels b1−b3), the solution projections are on the

same planes, but for the states, the velocity of which along this plane in the norm exceeds 0.5 of the total norm of the velocity. The time of revolution T_θ for the relevant planes is shown in Fig. 4 (panels c1−c3). It can be said that, for all projections, the distributions have a quasi-Gaussian form with the single maximum. In this sense, the circulation regimes are not separated. At the same time, if the states are considered, the evolution of which to a great extent is determined by the process specifying the relevant complex-valued NOF (e.g.,
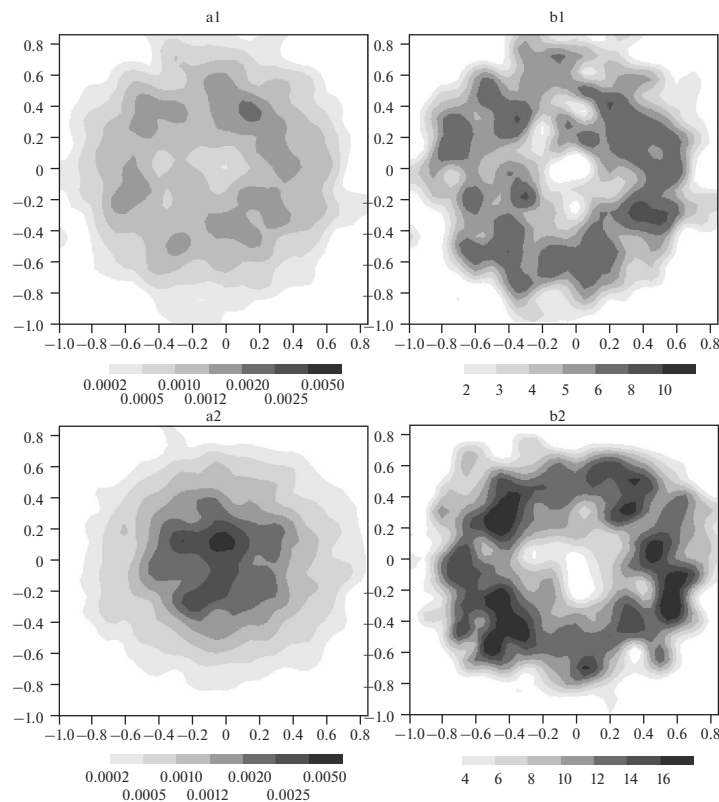
Fig. 6. (a) Projection of probability density onto the CEOF1 plane and (b) characteristic time for which the trajectory revolves around the origin of coordinates along the CEOF1 plane for the model states, the phase velocity of which is projected onto the plane of the corresponding CEOF1 with the correlation coefficient above 0.45. CEOF1 is obtained for the filtered data with filtration windows of 4−6 days (panels a1 and b1) and 10−15 days (panels a2 and b2).

propagation of the Rossby wave with the wavenumber 6, as in Fig. 4, panel a1), then the probability density has a nontrivial nature, acquiring a shape of "bagel." The trajectories of the system describe regular revolutions in the phase space around the origin of coordinates.

The period of revolution is close to the characteristic period of the eigen mode of the linear operator of the barotropic problem (6−8 days (Fig. 4, panel c1) for the plane in Fig. 2, panel a3, and 10−14 days (Fig. 4, panel c2) for the plane in Fig. 2, panel b3). With an increase in the filtration window up to 18−24, i.e., with the consideration of lower-frequency oscillations, the circulation nature becomes more complicated, regular motions around the mean state cease to dominate, and asymmetry appears in the distribution of rotational motion periods.

A similar pattern also takes place for periodic orbits of the system (Fig. 5). For short-period modes of variability, the specific features of the probability-density structure and characteristic times are marked correctly. This means that the model trajectories really execute regular rotational movements around the origin of coordinates along the cluster of the corresponding periodic trajectories. For the low-frequency oscillations, the approximation quality worsens, mainly due to the

fact that the number of orbits with a period of more than 10 days, obtained using the numerical algorithm, is insufficient for correctly describing the dynamics.

For the T12 barotropic model, a similar pattern is observed: regular oscillation movements for 10−15-day modes and the complex low-frequency dynamics take place; oscillations with a characteristic time of 7−9 days are absent in the model. The orbits approximate all details of dynamics with a higher accuracy than for the T21 model. The Marshall−Molteni model demonstrates a similar dynamics; however, its approximation by periodic motions is possible only for higher frequency modes (7- to 9-day modes and shorter).

A similar analysis of the data of the real climatic system shows that the real atmospheric circulation also has a similar structure. Using the entire available NCEP/NCAR archive (1948−2015, November−March), the complex-valued NOFs for filtered data with the filtration windows of 4−6 and 10−15 days were constructed for the stream function of the 250-mb surface. The states were calculated for which the norm of the velocity projection onto the chosen plane exceeds 0.45 of the velocity norm. Then, for these states, the two-dimensional probability densities were constructed (for CEOF1 and the filter of 4−6 days (Fig. 6, panel a1)

and for CEOF1 and the filter of 10−15 days (Fig. 6, panel a2)); the characteristic time of revolution around the mean state is estimated (Fig. 6, panels b1 and b2, respectively).

CONCLUSIONS

In this work, the relation between eigen modes (eigenvectors linearized with respect to a mean state) of models of atmosphere dynamics, the leading modes of variability of model circulation, and the unstable periodic solutions of model equations (solutions returning to the initial state in the specified time) are considered.

In order to search for periodic trajectories, the effective computational algorithm is implemented. It is shown that, in phase spaces of models, there exists a set of periodic solutions sufficient for approximating the base statistical characteristics of models (mean state, variance, and leading variability modes). In this case, for the Marshall−Molteni model, the most complicated model under study, the low-frequency variability modes (with periods of more than two weeks) are ill-approximated, which is probably associated with the insufficient number of found orbits.

Clusters of orbits can be related to different oscillation processes in the phase space of the atmosphere models under study: Rossby waves of different spatial scales and oscillation structures of PNA and NAO types. For high-frequency variability modes, the system dynamics is implemented in the form of a regular periodic motion in planes of complex-valued natural orthogonal functions; in this case, the system trajectory repeats the circular motion of orbits with the same period. The probability density of points executing these rotations has the shape of a torus, which is well-approximated using periodic orbits. As a rule, the plane of regular motion coincides with the plane of the most unstable eigen modes of the barotropic dynamics. Of interest is the fact that the linear nature of dynamics is observed not near the mean state, but at a certain distance from it. This may be interpreted within the linear dynamical-stochastic approach to describing atmospheric circulation, when the dynamical operator is approximated by a stable linear operator and a stochastic nonlinear term, simulating the nonlinear interactions for the system. When the system state is far from the mean state, the linear part of the tendency dominates and, for the states close to the mean state, the opposite occurs.

In the case of the lower frequency processes (from 14 days and more), the dynamics has a more elaborate nature, not described by simple two-dimensional rotations. Along with linear processes (e.g., wave one, propagating from east to west), more complicated modes with nonlinear dynamics also are present (circular mode, NAO). In this case, the possibility of approximating the dynamics by periodic trajectories remains the same. An analysis of observation data

shows that the same conclusions can probably also be made for the real atmospheric circulations. Projections of the solution onto the high-frequency modes of variability assumes the presence of periodic (quasi-periodic) trajectories in the real atmosphere.

REFERENCES

1. C. G. Rossby et al., "Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action," J. Mar. Res. **2,** 38−55 (1939).

2. J. G. Charney, "The dynamics of long waves in a baroclinic westerly current," J. Meteorol. **4,** 135−162 (1947).

3. E. T. Eady, "Long waves and cyclone waves," Tellus **1,** 33−52 (1949).

4. A. J. Simmons and B. J. Hoskins, "Baroclinic instability on the sphere: Normal modes of the primitive and quasigeostrophic equations," J. Atmos. Sci. **33,** 1454−1477 (1976).

5. J. S. Frederiksen, "A unified three-dimensional instability theory of the onset of blocking and cyclogenesis," J. Atmos. Sci. **39,** 969−982 (1982).

6. V. P. Dymnikov, "On the connection between the orthogonal components of meteoelement fields to the eigenfunctions of the dynamic operators," Izv., Acad. Sci., USSR, Atmos. Oceanic Phys. (Engl. Transl.) **24** (7), 675−683 (1988).

7. K. Hasselmann, "PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns," J. Geophys. Res. **93,** 11015−11021 (1988).

8. H. H. von Storch, T. Bruns, I. Fischer-Bruns, and K. Hasselmann, "Principal oscillation pattern analysis of the 30- to 60-day oscillation in general circulation model equatorial troposphere," J. Geophys. Res. **93,** 11022−11036 (1988).

9. H. H. von Storch, U. Weese, and J. S. Xu, "Simultaneous analysis of space-time variability: Principal oscillation patterns and principal interaction patterns with applications to the Southern Oscillation," Z. Meteorol. **40,** 99−103 (1990).

10. J.-S. Xu, "Predicting the state of the Southern Oscillation using principal oscillation pattern analysis," J. Clim. **3,** 1316−1329 (1990).

11. R. Schnur, G. Schmitz, N. Grieger, and H. von Storch, "Normal modes of the atmosphere as estimated by principal oscillation patterns and derived from quasigeostrophic theory," J. Atmos. Sci. **50,** 2386−2400 (1993).

12. J. G. Charney and J. G. DeVore, "Multiple flow equilibria in the atmosphere and blocking," J. Atmos. Sci. **36,** 1205−1216 (1979).

13. B. Legras and M. Ghil, "Persistent anomalies, blocking and variations in atmospheric predictability," J. Atmos. Sci. **42,** 433−471 (1985).

14. K. Mo and M. Ghil, "Cluster analysis of multiple planetary flow regimes," J. Geophys. Res.: Atmos. **93,** 10927−10952 (1988).

15. V. P. Dymnikov and E. V. Kazantsev, "Structure of an attractor generated by the system of equations of a barotropic atmosphere," Izv. Akad. Nauk, Fiz. Atm. Okeana **29,** 581−595 (1993).

16. K. Vautard, K. Mo, and M. Ghil, "Statistical significance test for transition matrices of atmospheric Markov chains," J. Atmos. Sci. **47,** 1926−1931 (1990).

17. M. Kimoto and M. Ghil, "Multiple flow regimes in the Northern Hemisphere winter. Part II: Sectorial regimes and preferred transitions," J. Atmos. Sci. **50,** 2645−2673 (1993).

18. D. Kondrashov, K. Ide, and M. Ghil, "Weather regimes and preferred transition paths in a three-level quasigeostrophic model," J. Atmos. Sci. **61,** 568−587 (2004).

19. A. J. Majda, C. L. Franzke, A. Fischer, and D. T. Crommelin, "Distinct metastable atmospheric regimes despite nearly Gaussian statistics: A paradigm model," Proc. Natl. Acad. Sci. U. S. A. **103,** 8309−8314 (2006).

20. C. Franzke, D. Crommelin, A. Fischer, and A. J. Majda, "A hidden Markov model perspective on regimes and metastability in atmospheric flows," J. Clim. **21,** 1740−1757 (2008).

21. R. Bowen, "Periodic points and measures for axiom A diffeomorphisms," Trans. Am. Math. Soc. **154,** 377−397 (1971).

22. R. Bowen, "Periodic orbits for hyperbolic flows," Am. J. Math. **94,** 1−30 (1972).

23. D. Auerbach, P. Cvitanovic, J. -P. Eckmann, G. Gunaratne, and I. Procaccia, "Exploring chaotic motion through periodic orbits," Phys. Rev. Lett. **58,** 2387−2389 (1987).

24. D. Ruelle, "Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics," J. Stat. Phys. **95,** 393−468 (1999).

25. V. Dymnikov and A. Filatov, *Mathematics of Climate Modelling* (Birkhauser, Boston, 1997).

26. G. Gallavotti, "Chaotic dynamics, fluctuations, nonequilibrium ensembles," Chaos **8** (2), 384−392 (1998).

27. E. Kazantsev, "Unstable periodic orbits and attractor of the barotropic ocean model," Nonlinear Proc. Geophys. **5,** 281−300 (1998).

28. F. M. Selten and G. Branstator, "Preferred regime transition routes and evidence for an unstable periodic orbit in a baroclinic model," J. Atmos. Sci. **61,** 2267−2282 (2004).

29. A. Gritsun, "Unstable periodic trajectories of a barotropic model of the atmosphere," Russ. J. Num. Anal. Math. Modell. **23** (4), 345−367 (2008).

30. A. S. Gritsun, "Statistical characteristics of barotropic atmospheric system and its unstable periodic solutions," Dokl. Earth Sci. **435,** 1688−1691 (2010).

31. A. Gritsun, "Connection of periodic orbits and variability patterns of circulation for the barotropic model of atmospheric dynamics," Dokl. Earth Sci. **438,** 636−640 (2011).

32. A. Gritsun, "Statistical characteristics, circulation regimes and unstable periodic orbits of a barotropic atmospheric model," Philos. Trans. R. Soc., A **371,** 20120336 (2013).

33. J. Marshall and F. Molteni, "Toward a dynamical understanding of planetary scale flow regimes," J. Atmos. Sci. **50,** 1792−1818 (1993).

34. G. P. Kurbatkin, V. N. Sinyaev, and A. G. Yantsen, "Spectral model of a long-term prognosis with average climatic restrictions," Izv., Acad. Sci., USSR, Atmos. Oceanic Phys. **9** (11), 3−8 (1973).

35. C. Franzke, A. Majda, and E. Vanden-Eijnden, "Low-order stochastic mode reduction for a realistic barotropic model climate," J. Atmos. Sci. **62,** 1722−1745 (2005).

36. V. P. Dymnikov and A. S. Gritsun, "Barotropic instability and the structure of low-frequency variability of circulation generated by a two-layer baroclinic atmosphere model," Izv. Akad. Nauk, Fiz. Atmos. Okeana **32** (5), 535−538 (1996).

37. H. von Storch and F. Zwiers, *Statistical Analysis in Climate Research* (Cambridge University Press, Cambridge, 1999).

38. W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, and M. Metcalf, *Numerical Recipes in Fortran 90* (Cambridge University Press, Cambridge, 1986).

Translated by M. Samokhina

Coupled Earth's Thermosohere-Ionosphere Global Dynamics Model

V. P. Dymnikov[a, b], D. V. Kulyamin[a, b], *, and P. A. Ostanin[c]

[a]*Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, 119991 Russia*
[b]*Fedorov Institute of Applied Geophysics, Federal Service for Hydrometeorology and Environmental Monitoring of Russia, Moscow, 129128 Russia*
[c]*Moscow Institute of Physics and Technology, Dolgoprudnyi, 141701 Russia*
*e-mail: kulyamind@mail.ru

Abstract—A coupled Earth's thermosohere-ionosphere global dynamics model (for altitudes of 90−500 km) is presented. This model is based on a three-dimensional thermospheric general circulation model and a dynamical model of the ionospheric F-layer, which takes into account plasma-chemical processes, ambipolar diffusion, and advective ion transport due to neutral wind. General upper atmosphere characteristics have adequately been reproduced and the thermosphere−ionosphere interaction has quantitatively been estimated based on this coupled model. The sensitivity of thermospheric characteristics to ionospheric parameters and the sensitivity of the electron-concentration field distribution in the ionospheric F-layer to thermospheric parameters have been studied within a specified diurnal cycle.

INTRODUCTION

One of the main directions in the development of current climate and atmospheric general circulation models that have been under development at the Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences (INM RAS), in the last four decades is in the transition to complex Earth system models, which, in particular, implies the introduction of descriptions of the atmospheric upper layers, including the Earth's ionosphere. The core of current atmospheric dynamics models—the solution of gas hydrothermodynamic equations in a thin spherical layer approximation—remains applicable up to heights of 500−600 km. However, the development of an Earth system model that includes both the thermosphere and the ionosphere is an independent problem. Within the framework of the solution of this problem, individual general circulation models of the Earth's lower and middle atmospheres, thermosphere, and ionosphere have been developed and some of them are still under development at the INM RAS [1−5].

It is known that the state of the thermosphere−ionosphere system is of interest in solving a number of problems associated with space exploration [6, 7]. The state of this system determines conditions for long-range radiocommunications and spacecraft motion, which should be taken into account in using satellite geolocation and radiocommunication and solving other applied problems. At the same time, both thermospheric and ionospheric characteristics are closely related to one another and the dynamics of the atmo-spheric lower layers; in the last decade, both ionospheric and thermospheric studies have shown that the role of dynamic processes and lower atmosphere effects is significant [6, 8]. At present, the problems associated with a correct description of the global state of the upper atmosphere and the mechanisms of its variability produce a need for the development of coupled thermospheric and ionospheric models and their introduction into Earth system models.

The conventional approach to the description of both ionospheric and thermospheric characteristics is associated with the construction of individual empirical models based on processed observational data [6, 7, 9]. Such upper atmosphere models are mainly the generalizations of local data (satellite measurements along motion paths, ionosondes, and others) and mostly limited to be used in describing variations in the global state of the thermosphere−ionosphere system.

At the same time, at present, the levels of the development and application of global numerical thermosphere−ionosphere models are not as high when compared to empirical, semiempirical and lower atmosphere models. The most advanced current coupled thermosphere−ionosphere models and Earth system models (which includes upper atmosphere) are under development mainly at the world's top-ranked research institutes [9]. Such models include the TIEM-GCM, WACCM, and other model versions developed at the NCAR (United States) [10, 11]; the CTIP-CMAT, WAM, and IDEA models [12, 13]; the domestic GCM-TIP models [14]; and others [15, 16]. The

development of global models of such a level for the upper atmosphere will make it possible to reasonably consider the mechanisms of the formation of variations in observed ionospheric and thermospheric characteristics and use model results for solving applied problems.

The specifics of the description of the global state and dynamics of the Earth's thermosphere (a height range of 90−500 km) involves the key role of shortwave solar radiation absorption determining a rapid temperature increase with height and the dominance of thermal tides in large-scale dynamics and a highly rarefied and inhomogeneous gas composition, which is responsible for significant effects of both molecular-diffusion and thermal-conductivity processes [1, 2]. Let us also note that the dynamical interaction between the neutral thermosphere and the ionosphere is one of the key processes in the formation of the structure of thermospheric circulation, especially in the lower layers [1, 5]. As for the description of the formation of the Earth's ionosphere, in the lower D- and E-layers (60−130 km) the key processes are the ionization and plasma-chemical interactions, and dynamic transport times (when compared to these processes) are small; in the upper F-layers (130−600 km), the dynamics of plasma and its interaction with both magnetic and electric fields become the key processes. At the same time, the processes of ionization in the lower layers of the neutral atmosphere are much more complicated than in its upper layers. The role of the thermosphere in the formation of the ionosphere is directly related to the direct participation of neutrals in ionization and chemical transformations and their dynamic interaction with ions. As a first approximation, the above-formulated problems may be studied based on the method of separating regions that interact through the corresponding boundary conditions and parameterizations [1−5].

The objective of this work is to formulate and verify this new coupled thermosphere−ionosphere global dynamic model (for heights of 90−500 km) based on the earlier created models of thermospheric general circulation [1] and the ionospheric F-layer [3]. In the future, the development of global upper atmospheric models at the INM RAS implies the creation of an earth system model that involves a consistent description of all atmospheric layers from the land surface up to heights of about 500 km.

In Section 1, an improved version of the thermospheric general circulation model is presented, the key parameterizations of basic physical processes that are characteristic of the thermosphere are briefly described, the calculation features of the numerical implementation of the thermospheric model are considered, and the role of the thermosphere−ionosphere interaction is discussed in more detail. In Section 2, the dynamic ionospheric F-layer model is presented, basic equations of the model are formulated, and the method of its numerical implementation is briefly described. Section 3 gives a description of the method of constructing the coupled thermosphere−ionosphere model. The results of control numerical experiments with the described models are given and the role of thermosphere−ionosphere interaction is considered in Section 4. Finally, basic results are formulated and discussed.

1. THERMOSPHERIC GENERAL CIRCULATION MODEL

The version of the global three-dimensional model of the Earth's thermospheric dynamics (for heights of 90 to 500 km) used in this work, generally, corresponds to the model earlier developed by the authors [1, 2]. At the first stage of this work, the main problem was to numerically implement the dynamic core of the model and adequately reproduce the basic features of thermospheric general circulation [1]. At the second stage, the priority problem was to adequately reproduce radiation processes in the thermosphere, which mainly determine the structure of thermospheric global circulation (a temperature increase with height, tides dominating in dynamics, etc.) and the energy of photochemical transformations. This improved thermospheric model version included consistent descriptions of radiation processes on the basis of the created parameterizations [2]. At this stage, some improvements were introduced into the numerical implementation of the model and the reproduction of thermospheric states under different external parameters was studied in detail.

1.1. Model Description

Let us dwell on the description of the thermospheric general circulation model that forms the basis for the coupled model. In this model, a system of nonlinear primitive equations of atmospheric hydrothermodynamics, which is written in a spherical coordinate system with a normalized isobaric vertical coordinate, is solved. The system of basic model equations has the form

$$\frac{du}{dt} - \left(f + \frac{u}{a}\tan\varphi\right)v + \frac{1}{a\cos\varphi}\left(\frac{\partial\Phi}{\partial\lambda}\right)$$
$$= g^2\frac{\partial}{\partial p}\mu\rho\frac{\partial u}{\partial p} - F_{in}^u,$$

$$\frac{dv}{dt} + \left(f + \frac{u}{a}\tan\varphi\right)u + \frac{1}{a}\left(\frac{\partial\Phi}{\partial\varphi}\right)$$
$$= g^2\frac{\partial}{\partial p}\mu\rho\frac{\partial v}{\partial p} - F_{in}^v, \quad \frac{\partial\Phi}{\partial p} = -\frac{RT}{p}, \quad (1)$$

$$\frac{dT}{dt} - \dot{p}\frac{RT}{pc_p} = \frac{g^2}{c_p}\left[\frac{\partial}{\partial p}\chi\rho\frac{\partial T}{\partial p}\right] + \varepsilon,$$

$$\frac{1}{a\cos\varphi}\left(\frac{\partial u}{\partial\lambda} + \frac{\partial v\cos\varphi}{\partial\varphi}\right) + \frac{\partial\dot{p}}{\partial p} = 0,$$

where $\dfrac{d}{dt} = \dfrac{\partial}{\partial t} + \dfrac{u}{a\cos\varphi}\dfrac{\partial}{\partial\lambda} + \dfrac{v}{a}\dfrac{\partial}{\partial\varphi} + \dot{p}\dfrac{\partial}{\partial p}$, $\dot{p} = dp/dt$.

In system (1), λ, φ, t, p are the longitude, latitude, time, and pressure, respectively; u, v are the horizontal

velocities; $\rho = \dfrac{p}{RT}$ is the gas density; $\Phi = gz$ is the geopotential field (z is the height from the land surface); a is the Earth's radius; and f is the Coriolis parameter.

In the right-hand sides of the motion and heat-influx equations, the rates of change of momentum and internal energy are taken into account due to physical subgrid-scale processes that determine the global thermospheric state and are specified by the corresponding parameterizations [1]. Summand ε describes the total heating due to radiation processes, which is calculated using an individually developed radiation block (details of its realization are described in [2]).

The periodicity of the solution (in longitude) and its limitations at the poles are taken as boundary conditions for this system of equations. In this case, the lower boundary is the surface of constant pressure $p_0 = 10^{-3}$ hPa and the upper boundary corresponds to the level $p = 4 \times 10^{-10}$ hPa (about 500 km). For simplicity, the vertical dynamic boundary conditions are chosen in a rigid-lid approximation. The lower boundary condition for the heat conductivity equation (for total temperature) is specified in a first approximation by the constant value of the mean mesospheric temperature $T_0 = 180$ K. The radiation condition is used as an upper boundary condition.

The augends in the right-hand sides of both velocity and temperature equations (1) describe vertical-diffusion and heat-conductivity processes (with the μ and χ coefficients, respectively). Using a simple approximation, the model takes into account molecular viscosity and heat conductivity, which play the key role in the thermospheric upper layers, and the effect of turbulent mixing associated with the collapse of internal gravity waves (IGWs) propagating from the mesosphere [2]. The radiation block takes into account air heating and cooling during the transport of radiation: the key processes are heating due to the absorption of solar ultraviolet radiation and cooling due to the infrared radiation of carbon dioxide in the thermospheric lower layers [2]. Note that, in this case, solar radiation fluxes are parameterized and the effect of solar activity is conventionally determined from the parameter of solar ultraviolet radiation $F_{10.7}$ [2].

Let us dwell upon the consideration for the effect of the ionosphere in simulating the thermosphere (terms F_{in} in the right-hand sides of system (1)). The thermosphere–ionosphere interaction occurs directly due to the momentum and energy exchange between charged and neutral particles during their collisions. In the general case, this force is the collision integral in deriving motion equations from the Boltzmann equation [6, 7]. Since the atmosphere in its upper layers at heights of 100–500 km is a weakly ionized plasma, for neutral particles, a significant contribution is made only by collisions with ions (ion-neutral resis-

tance force). For the ionosphere, the local quasineutrality condition relates electron concentration n_e with total ion concentration $n_i = n_e$.

It is well known from preliminary estimates that, for the general neutral-gas dynamics in the upper atmosphere, one significant contribution of collisions with ions is the interaction in the dynamo region (90–150 km, mainly the ionospheric E-layer). In this case, in a first approximation for ions, the ion-neutral collisions are locally balanced by the Lorentz force with good accuracy [1, 6, 7].

$$-en_i\left(\overline{E_0} + \left[\overline{u_i} \times \overline{B}\right]\right) = \overline{F_{in}} = n_i m_i \nu_{in}(\overline{u} - \overline{u_i}), \qquad (2)$$

where $\overline{E_0}$ is the external electric field intensity, $\overline{B}$ is the Earth's magnetic field, $\overline{u}$ is the three-dimensional vector of thermospheric velocity (neutral gas), $\overline{u_i}$ is the drift velocity of ions in the Earth's reference system, ν_{in} is the frequency of collisions of ions and neutrals, e is the average ion charge equal to electron charge, and m_i is the average atomic mass of ionic gas. Then, the ion–neutral interaction force is $\overline{F_{ni}} = \overline{F_{in}}$ for the thermosphere (according to Newton's third law). Note that this approximation is applicable mainly to the thermospheric lower layers and does not reflect the dynamics of ionospheric plasma in the F-layer.

Since, the thermosphere (together with the ionosphere) is, in fact, the medium of conductivity of charged particles in the Earth's magnetic field, for a small volume of conductive gas, the Lorentz force in these approximations may be expressed through conductivity, so that

$$\overline{F_{in}} = \left[\overline{J} \times \overline{B}\right], \qquad (3)$$

where $\overline{J}$ is the current density. In these approximations, let us use the Ohm law in a generalized form, and the current density in the medium may be determined as [6, 7]:

$$\overline{J} = \overline{\overline{\sigma}}\left(\overline{E_0} + \left[\overline{u} \times \overline{B}\right]\right). \qquad (4)$$

Term $\left[\overline{u} \times \overline{B}\right]$ describes the dynamo–effect field (induced by atmospheric motions in the Earth's magnetic field), and $\overline{\overline{\sigma}}$ is the conductivity tensor determined by instantaneous medium characteristics.

The Earth's magnetic field $\overline{B}$ is estimated with an adequate accuracy as a dipole field (in the general case) with the axis that does not coincide with the Earth's rotation axis. The vector of geomagnetic field $\overline{B}$ in a spherical coordinate system has the components

$$\vec{B} = \begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} = \begin{bmatrix} B\cos I \sin D \\ B\cos I \cos D \\ -B\sin I \end{bmatrix}, \qquad (5)$$
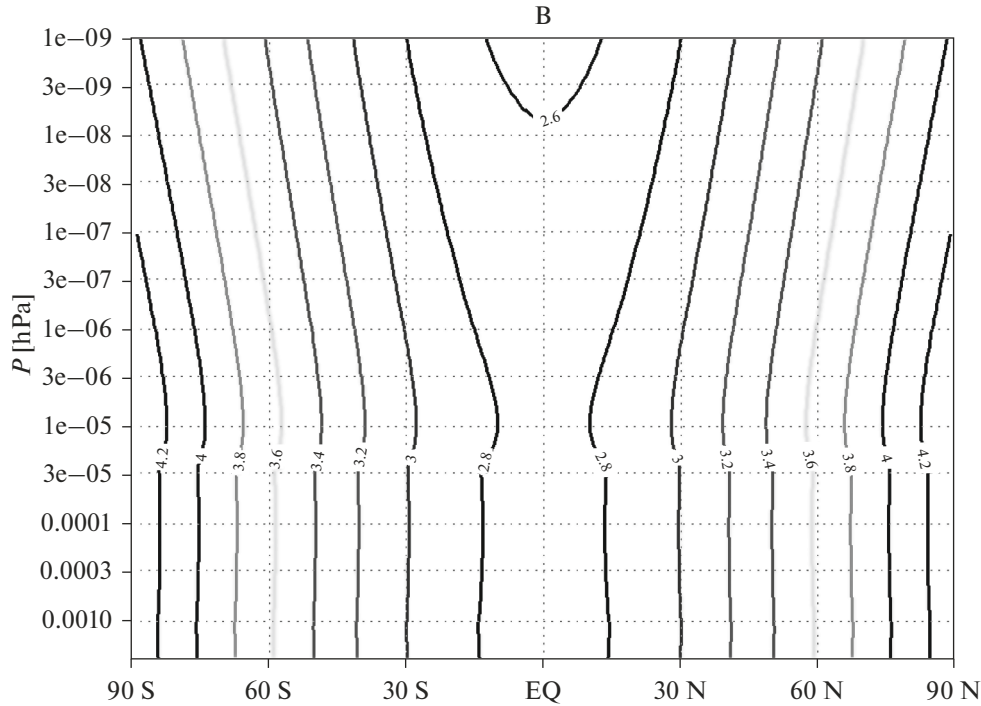
Fig. 1. Vertical magnetic-field profile $B(p, \varphi)$ $[10^{-5}$ T] used in the numerical experiments with the given thermospheric model version (the ordinate is the height in the p-coordinate system [hPa] on a logarithmic scale and the abscissa is latitude).

where I is the inclination angle (between a horizontal plane and the magnetic-field vector) and D is the declination angle (between the meridian and the projection of the magnetic-field vector onto the horizontal plane). The value of the magnetic field at each point may be calculated as $B(r, \theta) = \frac{P_m}{r^3} \sqrt{(1 + 3 \sin^2 \theta)}$; $(90 - \theta)$ is the angle between the magnetic axis and observation point (magnetic latitude) and $P_m = 7.8 \times 10^{15}$ T m^3 is the Earth's magnetic dipole moment.

In the numerical experiments given below, it is assumed for the thermospheric model that $E_0 = 0$, and the Earth's rotation axis and the geomagnetic axis coincide (thus, $\theta \approx \varphi$, declination D is zero, and the inclination angle may be expressed as $I = \arctan(2 \tan \varphi)$). Then, the approximate expression for the magnetic field is written as follows:

$$\bar{B} = (0, B \cos I, -B \sin I),$$
$$B = \frac{P_m}{(a + z)^3} \sqrt{(1 + 3 \sin^2 \varphi)}, \tag{6}$$

where height z above Earth's surface is calculated from pressure in accordance with the hydrostatic condition (1). The vertical magnetic-field profile $B(\varphi, p)$, which is used in the numerical experiments with the thermospheric model version under consideration, is shown in Fig. 1.

With consideration for a strong vertical diffusion and the hydrostatic approximation, the contribution of the Lorentz force to the equations of motion of the neutral component is considered only for the horizontal wind-velocity components. After some transformations, according to (3) and (4), the momentum influx for the horizontal wind-velocity components, which is caused by collisions with charges, may be expressed as

$$\overline{F_{in}^{uv}} = \frac{1}{\rho} \overline{F_{ni}} = -\bar{\bar{D}} \bar{u}, \tag{7}$$

where $\bar{\bar{D}}$ is the two-dimensional ion-resistance tensor written only for the horizontal components: $\bar{\bar{D}} = \begin{pmatrix} D_{xx} & D_{xy} \\ -D_{xy} & D_{yy} \end{pmatrix}$. Its coefficients may be calculated as

$$D_{xx} = \sigma_P B^2, \quad D_{xy} = \sigma_H B^2 \sin I,$$
$$D_{yy} = \sigma_P B^2 \sin^2 I, \tag{8}$$

where σ_P is the Pedersen longitudinal electroconductivity and σ_H is the Hall transverse electroconductivity. According to (4), in these approximations, the conductivities may be expressed in the following form [6, 7]:

$$\sigma_P = \frac{n_i e}{B} \left[\frac{\nu_{in}/\omega_i}{1 + (\nu_{in}/\omega_i)^2} \right], \quad \sigma_H = \frac{n_i e}{B} \left[\frac{(\nu_{in}/\omega_i)^2}{1 + (\nu_{in}/\omega_i)^2} \right]. \tag{9}$$
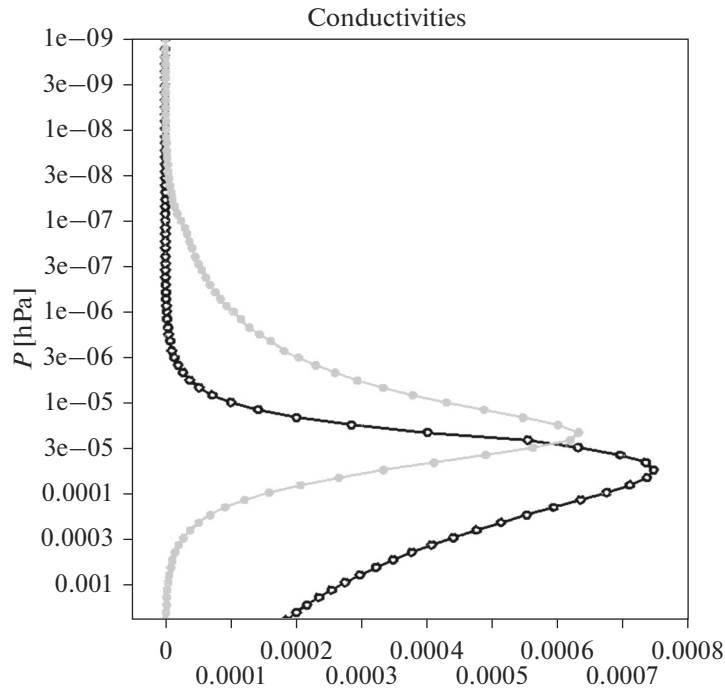
Fig. 2. Vertical profiles of both longitudinal σ_P (gray line) and transverse σ_H (black line) electroconductivities in the p-coordinate system [hPa] on a logarithmic scale; the values are given in S/m.

In these expressions, $\omega_i = \dfrac{eB}{m_i}$ is the cyclotron ion frequency. In a first approximation, one can estimate $m_i \approx 2 \times 10^{-26}$ kg and $B \approx 3 \times 10^{-5}$ T; thus, $\omega_i \approx 2 \times 10^2$ s^{-1} is the average estimate for the ionosphere.

In the general case, frequency ν_{in} is determined individually for the interaction of different gas components with charged particles. In a first approximation, for the thermosphere, this parameter may be generalized by the following average estimate [1]:

$$\nu_{ni} = 5 \times 10^{-16}(n_i + n_n) \text{ s}^{-1}, \qquad (10)$$

where $n_n = p/kT$ is the total concentration of neutral particles, which is calculated for the thermosphere according to the state equation, and k is the Boltzmann constant.

The Pedersen and Hall conductivities calculated in this thermospheric model version using the presented estimates are given in Fig. 2. These profiles show that the dynamic ionosphere–thermosphere interaction is most important for the lower-thermosphere heights, which was noted above.

The resistance force for the horizontal wind-velocity components in the right-hand sides of the motion equations is written in the form

$$F_u'' = -\frac{RT}{p}\left(D_{xx}u + D_{xy}v\right),$$
$$F_v'' = -\frac{RT}{p}\left(D_{yy}v - D_{xy}u\right). \qquad (11)$$

Thus, the components of the ion–neutral interaction, which are determined by the summands with D_{xx} and D_{yy}, act as a force of aerodynamic resistance, and the summands with D_{xy} act as a torque force (similarly to the Coriolis force).

For the ion–neutral interaction, one can take into account electrodynamic power transmitted as the Joule heat. In the above-described approximations, it may be estimated according to the generalized form of the Joule–Lenz law; the heat-influx expression will have the form

$$F_{in}^T = \frac{1}{\rho c_p}\left(\overline{JE}\right) = \frac{RT}{c_p p}\left(D_{xx}u^2 + D_{yy}v^2\right). \qquad (12)$$

The preliminary estimates and numerical experiments with the thermospheric model showed that, with the parameters used at this stage, this process makes no significant contribution to the formation of general circulation and, therefore, is not taken into account in this model version.

1.2. Numerical Implementation

A regular (in longitude and latitude) grid with steps $\Delta\varphi = 2°$ (in latitude) and $\Delta\lambda = 2.5°$ (in longitude) is used in the thermospheric general circulation model. The model has 80 vertical levels in the p-coordinate system. The step between the vertical levels is specified approximately uniform on the logarithmic scale, so that the distance between the levels is, on average, $\Delta z = \frac{RT}{g}\ln\left(\frac{p + \Delta p}{p}\right) \approx 4–8$ km. The details of the numerical implementation of the dynamic core are given in [1]. The spatial finite-difference approximation is performed on shifted grid C, for which the scalars (Φ, T) are determined in the cell centers and velocity values are determined in the centers of the corresponding side faces. The spatial derivatives in linear terms are approximated by central differences. The spatial approximation of the dynamic block of the model is performed using the finite-difference scheme of second-order accuracy and is based on the advective notation of summands. This scheme corresponds to the schemes of transport implementation in the INM RAS models of atmospheric general circulation and climate [4]. The grid is constructed so that only variable v is determined at the poles. The spatial Fourier filtration of high-frequency harmonics along a circle of latitude for all fields of prognostic variables is used in the vicinity of the poles (beginning from 69° N latitude).

In the finite-difference scheme, high-frequency spatial modes are suppressed using the eighth-order horizontal diffusion in order to eliminate a possible false energy cascade in the shortwave spectrum region. The diffusion coefficient is chosen so that the amplitude of the highest frequency spatial harmonic is four times smaller than that before filtration.

The time integration method of the model and the details of the implemented parameterizations of physical processes are described in [1, 3].

2. IONOSPHERIC F-LAYER MODEL

One of the central tasks of this work is to combine the global dynamic model of the ionospheric F-layer (developed by the authors [3]) and the thermospheric general circulation model presented in the previous section into a coupled model.

2.1. Problem Statement

The first version of the global dynamic ionospheric F-layer model is based on the solution of the continuity equation in order to determine ion concentrations n_i and on the following approximations [3]:

(1) the single-ion formulation of the model (assumed photochemical dominance of the ionization of atomic oxygen O and recombination of its ion O+ with major thermospheric constituents at these heights),

(2) the above-noted plasma quasi-neutrality ($n_i = n_e$),

(3) the joint motion of electrons and ions during ambipolar diffusion along the magnetic force lines as a key dynamic process in ionospheric plasma,

(4) the dominance of electromagnetic transverse drift,

(5) an approximated shape of the Earth's magnetic field—the above-described dipole field with coincident magnetic and geographic poles.

In these assumptions, the total balance of forces affecting the ionospheric plasma is determined by the ratio between the pressure gradients of electron and ion gases, the Lorentz force, gravitation, and ion–neutral collisions [3, 6, 7]. Using the projections of the balance of forces in plasma onto the direction coinciding with the direction of geomagnetic field $\overline{B}$ and onto the plane perpendicular to this direction, one can obtain an expression for both parallel $u_{i\parallel}$ and perpendicular $u_{i\perp}$ ion-velocity components. In the direction parallel to the Earth's magnetic field, the expression for the velocity component corresponds to the equation of ambipolar diffusion along magnetic lines [3]:

$$u_{i\parallel} = u_{\parallel} - D\left(\frac{1}{n_i}\nabla_{\parallel}(n_i) + \frac{1}{T_p}\nabla_{\parallel}T_p - \frac{1}{H}\right), \quad (13)$$

where $D = \frac{k(T_i + T_e)}{m_i\nu_{in}}$ is the coefficient of ambipolar diffusion; $H = \frac{k(T_i + T_e)}{m_i g}$ is the height scale for ionospheric plasma; and $T_p = \frac{1}{2}(T_i + T_e)$ is the mean plasma temperature, where T_e, T_i are the temperatures corresponding to each component.

In the transverse direction, one can use the balance of the Lorentz force and collision term (similar to that used in the thermospheric model to describe the ion–neutral interaction (2)). However, estimates for the heights of the ionospheric F-layer show that the dominating component in the transverse direction is an electromagnetic drift determined by external fields so that $u_{i\perp} = \frac{1}{B^2}\left[\overline{E_0} \times \overline{B}\right]$. In this version of the F-layer model block, only the electromagnetic drift is taken into account.

Thus, the continuity equation for the electron concentration in the F-layer, which is solved in the ionospheric model, has the form

$$\frac{\partial n_i}{\partial t} + \text{div}\left(n_i\frac{1}{B^2}\left[\overline{E_0} \times \overline{B}\right]_{\perp}\right) + \text{div} \times \left(n_i\overline{u_{\parallel}}\right.$$
$$\left. - D\left(\nabla_{\parallel}(n_i) + n_i\frac{1}{T_p}\nabla_{\parallel}T_p - \frac{n_i m_i}{2k T_p}\overline{g_{\parallel}}\right)\right) \quad (14)$$
$$= P - k_i n_i.$$

In the right-hand side of this equation, P is the rate of solar-radiation ionization (in this version, ionization of atomic oxygen O to ion O+), k_i is the total rate of recombination (in our case, the recombination only with molecular oxygen and nitrogen is taken into account).

It is conventional for solving systems of type (14) to use the curvilinear coordinate system related with directions of the Earth's magnetic field, which is implemented in a number of current models in the world [12–15]. Such an approach makes it possible to solve quasi-one-dimensional equations in a parallel direction; however, in this case, a number of difficulties arise [3] which impose significant limitations on the development of coupled thermosphere–ionosphere models. Therefore, in this ionospheric model version, the system of spherical coordinates (equations are formulated in a thin spherical-layer approximation), which is in agreement with the formulation of the above-described thermospheric model [1], is taken as a basic coordinate system. With the above-introduced approximations taken into account, the continuity equation for ions in the coordinate form will be written as [3]

$$\frac{\partial n_i}{\partial t} = EYZ(n_i) + DTr(n_i) + Tr(n_i) + [P - k_i n_i], \quad (15)$$

where

$$EYZ(n_i) = \frac{1}{a^2 \cos\varphi} \frac{\partial}{\partial\varphi} \left[D\cos^2 I \frac{\partial n_i}{\partial\varphi} \cos\varphi \right]$$
$$+ \frac{\partial}{\partial z} \left[D\sin^2 I \frac{\partial n_i}{\partial z} \right]$$
$$- \frac{1}{a\cos\varphi} \frac{\partial}{\partial\varphi} \left[D\sin I \cos I \frac{\partial n_i}{\partial z} \cos\varphi \right]$$
$$- \frac{1}{a} \frac{\partial}{\partial z} \left[D\sin I \cos I \frac{\partial n_i}{\partial\varphi} \right],$$

$$DTr(n_i) = \frac{1}{a\cos\varphi} \frac{\partial}{\partial\varphi} \left[\left(\frac{1}{a} D\cos^2 I \frac{1}{T_p} \frac{\partial T_p}{\partial\varphi} \right. \right.$$
$$\left. - D\sin I \cos I \left(\frac{1}{T_p} \frac{\partial T_p}{\partial z} + \frac{1}{H} \right) n_i \cos\varphi \right) \right]$$
$$+ \frac{\partial}{\partial z} \left[\left(-\frac{1}{a} D\sin I \cos I \frac{1}{T_p} \frac{\partial T_p}{\partial\varphi} + D\sin^2 I \right. \right.$$
$$\left. \left. \times \left(\frac{1}{T_p} \frac{\partial T_p}{\partial z} + \frac{1}{H} \right) \right) n_i \right],$$

$$Tr(n_i) = \frac{1}{a\cos\varphi} \frac{\partial}{\partial\lambda} \left[\frac{1}{B} (E_y \sin I + E_z \cos I) n_i \right]$$
$$- \frac{1}{a\cos\varphi} \frac{\partial}{\partial\varphi} \left[\left(\frac{1}{B} E_x \sin I + u_y \cos^2 I \right. \right.$$
$$\left. - u_z \sin I \cos I \right) n_i \cos\varphi \right]$$
$$- \frac{\partial}{\partial z} \left[\left(\frac{1}{B} E_x \cos I - u_y \sin I \cos I + u_z \sin^2 I \right) n_i \right].$$

The main difficulty in solving these equations is the presence of mixed derivatives (due to the lack of coincidence of magnetic lines with coordinate axes). Note that, as far as the authors know, a similar approach is used in [10, 16].

2.2. Solution Method

In solving system (15) of the ionospheric model, the following requirements for numerical schemes should be met [3]:

(1) the equation of the F-layer model is the mass-balance expression for ions, so that the corresponding integral balance relations should be satisfied in the numerical implementation;

(2) the ion-dynamics equations contain a description of the processes of ambipolar diffusion and transport along the force lines of the Earth's magnetic field in the projection onto coordinate directions; therefore, the problem has some geometric features associated with a preferred direction of motion (along a line), which should be taken into account in discretizing the equations on the difference grid;

(3) for the heights under study, the characteristic values of the diffusion coefficient and other parameters associated with air density in the atmosphere exponentially vary with height by 6–7 orders of magnitude, which should be taken into account in constructing solution methods in the region of interest;

(4) the characteristic times of plasma-chemical processes are very short (on the order of seconds), which, together with the previous feature, implies that the system is stiff and it is necessary to use special schemes;

(5) in the system of equations under consideration, the ion concentration should be positive.

The developed numerical scheme is based on the method of splitting into the physical processes: ambipolar diffusion and advective plasma transport. The difference scheme constructed with consideration for the above requirements is described in detail and its accuracy is verified in [3].

For time approximation in the ionospheric F-layer model, according to the above requirement for efficient solution methods, implicit difference schemes are used; in this case, the target values of time steps in constructing the model are the values corresponding to those used in the thermospheric model block (1–5 min). The modified CABARET scheme is used to take into account the processes of the three-dimensional advective transport of ions [17, 18].

3. COMBINING THE THERMOSPHERIC GENERAL CIRCULATION MODEL AND THE IONOSPHERIC F-LAYER MODEL INTO A COUPLED MODEL

Consistent calculations of the model units described above at each time-integration step with the interactive

exchange of the key parameters form the basis for the coupled thermosphere–ionosphere model is used (for a height range of 90–500 km). For the numerical implementation of the coupled model, the difference grid of the thermospheric model (grid steps 2° in latitude, 2.5° in longitude, and 80 vertical levels in the p-coordinate system) is used [1]; for the ionospheric model [3], the horizontal grid coincided and, vertically, the uniform grid was used in the z coordinate (values of the parameters necessary for the exchange were interpolated from constant-pressure levels and backward).

Since the thermospheric state is determining for the formation of the ionosphere, the thermospheric parameters were transmitted to the block of the ionospheric F-layer model at each time step: first and foremost, the neutral temperature and pressure were transmitted, and the solar-motion parameters were agreed. To calculate the advective ionospheric-plasma transport caused by neutral wind, an individual block is implemented to calculate the projections of neutral velocities onto the direction along magnetic force lines and the plane perpendicular to them.

Let us dwell on the consideration for the role of the ionosphere in simulating the thermosphere. The thermosphere–ionosphere interaction occurs directly due to the momentum–energy exchange between charged and neutral particles during their collisions. As was described in Section 1, in the thermospheric general circulation model under consideration, this exchange is taken into account using an individual parameterization of ion–neutral resistance (11), for which the external parameter is the ion concentration necessary for calculations of conductivities in the corresponding coefficients (9). Thus, in the first version of the coupled model, during consistent simulations of the dynamics of the ionospheric F-layer and thermospheric general circulation, electron-concentration values were transmitted into the thermospheric block at each time step. In this case, since the processes of the E-layer formation are not taken into account in the F-layer model and, therefore, the ionospheric lower layers are calculated with a low accuracy, electron-concentration values were corrected in the parameterization of ion–neutral friction for heights below 150 km (the nudging procedure was implemented, in which a specified ion-concentration profile was used as a basic one [1]).

4. RESULTS OF NUMERICAL EXPERIMENTS

Since the accuracy of implementing and identifying individual blocks of the thermospheric and ionospheric F-layer models on the basis of observational data was studied in [1, 3, 5], this work presents the results of reproducing some key characteristics of the global thermospheric and ionospheric dynamics on the basis of coupled-model data with emphasis given to the thermosphere–ionosphere interaction. Thus, in fact, the sensitivity of the global state of each of these systems to important characteristics of the other one is considered.

In this work, the system behavior was considered within a specified diurnal cycle under the equinox conditions. At the first stage, the sensitivity of the thermosphere to the key ionospheric parameters was considered based on the individual thermospheric circulation model [2], the sensitivity of the F-layer characteristics to thermospheric parameters was considered based on the model of ionospheric dynamics [5], and the system sensitivity to solar activity variations was also considered.

Note that both models have demonstrated a significant sensitivity of the characteristics of thermospheric circulation and electron-concentration distribution to solar activity, which implies the importance of consistent calculations of the solar cycle and its variations on the basis of the coupled model. The upper layers (200–500 km) are most sensitive to solar activity. In the thermosphere, the temperature of neutrals significantly increases with increasing solar activity (by 20–30% at a twofold increase of F10.7) and, in the F-layer, the electron concentration increases by 40–70% at a twofold increase of F10.7.

Since the electron concentration is the key parameter for the ion–neutral interaction, which plays a dominant role in the lower thermosphere, the latter proved to be most sensitive to concentration field variations. The results of numerical experiments with the thermospheric general-circulation model show that global electron-concentration variations do not change the thermospheric general-circulation structure; however, they significantly affect both neutral-wind velocity and temperature mainly within a height range of 0–200 km. A decrease in electron concentration causes an additional cooling in the polar region and, accordingly, an overestimation of zonal jet velocities in the middle latitudes; both meridional and vertical transports vary slightly (a global concentration decrease of 10% causes an additional cooling of the lower thermosphere in the subpolar latitudes and a warming of, on average, 10–15% at heights of above 200 km; see Figs. 1–3 in [1]). The amplitudes of reproduced diurnal tides in the lower thermosphere also correspondingly vary.

The influence of thermospheric parameters on the distribution of daytime electron concentration has been studied based on the ionospheric F-layer model. Since variations in the density and concentration of neutrals are determined, to a great extent, by the neutral-temperature variability, the thermospheric effect on the ionosphere is determined by this parameter within the framework of the models under consideration. The numerical experiments showed a noticeable nonlinear dependence of electron-concentration profiles under temperature variations (a temperature rise of 10% causes a concentration increase of approximately 5–15% in the upper ionosphere and a rise in
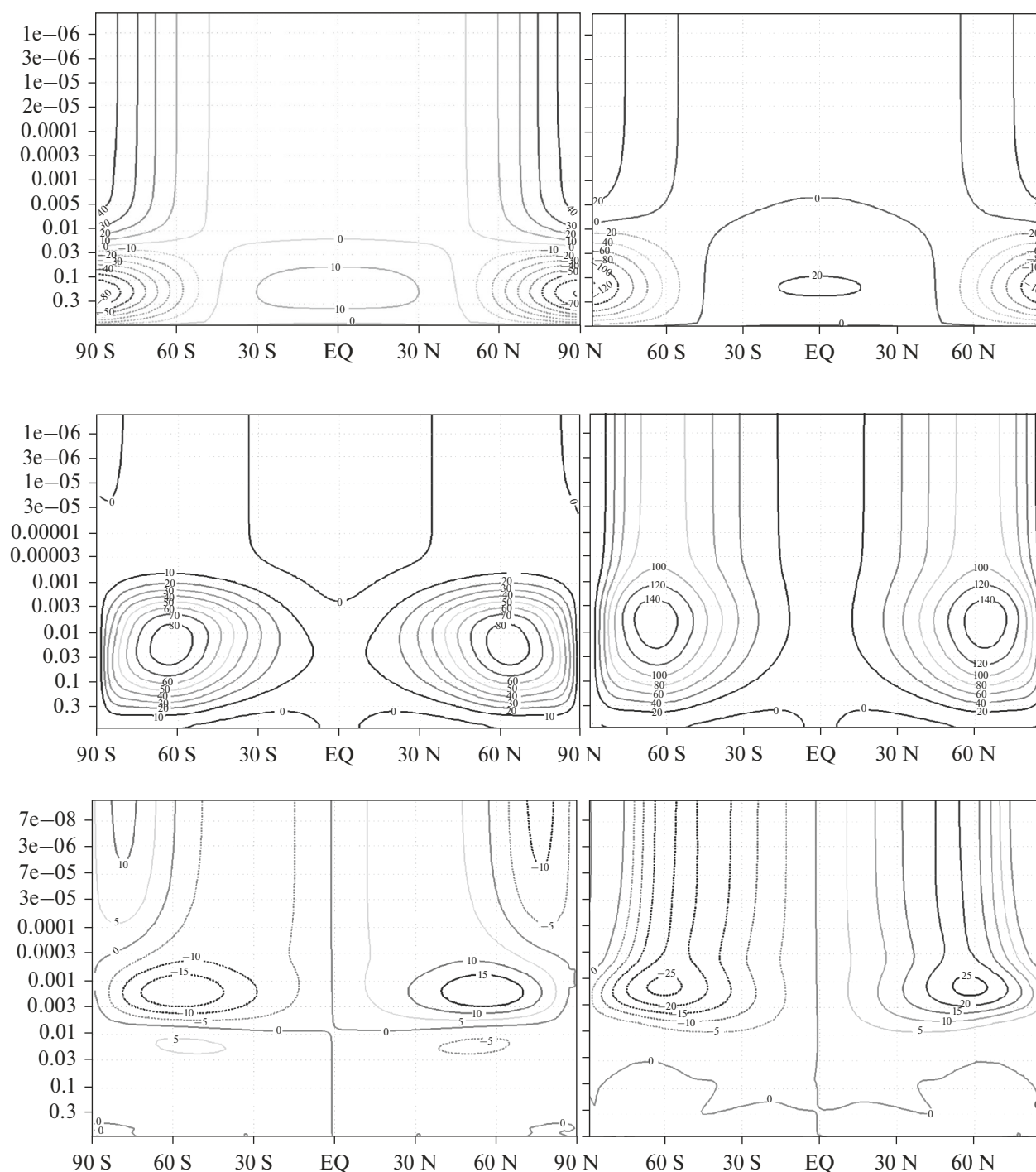
Fig. 3. Latitude−altitude distributions of the zonally averaged temperature deviations from the global profile (K) (top), zonal velocity (m/s) (middle), and meridional velocity (m/s) (bottom) according to data obtained in the test numerical experiments with the MINM RAS thermospheric general circulation model (on the left) and the coupled thermosphere−ionosphere circulation model (on the right) for equinox conditions.

the F-layer maximum of 50−80 km; the high-latitude region is most sensitive).

The first experiments with consideration for the advective ion transport determined by neutral wind showed a noticeable effect of the velocity component parallel to the magnetic field (mainly in both latitudi-

nal and vertical transports) and a slight effect of the perpendicular component on redistributions of the electron-concentration field at neutral-velocity values reproduced in the thermosphere. However, the effect of thermospheric dynamics on the ionosphere should be considered separately in more detail.
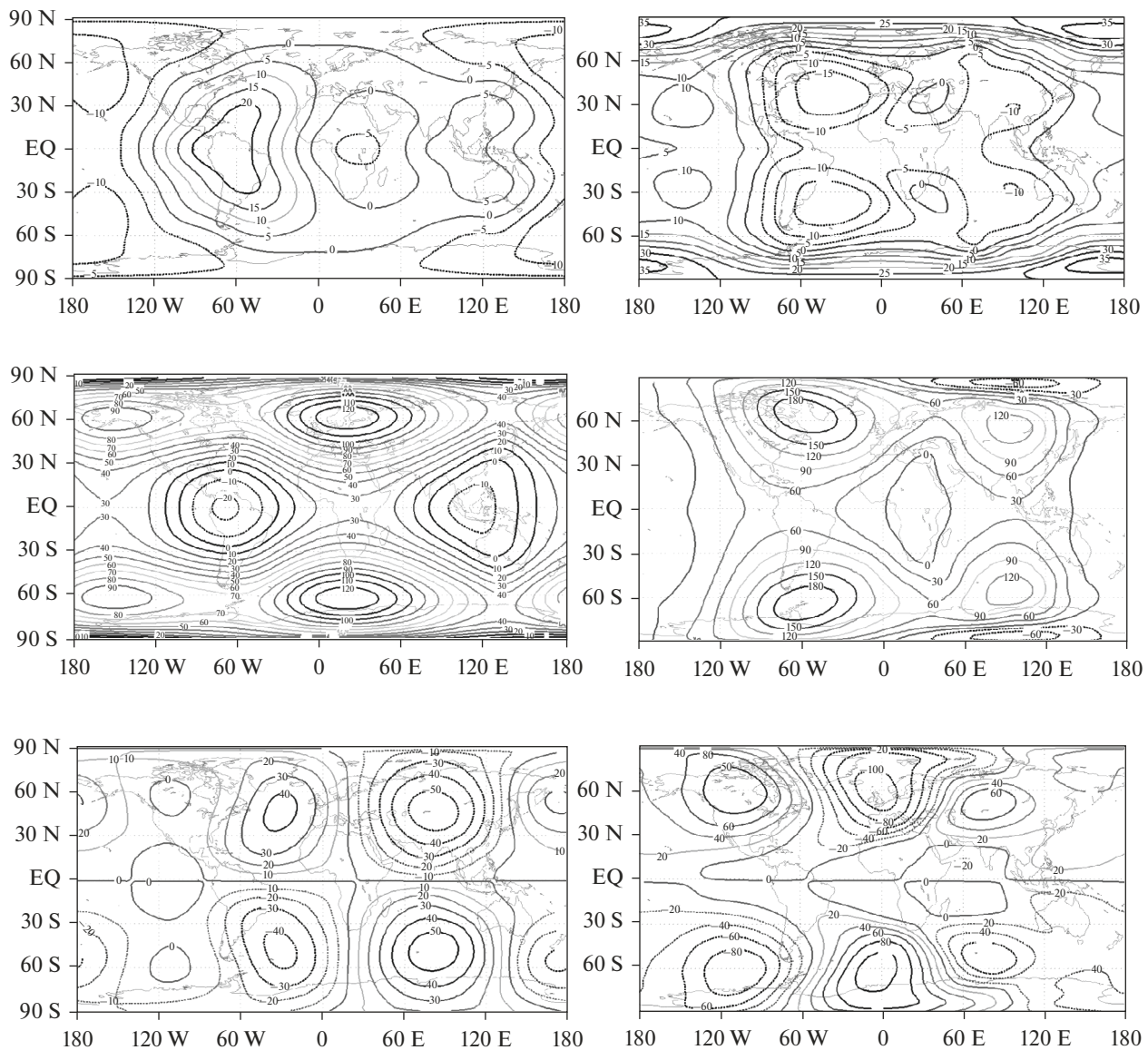
Fig. 4. Latitude−longitude distributions of temperature deviations (K) (top), zonal velocity (m/s) (middle), and meridional velocity (m/s) (bottom) at a height level of 4.e-6 Pa (~120 km) for 12:00 (Moscow time) according to data obtained in the test numerical experiments with the thermospheric general circulation model (on the left) at a specified electron-concentration distribution and with the coupled thermosphere−ionosphere circulation model (on the right) for equinox conditions.

Using the coupled thermosphere−ionosphere model, a number of test experiments have been carried out to verify and debug this model and to preliminarily study and compare interactions between thermospheric and ionospheric characteristics (for a period of 10−30 days) under the conditions of spring equinox. It is shown that the algorithm of numerical implementation of the coupled thermosphere−ionosphere model is operative; under specified external forcings, the diurnal cycle is established after a few days of calculations.

The preliminary results of simulating average characteristics of the thermosphere and the ionospheric F-layer within the framework of the problem of reproducing a diurnal cycle, on the whole, correspond to

generally accepted estimates and results obtained earlier using these model blocks individually. As an example, Figs. 3 and 4 give the results of reproducing some key characteristics of thermospheric circulation on the basis of data obtained with the coupled thermosphere−ionosphere model and the thermospheric model [1] in comparison: the zonally averaged profiles of temperature deviations, both mean-zonal and meridional velocities, and the latitude−longitude variations in these parameters for heights of the order of 250 km. It is seen that the widest circulation variations from the inclusion of F region calculation occur in the lower layers. A significant intensification of jet flows is associated with a considerable decrease (in value) in
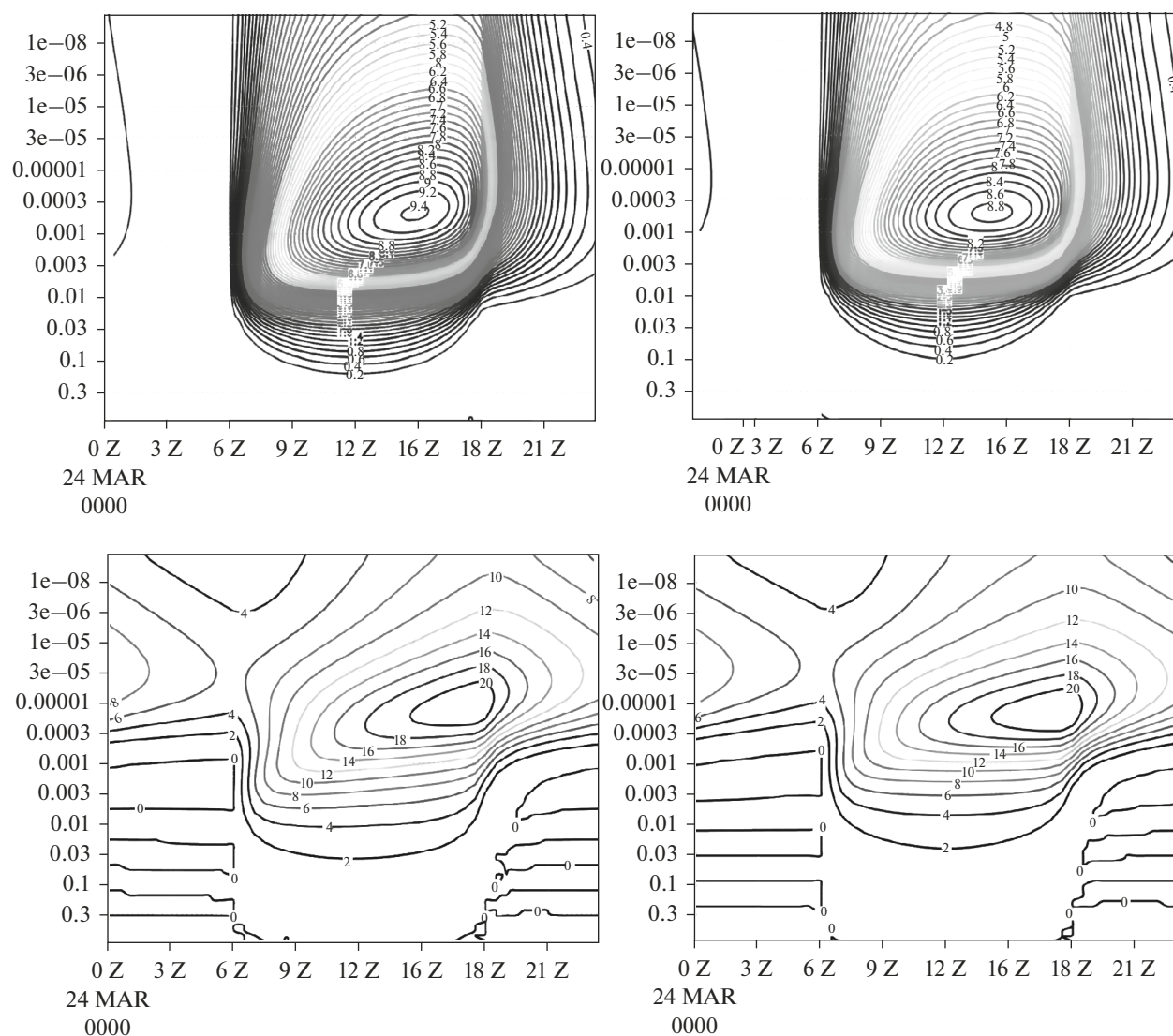
Fig. 5. Time variations in the profiles of electron concentration 10^5 (cm^{-3}) of the zonally averaged deviations at 60° N (top) and the equator (bottom) for 24 h according to data obtained in the test numerical experiments with the ionospheric F-layer dynamics model at a specified temperature (on the left) and the coupled thermosphere–ionosphere circulation model (right) for equinox conditions.

the electron-concentration distribution field established within the diurnal cycle (according to data obtained with the F-layer model), when compared to that specified earlier (mainly justified by the fact that, in this ionospheric model version, the night concentration is zero, which results in decreased ion–neutral friction and increased night dynamics). These variations are also seen in latitude–longitude distributions. Due to a more correct consideration for diurnal variations in the electron-concentration distribution, the diurnal zonal-velocity and temperature variations, on the whole, have become closer to experimental data obtained with other models and observational data, especially in analyzing the reproduction of tide characteristics (amplitudes of basic large-scale disturbances have significantly increased, and the semidiurnal mode has become pronounced).

Figure 5 gives the results of reproducing electron concentrations according to data obtained with the coupled thermosphere–ionosphere model and the individual ionospheric F-layer model using specified thermospheric parameters [1]. An analysis of the results of reproducing the spatial electron-concentration distribution showed that the coupled model adequately reproduces the characteristic behavior of the F-layer within the diurnal cycle; in this case, the consistent consideration for variations in thermospheric parameters (primarily, temperature) significantly contributes to the formation of the daytime-ionosphere structure (the pole-to-equator gradient noticeably increases and the profile structures vary in both subpolar and equatorial regions). Note that the results obtained with the coupled model correspond to data

on the sensitivity of electron-concentration profiles to temperature [3].

A detailed analysis of characteristic structures of the coupled thermosphere–ionosphere dynamics will be given in a separate paper.

CONCLUSIONS

Let us briefly formulate the basic results:

(1) A coupled Earth's thermosohere-ionosphere global dynamics model (for heights of 90–500 km) has been developed based on the thermospheric general circulation model integrated with the ionospheric F-layer model.

(2) The implementation of each model unit and the methods of coupling the models have been described.

(3) The sensitivity of the thermospheric key characteristics to ionospheric parameters and the sensitivity of electron concentration to thermospheric parameters have been studied based on the individual models.

(4) Control numerical experiments have been carried out to reproduce the thermospheric and ionospheric general circulation within a specified diurnal cycle using the coupled model. It is shown that the basic characteristics of the upper atmosphere and the thermosphere–ionosphere interaction have adequately been reproduced using the coupled model.

FUNDING

REFERENCES

1. D. V. Kulyamin and V. P. Dymnikov, "A three-dimensional model of general thermospheric circulation," Russ. J. Numer. Anal. Math. Modell. **28** (4), 353–380 (2013).

2. D. V. Kulyamin, V. Ya. Galin, and A. I. Pogoreltsev, "The thermosphere general circulation modeling with the parametrization of radiative processes," Russ Meteorol. Hydrol. **40,** 392–399 (2015).

3. D. V. Kulyamin, P. A. Ostanin, and V. P. Dymnikov, "Modeling the F-layer of the Earth's ionosphere. Solving the ambipolar diffusion equations," Mat. Model. **31** (4), 57–74 (2019).

4. D. V. Kulyamin and V. P. Dymnikov, "Modeling of the lower ionosphere climate," Izv., Atmos. Oceanic Phys. **51,** 272–291 (2015).

5. D. V. Kulyamin and E. M. Volodin, "INM RAS coupled atmosphere–ionosphere general circulation model INMAIM (0–130 km)," Russ. J. Numer. Anal. Math. Modell. **33** (6), 351–357 (2018).

6. R. W. Schunk and A. Nagy, *Ionospheres: Physics, Plasma Physics, and Chemistry* (Cambridge University Press, London, 2009).

7. M. C. Kelley, *The Earth's Ionosphere,* vol. 43 of *Int. Geophys. Ser.* (Academic Press, San Diego, CA, 1989).

8. E. Yigit et al., "A review of vertical coupling in the atmosphere–ionosphere system: Effects of waves, sudden stratospheric warmings, space weather, and of solar activity," J. Atmos. Sol.-Terr. Phys. **141,** 1–12 (2016).

9. R. W. Schunk, *Solar Terrestrial Energy Program (STEP): Handbook of Ionospheric Models* (Center for Atmospheric and Space Sciences, Logan, UT, USA, 1996).

10. A. D. Richmond, E. C. Ridley, and R. G. Roble, "A thermosphere/ionosphere general circulation model with coupled electrodynamics," Geophys. Res. Lett. **19** (6), 601–604 (1992).

11. S. E. McDonald, F. Sassi, and A. J. Mannucci, "SAMI3/SD-WACCM-X simulations of ionospheric variability during Northern winter 2009," Space Weather **13** (9), 568–584 (2015).

12. T. J. Fuller-Rowell et al., "Dynamics of the low-latitude thermosphere: quiet and disturbed conditions," J. Atmos. Terr. Phys. **59** (13), 1533–1540 (1997).

13. H. Wang et al., "First forecast of a sudden stratospheric warming with a coupled whole atmosphere/ionosphere model IDEA," J. Geophys. Res.: Space Phys. **119** (3), 2079–2089 (2014).

14. F. S. Bessarab et al., "E-region ionospheric storm on May 1–3, 2010: GSM TIP model representation and suggestions for IRI improvement," Adv. Space Res. **55** (8), 2124–2130 (2015).

15. A. V. Tashchilin and E. B. Romanova, "Numerical modeling of ionospheric plasma diffusion in the dipole geomagnetic field with a transversal drift," Mat. Model. **25** (1), 3–17 (2013).

16. S. A. Ishanov et al., "α-β" Iteration algorithm in problems of ionosphere plasma modeling," Mat. Model. **21** (1), 33–45 (2009).

17. V. M. Goloviznin et al., *New Algorithms of Computational Hydrodynamics for Multicore Computer Systems* (Izd-vo Mosk. Univ., Moscow, 2013) [in Russian].

18. S. V. Kostrykin, D. V. Kulyamin, P. A. Ostanin, and V. P. Dymnikov, "A model of the F-layer of the Earth's ionosphere based on transfer and ambipolar diffusion equations," Mat. Model. (2020) (in press).

Translated by B. Dribinskaya

Variational Data Assimilation in Problems of Modeling Hydrophysical Fields in Open Water Areas

V. I. Agoshkov[a, b], V. B. Zalesny[a], and T. O. Sheloput[a], *

[a]Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, 119333 Russia
[b]Moscow State University, Moscow, 119991 Russia
*e-mail: sheloput@phystech.edu

Abstract—The formulation of boundary conditions at liquid (open) boundaries is a topical problem in mathematically modeling the hydrothermodynamics of open water areas. Variational data assimilation is one method allowing one to take into account liquid boundaries in models. According to the approach considered in this paper, observational data at a certain time are given and the problem is treated as an inverse one with open boundary flows as additional unknowns. This paper presents a formulation of the general problem of the variational assimilation of observational data for a model of the hydrothermodynamics of open water areas based on the splitting method. Algorithms for the variational assimilation of temperature and sea-level data at the liquid boundary are formulated and the results of numerical experiments on the use of the algorithms in the Baltic Sea circulation model are presented.

INTRODUCTION

Modeling individual water areas (seas, gulfs, and open water areas of the ocean) and coastal areas is a topical and developing branch of mathematical geophysics. In particular, this is related to the necessity of estimating the effect of the human impact on the environment. To take into account the specificity of the water area under consideration and correctly reproduce physical phenomena in it, it is not sufficient to carry out calculations by the World Ocean general circulation model. However, when creating models of open water areas, it is necessary to solve the problem of setting the boundary conditions at liquid boundaries (by a liquid or open boundary, we mean a boundary of the water—water type separating the considered region from the World Ocean). The result directly depends on the way the boundary conditions are specified at liquid boundaries both in long-term calculations and in problems of real-time forecasting.

There are different approaches to solving the problem of taking into account liquid boundaries in models of the hydrothermodynamics of open water areas. Carrying out the calculation over the whole water area of the World Ocean—on a rough grid outside the considered water area and on a finer grid inside it—can be considered one of the approaches. In this way, a significant increase in computational efforts is encountered; however, outer liquid boundaries are absent. Some-times results of calculations over the whole water area of the World Ocean on a rough grid are used to specify boundary conditions at a liquid boundary. These ideas were developed through the method of nested grids—with feedback, with incomplete feedback, and without feedback [1]. One more widespread technique is using averaged data on flows through a liquid boundary [2] of specifying flows using observational data [3]. An individual cycle of works [4, 5] is devoted to the formulation of radiation conditions at liquid boundaries and their modifications as applied to models of hydrodynamics. Data assimilation also can be considered a promising method for solving the problem of formulating boundary conditions at a liquid boundary.

There are a lot of works considering different methods of variational data assimilation as applied to problems of modeling oceans and seas [6–10]. In [6], a survey of known investigations on this subject was presented. Data assimilation is a procedure to converge the modeling result and observational data. Models that do not involve data assimilation yield in most cases a statistically correct solution; however, in attempting to compare the observational data and the result obtained at a certain time, one can observe considerable differences between them. These contradictions are caused by the fact that general ocean-circulation models, which are often used for marine regions with a horizontal scale of more than 10 km, take into

account only the general effect from phenomena whose spatial scale is considerably less than the scale of the water area under consideration [11]. The differences can be also caused by errors in specifying the initial conditions, boundary conditions, and external data about the atmosphere, as well as by errors in the numerical solution of equations on the grid. Moreover, the modeled effects are significantly nonlinear, which results in the fact that errors can grow in time and pull the solution away from observational data [6]. Variational assimilation methods present a mathematical estimate of how the control vector (which can include different parameters of the model—coefficients, forcings, and initial and boundary conditions known not exactly) should be changed to obtain a solution maximally close to the observational data. It should be noted that the control vector obtained in this way can take physically invalid values. This can occur because real observational data reflect all phenomena occurring in nature, including those not taken into account in the mathematical model. To resolve this contradiction, the variational approach was proposed in [12]. Its idea is that not only the functions in the boundary conditions are specified inexactly, but also the boundary conditions themselves and even the model equations are formulated roughly. The essence of the method proposed in [12] is to find a generalized solution in the sense of least squares; it minimizes the residual functional, which includes summands responsible for closeness to observational data, together with selected weights.

The idea of the method proposed in [13−17] is that observational data at a certain time are given and the problem is treated as an inverse one in which functions of flows through the open boundary are additional unknowns. In [16], a class of inverse problems and problems of variational assimilation of observational data was formulated. They are related to mathematical modeling of hydrophysical fields in water areas (seas and oceans) in the presence of liquid boundaries as applied to the hydrothermodynamics model based on a system of equations in the Boussinesq and hydrostatics approximation [18, 19]. To approximate the model in time, the splitting method was used. This made it possible to consider the problem of data assimilation for the nonlinear hydrothermodynamics model on each time interval by subsequently solving simpler assimilation problems with the involvement of observational data corresponding to varying variables. A class of inverse problems about boundary functions determining boundary conditions at liquid boundaries was set and algorithms for their numerical solution based on variational assimilation of observational data were proposed. In [14], the inverse problem about the determination of the unknown function in boundary conditions at a liquid boundary for the simplest model of tides was studied. The results of the work were used in [20], where the results of numerical experiments as applied to the water area of the Sea of Okhotsk were

presented. In [15], inverse problems with the assimilation of data on the sea level from [16] were reformulated and studied as inverse problems for second-order evolutionary equations.

This work presents the formulation of the general problem of a variational assimilation of observational data for a model of the hydrothermodynamics of open water areas based on the splitting method. An algorithm of variational assimilation of data on temperature at a liquid boundary is formulated and the results of numerical experiments on using the algorithm in the Baltic Sea hydrothermodynamics model [18] as compared to the algorithm considered in [21] are presented. The algorithm of variation assimilation of data on the level at a liquid boundary is formulated and results of numerical experiments on the assimilation of observational data from satellites and level measuring stations in model [18] are presented.

1. FORMULATION OF THE PROBLEM

In this section, general symbols and equations used in the problems under consideration are introduced, and the class of inverse problems and problems of variational assimilation of observational data related to mathematical simulation of hydrophysical fields in water areas (seas and oceans) in the presence of liquid boundaries is formulated as applied to the hydrothermodynamics model based on the system of equations in the Boussinesq and hydrostatics approximation.

1.1. Let us consider a geographic (geodetic) coordinate system (λ, θ, r), where $\lambda \in [0, 2\pi]$ is the geographic longitude increasing from west to the east; $\theta \in [-\pi/2, \pi/2]$ is the geographic latitude increasing from south to north; and r is the distance from the point to the center of the Earth, the mean radius of which is taken to be equal to R_E. Instead of r, we introduce the coordinate $z = R_E - r$ of the Oz axis directed along the normal from the surface of the sphere S_R with the radius R_E to its center, i.e., in the direction of the force of gravity. The unit vectors in the λ-, θ-, and z-directions are denoted as e_λ, e_θ, and e_z, respectively. Then, the velocity vector in the ocean is written in the form $\vec{U} = u e_\lambda + v e_\theta + w e_z \equiv (\vec{u}, w)$, where $\vec{u} = (u, v)$ is the horizontal vector of the velocity in the coordinate form and w is the vertical velocity.

Let Ω denote a part of the surface of the sphere S_R, which is also called the reference surface. The ocean surface is specified by the equation $z = \xi(\lambda, \theta, t)$, where $(\lambda, \theta, R_E) \in \Omega$, and t is the time variable, $t \in [0, \overline{t}]$ $(\overline{t} < \infty)$. The bottom relief function is defined as $z = H(\lambda, \theta)$ at $(\lambda, \theta, R_E) \in \Omega$, where $H(\lambda, \theta) \geq \epsilon$, $\epsilon = \text{const} > 0$.

In what follows, we also use the following notation: $\lambda \equiv x$, $\theta \equiv y$, and $\vec{x} \equiv (x, y, z)$. The elementary volume in the domain $D(t)$:

$D(t) = \{(x, y, z) : (\lambda, \theta, R_E) \in \Omega, \xi(\lambda, \theta, t) < z < H(\lambda, \theta)\}$, $t \in [0, \overline{t}]$, is $dD = (R - z)^2 \cos y\,dxdydz$, and the element of surface Ω has the form $d\Omega = R_E^2 \cos y\,dxdy$. We introduce the following differential operations of the gradient, divergence, and total derivative in one of the spherical coordinate systems at $r \equiv r(z) \equiv R - z \cong R$, $n \equiv 1/r$, and $m \equiv 1/(r \cos y)$ (keeping for these operations the well-known notation from vector analysis):

$$\mathbf{Grad}\varphi \equiv \left(\mathbf{grad}\varphi, \frac{\partial\varphi}{\partial z}\right), \quad \mathbf{grad}\varphi \equiv \left(m\frac{\partial\varphi}{\partial x}, n\frac{\partial\varphi}{\partial y}\right),$$

$$\mathbf{Div}\vec{U} \equiv \mathbf{div}\vec{u} + \frac{1}{r^2}\frac{\partial r^2 w}{\partial z}, \quad \mathbf{div}\vec{u} \equiv m\frac{\partial u}{\partial x}$$

$$+ m\frac{\partial}{\partial y}\left[\frac{n}{m}v\right], \quad \frac{d\varphi}{dt} = \frac{\partial\varphi}{\partial t} + (\vec{U}, \nabla\varphi),$$

$$(\vec{U}, \nabla) \equiv (\vec{u}, \mathbf{grad}) + w\frac{\partial}{\partial z},$$

$$(\vec{u}, \mathbf{grad}) = um\frac{\partial}{\partial x} + vn\frac{\partial}{\partial y}.$$

The following second-order differential operators are also used below: $A_\Phi\Phi \equiv -\mathbf{Div}(a_\Phi\mathbf{Grad}\Phi)$, where $a_\Phi = \mathrm{diag}((a_\Phi)_{ii})$, $(a_\Phi)_{11} = (a_\Phi)_{22} \equiv \mu_\Phi$, and $(a_\Phi)_{33} \equiv \nu_\Phi$, and the subscript Φ can take values u, v, T, and S (i.e., designations of components of the horizontal velocity vector, temperature T, and salinity S). It is also taken that $\mu_u = \mu_v \equiv \mu$ and $\nu_u = \nu_v \equiv \nu$, and μ, ν, μ_T, μ_S, ν_T, and ν_S are supposed to be given positive and smooth functions. Then, the fourth-order operator $(A_k)^2$ will be also considered; here, the second-order operator A_k was introduced above at $A_\varphi = A_k$ and is defined by the matrix $\hat{k} = \mathrm{diag}\{k_{ii}\}$ with nonnegative diagonal elements k_{ii}, which are constant or sufficiently smooth functions. In what follows, $l = l(y)$ denotes the Coriolis parameter $l = 2\omega\sin y$, where ω is the angular velocity of the Earth's rotation and $f(u) = l + mu\sin y \equiv l + f_1(u)$.

Note that the level function $\xi = \xi(x, y, t)$ is also one of unknown functions to be determined; therefore, the domain $D(t)$ is a *domain with an unknown boundary* (or a *domain with a moving boundary*). Let us also consider a fixed domain $D = \{(\lambda, \theta, z) : (\lambda, \theta, R_E) \in \Omega, 0 < z < H(\lambda, \theta)\}$. We represent the domain boundary $\Gamma \equiv \partial D$ as a union of four disjoint parts Γ_S, $\Gamma_{w,\mathrm{op}}$, $\Gamma_{w,c}$, Γ_H, where $\Gamma_S \equiv \Omega$ is the still water surface, $\Gamma_{w,\mathrm{op}}$ is the liquid (open) part of the vertical lateral boundary, $\Gamma_{w,c}$ is the solid part of the vertical lateral boundary, and Γ_H is the ocean bottom. The characteristic functions of Γ_S, $\Gamma_{w,\mathrm{op}}$, $\Gamma_{w,c}$, and Γ_H—parts of the boundary Γ—are denoted as m_S, $m_{w,\mathrm{op}}$, $m_{w,c}$, and m_H, respec-

tively. Note that some of the parts Γ_S, $\Gamma_{w,\mathrm{op}}$, $\Gamma_{w,c}$, and Γ_H can be absent.

Hereinafter, we suppose that Ω is a multiply connected manifold on S_R, and the boundaries $\partial\Omega$, Γ are supposed to be piecewise smooth of class $C^{(2)}$ and locally satisfying the Lipschitz condition. The unit vector of the outer normal to Γ is denoted as $\vec{N} \equiv (N_1, N_2, N_3)$. Note that $\vec{N} = (0, 0, -1)$ on Γ_S and $\vec{N} = (N_1, N_2, 0)$ on $\Gamma_w = \Gamma_{w,\mathrm{op}} \bigcup \Gamma_{w,c}$; the vector $\vec{n} \equiv (N_1, N_2) \equiv (n_1, n_2)$ is the unit vector of the outer normal to $\partial\Omega$. The expression of the components N_1, N_2, and N_3 is determined by the chosen parametric representation of one or another part of the boundary.

When considering the velocity vector at boundary Γ, we denote its normal component as U_n: $U_n = \vec{U} \cdot \vec{N} = uN_1 + vN_2 + wN_3$. Further, let

$$U_n^{(+)} \equiv \frac{|U_n| + U_n}{2}, \quad U_n^{(-)} \equiv \frac{|U_n| - U_n}{2}.$$

We remark that $U_n = U_n^{(+)} - U_n^{(-)}$ on Γ.

1.2. Let us write the system of equations in the Boussinesq and hydrostatics approximation in the domain D in variables (λ, θ, z) at $t \in (0, \overline{t})$ [22], but taking the Lamé coefficients as corresponding to a spherical coordinate system [16, 23]:

$$\begin{cases} \dfrac{d\vec{u}}{dt} + \begin{bmatrix} 0 & -f \\ f & 0 \end{bmatrix}\vec{u} - g\mathbf{grad}\xi + A_u\vec{u} + (A_k)^2\vec{u} \\[2mm] = \vec{f} - \dfrac{1}{\rho_0}\mathbf{grad}P_a - \dfrac{g}{\rho_0}\mathbf{grad}\displaystyle\int_0^z \rho_1(T, S)dz', \\[4mm] \dfrac{\partial\xi}{\partial t} - m\dfrac{\partial}{\partial x}\left(\displaystyle\int_0^H \Theta(z)u\,dz\right) - m\dfrac{\partial}{\partial y}\left(\displaystyle\int_0^H \Theta(z)\dfrac{n}{m}v\,dz\right) = f_3, \\[4mm] \dfrac{dT}{dt} + A_T T = f_T, \quad \dfrac{dS}{dt} + A_S S = f_S, \end{cases} \quad (1)$$

where

$$\rho_1(T, S) = \rho_0\beta_T(T - T^{(0)}) + \rho_0\beta_S(S - S^{(0)})$$
$$+ \gamma\rho_0\beta_{TS}(T, S) + f_P,$$

$\vec{f} = (f_1, f_2)$, f_T, f_S, f_P are given functions of internal sources; $g = \mathrm{const} > 0$; ρ_0, $T^{(0)}$, and $S^{(0)}$ are unperturbed values of water density, temperature, and salinity; β_T and β_S are coefficients (assumed to be constant); $\beta_{TS}(T, S)$ and P_a are given functions; and γ is a numerical parameter. Hereinafter, the following weight function is used: $\Theta(z) \equiv r(z)/R$. In the case considered below, $\vec{f} \equiv g\mathbf{grad}G$ at a certain scalar func-

tion $G = G(\lambda, \theta, t)$, for example, $G \equiv \xi^+$—the static tide described in more detail in [23].

Let us pay attention to the fact that the coordinates (λ, θ, z) are geodetic in their physical meaning; however, by virtue of the approximate expression of the Lamé coefficients, system (1) takes the form in one of the spherical coordinate systems [16].

When considering (1) in $D \times (0, \bar{t})$, the following boundary and initial values can be specified [16]:

(i) Boundary conditions on Γ_S:

$$
\begin{cases}
\left(\int_0^H \Theta \vec{u}\, dz \right) \vec{n} + m_{w,\text{op}}\sqrt{gH}\xi = m_{w,\text{op}}\sqrt{gH} d_s \ \text{ on } \partial\Omega, \\[2mm]
U_n^{(-)} u - \nu \dfrac{\partial u}{\partial z} - k_{33}\dfrac{\partial}{\partial z} A_k u = \tau_x^{(a)} \big/ \rho_0, \\[2mm]
U_n^{(-)} v - \nu \dfrac{\partial v}{\partial z} - k_{33}\dfrac{\partial}{\partial z} A_k v = \tau_y^{(a)} \big/ \rho_0, \\[2mm]
A_k u = 0,\ A_k v = 0,\ U_n^{(-)} T - \nu_T \dfrac{\partial T}{\partial z} + \gamma_T (T - T_a) = Q_T, \\[2mm]
U_n^{(-)} S - \nu_S \dfrac{\partial S}{\partial z} + \gamma_S (S - S_a) = Q_S,
\end{cases}
\tag{2}
$$

where $\tau_x^{(a)}$, $\tau_y^{(a)}$ are components of tangential stress vectors of wind along the Ox and Oy axes, respectively, on the surface $z = 0$; γ_T, γ_S, T_a, and S_a, are given functions; and Q_T, Q_S are stream functions specified on the whole Γ.

(ii) Boundary conditions on $\Gamma_{w,c}$ (on the solid lateral wall):

$$
U_n = 0, \quad A_k \tilde{U} = 0, \quad \frac{\partial \tilde{U}}{\partial N_u}\tau_w + \left(\frac{\partial}{\partial N_u} A_k \tilde{U} \right)\tau_w = 0,
$$
$$
\frac{\partial T}{\partial N_T} = 0, \quad \frac{\partial S}{\partial N_S} = 0,
\tag{3}
$$

where $\tau_w = (-N_2,\ N_1,\ 0)$, $\tilde{U} \equiv (u, v, 0) \equiv (\vec{u}, 0)$, $\varphi/\partial N_\varphi \equiv \vec{N} \cdot \hat{a}_\varphi \cdot \mathbf{Grad}\varphi$, $\varphi = u, v, T, S$.

(iii) Boundary conditions on $\Gamma_{w,op}$ (on the liquid part of the lateral wall):

$$
\begin{cases}
U_n^{(-)}(\tilde{U} \cdot \vec{N}) + \dfrac{\partial \tilde{U}}{\partial N_u}\cdot \vec{N} + \left(\dfrac{\partial}{\partial N_u} A_k \tilde{U} \right) \cdot \vec{N} = U_n^{(-)} d, \\[2mm]
A_k \tilde{U} = 0, \quad U_n^{(-)}(\tilde{U} \cdot \tau_w) + \dfrac{\partial \tilde{U}}{\partial N_u}\cdot \tau_w \\[2mm]
+ \left(\dfrac{\partial}{\partial N_u} A_k \tilde{U} \right) \cdot \tau_w = 0, \quad U_n^{(-)} T + \dfrac{\partial T}{\partial N_T} \\[2mm]
= d_T + Q_T, \quad U_n^{(-)} S + \dfrac{\partial S}{\partial N_S} = d_S + Q_S.
\end{cases}
\tag{4}
$$

(iv) Boundary conditions on Γ_H (on the bottom):

$$
\begin{cases}
U_n = 0, \quad A_k \tilde{U} = 0, \quad \dfrac{\partial T}{\partial N_T} = 0, \quad \dfrac{\partial S}{\partial N_S} = 0, \\[2mm]
\dfrac{\partial \tilde{U}}{\partial N_u}\cdot \tau_x + \left(\dfrac{\partial}{\partial N_k} A_k \tilde{U} \right) \cdot \tau_x = -\tau_x^{(b)}\rho_0, \\[2mm]
\dfrac{\partial \tilde{U}}{\partial N_u}\cdot \tau_y + \left(\dfrac{\partial}{\partial N_u} A_k \tilde{U} \right) \cdot \tau_y = -\tau_y^{(b)}\rho_0,
\end{cases}
\tag{5}
$$

where τ_x, τ_y is the system of unit orthogonal tangent vectors on the surface $z = 0$; and $\tau_x^{(b)}$, $\tau_y^{(b)}$ are projections of the stress vector of bottom friction on the Ox and Oy axes, respectively.

(v) Initial conditions for u, v, T, S, and ξ:

$$
u = u^0, \quad v = v^0, \quad T = T^0, \quad S = S^0,
$$
$$
\xi = \xi^0 \quad \text{at} \quad t = 0,
\tag{6}
$$

where u^0, v^0, T^0, S^0, and ξ^0 are given functions.

Let functions d, d_s, d_T, d_S, and f_3 be given. Then, *the problem of large-scale dynamics of the ocean in terms of functions u, v, ξ, T, and S is formulated as follows: find u, v, ξ, T, and S satisfying (1)−(6).*

Note that the diffusion operators in the equations for u and v do not take into account some differential lesser order operators that are significant near poles. Therefore, generally speaking, system (1) must be considered in a domain with the excluded points of poles. In this work, we restrict ourselves by the consideration of system of equations (1).

To approximate problem (1)−(6) in time, the splitting method is used. Let the grid $0 = t_0 < t_1 < \ldots < t_{J-1} < t_J = \bar{t}$, $\Delta t_j = t_j - t_{j-1}$ be introduced on $[0, \bar{t}]$. We present one of the schemes obtained by the *weak approximation method*. It approximates the initial problem by considering subproblems entering into this scheme in classical formulations under the assumption that all components of the solutions possess the necessary smoothness with respect to all independent variables. This scheme consists of the steps (stages) listed below. The formulation involves indices: the subscript in the unknown variables (T, ξ, etc.) denotes the step number of the splitting method; the superscript denotes the number of the time step.

STEP 1. The solution of the jth subproblem at this step of the splitting method satisfies the system of relationships of the form (at $\bar{u}_1^j \equiv \vec{u}, \ldots, \xi_1^j \equiv \xi$):

$$\begin{cases} \dfrac{\partial \vec{u}}{\partial t} + \begin{bmatrix} 0 & -\ell \\ \ell & 0 \end{bmatrix} \vec{u} - g \cdot \mathbf{grad}\xi = 0, \\[2mm] \dfrac{\partial T}{\partial t} = 0, \ \dfrac{\partial S}{\partial t} = 0 \ \ \text{in} \ \ D \times (t_{j-1}, t_j), \\[2mm] \dfrac{\partial \xi}{\partial t} - \mathbf{div}\left(\int\limits_0^H \Theta \vec{u} dz \right) = f_3 \ \ \text{in} \ \ \Omega \times (t_{j-1}, t_j), \\[2mm] \left(\int\limits_0^H \Theta \vec{u} dz \right) \cdot \vec{n} + m_{w,\mathrm{op}} \sqrt{gH} \xi \\[2mm] = m_{w,\mathrm{op}} \sqrt{gH} d_s \ \ \text{on} \ \ \partial\Omega \times (t_{j-1}, t_j), \\[2mm] \vec{u} = \vec{u}^0, \ \ \xi = \xi^0, \ \ T = T^0, \\[2mm] S = S^0 \ \ \text{at} \ \ t = t_0 \ (j = 1), \\[2mm] \vec{u} = \vec{u}_5^{j-1}, \ \ \xi = \xi_5^{j-1}, \ \ T = T_5^{j-1}, \\[2mm] S = S_5^{j-1} \ \ \text{at} \ \ t = t_{j-1} \ (j = 2, 3, \ldots, J), \end{cases} \quad (7)$$

where $\vec{u}^0 = (u^0, v^0)$, $\vec{u}_5^{j-1} = \left(u_5^{j-1}, v_5^{j-1} \right)$.

From (7), we conclude that for this problem we have

$$T = T_1^0, \ \ S = S_1^0 \ \ \text{in} \ \ D \times (t_0, t_1) \ (j = 1),$$
$$T = T_5^{j-1}, \ \ S = S_5^{j-1} \ \ \text{in} \ \ D \times (t_{j-1}, t_j) \quad (8)$$
$$\text{for} \ \ j = 2, 3, \ldots, J.$$

Finding $\vec{u}$ and ξ is reduced to solving problems of the form

$$\begin{cases} \dfrac{\partial \underline{U}}{\partial t} + \begin{bmatrix} 0 & -\ell \\ \ell & 0 \end{bmatrix} \underline{U} - g \cdot \mathbf{grad}\xi = 0 \ \ \text{in} \ \ \Omega \times (t_{j-1}, t_j), \\[2mm] \dfrac{\partial \xi}{\partial t} - \mathbf{div}\left(H_1 \underline{U} \right) = f_3 \ \ \text{in} \ \ \Omega \times (t_{j-1}, t_j), \\[2mm] H_1(\underline{U} \cdot \vec{n}) + m_{w,\mathrm{op}} \sqrt{gH} \xi \\[2mm] = m_{w,\mathrm{op}} \sqrt{gH} d_s \ \ \text{on} \ \ \partial\Omega \times (t_{j-1}, t_j), \\[2mm] \underline{U} = \underline{U}^0, \ \ \xi = \xi^0 \ \ \text{at} \ \ t = t_0; \\[2mm] \underline{U} = \underline{U}_5^{j-1}, \ \ \xi = \xi_5^{j-1} \ \ \text{at} \ \ t = t_{j-1} \ (j = 2, 3, \ldots, J), \end{cases} \quad (9)$$

where

$$H_1 \equiv \int\limits_0^H \Theta(z) dz, \ \ \underline{U} = \dfrac{1}{H_1} \int\limits_0^H \Theta \underline{u} dz$$
$$\equiv (U_1(x, y, t), U_2(x, y, t))$$
$$= \left(\dfrac{1}{H_1} \int\limits_0^H \Theta u dz, \dfrac{1}{H_1} \int\limits_0^H \Theta v dz \right),$$

and the following problem:

$$\begin{cases} \dfrac{\partial \underline{W}}{\partial t} + \begin{bmatrix} 0 & -\ell \\ \ell & 0 \end{bmatrix} \underline{W} = 0 \ \ \text{in} \ \ D \times (t_{j-1}, t_j), \\[2mm] \underline{W} = (\vec{u}^0 - \underline{U}^0) \ \ \text{at} \ \ t = t_0 \ (j = 1), \\[2mm] \underline{W} = \left(\vec{u}_5^{j-1} - \underline{U}_5^{j-1} \right) \ \ \text{at} \ \ t = t_{j-1} \ (j = 2, 3, \ldots, J). \end{cases} \quad (10)$$

Solving (9) and (10), we obtain $\xi = \xi_1^j$, as well as

$$\vec{u}_1^j \equiv \vec{u} = \underline{U} + \underline{W} \ \ \text{in} \ \ D \times (t_{j-1}, t_j), \quad (11)$$

which completes the implementation of this step of the splitting method.

STEP 2. The solution of the jth subproblem satisfies the system of the form (at $\vec{u} \equiv \vec{u}_2^{j-1}, \ldots, \xi \equiv \xi_2^j$)

$$\begin{cases} \dfrac{\partial T}{\partial t} + (\vec{U}, \mathbf{Grad})T + \mathrm{A}_T T = f_T \ \text{in} \ D \times (t_{j-1}, t_j), \\[2mm] U_n^{(-)}T - \nu_T \dfrac{\partial T}{\partial z} + \gamma_T(T - T_a) = Q_T^j, \ z = 0, \ t \in (t_{j-1}, t_j), \\[2mm] \dfrac{\partial T}{\partial N_T} = 0 \ \text{on} \ \Gamma_H \cup \Gamma_{w,c}, \ t \in (t_{j-1}, t_j), \\[2mm] U_n^{(-)}T + \dfrac{\partial T}{\partial N_T} = d_T \ \text{on} \ \Gamma_{w,\mathrm{op}}, \ t \in (t_{j-1}, t_j), \\[2mm] \dfrac{\partial \vec{u}}{\partial t} + \lambda_T \vec{u} = g \cdot \mathbf{grad}\left(\int\limits_0^z \beta_T T dz' \right) \ \text{in} \ D \times (t_{j-1}, t_j), \\[2mm] \dfrac{\partial \xi}{\partial t} = 0 \ \text{in} \ \Omega \times (t_{j-1}, t_j), \\[2mm] \dfrac{\partial S}{\partial t} = 0 \ \text{in} \ D \times (t_{j-1}, t_j); \ \vec{u} = u_1^j(t_j), \\[2mm] \xi = \xi_1^j, \ T = T_1^j, \ S = S_1^j \ \text{at} \ t = t_{j-1}, \end{cases} \quad (12)$$

where $\vec{U} \equiv \vec{U}^j = (u^j, v^j, w^j)$, $U_n^{(-)}$ is calculated based on $\vec{U}^j$, and λ_T and λ_S are sufficiently small given constants introduced to justify the splitting scheme (see [16]).

Note that in this subproblem we have

$$\xi \equiv \xi_2^j = \xi_1^j \ \ \text{in} \ \ \Omega \times (t_{j-1}, t_j),$$
$$S \equiv S_2^j = S_1^j \ \ \text{in} \ \ D \times (t_{j-1}, t_j) \quad (13)$$

and, after solving the problem for $T \equiv T_2^j$, $\vec{u} \equiv \vec{u}_2^j$ is calculated by the formula

$$\vec{u}_2^j = \vec{u}_1^j(t_j) \cdot e^{-\lambda_T(t - t_{j-1})} + g \cdot \beta_T \int\limits_{t_{j-1}}^t \mathbf{grad}\left(\int\limits_0^z T_2^j dz' \right) \\ \times e^{-\lambda_T(t' - t_{j-1})} dt' \ \ \text{in} \ \ D(t_{j-1}, t_j). \quad (14)$$

Thus, the implementation of this step is reduced in fact to solving a subsystem of equations to calculate $T \equiv T_2^j$.

STEP 3. This step is similar to the previous one and is reduced to solving the problem for $S \equiv S_3^j$:

STEP 4. The subproblem solved here has the form

$$\begin{cases} \dfrac{\partial S}{\partial t} + (\vec{U}, \mathbf{Grad})S + A_S S = f_S & \text{in } D \times (t_{j-1}, t_j), \\[2mm] U_n^{(-)} S - \nu_S \dfrac{\partial S}{\partial z} + \gamma_S (S - S_a) = Q_S^j, \\[2mm] z = 0, \quad t \in (t_{j-1}, t_j), \\[2mm] \dfrac{\partial S}{\partial N_S} = 0 \quad \text{on } \Gamma_H \cup \Gamma_{w,c}, \quad t \in (t_{j-1}, t_j), \\[2mm] U_n^{(-)} S + \dfrac{\partial S}{\partial N_S} = d_S \quad \text{on } \Gamma_{w,\text{op}}, \quad t \in (t_{j-1}, t_j), \\[2mm] S = \left(S_2^j = S_1^j \right) \quad \text{at } t = t_{j-1}. \end{cases} \tag{15}$$

Then it is taken that

$$\begin{cases} \vec{u}_3^j = \vec{u}_2^j(t_j) \cdot e^{-\lambda_S(t - t_{j-1})} + g \cdot \beta_S \displaystyle\int_{t_{j-1}}^{t} \mathbf{grad}\left(\int_0^z S\,dz' \right) \\[2mm] \times\, e^{-\lambda_S(t' - t_{j-1})} dt' \quad \text{in } D \times (t_{j-1}, t_j), \\[2mm] \xi_3^j = \xi_2^j = \xi_1^j \quad \text{in } \Omega \times (t_{j-1}, t_j), \\[2mm] T_3^j = T_2^j \quad \text{in } D \times (t_{j-1}, t_j). \end{cases} \tag{16}$$

Functions $\vec{U}, U_n^{(-)}$ are determined in the same way as in the previous step.

$$\begin{cases} \xi_4 = \xi_1^j \quad \text{in } \Omega \times (t_{j-1}, t_j), \quad T_4 = T_2^j \\[2mm] \text{in } D \times (t_{j-1}, t_j), \quad S_4 = S_3^j \quad \text{in } D \times (t_{j-1}, t_j), \\[2mm] \dfrac{\partial \vec{u}_4}{\partial t} + A_k^2 \vec{u}_4 = 0 \quad \text{in } D \times (t_{j-1}, t_j), \quad A_k \vec{u}_4 = 0, \\[2mm] k_{33} \dfrac{\partial}{\partial z} A_k \vec{u}_4 = 0 \quad \text{on } \Gamma_S \times (t_{j-1}, t_j), \\[2mm] (\tilde{U}_4 \cdot \vec{N}) = 0, \quad A_k \tilde{U}_4 = 0, \\[2mm] \left(\dfrac{\partial}{\partial N_k} A_k \tilde{U}_4 \right) \cdot \tau_w = 0 \quad \text{on } \Gamma_{w,c} \times (t_{j-1}, t_j), \\[2mm] \left(\dfrac{\partial}{\partial N_k} A_k \tilde{U}_4 \right) \cdot \vec{N} = 0, \quad A_k \tilde{U}_4 = 0, \\[2mm] \dfrac{\partial \tilde{U}_4}{\partial N_k} \cdot \tau_w + \left(\dfrac{\partial}{\partial N_k} A_k \tilde{U}_4 \right) \cdot \tau_w = 0 \quad \text{on } \Gamma_{w,\text{op}} \times (t_{j-1}, t_j), \\[2mm] (\tilde{U}_4 \cdot \vec{N}) = 0, \quad A_k \tilde{U}_4 = 0, \quad \dfrac{\partial \tilde{U}_4}{\partial N_k} \cdot \tau_w \\[2mm] + \left(\dfrac{\partial}{\partial N_k} A_k \tilde{U}_4 \right) \cdot \tau_w = 0 \quad \text{on } \Gamma_H \times (t_{j-1}, t_j), \\[2mm] \text{under the condition } \tilde{U}_4 = \tilde{U}_3^j \quad \text{at } t = t_{j-1}. \end{cases}$$

STEP 5. The system of equations corresponding to this stage of solution of the problem has the form (at $\vec{u} \equiv \vec{u}_5^j, \ldots, \xi \equiv \xi_5^j$)

$$\begin{cases} \dfrac{\partial \vec{u}}{\partial t} + (\vec{U}, \mathbf{Grad})\vec{u} + A_u \vec{u} - (\lambda_T + \lambda_S)\vec{u} \\[2mm] + \begin{bmatrix} 0 & -\ell_1(\vec{u}^j) \\ \ell_1(\vec{u}^j) & 0 \end{bmatrix} \vec{u} = \vec{f} - \dfrac{1}{\rho_0} \mathbf{grad}\left(P_1(\bar{T}^j, \bar{S}^j) \right) \text{ in } D \times (t_{j-1}, t_j), \\[2mm] \dfrac{\partial \xi}{\partial t} = 0 \text{ in } \Omega \times (t_{j-1}, t_j), \dfrac{\partial T}{\partial t} = 0, \dfrac{\partial S}{\partial t} = 0 \text{ in } D \times (t_{j-1}, t_j), \\[2mm] U_n^{(-)} \vec{u} - \nu \dfrac{\partial \vec{u}}{\partial z} = \dfrac{\underline{\tau}^{(a)}}{\rho_0}, \ \underline{\tau}^{(a)} = (\tau_x^{(a)}, \tau_y^{(a)}), \ z = 0, \ t \in (t_{j-1}, t_j), \\[2mm] \tilde{U} \cdot \vec{N} = 0, \dfrac{\partial \tilde{U}}{\partial N_U} \cdot \tau_w = 0 \text{ on } \Gamma_{w,c}, \ t \in (t_{j-1}, t_j), \\[2mm] U_n^{(-)} \left(\tilde{U} \cdot \vec{N} \right) + \dfrac{\partial \tilde{U}}{\partial N_U} \cdot \vec{N} = U_n^{(-)} d \text{ on } \Gamma_{w,\text{op}}, \ t \in (t_{j-1}, t_j), \\[2mm] \dfrac{\partial \vec{u}}{\partial N_u} = -\dfrac{\underline{\tau}^{(b)}}{\rho_0} \text{ on } \Gamma_H, \ t \in (t_{j-1}, t_j), \\[2mm] \vec{u} = u_4^j(t_j), \xi = \xi_4^j(t_j), T = T_4^j(t_j), S = S_4^j(t_j) \text{ at } t = t_{j-1}, \end{cases} \tag{17}$$

where $\underline{\tau}^{(a)} = \left(\tau_x^{(a)}, \tau_y^{(a)} \right)$, $\underline{\tau}^{(b)} = \left(\tau_x^{(b)}, \tau_y^{(b)} \right)$, $\tilde{U} \equiv \tilde{U}_4^j = \left(u_4^j, v_4^j, 0 \right)$, and $\vec{U}, U_n^{(-)}$ are determined as above.

From (17), we have

$$\begin{aligned} \xi_5^j &= \xi_4^j = \xi_3^j = \xi_2^j = \xi_1^j \quad \text{in } \Omega \times (t_{j-1}, t_j), \\ T_5^j &= T_4^j = T_3^j = T_2^j, \\ S_5^j &= S_4^j = S_3^j \quad \text{in } D \times (t_{j-1}, t_j), \end{aligned} \tag{18}$$

and implementation (17) is reduced to solving the problem for $\vec{u} \equiv \vec{u}_5^j = (u_5^j, v_5^j)$. Then, it is taken that

$$\Phi^j \equiv (u^j, v^j, \xi^j, T^j, S^j) \equiv (u_5^j, v_5^j, \xi_1^j, T_2^j, S_3^j) \quad (19)$$

is an approximate solution of problem (1)–(6) on the interval (t_{j-1}, t_j). The calculation of w^j and P_1 is carried out by formulas presented in [16].

Note that the splitting scheme written above can be considered an initial problem for which a class of inverse problems is formulated below.

1.3. Let functions d_T on $\Gamma_{w,op} \times (0, \bar{t})$ and d_s on $\Gamma_{w,op} \cap \partial\Omega \times (0, \bar{t})$ be taken as additional unknowns. In this work, functions d, d_S, and f_3 are supposed to be given (the consideration of problems with unknown d and f_3 can be found in [15, 16]; function d_S is retrieved similarly to d_T).

Let us introduce closure conditions (additional equations). Suppose that observational data on temperature T_{obs} are available on a certain portion $\Gamma_{o,T}$ of the boundary $\Gamma = \partial D$ and during a time interval from $(0, \bar{t})$ (we mention the possibility of the case where $\Gamma_{o,T} \subset \Gamma_{w,op}$, $\Gamma_{w,op} \subset \Gamma_{o,T}$, $\Gamma_{o,T} \equiv \Gamma_{w,op}$, $\Gamma_{o,T} = \Gamma$, $mes(\Gamma_{o,T} \cap \Gamma_{w,op}) > 0$). The characteristic function of the point set on which observations T_{obs} are available is denoted as $\chi_T \equiv \chi_T(x, y, z, t)$ (for definiteness, we assume that $T_{obs} \equiv 0$ on $(\Gamma \setminus \Gamma_{o,T}) \times (0, \bar{t})$). Similarly, for observations $\xi_{obs,\partial\Omega}$ on $\partial\Omega \times (0, \bar{t})$, the characteristic function $\chi_{\xi,\partial\Omega}$ is introduced.

Let us consider the following closure equations:

$$\begin{aligned} \chi_T T &= \chi_T T_{obs} \quad \text{on} \quad \Gamma \times (0, \bar{t}), \\ \chi_{\xi,\partial\Omega} \xi &= \chi_{\xi,\partial\Omega} \xi_{obs} \quad \text{on} \quad \partial\Omega \times (0, \bar{t}). \end{aligned} \quad (20)$$

Now, *the general inverse problem is formulated as follows: find* $\Phi = (u, v, \xi, T, S)$ *and* $V = (d_T, d_s)$ *such that Eqs. (1)–(6) and (20) are satisfied.*

In this form, problem (1)–(6), (20) is an exact control problem and, as a rule, it has no solutions. Therefore, it is expedient to pass from (1) to a weak formulation of these equations with introduction of a regularization procedure [24]. For this purpose, we introduce a functional of the form

$$\Im_{\alpha,\gamma}(V, \Phi) = \sum_{j=1}^{J} \Im_{\alpha,\gamma}^{(j)}(V, \Phi)$$

$$= \sum_{j=1}^{J} \left(\Im_{\alpha}^{(j)}(V) + \Im_{\gamma}^{(j)}(\Phi) \right), \quad (21)$$

where $(j = \overline{1, J})$:

$$\Im_{\alpha}^{(j)}(V) = \frac{\alpha_1}{2} \int\limits_{t_{j-1}}^{t_j} \int\limits_{\Gamma_{w,op}} \left(d_T - d_T^{(0)} \right)^2 d\Gamma dt$$

$$+ \frac{\alpha_2}{2} \int\limits_{t_{j-1}}^{t_j} \int\limits_{\partial\Omega} m_{w,op} g\left(d_s - d_s^{(0)} \right)^2 d\Gamma dt,$$

$$\Im_{\gamma}^{(j)}(\Phi) = \frac{\gamma_1}{2} \int\limits_{t_{j-1}}^{t_j} \int\limits_{\Gamma} \chi_T (T - T_{obs})^2 d\Gamma dt$$

$$+ \frac{\gamma_2}{2} \int\limits_{t_{j-1}}^{t_j} \int\limits_{\partial\Omega} \chi_{\xi,\partial\Omega} (\xi - \xi_{obs})^2 d\Gamma dt,$$

where α_1, α_2 and γ_1, γ_2 are given nonnegative (dimensional) constants, which are chosen in one way or other depending on the problem under consideration and procedures of its solution; α_1, α_2 are the regularization parameters; and γ_1, γ_2 are the weight parameters.

Functions $d_T^{(0)}$, $d_s^{(0)}$ are supposed to be given, e.g., $d_T^{(0)} \equiv 0$, $d_s^{(0)} \equiv 0$; they are introduced for generality of the consideration. At the same time, if these functions are reasonable approximations to the unknown d_T, d_s, finding the minimum $\Im_{\alpha,\gamma}$ at $\alpha_k > 0$, $k = 1, 2$ will lead to the requirement of a definite closeness of the unknown d_T and d_s to $d_T^{(0)}$ and $d_s^{(0)}$, respectively.

Let us now formulate the *general problem of variational assimilation of observational data* for model (7)–(17) which is an approximation to (1)–(6): find Φ^j, V^j such that relationships (7)–(17) are satisfied and, at the same time,

$$\inf_{V^j} \Im_{\alpha,\gamma}^{(j)}(V^j, \Phi^j), \quad j = 1, 2, 3, .., J. \quad (22)$$

Pay attention to the fact that in the formulation of the general problem *assimilation of observational data is implemented by the corresponding subproblem*. For example, the assimilation of data on the level is performed at Step 1 of the splitting method; assimilation of temperature is performed at Step 2 of this method. Such an approach to solving the whole problem is considerably simpler when compared to the version in which assimilation is implemented by the complete system of equations, when one has to solve both the full system of equations of the problem and the complete system of adjoint problems related to each other at each iteration of the external iterative process.

2. PROBLEM OF VARIATIONAL ASSIMILATION OF DATA ON TEMPERATURE

In this section, the problem of variational assimilation of data on temperature at a liquid boundary and the algorithm of its solution are formulated. The section also presents results of numerical experiments on

using the algorithm in the model of Baltic Sea hydro-thermodynamics.

Let us write the problem of heat transport−diffusion (12) from Step 2 of the splitting method in the following form:

$$
\begin{cases}
T_t + \vec{U} \cdot \mathbf{Grad}\, T - \mathbf{Div}(\hat{a}_T \mathbf{Grad}\, T) \\
= f_T \quad \text{in} \quad D \times (t_{j-1}, t_j), \\
U_n^{(-)} T + \dfrac{\partial T}{\partial N_T} + \gamma_T (T - T_a) \\
= m_{w,\text{op}} d_T + Q_F \quad \text{on} \quad \Gamma \times (t_{j-1}, t_j), \\
T = T_{j-1} \quad \text{at} \quad t = t_{j-1} \quad \text{in} \quad D,
\end{cases} \tag{23}
$$

where f_T, γ_T, T_a, and Q_F are given functions; $\partial T / \partial N_T \equiv \vec{N} \cdot \hat{a}_T \cdot \mathbf{Grad}\, T$, $\gamma_T \equiv 0$, $T_a \equiv 0$ on $\Gamma \backslash \Gamma_S$ $\forall t \in (t_{j-1}, t_j)$; $U_n^{(-)} = 0$ on $(\Gamma_{w,c} \cup \Gamma_H \cup \Gamma_S) \times (t_{j-1}, t_j)$; and $d_T \equiv 0$ on $(\Gamma \backslash \Gamma_{w,\text{op}}) \times (t_{j-1}, t_j)$. After solving this problem, function T_j is taken as an approximation to the component T (the temperature) of the exact solution on $D \times (t_{j-1}, t_j)$.

Let observational data on temperature on portion $\Gamma_{o,T}$ of the boundary $\Gamma = \partial D$ be available for almost all $t \in (t_{j-1}, t_j)$; we denote them as T_{obs}. Then, considering the function d_T as an additional unknown (or as a function known approximately and requiring a refinement), one can formulate the problem of temperature data assimilation at a given step of the splitting scheme as an inverse problem, i.e., as *a problem about finding T and d_T satisfying system (23) and closure equation*

$$
\chi_T T = \chi_T T_{\text{obs}} \quad \text{on} \quad \Gamma \times (t_{j-1}, t_j), \tag{24}
$$

where χ_T is the characteristic function of the boundary $\Gamma_{o,T} \subseteq \Gamma$.

For the approximate solution of the inverse problem, we reformulate it as a problem of minimizing a functional:

$$
J_\alpha(d_T, T(d_T)) \equiv \frac{\alpha}{2} \int_{t_{j-1}}^{t_j} \int_\Gamma m_{w,\text{op}} (d_T - d_T^{(0)})^2 d\Gamma dt
$$
$$
+ \frac{1}{2} \int_{t_{j-1}}^{t_j} \int_\Gamma \chi_T (T - T_{\text{obs}})^2 d\Gamma dt \to \inf_{d_T}, \tag{25}
$$

where $T = T(d_T)$ is a solution of problem (23) and α is a small nonnegative parameter which has the sense of A.N. Tikhonov's regularization parameter [25].

If $\Gamma_{o,T}$ coincides with $\Gamma_{w,\text{op}}$, one can demonstrate [21, 26] that inverse problem (23), (24) is uniquely and densely solvable. In this case, based on the theory presented in [24], one can assert that constructing an approximate solution of problem (25) at a small $\alpha > 0$ is sufficient for an approximate solution of problem (23), (24). Introducing the adjoint prob-

lem, we write the gradient descent method for solving the extremum problem (25):

$$
\begin{cases}
(T_k)_t + \vec{U} \cdot \mathbf{Grad}\, T_k - \mathbf{Div}(\hat{a}_T \mathbf{Grad}\, T_k) \\
= f_T \quad \text{in} \quad D \times (t_{j-1}, t_j), \\
U_n^{(-)} T_k + \dfrac{\partial T_k}{\partial N_T} + \gamma_T (T_k - T_a) = m_{w,\text{op}} d_{T,k} \\
+ Q_F \quad \text{on} \quad \Gamma \times (t_{j-1}, t_j), \\
T_k = T_{j-1} \quad \text{at} \quad t = t_{j-1} \quad \text{in} \quad D;
\end{cases} \tag{26}
$$

$$
\begin{cases}
-(T_k^*)_t - \mathbf{Div}(\vec{U} T_k^*) - \mathbf{Div}(\hat{a}_T \mathbf{Grad}\, T_k^*) \\
= 0 \quad \text{in} \quad D \times (t_{j-1}, t_j), \\
T_k^* = 0 \quad \text{at} \quad t = t_j \quad \text{in} \quad D, \\
U_n^{(+)} T_k^* + \dfrac{\partial T_k^*}{\partial N_T} + \gamma_T T_k^* \\
= \chi_{\text{obs}}(T_k - T_{\text{obs}}) \quad \text{on} \quad \Gamma \times (t_{j-1}, t_j);
\end{cases} \tag{27}
$$

$$
d_{T,k+1} = d_{T,k} - \tau_k m_{w,\text{op}} \left[\alpha (d_{T,k} - d_T^{(0)}) + T_k^* \right]. \tag{28}
$$

In view of dense solvability of the problem when $\Gamma_{w,\text{op}} = \Gamma_{o,T}$ at a sufficiently small $\tau_k = \tau = \text{const} > 0$, we have [24]

$$
\int_{t_{j-1}}^{t_j} \int_{\Gamma_{o,T}} (T_k(\alpha) - T_{\text{obs}})^2 d\Gamma dt \to 0
$$
$$
\text{as} \quad \alpha \to 0, \quad k \to \infty.
$$

Moreover, if there exist a solution d_T, T of inverse problem (23), (24), it is unique and the following statement about convergence of algorithm (26)−(28) is valid:

$$
\int_{t_{j-1}}^{t_j} \int_{\Gamma_{w,\text{op}}} (d_{T,k}(\alpha) - d_T)^2 d\Gamma dt
$$
$$
+ \int_{t_{j-1}}^{t_j} \int_D (T_k(\alpha) - T)^2 dD dt \to 0
$$
$$
\text{as} \quad \alpha \to 0, \quad k \to \infty.
$$

When the conditions of dense solvability $\inf J_\alpha \cong 0$ as $\alpha \to +0$ are satisfied, one can just take [24]

$$
\tau_k \cong \frac{J_\alpha(d_{T,k})}{\left\| J_\alpha'(d_{T,k}) \right\|^2} \cong \frac{1/2 \int_{t_{j-1}}^{t_j} \int_{\Gamma_{\text{obs}}} (T_k - T_{\text{obs}})^2 d\Gamma dt}{\int_{t_{j-1}}^{t_j} \int_{\Gamma_{w,\text{op}}} (T_k^*)^2 d\Gamma dt} \tag{29}
$$

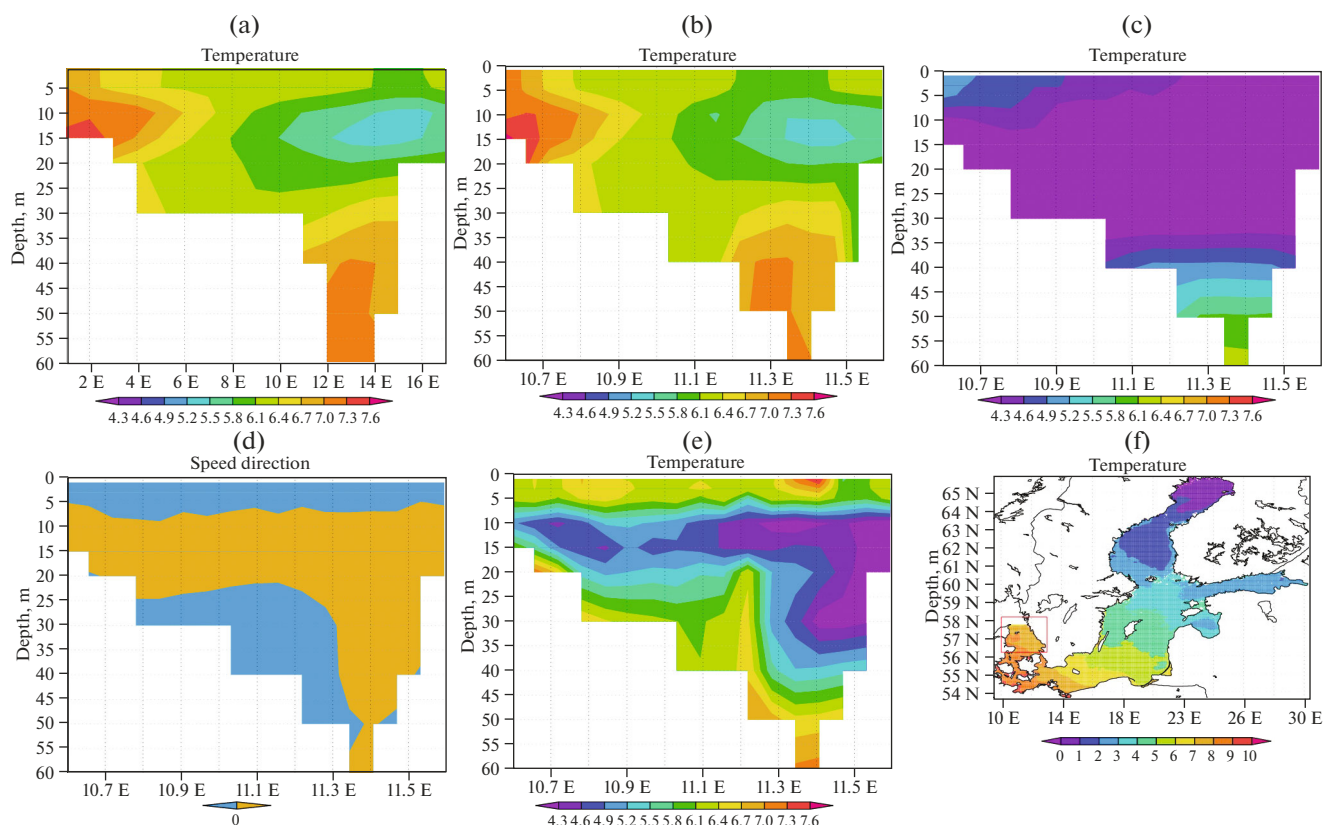i.e., the optimal set of parameters of the iterative process in this problem at $\alpha \cong +0$.

Fig. 1. Temperature profile at the liquid boundary on April 7, 2007: (a) observational data, (b) calculation with the use of assimilation, and (c) calculation without assimilation. (d) Direction of the normal component of the velocity on April 7; (e) temperature profile at the liquid boundary on April 7, 2007, calculation with assimilation on the boundary fragment with the incoming flow [21]; and (f) position of the liquid boundary in model [18].

Let us present the results of numerical experiments as applied to a real sea water area. Algorithm (26)−(28) was incorporated into the model of Baltic Sea hydrothermodynamics [18] and used in calculations on every second time layer of the model at Step 2 of the splitting method described above. Calculations of Baltic Sea hydrothermodynamics for the time interval of April 1−30, 2007, were of the following types: calculation without the use of the assimilation procedure, calculation with the use of assimilation (26)−(28), and calculation with the assimilation procedure described in [21]. The assimilation procedure described in this work differs from the procedure presented in [21] in the notation of the boundary condition. In particular, data assimilation in algorithm (26)−(28) is carried out on the whole liquid boundary; in [21], it is carried out only on the boundary portion where the normal component of the velocity is directed inside the domain. For the observational data, the data of reanalysis by the HIROMB (High-Resolution Operational Model for the Baltic) model, version 3.0, were used. They were presented by the Swedish Meteorological and Hydrological Institute (SMHI). The data were loaded from the website [27].

Figure 1a presents the field of temperature T_{obs} by reanalysis data from the website [27] (section in depth on the open boundary between the North and Baltic Seas; the position of the boundary itself can be seen in Fig. 1f) on April 7, 2007, as observational data; Figs. 1b and 1c present the temperature fields at the time of seven model days since the beginning of the calculation, with the use of the assimilation procedure (26)−(28) and without it, respectively. In Fig. 1d, the blue color marks points where the velocity vector of the water flow is directed inside the water area (from the North Sea to the Baltic Sea); in Fig. 1e, the temperature field at the liquid boundary according to results of the calculation with the use of assimilation from [21]. It is seen from the figures that the developed algorithm (26)−(28) brings the calculation result of the model closer to the data at the open boundary; for the algorithm [21], this is true only on the boundary portion where the normal component of the velocity is directed inside the domain.

For comparison, Fig. 2 presents temperature distributions through the depth on April 27, 2007, at a latitude of 57.4°: reanalysis data according to the HIROMB model (Fig. 2a), results of the calculation without assimilation (Fig. 2b), with the use of assimi-
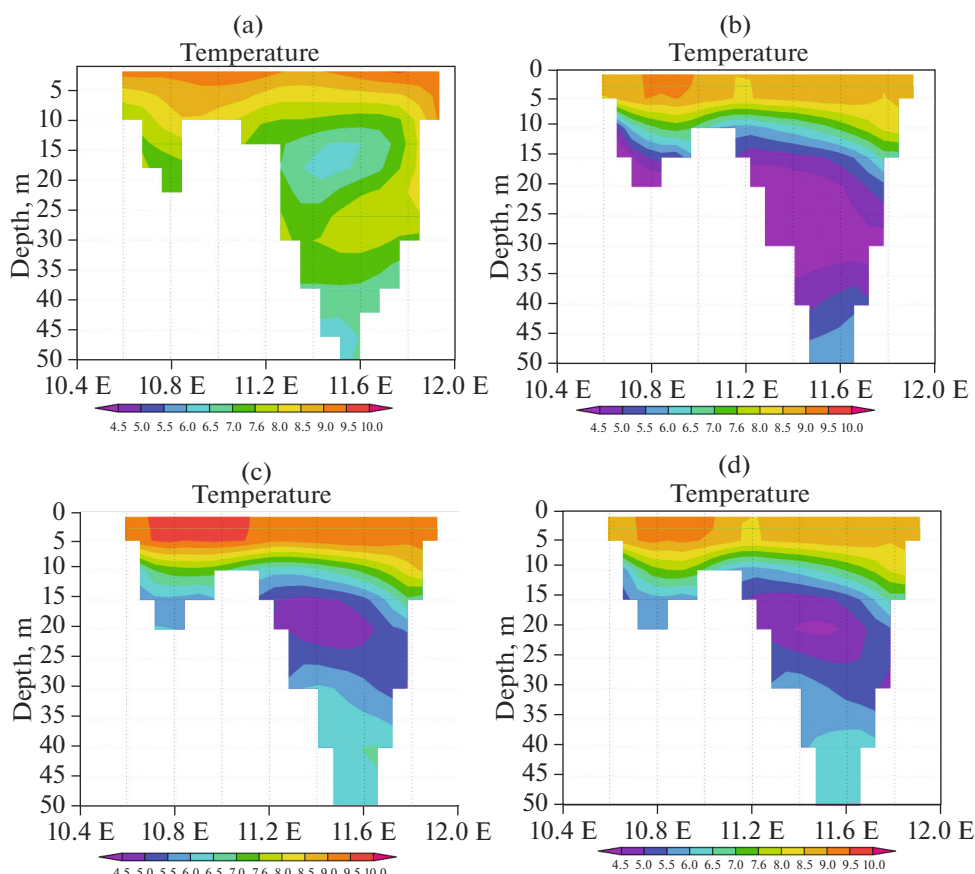
Fig. 2. Temperature profile at the liquid boundary on April 27, 2007, at a latitude of 57.4°: (a) observational (reanalysis) data, (b) calculation without assimilation, (c) calculation with assimilation, and (d) calculation with assimilation on the boundary fragment with the incoming flow [21].

lation (26)−(28) (Fig. 2c), and with the use of assimilation described in [21] (Fig. 2d). As is seen from the figures, the assimilation algorithm brings the calculation results closer to the reanalysis results not only at the liquid boundary where the assimilation is carried out, but also on other vertical sections. Nevertheless, the construction of temperature distributions on sections spaced from the liquid boundary farther than the section presented in Fig. 2 demonstrates that the assimilation algorithm (26)−(28) has almost no effect on the temperature distribution through the depth to the south from 56° N. This might be related to the short calculation interval (one month). Longer term calculations need further investigations into the effect exerted by the considered algorithm on temperature distributions through the depth obtained by simulation results at different latitudes.

Figures 3b−3d present daily average surface temperature distributions according to simulation results for April 14, 2007: Fig. 3b, results of the calculation without assimilation; Fig. 3c, with the use of assimilation (26)−(28); and Fig. 3d, with the use of assimilation from [21]. These distributions can be compared with the result of interpolation of satellite observation data [28, 29] to the grid of the model; this result is presented in Fig. 3a. As one can see, the temperature field in Fig. 3c is closer to that presented in Fig. 3a than the temperature fields in Figs. 3b and 3d. The difference between Figs. 3c and 3d is not so great and is concentrated near the liquid boundary. Note that the performed experiments did not include the assimilation of data on the sea surface temperature; for this reason, simulation results presented in Figs. 3b−3d differ from the observational data from satellites.

In addition, to compare the assimilation algorithms, daily average temperature distributions over the surface on April 18, 2007, according to simulation results for March−April, 2007 are presented: with the use of algorithm (26)−(28) (Fig. 4b) and with the use of the algorithm from [21] (Fig. 4c). Figure 4a shows the daily average temperature distribution obtained from reanalysis according to the NEMO-Nordic model (SMHI) [30, 27]. The grid of the NEMO-Nordic model covers the water areas of the Baltic and North seas; this model also implements the assimilation of the sea surface temperature and of temperature and salinity profiles throughout the depth. Nevertheless, building principles of the NEMO model are quite
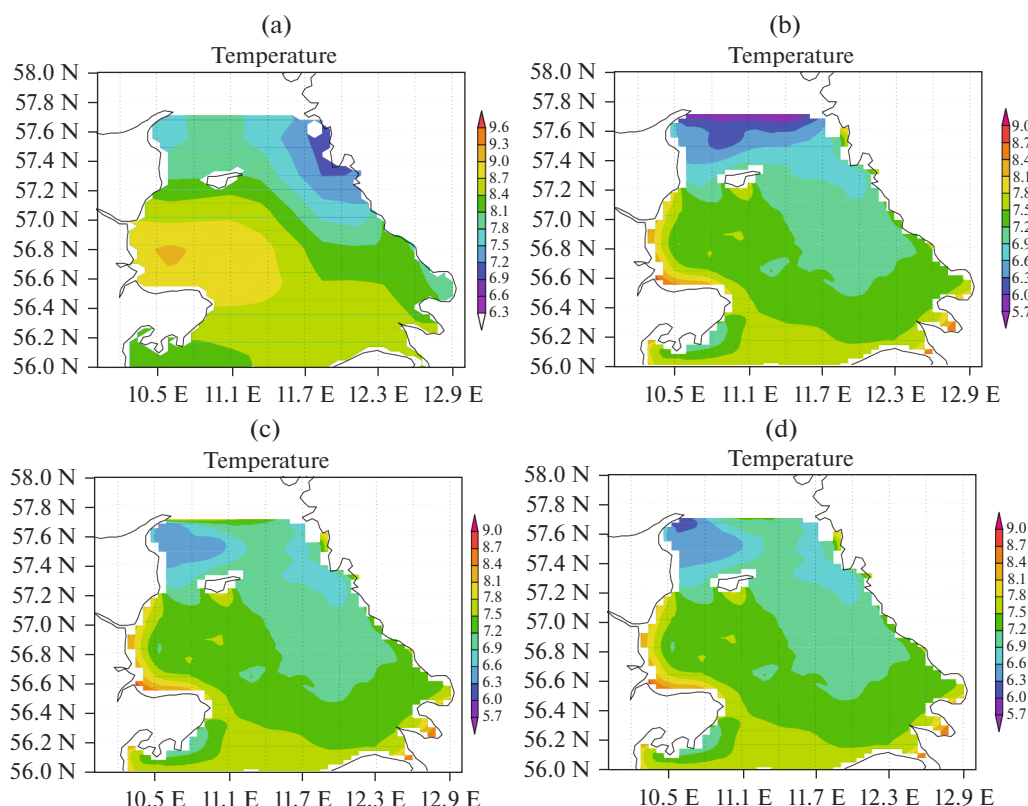
Fig. 3. Sea surface temperature on April 14, 2007: (a) observational data from satellites, (b) calculation without assimilation, (c) calculation with assimilation, and (d) calculation with assimilation on the boundary fragment with the incoming flow [21].
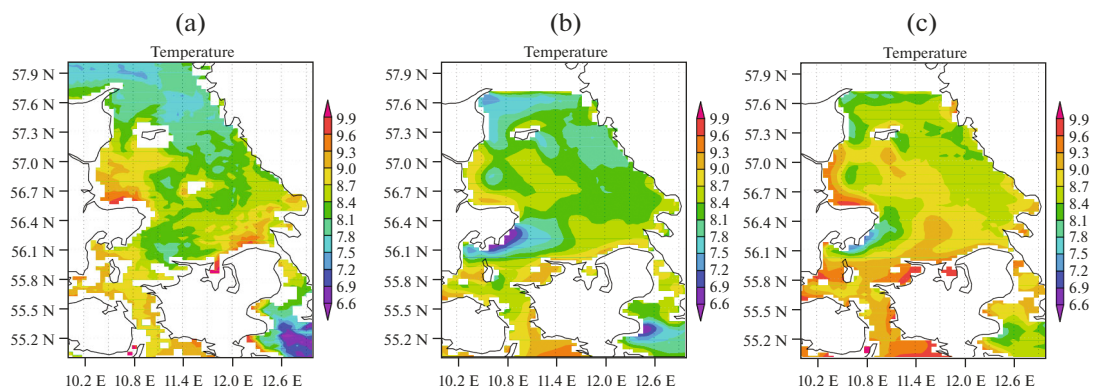

Fig. 4. Sea surface temperature on April 18, 2007: (a) reanalysis by the NEMO-Nordic model, (b) calculation with assimilation, and (c) calculation with assimilation on the boundary fragment with the incoming flow [21].

different; for this reason, the distributions presented in Fig. 4 are not similar.

Based on results of the numerical experiments, one can make the following conclusions. The developed algorithm really successfully brings the calculation result of the model closer to data at the open boundary, which corroborates the theoretical results. In short-term calculations, the assimilation effect manifests itself only in a domain near the open boundary. The algorithm exerts a weak effect on the temperature distributions through the depth, and the longer the calculation with assimilation is, the stronger the effect. Longer calculations need further investigations into the effect of the considered algorithm on depth temperature distributions obtained by simulation results at different latitudes. The effectiveness of using the algo-

rithm depends on the quality of observational data. In this work, reanalysis data from the Swedish model of North and Baltic Sea hydrothermodynamics were used—those data were available for every 6 h and possessed sufficient smoothness. Experiments with other data were not carried out.

3. PROBLEM OF VARIATIONAL ASSIMILATION OF DATA ON THE LEVEL

Let us consider problem (7) from Step 1 of the splitting method. As was already mentioned, solving this problem can be reduced to an independent solution of problems (9) and (10); the level function $\xi \equiv \xi_1^j$ obtained in problem (9) varies neither in problem (10) nor at other steps of the splitting method; i.e., $\xi^j = \xi_5^j = \xi_1^j$. Therefore, the assimilation of data about the level must be carried out directly in problem (9).

Further, let function d_s be an additional unknown (control) which must be determined by the data assimilation procedure. Let also observational data ξ_{obs} on the sea level be available on a certain portion of the boundary $\partial\Omega$ on a time interval (t_{j-1}, t_j). By $\chi_{\xi,\partial\Omega}$, we denote the characteristic function of the set where ξ_{obs} is given. Let us formulate the closure condition:

$$\chi_{\xi,\partial\Omega}\xi = \chi_{\xi,\partial\Omega}\xi_{obs} \quad \text{on} \quad \partial\Omega \times (t_{j-1}, t_j). \quad (30)$$

Writing the implicit approximation scheme of problem (9) on the interval (t_{j-1}, t_j), $\tau = t_j - t_{j-1}$, in time, we obtain the semidiscrete problem, which is an approximation to (9):

$$\begin{cases} \dfrac{\underline{U}}{\tau} + \begin{bmatrix} 0 & -\ell \\ \ell & 0 \end{bmatrix}\underline{U} - g \cdot \mathbf{grad}\xi = \tilde{\tilde{f}} \quad \text{in} \quad \Omega, \\[2mm] \dfrac{\xi}{\tau} - \mathbf{div}\left(H\underline{U}\right) = (\bar{F})_3 \quad \text{in} \quad \Omega, \\[2mm] H\underline{U} \cdot \vec{n} + m_{w,\text{op}}\sqrt{gH}\xi = m_{w,\text{op}}\sqrt{gH}d_s \quad \text{on} \quad \partial\Omega, \end{cases} \quad (31)$$

where $\bar{F} = (f_1^j + u^{j-1}/\tau,\ f_2^j + v^{j-1}/\tau,\ \xi^j/\tau)^T$, $\tilde{\tilde{f}} = ((\bar{F})_1, (\bar{F})_2))^T \equiv (\tilde{f}_1, \tilde{f}_2)$. Thus, the inverse problem is formulated as follows: find $\underline{U}$, ξ, and d_s such that (31) and (30) are satisfied.

Let us reformulate the problem as an optimal control problem. For this purpose, we introduce the functional

$$J_\alpha = J_\alpha(d_s, \xi(d_s)) = \frac{\alpha}{2}\int_{\partial\Omega} m_{w,\text{op}}\sqrt{gH}(d_s - d_s^{(0)})^2 d\Gamma$$
$$+ \frac{1}{2}\int_{\partial\Omega} \chi_{\xi,\partial\Omega}\sqrt{gH}(\xi - \xi_{obs})^2 d\Gamma, \quad (32)$$

where $d_s^{(0)}$ is a given function. The optimal control problem is formulated as follows: find the function d_s minimizing the functional J_α such that (31) is satisfied.

Inverse problem (31), (30) was studied theoretically in [14]. For the theoretical investigation, system (31) was reduced to the elliptical third boundary value problem for ξ; then the generalized formulation of the problem was introduced and conditions for unique and dense solvability were obtained (in particular, if the observational data are given at a liquid boundary, the problem is uniquely and densely solvable). The problem of minimizing functional J_α can be solved, e.g., using the gradient descent method, which will have the form of the following iterative process (for a system reduced to a semidiscrete form):

$$\begin{cases} \dfrac{\underline{U}_k}{\tau} + \begin{bmatrix} 0 & -\ell \\ \ell & 0 \end{bmatrix}\underline{U}_k - g \cdot \mathbf{grad}\xi_k = \tilde{\tilde{f}} \quad \text{in} \quad \Omega, \\[2mm] \dfrac{\xi_k}{\tau} - \mathbf{div}\left(H\underline{U}_k\right) = (\bar{F})_3 \quad \text{in} \quad \Omega, \\[2mm] H\underline{U}_k \cdot \vec{n} + m_{w,\text{op}}\sqrt{gH}\xi_k = m_{w,\text{op}}\sqrt{gH}(d_s)_k \quad \text{on} \quad \partial\Omega, \end{cases} \quad (33)$$

$$\begin{cases} \dfrac{\underline{U}_k^*}{\tau} - \begin{bmatrix} 0 & -\ell \\ \ell & 0 \end{bmatrix}\underline{U}_k^* + g \cdot \mathbf{grad}\xi_k^* = 0 \quad \text{in} \quad \Omega, \\[2mm] \dfrac{\xi_k^*}{\tau} + \mathbf{div}\left(H\underline{U}_k^*\right) = 0 \quad \text{in} \quad \Omega, \\[2mm] -H\underline{U}_k^* \cdot \vec{n} + m_{w,\text{op}}\sqrt{gH}\xi_k^* \\[1mm] = \chi_{\xi,\partial\Omega}\sqrt{gH}(\xi_k - \xi_{obs}) \quad \text{on} \quad \partial\Omega, \end{cases} \quad (34)$$

$$(d_s)_{k+1} = (d_s)_k - \tau_k m_{w,\text{op}}\left[\alpha((d_s)_k - d_s^{(0)}) + \xi_k^*\right], \quad (35)$$

where k is the iteration number. In view of dense solvability of the problem in the case when observation data are given on a liquid boundary, at a sufficiently small $\tau_k = \tau = \text{const} > 0$, we have [24]

$$\int_{\partial\Omega} \chi_{\xi,\partial\Omega}\sqrt{gH}(\xi_k(\alpha) - \xi_{obs})^2 d\Gamma \to 0$$
$$\text{as} \quad \alpha \to 0, \quad k \to \infty.$$

In addition, the parameter of the iterative process in this case can be calculated using the following formula (see [24]):

$$\tau_k = \frac{J_\alpha}{\|J_\alpha'\|^2} = \frac{1/2\int_{\partial\Omega} \chi_{\xi,\partial\Omega}\sqrt{gH}(\xi - \xi_{obs})^2 d\Gamma}{\int_{\partial\Omega} m_{w,\text{op}}\sqrt{gH}(\xi^*)^2 d\Gamma}.$$

Particular attention should be given to the question about the availability of real observational data. As was shown in [14], if the observational data are given on the whole liquid boundary, the problem is uniquely and densely solvable and the iterative process converges; the functions $\underline{U}_k(\alpha)$, $\xi_k(\alpha)$, and $(d_s)_k(\alpha)$ at a sufficiently small α and large k can be taken as approximate solutions of inverse problem (31), (30). In practice, however, only the following information about the level is available: satellite altimetry data and data of
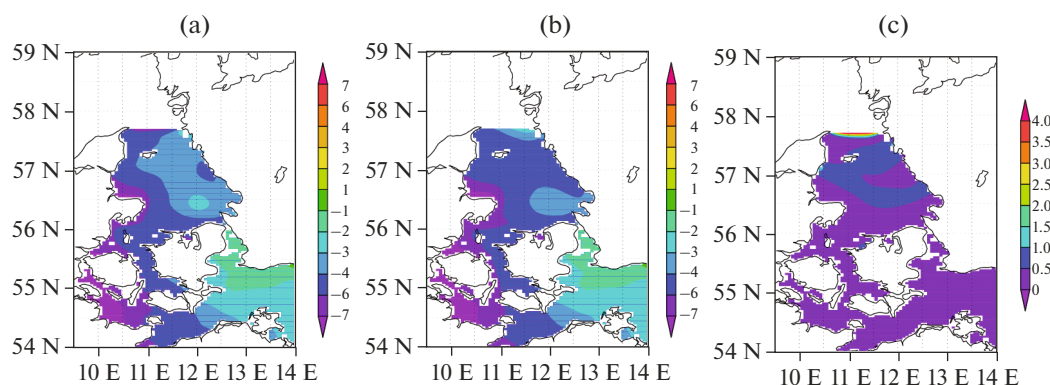
Fig. 5. Sea level (cm) averaged over the whole calculation period from March 1, to March 14, 2017: (a) calculation with the use of assimilation, (b) calculation without assimilation, and (c) absolute value of the difference between them.

observations at level measuring stations. This work involves the DUACS DT2014 (Data Unification and Altimeter Combination System, Delayed Time mode) satellite altimetry information from about level anomalies and dynamic topography [31]. In the whole array of satellite data, only those that are measurements at the liquid boundary are suitable for the variational assimilation algorithm (33)−(35). According to analysis of the data, there are critically few such measurements (one or two per day). Among other things, tracks of satellites never go directly along the liquid boundary, but only cross it at a certain point. Therefore, there are one or two measurements of the level per day at some definite points at the liquid boundary. Coordinates of these points depend on time. As for observational data from level measuring stations, they are obtained at fixed points near the shore (at measuring coastal stations) with a definite frequency known beforehand.

Iterative algorithm (33)−(35) was implemented as a plugin of the Baltic Sea hydrothermodynamics model. Observational data are assimilated at the time step at which these data are available; at other steps, assimilation is absent and the computation occurs in the usual regime. The work involved data from Skagen (Denmark) and Göteborg-Torshamnen (SMHI, Sweden) level measuring stations and DUACS DT2014 satellite altimetry data. There are two sets of satellite altimetry data on open access [27]: processed data by satellite tracks (L3) and a result of using the optimal interpolation method for constructing the level on the grid (L4) [31]. For the assimilation, L3 data were used. Data at grid nodes of the model at the liquid boundary in this work were obtained using usual quadratic interpolation. One can univocally assert that available points with observational data are insufficient to obtain an approximation close to the real level to any extent. Moreover, the method presupposes that the inverse problem about the retrieval of the function d_s is solved using observational data on the interval $\left(t_{j_k}, t_{j_k+1}\right)$, where k is the number of the time step at which the

observational data are available, and the further calculation at $t \in \left(t_{j_k+1}, t_{j_k+1}\right)$ is performed with the found (retrieved) d_s. In the case with rare data, this approach is not quite correct, because d_s is a time-varying function and it is an error to think that it is constant on large time intervals. Moreover, one should take into account that function d_s is retrieved not ideally but with errors. Therefore, it is expedient to set d_s to zero on time intervals where the data are absent. However, this is also incorrect, because does not describe the liquid boundary from the physical point of view. In this work, the boundary condition from (31) with d_s is used for several time steps after assimilation; then, the calculation is carried out with the boundary condition that was used in the model before the inclusion of the assimilation block.

Thus, let us consider results of the numerical experiments. The calculation was carried out for 14 days, from March 1 to March 14, 2017. In data from the satellites, there were a total of 18 time intervals with measurements of the sea level near the liquid boundary during the whole calculation period; therefore, assimilation was performed 18 times. For comparison, Fig. 5 presents sea levels averaged over the whole calculation period and obtained by the model with the use of assimilation and without it, as well as the difference between them. As is seen from the figures, the inclusion of the assimilation block had an effect on the average result near the liquid boundary, but it remained almost unchanged farther the islands.

One advantage of algorithm (33)−(35) is rapid convergence—the acceptable accuracy is reached in 8−15 iterations; after 70 iterations, the residual drops to the minimum possible value. However, in view of rough interpolation of observational data, such accuracy is not required in the problems under consideration; for this reason, the number of iterations in the numerical experiments was restricted to eight. According to the experiments, the iterative method converges at the first 8−15 iterations with the rate of

a geometrical progression, which corresponds to the theory in [24].

Note that the method admits the introduction of different weight coefficients taking into account the significance of certain observational data into the functional. Owing to such coefficients, one can reduce the negative influence of the rough interpolation on the result. Note also that the question about the efficiency of using the algorithm proposed in this section with real observational data requires further investigations. Nevertheless, for the algorithm's observational data, one can also use data of calculations by models of large water areas, as was done in the previous section; however, the efficiency of using the algorithm in this capacity has not yet been studied (in this case, it is necessary to compare the algorithm with known methods of nested grids). Note that an increase in the quality of modeling and forecasting sea currents, along with observational data assimilation, depends on parameterization of subgrid processes. In doing this, special attention should be paid to the parameterization of turbulent exchange processes [32].

REFERENCES

1. A. I. Kubryakov, "The use of overset grid technology for the construction of a system for hydrophysical field moitoring in the coastal areas of the Black Sea, in *Ecological Safety of Coastal and Shelf Zones and Integrated Exploitation of the Shelf Resources* (EKOSI-Gidrofizika, Sevastopol, 2004), no. 11, pp. 31−50.

2. I. A. Chernov and A. V. Tolstikov, "Numerical modeling of large-scale dynamics of the White Sea," Trudy Karel. Nauch. Tsentra RAN, No. 4, 137−142 (2014).

3. S. A. Myslenkov, "The use of satellite altimetry for the numerical simulation of water transfer in the Northern Atlantic Ocean," Tr. GU Gidromettsentr Rossii **345,** 119−125 (2011).

4. I. Orlanski, "A simple boundary condition for unbounded hyperbolic flows," J. Comput. Phys. **21** (3), 251−269 (1976).

5. P. Marchesiello, J. C. McWilliams, and A. Shchepetkin, "Open boundary conditions for long-term integration of regional oceanic models," Ocean Model. **3,** 1−20 (2001).

6. C. A. Edwards, A. M. Moore, I. Hoteit, and B. D. Cornuelle, "Regional ocean data assimilation," Annu. Rev. Mar. Sci. **7,** 6:1−6:22 (2015).

7. G. I. Marchuk and V. B. Zalesny, "Modeling of the world ocean circulation with the four-dimensional assimilation of temperature and salinity fields," Izv., Atmos. Ocean. Phys. **48** (1), 15−29 (2012).

8. V. B. Zalesny, V. I. Agoshkov, V. P. Shutyaev, F. Le Dimet, and B. O. Ivchenko, "Numerical modeling of ocean hydrodynamics with variational assimilation of observational data," Izv., Atmos. Ocean. Phys. **52,** 431−442 (2016).

9. H. Ngodock and M. Carrier, "A weak constraint 4D-Var assimilation system for the navy coastal ocean model using the representer method," in *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications* (Springer, Berlin, Heidelberg, 2013), Vol. II.

10. V. I. Agoshkov, V. P. Shutyaev, E. I. Parmuzin, N. B. Zakharova, T. O. Sheloput, and N. R. Lezina, "Variational Data Assimilation in the Mathematical Model of the Black Sea Dynamics," Physical Oceanography. **26** (6), 515−527 (2019).

11. H. S. Tang, K. Qu, and X. G. Wu, "An overset grid method for integration of fully 3D fluid dynamics and geophysics fluid dynamics models to simulate multiphysics coastal ocean flows," J. Comput. Phys. **273,** 548−571 (2014).

12. A. F. Bennett and P. C. McIntosh, "Open ocean modeling as an inverse problem: tidal theory," J. Phys. Oceanogr. **12** (10), 1004−1018 (1982).

13. V. I. Agoshkov, "Application of mathematical methods for solving the problem of liquid boundary conditions in hydrodynamics," Z. Angew. Math. Mech. **76** (S1), 337−338 (1995)

14. V. I. Agoshkov, "Inverse problems of the mathematical theory of tides: boundary-function problem," Russ. J. Numer. Anal. Math. Model. **20** (1), 1−18 (2005).

15. V. I. Agoshkov, "Statement and study of some inverse problems in modelling of hydrophysical fields for water areas with 'liquid' boundaries," Russ. J. Numer. Anal. Math. Model. **32** (2), 73−90 (2017).

16. V. I. Agoshkov, *Methods of Solving Inverse Problems and the Problems of Variational Assimilation of Observation Data In Studies of Large-Scale Dynamics of Oceans and Seas* (IVM RAN, Moscow, 2016) [in Russian].

17. V. I. Agoshkov, N. R. Lezina, and T. O. Sheloput, "Domain decomposition method for the variational assimilation of the sea level in a model of open water areas hydrodynamics," J. Mar. Sci. Eng. **7** (6), 195 (2019).

18. V. B. Zalesny, A. V. Gusev, V. O. Ivchenko, R. Tamsalu, and R. Aps, "Numerical model of the Baltic Sea circulation," Russ. J. Numer. Anal. Math. Model. **28** (1), 85−100 (2013).

19. V. B. Zalesnyi, A. V. Gusev, and V. I. Agoshkov, "Modeling Black Sea circulation with high resolution in the coastal zone," Izv., Atmos. Ocean. Phys. **52,** 277−293 (2016).

20. E. V. Dement'eva, E. D. Karepova, and V. V. Shaidurov, "Recovery of a boundary function from observation data for the surface wave propagation problem in an open basin," Sib. Zh. Industr. Matem. **16** (1), 10−20 (2013).

21. V. I. Agoshkov and T. O. Sheloput, "The study and numerical solution of some inverse problems in simulation of hydrophysical fields in water areas with 'liquid' boundaries," Russ. J. Numer. Anal. Math. Model. **32** (3), 147−164 (2017).

22. G. I. Marchuk, V. P. Dymnikov, and V. B. Zalesnyi, *Mathematical Models in Geophysical Fluid Dynamics and Numerical Methods of Their Implementation* (Gidrometeo-izdat, Leningrad, 1987) [in Russian].

23. V. I. Agoshkov and M. V. Assovskii, *Mathematical Modelling of the Dynamics of the World's Oceans with Tide-Forming Forces Taken Into Account* (IVM RAN, Moscow, 2016) [in Russian].

24. V. I. Agoshkov, *Optimal Control Methods and Adjoint Equations in Mathematical Physics Problems* (IVM RAN, Moscow, 2016) [in Russian].

25. A. N. Tikhonov and V. Ya. Arsenin, *Methods of Solution of Ill-Posed Problems* (Nauka, Moscow, 1986) [in Russian].

26. V. I. Agoshkov and T. O. Sheloput, "The study and numerical solution of the problem of heat and salinity transfer assuming 'liquid' boundaries," Russ. J. Numer. Anal. Math. Model. **31** (2), 71−80 (2016).

27. Copernicus. Marine Environment Monitoring Service. http://marine.copernicus.eu/

28. J. L. Hoyer and J. She, "Optimal interpolation of sea surface temperature for the North Sea and Baltic Sea," J. Mar. Syst. **65** (1−4), 176−189 (2007).

29. J. L. Høyer, P. Le Borgne, and S. Eastwood, "A bias correction method for Arctic satellite sea surface temperature observations," Remote Sens. Environ. **146,** 201−213 (2014).

30. R. Hordoir et al., "Influence of sea level rise on the dynamics of salt inflows in the Baltic Sea," J. Geophys. Res.: Oceans **120** (10), 6653−6668 (2015).

31. M. I. Pujol et al., "DUACS DT2014: the new multimission altimeter data set reprocessed over 20 years," Ocean Sci. **12,** 1067−1090 (2016).

32. V. B. Zalesnyi, S. N. Moshonkin, V. L. Perov, and A. V. Gusev, "Ocean Circulation Modeling with K-Omega and K-Epsilon Parameterizations of Vertical Turbulent Exchange," Physical Oceanography. **26** (6), 455−466 (2019).

Translated by A. Nikol'skii

Study of the Ocean Boundary Layer Convection under Inhomogeneous Ice with the Help of the Large-Eddy Simulation Model

A. V. Glazunov[a, b, *] and N. G. Iakovlev[a, **]

[a]*Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, 119333 Russia*
[b]*Moscow State University, Moscow, 119991 Russia*
**e-mail: and.glas@gmail.com*
***e-mail: nick_yakovlev@mail.ru*

Abstract—Upper ocean convection under narrow ice leads is considered and simulated. The effects associated with the localization of the buoyancy source and with the influence of the Coriolis force are discovered. An explanation is proposed for a mechanism that forms a stably salt-stratified isothermal layer during cold seasons at high latitudes. Observational data are qualitatively consistent with the simulation results. The existing parameterizations of under-ice convection in modern climate models are discussed and their possible defects are indicated.

Keywords: upper layer of the ocean, under-ice convection, Arctic, Antarctic, ice cracks and leads, large-eddy simulation, ocean turbulence

DOI: 10.1134/S000143382003007X

INTRODUCTION

One important feature of sea ice is that it moves, resulting in the constant appearance of open water due to the divergence and shear of the drift speed field. Although the sizes of an individual lead are relatively small, their effect on the heat and mass exchange between the atmosphere and ocean is of great importance. According to observational data, the absolute values of the sea-ice drift velocities and deformation rates in the Arctic Ocean (AO) have a positive trend in time [1]; therefore, the role of open water is expected to be increasing in a changing climate. Open-water areas are usually long cracks or leads 50 to 1000 m in width and 1 to 50 km in length [2]. Current satellite data [3] show that narrow cracks are more frequent and contribute considerably to the heat and mass fluxes.

The open-water areas produce a strong inhomogeneity in the ocean boundary layer under the ice and change the convection processes during winter cooling and ice formation, thus favoring the formation of a specific hydrostatically stable structure of temperature and stability. The December monthly mean climatological data of Polar Science Center Hydrographic Climatology version 3 (PHC 3.0) [4] clearly show such a structure (Fig. 1a) with an isothermal layer in the upper 50 m and a weakly stable salt stratification. Deep-water areas of the central Arctic were considered in order to minimize the direct effect of river inflow. The temperature below 50 m does not always increase

monotonically, which may be related to the properties of ocean dynamics. The PHC 3.0 data (Fig. 1a) show a distinct isothermal layer down to a depth of 50 m, whereas World Ocean Atlas 2013 (WOA13) data [5] (Fig. 1b) demonstrate a more complex pattern with a high temperature variance in the upper layer of the ocean and sometimes the salinity-uniform upper layer. The distinction between the two atlases may be due to different datasets of measurements and to the different methods used to process them. It is worth mentioning measurements from ice-tethered profilers (ITPs; https://www.whoi.edu/website/itp/overview). These measurements show that the halocline in the upper ocean is usually encountered in the central Arctic in areas covered with thick and rather slow pack ice (see, e.g., ITP9 and ITP16). There is a high temperature variance in the upper approximately 50-m layer of the ocean (average temperature being close to isothermal), with a relatively low variance of salinity, in agreement with WOA13 [5].

It is argued that the persistent halocline drives a stable density stratification of the upper ocean and plays a major role in blocking the heat flux toward a cold ocean surface and, hence, accounts for the stability of ice cover; apart from the hydrologic and climatic implications, the structure of the halocline is crucial to the functioning of the AO biochemical system [6].

It is a common belief that a pattern like the upper-ocean halocline does not occur in the Southern Ocean
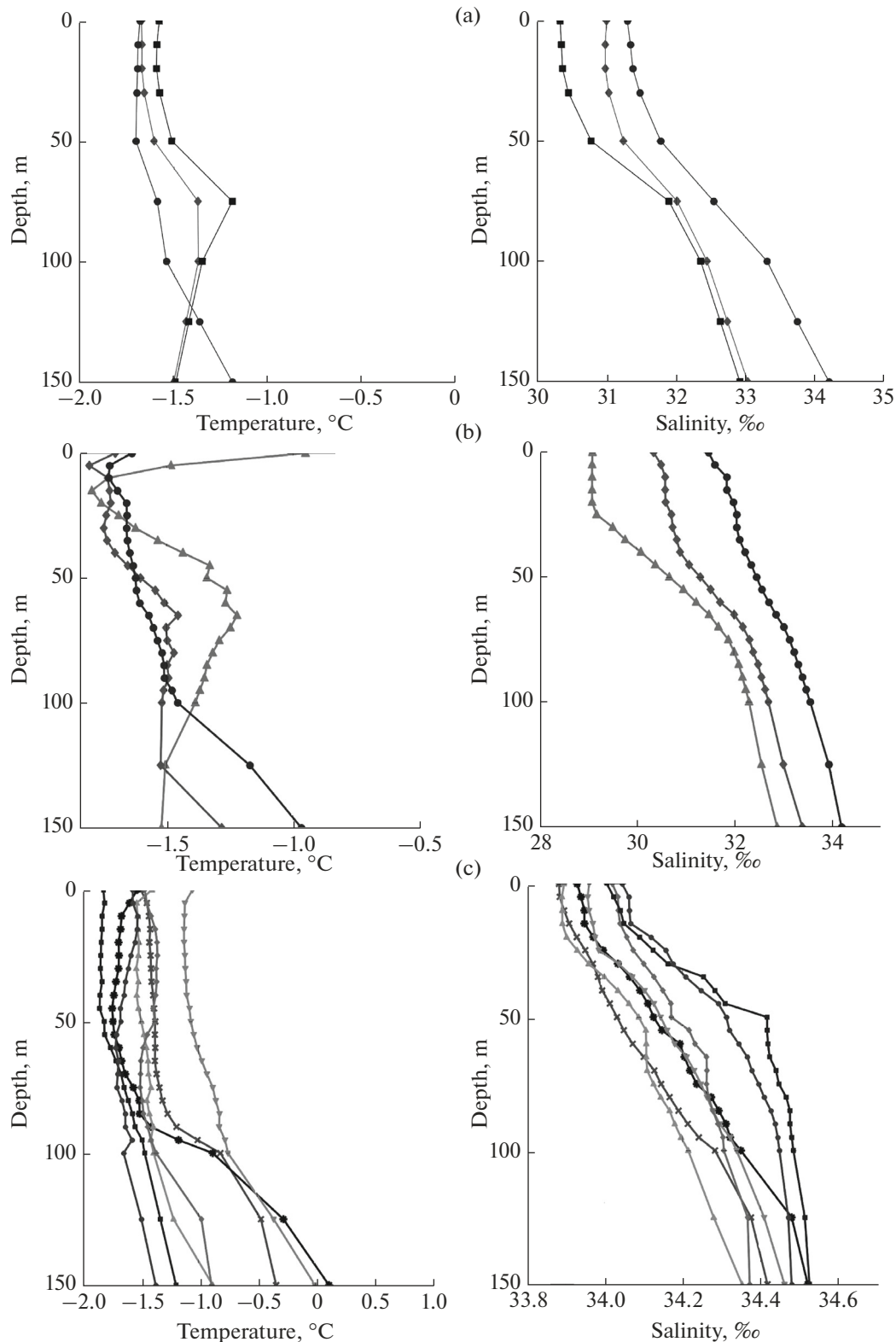
Fig. 1. Measured temperature and salinity profiles in the upper 150 m (climatological monthly means): (a) from PHC3.0 [4] in the Arctic Ocean in December (at the North Pole (black circles); 82.5 N, 145.5 W (gray diamonds); and in the Beaufort Sea 75.5 N, 145.5 W (boxes)), (b) the same profiles as in (a) from WOA13 [5] (at the North Pole (black circles); 82.5 N, 145.5 W (gray diamonds); and in the Beaufort Sea 75.5 N, 145.5 W (gray triangles)), and (c) off the Antarctic coast at the outer margin of the continental slope in June from WOA13 [5] (70.5 S, 52.5 W (gray boxes); 73.5 S, 42.5 W (black circles); 66.5 S, 23.5 E (black flakes); 64.5 S, 82.5 E (crosses); 63.5 S, 111.5 E (upside down triangles); 71.5 S, 118.5 W (triangles with the vertex upward); and 74.5 S, 169.5 W (diamonds).

off the coast of the Antarctic because there is no intense river inflow in the region, but, according to WOA13 data (Fig. 1c), the under-ice halocline (deeper with roughly half as large vertical salinity gradients as those in the AO) is observed there during cold months, so one can suggest a common physical mechanism that drives convection under the inhomogeneous ice.

Penetrative convection induced by the sinking of an instantaneously formed isolated plume (the anomaly having a geometric shape and approximately constant density that is higher than ambient) was modeled in laboratory conditions [7]. Scaling analysis and experimental measurements revealed two convection regimes: stratification at large values of N/f (here N is the Brunt–Väisälä frequency and f the inertial frequency) and rotation when this parameter was small. The transition between the stratification and rotation regimes occurs at $N/f \approx 0.6$ and is independent of the initial plume buoyancy. This condition takes place when $N < 10^{-4}$ s^{-1}. It is therefore concluded [7] that considering the rotation-controlled regime makes no sense, because $N \sim 10^{-2}$ s^{-1} in the Arctic Ocean. In fact, this estimate is true only for summer conditions; in winter the upper layer is well-mixed and the Brunt–Väisälä frequency may be very small.

In [8], a study was undertaken to investigate convection generated by a continuous line shaped of buoyancy in a rotating tank. Such a laboratory design is a close analog of under-ice lead convection. Although the experimental parameters matched the stratification regime, there were lenticular vortices that were explained as emerging from the instability at average flow shears developing under the influence of the Coriolis forces for a time well above the time required for fluid particles to sink to depths of neutral buoyancy. From the above, the argument that rotation is unimportant for convective plumes beneath narrow ice leads calls for an additional verification by direct simulation.

The papers cited above discuss large-scale buoyancy anomalies, and measuring methods are limited by relatively small Reynolds numbers achievable in laboratory conditions. Unlike convection of this type, upper ocean convection under leads is a continuous random process in which the buoyancy anomalies are generated and distributed in spatial scales not only by a local surface source, but also by an average buoyancy field, and they are transformed when carried by a turbulent current. Turbulent temperature and salinity fluctuations have a broad continuous variance spectrum and do not retain their shape when transported. Such anomalies can only conventionally be called plumes, and their sizes and times should be understood as the corresponding turbulent length scales and average statistical times of descent of the fluid particles in the convection zone from the surface to the base of the upper ocean layer. Nonetheless, given a long-standing tradition, we will further use the term "plume," implying any observed positive density anomaly at a fixed time and not distinguishing the organized patterns from turbulent fluctuations.

Field observations with autonomous underwater vehicles [9] show that a large role is played by a relative drift speed. When the difference between the water and ice velocities is greater than 10 cm/s, the convective plume that forms beneath the open-water area desintegrates rapidly, with the excessive salt distributed uniformly throughout the entire upper ocean layer.

Because conducting field and laboratory experiments is difficult, numerical modeling remains an important tool for research. Use is made of numerical models of different complexity, from fairly complicated nonhydrostatic [10] to hydrostatic ones [11], in which the convection process is interpreted as the vertical turbulent heat and salt diffusion, with coefficients calculated from the Mellor–Yamada model [12] with second-order closure (see [13]).

Realistic three-dimensional nonhydrostatic large-eddy simulation (LES) models were used to reproduce turbulence in the upper ocean beneath sea-ice leads [14]. These simulations found that ice drift enhances turbulent entrainment at the base of the upper ocean when there is convection under the continuous ice cover, as well as convection beneath the lead. This appears as an increase in the turbulent salt fluxes at the bottom of the mixed layer (see Figs. 5 and 11 from [14]). The enhancement of entrainment and the increase in the growth rate of the mixed-layer thickness are caused by the added shear generation of turbulent kinetic energy (TKE). This process is typical of the convective boundary layer and can be reproduced by standard locally one-dimensional upper-ocean models that contain the TKE equation. This conclusion agrees with field experiments [2].

The under-ice salt convection in large-scale models for the Southern Ocean was described using parameterization [15, 16], where all the rejected salt was distributed uniformly in a surface layer, the depth of which was chosen empirically. Another parameterization scheme of penetrative convection [17] and its modification with the open-water fraction [13] were proposed for the Arctic Ocean, which has its own features. In these schemes, the salt mass ∂S rejected during ice formation is distributed vertically in the upper mixed layer not uniformly, but according to the empirical power law as a function of depth.

Parameterizations of convective mixing in the Weddel Sea were tested in [18] based on turbulent diffusion, as well as on convective adjustment and penetrative convection models. Turbulent diffusion combined with penetrative convection gave a remarkable result [19]. The inhomogeneity of ice cover, however, was disregarded there, and the results were true most likely for convection in open water.

The parameterization used in the World Ocean model [20] is based on a combination of various ideas about the penetrative convection in the atmosphere and ocean. It is shown that the parameterization of the

penetrative salt plumes under the ice improves the simulation of the state of the ocean and reduces the error of the surface salinity in the AO.

In our paper, a three-dimensional nonhydrostatic LES model is used to investigate convection in the inhomogeneous boundary layer under sea ice (either drifting or motionless relative to the ocean). It will be shown that under-ice convection beneath leads in extreme ambient conditions may be much different from convection in a horizontally uniform upper ocean. Such convection through leads may result in salinity and temperature profiles uncharacteristic of the convective upper ocean. Particular attention will be given to checking whether or not the Earth's rotation influences the generation of convection beneath narrow leads in the ice.

NUMERICAL MODEL

The model is written in the Boussinesq approximation and designed to reproduce turbulence dynamics of a three-dimensional incompressible fluid. The system of equations contains evolution equations for three components of the filtered flow velocity $\overline{u}_i$, $i = 1,3$ and filtered temperature $\overline{T}$ and salinity $\overline{s}$. Here $\overline{a} \equiv F_{\overline{\Delta}}(a)$ is a given spatial filter commuting with differentiation operators. The momentum balance equations in tensor form are

$$
\frac{\partial \overline{u}_i}{\partial t} + \frac{\partial \overline{u}_i \overline{u}_j}{\partial x_j} = -\frac{\partial \tau_{ij}}{\partial x_j} + \nu \frac{\partial^2 \overline{u}_i}{\partial x_j^2}
$$
$$
- \frac{\partial \overline{p}}{\partial x_i} - g\delta_{i3}\frac{\rho'}{\rho_0} + \overline{F_i^e}, \quad \frac{\partial \overline{u}_i}{\partial x_i} = 0, \tag{1}
$$

where ρ_0 is the average density, $\rho' = \overline{\rho} - \rho_0$ denotes density fluctuations in the model-resolvable range of the spatial scales, F_i^e represents the external forcing of flow and the Coriolis force (the horizontal and vertical components of the angular velocity of Earth' rotation are both taken into account; see [21]), $\overline{p}$ is the normalized pressure, and $\tau_{ij} = \overline{u_i u_j} - \overline{u}_i \overline{u}_j$ is the subgrid/subfilter stress tensor to be parameterized. Because of the large Reynolds numbers typical of under-ice turbulent convection, the term with kinematic viscosity ν can be neglected.

The transfer equations for the filtered scalars $\overline{a}$ (temperature $\overline{T}$ and salinity $\overline{s}$) are written as

$$
\frac{\partial \overline{a}}{\partial t} + \overline{u}_i \frac{\partial \overline{a}}{\partial x_i} = -\frac{\partial \vartheta_i^s}{\partial x_i} + \mu \frac{\partial^2 \overline{a}}{\partial x_i^2} + \overline{Q_a}, \tag{2}
$$

where $\overline{Q_a}$ represents volume sources; $\vartheta_i^s = \overline{u_i a} - \overline{u}_i \overline{a}$ denotes parameterized subgrid-scale fluxes; and μ is the coefficient of molecular diffusion and of molecular heat conductivity, which is also negligible compared to the convective heat and salt transfer.

The calculation of the buoyancy forces involves the nonlinear equation of state [22]. This equation is directly used to calculate $\overline{\rho} = \overline{\rho}(\overline{T}, \overline{s})$. The pressure dependence of density, because of the small penetration depths of convection, and the dependence of the filtered density $\overline{\rho}$ on subgrid-scale temperature and salinity fluctuations, which results from the nonlinearity of the equation of state, are neglected.

It is assumed that temperature $\overline{T}$ does not reach values less than the seawater freezing point $T_f(s, P)$ [22] with the current modeled salinity $\overline{s}(x, y, z)$ and average pressure $P(z)$ at the corresponding depth. Seawater cooling below this limit is compensated by the heat released during formation of salt-free shuga. It is supposed that all the shuga that has formed rises to the surface over a period of time that is significantly less than characteristic times of the dynamics of the explicitly modeled plumes and does not interact with the environment thermally or dynamically. Freshwater loss is taken into account as an intravolume salt source $\overline{Q_s}$. The change in the total volume of the two-phase environment and the liquid phase is neglected, as is required by the formulation of the problem in the Boussinesq approximation. Under conditions typical of lead ice formation in the Arctic and Antarctic, these assumptions are acceptable, because freezing and melting occur mainly near the surface and are subgrid-scaled relative to turbulent convection. Under these conditions, salt rejection is the primary near-surface source of buoyancy.

Tensor τ_{ij} is calculated using the mixed model [23]:

$$
\tau_{ij}^{\text{mix}} = \tau_{ij}^{\text{smag}} + \tau_{ij}^{ssm}
$$
$$
= -2(C_s\overline{\Delta})^2|\overline{S}|\overline{S}_{ij} + \left(\overline{\overline{u}_i\,\overline{u}_j} - \overline{\overline{u_i u_j}}\right), \tag{3}
$$

where $\overline{S}_{ij}$ is the filtered tensor of deformation rates and C_s is the spatially and temporally variable dimensionless parameter, which depends on the local flow characteristics and is determined dynamically [24]. The model of turbulent diffusion is used as closure for the scalars:

$$
\vartheta_i^s = -K_h^{\text{subgrid}}\frac{\partial \overline{s}}{\partial x_i}, \tag{4}
$$

where the coefficient $K_h^{\text{subgrid}} = \left(1/Sc^{\text{subgrid}}\right)(C_s\overline{\Delta})^2|\overline{S}|$ is proportional to the eddy viscosity. The turbulent subgrid-scale Schmidt number Sc^{subgrid} has a constant value $Sc^{\text{subgrid}} = 0.8$.

With no eddy activity in the resolvable range of scales, the subgrid-scale turbulent viscosity, diffusion, and heat conductivity are significantly reduced owing to a decrease in coefficient C_s. This property of the model increases its effective resolution and prevents large salinity gradients at the halocline boundary from blurring in stable stratification.

Details of dynamic closure and specifics of the LES implementation are described in [25–27].

NUMERICAL EXPERIMENTS AND RESULTS

3.1. Setup of Numerical Experiments

The initial state of the upper ocean with a depth of about 70 m was specified to be nearly homogeneous (with a freezing-point temperature and a corresponding salinity). Below the homogeneous upper ocean layer was a stably stratified layer with a linear distribution of temperature and salinity (Figs. 3a, 3b). This profile approximately corresponds to conditions of the Weddel Sea off the Antarctic coast (Fig. 1c) with a homogeneous ice cover. We considered the adjustment of the upper ocean to an instantaneously formed lead, which was kept open sufficiently long. The lead was specified as a 100-m-wide strip aligned with the y axis. The model was run for 24 h. The computational domain (doubly periodic) was $1024 \times 512 \times 180$ m^3. Basic simulations were run with a 2-m grid spacing (it has been verified in advance that the results remain qualitatively unchanged when the grid spacing is reduced).

The under-ice heat flux from the ocean to the ice was assumed to be 2 W/m^2 (this value was taken approximately equal to its climatic mean and is accepted in classical sea-ice thermodynamics modeling studies [28]; such a flux has no large influence on results of simulations for short times, because its magnitude is significantly less than the absolute values of the sensible and latent heat fluxes over open water). The roughness length z_0 for the underwater sea-ice surface was set equal to 0.5 cm. At the open lead surface, the total flux of latent and sensible heat and the total flux of incoming and outgoing longwave radiation (no shortwave radiation flux during polar winter) were calculated from standard formulas used in large-scale models of the polar ocean (see, e.g., the Arctic Ocean Model Intercomparison Project, AOMIP, at https://www.whoi.edu/page.do?pid=29836). The near-surface wind speed was assumed to be 15 m/s, and the air–water temperature difference was 30°C; characteristic values of humidity and cloudiness were specified. Simulations were run with various directions of the mean flow in the upper ocean relative to the lead and with no mean flow. In addition, a test simulation in which the Coriolis force was neglected and a simulation in which the heat flux was evenly distributed over the entire surface of the computational domain but its integral magnitude was set equal to the flux concentrated in a narrow strip for simulations with a lead were performed to determine characteristics of lead convection.

3.2. Results of Simulations with Different Scenarios and Atypical Features of Lead Convection

The under-ice convection in the LES model is illustrated graphically in Figs. 2a–2d (all simulations in this section were run with the Coriolis force). Salinity fields after 24 h of the model time from the start of computation are shown in (a) horizontal and (b–d) vertical cross sections (fragments of the computational domain). The most vigorous mixing at the base of the upper ocean occurs when the surface heat flux is concentrated in a narrow strip and the sinking convective plumes are not destroyed by velocity shear (Fig. 2b). This regime is possible if the relative mean velocity is close to zero or directed along the lead. With an under-ice cross-lead flow, the velocity shear in the upper ocean destroys plumes and inhibits the intense generation and breaking of the internal waves at the base of the upper ocean, thereby weakening the entrainment and decreasing the growth rate of the mixed-layer thickness (Fig. 2c).

Some average characteristics of the upper ocean are shown in Figs. 3a–3f. The profiles of the first and second moments confirm characteristic features of under-ice convection in the different regimes described above. The least growth of the upper ocean thickness occurs if the heat flux from the ocean to the atmosphere is distributed uniformly throughout the surface (dashed curves). The bulk of the mixed layer is neutrally stratified, and unstable stratification forms near the surface. This is evident from Fig. 3c, in which the Brunt–Väisälä frequency is displayed by taking into account the sign of the density gradient $N = \sqrt{|N^2|}\,\mathrm{sign}\left(N^2\right)$, with negative values corresponding to unstable stratification. Note that local one-dimensional upper ocean models neglecting horizontal inhomogeneity have been calibrated so as to reproduce just this solution, which significantly differs from the solution with leads.

The heat exchange through a limited open-water area leads to intense frazil ice formation, thus increasing (with the same heat exchange across the surface) the salt flux (Fig. 3e) and, consequently, the buoyancy flux (Fig. 3f). With leads, the magnitude of the fluxes (both heat and salt) at the depths below the level of intense freezing depends little on the occurrence and direction of the ice drift (Figs. 3d, 3e; black and gray curves and open circles). In our simulations, these fluxes have approximately doubled versus the fluxes in the simulation with the horizontally homogeneous scenario (dashed curves). The position of the minimum in the buoyancy flux profile (Fig. 3f) matches the position of the entrainment layer. The convective layer has the largest depth (about 100 m for 24 h of the simulation) with no ice drift or along-lead ice drift (black and gray curves).

Note one more nonstandard property of lead convection when there is no transverse ice drift. This process results in the formation of a stably stratified average density profile. The average Brunt–Väisälä frequency in the top fraction of the upper ocean (black and gray curves in Fig. 3c) reaches about 1 cycle per hour, a value comparable to its counterpart in the halocline. Salinity increases on average with depth (Fig. 3b), even though the salt source associated with freezing is near the surface. Thus, there is a counter-gradient average salt transport in the vertical direction.
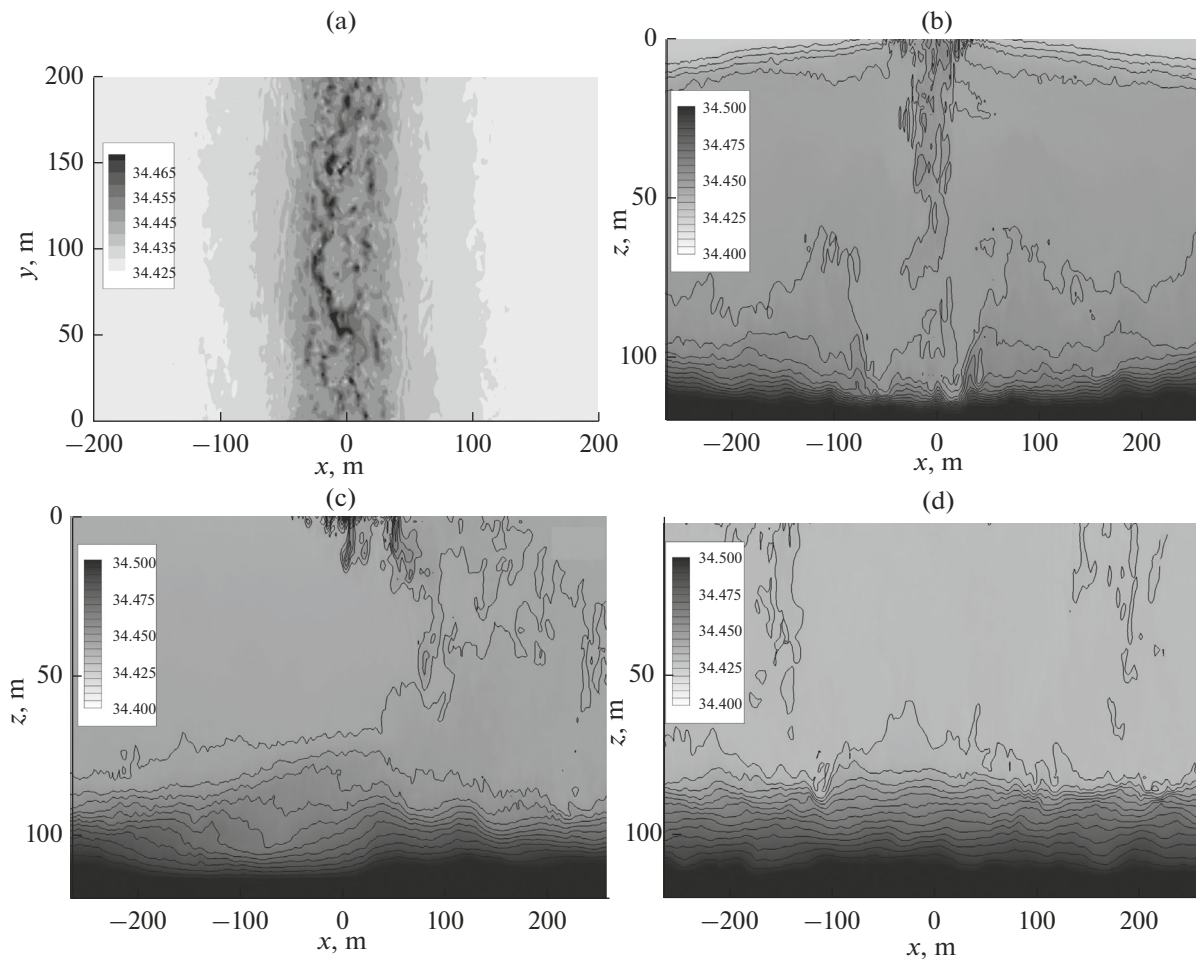
Fig. 2. Salinity in LES (instantaneous state after 24 h of the simulation); (a) horizontal cross section at 2-m depth, the lead 100 m in width is aligned with axis Y (fragment of the computational domain); (b, c, d) vertical cross section in the XZ plane ((b) difference of average ice and water velocities is zero, (c) relative current velocity is normal to the lead and is 3 cm/s, and (d) the heat flux is uniformly distributed over the surface (the area-averaged heat flux is the same in all three cases)).

This process cannot be described using the turbulent diffusion approximation.

The relatively freshened surface layer forms because salt is transported downward under the lead, where the sinking heavy and salty plumes inject the salt immediately to great depths. This is evidenced by the fact that in the first hours of convection development the salt flux is constant with depth from the surface to approximately one-third of the upper layer thickness and has a distinct maximum of the gradient near its bottom boundary, where salt accumulates (Fig. 3e; asterisks). Salinity near the surface under the ice would increase owing to the upward transport of salty water by a large-scale return current and, partly, to turbulent mixing, which is small at a distance from the zone of intense convection. In the case under consideration, the large-scale upward salt transport is weak and salinity beneath the ice remains unchanged throughout the simulation.

3.3. Influence of Earth's Rotation on Lead Convection

We will show that, in the case of motionless ice, despite small characteristic times of the plume descent to depths of the base of the upper ocean layer, the Coriolis force influences the convection process generally and is one of the causes of the formation of a near-surface stable freshwater layer. For this, we use results of an additional simulation in which the Coriolis force is turned off, but in all other respects it is analogous to the simulation shown in Fig. 2b.

The plume lifetimes were estimated using Lagrangian tracers, which were injected into the lead zone to the model grid cells nearest to the surface at each time step. Lagrangian trajectories were calculated by methods described in [25, 26]. The time required for the particles to reach some depth firstly was measured for each of them. Histograms of these times for all particles that crossed the horizontal plane at 80-m depth at hour 24 of the LES are shown in Fig. 4.
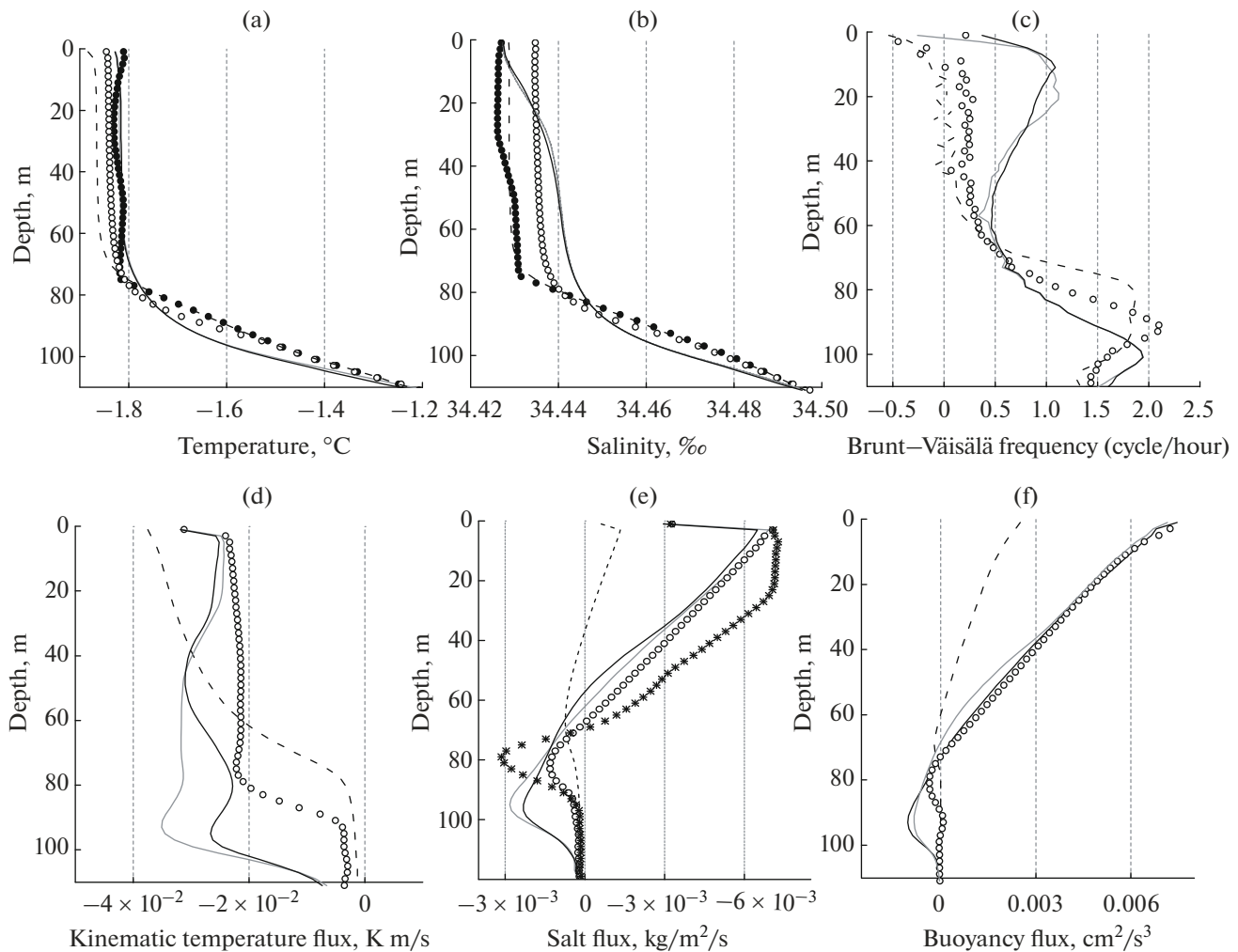
Fig. 3. Horizontally averaged (over the entire computational domain) and time-averaged (from hour 22 to hour 24 of the model integration) characteristics of the convective upper layer of the ocean in three numerical experiments. The ice is stationary relative to water (solid black curves); the 3-cm/s current is parallel to the lead (gray curves); heat flux is evenly distributed over the surface (dashed line); cross-lead current (open circles, see Fig. 2c). Black circles in panel (a) and (b) are the initial temperature and salinity profiles. (d) Salt flux averaged over hour 5 of the simulation with stationary ice (asterisks).
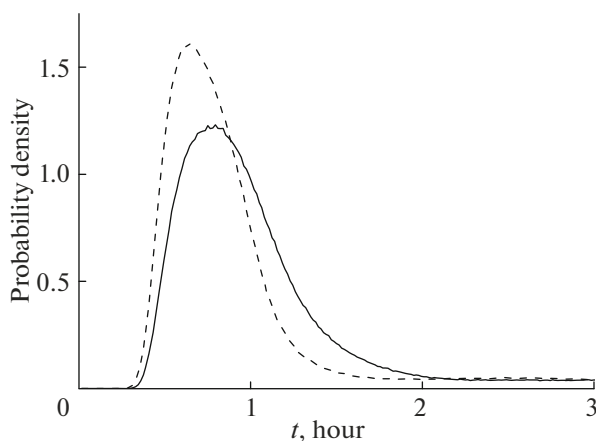
Fig. 4. Distributions histograms of the descent time of Lagrangian tracers from the surface to 80-m depth. Simulation with the Coriolis force (solid curve) and with no Coriolis force (dashed curve).

The time of the descent of particles to a depth comparable to the upper-ocean thickness is far less than the characteristic time $t_S \sim f^{-1}$, determined by the Coriolis force. For simulations with this force (solid line in Fig. 4), the average descent time is about 1.2 h and the distribution function peaks at 0.8 h. Accordingly, the Coriolis force cannot exert a direct influence on the sinking particles. Nevertheless, when the Coriolis force is ignored (dashed curve in Fig. 4), the descent time of the plumes is reduced.

The distinction we have found can be explained in the following way. The systematic descent of the plumes drives the long-lived large-scale circulation shown in Fig. 5 (the flow velocity averaged along the lead axis and over time for the 12th hour of the simulation). The flow directed to the lead axis forms near the surface. Under the influence of the Coriolis force,
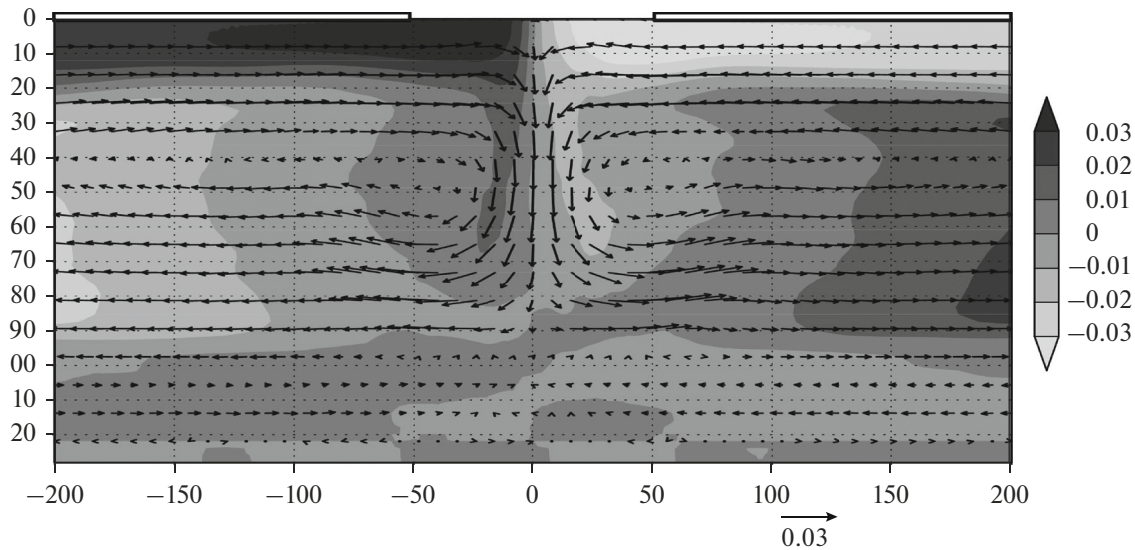
Fig. 5. Average current velocity within 12 h after the beginning of convection. The cross-lead horizontal velocity component U and the vertical component W are shown by arrows; the along-lead component V is shown by gray gradations. Velocities are in meters per second (m/s). The distance from the middle of the lead and depth (m) are plotted on the axes.

this flow acquires the along-lead velocity component characterized by a shear.

The average velocity shear produces small-scale vortices in the horizontal plane, which are depicted in Fig. 6a. In the upper ocean, these vortices have a vertically quasi-homogeneous axisymmetric structure and rotate clockwise, that is, opposite to the direction of the geostrophic balance in the Southern Hemisphere. Vortex-trapped salty heavy anomalies sink along spiral paths. The downward vertical velocity reaches maximum values on the periphery of these vortices. The bulk of the energy produced by buoyancy forces seems to be spent for maintaining these vortices, with the kinetic energy of the vertical velocity transformed into the energy of the horizontal vortex motion, which slightly slows down the convection process under the lead. With no Coriolis force, analogous organized patterns do not appear (Fig. 6b) and the flow in the convection zone is random and turbulent.

In addition to affecting the structure of the plumes directly beneath a lead, the Coriolis force changes the shape of a large-scale convective cell, transferring part of the kinetic energy to the along-lead velocity component. This localizes the process and prevents horizontal mixing and an increase in salinity near the surface under the ice away from the convection zone. The salinity averaged along the lead axis and over time for hour 24 of the simulation (the averaged salinity field in Fig. 7 at the bottom matches its instantaneous realization in Fig. 2b) is shown in Fig. 7. It can be seen that salinity in the simulation with rotation is distributed uniformly in the horizontal, except for a narrow convective zone under the

lead, while in the simulation with the Coriolis force a thicker freshwater layer remains near the surface at some distance from the lead.

CONCLUSIONS

We have numerically reproduced and analyzed under-ice convection in the upper ocean through leads with a width close to the mixed-layer thickness. We have modeled a process in which a major external buoyancy source is determined by ice rejection during frazil ice formation in the lead. The open-water area was about 10% of the total surface area, and heat fluxes to the atmosphere through a lead were consistent with their winter extremes at high latitudes. The experiments do not span the entire range of possible parameters of the process, but they are able to discover significant buoyancy-source localization effects at the qualitative levels. These effects are listed below and are subject to a further investigation and parameterization with consideration for the statistics of the distribution of leads and meteorological parameters of the atmosphere and ocean at high latitudes.

(1) The largest rate of growth of the average thickness of the convective upper ocean is achieved when the drift velocity is small or directed along the lead. In this case, the sinking plumes have the largest kinetic energy and the entrainment of the underlying dense water to the boundary layer is concentrated in a narrow band under the lead. The transverse drift and velocity shear break down and diffuse the narrow convection zone, thus reducing the growth rate of the upper ocean thickness. This effect is opposite to the conclusions made in [14] and [2] and is not typical of
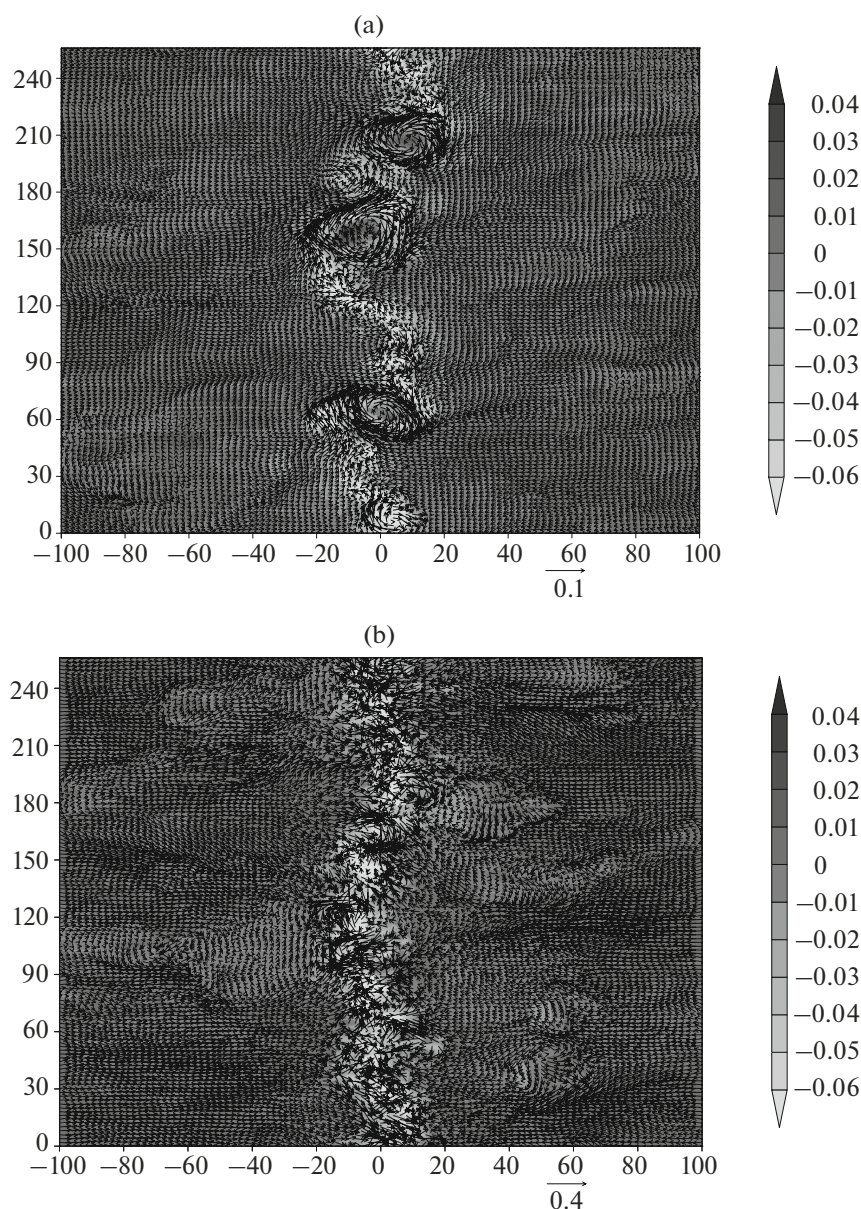
(a)

(b)

Fig. 6. Instantaneous current velocity under the lead within 12 h after the beginning of convection in a horizontal cross section at 30-m depth. Arrows show the horizontal velocity component, and gray gradations show the vertical component (negative values correspond to the downward velocity). (a) Simulation with the Coriolis force and (b) without the Coriolis force.

the horizontally homogeneous convective boundary layers, where the added shear generation of turbulent kinetic energy favors mixing.

(2) With a low drift velocity, the convection, injecting the heavy and salty water directly to large depths, results in the formation of a stable density profile (on average across the area comprising the lead and its ice-covered surroundings). The largest stability is in the top of the upper ocean, where a freshened layer forms. An explanation for this effect is as follows. Compared to the horizontally homogeneous upper ocean, convection through a lead increases the kinetic energy that

is spent for entraining and deepening the upper layer, thereby decreasing the upward transport of higher salinity and density anomalies by large-scale return currents. Close to the freezing point, temperature anomalies have a near-zero buoyancy, which makes their transport upward more efficient and, thus, induces the temperature mixing and the formation of an isothermal layer. In additional tests (not shown), the model result within 24 h after the start of convection was taken as the initial state and the heat flux through a lead was nullified. In such simulations without forcing, the residual turbulence and the resulting large-scale circulation were insufficient to mix salinity
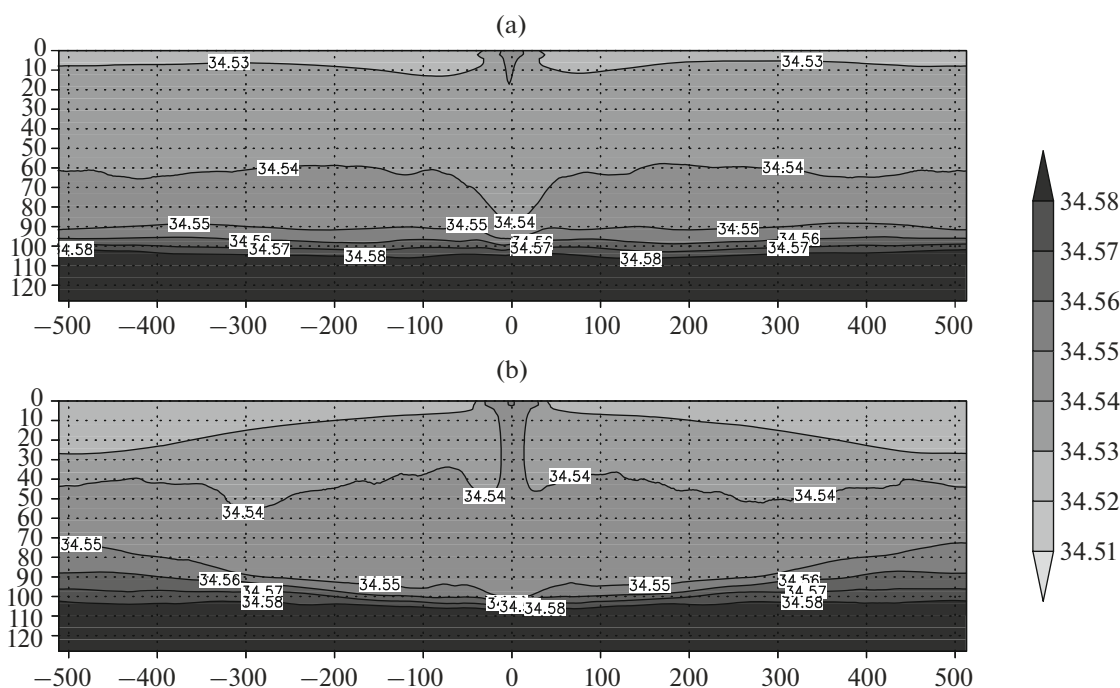
Fig. 7. Average salinity within 24 h after the beginning of convection. (top) Simulation without the Coriolis force and (bottom) with the Coriolis force.

in the upper ocean completely and stable stratification in its upper part retained. Thus, periodic lead formation and freezing will lead to a systematic effect. It seems quite possible that the simulated convection mechanism explains the formation of the observed cold halocline at high latitudes in the winter period.

(3) Despite short characteristic lifetimes of small-scale convective plumes, their structure and dynamics are subject to a significant influence of the Earth's rotation. This influence is achieved owing to the impact of the Coriolis force on the secondary large-scale circulation. In particular, small-scale helical vortices with a vertical axis of rotation, rotating clockwise in the Southern Hemisphere (their counterparts in the Northern Hemisphere rotate in opposite direction), were observed in the convective zone under the lead. The nature of these vortices is associated with a horizontal velocity shear of the large-scale flow, which forms throughout a convection time that exceeds the lifetime of individual plumes (see Fig. 5 and comment on it in Section 3.3). Similar mechanisms of vortex generation were discussed in [8], but helical vortices obtained in our simulations were not recorded in laboratory experiments, perhaps because of the specifics in the organization of measurements.

(4) The Coriolis force, redistributing the kinetic energy of large-scale circulation to the along-lead velocity component, localizes the process and prevents horizontal mixing and an increase in salinity near the surface under the ice at some distance from the convection zone.

Note that the effect of the Coriolis force in usual convective boundary layers of both the atmosphere and the ocean is small because the size of convective cells is compared to mixed-layer thickness and circulation time in these cells (close to the convective Deardorff timescale [29]) under typical conditions is far less than $1/f$.

Overall, these effects lead to salinity and temperature profiles averaged over large horizontal scales and the evolution of the profiles differ significantly on their counterparts in the horizontally homogeneous situation. Local one-dimensional upper ocean models based on turbulent diffusion will lead to systematic errors (most considerable at the water–ice interface), which may negatively affect the simulation of sea-ice evolution in the climate system models.

REFERENCES

1. P. Rampal, J. Weiss, and D. Marsan, "Positive trend in the mean speed and deformation rate of Arctic sea ice, 1979−2007," J. Geophys. Res. **114,** C05013 (2009).
https://doi.org/10.1029/2008JC005066

2. J. H. Morison, M. G. McPhee, T. B. Curtin, and C. A. Paulson, "The oceanography of winter leads," J. Geophys. Res. **97** (C7), 11199−11218 (1992).

3. S. Marcq and J. Weiss, "Influence of sea ice lead-width distribution on turbulent heat transfer between the ocean and the atmosphere," The Cryosphere **6,** 143−156 (2012).
https://doi.org/10.5194/tc-6-143-2012

4. M. Steele, R. Morley, and W. Ermold, "PHC: A global ocean hydrography with a high-quality Arctic Ocean," J. Clim. **14** (9), 2079−2087 (2001).

5. T. P. Boyer, J. I. Antonov, O. K. Baranova, C. Coleman, H. E. Garcia, A. Grodsky, D. R. Johnson, R. A. Locarnini, A. V. Mishonov, T. D. O' Brien, C. R. Paver, J. R. Reagan, D. Seidov, I. V. Smolyar, and M. M. Zweng, *World Ocean Database. NOAA Atlas NESDIS 72,* Ed. by S. Levitus (NOAA, Silver Spring, MD, 2013).
https://doi.org/10.7289/V5NZ85MT

6. E. C. Carmack, M. Yamamoto-Kawai, T. W. N. Haine, S. Bacon, B. A. Bluhm, C. Lique, H. Melling, I. V. Polyakov, F. Straneo, M.-L. Timmermans, and W. J. Williams, "Freshwater and its role in the Arctic Marine System: Sources, disposition, storage, export, and physical and biogeochemical consequences in the Arctic and global oceans," J. Geophys. Res.: Biogeosci. **121** (3), 675−717 (2016).
https://doi.org/10.1002/2015JG003140

7. K. Helfrich, "Thermals with background rotation and stratification," J. Fluid Mech. **259,** 265−280 (1994).

8. J. Bush and A. Woods, "Vortex generation by line plumes in a rotating stratified fluid," J. Fluid Mech. **388,** 289−313 (1999).

9. J. Morison and M. McPhee, "Lead convection measured with an autonomous underwater vehicle," J. Geophys. Res. **103** (C2), 3257−3281 (1998).

10. D. C. Smith IV and J. Morison, "Nonhydrostatic haline convection under leads in sea ice," J. Geophys. Res. **103** (C2), 3233−3247 (1998).

11. G. L. Mellor, *User Guide for A Three-Dimensional, Primitive Equation, Numerical Ocean Model* (Princeton University, Princeton, NJ, 2004).

12. G. L. Mellor and T. Yamada, "A hierarchy of turbulence closure models for planetary boundary layers," J. Atmos. Sci. **31** (7), 1791−1806 (1974).

13. M. Jin, J. Hutchings, Yu. Kawaguchi, and T. Kikuchi, "Ocean mixing with lead-dependent subgrid scale brine rejection parameterization in a climate model," J. Ocean Univ. China (**11** (4), 473−480 (2012).

14. E. D. Skyllingstad and W. D. Denbo, "Turbulence beneath sea ice and leads: a coupled sea ice/large eddy simulation study," J. Geophys. Res. **106** (C2), 2477−2497 (2001).
https://doi.org/10.1029/1999JC000091

15. P. Duffy and K. Caldeira, "Sensitivity of simulated salinity in a three-dimensional ocean model to upper ocean transport of salt from sea-ice formation," Geophys. Res. Lett. **24** (11), 1323−1326 (1997).

16. P. Duffy, M. Eby, and A. Weaver, "Effects of sinking of salt rejected during formation of sea ice on results of an ocean-atmosphere-sea ice climate model," Geophys. Res. Lett. **26** (12), 1739−1742 (1999).

17. A. T. Nguyen, D. Menemenlis, and R. Kwok, "Improved modeling of the Arctic halocline with a subgrid-scale brine rejection parameterization," J. Geophys. Res. **114** (C11014) (2009).

18. R. Timmermann and A. Beckmann, "Parameterization of vertical mixing in the Weddell Sea," Ocean Modell. **6** (1), 83−100 (2004).

19. T. Paluszkiewicz and R. D. Romea, "A one-dimensional model for the parameterization of deep convection in the ocean," Dyn. Atmos. Oceans **26,** 95−130 (1997).

20. N. G. Iakovlev, E. M. Volodin, D. V. Sidorenko, and A. S. Gritsun, "Role of penetrative convection under the ice in the formation of the state of the world ocean," Izv., Atmos. Ocean. Phys. **54,** 594−607 (2018).
https://doi.org/10.1134/s0002351518060147

21. A. V. Glazunov, "On the effect that the direction of geostrophic wind has on turbulence and quasiordered large-scale structures in the atmospheric boundary layer," Izv., Atmos. Ocean. Phys. **46,** 727−747 (2010).

22. IOC, SCOR, and IAPSO. *The International Thermodynamic Equation of Sea Water-2010: Calculation and Use of Thermodynamic Properties,* no. 56 of *Intergovernmental Oceanographic Commission, Manuals and Guides* (UNESCO, Paris, 2010).

23. J. Bardina, J. H. Ferziger, and W. C. Reynolds, "Improved subgrid scale models for large-eddy simulation," Am. Inst. Aeronaut. Astronaut. Pap. **80,** 1357 (1980).

24. M. Germano, U. Piomelli, P. Moin, and W. H. Cabot, "A dynamic subgrid-scale eddy viscosity model," Phys. Fluids A **3** (7), 1760−1765 (1991).

25. A. V. Glazunov, "Large-eddy simulation modeling of turbulence with the use of a mixed dynamic localized closure. Part 1. Formulation of the problem, model description, and diagnostic numerical tests," Izv., Atmos. Ocean. Phys. **45** (1), 5−24 (2009).

26. A. V. Glazunov, U. Rannik, V. Stepanenko, V. Lykosov, M. Auvinen, T. Vesala, and I. Mammarella, "Large-eddy simulation and stochastic modeling of Lagrangian particles for footprint determination in the stable boundary layer," Geosci. Model Dev. **9** (9), 2925−2949 (2016).

27. A. V. Glazunov, "Numerical modeling of turbulence and fine-particle pollutant transport in urban canyons," Vychisl. Metody Program. **19,** 17−37 (2018).

28. G. A. Maykut and N. Untersteiner, "Some results from a time dependent thermodynamic model of sea ice," J. Geophys. Res. **76,** 1550−1575 (1971).

29. J. W. Deardorff, "Convective velocity and temperature scales for the unstable planetary boundary layer and for Rayleigh convection," J. Atmos. Sci. **27** (8), 1211−1213 (1970).

Translated by N. Tret'yakova

Water Exchange between the Northern and Middle Caspian

G. S. Dyakonov[a, *] and R. A. Ibrayev[a, b, c]

[a]*Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, 117997 Russia*
[b]*Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, 119333 Russia*
[c]*Moscow Institute of Physics and Technology, Dolgoprudnyi, 141700 Russia*
*e-mail: gleb.gosm@gmail.com

Abstract—The Caspian Sea water dynamics on the boundary of its northern and central basins is considered. A high-resolution numerical model (with a grid step of ~2 km) has been used to reproduce the mesoscale structure of currents. The results of two experiments are presented: a realistic reconstruction of the sea circulation in 2003 and an idealized one with an artificial forcing. The realistic calculation considers the water exchange between the two basins, which occurs primarily due to coastal jet currents near Tyub-Karagan Peninsula in the east and near the Agrakhan Peninsula in the west. The formation and evolution of these currents under various synoptic situations are analyzed, and their flow rates are quantitatively estimated. The effect of contrasting water intrusions caused by these currents on the salinity field in both basins with highly different haline regimes is estimated. In particular, the western current (WC) is shown to increase the sea surface salinity (SSS) in the Northern Caspian and decrease its value in the Middle Caspian by almost 1−2 psu/year. The water exchange in the east is asymmetric: the northern current increases salinity in the northern basin by 2−3 psu/year, while the southern current decreases the average SSS in the central basin by only around 0.5 psu/year; in this case, the southern current occurs stably only in June and July. The idealized experiment simulates the Caspian water dynamics for winds of various points: from 0° to 350° with a step of 10°. The flow rates of both western and eastern currents have a sinusoidal dependence on the wind direction: their maxima are reached with winds of 40° and 220° points for the WC and with 120° and 300° points for the eastern current (EC). This study also analyzes the establishment of quasi-steady-state sea circulation at a constant wind.

Keywords: Caspian Sea, numerical modeling, wind-driven currents, gradient currents, water exchange

DOI: 10.1134/S0001433820030044

INTRODUCTION

The Caspian Sea consists of three basins: Northern, Middle, and Southern (hereafter, denoted as NCB, MCB, and SCB), partially separated from each other by peninsulas. Due to large differences in the bottom topography, nonuniform spatial distribution of the river runoff, and the large meridional extent of the sea, the thermohaline structure and circulation mode of waters in these basins are different [1]. The salinity in the extremely shallow-water NCB, which is the Volga River estuary, varies from 0 to 12 psu and is generally a few per mil lower than in the Middle and Southern Caspian [2]. In winter, the entire surface of the NCB freezes and the water temperature is several degrees lower than in the deepwater MCB. In summer, on the contrary, the NCB warms up well, while a powerful coastal upwelling raises cold deep waters to the surface of the MCB [3]; the differences in average temperatures in the two basins have opposite signs. As a result, the water masses penetrating from one basin to another have contrasting values of temperature, salinity, and density with respect to surrounding waters and thus highly affect the heat and salt balances

of the relevant basin, modulating the thermohaline structure of its waters.

The sea boundary of the mixing zone of the Volga and Caspian waters (isohaline of 11.6 psu) [4] goes approximately along the Tyub-Karagan Peninsula−Chechen' Island line (see Fig. 1); it is here that the NCB and MCB waters are exchanged. The water dynamics in this region is determined by wind, river runoff (primarily, the Volga river runoff), coastal contours, and bottom topography [1]. Here, the sea depth is only 10−20 m, increasing from north to south. In the absence of wind, the runoff currents carry fresh waters of the Volga river southward along the MCB western coast [5, 6], which generally determines the zonal distribution of sea surface salinity (SSS) in this basin. Along with other factors (wind, baroclinicity, and evaporation), this current creates a predominantly cyclonic circulation in the MCB and entire sea as a whole [7]. In this case, there is also evidence of a water-balance flow from the MCB to the NCB [8], which leads to the replenishment of the salt reserve of the NCB. This flow occurs predominantly in the eastern part of the interface between the two basins—along
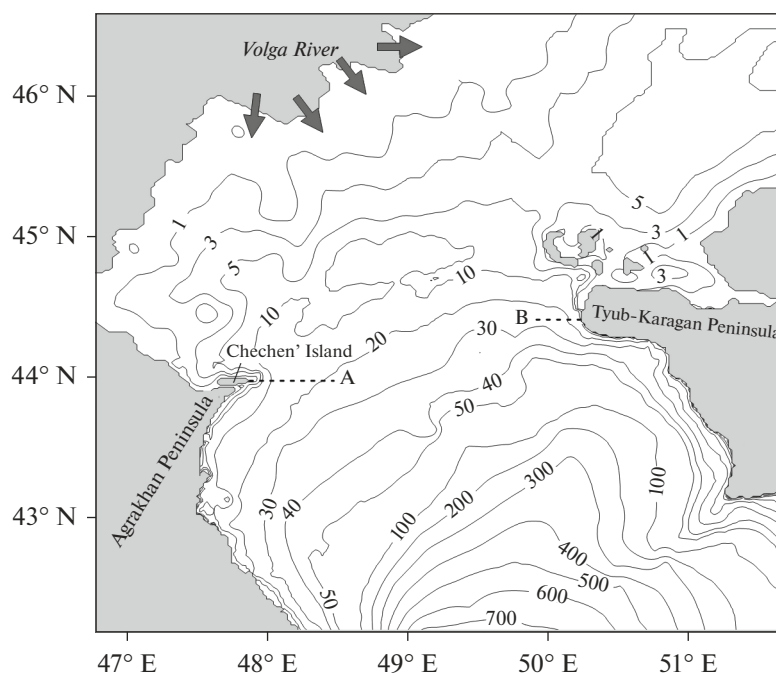
Fig. 1. Map and bathymetry (m) of the Caspian Sea region under consideration. The dotted lines indicate the sections for which the flow rates were calculated (Section 3): (A) for the western current and (B) for the eastern current (in the body of the text, WC and EC, respectively). The depth is given at a sea level of −28 m BHS (Baltic Height System).

the coast of the Tyub-Karagan Peninsula. However, this region has also a reverse water transport (from north to south). For example, such an episode was recorded in satellite images of early July 2003, when the northwestern wind created a jet stream of NCB waters (Fig. 2) with rates of almost 15 cm s^{-1} and a width of 20 km [9, 10]. Unlike the Volga river waters spreading along the western coast, they fall into the central part of the MCB and actively mix with its surface waters, significantly reducing their salinity and weakening the ver-
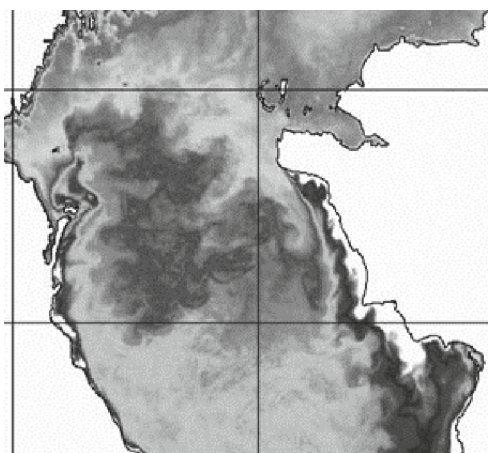
tical mixing in this deepwater basin [11]. Both currents (western and eastern) are reproduced well in modern models of the Caspian [12−15]. The rate of the water exchange between Caspian Sea basins is estimated both using model calculations [16] and from an analysis of observational data [17].

In this paper we study the structure of currents at the MCB and NCB boundary and their intensity and temporal variability. First and foremost, coastal jet currents are considered because they are what create regular penetrations of contrasting waters into both basins, highly affecting the sea surface temperature and salinity there. This study is based on an analysis of sea circulation reproduced in the SZ-COMPAS numerical model [11] developed specifically for the Caspian Sea using a seamless prediction approach [18, 19]. SZ-COMPAS results from further development of the Model for Enclosed Sea Hydrodynamics (MESH), which was used in our previous studies of the Caspian [5, 12]. SZ-COMPAS was verified and used earlier to investigate the water balance of the Caspian Sea and the dynamics of its waters [11, 20, 21]. This paper consists of four sections. The first section briefly describes the model and the numerical experiment. The second section considers the dynamics of currents on the interface between the NCB and MCB. The third section estimates the intensity (flow rate) of jet currents between these basins, as well as their impact on the haline regime of the sea. Finally, the fourth section considers the effect of wind direction on these currents.

Fig. 2. Sea surface temperature according to an infrared image of the NOAA-16 satellite on July 3, 2003 [9].

1. NUMERICAL MODEL
OF THE CASPIAN SEA DYNAMICS

The sea currents are calculated using the model described in [11, 20]. The horizontal resolution is 2 km, which is sufficient to describe the mesoscale dynamics of water: in the Caspian Sea, the baroclinic Rossby radius of deformation is estimated at 17−22 km in deepwater regions and 3−8 km on the shelf in the eastern part of the MCB [22]. The vertical resolution is from 2 to 30 m. The horizontal turbulent viscosity is described using a fourth-order Smagorinsky parameterization scheme [23] with a minimum dimensionless parameter C = 2 as recommended by the authors. Among the numerical models of the Caspian Sea known to us, only two [13, 15] have higher resolution (around 1.5 km); however, our model has a significantly lower dissipation and, therefore, a higher effective resolution, which make it possible to reproduce a wide range of motions: from large-scale to mesoscale. The vertical turbulent viscosity is parameterized using the Munk−Anderson scheme with a maximum coefficient of $K_m = 10^{-3}$ m^2 s^{-1}.

The model is initialized by three-dimensional climatic fields of temperature and salinity [2] interpolated to the model computational grid and is "spun up" for 3 years until these fields become realistically distributed in coastal areas. The resulting fields are used as initial conditions for the experiments considered below.

The boundary conditions on the sea surface are set from ECMWF ERA-Interim data [24] with a high spatial resolution of 80 km for global reanalysis. The ERA-Interim fields are also interpolated to the model grid. In this case, all atmospheric fields, with the exception of precipitation and solar radiation, have significantly different contrasting values depending on the type of underlying surface (sea or land); therefore, the interpolated fields in the coastal zone may have values untypical for marine cells. This phenomenon is known as *Land Contamination* [25, 26]. To avoid this situation, the ERA-Interim fields are interpolated taking into account the land/sea mask indicating the type of underlying surface used in this reanalysis.

Since calculations using the Caspian Sea model with a resolution of 2 km for a large number of years require extremely large computational resources, we decided to analyze the sea circulation for 1 year when the wind conditions in the Caspian basin were typical for this region. To do this, we used monthly mean ERA-Interim data averaged over 30 years (1979–2008). Based on a comparison with this "climate" of annual average fields for individual years, the year 2003 was chosen. Here, we analyzed first the fields of the resulting wind and wind-speed module: generally, they are very consistent with the reanalysis long-term averages. In turn, the ERA-Interim "climate" for 1979−2008 was compared with data of the Caspian Sea hydrometeorological atlases to conclude that this reanalysis provides sufficiently reliable data on the

atmospheric forcing in the given region [27]. Thus, the Caspian Sea circulation obtained in this study can be taken as rather typical.

2. WATER-EXCHANGE STRUCTURE

For a detailed study of the structure and variability of water exchange in the MCB and NCB, we conducted a numerical calculation of the Caspian Sea circulation for 2003. An analysis of the results of this calculation showed that the mixing of waters of these two basins is largely conditioned by alongshore jet currents that arise near the coast of the Agrakhan Peninsula and Chechen' Island in the west (hereafter denoted as the western current (WC)) and near the Tyub-Karagan Peninsula in the east (hereafter denoted as the eastern current (EC)).

The WC (Fig. 3a) is more stable due to its association with the Volga River runoff. The water transported by this current from the NCB lowers the salinity of the upper sea layer in the west of the MCB, conditioning the zonal distribution of SSS in this basin, with the exception of the summer period. At the same time, under the influence of wind, the WC can multiply amplify or go in the opposite direction (flow rate). According to model calculations, the consumption of this jet in 2003 was up to 300 mSv (see Section 3). It will be shown below that this current is most intense under northeastern winds that are parallel to the coastline; however, it can also exist under winds of other points. For example, eastern winds increase the level in the western part of the NCB, which leads to the generation of southward WC even under unfavorable local wind conditions. If the easterly wind is replaced by a northeasterly, the wind effect coincides in direction with the effect of the level gradient, and the current arising along the MCB west coast is most intense. An example of this situation is shown in Fig. 3a: the WC velocities exceed 100 cm/s under a local wind speed of 5−7 m/s. It can also be seen from this figure that the easterly wind created a northwestward alongshore current along the Tyub-Karagan Peninsula, which separates from the coast, follows the isobaths, and merges with the WC near Chechen' Island, closing the branch of the cyclonic circulation of the Middle Caspian. This pattern is typical for the winter circulation of the Caspian. In this case, the EC supplies saline MCB waters to the NCB, replenishing its salt balance. The numerical experiments indicated that this balance is extremely sensitive to the parametrization of bottom friction in the model: with an excessively viscous scheme for describing this factor, the average salinity of the NCB almost vanishes within 10 years [11].

Figure 3b shows the opposite situation: the northwesterly wind directs the EC southward, conditioning the flow of relatively fresh NCB waters to the MCB. This jet is clearly visualized on a satellite image taken at the same time instant (Fig. 2). In this case, the EC is alongshore; however, under a northerly wind, this
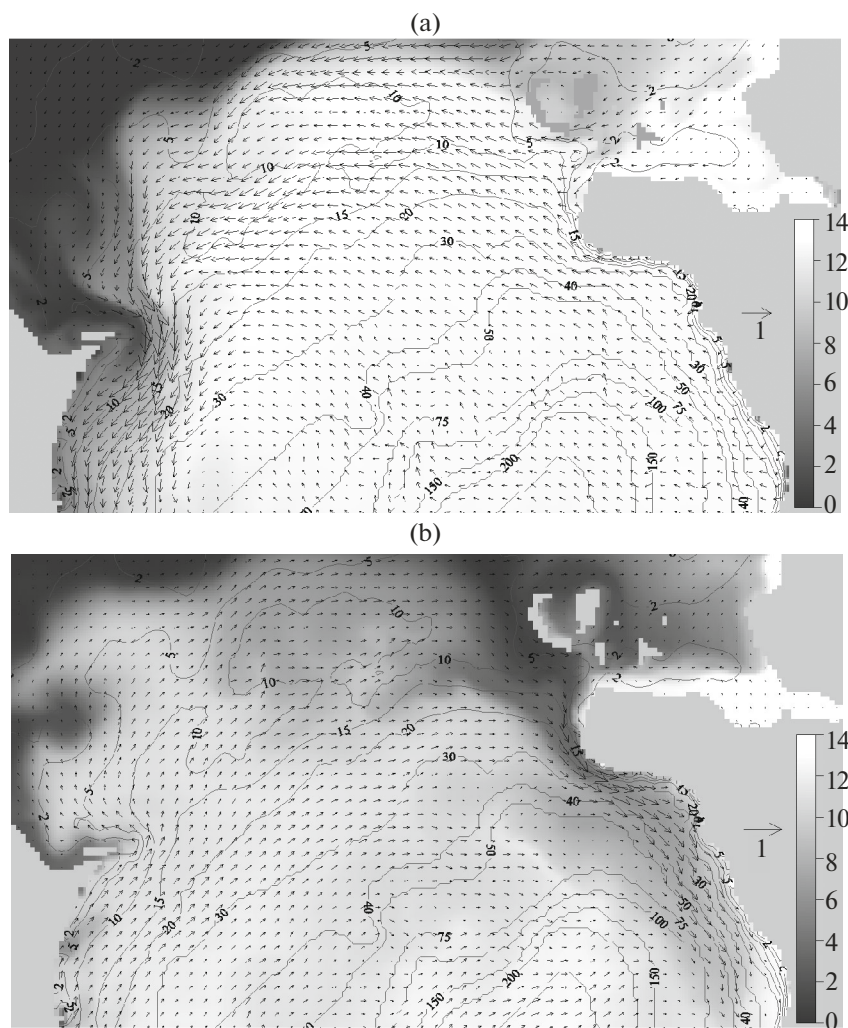
Fig. 3. Currents (m/s, every third vector plotted) and salinity (psu) on the surface near the interface between the Middle and Northern Caspian obtained by the model for 2003: (a) March 8 and (b) July 3. The contour lines indicate the sea depth.

current is southward or southwestward, and fresh waters appear in the central deepwater part of the MCB, where, due to high buoyancy, they significantly affect the intensity of vertical mixing and the convection depth [11]. According to the calculation results, the EC flow rate in 2003 reached 70 mSv and the velocities normally did not exceed 50−60 cm/s at the surface.

The open part of the sea, between Chechen' Island and the Tyub-Karagan Peninsula, is dominated by drift currents, while eddy structures are rare here. Under northeasterly and southwesterly winds, these currents are directed along isobaths and contribute little to the mixing of MCB and NCB waters. At northwesterly and southeasterly wind points, frontal transport along the entire interface occurs between the basins (Figs. 4a, 4b); however, due to the low stability of wind conditions, these anomalous water penetrations rarely manage to shift the SSS front further than by 20−30 km. Under these conditions, the direction of water motion in the EC zone coincides with the general drift direction,

while in the zone of the WC, offshore Chechen' Island, the current is more governed by the level gradient than by local wind. Under the northwesterly wind (Fig. 4a), the WC is relatively stable; under the southeasterly wind (Figs. 4b, 4c), the currents in this region are oscillatory, which contributes to the formation of eddies. In this case, the wind creates a positive level anomaly in the western part of the NCB, which, as noted above, generates a southward (i.e., directed against the wind) WC (Fig. 4c). Thus, under the southeasterly wind, the WC pulsates, changing its direction from north to south and vice versa, which gives rise to mesoscale eddies with a diameter of 15−30 km (Fig. 5a). In the example considered in Fig. 5a, the lifetime of these eddies did not exceed 5−10 h. The EC generates similar vortices: Fig. 5b shows an example of an EC-induced cyclone on the MCB surface south of the Tyub-Karagan Peninsula. This eddy soon separated from the jet stream of the EC and persisted significantly longer (almost 8−9 days), because this part of the sea is deeper: 50−75 m.
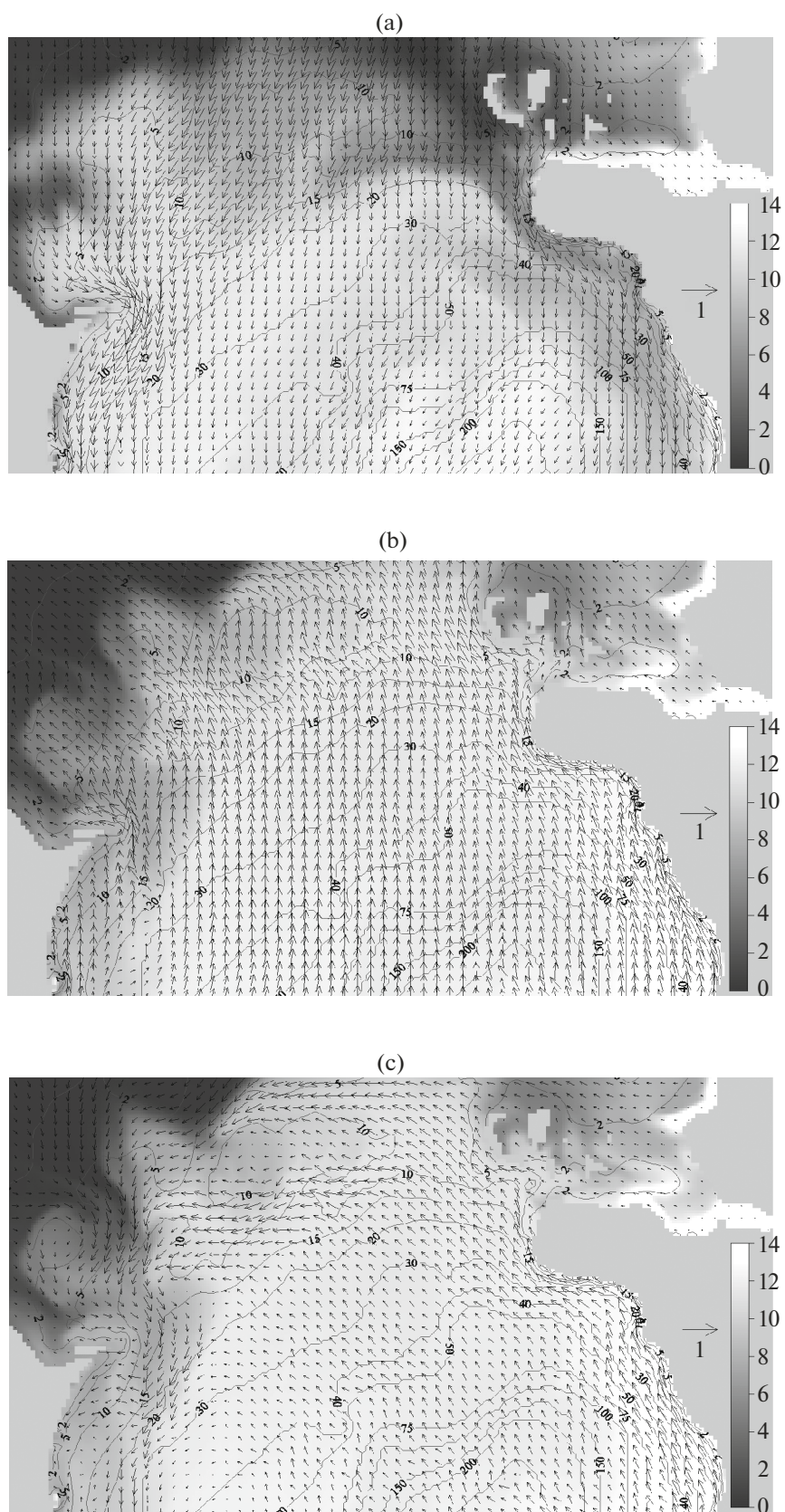
(a)

(b)

(c)

Fig. 4. Same as Fig. 3, except for (a) July 8 (northwesterly wind) and (b) at 12:00 a.m. on September 28 and (c) 12:00 a.m. on September 29 (southeasterly wind). Time is in GMT.
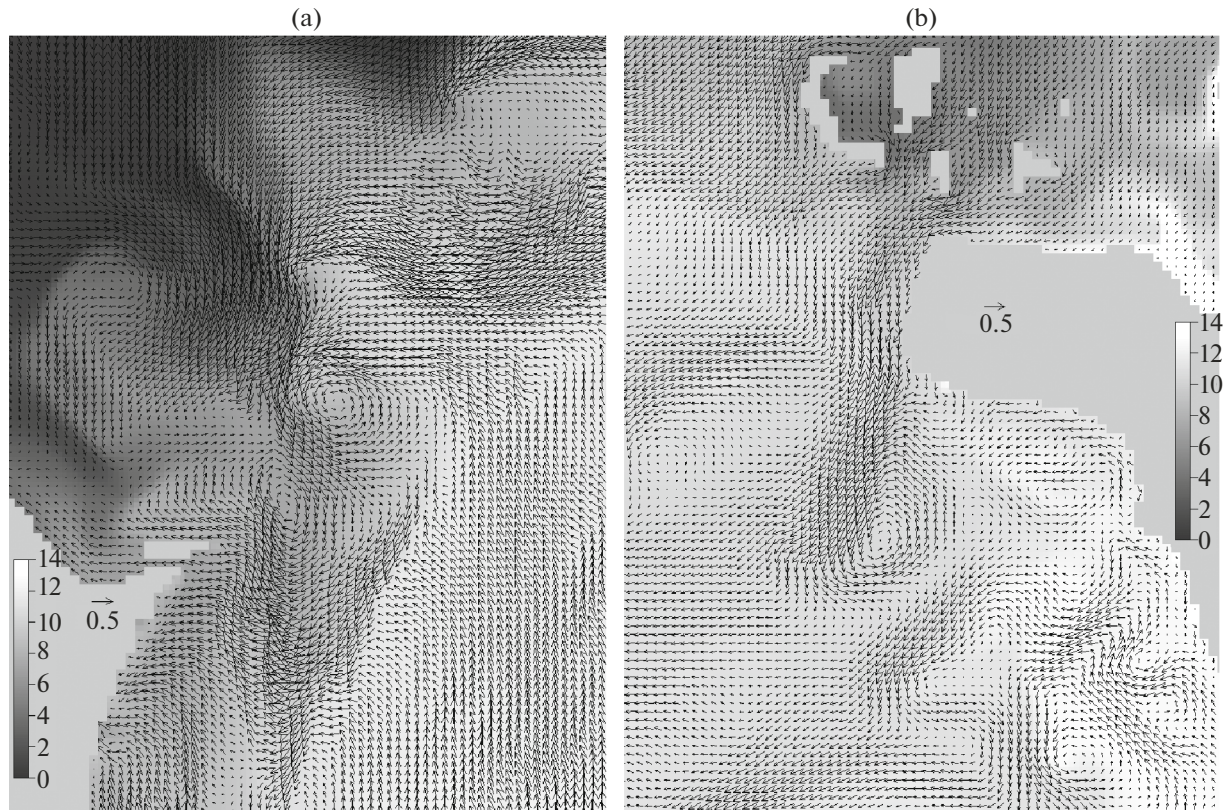
Fig. 5. Currents (m/s) and salinity (psu) on the surface obtained by the model: (a) near Chechen' Island at 6:00 p.m. on September 28, 2003, and (b) near the Tyub-Karagan Peninsula at 6:00 p.m. on August 27, 2003. Time is in GMT.

3. CURRENT FLOW RATES AND THEIR IMPACT ON SALINITY

To study the effect of currents in the given region on the thermohaline structure of Caspian basins, one needs to track the trajectories of waters they carry and the processes of their mixing with surrounding waters, which itself is a difficult problem. However, the significance of these currents for the thermohaline regime of the MCB and NCB can be estimated more simply by calculating the flows between them and somehow comparing them with the amount of heat and salt available in these basins. As we have already noted in the previous section, the waters with contrasting characteristics penetrate into the MCB from the NCB and vice versa mainly during their transport by the alongshore current: WC or EC. In this section, we quantitatively estimate the amounts of these penetrations, as well as their impact on the saline regime of both basins. To do this, we consider the localization zones of WC and EC. These jets are most coherent in areas where they are parallel to isobaths and the shore (see Figs. 3a, 3b); therefore, we calculate their flow rates on the vertical sections indicated in Fig. 1. We find the flow rate as the integral of the meridional velocity component over the section surface. We also evaluate the current effect of the MCB and NCB surface salinity.

To do this, we calculate the quantity (with a dimension of psu/s)

$$dS^{\pm} = \frac{1}{A^{bas}} \int \frac{|v(x)| \pm v(x)}{2} \left(S_{flow}(x) - \overline{S}^{bas} \right) dx,$$

where v and S_{flow} are the meridional velocity component and SSS at the section latitude, respectively; $\overline{S}^{bas}$ is the average SSS in the basin which receives this current (MCB or NCB); and A^{bas} is the surface area of this basin. All functions are also time dependent; the integral is taken over the section width. dS^+ and dS^- characterize the northward and southward flows, respectively. Formally, this quantity is the rate of change in the average SSS $\overline{S}^{bas}$ due to water transport by the current from the neighboring basin under the assumption that they displace water with medium salinity $\overline{S}^{bas}$ and remain on the surface. Since the penetrating waters can go down to the neutral buoyancy depth, $dS^{\pm}$ has no direct physical sense. However, this variable can be taken as a quantitative measure for the contrast of salinity of the waters carried by the current relative to the waters of the basin that receives them, as well as to the extent that the water exchange between the MCB and NCB can affect their haline structure.
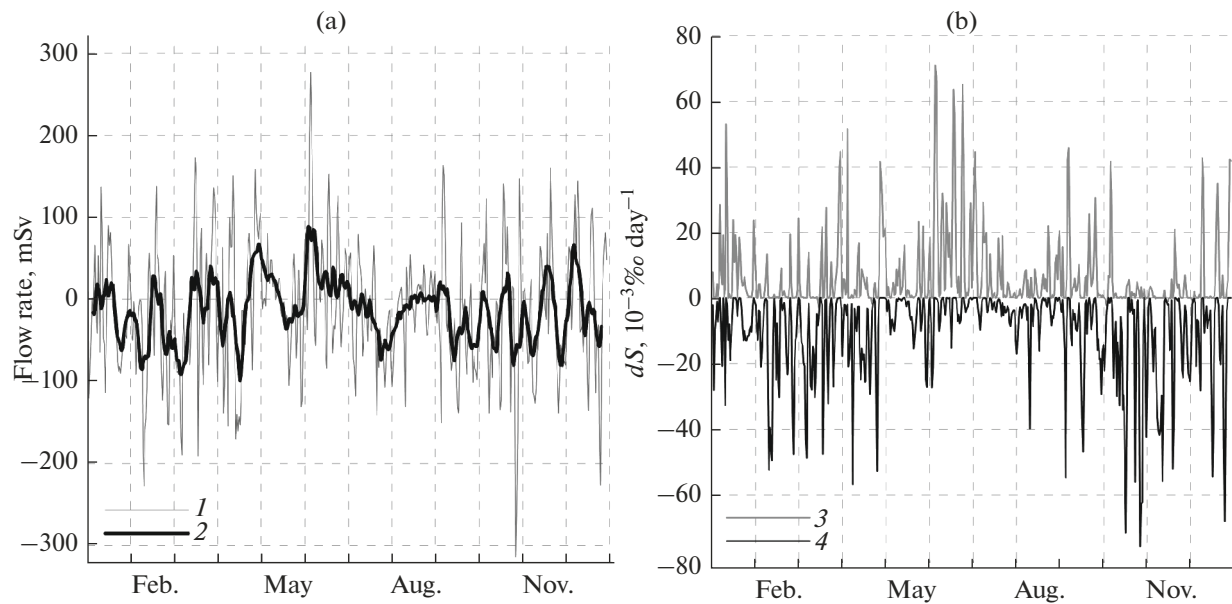
Fig. 6. Parameters of the WC on section A (see Fig. 1) calculated by the model for 2003: (a) flow rate, mSv and (b) $dS^{\pm}$, 10^{-3} psu/day, related to this current. (*1*) Flow rate, (*2*) moving average, (*3*) dS^{+} (flow in the NCB), and (*4*) dS^{-} (flow in the MCB). The northward current corresponds to a positive flow rate and a salinity change in the NCB (curve *3*) and the southward current corresponds to a negative flow rate and a salinity change in the MCB (curve *4*).

These two quantities (flow rate and $dS^{\pm}$) calculated by the model for 2003 are shown in Figs. 6 (for the WC) and 7 (for the EC). It can be seen that the WC is directed predominantly southward in most of the year, which is due to the predominance of winds of eastern points, as well as by the Volga River runoff. In this case, the first half of May and June is often characterized by the opposite situation. Since there is a quasi-constant negative salinity anomaly in the west of the MCB, the WC transports relatively fresh waters in both directions (with a salinity significantly lower than in the MCB), and the effect of this transport on the salinity of both basins is comparable in amplitude. However, the WC effect acts exclusively to increase salinity in the NCB and decrease it in the MCB. By order of magnitude, this action can be estimated at 1−2 psu/year. Interestingly, the WC-related desalination of MCB waters during the flood period is minimal during the year. This is due to the fact that river water flows southward with the EC in June−July and the WC is unstable in May and has no clearly expressed direction.

In January−April, the EC is more often northward, thus creating a water-balance flow from the MCB to the NCB; its existence is discussed in [8]. From mid-September to the end of the year, the EC has no stable direction. In June and July as well as in mid-August, the prevailing northwesterly wind leads to a stable southward transport of NCB waters near the Tyub-Karagan Peninsula (Fig. 7a). It is the long-term nature of this current that ensures the maximum decrease in surface salinity in the MCB, while the direction of these waters in other periods of the year often reverses,

failing to completely mix with waters of a particular basin. It can be seen from Fig. 7b that due to this transport the average SSS in the MCB decreases annually by almost 0.5 psu/year. The remaining period of the year is characterized by a regularly arising northwestward EC that carries the most saline waters of the MCB, which annually replenishes the salt budget of the NCB by 2−3 psu. Therefore, the salinity in the NCB northwest of the Tyub-Karagan Peninsula is often higher than the average SSS in the MCB, which explains the positive values of dS^{-} in Fig. 7b. For the same reason, dS^{-} in June lags by almost a week behind the flow-rate schedule: intensive southward intrusions begin in early June, when the salinity in the NCB northwest of the Tyub-Karagan Peninsula is almost the same as that in the MCB and the flow of waters with abnormally low salinity into it begins only when the salt water is displaced southward.

4. ALONGSHORE CURRENTS UNDER WINDS OF DIFFERENT POINTS

The current patterns in the North Caspian under a uniform wind of different directions were studied in [4]. To analyze the effect of wind direction on the western and eastern alongshore currents (WC and EC) considered above, we conducted a series of idealized model experiments in addition to the realistic calculation. The model configuration is identical to that described above, but artificial forcing is used: a uniform wind field is given in the entire Caspian with a speed of
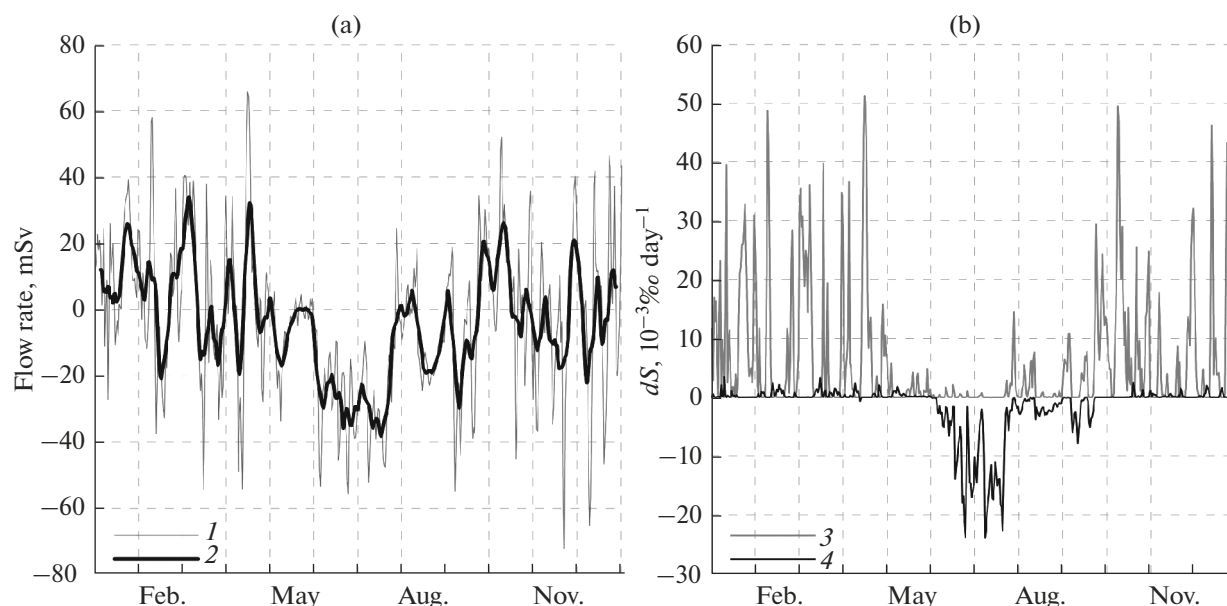
Fig. 7. Same as Fig. 6, except for the EC on section B (see Fig. 1).

5 m/s and a direction varying from 0° to 350° with a step of 10°. Thus, a total of 36 experiments were performed. Under real conditions, the wind direction extremely rarely persists longer than several days; therefore, each calculation is limited to 5 days, which is sufficient to establish a quasi-steady-state circulation in the upper layer of the sea. All other components of the atmospheric impact are disabled and, in addition, constant runoffs of the Volga and Ural rivers are set with a rate of 8000 and 150 $m^3 s^{-1}$, respectively, which is close to their climatic average annual flow rates. The river runoffs are necessary to obtain a realistic level anomaly and a related riverine component of the currents. In each of the 36 experiments, the flow rates of the WC and EC are calculated for sections A and B, respectively (Fig. 1) and the resulting currents and the process of their establishment are analyzed.

The dependence of the intensity of both currents on wind direction is close to sinusoidal: the maximum flow rate of the WC is reached at winds of 40° (northeasterly) and 220° (southwesterly); the maximum flow rate of the EC is reached at winds of 120° (southeasterly) and 300° (northwesterly). No critical wind directions with sharp jumps in the flow rate of either current (caused, for example, by the flow around peninsulas) were observed. The indicated extreme points are obviously due to the local direction of the coastline near the sections where the intensity of the currents is calculated. The WC and EC flow evolution for some wind points is shown in Fig. 8. The maximum flow rate of a southward WC (at a NE wind) is slightly higher than the similar value at a reverse transport, which is caused primarily by the runoff of rivers flowing into the MCB mostly with the WC. The situation is opposite for the EC: the northward flow is slightly

larger than the southward flow, which is associated with bottom friction (its effect is significantly higher in the shallow-water NCB than in the deepwater MCB).

It also follows from Fig. 8 that the adaptation of the EC and WC to wind is qualitatively different due to the basin geometry. Unlike the WC, the EC is not exactly alongshore: the EC follows the coastline only for a small extent and then separates from the coast to be found in the open sea (see Fig. 5b). Therefore, the geostrophic adjustment of the EC is of a nature typical for the open sea: several inertial pulsations are followed by the jet intensity levelling off (Fig. 8b). The WC, on the contrary, is alongshore in the MCB almost entirely, which reduces the effect of inertial pulsations, and the flow rate of the WC increases until the wind effect is compensated by the level difference in the MCB and NCB.

It is of interest to consider the evolution of the WC and EC under the winds of nonextreme points. The southeasterly wind first induces a northward WC (curve 120° in Fig. 8a); however, as was noted in Section 2, this wind increases the level in the west of the NCB and the resulting pressure gradient reverses the WC near Chechen' Island against the local wind. In 5 days, the flow rate of the WC reaches values that are only 40% less than with the extreme northeasterly (with a point of 40°) wind. Nevertheless, the WC during the first days is northward, although a faster reversal at winds with speeds exceeding 5 m/s should be expected. At a southwesterly wind, the EC established after 2 days seaward the Tyub-Karagan Peninsula is strictly eastward, has an almost vanishing flow through the respective section (curve 220° in Fig. 8b), and is divided by the peninsula into two (southern and northern) branches. In this case, the EC in the first
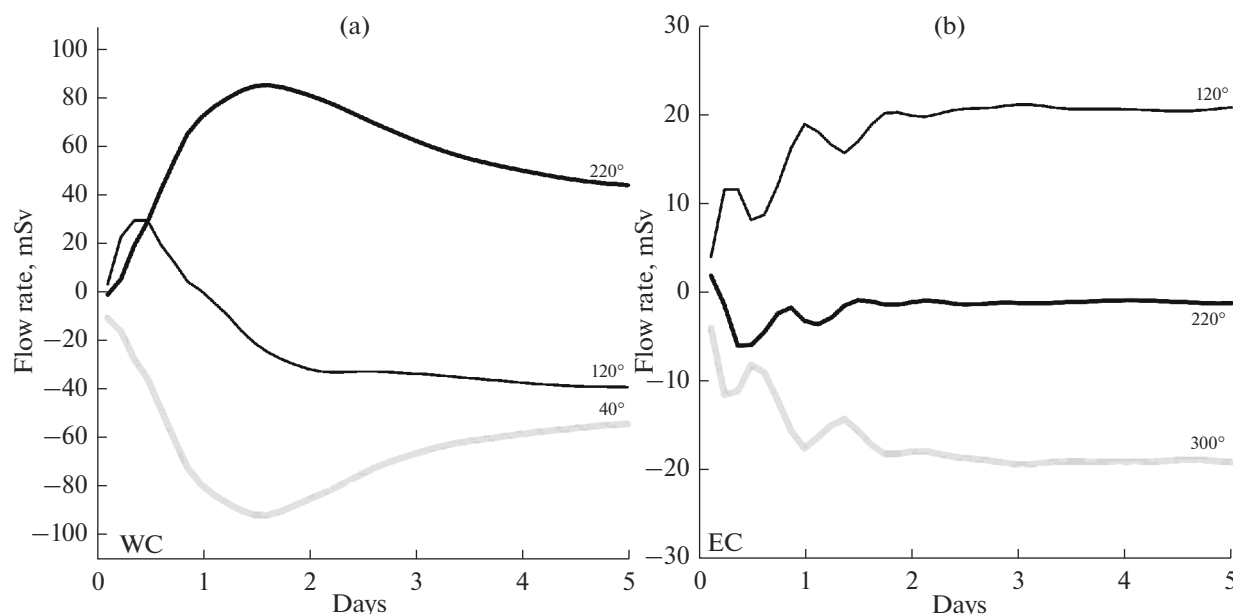
Fig. 8. Flow rate of WC (a) and EC (b) for different wind directions in the idealized experiment described in Section 4. The captions for the curves indicate the wind direction: 40° (northeasterly wind), 120° (southeasterly wind), 220° (southwesterly wind), and 300° (northwesterly wind).

hours is directed downwind, but against the local gradient of bottom topography, which causes the rapid southward reversal of this current.

CONCLUSIONS

The water exchange between the Middle and Northern Caspian basins (MCB and NCB) occurs mainly by coastal currents: at Tyub-Karagan peninsula in the east and the Agrakhan Peninsula in the west. Both currents can have both southern and northern directions and normally have a jet character, which is caused by the effect of the coast and bottom topography. These jets are most coherent under alongshore winds. In the west, the current is more often southward and carries the North Caspian and Volga waters to the MCB, forming a negative salinity anomaly along its western coast. According to the model calculation results presented in this paper, the flow rate of the WC reached 300 mSv in 2003, and its effect on salinity can be estimated at 1–2 psu/year for both basins (an increase in the NCB and a decrease in the MCB).

The EC with a maximum flow rate of 70 mSv in the same year is the main factor that compensates for the decrease in salinity of the NCB due to the river water inflow into it: according to model estimates, the inflow of Middle Caspian waters along the coast of the Tyub-Karagan Peninsula increases salinity in the NCB by 2–3 psu annually. At the same time, in June–July, the northwesterly winds prevailing in summer create here a stable southward transport, which decreases the surface salinity in the MCB by almost 0.5 psu. It should be noted that this estimate is probably somewhat overestimated, since the annual inflow

of relatively fresh waters led to an underestimated average long-term salinity in the MCB by 0.5–0.7 psu according to the numerical reconstruction of the Caspian circulation in an experiment similar to the one presented in this paper, but conducted over several decades [11].

This study also shows that these currents are caused not only by wind, but also by the sea level gradient, which often directs the water flow against the local wind action. Although the sea depth is small near Chechen' Island, vortices and vortex dipoles with a characteristic diameter of 15–30 km are regularly observed here, which is associated with the jet current instability during wind direction changes or upon a jet reversal by the pressure gradient. However, the lifetime of these mesoscale structures does not exceed several hours. In the east, near the Tyub-Karagan Peninsula, the southward current generates vortices of the same size in the MCB, but existing for more than a week, even after separation from the current that created them.

ACKNOWLEDGMENTS

FUNDING

REFERENCES

1. *Hydrometeorology and Hydrochemistry of the Soviet Seas,* Vol. 6: *The Caspian Sea,* No. 1: *Hydrometeorological Conditions,* Ed. by F. S. Terziev, A. N. Kosarev, and A. A. Kerimov (Gidrometeoizdat, St. Petersburg, 1992) [in Russian].

2. A. N. Kosarev and V. S. Tuzhilkin, *Climatic Thermohaline Fields in the Caspian Sea* (Sorbis, Moscow, 1995) [in Russian].

3. A. I. Ginzburg, A. G. Kostyanoi, D. M. Solov'ev, and N. A. Sheremet, "Upwelling structure at the western coast of the Central Caspian Sea (according to satellite observations)," Issled. Zemli Kosmosa, No. 4, 76−85 (2005).

4. *The Volga Estuary: Hydromorphological processes, Pollutant Regimen, and the Effects of Caspian Sea Level Variations,* Ed. by V. F. Polonskii, V. N. Mikhailov, and S. V. Kir'yanov (GEOS, Moscow, 2008) [in Russian].

5. R. A. Ibraev, *Mathematical Modeling of Thermohydrodynamics Processes in the Caspian Sea* (GEOS, Moscow, 2008) [in Russian].

6. A. B. Kara, A. J. Wallcraft, E. J. Metzger, and M. Gündüz, "Impacts of freshwater on the seasonal variations of surface salinity and circulation in the Caspian Sea," Cont. Shelf Res. **30** (10−11), 1211−1225 (2010). https://doi.org/10.1016/j.csr.2010.03.011

7. V. N. Zyryanov, "Hydrodynamic basis of formation of large-scale water circulation in the Caspian Sea: 2. Numerical calculations," Water Resour. **43,** 292−305 (2016). https://doi.org/10.1134/S0097807816020184

8. L. P. Ostroumova and V. F. Polonskii, "Water balance model of the North Caspian and its separate parts and estimation of river water inflow to the Middle Caspian Sea," Trudy GOIN, No. 213, 343−355 (2011) [in Russian].

9. O. Yu. Lavrova, A. G. Kostyanoi, S. A. Lebedev, M. I. Mityagina, A. I. Ginzburg, and N. A. Sheremet, *Complex Satellite Monitoring of the Russian Seas* (IKI RAN, Moscow, 2011) [in Russian].

10. S. A. Lebedev, A. G. Kostyanoi, and A. I. Ginzburg, in *Mater. III Int. Sci.-Pract. Conf. "Aspects of Applied Geology, Geophysics, and Geoecology Using Modern Information Technology" (Maikop, 11−14 May 2015)* (IP Kucherenko V.O., Maikop, 2015), pp. 146−179.

11. G. S. Dyakonov and R. A. Ibrayev, "Long-term evolution of Caspian Sea thermohaline properties reconstructed in an eddy-resolving ocean general circulation model," Ocean Sci. **15,** 527−541 (2019). https://doi.org/10.5194/os-15-527-2019

12. R. A. Ibrayev, E. Özsoy, C. Schrum, and H. I. Sur, "Seasonal variability of the Caspian Sea three-dimensional circulation, sea level and air-sea interaction," Ocean Sci. **6,** 311−329 (2010). https://doi.org/10.5194/os-6-311-2010

13. S. K. Popov, V. I. Batov, V. V. Elisov, and A. L. Lobov, "Advanced technology of the Caspian Sea currents and level forecasting," Zashch. Okr. Sredy Neftegaz. Kompl., No. 5, 53−59 (2013).

14. M. Gündüz and E. Özsoy, "Modeling seasonal circulation and thermohaline structure of the Caspian Sea," Ocean Sci. **11,** 259−292 (2014). https://doi.org/10.5194/os-10-459-2014

15. N. A. Diansky, V. V. Fomin, T. Yu. Vyruchalkina, and A. V. Gusev, "Simulation of the Caspian Sea circulation with calculation of atmospheric forcing using the WRF model," Tr. Karel. Nauch. Tsentra RAN, No. 5, 21−34 (2016). https://doi.org/10.17076/lim310

16. S. K. Popov, A. L. Lobov, G. A. Monakhova, and K. I. Asaeva, "Use of a dynamic hydrometeorological model for the calculations of water exchange and pollutant transfer between different parts of the Caspian Sea," Trudy GOIN, No. 216, 340−357 (2015).

17. S. A. Lebedev and A. G. Kostianoy, "Interannual variability of water exchange anomalies between the Northern, Middle and Southern Caspian based on satellite altimetry data," Ecol. Montenegr. **25,** 106−115 (2019).

18. J. Shukla, *Seamless Prediction of Weather and Climate: A New Paradigm for Modeling and Prediction Research. US NOAA Climate Test Bed Joint Seminar Series* (NOAA, Camp Springs, MD, 2009).

19. B. Hoskins, "The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science,". Q. J. R. Meteorol. Soc. **139,** 573−584 (2013). https://doi.org/10.1002/qj.1991

20. G. S. Dyakonov and R. A. Ibrayev, "Reproduction of interannual variability of the Caspian Sea level in a high-resolution hydrodynamic model," Oceanology (Engl. Transl.), **58,** 8−18 (2018). https://doi.org/10.7868/S0030157418010021

21. G. S. Dyakonov and R. A. Ibraev, "Dynamics of Caspian Sea waters over the Apsheron Sill in 2003,"Phys. Oceanogr. V. 26 Iss. 6. P. 633−645 (2019). https://doi.org/10.22449/1573-160X-2019-6-557-568

22. V. S. Arkhipkin, A. L. Bondarenko, D. L. Vedev, and A. N. Kosarev, "Specific water circulation at the eastern coast of Central Caspian region," Vodn. Resur., No. 6, 36−43 (1992).

23. S. M. Griffies and R. W. Hallberg, "Biharmonic friction with a Smagorinsky-like viscosity for use in large-scale eddy-permitting ocean models," Mon. Weather Rev. **128,** 2935−2946 (2000). https://doi.org/10.1088/1742-6596/16/1/048

24. D. P. Dee et al., "The ERA-Interim reanalysis: configuration and performance of the data assimilation system," Q. J. R. Meteorol. Soc. **137,** 553−597 (2011). https://doi.org/10.1002/qj.828

25. M. Gündüz, B. Kara, A. Wallcraft, and J. Metzger, "Application of creeping sea-fill methodology to the wind speed over the Caspian Sea," Rapp. Comm. Int. Mer Medit. **38,** 155 (2007).

26. A. B. Kara, A. J. Wallcraft, and H. E. Hurlburt, "A correction for land contamination of atmospheric variables near land-sea boundaries," J. Phys. Oceanogr. **37,** 803−818 (2007). https://doi.org/10.1175/JPO2984.1

27. G. S. Dyakonov, R. A. Ibraev, and P. O. Shishkova, "Assessment of data quality of the ERA-Interim reanalysis in the Caspian Sea region," Russ. Meteorol. Hydrol. 2020 (in press).

28. V. V. Voevodin, S. A. Zhumatii, S. I. Sobolev, A. S. Antonov, P. A. Bryzgalov, D. A. Nikitenko, K. S. Stefanov, and V. V. Voevodin, "Practical use of *Lomonosov* supercomputer," Otkrytye sistemy, No. 7, 36−39 (2012).

Translated by V. Arutyunyan

Numerical Simulation of Ocean Circulation with Ultrahigh Spatial Resolution

G. K. Korotaev[a], * and V. P. Shutyaev[a, b], **

[a]Marine Hydrophysical Institute, Russian Academy of Sciences, Sevastopol, 299011 Russia
[b]Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, 119333 Russia
*e-mail: gkorotaev@gmail.com
**e-mail: shutyaev@inm.ras.ru

Abstract—This paper provides an overview of publications related to the study of submesoscale ocean variability. The significant progress achieved recently in understanding the special role of submesoscale processes in the formation of ocean circulation and ocean—atmosphere interaction, including modulation of the heat and gas exchange of the two media and their impact on the flux of biogenic elements into the top layer of the ocean, is emphasized. The necessity of conducting a study of submesoscale ocean variability in order to improve the quality of short-term marine forecasts is noted. Marine forecasts of the new level will be based on models with ultrahigh spatial resolution. The models will allow one to describe the reverse cascade of energy in the submesoscale range. In connection with the fact that new instruments capable of conducting 2—10 km surface observations of seas and oceans will be launched in the next five years, the possibilities of using the Black Sea as a testing area for the development of numerical models of circulation in a basin with a resolution of no less than a kilometer and improving methods of analysis and assimilation of observational technologies of the next generation are discussed.

Keywords: numerical simulation, ultrahigh resolution, submesoscale variability, data assimilation, Black Sea polygon

DOI: 10.1134/S000143382003010X

1. INTRODUCTION

The development of a numerical simulation of ocean circulation began right after computers appeared. Numerical methods for solving primitive equations describing the large-scale circulation of the World Ocean and its individual basins gained momentum as early as in the 1960s. At the initial stage, numerical models had a rather rough spatial resolution. At the same time, the outstanding features of the large-scale structure of ocean fields were successfully reproduced in numerical experiments. They include the main ocean thermocline, main ocean gyres, western boundary currents, abyssal circulation, and equatorial countercurrents [1].

In the early 1970s, revolutionary changes occurred in oceanology due to the discovery of the synoptic variability of the World Ocean [2]. It has been found that, along with large-scale currents, intense vortex structures and planetary waves with a typical scale determined by the internal radius of the Rossby circulation are observed in the ocean. The kinetic energy density of synoptic processes is significantly higher than for middle currents almost everywhere. Therefore, the current state of the marine environment is determined by synoptic processes and cannot be taken from climatic atlases.

Alongside the refinement of physical knowledge about ocean circulation, the performance of computing machinery has considerably improved. Its development made it possible to decrease grid steps when approximating primitive equations and resolve synoptic processes explicitly using the so-called eddy-resolving numerical models.

The passage from models with a rough spatial resolution to eddy-resolving made it possible to noticeably increase the practical significance of numerical simulation. Numerical calculations of ocean fields with a real atmospheric forcing reflect the observed state of the ocean rather well. This fact, as well as the rapid development of real-time oceanographic observations, formed the basis for the creation of systems of real-time marine forecasts. Present-day real-time marine forecasts are sufficiently accurate and satisfy the requirements of a great circle of customers [3].

The considerable increase in the accuracy of numerical calculations of the ocean state in passing to eddy-resolving models had a deep physical basis. The matter is that present-day methods of numerically

integrating primitive equations are based on a finite-dimensional approximation of partial differential equations. In passing to numerical analogs of differential equations, one has to parameterize part of the lost degrees of freedom related to the so-called subgrid scales. In doing this, as a rule, the simplest parameterization based on introducing coefficients of turbulent viscosity and diffusion is used. Such a parameterization turns out to be more or less satisfactory if energy is transferred from large scales to small ones. However, during the motion of a thin liquid layer on a fast-rotating sphere, a reverse energy flow from small to large scales is observed. Under these circumstances, traditional parameterizations of turbulent viscosity and diffusion turn out to be too rough. By virtue of this, numerical models that do not describe synoptic scales explicitly cannot reliably reproduce even the structure of large-scale currents.

Note, however, that the synoptic variability of the ocean, just like large-scale circulation, is described by primitive equations. For this reason, earlier developed numerical methods of integrating primitive equations in the implementation of eddy-resolving models were improved with allowance for the possibility of parallel computations. Note also the use of irregular horizontal grids and various vertical coordinates. However, subgrid scales in eddy-resolving models are still considered an intermediate range through which the energy is translated to even lesser scales, where it is finally dissipated. The high accuracy of marine forecasts for a period of up to 10 days is ensured to a great extent by the good initialization of the models by the assimilation of available satellite and contact real-time observations.

The further development of computing machinery resulted in the fact that numerical models of sea and ocean circulation with ultrahigh spatial resolution are increasingly widely used in the last 10−15 years [4−7]. Present-day computational clusters allow one to explicitly reproduce the submesoscale (i.e., having a spatial scale less than the internal Rossby deformation radius) variability of the marine environment in numerical calculations. Numerical experiments testify to the widespread occurrence of the submesoscale variability of ocean fields [8]. Satellite high spatial resolution observations of the ocean surface convincingly corroborate conclusions obtained based on the analysis of numerical calculations. They demonstrate the pervasiveness of oceanic structures whose spatial scales are considerably less than the Rossby deformation radius.

The most significant conclusion obtained based on numerical experiments is, however, the determination of the energy transfer from submesoscales upwards along the spectrum. Therefore, it is quite possible that inaccuracies of eddy-resolving models are at least partially caused by the incorrect parameterization of subgrid scales. Thus, more careful allowance for the influence of submesoscales in models with ultrahigh spatial resolution can improve the quality of real-time marine forecasts.

This work is devoted to the analysis of trends in the development of models with ultrahigh spatial resolution. In Section 2, current knowledge about the nature of submesoscale variability of ocean fields is summarized. In Section 3, the possibility of retrieving the three-dimensional structure of the synoptic and submesoscale variability by observations of the ocean surface is discussed. Section 4 presents a survey of observational systems which allow one to diagnose the submesoscale variability of the ocean and yield information for the initialization of prognostic ultrahigh spatial resolution models. In Section 5, principles of assimilation of observations that will become available in the next 5 years in ultrahigh resolution models of sea and ocean circulation are discussed. In Section 6, perspectives of studying the submesoscale variability over the area of the Black Sea are discussed. In conclusion, fundamental principles discussed in the paper are briefly summarized.

2. CHARACTERISTIC OF SUBMESOSCALE VARIABILITY OF THE OCEAN

Model numerical experiments and observations from artificial Earth satellites (AESs) convincingly demonstrate the considerable variability in the ocean on scales which are less than synoptic scales. In the open ocean, the characteristic scale of synoptic variability is determined by the Rossby deformation radius, the magnitude of which in middle latitudes ranges within 40−50 km. Therefore, 80−100 km can be treated as the lower boundary of the range of spatial scales on which the synoptic variability of the ocean is observed in middle latitudes. Since the Rossby deformation radius is inversely proportional to the Coriolis parameter, which equals zero at the equator and reaches the maximum value at the poles, its magnitude decreases when approaching the pole and increases along the direction to the equator.

The submesoscale variability includes processes with a characteristic scale of several tens of kilometers in middle latitudes. At the upper boundary, they are adjacent to the synoptic variability of the ocean; at the lower boundary, they are adjacent to internal waves. In time, periods of submesoscales exceed the inertial period.

On the ocean surface, submesoscale phenomena manifest themselves in the sea surface-temperature distribution constructed by data of IR scanners of high spatial resolution, in observations of the ocean color by scanning spectrophotometers of the visible range, and in measurements of the radio signal scattering by synthetic aperture radars. Numerical experiments demonstrate that submesoscale structures are also distinguished in deviations of the sea-level surface.

The main ideas about characteristics of submesoscale variability and mechanisms of its formation were

obtained based on numerical simulation [4, 8]. Numerical experiments demonstrate that submesoscale formations are concentrated in the upper mixed layer (UML) of the ocean [9]. Below the UML, their energy attenuates rather rapidly. This is the manifestation of the difference from the ocean synoptic variability, which covers the whole main pycnocline.

As for the cause of the formation of submesoscale structures, two hypotheses were put forward. The first [10] associates their formation with the action of synoptic vortices on the temperature field of the sea surface. Indeed, let the UML temperature vary initially along one of directions (e.g., from the south to the north due to the nonuniformity of heating). Currents induced by synoptic vortices can create a deformation field; as a result, narrow frontal zones appear on the background of initially even a slow change in the ocean surface temperature. Further, the instability of these fronts will lead to the appearance of various structures typical for two-dimensional turbulence.

The second hypothesis relates the formation of submesoscale phenomena with UML instability [11]. Indeed, in the case of a change in the UML depth in the horizontal direction, there appears a slope of its lower boundary. The slope of the UML boundary causes the accumulation of available potential energy, which is released under certain conditions with the development of baroclinic instability. The scale of structures appearing with the development of baroclinic instability is determined by the magnitude of the Rossby radius calculated based on the thickness of the UML and temperature difference at its lower boundary. Certainly, the magnitude of the Rossby radius calculated in this way is less than the first baroclinic mode Rossby radius determining the scale of the synoptic variability. That is why the size of submesoscale formations is less than that of synoptic vortices.

The second mechanism of the formation of submesoscale structures turns out to be more corresponding to observations. Indeed, numerical calculations [12, 13], as well as the analysis of observations of flow velocities, temperature, and salinity of sea water [14], point to an increase in the intensity of submesoscale structures in middle latitudes in winter. Numerical experiments demonstrate that such behavior is related to more favorable conditions for the development of baroclinic instability in winter, when the UML deepening occurs. Indeed, an increase in the UML depth is accompanied by an increase in available potential energy, which can be released as a result of baroclinic instability. In addition, the Rossby deformation radius calculated based on the UML depth is less in summer and, correspondingly, vortices of a lesser scale will be generated due to baroclinic instability. The vortices can be suppressed by the atmospheric action [15].

In one way or other, both abovementioned mechanisms lead to the development of turbulence in the upper mixed layer. Since the vertical scale of submeso-scale structures is significantly less than the horizontal one, the energy cascade in this range of scales is adequate to regularities of quasi-geostrophic two-dimensional turbulence [16]. Rather sharp fronts are an integral part of submesoscale variability. They are associated with one or another type of instability and formation of submesoscale vortices. According to the general regularity of two-dimensional turbulence, small-scale vortices are united into larger ones and provide the energy transfer upwards on the scales.

According to Rhines' theory [17], nonlinear interactions cause the energy transfer to a scale corresponding to the Rossby radius determined by UML parameters. In the neighborhood of this scale, energy must be transferred to a lower motion mode, i.e., to synoptic scales, which is corroborated in part by numerical calculations [9]. Thus, submesoscales play a significant part in the energy redistribution in upper layers of the ocean. By virtue of the energy transfer upwards over the spectrum to larger scales, the development of submesoscale variability leads to the appearance of so-called negative viscosity; i.e., their parameterization by viscosity and diffusion operators is not adequate. In models of ocean circulation, generally speaking, one should considerably reduce the grid step to resolve submesoscale variability explicitly. Note, however, that the energy in the shortwave part of the spectrum in the submesoscale range is transferred to small scales [18]—probably, in view of their ageostrophicity and break of regularities peculiar to geostrophic turbulence.

Numerical experiments also demonstrate one more feature of submesoscale processes. When they are developed at the lower boundary of a homogeneous layer, there appear considerable vertical velocities favoring the inflow of colder waters to the UML and having a significant effect on the heat balance of the upper layer of the ocean [18]. In [18], it was also noted that the heat exchange between the ocean and atmosphere can considerably change when taking into account submesoscales. Therefore, submesoscale variability of the ocean must be taken into account also in climatic models.

Large vertical velocities caused by submesoscale variability also favor the inflow of biogenic elements into the UML [19]. The inflow of biogens favors local phytoplankton bloom and the reflection of submesoscale structures in satellite observations of the ocean color.

3. RETRIEVAL OF THE THREE-DIMENSIONAL STRUCTURE OF SYNOPTIC AND SUBMESOSCALE VARIABILITY BY OBSERVATIONS OF THE OCEAN SURFACE

In a series of works beginning from [10], the possibility of retrieving the vertical structure of variability of density and current velocity fields on synoptic scales

and submesoscales by observations on the ocean surface was investigated. In those works, it was assumed that submesoscale formations, as well as synoptic ones, could be described within the framework of the quasi-geostrophic model.

Let us briefly recall the main relationships of the quasi-geostrophic model in a stratified ocean. First, it is assumed that horizontal components of current velocities are approximately in geostrophic balance. This allows one to introduce the current function so that the zonal meridional components of the current velocity are respectively specified by the expressions

$$u = -\frac{\partial \psi}{\partial y}, \quad v = \frac{\partial \psi}{\partial x}. \tag{1}$$

Relationships (1) are derived taking into account that Coriolis parameter f slightly varies on synoptic scales and submesoscales. By virtue of geostrophic relationships, pressure P, normalized to mean density, is expressed in terms of the current function as

$$P = f\psi.$$

Using the hydrostatic equation, we find that the deviation of sea-water density ϱ from its local mean profile $\varrho_0(z)$ normalized to the mean density is related to the current function by the relationship

$$g\varrho = f\frac{\partial \psi}{\partial z}.$$

Introducing the Brunt−Väisälä frequency $N^2 = g\frac{d\varrho_0}{dz}$, we find the expression for the potential vortex ξ in the quasi-geostrophic approximation:

$$\xi = \Delta\psi + \frac{\partial}{\partial z}\left(\frac{f^2}{N^2}\frac{\partial \psi}{\partial z}\right).$$

Let us now know the distribution of potential vorticity in a certain part of the ocean $\xi_0(x, y, z)$. We also assume that the density distribution of the normalized anomaly of the sea water density $\varrho(x, y)$ on its surface or the deviation of the level surface, i.e., $\psi_0(x, y)$, is known.

Then, solving the equation

$$\Delta\psi + \frac{\partial}{\partial z}\left(\frac{f^2}{N^2}\frac{\partial \psi}{\partial z}\right) = \xi_0(x, y, z)$$

with the boundary condition on the ocean surface $f\frac{\partial \psi}{\partial z} = g\varrho(x, y)$ or $\psi = \psi_0(x, y)$ and setting the zero density anomaly on the ocean bottom, as well as, from additional considerations, conditions at lateral boundaries of the region, we find the distribution of the current function in whole integration domain.

The question about the retrieval of the potential vorticity distribution over the depth is the most significant. This problem is solved most simply in the case

of frontogenesis development in the upper sea layer. In this process, one can observe considerable gradients of sea water temperature and salinity and, therefore, sea water density gradients on the sea surface. After distinguishing the large-scale mean value and anomaly caused by synoptic and submesoscale variability in the equation of potential vortex conservation and a series of approximate transformations, it becomes possible to express $\xi_0(x, y, z)$ in terms of the density anomaly on the sea surface:

$$\xi_0(x, y, z) = \alpha(z)\frac{\partial \psi}{\partial z}\bigg|_{z=0},$$

where $\alpha(z)$ is expressed in terms of derivatives of the mean value of the potential vortex and mean density. Then, calculating the current function, one can construct the distribution of relative vorticity and vertical flow velocity at any horizon. Numerical calculations of sea dynamics on the basis of primitive equations demonstrate [10] that the relative vorticity and vertical velocity fields retrieved by the technique described above correspond very well to their exact values.

However, density anomalies on the sea surface are insignificant if the UML instability is the main energy source in exciting the submesoscale variability. Within this context, it was proposed in [9] to determine the distribution $\xi_0(x, y, z)$ over the vertical by use of expansions in eigenfunctions of the operator $\frac{\partial}{\partial z}\left(\frac{f^2}{N^2}\frac{\partial}{\partial z}\right)$. The expansion coefficients depend not only on the level, but also on the scale of the process. The transfer function on scales larger than $50-60$ km has significant values up to a depth of about 800 m. With an increase in the wavenumber, it is concentrated within limits of about 400 m. The solution of the potential vortex equation with a given level surface and retrieved distribution $\xi_0(x, y, z)$ allows one to calculate the relative vorticity distribution and vertical velocity. As is shown in [9], the accuracy of fields retrieved in this way is sufficiently high.

4. PROMISING METHODS FOR OBSERVATIONS OF SUBMESOSCALE PROCESSES

Increasing the resolution of ultrahigh spatial resolution numerical models for preparing real-time marine forecasts evidently requires an adequate reorganization of the observation system. At present, the only real possibility to observe submesoscales on a considerable part of the ocean is based on carrying out observations from AESs. The most convenient fields for observations of surface manifestations of submesoscales on the ocean surface are the sea surface temperature (SST) and sea level. The difficulty of using satellite observations of the SST is related to the influence of cloudiness. Outstandingly useful information about submesoscale processes is given by scanners of

the IR and visible ranges of the electromagnetic wave spectrum. However, the influence of cloudiness significantly diminishes the possibility of constructing regular observations of submesoscales based only on such measurements. In recent times, observations of the SST are also performed from geostationary AESs with a high frequency in time and with all the higher spatial resolution. In the long run, this will make it possible to observe manifestations of submesoscales in the SST distribution on considerable areas of the ocean surface and use them for studying submesoscale variability.

Submesoscale variability can be also effectively monitored by observations of the sea surface topography. However, traditional altimeters, which turned out to be very useful for observations of synoptic variability of the ocean, do not provide the necessary spatial resolution and coverage of the ocean surface. In connection with this, it is planned to implement in the short run a new approach to observations of the ocean topography. The approach is based on the interferometry technique [20]. Novel wide-swath altimeters will carry out measurements in a wider range and, according to calculations, must provide a resolution of processes in a wavelength range from 15 to 100 km.

The Surface Water and Ocean Topography (SWOT) Mission [20] is discussed in most detail. The wide-swath SWOT altimeter will measure the sea surface topography in two 50-km-wide bands separated by a 20-km band where observations are absent. In addition, a traditional altimeter, which will measure the sea surface topography to nadir, is planned to be mounted on the satellite. The bands along which the observations are planned will be exactly repeated every 21 days. However, in view of their partial overlapping in different regions of the ocean, a sufficiently complete coverage of the sea surface can be expected for 10 days.

Another similar mission, COMPIRA [21], is planned by the Japan Aerospace Exploration Agency. It will measure the ocean surface topography in two 80-km-wide bands to the left and right from the direction of the AES flight. The declared spatial resolution is 5 km. In the middle, in a 20-km-wide band, observations are not carried out, like in the SWOT program. The AES orbits will be repeated every 10 days. In contrast to the SWOT program, which provides the global coverage of the World Ocean, observations of the COMPIRA mission will be carried out only on the water area of the Pacific Ocean and the northern half of the Indian Ocean.

One more project oriented toward observations of the submesoscale dynamics from AESs, SEASTAR, is based on using radar observations for retrieving the velocity of surface flows with a resolution of about 1 km [22]. It is planned to carry out observations in a band with a width of several hundred kilometers with the initial resolution of several 10 m. Then, as a result of processing, the velocity field of surface flows will be constructed with an accuracy of 10 cm/s in the absolute value and about 20° in direction. Note also that useful information about submesoscale processes in the upper layer of the ocean is also presented by traditional synthetic aperture radars.

In carrying out the plans to improve satellite observation systems, substantial progress will be achieved in the detailed elaboration of the role of submesoscale processes. Certainly, each of the systems presented above has disadvantages. Generally speaking, the spatial resolution of wide-swath altimetry is insufficient. Radiolocation makes it possible to measure the velocity of surface currents with a relatively low accuracy, etc. By virtue of this, the most considerable effect from promising new satellite missions will undoubtedly be achieved due to the formation of an integrated database and complex analysis of the observations listed above. In this respect, the assimilation of new-level observations in ocean circulation models with ultra-high spatial resolution will allow one to compensate for the disadvantages of individual observation systems. Correspondingly, one should expect an increase in the accuracy of real-time marine forecasts due to the assimilation of new observation data flows in numerical models.

5. PROBLEMS OF OBSERVATION ASSIMILATION WHEN RESOLVING SUBMESOSCALES

New possibilities of carrying out calculations of sea and ocean circulation with an explicit resolution of submesoscales must potentially improve the quality of marine forecasts. Recently, at the same time, there have been investigations demonstrating that an increase in the spatial resolution of numerical models itself not only does not improve forecast quality, but even decreases it [23]. In [23], the results of forecasts by the circulation model in the Tasman Sea with a resolution of 2.5 and 10 km were compared. In doing this, both models were initialized based on currently existing satellite and contact observations in a common manner. It turned out that an increase in the model resolution led to the degradation of forecast accuracy. The authors of the investigation associated this fact with the unsatisfactory reproduction of the energy flow from submesoscales upward the spectrum in the higher resolution model. It seems natural that the spatial resolution of the model must be increased while also taking into account observations that make it possible to determine the position of submesoscale structures in the space more exactly, which could reliably reproduce energy flows over the scale spectrum with a certain accuracy. The promising satellite missions listed in the previous section make it possible to obtain a considerably higher spatial resolution of observations. Therefore, if they are assimilated in ocean circulation models, one can expect an improvement of the accuracy of marine forecasts.

At the first stage, it is possible to use simple approaches based on the Kalman filter and conceptually close to the four-dimensional Gandin analysis [24]. They showed good results in the assimilation of observations in eddy-resolving models. However, the best accuracy is yielded by new generations of assimilation schemes based on four-dimensional variational data assimilation (4D-Var) and ensemble Kalman filter (EnKF) [25—28].

In observation assimilation by the EnKF method, as in the classical Kalman filter, the key role is played by the covariance matrix. This matrix is calculated based on the Monte Carlo method at each time step by averaging over the finite number of model trajectories calculated with random initial conditions. By using it, the ensemble Kalman filter assimilates observations subsequently at each specific time instant.

Variational assimilation is based on the idea of minimizing a certain functional related to observation data on trajectories (solutions) of the considered model. Therefore, the problem of data assimilation is stated as an optimal control problem. Theoretical foundations of studying and solving such problems were laid in the classical works [29, 30] and others. In solving minimization problems, it becomes necessary to calculate the gradient of the initial functional. Using the theory of conjugate equations [29, 30] was an important step in this direction. A lot of research has widely applied the equations for the study and numerical solution of problems of data assimilation, beginning with well-known works [31—34]. In the 4D-Var method, in contrast to the EnKF method, where observations are assimilated subsequently, the model solution is optimized immediately on the whole time interval in the assimilation window.

The 4D-Var system, which is based on [33], was used for the first time in the European Center for Medium-Range Weather Forecasts. Another approach to the variational assimilation of observations was developed and justified by researchers of the Institute of Numerical Mathematics, Russian Academy of Sciences. One characteristic feature of the proposed algorithms is the use of splitting with respect to physical processes and geometric coordinates, which made it possible to simplify the considered problems at each splitting step and ensure efficiency in implementing variational assimilation algorithms [35, 36].

It is well known that 4D-Var methods and a Kalman filter in the case of the linear model, linear operator, and Gaussian observation errors yield identical results in the end of the assimilation window if the model errors are not taken into account. Further, under the same assumptions about linearity and Gaussianity at a sufficiently large number of elements of the ensemble, the EnKF method well approximates the Kalman filter [37]. Nonlinearities of the model and observation operator (and, as a consequence, non-Gaussianity of errors) are the eventual cause of the discrepancy of results when using the 4D-Var and EnKF [38]. When errors of observations and the initial approximation (background) remain Gaussian and the dynamics model is nonlinear, the 4D-Var method yields a maximum likelihood estimator—the mode of the distribution function of the conditional posterior probability. At the same time, in the general case, it is not clear how finding such mode is related to the result of the EnKF method [38].

In the problem of describing the ocean dynamics with a submesoscale resolution, the dimension of the system state vector is so large that it is necessary to seek an accommodation between computational capabilities and theoretically optimal approaches. For example, the EnKF method has sampling errors due to the restricted size of the ensemble; in the 4D-Var method, due to the large dimension, one has to search for approximations of covariance matrices of the background, which also leads to errors which are difficult to estimate in a comparative analysis of the methods.

At large dimensions of the state vector, the work with covariance matrices in the EnKF method becomes a serious computational problem. Using a bounded number of ensemble elements deteriorates the approximation of the Kalman filter. At the same time, when using iterative gradient methods in the 4D-Var method, one has to construct and solve linearized direct and adjoint problems, which is often a big problem for complex models of the ocean dynamics [39].

The performed numerical comparisons of the 4D-Var and EnKF [38, 40, 41] demonstrate that the EnKF yields more accurate results for small time intervals. For observations with data gaps, when ensemble perturbations grow linearly and become non-Gaussian, the 4D-Var leads to lesser errors than the EnKF [40]. It should be noted that the EnKF turns out to be more preferable from the viewpoint of algorithm parallelization, because the computations can be performed independently for each member of the ensemble [41].

The synthesis of a broad discussion about the comparison of the 4D-Var and EnKF methods [38, 40, 41] resulted in the acceptance of the necessity to develop new approaches to data assimilation. These approaches should combine the 4D-Var and EnKF methods and contain their best advantages [42]. This was how the hybrid approach combining the ensemble Kalman method and variational data assimilation—Hybrid 4DVar [42, 43], as well as the ensemble method of four-dimensional variational data assimilation—4DEnVar—appeared [44, 45].

Present-day data-assimilation methods are already used in investigations based on the twin method for estimating the potential contribution of wide-swath altimetric observations [46]. The results of the calculations turn out to be promising; however, in the process of investigations, new problems were identified. They are related to features of satellite observations whose errors
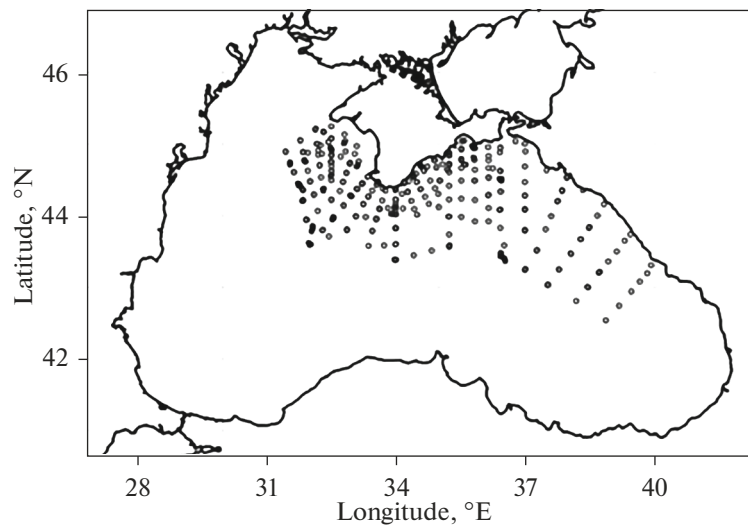
Fig. 1. Hydrological observations in the economic zone of the Russian Federation in the Black Sea.

now cannot be treated as noncorrelated [47] and to the overlapping of internal waves on the considered scale interval [48].

There is the necessity of further intense investigations of submesoscale processes based on numerical simulation with ultrahigh spatial resolution and the use of present-day methods of data assimilation to create a basis for marine forecast systems at a radically new level.

6. THE BLACK SEA AS A TEST AREA FOR INVESTIGATING SUBMESOSCALE VARIABILITY

International cooperation working out the development strategy for the SWOT program proposes, first and foremost, performing detailed investigations (including numerical simulation) on small test areas. In Russia, such investigations can be carried out in the Black Sea. The economic zone of the Russian Federation between Sochi and Sevastopol combines a wide network of interdisciplinary regular observations of the Institute of Oceanology near Gelendzhik and the Marine Hydrophysical Institute near Katsiveli. Regular complex shipboard surveys are carried out every year on the whole abovementioned water area (Fig. 1). Moreover, daily forecasts and analyses of the state of the marine environment are given for this region by the Marine Forecast Center of the Marine Hydrophysical Institute, and the systematic receipt and processing of all available observations from AESs are carried out. According to observations from AESs, the existence of a developed submesoscale variability in the Black Sea has been shown in a series of papers [49−51]. Note that submesoscale vortices identified by satellite observations are observed both along the coast and in the open sea [50].

The Black Sea is also a very convenient basin for carrying out numerical calculations with ultrahigh spatial resolution [7, 52], because it is almost closed and has a relatively simple configuration. Calculations with high spatial resolution demonstrate that high variability is observed in the Black Sea in a wide scale range (Fig. 2). This figure presents the topography of the sea surface level for September 4, 2008. It was obtained in calculations by the NEMO model with a grid step of 1.157 km. In Fig. 2, synoptic structures are easily seen both in the coastal and in the open part of the sea. Variability of the Black Sea in the economic zone of the Russian Federation is presented in more detail in Fig. 3. In this figure, the map of the potential vorticity distribution in the end of February 2008 at a depth of 4 m in the northeastern part of the Black Sea is constructed. In the vorticity distribution, thin particularities of the jet of the Black Sea Rim Current (BSRC), vortices with a scale of less than 10 km, mushroom-shaped structures, and filaments are discerned. Thus, the northeastern part of the Black Sea is a suitable object for studying submesoscale variability.

Carrying out calculations with the assimilation of a new generation of satellite observations must be an important field of the investigations. Researchers at the Institute of Numerical Mathematics developed and justified algorithms for the numerical solution of problems of the variational assimilation of satellite observation data on temperature and ocean surface level, as well as data on temperature and salinity from the Argo buoy system, with the use of the global three-dimensional ocean hydrothermodynamics model developed at the institute [35, 36, 53]. One characteristic feature of the proposed algorithms is the use of splitting with respect to physical processes and geometric coordinates, which made it possible to simplify the considered problems at each splitting step and
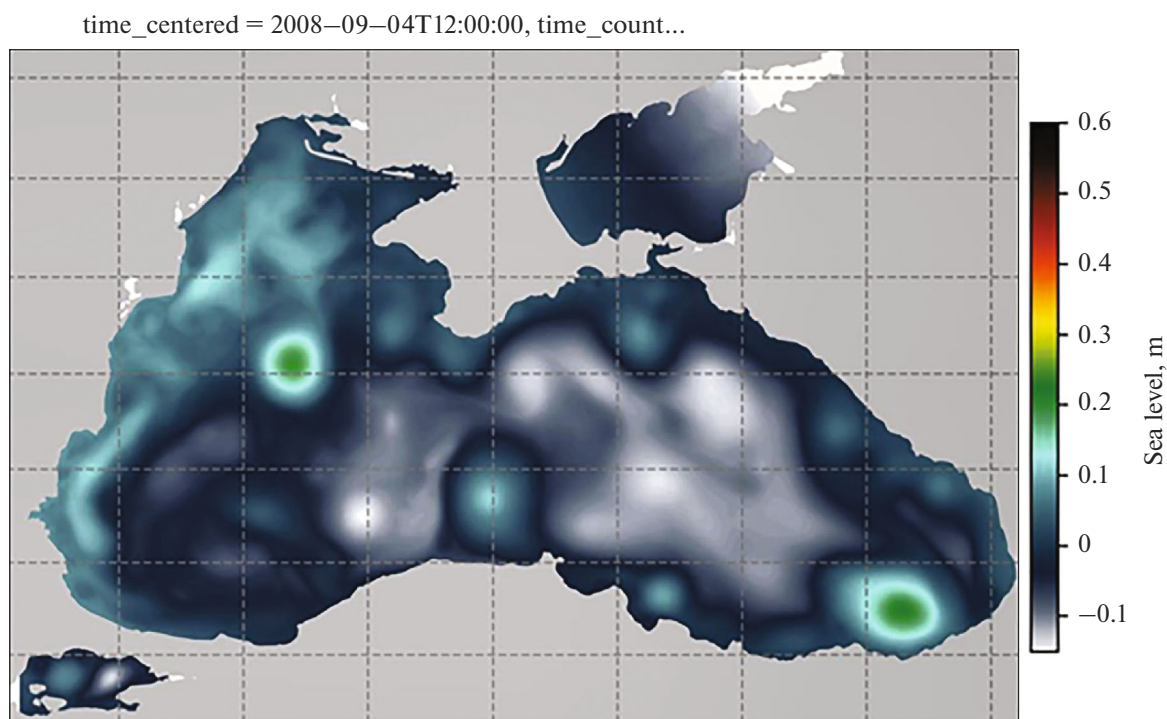
time_centered = 2008−09−04T12:00:00, time_count...

Fig. 2. Surface-level map of the Black Sea in September 2008 according to results of numerical simulation carried out by A.I. Mizyuk.
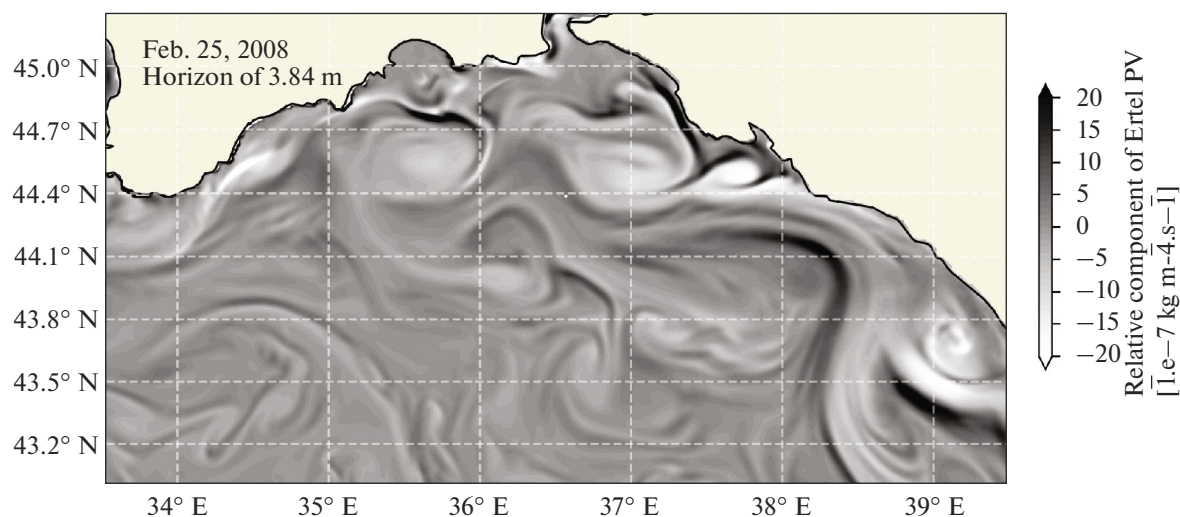

Fig. 3. Distribution of the potential vorticity in the northeastern part of the Black Sea in the end of February 2008 according to the results of numerical simulation carried out by A.I. Mizyuk.

ensure efficiency in the implementation of variational assimilation algorithms. The variational assimilation algorithms have been already carried to implementation in the model of the Black Sea dynamics [54].

When carrying out investigations in the Black Sea, it is also possible to use the 4D-var assimilation algorithm, which is one of the components of the NEMO system.

Methods of variational assimilation, however, are computer-intensive. In this respect, it also seems attractive to develop more efficient assimilation schemes. In particular, one can consider using the quasi-geostrophic approximation when assimilating observations of synoptic and submesoscale processes. The above-discussed methodology of retrieving the vertical structure of density and flow velocity fields on

synoptic scales and submesoscales by observations on the ocean surface allows one to look at the problem of assimilating observations of the ocean surface in a new light. Based on [9, 10], we assume that forecast errors can be described using the quasi-geostrophic approximation. In this case, one can describe all residuals of the forecast using the current function satisfying the equation

$$\Delta\psi + \frac{\partial}{\partial z}\left(\frac{f^2}{N^2}\frac{\partial\psi}{\partial z}\right) = \delta\xi_0\,(x, y, z). \qquad (2)$$

The right-hand side of the equation now contains the unknown forecast error of the potential vortex. Using the algorithm from [9], one can express $\delta\xi_0\,(x, y, z)$ approximately in terms of the known residual of the sea-level forecast. Then, solving Eq. (2) with the current function on the sea surface in proportion to the residual of the level forecast and the condition of its attenuation with the distance from the surface, we first find residuals for the flow velocities and sea-water density (and, in terms of the T−S relationship, for temperature and salinity). Further, using these residuals, we correct all fields calculated by the model and obtain new initial conditions for the subsequent forecast. In general, the assimilation method constructed based on the quasi-geostrophic approximation and methodology of [9, 10] can turn out to be attractive due to its efficiency. However, it is associated with certain difficulties related to the fact that the transfer function obtained in [9] for the calculation of the potential vortex anomaly by observations of the ocean surface topography depends on process scales, which requires additional investigations.

It is also expedient to estimate the possibilities of the traditional four-dimensional analysis when assimilating observations in models with ultrahigh spatial resolution [55]. This method is successfully used when assimilating observations in eddy-resolving models [56], which is explained by the contribution of the low mode to the formation of synoptic variability [57, 58]. Since the considerable part of submesoscale variability can be described by the quasigeostrophic approximation (see Section 3), the vertical structure of motions related to the UML can be considered based on the analysis of eigenfunctions of operator (2). Taking into account that the Brunt−Väisälä frequency in the mixed layer is significantly less than in the main pycnocline, one can distinguish a family of eigenfunctions whose amplitude is small beyond the UML. One can assume that, as for synoptic scales, the energy of submesoscale variability related to the mixed layer is concentrated in the mode which maintains the sign within the UML. Based on this, one can construct an efficient scheme for the assimilation of wide-swath altimetry data; the scheme is similar to that described in [55] and corrects submesoscales.

7. CONCLUSIONS

There has recently been substantial progress in the development of oceanology related to understanding the special role of submesoscale processes in the formation of the ocean circulation and the interaction between the ocean and atmosphere, including the modulation of heat and gas exchange of two media, as well as their impact on the flux of biogenic elements into the top layer of the ocean. According to currently existing understanding, the exact inclusion of submesoscale processes can change estimates of flows at the ocean−atmosphere interface almost by 20%, which can require a considerable modification of the knowledge about the role of the ocean in present-day climatic changes. Investigations of submesoscale variability are also necessary to increase the quality of short-term marine forecasts. The latest analyses of temporal and spatial spectra in the submesoscale interval and numerous numerical calculations with ultrahigh spatial resolution demonstrate that this type of variability is associated with the inverse energy cascade, not allowing one to parameterize them in terms of turbulent viscosity and diffusion. Therefore, a further increase in the accuracy of marine forecasts (and, correspondingly, medium-term weather forecasts) is related to a detailed investigation of submesoscale variability. In the next five years, several space agencies are planning to launch new devices capable of carrying out area observations with a resolution of 2−10 km. To develop methods of analysis and assimilation of these observations, it is necessary to carry out special experiments under well-monitored conditions. The Black Sea basin, with its developed observational base and convenient configuration for numerical simulation, can become a test area for studying submesoscales and developing methods of marine forecasts based on using observational technologies of the next generation.

REFERENCES

1. G. I. Marchuk, V. P. Dymnikov, and V. B. Zalesny, *Mathematical Models in Geophysical Hydrodynamics and Numerical Methods for Their Implementation* (Gidrometeoizdat, Leningrad, 1987) [in Russian].

2. L. M. Brekhovskikh, G. N. Ivanov-Frantskevich, M. N. Koshlyakov, K. N. Fedorov, L. M. Fomin, and A. D. Yampol'skii "Polygon hydrophysical experiment

in the tropic zone of the Atlantic," Izv. Akad. Nauk SSSR, Fiz. Atmos. Okeana, No. 6, 1434−1437 (1971).

3. *Ocean Weather Forecasting,* Ed. by E. P. Chassignet and J. Verron, (Springer, Dordrecht, 2006).

4. E. P. Chassignet and X. Xu, "Impact of horizontal resolution (1/12° to 1/50°) on Gulf Stream separation, penetration, and variability," J. Phys. Oceanogr., No. 47, 1999−2021 (2017).

5. A. V. Gusev, V. B. Zalesny, and V. V. Fomin, "Technique for simulation of Black Sea circulation with increased Rresolution in the area of the IO RAS Polygon," Oceanology (Moscow, Russ. Fed.) **57,** 880−891 (2017).

6. N. A. Diansky, V. V. Fomin, N. V. Zhokhova, and A. N. Korshenko, "Simulations of currents and pollution transport in the coastal waters of Big Sochi," Izv., Atmos. Ocean. Phys. **49,** 611−621 (2013).

7. A. I. Mizyuk, M. V. Senderov, G. K. Korotaev, and A. S. Sarkisyan, "Features of the horizontal variability of the sea surface temperature in the Western Black Sea from high resolution modeling," Izv., Atmos. Ocean. Phys. **52,** 570−578 (2016).

8. Z. Su, J. Wang, P. Klein, A. F. Thompson, and D. Menemenlis, "Ocean submesoscales as a key component of the global heat budget," Nat. Commun., No. 9, 775 (2018).

9. A. L. Ponte and P. Klein, "Reconstruction of the upper ocean 3D dynamics from high-resolution sea surface height," Ocean Dynamics **63,** 777−791 (2013).

10. G. Lapeyre and P. Klein, "Dynamics of the upper oceanic layers in terms of surface quasigeostrophy theory," J. Phys. Oceanogr. **36,** 165−176 (2006).

11. G. Boccaletti, R. Ferrari, and B. Fox-Kemper, "Mixed layer instabilities and restratification," J. Phys. Oceanogr. **37,** 2228−2250 (2007).

12. X. Capet, J. C. McWilliams, M. J. Mokemaker, and A. F. Shchepetkin, "Mesoscale to submesoscale transition in the California current system. Part I: flow structure, eddy flux, and observational tests," J. Phys. Oceanogr. **38,** 29−43 (2008).

13. J. A. Mensa, Z. Garraffo, A. Griffa, T. M. Özgökmen, A. Haza, and M. Veneziani, "Seasonality of the submesoscale dynamics in the Gulf Stream region," Ocean Dynamics **63,** 923−941 (2013).

14. J. Callies, R. Ferrari, J. M. Klymak, and J. Gula, "Seasonality in submesoscale turbulence," Nat. Commun, No. 6, 6862 (2015).

15. J. Marshall and F. Schott, "Open-ocean convection: observations, theory, and models," Rev. Geophys., No. 37 (1), 1−64 (1999).

16. J. C. McWilliams, "Submesoscale currents in the ocean," Proc. R. Soc. London, Ser. A **472,** 2016017 (2016).

17. P. B. Rhines, "Geostrophic turbulence," Annu. Rev. Fluid Mech., No. 11, 401−441 (1979).

18. P. Klein, G. Lapeyre, L. Siegelman, B. Qiu, L.-L. Fu, H. Torres, et al., "Ocean-scale interactions from space," Earth Space Sci., No. 6, 795−817 (2019).

19. D. J. McGillicuddy Jr., "Mechanisms of physical-biological-biogeochemical interaction at the oceanic mesoscale," Annu. Rev. Mar. Sci., No. 8, 125−159 (2016).

20. R. Morrow, L. Fu, F. Ardhuin, M. Benkiran, B. Chapron, E. Cosme, F. d'Ovidio, J. T. Farrar, S. T. Gille, G. Lapeyre, P. Le Traon, A. Pascual, A. Ponte, B. Qiu, N. Rascle, C. Ubelmann, J. Wang, and E. D. Zaron, "Global observations of fine-scale ocean surface topography with the Surface Water and Ocean Topography (SWOT) Mission," Front. Mar. Sci. **6,** 232 (2019). https://doi.org/10.3389/fmars.2019.00232

21. A. Uematsu, R. Nakamura, Y. Nakajima, and Y. Yajima, in *2013 IEEE International Geoscience and Remote Sensing Symposium−IGARSS* (Melbourne, 2013), pp. 2943−2946.

22. C. Gommenginger, B. Chapron, A. Hogg, C. Buckingham, B. Fox-Kemper, L. Eriksson, F. Soulat, C. Ubelmann, F. Ocampo-Torres, B. Buongiorno Nardelli, D. Griffin, P. Lopez-Dekker, P. Knudsen, O. Andersen, L. Stenseng, N. Stapleton, W. Perrie, N. Violante-Carvalho, J. Schulz-Stellenfleth, D. Woolf, J. Isern-Fontanet, F. Ardhuin, P. Klein, A. Mouche, A. Pascual, X. Capet, D. Hauser, A. Stoffelen, R. Morrow, L. Aouf, Ø. Breivik, L.-L. Fu, J. A. Johannessen, Ye. Aksenov, L. Bricheno, J. Hirschi, A. C. H. Martin, A. P. Martin, G. Nurser, J. Polton, J. Wolf, H. Johnsen, A. Soloviev, G. A. Jacobs, F. Collard, S. Groom, V. Kudryavtsev, J. Wilkin, V. Navarro, A. Babanin, M. Martin, J. Siddorn, A. Saulter, T. Rippet, B. Emery, N. Maximenko, R. Romeiser, H. Graber, A. Alvera Azcarate, C. W. Hughes, D. Vandemark, J. da Silva, P. J. Van Leeuwen, A. Naveira-Garabato, J. Gemmrich, A. Mahadevan, J. Marquez, Y. Munro, S. Doody, and G. Burbidge, "SEASTAR: mission to study ocean submesoscale dynamics and small-scale atmosphere-ocean processes in coastal, shelf and polar seas," Front. Mar. Sci. **6,** 457 (2019).

23. A. Sandery and P. Sakov, "Ocean forecasting of mesoscale features can deteriorate by increasing model resolution towards the submesoscale," Nat. Commun., No. 8, 1566 (2017).

24. V. V. Knysh, G. K. Korotaev, A. I. Mizyuk, and A. S. Sarkisyan, "Assimilation of hydrological observation data for calculating currents in seas and oceans," Izv., Atmos. Ocean. Phys. **48,** 57−73 (2012).

25. M. Asch, M. Bocquet, and M. Nodet, *Data Assimilation: Methods, Algorithms, and Applications* (SIAM, Philadelphia, 2016).

26. S. J. Fletcher, *Data Assimilation for the Geosciences: from Theory to Application* (Elsevier, Amsterdam, 2017).

27. A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen, "Data assimilation in the geosciences: an overview of methods, issues, and perspectives," WIREs Clim. Change **9** (5), e535 (2018).

28. V. P. Shutyaev, "Methods for observation data assimilation in problems of physics of atmosphere and ocean," Izv., Atmos. Ocean. Phys. **55,** 17−31 (2019).

29. G. I. Marchuk, "On the formulation of certain inverse problems," Dokl. Akad. Nauk SSSR **156** (3), 503−506 (1964).

30. J. L. Lions, *Contrôle optimal des systèmes gouvernés par des équations aux dérivées partielles* (Dunod, Paris, 1968).

31. V. V. Penenko and N. V. Obraztsov, "A variational initialization method for the fields of the meteorological elements," Meteorol. Gidrol. **11,** 1−11 (1976).

32. G. I. Marchuk and V. V. Penenko, in *Modelling and Optimization of Complex Systems. Proc. of the IFIP-TC7 Conf.* (Springer, New York, 1978), pp. 240−242.

33. F. X. Le Dimet and O. Talagrand, "Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects," Tellus A **38,** 97−110 (1986).

34. J. Lewis and J. Derber, "The use of adjoint equations to solve a variational adjustment problem with advective constraints," Tellus A **37,** 309−322 (1985).

35. V. I. Agoshkov, V. M. Ipatova, V. B. Zalesny, E. I. Parmuzin, and V. P. Shutyaev, "Problems of variational assimilation of observational data for ocean general circulation models and methods for their solution," Izv., Atmos. Ocean. Phys. **46** (6), 677−712 (2010).

36. V. B. Zalesny, V. I. Agoshkov, V. P. Shutyaev, F. Le Dimet, and V.O. Ivchenko, "Numerical modeling of ocean hydrodynamics with variational assimilation of observational data," Izv., Atmos. Ocean. Phys. **52,** 431−442 (2016).

37. G. Evensen, "The ensemble Kalman filter: theoretical formulation and practical implementation," Ocean Dynamics **53** (4), 343−367 (2003).

38. A. Caya, J. Sun, and C. Snyder, "A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation," Mon. Weather Rev. **133** (11), 3081−3094 (2005).

39. K. Mogensen, M. A. Balmaseda, A. T. Weaver, M. Martin, and A. Vidard, *NEMOVAR: a Variational Data Assimilation System for the NEMO Ocean Model, ECMWF Technical Memorandum,* No. 120 (2009).

40. E. Kalnay, H. Li, T. Miyoshi, S. -C. Yang, and J. Ballabrera-Poy, "4D-Var or ensemble Kalman filter?" Tellus A **59,** 758−773 (2007).

41. D. Fairbairn, S. R. Pring, A. C. Lorenc, and I. Roulstone, "A comparison of 4DVar with ensemble data assimilation methods," Q. J. R. Meteorol. Soc. **140,** 281−294 (2014).

42. N. Gustafsson, J. Bojarova, and O. Vignes, "A variational ensemble data assimilation for the HIgh Resolution Limited Area Model (HIRLAM)," Nonlin. Proc. Geophys. **21,** 303−323 (2014).

43. A. M. Clayton, A. C. Lorenc, and D. M. Barker, "Operational implementation of a hybrid ensemble/4D-Var global data assimilation at the Met Office," Q. J. R. Meteorol. Soc. **139,** 1445−1461 (2013).

44. G. Desroziers, J.-T. Camino, and L. Berre, "4DEnVar: link with 4D state formulation of variational assimilation and different possible implementations," Q. J. R. Meteorol. Soc. **140,** 2097−2110 (2014).

45. N. Gustafsson and J. Bojarova, "Four-dimensional ensemble variational (4D-En-Var) data assimilation for the HIgh Resolution Limited Area Model (HIRLAM)," Nonlin. Proc. Geophys **21,** 745−762 (2014).

46. S. Verrieri, P.-Y. Le Traoni, E. Remy, and J. M. Lellouche, "Assessing the impact of SAR altimetry for global ocean analysis and forecasting," J. Oper. Oceanogr. **11** (2), 82−86 (2018).

47. M. Yaremchuk, J. M. D'Addezio, G. Panteleev, and G. Jacobs, "On the approximation of the inverse error covariances of high-resolution satellite altimetry data," Q. J. R. Meteorol. Soc. **144,** 1995−2000 (2018).

48. T. F. Duda, W. G. Zhang, K. R. Helfrich, A. E. Newhall, Y.-T. Lin, J. F. Lynch, P. F. J. Lermusiaux, P. J. Haley Jr., and J. Wilkin, "Issues and progress in the prediction of ocean submesoscale features and internal waves," in OCEANS'14 MTS/IEEE Conf. Proc. (Taipei, 2014), pp. 1−8.

49. S. Karimova, in *Remote Sensing of the Changing Oceans,* Ed. by D. Tang (Springer, Berlin, 2011).

50. A. G. Kostianoy, A. I. Ginzburg, O. Yu. Lavrova, and M. I. Mityagina, in *The Ocean in Motion,* Ed. by M. G. Velarde (Springer, Berlin, 2018), pp. 397−413.

51. A. Zatsepin, A. Kubryakov, A. Aleskerova, D. Elkin, and O. Kukleva, "Physical mechanisms of submesoscale eddies generation: evidences from laboratory modeling and satellite data in the Black Sea," Ocean Dynamics **69** (2), 253−266 (2019).

52. G. I. Marchuk, B. E. Paton, G. K. Korotaev, and V. B. Zalesny, "Data-computing technologies: A new stage in the development of operational oceanography," Izv., Atmos. Ocean. Phys. **49,** 579−591 (2013).

53. V. I. Agoshkov, E. I. Parmuzin, and V. P. Shutyaev, "Numerical algorithm for variational assimilation of sea surface temperature data," Comput. Math. Math. Phys. **48** (8), 1293−1312 (2008).

54. V. I. Agoshkov, V. P. Shutyaev, E. I. Parmuzin, N. B. Zakharova, T. O. Sheloput, and N. R. Lezina, "Variational assimilation of observation data in a mathematical model of Black Sea dynamics," Morsk. Gidrofiz. Zh., No. 6, 15−24 (2019).

55. V. V. Knysh, S. G. Demyshev, G. K. Korotaev, and A. S. Sarkisyan, "Method and results of assimilation of climatic data on temperature, salinity, and sea level into a numerical model of the Black Sea," Izv., Atmos. Ocean. Phys. **43,** 363−377 (2007).

56. G. K. Korotaev, Yu. B. Ratner, M. V. Ivanchik, A. L. Kholod, and A. V. Ivanchik, "Operational system for diagnosis and forecast of hydrophysical characteristics of the Black Sea," Izv., Atmos. Ocean. Phys. **52** (5), 542−549 (2016).

57. G. K. Korotaev, Baroclinic sea layer: from Lineikin's theory to operational oceanography, in *Current Problems in Oceanic and Atmospheric Dynamics: Collection of Papers Devoted to the 100-th Anniversary of Prof. P.S. Lineikin*, Ed. by A. V. Frolov and Yu. D. Resnyanskii (TRIADA LTD, Moscow, 2010), pp. 79−97.

58. R. Ferrari and C. Wunsch, "Ocean circulation kinetic energy: Reservoirs, sources, and sinks," Annu. Rev. Fluid Mech., No. 41, 253−282 (2009).

Translated by A. Nikol'skii

Connections between the Long-Period Variability Modes of Both Temperature and Depth of the Upper Mixed Layer of the North Atlantic and the Climate Variability Indices

P. A. Sukhonos[a], * and **N. A. Diansky[b, c, d], ****

[a]*Institute of Natural and Technical Systems, Sevastopol, 299011 Russia*
[b]*Zubov State Oceanographic Institute, Moscow, 119034 Russia*
[c]*Moscow State University, Moscow, 119991 Russia*
[d]*Marchuk Institute of Numerical Mathematic, Russian Academy of Sciences, Moscow, 119333 Russia*
*e-mails: pasukhonis@mail.ru
**e-mails: nikolay.diansky@gmail.com

Abstract—The interannual–multidecadal variability of the temperature and depth of the upper mixed layer (UML) in the North Atlantic (NA) is analyzed on the basis of the ORA-S3 ocean reanalysis data for 1959–2011. A large part of the UML in the NA is characterized by warming and thinning in all seasons in the period under study. After removing a linear trend, the UML temperature and depth anomalies in individual seasons are decomposed into empirical orthogonal functions (EOF). It is found that the three leading EOFs describe more than 50% of the total variability of the UML temperature and depth. The structure of the first EOF is horseshoe-shaped; this EOF represents coherent changes in the UML temperature and depth throughout the NA, which manifest themselves the year round. This mode corresponds to the Atlantic multidecadal oscillation. The spatial structure of the second EOF in the winter–spring period is a tripole and is caused by the North Atlantic Oscillation (NAO). The time coefficient of the second EOF of the UML temperature in the NA and the NAO index strongly correlate both synchronously and when the NAO index is 11 years ahead. The second EOF of the UML temperature in the summer–autumn period is associated with the Atlantic meridional mode. The third EOF is typical for the UML temperature fluctuations in January and corresponds to the East Atlantic Pattern.

INTRODUCTION

The results of the fundamental study by J. Bjerknes [1] show that the interannual variability of the sea surface temperature (SST) is caused by heat fluxes on the sea surface, which, in turn, are caused by changes in the atmospheric circulation, while decadal or longer SST fluctuations are associated with changes in ocean circulation. In particular, the interannual variability of the subtropical gyre can be a response to the long-term atmospheric forcing (changes associated with the intensity and position of the subtropical maximum). The structures of coherent low-frequency variability in the ocean–atmosphere system, which partly confirm the results by J. Bjerknes, were later obtained with the use of long-term data arrays [2–6]. The authors of recent work [7] note that the interdecadal variability in the North Atlantic (NA) is a direct response of the upper mixed layer (UML) of the ocean to the stochastic atmospheric forcing without the participation of the

thermohaline circulation of ocean waters. The discussion about the role of the ocean in the formation of interdecadal variability continues [8, 9]. Hence, it is quite difficult to identify modes of interannual and multidecadal variability of the NA UML parameters associated with the atmosphere–ocean interaction.

Let us consider main climate signals that can be determined for the NA water area.

The natural long-period SST fluctuation in the NA is called the Atlantic Multidecadal Oscillation (AMO) [2, 10]. This is one of the main climate signals in the World Ocean temperature field on interannual-to-multidecadal scales, and its origin is not associated with the El Niño–Southern Oscillation (ENSO) [11]. Many studies show the AMO effect on climate conditions in the Northern Hemisphere.

The main climate signal in the air pressure field over the Atlantic–European sector is North Atlantic Oscillation (NAO) (see, for example, [12] and the bib-

liography therein). The NAO has several definitions, but it is generally a meridional dipole structure in the air pressure field over the NA. The climate signal that is second in significance is the East Atlantic pattern (EA). The EA is a well-defined monopole in the air pressure field south of Iceland. The NAO and EA strongly affect the atmospheric circulation and long-term weather changes in Europe [13].

The Atlantic meridional mode (AMM) is clearly pronounced in the interannual–decadal variability of the hydrophysical parameters of the tropical Atlantic. It differs from the "zonal" mode of the ENSO type in its physical nature [14]. This mode manifests itself in the form of an anomalous meridional SST gradient through the central latitude of the Intertropical Convergence Zone (ICZ) [15]. The SST anomalies in the tropical Atlantic show significant consistency with NAO and with the variability of the sea level pressure over Iceland and the Azores separately on two sides of the ICZ [16]. Assuming that the NAO affects the meridional modes, the authors of [14] suggest that the AMM can act as an effective conductor for the effect of the extratropical atmosphere on the tropics. In addition, the AMM and AMO strongly correlate with hurricane activity in the NA on the decadal scale. The AMM also strongly correlates with hurricane activity in the NA on an interannual scale [17]. Thus, tropical and extratropical modes of climate variability are interrelated.

To identify the self-consistent spatiotemporal structures in the fields of hydrophysical parameters, decomposition into the empirical orthogonal functions (EOFs) can be used successfully [18]. The SST anomalies are decomposed into EOFs in many works for various time periods and different NA regions. However, different authors use different data processing techniques. First and foremost, this refers to the spatiotemporal averaging of the source data. This can be one of the reasons for the inconsistency of the results. At the same time, the main, most energetic, EOF modes show a strong tendency to have the simplest spatial structure inside a region analyzed. This property leads to a strong dependence of EOFs on the shape of the spatial boundaries of a region. In addition, the results of the EOF analysis depend on the length of the time series, since individual modes of upper ocean layer temperature variability can make different contributions to the total dispersion at different time periods and, therefore, are time dependent. Thus, EOFs should be extremely carefully interpreted as physical/dynamic modes of variability; this interpretation should always be accompanied by the physical analysis of their generation.

The EOF analysis was apparently first applied to the SST field in the NA in [19]. The low-frequency winter climate variability over the NA was analyzed in [2] on the basis of observations over 90 years. The main spatiotemporal regularities in the SST and sea-level pressure variability in the Atlantic Ocean for 1856–1991

are described in [20]. Note that interannual fluctuations in hydrophysical fields are shown in simultaneous variations in the annual averages and characteristics of seasonal variability. This means that the annual variability of parameters of hydrophysical fields changes from year to year, which is confirmed in [21] for large-scale SST anomalies in the NA. Therefore, we analyze the spatiotemporal structures of the interannual and multidecadal variability of the monthly average UML temperatures and depth in the NA separately for different seasons. The results are based on the analysis of EOFs calculated from the detrended data of ORA-S3 ocean reanalysis. We set the goal to find the correlations between the EOF of the UML temperature and depth and the above-described climate signals.

DATA AND PROCESSING TECHNIQUE

The data on the monthly average UML temperature and depth are taken from the ORA-S3 ocean reanalysis array for the period from January 1959 to December 2011 [22]. The spatial resolution of these data is $1°$ and, in the equatorial zone ($\pm 10°$ latitude), $0.3° \times 1°$ in latitude and longitude, respectively. In addition, we use data on the net ocean surface heat fluxes and the wind stress over the NA water area taken from the ERA-40 atmospheric reanalysis array [23] for the period from January 1959 to June 2002 and operational ERA-40 model analysis for July 2002 to December 2011. These heat and momentum fluxes are used as boundary conditions in the ORA-S3 ocean reanalysis model. The NA water area selected for this study is bounded by the coordinates $0°-70°$ N and $80°-10°$ W, which coincides with the water area of the AMO definition.

The UML depth is calculated from the selected reanalysis based on the semiempirical theory of turbulence [24]. According to this theory, the UML depth corresponds to the depth where the Richardson number attains a critical value of 0.3.

The average temperature is calculated for each month from 1959 to 2011 within the UML depth variable in space and time based on 3D data of the reanalysis selected. Then, the values of the temperature and UML depth are distinguished for individual months. Further, linear trends are removed from the UML temperature and depth time series at each node of the spatial grid. Linear trend parameters are calculated by the least-squares method. After this, the resulting arrays of UML temperature anomalies and depths are EOF decomposed separately for each calendar month over the period under study [25].

The monthly average AMO and AMM indices for 1948–2017 were taken from the website https://www.esrl.noaa.gov/psd/data/climateindices/list/. The monthly average NAO and EA for 1950–2017 were taken from http://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml.

Along with EOF decomposition, we use composite analysis, which consists of the following. The NAO, AMM, and EA indices, which exceed in absolute value one standard deviation, allowed us to identify anomalous years. These years are grouped into two samples which correspond to the positive and negative phases of each climate signal. Each phase for each index includes at least 7 abnormal years, which is no less than 15% of the length of the corresponding time series. For these groups of years, the average values, variances, and standard deviations of the UML temperature and depth are calculated at each node of the regular grid. Then, a "clean" climate signal in the NA UML temperature and depth is found for calendar months. For this, the difference between the sample averages at each grid node (the so-called difference composite) is determined. The statistical significance of the differences between composite anomalies in the periods under study is assessed by the standard algorithm with the use of Student's *t*-test. We also use correlation analysis.

RESULTS

Analysis of Linear Trends in UML Temperature and Depth

According to the data used, most of the NA is characterized by a positive linear trend in the UML temperature. The highest coefficients of the linear trend are observed in the region of the Gulf Stream transition into the North Atlantic Current. Their values are 0.05, 0.08, 0.04, and 0.04°C/year in January, April, July, and October, respectively. The coefficients of the linear trend in the cold season noticeably exceed those in the warm season, which corresponds to stronger warming in the winter period. Inside the subpolar gyre, the coefficients of the linear trend in the UML temperature are negative, except for October. The contribution of the linear trend variance into the total variance of the UML temperature exceeds 30% in the latitudinal band $0°−10°$ N to the east of $40°$ W (except for April), in the region of Gulf Stream transition into the North Atlantic Current (except for July) and in the vicinity of the East Greenland Current in July. The contribution of the linear trend variance into the total variance of the UML temperature throughout the NA water area is 13.8, 9.4, 15.4, and 20.7% in January, April, July, and October, respectively.

In tropical and subtropical latitudes, there are regions where the UML depth is characterized by an insignificant positive linear trend in the winter months in 1959−2011. In high latitudes, significant negative linear trends in the UML depth were revealed for the period under study. The coefficient of the linear trend in the UML depth is 30 m/year in the region of intense convection in the Labrador Sea in January, which results in an almost twofold decrease in the average UML depth (from 3 to 1.5 km). After removing the

linear trend, the standard deviation of the UML depth is equal to ~1 km in this region in January. The UML depths, their standard deviations, and the coefficients of the linear trend in the summer months are smaller than in the winter months. Thus, processes which occur in high latitudes make the main contribution to the low-frequency variability of the NA UML depth. A decrease in this parameter is generally noted and is most pronounced in the cold season.

Thus, the NA UML is characterized by warming and thinning during the period under study. The latter occurs mainly due to the weakening of convective mixing in high latitudes. This might well be due to an increase in the Arctic Ocean temperature, the intensification of melting of Greenland glaciers, and the flowing of desalinated waters out of the Arctic Ocean in the second half of the 20th century [26]. Possible climate changes in the Northern Hemisphere if the oceanic heat influx in the NA ceases were estimated in [27]. Then, after detrending the time series, we analyze the interannual and multidecadal variability of the UML temperature and depth based on the EOF decomposition.

Main EOFs of the UML Temperature and Depth in the NA

The spatial structures of the first EOF of the NA UML temperature for each month are consistent with each other and are of horseshoe shape. The values have the same sign in most of the water area, while the region of the opposite sign is in the western part of the subtropical gyre (Fig. 1a). However, there are some differences between the structures in different months. The size of the region of the opposite sign in the western part of the subtropical gyre is maximal in April and minimal in October. The contribution of the first EOF to the total variability of the UML temperature in the NA is 29.6, 40.4, 18.3, and 21.8% in January, April, July, and October, respectively. The high contribution in April is explained by a change from winter mixed to summer stratified state, when high temperature anomalies spontaneously occur in UML at its small depth. The time coefficients of the first EOF of the NA UML temperature show the same variability on the interdecadal−multidecadal scale for each month of the year (Fig. 1c). This is manifested in a strong correlation between these time series. As for the multidecadal variability, there are long periods of low (e.g., in the early 1970s to the early 1990s) and high UML temperatures (e.g., in the late 1990s and 2000s). The correlation coefficients between the time coefficients of the first EOF of the NA UML temperature and the AMO index are 0.82 in January, 0.88 in April, 0.85 in July, and 0.88 in October in 1959−2011.

The spatial structures of the first EOF of the UML depth in the NA for each month are consistent with each other. The highest values of the same sign

are concentrated in the inner part of the subpolar gyre (Fig. 1b). The area of this region is maximal in the autumn−winter months and minimal in summer. The contribution of the first EOF to the total variability of the UML depth in the NA is 51.5, 40, 32.5, and 59.7% in January, April, July, and October, respectively. The correlation coefficient between the time coefficients of the first EOF of the UML depth is equal to 0.73 in January and April. A low-frequency quasi-sixty-year oscillation is manifested in the time coefficients of the first EOF of the UML depth in the NA for each month of the year, which is similar to the behavior of the AMO index. The correlation coefficient between the time coefficient of the first EOF of the UML depth and the AMO index is 0.69 in January in the period under study (Fig. 1c). The correlation coefficient between the time coefficient of the first EOF of the UML temperature and the time coefficient of the first EOF of the UML depth is 0.75 in January.

Thus, despite some local differences between months, the first EOFs of the UML temperature and depth in the NA correlate well with the AMO. This points out to the large-scale nature of this climate signal. The UML temperature increases with a decrease in the UML depth in the subpolar gyre during the positive AMO phase. Note that the amplitude of this EOF in the subpolar gyre can be underestimated because of averaging of the ocean temperature over deep UML, the depth of which can exceed 2000 m in January. This becomes possible in the upper ocean layer when regions of local heating are formed in it due to the winter interaction of the active ocean layer with the atmosphere [28].

The spatial structures of the second and third EOFs of the UML depth in the NA are regions of different signs within the subpolar gyre. The contributions of the second and third EOFs to the total variability of the UML depth are low; they are 8.7, 8, 10.3, and 8.6% for the second EOF and 5.8, 6, 9, and 5.6% for the third EOF in January, April, July, and October, respectively. We will not dwell on these EOFs of the UML depth in detail below; only note that, according to the data used, the variance of the UML depth throughout the NA water area is 19134, 17442, 2333, and 4631 m^2 in January, April, July, and October, respectively. This variance is approximately 7−8 times higher in the winter months than in the summer months. Therefore, despite the higher relative contribution (in %) of the second and third EOFs to the total variability of the UML depth in the summer months as compared to the winter months, the parts of the variance described by these EOFs are small in the summer months.

Let us analyze the second EOF of the NA UML temperature. Figure 2a shows its spatial structure for January and April. The structure of the second EOF in these months shows changes in the UML temperature of different signs in different latitudinal zones of
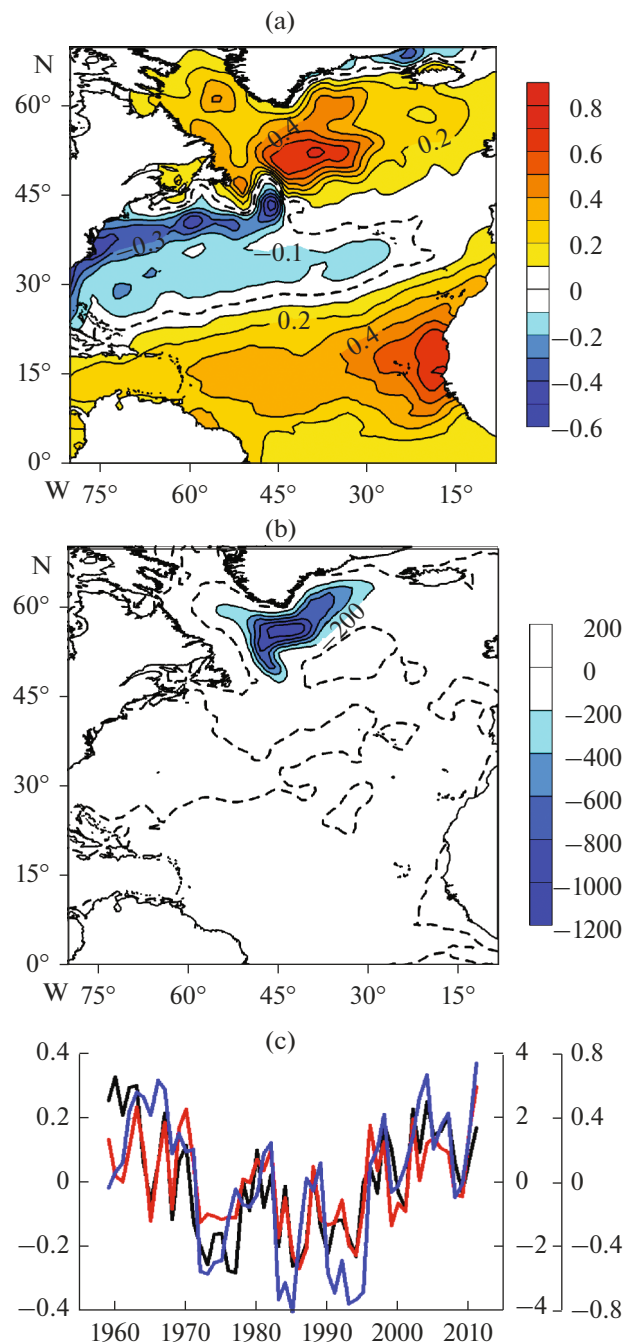

Fig. 1. Spatial structures of the first EOF of the (a) UML temperature and (b) depth in January; (c) corresponding time coefficients of the decomposition of the UML temperature (red curve, second scale on the right) and depth (blue curve, first scale on the right) and AMO index in January (black curve, left scale). The dashed curves in (a, b) show the zero isolines.

the NA. The changes in the spatial structure of this mode for January and April are insignificant and are mainly shown in an increase in the regions of the opposite signs in the western parts of the subtropical and subpolar gyres in April. This EOF describes from 11.3 (January) to 8.6% (April) of the total variability of
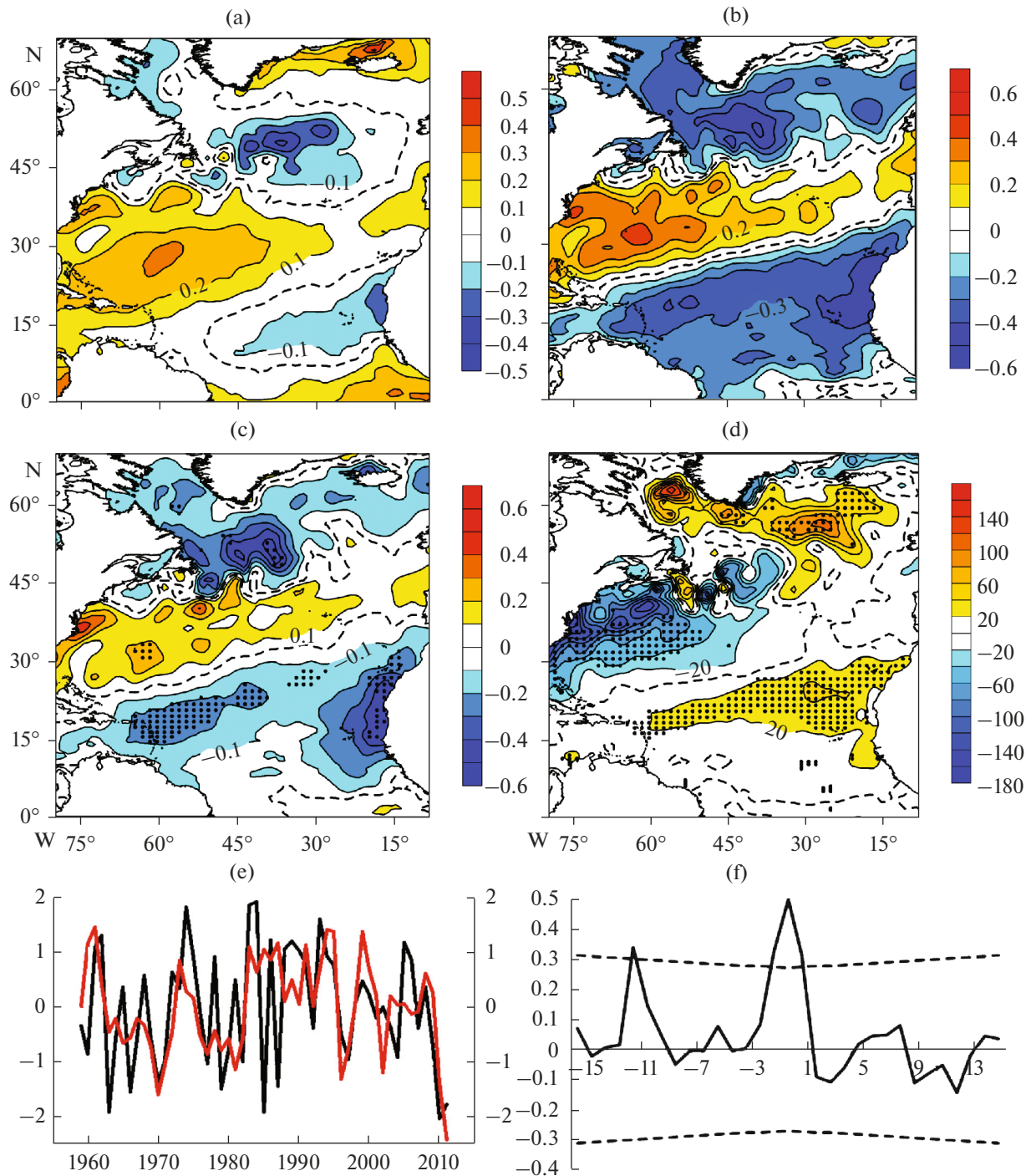
Fig. 2. (a) Spatial structure of the second EOF of the UML temperature in January; (b) spatial distribution of the correlation coefficients between the UML temperature and the NAO index in January; differences between (c) the UML temperature (°C) and (d) net ocean surface heat fluxes (W/m^2; positive values correspond to the heat loss) anomalies during the positive and negative NAO phases in January; (e) time coefficients of the second EOF of the UML temperature (red curve, right scale) and the NAO index (black curve, left scale) in January; and (f) cross-correlation function between the time series shown in (e). Negative shifts (years) correspond to the advance in the NAO index. Dashed curves in (a−d) show the zero isolines. Black dots in (c, d) show the grid nodes, where the difference is statistically significant at a level of 90%. The dashed curves in (f) show the 95% confidence interval.

UML temperature in the NA. The correlation coefficient between them is 0.48 in January and April for 1959−2011. The time coefficients of the second EOF of the UML temperature show strong interannual variability (Fig. 2e). The correlation coefficient between the time coefficient of the second EOF of the UML temperature and the NAO index is 0.51 after removing the linear trend in January.

The correlation analysis of the NAO index and the NA UML temperature at each grid node in January shows a close relationship between these characteristics, negative in the inner part of the subpolar gyre and in the West African upwelling region and positive in the inner part of the subtropical gyre (Fig. 2b).

The period under study includes 10 years with the positive phase (1974, 1983, 1984, 1986, 1989, 1990, 1993, 1994, 2005, and 2006) and 9 years with the negative phase of the NAO (1960, 1966, 1970, 1971, 1977, 1979, 1985, 1987, and 2010). The UML temperature during the positive NAO phase, as compared to the negative phase, is characterized by a statistically significant decrease by 0.4°C in the inner part of the subpolar gyre and in the West African upwelling region and an increase in the inner part of the subtropical gyre (Fig. 2c).

The spatial structures of the second EOF of the UML temperature (Fig. 2a), the field of the correlation coefficients between the UML temperature and the NAO index (Fig. 2b), and the difference composite (Fig. 2c) are very similar. A similar structure is also confirmed by the composite analysis of the net ocean surface heat fluxes during the positive and negative NAO phases (Fig. 2d). Regions with high heat loss from the ocean surface are located in the inner part of the subpolar gyre ($+70$ W/m^2) and in the Northern Equatorial Current ($+35$ W/m^2). A decrease in the heat release to the atmosphere is noted in the western part of the subtropical gyre (-45 W/m^2). Thus, the composite and correlation analysis confirm that the intensification of NAO is accompanied by a drop in the UML temperature in the regions of trade and westerly winds and its increase in subtropical latitudes. However, the signs of the heat flux and UML temperature anomalies are inconsistent to the east of Newfoundland along 45° N, where the extreme point of the northern cell of the EOF is located. This can be explained by a significant role of advection factors of formation of UML temperature anomalies here, since this region is influenced by the Gulf Stream and the North Atlantic Current.

The analysis of the cross-correlation function between the time coefficient of the second EOF of the UML temperature and the NAO index after detrending in January showed that the highest correlation coefficients of about 0.5 are observed at a zero shift between these time series (Fig. 2f). In addition, these time series strongly correlate when the NAO index is 11 years ahead.

The spatial structures of the second EOF of the UML temperature in the NA in July and October are similar. Figure 3a shows the spatial structure of the second EOF based on July data. The structure of this EOF is horseshoe-shaped, oriented from west to east, with values of one sign in the eastern part of the equatorial Atlantic, latitudinal band 35°−50° N, and the East Greenland Current and values of the opposite sign in the rest of the water area. The second EOF describes from 12 (July) to 10.5% (October) of the total variability of the UML temperature in the NA. The coefficient of correlation between the time coefficient of this EOF and the AMM index after detrending is equal to 0.47 for 1959−2011 (Fig. 3e). The AMM index in July shows pronounced 10-year variability: it decreases by the early 1970s and then increases.

The detrended AMM index and the UML temperature strongly correlate at each grid node in the NA, especially to the south of 25° N, in July (Fig. 3b); the correlation coefficients exceed 0.6 here. Thus, more than 35% of the total variability of the UML temperature in the tropical Atlantic in summer is due to the effect of the AMM.

The period under study includes 7 years with the positive AMM phase (1962, 1988, 1989, 1995, 2004, 2005, and 2010) and 9 years with the negative AMM phase (1972, 1973, 1974, 1984, 1986, 1991, 1993, 1994, and 2002). During the positive AMM phase, as compared to the negative phase, the UML temperature is characterized by a statistically significant increase in the inner part of the subpolar gyre, the eastern part of the subtropical gyre, and tropical latitudes (except for the eastern part of the equatorial Atlantic) (Fig. 3c). Negative UML temperatures are observed in the East Greenland Current. The analysis of the difference composite of the wind-stress module during the positive and negative AMM phases shows a significant decrease in this parameter during the positive AMM phase in the trade-wind region (-0.012 N/m^2) (Fig. 3d). Thus, it is confirmed that the intensification of the AMM in July is accompanied by UML warming in the tropical Atlantic and a decrease in the wind-stress modulus in the trade wind region.

Let us now consider the third EOF of the NA UML temperature. Figure 4a shows the spatial structure of this mode for January. One can see changes in the UML temperature of different signs: one sign in the vicinity of ICZ and to the north of 30° N, and the opposite sign in the latitudinal band 15°−30° N. This EOF describes 8.2% (January) of the total variability of the NA UML temperature; its time coefficient is characterized by pronounced interdecadal variability. The correlation coefficient between this coefficient and the EA index after removing the linear trend is equal to 0.31 in January for 1959−2011 (Fig. 4b). The correlation between these time series after the removal of the parabolic trend is 0.33.

The period under study includes 10 years with the positive EA phase (1970, 1971, 1973, 1988, 1991, 2001, 2002, 2003, 2007, and 2009) and 8 years with the negative one (1963, 1965, 1968, 1969, 1976, 1981, 2000, and 2005). A statistically significant decrease of 0.3°C in the UML temperature during the positive EA phase, as compared to the negative phase, is observed within the region 35°−45° N and 35°−20° W. Positive UML temperatures are observed to the north of South
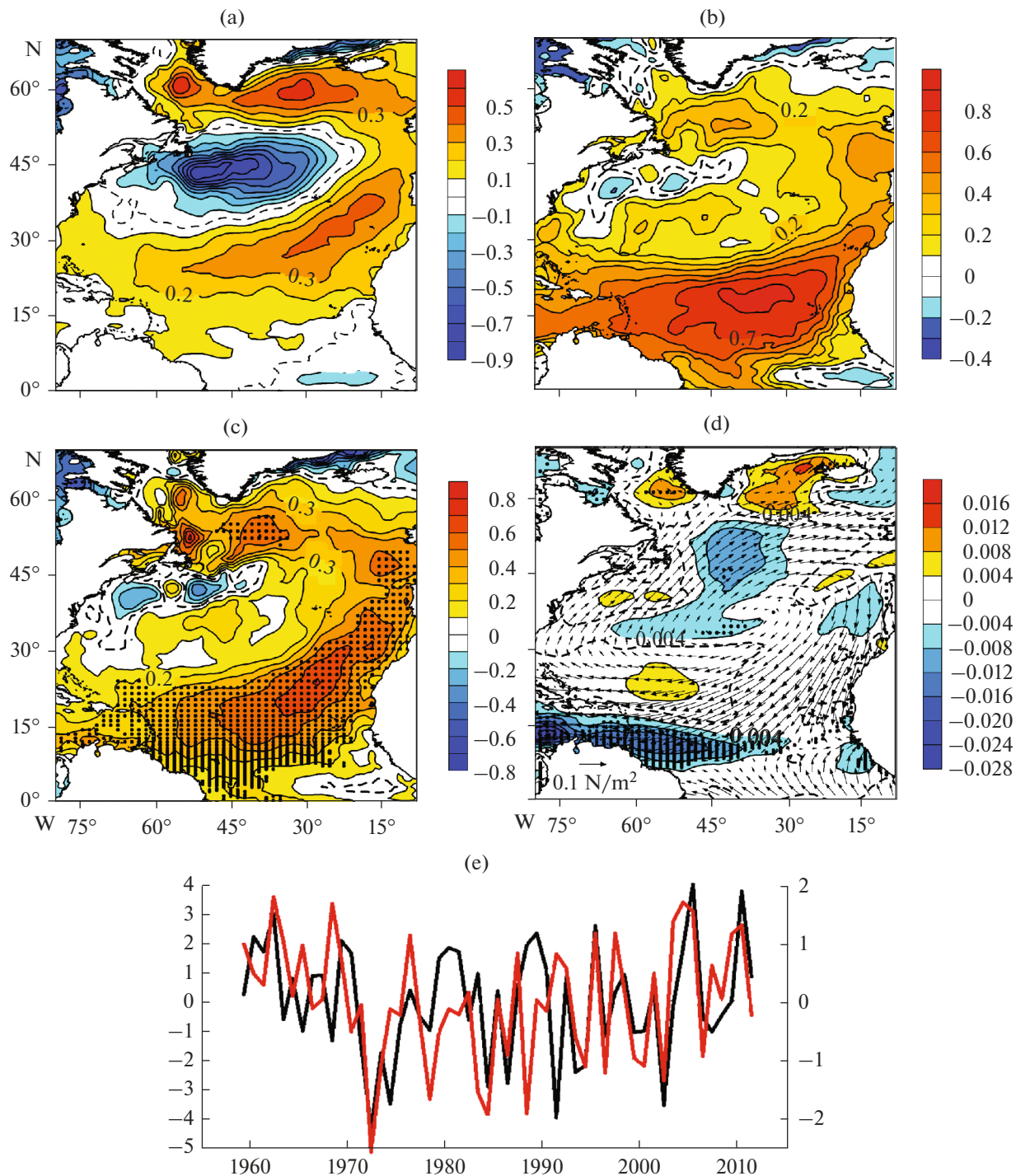
Fig. 3. (a) Spatial structure of the second EOF of the UML temperature in July; (b) spatial distribution of correlation coefficients between the UML temperature and the AMM index in July; differences between (c) the UML temperature (°C) and (d) the wind-stress-modulus (N/ m²) anomalies during the positive and negative AMM phases in July; and (e) time coefficient of the second EOF of the UML temperature (red curve, right scale) and the AMM index (black curve, left scale). Dashed curves in (a−d) show the zero isolines. Black dots in (c, d) show the grid nodes, where the difference is statistically significant at a level of 90%. The vectors in (d) show the 1959−2011 averaged wind stress in July.

America (+0.2°C) (Fig. 4c). Thus, the intensification of EA is accompanied by the cooling of UML in the vicinity of the Azores and warming in the Greater Antilles. This result is confirmed by the correlation coefficients between the time series of the UML temperature and the EA index after the removal of the linear trend at each grid node in January (Fig. 4d). Note, however, that the third EOF describes a much
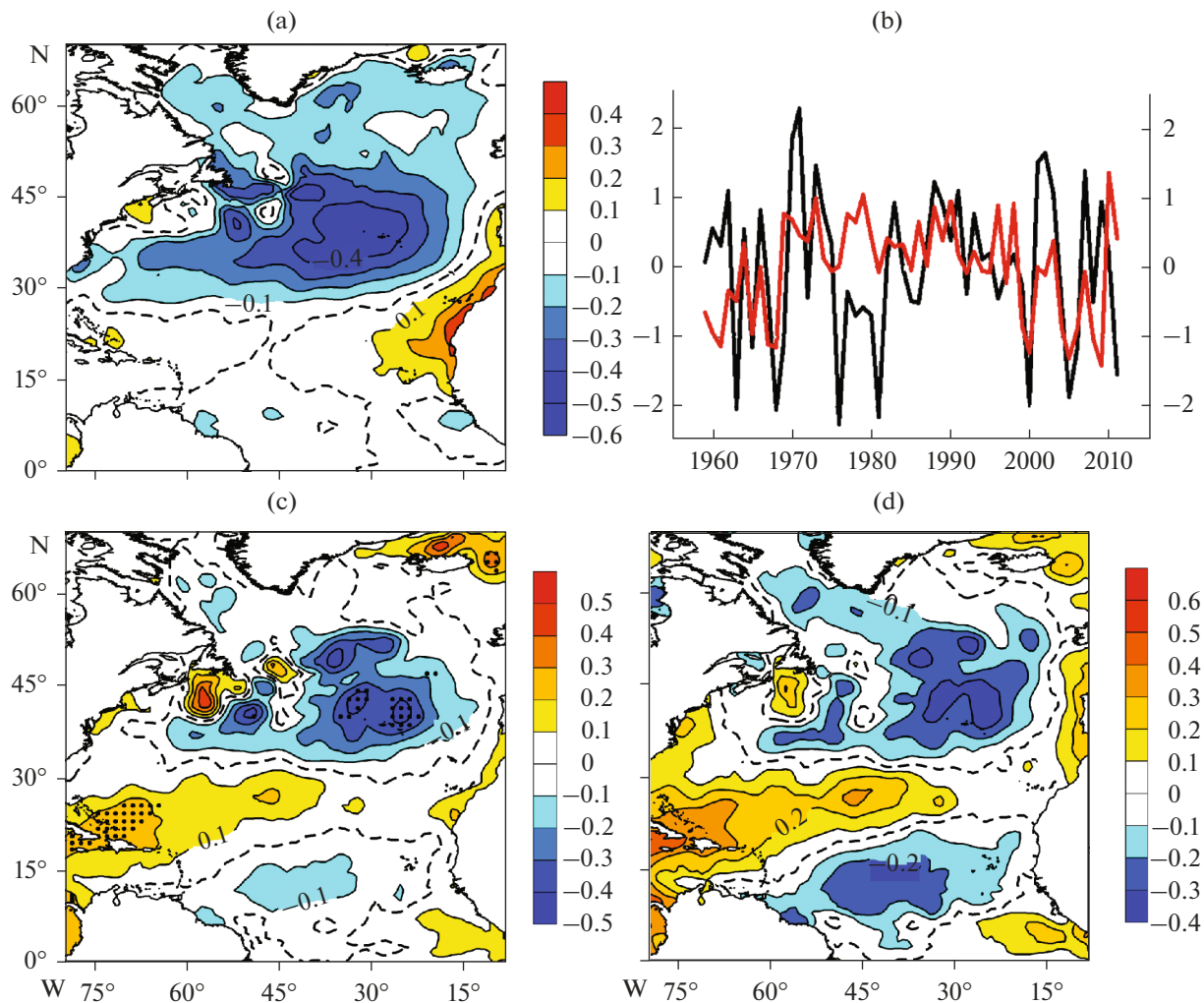
Fig. 4. (a) Spatial structure of the third EOF of the UML temperature in January; (b) time coefficient of the third EOF of the UML temperature (red curve, right scale) and the EA index (black curve, left scale) in January; (c) difference between the UML temperature anomalies (°C) during the positive and negative EA phases in January; and (d) spatial distribution of the correlation coefficients between the UML temperature and the EA index in January. Dashed curves in (a, c, d) show the zero isolines. Black dots in (c) show the grid nodes, where the difference is statistically significant at a level of 90%.

smaller fraction of the total winter variability of the UML temperature.

The spatial structures of the third EOF of the UML temperature in April, July, and October are regions with different signs within the NA. The contribution of this EOF to the total variability of the UML temperature is low: 6, 10, and 7.7% in April, July, and October, respectively. The variance of UML temperature is 0.38, 0.28, 0.22, and 0.26°C² in January, April, July, and October, respectively, for the entire NA water area. This parameter is approximately 1.5 times higher in the winter months than in the summer months. Therefore, despite the large relative contribution (in %) of the third EOF to the total variability of the UML temperature from the summer months as compared to the winter months, the part of the variance described by this EOF in January is almost 1.5 times larger than the corresponding part of the dispersion of this EOF in July.

DISCUSSION

The variability of the NA UML depth was previously analyzed for 1960–2004 in [29]. The authors showed that the UML depth increased by 10–40 m in the central part of the NA during the winter–spring period for those 45 years. In that work, the UML depth was calculated using the temperature criterion. According to this criterion, the UML depth is defined as the depth at which the temperature changes by 0.2°C with respect to its value at a depth of 10 m. We emphasize that the difference and gradient criteria for UML depth estimation require the careful selection of the threshold values, since the resulting UML depth (and its long-period variability) strongly depend on the methodology for its estimation. High values of the temperature criterion apparently cover deeper gradients in the thermocline instead of the lower UML boundary, which is especially important under condi-

tions of low-temperature stratification in the northern NA (see, for example, Figs. 2g and 2i in [30]). It should be noted that the temperature difference criterion for UML depth estimation does not take into account the salinity contribution to the density. Therefore, it is more correct to use the density difference criterion. Note that the technique for determining the depth of the UML lower boundary by the Richardson number is more justified from the physical point of view. According to our data, the UML depth actually increased in the subtropics in the winter–spring period from 1960 to 2004. However, since the early 2000s, during the positive AMO phase, the intensity of the subtropical convective cell has weakened [31] and the UML depth in January has decreased on a multidecadal scale [32]. That resulted in a decrease in the long-term (1959–2011) winter deepening of the UML in the subtropics. As for the interannual–multidecadal variability of the UML depth for the entire NA (after removing the linear trend), the role of ocean processes at high latitudes is of great importance. Note that linear trends are separately analyzed and removed in this work, after which the UML parameters are analyzed. This is due to the fact that tendencies in the anthropogenic forcing and natural variability coincided in the UML in 1960–2004.

The maximal UML depth in the central part of the Labrador Sea was assessed in [33] on the basis of available observational data for 1993–2014. The winter maxima of the UML depth significantly decreased from the mid-1990s to the mid-2000s against the background of intense interannual variability. This fact is consistent with our assessments of the UML depth by the Richardson criterion.

According to our results, the AMO index can also be defined as the time coefficient of the first mode of EOF decomposition of the monthly average UML temperature or depth. This statement is probably true for the EOF decomposition of long time series (longer than or equal to the AMO period). For example, the Pacific Decadal Oscillation index is defined as the time coefficient of the first mode of monthly average SST decomposition in the North Pacific Ocean (to the north of 20° N) [34]. Our results also show the subpolar gyre to be a key region for the AMO formation and the importance of processes at the lower UML boundary in the evolution of variability on this scale. This conclusion does not agree with the results [7], obtained with the use of an extremely simplified model of the ocean with the constant-depth UML. This is an indirect confirmation of the important role of thermohaline circulation in the AMO formation defended in [8]. In addition, the Arctic ocean processes play a large role in maintaining the AMO [35, 36].

The spatial structure of the second EOF of the UML temperature, consistent with the NAO index, is tripole, where correlations with SST are positive in the Sargasso Sea and negative in the northwestern part of the tropical Atlantic and the vicinity of the Labrador

Sea (see, e.g., [4, 5, 37, 38] and others). This structure in the ocean–atmosphere system is associated with wind heat advection over the ocean [39]. As is shown in [4, 5], this mode in winter UML temperature anomalies is generated under the atmospheric forcing, which confirms the forced nature of this mode. The maximal correlation coefficient between the time coefficient of the second EOF of the UML temperature and the NAO index after detrending increases to approximately 0.75 when the atmospheric forcing is 0.5 months ahead [5], which is explained with the help of a simple analytical model for the evolution of UML temperature anomalies. The half-month shift of the delay of large-scale UML temperature anomalies in the midlatitudes is defined as a quarter of the period from the most significant period of fluctuations in the atmospheric forcing, which is about 2 months in the most energy-carrying range of NAO variability, since the low-frequency variability of the atmosphere in the monthly average fields is clearly manifested just in this period.

The increase in the correlation observed when the NAO index is 11 years ahead (Fig. 2f) cannot be explained by these simple considerations, since a minimum is observed in the NAO index spectrum in 10- to 40-year periods (see, for example, [40]). Therefore, the relationship between the UML temperature anomalies and the NAO 11 years ahead requires an explanation. Note that the typical time for the subtropical gyre to adapt to changing atmospheric forcing is about 10 years [41].

The authors of [42] suggested a correlation between the monopole mode south of Iceland (the third EOF of the UML temperature in January in the present work) and the ocean effect on atmospheric processes based on relatively short time series (1950–1987). Our results are based on long-term data; they show the coincidence of the third EOF of the UML temperature in January with the EA. Moreover, the EA significantly affects the UML temperature in several small NA regions. However, its contribution to the total winter variability of the UML temperature is the smallest in comparison with other modes under study.

Atmospheric circulation factors play an important role in the formation of the NA UML temperature variability in winter. In summer, winds and currents are weak in the NA and the UML depth is reduced. Hence, the atmospheric circulation indices, such as NAO and EA, correlate weakly with the time coefficients of the EOF decomposition of the UML temperature for the summer months. The summer NAO is characterized by lower amplitude when compared with the winter period and the northeastward displacement of its centers of action beyond the NA boundaries [43]. Therefore, in summer, this climate signal can no longer describe a large fraction of the variance of the NA UML parameters, and the role of tropical variability modes increases.

Changes in the intensity of the trade winds in the tropical Atlantic precede SST anomalies (and, therefore, the anomalous SST gradient at the central latitude of the ICZ): weaker (stronger) trade winds are accompanied by warmer (colder) SST anomalies [15]. In addition, this means that the "meridional" mode of the NA UML temperature variability is generated under an external forcing. The NAO can act as one of the sources of this forcing. However, another explanation of the meridional mode is a positive feedback between the wind speed, evaporation, and SST anomalies [44, 45] (although an external forcing is also required to maintain the meridional mode in this case).

The close relationship between the EOF decomposition modes of the UML temperature and individual processes in the ocean–atmosphere system is of interest. Since EOFs are mutually orthogonal by definition, there must also be a quasi-orthogonal AMO, NAO, EA, and AMM associated with them. This is partially confirmed by the small values of the synchronous correlation coefficients between the indices of the climate signals under study in different seasons.

CONCLUSIONS

We have analyzed the linear trends and interannual–multidecadal variability of the NA UML temperature and depth in different seasons. The results are based on the EOF decomposition of ORA-S3 ocean reanalysis data for 1959–2011.

Warming of the NA UML is noted, along with a decrease in its depth in the period under study. A positive linear trend in the UML temperature is pronounced in all months of the year in most of the NA water area, although negative trends are observed in some regions. Significant linear trends in the UML depth are mainly concentrated at high latitudes and better pronounced in the winter months. In summer, linear trends in the UML depth variability are also observed, but their coefficients are small.

The analysis of the main modes of NA UML temperature and depth after detrending shows the following. The three leading EOFs describe more than 50% of the total variability of the UML temperature and depth. The first EOF shows a coherent multidecadal variability of these parameters throughout the NA water area. Despite some differences in its spatial structure in individual months, this EOF is a manifestation of the AMO. The second EOF is characterized by a spatial structure with opposite signs in different latitudinal zones of the NA for UML temperature fluctuations in January and April. The contribution of this EOF to the total variability of the UML temperature is about half the contribution of the first EOF. This EOF is caused by the NAO. A significant correlation was found between the time coefficient of the second EOF of the UML temperature and the NAO index without the linear trend, both synchronously

and when the NAO is 11 years ahead. The second EOF for UML temperature variations in July and October is characterized by a spatial structure, where changes in the UML temperature have one sign in the eastern part of the equatorial Atlantic, North Atlantic, and East Greenland currents and the opposite sign in the rest of the NA water area. This EOF was found to correspond to the AMM. The third EOF of the UML temperature fluctuations coincides with the EA in January. However, its contribution to the total UML temperature variability is low.

Thus, only the lowest frequency mode shows the evolution of the AMO index, which can be associated with fluctuations in the thermohaline circulation in the NA. The second and third EOF modes are the response of the UML to the atmospheric forcing determined by the NAO, AMM, and EA. Moreover, the second UML temperature mode differs in nature in the cold and warm season.

REFERENCES

1. J. Bjerknes, *Atlantic Air-Sea Interaction,* vol. 10 of *Advances in Geophysics* (Academic Press, New York, 1964). https://doi.org/10.1016/S0065-2687(08)60005-9

2. C. Deser and M. L. Blackmon, "Surface climate variations over the North Atlantic Ocean during winter: 1900–1989," J. Clim. **6** (9), 1743–1753 (1993).

3. Y. Kushnir, "Interdecadal variations in North Atlantic sea surface temperature and associated atmospheric conditions," J. Clim. **7** (1), 141–157 (1994).

4. C. Deser and M. S. Timlin, "Atmosphere–ocean interaction on weekly timescales in the North Atlantic and Pacific," J. Clim. **10** (3), 393–408 (1997).

5. N. A. Dianskii, "Temporal relationships and special forms of the combined modes of altitude anomaly for the 500 mb isobaric surface and ocean surface temperature in the North Atlantic Ocean in winter," Izv., Atmos. Ocean. Phys. **34** (2), 197–213 (1998).

6. S. K. Gulev, M. Latif, N. Keenlyside, et al., "North Atlantic Ocean control on surface heat flux on multidecadal timescales," Nature **499** (7459), 464–467 (2013). https://doi.org/10.1038/nature12268

7. A. Clement, K. Bellomo, L. N. Murphy, et al., "The Atlantic Multidecadal Oscillation without a role for ocean circulation," Science **350** (6258), 320−324 (2015).
https://doi.org/10.1126/science.aab3980

8. R. Zhang, R. Sutton, G. Danabasoglu, et al., "Comment on "The Atlantic Multidecadal Oscillation without a role for ocean circulation"," Science **352** (6293), 1527 (2016).
https://doi.org/10.1126/science.aaf1660

9. K. Bellomo, L. N. Murphy, M. A. Cane, et al., "Historical forcings as main drivers of the Atlantic multidecadal variability in the CESM large ensemble," Clim. Dyn. **50** (9-10), 3687−3698 (2018).
https://doi.org/10.1007/s00382-017-3834-3

10. M. E. Schlesinger and N. Ramankutty, "An oscillation in the global climate system of period 65−70 years," Nature **367** (6465), 723−726 (1994).
https://doi.org/10.1038/367723a0

11. D. B. Enfield and A. M. Mestas-Nunez, "Multiscale variabilities in global sea surface temperatures and their relationships with tropospheric climate patterns," J. Clim. **12** (9), 2719−2733 (1999).

12. E. S. Nesterov, *The North Atlantic Oscillation: The Atmosphere and the Ocean* (Triada, Moscow, 2013) [in Russian].

13. G. W. K. Moore and I. Renfrew, "Cold European winters: Interplay between the NAO and the East Atlantic mode," Atmos. Sci. Lett. **13** (1), 1−8 (2012).
https://doi.org/10.1002/asl.356

14. J. C. H. Chiang and D. J. Vimont, "Analogous Pacific and Atlantic meridional modes of tropical atmosphere−ocean variability," J. Clim. **17** (21), 4143−4158 (2004).
https://doi.org/10.1175/JCLI4953.1

15. P. Nobre and J. Shukla, "Variations of sea surface temperature, wind stress, and rainfall over the tropical Atlantic and South America," J. Clim. **9** (10), 2464−2479 (1996).

16. B. Rajagopalan, Y. Kushnir, and Y. M. Tourre, "Observed decadal midlatitude and tropical Atlantic climate variability," Geophys. Res. Lett. **25** (21), 3967−3970 (1998).
https://doi.org/10.1029/1998GL900065

17. D. J. Vimont and J. P. Kossin, "The Atlantic Meridional Mode and hurricane activity," Geophys. Res. Lett. **34** (7), L07709 (2007).
https://doi.org/10.1029/2007GL029683

18. P. A. Vainovskii and V. N. Malinin, *Processing and Analysis Methods for Oceanological Information. Part 2. Multidimensional Analysis* (RGGMI, St. Petersburg, 1992) [in Russian].

19. B. C. Weare and R. E. Newell, "Empirical orthogonal analysis of Atlantic Ocean surface temperatures," Q. J. R. Meteorol. Soc. **103** (437), 467−478 (1977).
https://doi.org/10.1002/qj.49710343707

20. Y. M. Tourre, B. Rajagopalan, and Y. Kushnir, "Dominant patterns of climate variability in the Atlantic Ocean during the last 136 years," J. Clim. **12** (8), 2285−2299 (1999).

21. A. I. Ugryumov "On large-scale oscillations of water surface temperature in the North Atlantic," Meteorol. Gidrol. **5,** 12−22 (1973).

22. M. A. Balmaseda, A. Vidard, and D. L. T. Anderson, "The ECMWF Ocean Analysis System: ORA-S3," Mon. Weather. Rev. **136** (8), 3018−3034 (2008).
https://doi.org/10.1175/2008MWR2433.1

23. S. M. Uppala, P. W. Kållberg, A. J. Simmons, et al., "The ERA-40 re-analysis," Q. J. R. Meteorol. Soc. **131B** (612), 2961−3012 (2005).

24. R. C. Pacanowski and S. G. H. Philander, "Parameterization of vertical mixing in numerical models of tropical oceans," J. Phys. Oceanogr. **11** (11), 1443−1451 (1981).

25. N. E. Huang, Z. Shen, S. R. Long, et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. R. Soc. London, Ser. A **454** (1971), 903−995 (1998).
https://doi.org/10.1098/rspa.1998.0193

26. *IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change,* Ed. by T. F. Stocker, D. Qin, G.- K. Plattner, et al. (Cambridge University Press, Cambridge, UK and New York, NY, USA, 2013).
https://doi.org/10.1017/CBO9781107415324

27. V. V. Zuev, V. A. Semenov, E. A. Shelekhova, S. K. Gulev, and P. Koltermann, "Evaluation of the impact of oceanic heat transport in the North Atlantic and Barents Sea on the Northern Hemispheric climate," Dokl. Earth Sci. **445** (2), 1006−1010 (2012).

28. S. N. Moshonkin and N. A. Diansky, "Upper mixed layer temperature anomalies at the North Atlantics storm-track zone," Ann. Geophys. **13,** 1015−1026 (1995).

29. J. A. Carton, S. A. Grodsky, and H. Liu, "Variability of the oceanic mixed layer, 1960−2004," J. Clim. **21** (5), 1029−1047 (2008).
https://doi.org/10.1175/2007JCLI1798.1

30. K. Lorbacher, D. Dommenget, P. P. Niiler, et al., "Ocean mixed layer depth: A subsurface proxy of ocean-atmosphere variability," J. Geophys. Res. **111,** C07010 (2006).
https://doi.org/10.1029/2003JC002157

31. C. Wang and L. Zhang, "Multidecadal ocean temperature and salinity variability in the tropical North Atlantic: Linking with the AMO, AMOC, and subtropical cell," J. Clim. **6** (16), 6137−6162 (2013).
https://doi.org/10.1175/JCLI-D-12-00721.1

32. N. A. Diansky and P. A. Sukhonos, "Multidecadal variability of hydro-thermodynamic characteristics and heat fluxes in North Atlantic," in *Physical and Mathematical Modeling of Earth and Environment Processes. Springer Geology.* (Springer, Cham, 2018), pp. 125−137.
https://doi.org/10.1007/978-3-319-77788-7_14

33. D. Kieke and I. Yashayaev, "Studies of Labrador Sea Water formation and variability in the subpolar North Atlantic in the light of international partnership and collaboration," Prog. Oceanogr. **132,** 220−232 (2015).
https://doi.org/10.1016/j.pocean.2014.12.010

34. N. J. Mantua, S. R. Hare, Y. Zhang, et al., "A Pacific interdecadal climate oscillation with impacts on salmon production," Bull. Amer. Meteorol. Soc. **78** (6), 1069−1079 (1997).

35. G. V. Alekseev, N. I. Glok, A. V. Smirnov, and A. E. Vyazilova, "The influence of the North Atlantic on climate variations in the Barents Sea and their pre-

dictability," Russ. Meteorol. Hydrol. **41,** 544–558 (2016).

36. V. V. Ivanov and I. A. Repina, "The effect of seasonal variability of Atlantic water on the Arctic Sea ice cover," Izv., Atmos. Ocean. Phys. **54** (1), 65–72 (2018).

37. D. R. Cayan, "Latent and sensible heat flux anomalies over the northern oceans: Driving the sea surface temperature," J. Phys. Oceanogr. **22** (8), 859–881 (1992).

38. V. Slonosky and P. Yiou, "Does the NAO index represent zonal flow? The influence of the NAO on North Atlantic surface temperature," Clim. Dyn. **19** (1), 17–30 (2002).
https://doi.org/10.1007/s00382-001-0211-y

39. E. Zorita, V. Kharin, and H. von Storch, "The atmospheric circulation and sea surface temperature in the North Atlantic area in winter: their interaction and relevance for Iberian precipitation," J. Clim. **5** (10), 1097–1108 (1992).

40. R. J. Greatbatch, "The North Atlantic Oscillation," Stochast. Environ. Res. Risk Assessment **14** (4–5), 213–242 (2000).
https://doi.org/10.1007/s004770000047

41. M. Latif, A. Grötzner, M. Münnich, et al., in *Decadal Climate Variability,* vol. 44 of *NATO ASI Series (Series I: Global Environmental Change)*, Ed. by D. L. T. Anderson and J. Willebrand (Springer, Berlin, Heidelberg, 1996).
https://doi.org/10.1007/978-3-662-03291-6_6

42. S. Peng and J. Fyfe, "The coupled patterns between sea level pressure and sea surface temperature in the midlatitude North Atlantic," J. Clim. **9** (8), 1824–1839 (1996).

43. C. K. Folland, J. Knight, H. W. Linderholm, D. Fereday, S. Ineson, J. W. Hurrell, "The summer North Atlantic Oscillation: past, present, and future," J. Clim. **22** (5), 1082–1103 (2009).
https://doi.org/10.1175/2008JCLI2459.1

44. P. Chang, L. Ji, and H. Li, "A decadal climate variation in the tropical Atlantic Ocean from thermodynamic air–sea interactions," Nature **385** (6616), 516–518 (1997).
https://doi.org/10.1038/385516a0

45. S. P. Xie, "A dynamic ocean–atmosphere model of the tropical Atlantic decadal variability," J. Clim. **12** (1), 64–70 (1999).
https://doi.org/10.1175/1520-0442-12.1.64

Translated by O. Ponomareva

Numerical Modeling of Sea Waves

D. V. Chalikov[a, b, *]

aShirshov Institute of Oceanology of Russian Academy of Science, Moscow, 197110 Russia
bUniversity of Melbourne, Victoria, 3010 Australia
*e-mail: dmitry-chalikov@yandex.ru

Abstract—The paper describes characteristics of numerical models of surface waves based on full equations for the flow with free surface in potential approximation, as well as their efficiency and applications. A more detailed description is given for the model using the surface following coordinate system. In such coordinate system Laplace equation for the potential transforms into a full elliptical equation that is solved as Poisson equations with iterations in the right-hand side. The method of solution based on separation into analytical solution and the deviation from it is offered. Such approach significantly accelerates the calculations. The problem on the whole is solved for the periodical in both horizontal directions Fourier domain using calculation of the nonlinear terms on a dense grid. The simulation of the wave field development under the action of wind and dissipation was carried out. It is demonstrated that transformation of the basic integral characteristics has a satisfactory agreement with the known data. All possible applications of the approach developed are described.

1. INTRODUCTION

The terms used in current paper need explanations. The surface waves can be one-dimensional (unidirected) or two-dimensional, i.e. running in different directions. One-dimensional waves are invariant to lateral shift and they are described by two-dimensional equations which are solved in vertical plane $(x - z)$. The vertical structure of solution in conformal coordinates [1] is known, hence it is acceptably to call the conformal model of one-dimensional waves also as one-dimensional model. Two-dimensional wave field is described by three-dimensional equations and till now nobody could reduce this problem to two-dimensional one. Method of surface integral [2] operates with surface variables only, however the method is based on use of Green function and it is difficult to attribute it to two-dimensional methods.

The first models simulating the flow with free surface on the basis of a certain type of the Lagrangian approach were formulated in the papers [3–7]). The method was based on the tracing of a surface position on a fixed grid with different degrees of accuracy. However, such method was found to be insufficiently exact, since it could be used but for a short period of time. Still, the method is valuable as it can be used for investigation of the non-stationary flow with free (not necessarily single-value) surface on the basis of the equations for the two-phase flow (see [8]).

Over the past two decades the problem of the numerical modeling has been rapidly developed. As a proof, the following papers using different modifications of Petrov-Galerkin method, can be cited [9–11]. Most of such models are targeted at solution of various technical problems such tsunami; waves caused by landslides; interaction of waves and coast or engineering constructions, etc.

The degree of complexity in wave modeling depends on the assumptions taken. The condition of periodicity significantly simplifies the formulation of the model and numerical scheme, because in this case Fourier representation providing exact differentiation can be used. We are not inclined to assign a particular physical meaning the Fourier modes; therefore, we consider them as one of the ways of interpreting results. The Fourier transform is mostly used as a method of numerical solution of equations. The periodicity of variables can be interpreted as a property of a smaller object as compared to a big basin with the characteristics slowly changing in space. The periodicity allows us to use the exact grid Fourier method (Fourier-transform), which gives the possibility to carry out a long-term integration and investigate the transformation of the wave field due to the nonlinear interactions as well as the energy input and dissipation. For the non-periodical objects such as small continental basins or smaller parts of big basins with the known boundary conditions, the global Fourier representa-

tion is inapplicable, while the numerical scheme must use finite differences. In this case it is possible to allow for the specific boundary conditions (absorption, reflection or partial reflection), real bathymetry and bottom processes. On the whole, such models cannot be very exact. Still, such approach provides more detailed results than the spectral models.

The two-dimensional approaches (for example, a method based on the conformal transformation) are convenient for investigation of many specific problems (for example, for investigation of wave breaking), but they are not quite realistic because the strictly one-dimensional waves in the presence of disturbances quickly obtain a two-dimensional structure. The complications of the solution of the three-dimensional potential equations happen due to the fact that the three-dimensional problem cannot be reduced to the two-dimensional one. The method of surface integral [2, 12−18] formally uses surface variables only, but its realization is so complicated that it can be applied but to comparatively simple situations. The main advantage of the method is that it does not impose any limitations on the surface steepness and therefore can be used for investigation of such complex phenomenon as wave breaking. We presume that such method is hardly applicable to the modeling of the multimode wave field over a large area.

At present the most popular is HOS model (High Order Scheme, [19, 20 based on the papers [21, 22]) and on the formulation of the equations offered by Zakharov [23]. The model uses two coordinate systems: the surface following curvilinear coordinate system for kinematic and dynamic boundary conditions and the Cartesian coordinate system where the analytical solution of Laplace equation for the velocity potential is known. The transfer from one coordinate system to another is done using Tayler series in the Cartesian coordinate system. The accuracy and efficiency of this method depend on the number of used terms of Taylor series. For the waves of small amplitude and narrow wave spectrum the method accuracy can be quite high. The modeling of waves with the broad spectrum requires a great number of terms of Taylor series, which indeed can be the cause of numerical instability. The difficulties arise because disturbances of the velocity potential for smaller waves propagating over the surface of big waves attenuate with depth quite quickly at the depth comparable with the height of a big wave. The simulation of the wave field with HOS model using short Taylor series is carried out successfully due to filtration of short waves, which increases the stability of the scheme. The HOS model is widely used in investigations (see, for example [24−26]. It was announced recently (Ecole Centrale Nantes, LHEEA Laboratory CNRC) that HOS model became available for common use [27].

Another group of models, unlike HOS model, is based on the direct solution of the three-dimensional equation for the velocity potential written in the surface-following curvilinear coordinate system. The most universal approach is being successfully developed at Technical University of Denmark, TUD (see [28]). The *ModelWave3D*, developed at TUD, is intended for investigation of a wide scope of problems including the problem of wave interaction with the floating and submerged objects. Due to its universal character, the model can be used for oceanographic research work, such as simulation of the wave regime in small basins with real shape and bathymetry. The effect of bathymetry is introduced by use of the so-called sigma-coordinate straightening the bottom and free surface. On the coast line either the boundary conditions of absorption or reflection, or mixed conditions are used. The comparison of TUD model with HOS model for the relatively simple wave processes has given a good agreement.

The TUD model is indeed a prominent achievement of the engineering programming. However, the importance of the model as an achievement of geophysical hydrodynamics has not been formulated so far, since the model does not describe the input and dissipation of energy. Perhaps, it is suggested that the modeling on the basis of the adiabatic variant of the model should be carried out, though no data on maintaining stability and filtration of disturbances are provided. It is true that in the description of HOS model there is no detailed information of this kind either. The TUD model, being mostly targeted at the engineering problems, does not seem to be very convenient for the description of wave field transformation due to the nonlinearity, energy input and dissipation for a long period of time. Such processes can be reliably described using Fourier method only on the assumption of the process periodicity or on the basis of finite-difference schemes of high order.

All of the above-mentioned papers mostly suggested the quasi-stationarity, i.e. the constancy of total energy.

The processes connected with the change of total energy are a lot more difficult for modeling because waves are conservative and their energy changes for hundreds and thousands of periods. The only approach allowing us to simulate evolutions of the wave field in space and time is the spectral approach where waves as physical objects are absent, while the density of potential energy as a function of frequency, direction, physical coordinates and time are the only variable. The description of physical transformations of energy in the spectral approach is borrowed from other branches of wave mechanics that consider the waves as they are. However, only the spectral models allow us to describe wave processes in space and time in the ocean.

The *phase-resolving* or *direct* models considered in this paper cannot compete with the spectral models, since they are able to simulate wave processes over the

area of just several tens of square kilometers containing several thousands of big waves. Still, the direct wave modeling plays an ever increasing role in practical oceanography, because it simulates "the research laboratory" for studying the processes that cannot be simulated by the spectral models.

The comparison of laboratory measurements in the wave or wind-wave channels with the results of modeling plays an important role in investigation. This method, for example, was used in [24] for investigation of instability and breaking of single waves. In the paper [21] the solution on the basis of Cauchy integral was compared to the exact measurements in the channel, and the results showed quite a good agreement.

2. MAIN RESULTS OF DIRECT MODELING

2.1. One-Dimensional Modeling

The non-stationary conformal transformation for finite depth allows us to rewrite the equation of potential waves in the surface-following coordinate system. Due to the conformity, Laplace equation for the velocity potential obtains a standard representation on the basis of Fourier expansion on the surface. As a result, the system of equations without simplifications turns into a system of two evolutionary equations that can be solved with traditional methods, such as Fourier-transform method, for example. In this case the assumption of potentiality simplifies the problem so dramatically that the equations can be solved using no finite differences when space derivatives are calculated analytically, while the nonlinear terms are calculated on a dense grid with the estimated accuracy. For the limited order of nonlinearity this method is also exact. The model provides a rare example in hydrodynamics when the equations describe a real process with high accuracy. This assertion is exact unless the surface steepness in the Cartesian coordinates moves to infinity. Theoretically, in case of a simply connected area the conformal mapping exists, and the surface remains one-valued; though, to preserve the accuracy a quickly growing number of modes are required. However, the practice shows that the surface may incline by an angle exceeding 90°. Then follows the physical instability caused by a drop of a certain volume of liquid due to the pressure loss. If no specific artificial measures are taken, (see, for example, [29]), a quick instability of wave crest develops [30], which finally becomes apparent as a separation into two phases. It is easy to understand that these processes are non-potential. Therefore, if it is desirable that the calculations should be continued, then, as in many branches of geophysical hydrodynamics, it is necessary to undertake special measures to prevent numerical instability and parameterize the processes connected with the breaking event. After that the model loses its transparency and mathematical rigor but becomes closer to real processes.

The development of the methods of calculation of stationary solutions of equations of potential motion with free surface: deep-water gravitational waves (Stokes waves [1]), shallow-water gravitational waves, gravitational-capillary waves and capillary waves (Crapper waves, [31]) can be considered as the first result obtained with the conformal model. These results are described in the papers [1, 32]. The numerous methods of calculations of stationary solutions developed earlier were based on the poorly converging expansions and were quite inefficient. The schemes based on the conformal mapping allow us to calculate stationary solutions with a very high accuracy and speed exceeding that of the traditional schemes by two orders. The existence of three types of stationary solutions for full equations of potential motion with free surface provides a unique possibility for verification of non-stationary models. To carry out the calculations, one of the stationary solutions is selected as the initial conditions; after that, time integration is carried out with the full stationary model. If the model is correct (including the numerical scheme and programming) and the stationary solution shows stability to disturbances, then the wave assigned should move with a correct phase velocity without changing its shape.

Lately, it has been widely recognized that Stokes waves are not just a curious mathematical object or the instrument of verification of numerical models. They begin to play an important role in investigation of the physics of surface waves. The traditional approach to study the nonlinear properties of the multimode wave process is based on the assumption that the real wave field can be represented as a superposition of linear waves. Such approach used in a famous work by Hasselmann [33], is taken as basic for modelling of the nonlinear transformation of wave spectrum. This essentially spectral approach suggests that phases of all the modes satisfy the linear dispersion relation and are distributed evenly and at random. In reality, the phase velocities of some modes are equal to the phase velocity of bigger waves since the wave field is a mixture of free waves and bound modes. The bound modes are not waves but just the "bricks" used for construction of a nonlinear wave shape (therefore, the term "bound waves" obscures the essence). Even the visual observations of sea surface confirm that real waves are sharp, while their troughs are smooth.

It is a general opinion that the quick mechanism (as compared to the Hasselmann's one) of wave spectrum transformation is suggested by the theory of modulational instability known as the theory of Benjamin-Feir instability [34]. The essence of this theory is quite simple: in the one-dimensional train of waves the new modes develop, which results in development of compact wave spectrum. According to the results obtained by the authors of the theory as well as the results of the numerical calculations [35], the development of new modes is comparatively slow. The generally accepted scenario of the extreme wave development suggests

that one of the spectrum modes for some unknown reason begins to take energy from the spectral environment, slowly grows and finally reaches quite high values. There is no any proof that such process can occur in nature or laboratory. The numerical experiments with the exact three-dimensional model on the contrary show that generation of the extreme wave occurs very quickly—over the time of the order of one period.

The numerical investigation of the wave breaking process [37] proved that neither of the dynamic, kinematic or geometric characteristics can be considered as a reliable predictor of wave breaking; therefore, the evolution of the above characteristics describes the breaking itself but does not explain its mechanism. Meanwhile, even visual observations reveal that wave breaking is always preceded by sharpening of wave crests. This effect cannot be explained just by the crest instability [30] that relates to the waves with critical steepness, while sharpening of wave crests followed by energy focusing occurs even at small steepness.

The effects of one-dimensional and two-dimensional focusing were investigated numerically and in laboratory [37−39]. In the papers cited it was noted that at merging of wave crests with different wave numbers the wave height can exceed the height available at linear superposition. Such focusing, almost for sure, is the main cause of wave breaking. The calculations with the conformal model [40] have fully confirmed the laboratory results.

2.2 Three - Dimensional Modeling

Certainly, the method of conformal variables cannot be generalized for the three-dimensional motion. Most of the models constructed for the three-dimensional waves are based on simplified equations. On the whole it is not quite clear which effects are missing in such models.

From the point of view of physics the problem of direct (*phase-resolving*) modeling can be divided into two groups: the adiabatic and non-adiabatic modeling. A simple adiabatic model suggests the absence of energy input and dissipation. This approach being not quite free from the restrictions allows us to investigate wave processes using exact initial equations. Accounting for the effects of energy input and dissipation is most often related to the hypotheses contradicting the assumption of potentiality, i.e. the terms added to the equations can be referred to the phenomenological ones. Due to this fact the formulations of the non-adiabatic approach quite often appear to be different.

The direct modeling uses the methods of computational mathematics; therefore it inherits all the advantages and disadvantages of those methods: on one side it is the ability for a detailed description of the processes, while on the other side it is the whole complex of problems connected with the numerical instability and contradictions between the desired resolution and computational capabilities. The mathematical modeling yields huge volumes of information, the processing of which is quite often no less complicated than the modeling itself.

The direct modeling of wave processes takes a lot of computer time since it has to use the surface following coordinate system, which makes the equations more complex. A fortunate exception is the conformal model that appears a lot simpler than the initial equations in the Cartesian coordinate system. In many other cases we have to solve the full elliptical equation for the velocity potential following from Laplace equation. However, the problem of the three-dimensional wave modeling that seemed impossible twenty years ago, is now solved, though not so quickly but within the limits of reasonable time.

In the two-dimensional models, (conformal, for example) a strongly idealized situation is considered, since even the monochromatic waves in the presence of lateral disturbances quickly obtain a two-dimensional structure. The three-dimensional modeling is a lot more complicated. The difficulties arise not because one more dimension is added and the problem becomes more cumbersome by orders. The basic difficulty is the impossibility of reducing the three-dimensional problem to the two-dimensional one, as well as the transformation of Laplace equation in the surface following coordinates into the general elliptical equation. The difficulty is so great that for a long time the three-dimensional modeling was carried out using simplified models that unpredictably distorted the problem.

The main types of the three-dimensional models are given in Introduction. Here a model that luckily combines the comparative straightforwardness with reasonable efficiency is considered in more detail. The model is designed (see [32, 41]) for integration of the three-dimensional full equations of flow in potential approximation with free surface with the periodical boundary conditions in horizontal directions. The model was many times discussed in the publications, so, here just its brief description is given. The non-orthogonal surface-oriented coordinate system is used:

$$\xi = x, \quad \vartheta = y, \quad \zeta = z - \eta(\xi, \vartheta, \tau), \quad \tau = t, \quad (1)$$

Where $\eta(x, y, t) = \eta(\xi, \vartheta, \tau)$ is the moving single-valued periodical surface. The equations of motion have the following form:

$$\eta_\tau = -\eta_\xi \varphi_\xi - \eta_\vartheta \varphi_\vartheta + \left(1 + \eta_\xi^2 + \eta_\vartheta^2\right)\Phi_\zeta, \quad (2)$$

$$\varphi_\tau = -\frac{1}{2}\left(\varphi_\xi^2 + \varphi_\vartheta^2 - \left(1 + \eta_\xi^2 + \eta_\vartheta^2\right)\Phi_\zeta^2\right) - \eta - p, \quad (3)$$

$$\Phi_{\xi\xi} + \Phi_{\vartheta\vartheta} + \Phi_{\zeta\zeta} = \Upsilon(\Phi), \quad (4)$$

where Υ is the operator:

$$\Upsilon(\ ) = 2\eta_\xi (\ )_{\xi\zeta} + 2\eta_\vartheta (\ )_{\vartheta\zeta}$$
$$+ (\eta_{\xi\xi} + \eta_{\vartheta\vartheta})(\ )_\zeta - \left(\eta_\xi^2 + \eta_\vartheta^2\right)(\ )_{\zeta\zeta}. \quad (5)$$
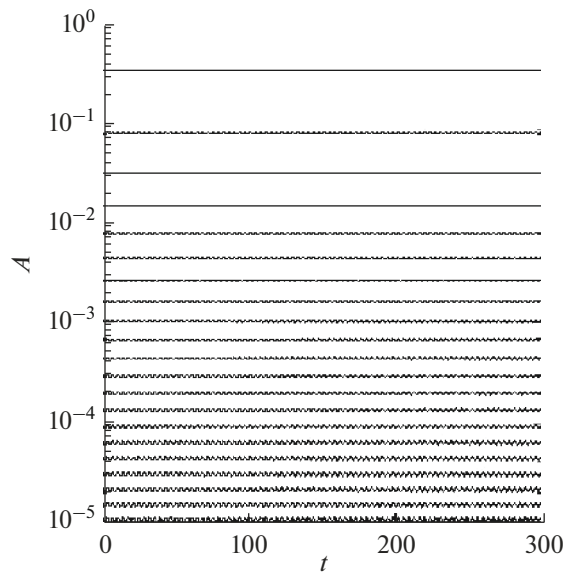
Fig. 1. Dependence of amplitudes A of Stokes waves on time t (expressed in wave periods).

The capital letter Φ is used for domain $\zeta < 0$, while a small letter φ refers to surface $\zeta = 0$. The term p describes the pressure on the surface $\zeta = 0$. The equations are written in non-dimensional form which follows from suggestion that gravity acceleration is equal to 1. In the paper [41] it is suggested that the velocity potential Φ (φ) should be represented as a sum of the analytical $\overline{\Phi}$, $(\overline{\varphi} = \overline{\Phi}(\xi, \vartheta, 0))$ and arbitrary nonlinear $\tilde{\Phi}$, $(\tilde{\varphi} = \tilde{\Phi}(\xi, \vartheta, 0))$ components.

$$\varphi = \overline{\varphi} + \tilde{\varphi}, \quad \Phi = \overline{\Phi} + \tilde{\Phi}. \tag{6}$$

The analytical component satisfies Laplace equation:

$$\overline{\Phi}_{\xi\xi} + \overline{\Phi}_{\vartheta\vartheta} + \overline{\Phi}_{\zeta\zeta} = 0, \tag{7}$$

with the known solution

$$\overline{\Phi}(\xi, \vartheta, \zeta, \tau) = \sum_{k,l} \overline{\varphi}_{k,l}(\tau) \exp(|k|\zeta) \Theta_{k,l}, \tag{8}$$

here k and l are wave numbers in directions and y, $|k| = \left(k^2 + l^2\right)^{1/2}$, $\overline{\varphi}_{k,l}$ are Fourier coefficients for the surface potential $\overline{\Phi}$ at $\zeta = 0$; $\Theta_{k,l}$ are basic functions of Fourier expansion. The nonlinear component of the potential satisfies Poisson equation:

$$\tilde{\Phi}_{\xi\xi} + \tilde{\Phi}_{\vartheta\vartheta} + \tilde{\Phi}_{\zeta\zeta} = \Upsilon(\Phi), \tag{9}$$

Equation (9) is solved with boundary conditions:

$$\varsigma = 0 : \tilde{\Phi} = 0,$$
$$\varsigma \to -\infty : \tilde{\Phi}_\zeta \to 0. \tag{10}$$

The advantage of separation of the velocity potential into two components is that the nonlinear component by one or two orders is less than the linear one; therefore, the accuracy required for solution of the equation (9) is achieved at a smaller number of iterations than it is necessary for the full velocity potential. The fields of elevation and surface potential satisfying the linear equations at assigned wave spectrum are taken as the initial conditions. A more detailed description of the model is given in the book [32].

The model described in the above mentioned papers was lately improved by small modifications of the finite-difference scheme and the schemes of parameterization of energy input and dissipation. The adiabatic version of the last variant of the model was verified by modeling the motion of Stokes wave with the limit steepness $AK = 0.40$. The initial conditions for Stokes wave were calculated by the algorithm developed in the paper [1].

Figure 1 demonstrates an evolution of amplitudes of the Stokes modes calculated with the three-dimensional model. As seen from the figure, at least the first 15 modes (including the modes with the amplitude of the order of 10^{-4}) are preserved with a very high accuracy; the modes with the amplitudes of $10^{-6} - 10^{-4}$ preserve their identity; just the modes with the amplitudes of $A < 10^{-6}$ become chaotic. Note that the structure of Stokes wave is quite complicated: it consists of an infinite number of modes, each of them moving with the speed of the first mode. Even the minor errors in the formulation of the problem or in the codes lead to the quick wave breaking. Obviously, it does not happen. Such verification is exact and non-trivial. The first cycle of the works based on the model $(1-10)$ was devoted to the study of adiabatic processes in the wave fields, i.e. the processes occurring with no energy input and dissipation. In the nonlinear hydrodynamic system with the finite number of degrees of freedom the flux of energy is mostly directed towards the small wave numbers. To prevent instability in this range, the selective filter is introduced. To maintain the quasi-stationary state, a very slow change of energy E (order's $\left(10^{-9} - 10^{-7}\right) E$ for one time step) is compensated with the integral energy input. In such regime the model is used for investigation of many problems of the wave mechanics, particularly, the problem of extreme wave statistics.

(1) The statistics of full wave height (from crest to trough) was found to be quite identical for the linear and nonlinear models [42]. The nonlinearity reveals itself only in the statistics of wave heights and trough depth, singly. This fact has no explanation as yet.

(2) It is revealed that in the process of development wave spectrum obtains strong inhomogeneity: deep holes in the two-dimensional spectrum alternate with well-pronounced peaks [41].

It is interesting to note that that with increase of resolution the low-energy spectrum areas widen. The resolution used appears to be excessive. This statement sounds too optimistic and hardly plausible. However, it can be justified by the following reasoning. It is evident that the multimode wave dynamics is different from the multiscale turbulent motion. The increase of resolution in the course of modeling of turbulence allows us to take into account smaller structures, but within reasonable limits it does not significantly influence a reproduction of the flow structure on the whole. With increase of resolution in the wave field other modes with their own phase velocities are introduced. Due to dispersion, the geometry and evolution of the wave field with the same spectrum and doubled resolution will be different. It means that the numerical wave problem does not satisfy the basic requirement of convergence at increase of resolution. The situation can be saved if the limit resolution exists; when it is exceeded, some of the modes do not contribute to the motion because their energy is too small. It can be suggested, for example, that very closely spaced modes cannot exist separately and quickly merge into one.

The existing three-dimensional models were mostly used for simulation of fast wave processes or the quasistationary regime at which wave spectrum did not change significantly. A unique example of calculation of the surface wave evolution is given in the paper [42] where on the basis of the two-dimensional model the development of wave field under the action of wind was calculated. This experience was partially successful but not fully convincing because developing waves are always two-dimensional, i.e. the wave field consists of the modes spreading at different angles to the wind direction.

When wind velocity exceeds the phase velocity of main energy-containing waves, the energy of waves grows, while wave spectrum shifts towards the low wave numbers. This process was simulated many times with the spectral models to specify the schemes for calculation of the rate of energy input from wind and dissipation. The modeling of wave development with the phase-resolving models is a lot more complicated problem, at any rate, because the calculation of wave evolution should be carried out at least for hundreds of peak wave periods. For the essentially three-dimensional problem where at each step Poisson's equations are solved with iterations, such calculations are carried out over tens of hours of computer time. To solve such a problem, the algorithms describing the input of energy and momentum and energy dissipation should be included in the equations (1–10). The testing and tuning of such schemes for the non-stationary cumbersome problem are quite a time-consuming process. It should be admitted that despite a huge number of papers in the physics of sea waves, the development of the problem remains unsatisfactory. For example, the phenomenon of wave dissipation through the wave breaking is an object of the two-phase hydrodynamics.
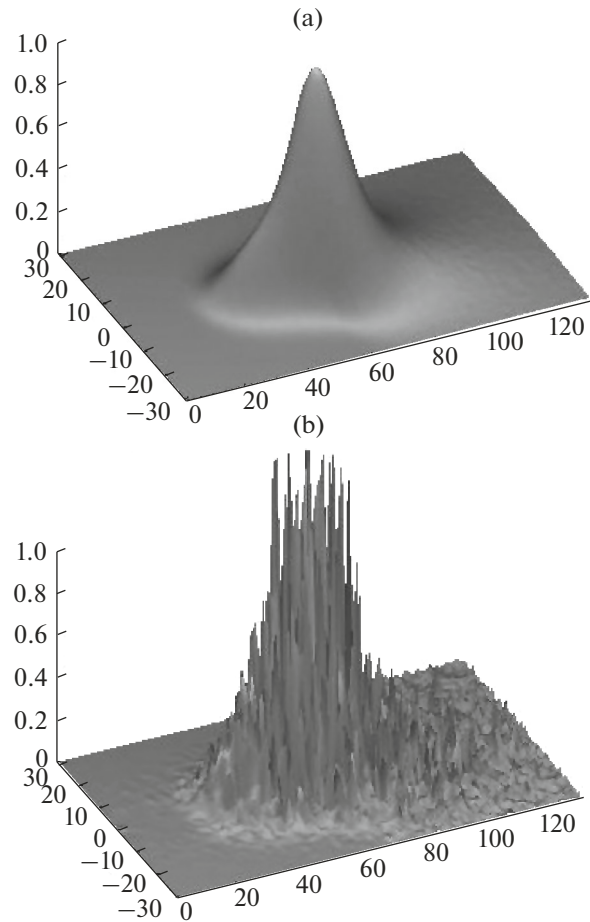

Fig. 2. Two-dimensional wave spectra in the wave-number space (the right axis corresponds to k while the left one—to l). The upper section demonstrates the spectrum corresponding to the initial conditions. The bottom section shows wave spectrum after integration for 318 peak wave periods. The spectra are normalized for a maximum value of the undisturbed spectrum.

There are no universally recognized reliable algorithms of parameterization of this phenomenon. The problem of wave and wind interaction is more complicated than the wave modeling problem itself, since it requires fulfilment of extremely complicated measurements or the numerical modeling of interaction of air flow and multimode wave field [43, 44].

An exchange of energy and moment between water and air occurs through inhomogeneous distribution of the dynamic pressure on the surface p_0 in (3). According to the generally accepted opinion, the complex Fourier components of pressure p_k and p_{-k} on the surface are connected with the components of elevation h_k and h_{-k}:

$$p_k + ip_{-k} = \frac{\rho_a}{\rho_w}(\beta_k + i\beta_{-k})(h_k + ih_{-k}), \qquad (11)$$

where β_k and β_{-k} are real and imaginary coefficients of the β-function introduced by Miles [45]. At present, it is generally assumed that the β-function depends on a single parameter $\Omega = \omega_k U \cos\psi = U/c_k \cos\psi$ (where ω_k and U are the nondimensional frequency and wind velocity; c_k is the phase velocity of k-th mode; ψ is an angle between the wind directions and that mode. Most of the schemes recommended for calculation of this function are based on the strongly simplified analytical formulas (see, for example, [46]). The linear assumption (11) is already significantly simplified; therefore, the use of further simplifications cannot be somehow justified. The study of the β-function should be based on the simultaneous recording of the fields of elevation η and surface pressure p. Such exceptionally difficult work was carried out experimentally, in field with use of the surface-following air pressure sensor [47, 48]. The volume of the data obtained by this method does not allow us so far to plot the β-function in a sufficiently wide range. Another method of study of the β-function is to carry out numerical modeling of wave and wind interaction [49]. This method enables us to obtain quite a big volume of data necessary to define the β-function, though the quality of such investigations certainly depends on the accuracy of the problem formulation, especially, the formulation of the problem of closure for Reynold's equations. The first version of the algorithm of calculation of the flux of energy to waves was introduced into the forecasting model WAVEWATCH (see [45, 46]. The next version of the model describing the interaction of the boundary layer and waves [43], gave most reliable results, also as a result of the processing of a huge volume of data covering a wide range of nondimensional frequency Ω. The β-function obtained in the calculations is used for modeling of wave development [50].

The potential waves in the absence of energy input are conservative in theory, i.e., in the process of their movement the sum of potential and kinetic energy should remain constant. This property is not observed for numerical wave models due to the approximation errors and energy transition to subgrid disturbances. The approximation errors are insignificant when Runge−Kutta method and Fourier transform are used, so the main cause of non-conservation of energy is the nonlinearity of the equations due to which the energy flux is generated in the range of high and low wave numbers. The growth of wave amplitudes at high wave numbers forms the areas with very high surface steepness and causes a fast and dramatic instability. Such phenomenon is eliminated by use of a high-selective filter simulating the non-linear viscosity [50]. The main purpose of this algorithm is maintaining monotonicity and smoothness of wave spectrum in the range of high wave numbers. Since very small amplitudes are subjected to filtration, it does not actually influence energy preservation but efficiently eliminates the numerical instability.

The basic process of the wave energy dissipation is wave breaking [51]. Whatever the form, the wave breaking is taken into account in all of the wave models; in particular, in wave forecasting models of a WAVEWATCH type. Since the waves themselves are missing in such models, no local criteria of breaking can be defined; hence, the wave dissipation is represented in the spectral models in a distorted way. The wave breaking occurs in a comparatively narrow range. In the spectral models the spectrum of breaking actually spreads over the entire spectral range, i.e. all the waves including the short ones are weakened. In reality the breaking lowers height and energy of main waves. The contradiction arises because waves in the spectral models are assumed to be linear, while in reality it is the nonlinear sharp wave that is subject to breaking.

The interface instability resulting in breaking is an important and poorly-developed branch of hydromechanics. This essentially nonlinear process should be considered in a general case for the two-phase flow. At present one can hardly find a full description of this process (though a certain progress is noticeable, See [8]). Still, without considering its basic effects it is impossible to fulfil the realistic wave modeling. The breaking onset is to some degree similar to the loss of hydrostatic instability. By analogy, it can be assumed that appearance of the non-single value in some of the surface areas can be an obvious sign of breaking. Such control can be carried out in the conformal coordinates where the dependence of elevation on the longitudinal coordinate is always single-valued; therefore, the non-single value in the Cartesian coordinates is easily recognizable. Then a part of the flow not supported by pressure begins to move downwards. The process ends in local smoothing of the surface and decrease of its steepness and curvilinearity. An excess of the physical horizontal speed of particles over the wave phase velocity can be considered as quite a pronounced criterion of breaking. However, the detailed numerical experiments [36] revealed that no simple and reliable criterion of breaking ever exists.

The mechanism of wave breaking at developed spectrum in the sea strongly differs from that in the wave field represented by several modes as it is assumed in the theoretical and laboratory investigations. To a certain degree, the wave breaking is similar to development of the extreme wave that also appears suddenly, with no pronounced prehistory. Neither of the events exhibits any signs of the modulational instability manifesting itself in growth of the linear wave taking energy from other modes. The situation regarding breaking and appearance of the extreme ('freak') waves is rather opposite: the energy of the single wave is preserved while the columnar energy is concentrated round the crest of the wave that becomes more sharpened and unstable. These particular considerations became basic when choosing a degree of the wave profile curvature as a criterion of breaking (see [41, 50]).

The description of breaking in the direct modeling should satisfy three conditions: (1) to prevent breaking and the subsequent numerical instability; (2) to realistically describe the losses of potential and kinetic energy; (3) to preserve the volume. The parameterization of breaking can be based on a highly-selective local diffusion operator with the appropriate diffusion coefficient [32]. Such algorithm does not change the volume and decreases the kinetic and potential energy of waves. It is assumed that the momentum leakage is transferred to the horizontal flow, while the energy is partially transferred to the turbulence and current [52]. Besides, the energy and momentum can be transferred to other wave modes. The choice of parameters in the scheme is based on the straightforward considerations: waves should approach breaking as close as possible but instead of breaking they should be subject to smoothing.

The algorithms of the energy input and dissipation were included in the model (1−10), after which the model was initiated for a long-term calculation. For the calculations a grid was used with the number of knots equal to 2048×1024, while the number of levels for calculation of the nonlinear component totaled 10. The stretched grid was used $\Delta z_{i+1} = \gamma \Delta z_i$, where $\gamma = 1.2$ is the stretch coefficient. The depth of domain H was taken equal to $2\pi k_p$, where k_p is a current value of the wave number of the spectrum peak. As the initial conditions JONSWAP spectrum [53] was taken with a maximum at wave number $k_p = 100$. The initial waves were almost unidirectional: the angle resolution in a range of the energy-containing waves was assigned proportional to $(\cos \psi)^{256}$ (ψ—an angle between the direction of wave mode and wind direction). The calculations were fulfilled for 1200000 time steps, with the step $\Delta t = 0.005$.

The full specific energy of the wave field $E = E_p + E_k$ (E_p is potential energy; E_k is kinetic energy) was calculated by the following formulas:

$$E_p = 0.25\overline{\eta^2}, \quad E_k = 0.5\overline{\overline{(\varphi_x^2 + \varphi_y^2 + \varphi_z^2)}}, \qquad (12)$$
$$E = E_p + E_k,$$

where one line denotes the averaging with respect to the coordinates ξ and ϑ, while a double line denotes the averaging with respect to the entire volume. An equation of evolution of the integral energy E looks in the following way:

$$\frac{dE}{dt} = \overline{I} + \overline{D_b} + \overline{D_t} + \overline{N}, \qquad (13)$$

where $\overline{I}$—the rate of energy input from wind to waves; $\overline{D_b}$—the rate of energy dissipation due to wave breaking; $\overline{D_t}$—the rate of dissipation due to the high-frequency filtration; $\overline{N}$ is the integral effect of the nonlinear inter-

actions described by the right-hand sides of the equations (2), (3), on condition that the surface pressure p is equal to zero. The differential form of the energy transformation components might be derived from equations (2)−(5), but no one seems to have succeeded in it yet. A more straightforward method free from the errors of approximation is to calculate the increase of energy by different groups beyond the main calculation cycle, using the fictitious time step. For example, the value of the energy input is calculated by a ratio:

$$\overline{\overline{I}} = \frac{1}{\Delta t}\left(\overline{\overline{E^{t+\Delta t}}} - \overline{\overline{E^t}}\right), \qquad (14)$$

Where $\overline{\overline{E^{t+\Delta t}}}$ is the integral energy obtained after one time step with the right-hand side of the equations (2), (3) including surface pressure only.

An evolution of the integral characteristics of the wave field is shown in Fig. 3a. The fast change of all the characteristics at $t < 500$ is explained by adaptation of the primary linear fields to the nonlinearity. The integral effect of the nonlinear interactions $\overline{\overline{I}}$ (straight line 1) was found to be close to zero. The high-frequency dissipation $\overline{\overline{D_t}}$ (curve 2) is significantly weaker than the dissipation due to breaking $\overline{\overline{D_t}}$ (curve 3). The value of $\overline{\overline{D_b}}$ is found to be strongly variable because the breaking process displays a tendency for intermittency. The total dissipation $\overline{\overline{D_b}} + \overline{\overline{D_t}}$ compensates most of the incoming energy, and only a small part of it increases energy. The energy balance $\overline{\overline{B}} = \overline{\overline{I}} + \overline{\overline{D_t}} + \overline{\overline{D_b}}$ (curve 5) is approaching zero by the end of integration. Some additional characteristics of the solution are presented in section 3b. Curve 1 describes an evolution of the weighted peak frequency ω_w:

$$\omega_w = \left(\frac{\int kSdkdl}{\int Sdkdl}\right)^{1/2}, \qquad (15)$$

where the integral is taken over the entire Fourier domain. The value of ω_w does not depend on the spectrum details, therefore its evolution well reproduces the spectrum development (downshifting). Curve 2 describes an evolution of frequency of a spectrum maximum. The stepped shape of this curve exhibits the basic property: the spectrum shifts not monotonically from the wave number to the nearest wave number, but rather step-wise; for this particular case—by three or four wave numbers through decrease of the previous maximum and growth of a new one. It is obviously closely connected with the discrete structure of the spectrum shown in Fig. 1. Curiously enough, such phenomenon was observed in the spectral model [54].
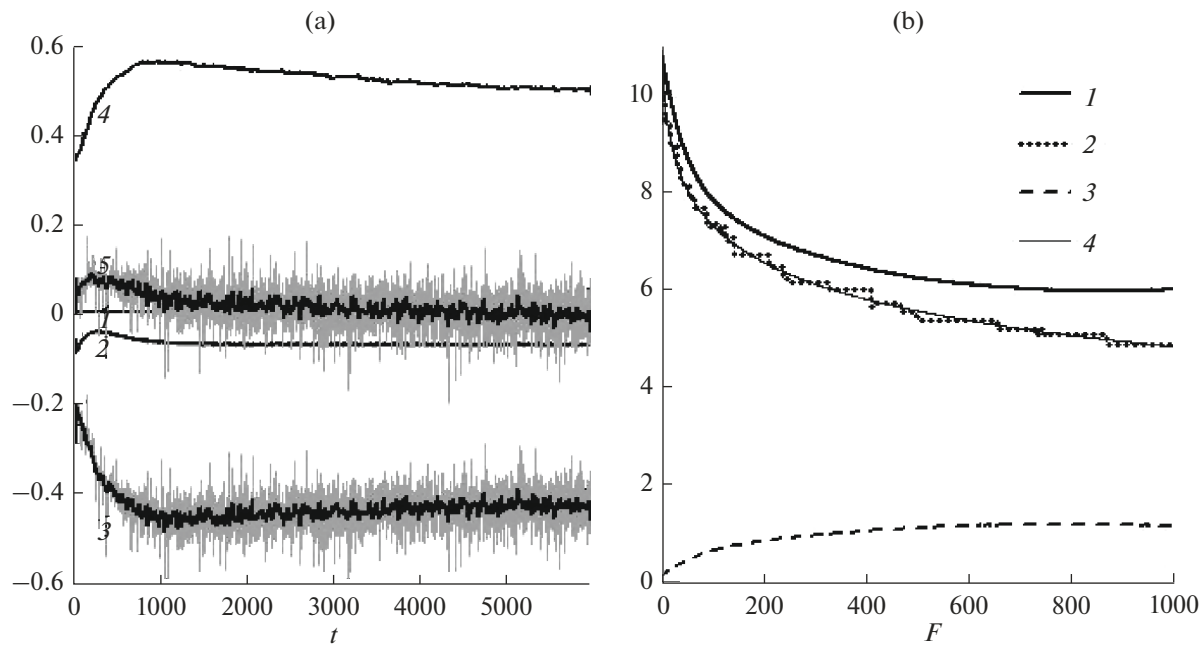
Fig. 3. (a) Rates of integral wave energy evolution multiplied by 10^9 due to: (*1*) nonlinear interactions; (*2*) dissipation in a high-frequency part of the spectrum; (*3*) wave breaking; (*4*) energy input. Curve *5* describes the balance of all the components. Grey segments show instantaneous values, while solid curves show the result of smoothing; (b) dependence of integral characteristics on fetch F: (*1*) weighted frequency at a maximum of spectrum; (*2*) frequency of a maximum of spectrum; (*3*) total wave energy multiplied by 10^6; (*4*) approximation (17).

The value of fetch F for the periodical problem can be estimated by time integration of the phase velocity $c_p = |k|^{-1/2}$ at peak of spectrum

$$F = \int_{t_0}^{t} c_p dt. \qquad (16)$$

The numerical experiments carried out on wave field development under the action of constant wind (except for the inconsistencies imposed by the condition of periodicity) agree with the JONSWAP field experiment [53]. The processing of the results showed that the frequency of spectral peak ω_p decreases as $F^{-1/3}$, while full energy linearly grows with fetch F. Neither of the dependencies can be exact: the first is singular at $F = 0$, while both dependencies do not take into account the oncoming of the regime of saturation. In our case the more exact formulation describing dependence ω_p of frequency on fetch is as follows:

$$\omega_p = \frac{75.6}{5.63 + F^{1/3}} \qquad (17)$$

(curve *4*). The formula $\omega_p \sim F^{-1/3}$ is valid in a narrower interval F. The coefficients in (17) certainly depend on the wind velocity. The linear dependence of energy E on fetch F is not altogether confirmed, since the model simulates a more extended interval of fetch as compared to that observed in the experiment.

3. CONCLUSION

The phase-resolving (direct) three-dimensional models are a powerful instrument of investigating the properties of real waves. Opposite to the theoretical investigations based on simplified equations, the results obtained on computer can be considered almost like observations in laboratory or field, with one valuable exception though: the numerical data practically contain no noise or disturbances that are always present in the world of reality. The model gives the possibility to "observe" almost any characteristics of the wave field with a high resolution in space and time. The wave modeling fell behind other branches of geophysical hydrodynamics, and particularly, technical hydrodynamics. The numerical modeling can be compared to the ideal laboratory modeling with the vast possibilities of parameter variations and "observations" of any properties of the object. Since the mathematical modeling is free from scale differences, the accuracy of simulation of natural processes is higher than that of the physical modeling. This situation is typical for other branches of hydrodynamics, especially for technical applications. Very often the numerical methods turn out to provide higher accuracy As

compared to the experiments. It also valid for wave experiments in wave channels. Such experiments are often devoted to the investigations of the nonlinear wave processes. Here we would like to make a note. Suppose we use the two-dimensional model and have a proof that the equations are solved exactly. We try to simulate with the model the physical experiment in the wave channel and obtain one of the following results:

(a) a rather good agreement with the results obtained. If not to pay attention to the ever-existing desire of authors to present their results in the best possible way, it can be concluded that the physical experiment was carried out in compliance with the theory of potentiality within the frame of the parameters compared.

(b) The agreement turned out to be quite poor. In this case the bugs in codes are possible. If no bugs are found, it means there is a mismatch between the model and experiment. If the model is well verified, it is difficult to improve it. Hence, the experiment may be good but it does not match the model. The problem is that the investigator never knows exactly the wave characteristics, since the surface is generated while the second variable, i.e., the velocity potential, remains unknown, with the control over the second fundamental variable, i.e. the potential, being absent. The mutual adjustment of the actual surface to the unknown potential is accompanied by appearance of additional unwanted modes that make the agreement worse or even absent. The uncertainty grows when the multimode or angle-distributed processes are investigated. The relatively small sizes of wave channels prevent from investigating the processes in big time and space scales. Therefore, to study the extreme waves or the processes of wave breaking, quite often a strong artificial interaction of waves with the specifically chosen amplitudes and phases is provided.

It should be also noted that the experimental data can be compared to the specially tuned model that exactly simulates experiment, which is the way of generating and dumping waves. The model using the conditions of periodicity is obviously not suitable for quantitative comparison with the experiment in a basin.

It follows from what was said above that in most cases we cannot expect a quantitative agreement of the results of modeling and experiment. The degree of agreement of the results allows us to judge of the experiment quality or applicability of potential approximation but not of the quality of the potential model whose formal accuracy is proved in other ways.

Our remark regards the simple processes where an object is the non-breaking waves themselves. The more complicated problems such as interaction of waves and wind or interaction of waves and sea constructions or floating objects are equally the subject of modeling and experiment complementing one another.

The direct modeling is too complicated to compete with the spectral modeling. However, it gives the possibility to investigate various problems of wave physics such as the mechanism of growth and attenuation; the mechanics and statistics of the extreme waves and wave breaking; the nonlinear interactions of waves and many others. Such results are targeted at improvement of the engineering and forecasting wave modeling. The direct modeling can be already used for simulation of the wave regime in small basins.

REFERENCES

1. D. Chalikov and D. Sheinin, Direct modeling of one-dimensional nonlinear potential waves, in *Advances in Fluid Mechanics,* Ed. by W. Perrie (Computational Mechanics Publications, MA, 1997), vol. 17, p. 207.

2. W. Craig and C. Sulem, "Numerical simulation of gravity waves," J. Comput. Phys. **108,** 73−83 (1993).

3. F. H. Harlow and E. Welch, "Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface," Phys. Fluids **8,** 2182−2189 (1965).

4. W. F. Noh and P. Woodward, SLIC (simple line interface calculation), in *Lecture Notes in Physics* (Springer, New York, 1976), vol. 59, p. 330−340.

5. C. W. Hirt and B. D. Nichols, "Volume of fluid method for the dynamics of free surface," J. Comput. Phys. **39,** 201−225 (1981).

6. A. Prosperetti and J. W. Jacobs, "A numerical method for potential flow with free surface," J. Comput. Phys. **51,** 365−386 (1983).

7. H. Miyata, "Finite-difference simulation of breaking waves," J. Comput. Phys. **5,** 179−214 (1986).

8. A. Iafrati, "Numerical study of the effects of the breaking intensity on wave breaking flows," J. Fluid Mech. **622,** 371−411 (2009).

9. X. Zhao, B.-J. Liu, S.-X. Liang, and Z.-C. Sun, "Constrained Interpolation Profile (CIP) method and its application," J. Ship Mech. **20.** 393−402 (2016).

10. D.-L. Young, N.-J. Wu, and T.-K. Tsay, "Method of fundamental solutions for fully nonlinear water waves," in *Advances in Numerical Simulation of Nonlinear Water Waves,* Ed. by Q. Ma (World Scientific, Singapore, 2010), pp. 325−355.

11. Q. W. Ma and S. Yan, "Qale-FEM method and its application to the simulation of free responses of floating bodies and overturning waves," in *Advances in Numeri-*

cal Simulation of Nonlinear Water Waves, Ed, by Q. Ma (World Scientific, Singapore, 2010), pp. 165−202.

12. J. T. Beale, "A convergent boundary integral method for three-dimensional water waves," Math. Comput. **70** (335), 977−1029 (2001).

13. M. Xue, H. Xu, Y. Liu, and D. K. P. Yue, "Computations of fully nonlinear three-dimensional wave and wave−body interactions. I. Dynamics of steep three-dimensional waves," J. Fluid Mech. **438**, 11−39 (2001).

14. S. Grilli, P. Guyenne, and F. Dias, "A fully nonlinear model for three-dimensional overturning waves over arbitrary bottom," Int. J. Numer. Methods Fluids **35**, 829−867 (2001).

15. D. Clamond and J. Grue, "A fast method for fully nonlinear water wave dynamics," J. Fluid. Mech. **447**, 337−355 (2001).

16. D. Clamond, D. Fructus, J. Grue, and O. Kristiansen, "An efficient method for three-dimensional surface wave simulations. Part II: Generation and absorption," J. Comput. Phys. **205**, 686−705 (2005).

17. D. Fructus, D. Clamond, J. Grue, and O. Kristiansen, "An efficient method for three-dimensional surface wave simulations. Part I: free space problems," J. Comput. Phys. **205**, 665−685 (2005).

18. C. Fochesato, F. Dias, and S. Grilli, "Wave energy focusing in a three-dimensional numerical wave tank," Proc. R. Soc A **462**, 2715−2735 (2006).

19. G. Ducrozet, F. Bonnefoy, D. Le Touze, and P. Ferrant, "3-D HOS simulations of extreme waves in open seas," Nat. Hazards Earth Syst. Sci. **7**, 109−122 (2007).

20. G. Ducrozet, H. B. Bingham, A. P. Engsig-Karup, F. Bonnefoy, and P. A. Ferrant, "Comparative study of two fast nonlinear free-surface water wave models," Int. J. Numer. Methods Fluids **69**, 1818−1834 (2012).

21. D. Dommermuth and D. Yue, "A high-order spectral method for the study of nonlinear gravity waves," J. Fluid Mech. **184**, 267−288 (1987).

22. B. West, K. Brueckner, R. Janda, D. Milder, and M. R. Milton, "A new numerical method for surface hydrodynamics," J. Geophys. Res. **92**, 11803−11824 (1987).

23. V. E. Zakharov, "Stability of periodic waves of finite amplitude on the surface of deep water," J. Appl. Mech. Tech. Phys. **9**, 190−194 (1968).

24. M. Tanaka, J. W. Dold, M. Lewy, and D. H. Peregrine, "Instability and breaking of a solitary wave," J. Fluid Mech. **187**, 235−248 (1987).

25. A. Toffoli, M. Onorato, E. Bitner-Gregersen, and J. Monbaliu, "Development of a bimodal structure in ocean wave spectra," J. Geophys. Res. **115** (C3) (2010). https://doi.org/10.1029/2009JC005495

26. J. Touboul and C. Kharif, "Two-dimensional direct numerical simulations of the dynamics of rogue waves under wind action," in *Advances in Numerical Simulation of Nonlinear Water Waves,* Ed. by Q. Ma (World Scientific, Singapore, 2010), pp. 43−74.

27. G. Ducrozet, F. Bonnefoy, D. Le Touzé, and P. Ferrant, "HOS-ocean: Open-source solver for nonlinear waves in open ocean based on High-Order Spectral method," Comput. Phys. Commun. **203**, 245 (2016). https://doi.org/10.1016/j.cpc.2016.02.017

28. A. P. Engsig-Karup, H. B. Bingham, and O. Lindberg, "An efficient flexible-order model for 3D nonlinear water waves," J. Comput. Phys. **228**, 2100−2118 (2009).

29. J. W. Dold, "An efficient surface-integral algorithm applied to unsteady gravity waves," J. Comput. Phys. **103**, 90−115 (1992).

30. M. S. Longuet-Higgins and M. Tanaka, "On the crest instabilities of steep surface waves," J. Fluid Mech. **336**, 51−68 (1997).

31. G. D. Crapper, "An exact solution for progressive capillary waves of arbitrary amplitude," J. Fluid Mech. **96**, 417−445 (1957).

32. D. Chalikov, Numerical investigation of wave breaking, in *Numerical Modeling of Sea Waves* (Springer, Cham, 2016), p. 137. https://doi.org/10.1007/978-3-319-32916-1

33. K. Hasselmann, "On the non-linear energy transfer in a gravity wave spectrum, Part 1," J. Fluid Mech. **12**, 481−500 (1962).

34. T. B. Benjamin and J. E. Feir, "The disintegration of wave trains in deep water," J. Fluid Mech. **27** (3), 417−430 (1967).

35. D. Chalikov, "Simulation of Benjamin-Feir instability and its consequences," Phys. Fluids **19** (1), 016602−016615 (2007).

36. D. Chalikov and A. V. Babanin, "Simulation of wave breaking in one-dimensional spectral environment," J. Phys. Oceanogr. **42** (11), 745−1761 (2012).

37. T. B. Johannessen and C. Swan, "A numerical transient water waves − part 1. A numerical method of computation with comparison to 2-D laboratory data," Appl. Ocean. Res. **19**, 293−308 (1997a).

38. T. B. Johannessen and C. Swan, "A laboratory study of the focusing of transient and directionally spread surface water waves," Proc. R. Soc. A **457**, 971−1006 (1997b).

39. T. B. Johannessen and C. Swan, "On the nonlinear dynamics of wave groups produced by the focusing of surface-water waves," Proc. R. Soc. A **459**, 1021−1052 (2003).

40. D. Chalikov and A. V. Babanin, "Nonlinear sharpening during superposition of surface waves," Ocean Dynamics **66** (8), 931−937 (2016a).

41. D. Chalikov, A. V. Babanin, and E. Sanina, "Numerical modeling of three-dimensional fully nonlinear potential periodic waves," Ocean Dynamics **64** (10), 1469−1486 (2014).

42. D. Chalikov and A. V. Babanin, "Simulation of one-dimensional evolution of wind waves in a deep water," Phys. Fluids **26** (9), 096697 (2014).

43. D. Chalikov and S. Rainchik, "Coupled numerical modelling of wind and waves and the theory of the wave boundary layer," Boundary-Layer Meteorol. **138**, 1−41 (2010).

44. P. P. Sullivan, J. C. McWilliams, and E. G. Oatton, "Large-eddy simulation of marine atmospheric bound-

ary layers above a spectrum of moving waves," J. Atmos. Sci. **71** (11), 4001−4027 (2014).

45. J. W. Miles, "On the generation of surface waves by shear flows," J. Fluid Mech. **3,** 185−204 (1957).

46. P. A. E. M. Janssen, "Quasi-linear theory of wind-wave generation applied to wave forecasting," J. Phys. Oceanogr. **21** (11), 1631−1642 (1991).

47. M. A. Donelan, A. V. Babanin, I. R. Young, M. L. Banner, and C. McCormick, "Wave follower field measurements of the wind input spectral function. Part I. Measurements and calibrations," J. Atmos. Oceanic Tech. **22,** 799−813 (2005).

48. M. A. Donelan, A. V. Babanin, I. R. Young, and M. L. Banner, "Wave follower field measurements of the wind input spectral function. Part II. Parameterization of the wind input," J. Phys. Oceanogr. **36,** 1672−1688 (2006).

49. D. V. Chalikov, "Numerical simulation of the boundary layer above waves," Boundary-Layer Meteorol. **34** (1−2), 63−98 (1986).

50. D. Chalikov, "Numerical modeling of surface wave development under the action of wind," Ocean Sci. **14,** 453−470 (2018).
https://doi.org/10.5194/os-14-453-2018

51. A. V. Babanin, *Breaking and Dissipation of Ocean Surface Waves* (Cambridge University Press, Cambridge, 2011).

52. D. Chalikov and M. Belevich, "One-dimensional theory of the wave boundary layer," Boundary-Layer Meteorol. **63,** 65−96 (1993).

53. K. Hasselmann, R. P. Barnett, E. Bouws, et al., "Measurements of wind-wave growth and swell decay during the Joint Sea Wave Project (JONSWAP)," Dtsch. Hydrogr. Z. **A8** (12), 1−95 (1973).

54. W. E. Rogers, A. V. Babanin, and D. W. Wang, "Observation-consistent input and whitecapping-dissipation in a model for wind-generated surface waves: description and simple calculations," J Atmos. Oceanic Tech. **29** (9), 1329−1346 (2012).

An Overview of Parameterezations of Heat Transfer over Moss-Covered Surfaces in the Earth System Models

V. M. Stepanenko[a, b, *], I. A. Repina[a, c], V. E. Fedosov[d], S. S. Zilitinkevich[e], and V. N. Lykossov[a, f]

[a]Research Computing Center, Moscow State University, Moscow, 119991 Russia
[b]Faculty of Geography, Moscow State University, Moscow, 119991 Russia
[c]Obukhov Institute of Atmospheric Physics, Russian Academy of Sciences, Moscow, 119017 Russia
[d]Faculty of Biology, Moscow State University, Moscow, 119991 Russia
[e]Finnish Meteorological Institute, P.O. BOX 503, FI-00101, Helsinki, Finland
[f]Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, 119333 Russia
*e-mail: stepanen@srcc.msu.ru

Abstract—Moss cover plays an important role in shaping the thermodynamic and biogeochemical functioning of ecosystems at high latitudes, where it is the dominant vegetation type. It reduces the heat transfer through the soil–atmosphere interface and between deep levels of the active layer, which leads to a decrease in the thickness of the seasonally thawed layer, and regulates the soil moisture. At the same time, in Earth system models, the heat and moisture exchange in the moss cover is represented by simplified approaches. The purpose of this paper is to summarize up-to-date knowledge on heat-transfer mechanisms acting in moss cover and methods for their quantitative description; we also aim to identify the data missing for constructing physically justified parameterizations and formulate tasks for the future theoretical and experimental work. This paper provides a brief overview of the heat-transfer parameterizations for the moss–lichen cover used in modern Earth system models; presents the results of experimental and theoretical studies of the thermal conductivity of soils and mosses; and strives to encompass available information on the dynamic and thermal roughness of plant communities, including mosses. In conclusion, recommendations are given on the development of parameterizations of thermodynamic processes in the moss cover and design of desirable model and field experiments.

INTRODUCTION

One of most serious consequences of climate warming is the degradation of permafrost. In such a case, the organic substance that was earlier fixed in frozen soil starts to decay and generate greenhouse gases, mainly CO_2 and CH_4 that escape into the atmosphere. In Earth system models, it is important to adequately take into consideration the positive feedback created by this process with the temperature growth in the troposphere. Here, the most important factor determining energy and mass fluxes across the underlying surface is the vegetation cover, which is mainly represented by mosses (bryophytes) and lichens at high latitudes (Fig. 1).

Higher plants that constitute the overwhelming majority of the ground primary production are subdivided into two large groups: vascular plants and bryophytes. The number of bryophyte forms is generally smaller than the vascular plant variety, and its distribution at high and middle latitudes is relatively regular, which suggests that the origin and distribution of this group of organisms are linked to comparatively cold epochs [1–3]. The major properties of the bryophytes allowing them to exist successfully in conditions of severe climate of high latitudes are (1) small sizes, (2) the ability to endure almost complete tissue dehydration without losing their viability, (3) slow physiological processes, (4) vegetation in dense sods, and (5) a high prevalence of vegetative reproduction and groundbreaking and ruderal strategies [4–6]. Due to these capabilities, the bryophytes dominate over the vascular plants in biomass and production in harsh environmental conditions of the Arctic and highlands. Thick moss sods are typical for zonal tundra and north taiga vegetation communities. In this regard, mosses apparently play an important role in the creation of the thermodynamic and biogeochemical functioning condition of these ecosystems. Hence, it is shown that, at high latitudes, the moss cover reduces heat exchange
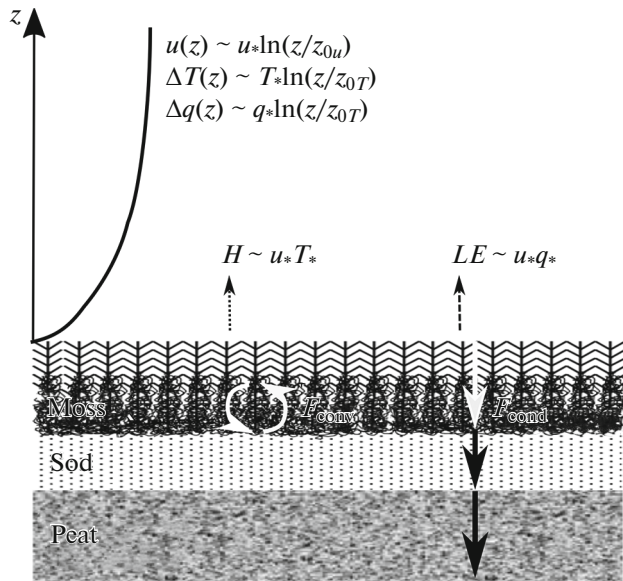
Fig. 1. Momentum, heat, and moisture transfer in the "near-ground layer–active soil layer" system in the case of moss presence using the example of a moorland ecosystem. Standard denotations of the near-ground layer theory are applied (for simplicity, for neutral stratification); F_{cond} and F_{conv} are the molecular heat transfer and the convective heat transfer in a moss cover, respectively.

between the soil surface and the atmosphere, which causes a decrease in the thickness of the seasonally thawed layer and limits the carbonic gas emission [7−10].

Mosses are also known by their low moisture conductivity, to say the least, due to poor moisture suction by roots. This circumstance is important for the soil moistening condition at high latitudes and, in particular, promotes keeping the highland moors waterlogged. However, for reasons of space of the paper, this role of moss is not considered here. The results of physical and mathematical simulation of the moisture transfer in mosses can be found, for example, in [11, 12].

The purpose of this paper was to estimate modern knowledge about the mechanisms of heat exchange in the moss cover and the ways to quantitatively describe them, as well as to identify missing data for the construction of physically justified parametrizations and formulate problems for the theoretical and experimental studies in the nearest future.

The paper is organized as follows. In Section 1, a short review of parametrizations of heat transfer in the moss and lichen cover is considered within the modern Earth system models. Section 2 is devoted to experimental data and theoretical approaches to the description of soil and moss thermal conductivity accumulated so far in soil science and biology. In Section 3, the empirical and theoretical knowledge of the dynamic and thermal roughness of the ecosystem, including mosses, are presented. In the Conclusions, the main outcomes of the study are formulated and the recommendations for the future research are presented.

1. PARAMETRIZATIONS OF HEAT TRANSFER IN THE MOSS COVER USED IN EARTH SYSTEM MODELS

Nowadays, the specificity of heat and moisture exchange in the moss cover in the Earth system models is represented by simplified approaches. For example, in the LSM land surface scheme (Land Surface Model, [13]), the moss is presented by a single (upper) layer in the finite difference grid of the soil model with a porosity of 0.9 and constant values of heat capacity and thermal conductivity of the moss body (the authors do not present references for these values). A high sensitivity of the soil thermal regime and energy fluxes in the underlying surface to the presence of this layer is shown. In the JULES (Joint UK Land Environment Simulator, [14]) active-layer model, the option of partial cover of the model cell with moss is considered. The moss thickness is set by the constant of 5 cm, and the thermal conductivity coefficient is calculated as a linear function of moisture content according to [9]. The introduction of the moss layer and other improvements of the model have considerably increased the quality of simulation of the permafrost seasonal-thawing thickness. In the CLASS (Canadian Land Surface Scheme [15]) Canadian land model, a 10-cm thickness of the moss layer is set in the moor block. The thermal conductivity coefficient is also considered a linear function of moisture according to the experimental data [16]. The same dependence is assumed in the active layer scheme of the ORCHIDEE model (Organizing Carbon and Hydrology In Dynamic Ecosystems [17]). In the JSBACH land model of the Institute of meteorology, Max Planck Institute of Meteorology [10], it is supposed that the heat transfer in the moss and lichen layer (4.5 cm in thickness) can be described using the geometrical averaging of the thermal-conductivity coefficients and the normalized thermal conductivity concept (via the Kersten numbers; see Section 2.1); here the thermal conductivity mechanism is considered molecular. When the bryophyte layer was considered in the model, the soil upper layer temperature to the north of the 50° north latitude decreased by 2.7 K on average. In the next sections of this paper, it is shown that the moss thermal-conductivity relationship only with the moisture content assumed in the Earth system models does not consider other important physical factors.

2. PHYSICAL MECHANISMS OF HEAT TRANSFER IN THE MOSS COVER

Data of natural measurements show that thermal-insulation properties of the moss cover are determined primarily by its thickness, density, and waterlogging [9]. Depending on the moistening degree and a variety of other environmental factors, the species composition, the moss projective cover, and the moss sod thickness vary greatly [18−20]. The moss can be absent on sites exposed to intensive erosion or snow and wind corra-

sion, or it can comprise 100% of the cover and thickness up to 20 cm and more in conditions favorable to peat accumulation. The ecological diversity allows a stable coexistence of 150–200 bryophyte species even under conditions of plain Arctic tundra [20]. The overwhelming majority of papers devoted to the moss-cover contribution to the heat exchange between the soil and the atmosphere at high latitudes (including the active layer models, see Section 1) do not consider the complexity of the spatial structure of the moss cover. In this overview, the aspect of horizontal inhomogeneity is not considered either, since this issue requires a special discussion.

Below, physical mechanisms of the molecular and turbulent thermal conductivity of the moss layer and their mathematical description are considered.

2.1. Molecular Thermal Conductivity

In the case of moss vegetation, the upper portion of the active layer of land can be divided into three parts: moss cover, sod, and organic layer. The two latter are low-porosity media, which implies a low velocity of interstitial moisture and gas flows and, consequently, predominately molecular mechanism of thermal conductivity. In the moss layer, due to its considerable porosity, air circulation is possible, which is considered in the next subsection 2.2. The available information about the molecular heat transfer in the moss cover is presented in this paragraph.

Few papers are devoted to the experimental definition of the moss thermal conductivity. Thus, N.A. Soudzilovskaia et al. [9] measured the thermal conductivities of 17 moss species in natural and laboratory conditions. It was shown that the thermal-conductivity coefficient weakly depended on the moss density and linearly depended on its moisture content. The authors concluded that biological processes did not render a noticeable influence on thermal conductivity. They showed the following species-specific differences of the thermal insulation effect: top-fruiting mosses forming tight cushionlike sods (*Polytrichum commune*, *P. strictum*, and *Sphagnum fuscum*) reduce the mean amplitude of the soil temperature variations much more than other considered species (from 10 to 4 K as compared with bare ground for *Sphagnum fuscum* over the considered measurement time interval). The authors described a similar effect for side-fruiting mosses that formed long carpets, *Pleurozium schreberi* and *Hylocomium splendens*, the major dominant species of the ground cover in zonal taiga forests and tundra. On the contrary, rather high thermal conductivity is characteristic for another typical moss of zonal tundra, *Aulacomnium turgidum* (the temperature amplitude decreased from 7 to 5 K as compared with the bare surface). Further, this effect can be considered and taken into account based on the classification of the bryophytic growth forms [18]. According to [9], the removal of the moss cover from the natural surface

weakly changes the mean surface temperature. On the other hand, in [21], the decrease in mean temperature of the soil under the moss cover by 0.9–2.1°C was shown (which agrees with the results of simulation presented in [10]) and a reduction in the seasonally thawed layer thickness by 9–20 cm is reported. Similar conclusions were made in [7] according to the results of a series of numerical experiments within the one-dimensional heat and moisture transfer simulation in the "soil–vegetation–snow cover" system developed at the Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, using the data from systematic observations at the Franklin Bluffs station (Alaska) and at four meteorological stations in the northern and central parts of Siberia. The authors of [22] came to a conclusion on the sharp increase in the heat and evaporation fluxes on the surface after the removal of the moss cover and, thus, proved the significance of heat and moisture insulating functions of moss in ecosystems of high latitudes. In [23], similar estimations were supplemented with the demonstration of influence of the moss-cover thickness on the depth of seasonal thawing of permafrost soils. In [24], based on the measured data, the influence of the moss cover on the temperature regime of the permafrost in the West Spitsbergen was considered; experimental data were supplemented with the results of calculations performed within the regional and global climate models at different snow depths and moss-cover thicknesses. It was shown that the decrease in the ground temperature under the moss cover by few degrees could offset the air temperature increase due to climate change and prevent permafrost from degradation. The materials of experimental research of the thermophysical properties of the bryophytes and their effect on the active layer of the Arctic tundra soils were presented in [25]. The empirical dependence of the moss thermal conductivity coefficient on its moisture during the cold and warm seasons of the year was obtained. In good agreement with the results of [9], the linear dependence of the moss thermal conductivity coefficient on its volumetric moisture content was also revealed in [16]. The dependence of the moss thermal conductivity and specific heat capacity on its moisture and density was also confirmed by the results of studies of moss-cover samples by the set stationary heat flux and calorimetric methods in [26]. However, as far as we know, there were no theoretical papers on determining moss molecular thermal conductivity.

Theoretical approaches to determining the molecular thermal conductivity coefficient in a porous natural environment are well-developed in soil science (see, for example, [27]). In 1912, O. Wiener [28] considered an medium consisting of n materials with different thermal conductivity coefficients and the given shares of total volume. It was found that the thermal conductivity coefficient reached the maximum when the materials formed layers along the temperature gradient, and it reached the minimum when the layers

were distributed in the perpendicular direction. In the first case, the thermal conductivity coefficient of medium is the sum of coefficients k_i weighed by volume fractions α_i:

$$k_{\max} = \sum_i \alpha_i k_i, \qquad (1)$$

in the second case, the reciprocal values of the coefficients are weighed in the same way:

$$k_{\min} = \left(\sum_i \frac{\alpha_i}{k_i} \right)^{-1}. \qquad (2)$$

Hence, the linear dependence of the moss thermal conductivity on the volumetric moisture content [9, 16] can serve an indication for the vertical orientation of pores in the moss cover due to the vertical arrangement of the moss structural elements. Between the abovementioned extreme values of thermal conductivity, given by (1) and (2), there is the *geometrical average* of the thermal-conductivity coefficients of materials composing the soil [29]:

$$k = \prod_i k_i^{\alpha_i}. \qquad (3)$$

The majority of modern agrological heat transfer models are based on this averaging [30−33]. Thus, the thermal-conductivity coefficient of the solid part of the soil is averaged over the values for quartz (the most heat-conductive mineral of the soil), other minerals, ice [34], and organic matter [35]; the thermal-conductivity coefficient of the soil saturated with liquid moisture is averaged over the corresponding coefficients for water and the solid part. In the case of unsaturated soil, the heat transfer also occurs due to water and water-vapor motion driven by the difference of matrix potential and the gradient of concentration, respectively. In addition, for the different liquid moisture contents, the topologies of liquid and gas distribution in the fixed soil skeleton also differ; at its simplest, they are expressed by the tortuosity and connectivity characteristics. As a result, the geometrical averaging of the thermal-conductivity coefficient for unsaturated soil gives an unsatisfactory agreement with observations. In this regard, O. Johansen [29] introduced the concept of a normalized thermal conductivity coefficient (the Kersten number):

$$Ke = \frac{k - k_d}{k_{sat} - k_d}, \qquad (4)$$

whose relation to the volumetric moisture content is subject to the experimental determination (k_d is the thermal-conductivity coefficient of dry soil and k_{sat} is the correspondent coefficient of the water saturated soil). Some researchers [29, 31, 35] proposed functional forms of this dependence using empirical coefficients. In [36], the Cote and Konrad dependence was applied to calculate the thermal conductivity of the moss and lichen cover [31]; however, no measure-

ments of thermal-conductivity coefficient were carried out.

As far as the authors of this review know, no theoretical (semiempirical) models of the molecular thermal conductivity in porous media have been verified with empirical data for the moss cover. The following parameters should be measured to validate these models: (1) moss-cover porosity, (2) volumetric moisture content, (3) volume fraction and mineral part content in the sod layer, and (4) thermal conductivity coefficient of moss components (leaves, stems, and roots). The empirical thermal conductivity coefficient (or temperature conductivity) of the moss layer can be estimated both in a laboratory using specialized instruments and in natural conditions by the damping of daily temperature variations with depth.

2.2. Turbulent Thermal Conductivity

The experimental data about heat transfer due to air circulation in the top part of the moss layer are also scarce; apparently, they are presented only in two papers. In [37], the transfer of heat and water vapor in the *Racomitrium Canescens* moss layer, which was used to cover building roofs to decrease the urban heat-island effect, was studied in laboratory conditions. The shear turbulence in the airflow was controlled by the Reynolds number and the convective turbulence was controlled by the Grashoff number. It was shown that the ratio of the maximal value of the moss thermal-conductivity coefficient to its minimal value, attained by varying the relation between the Reynolds and the Grashoff numbers, was about six. It is also important that the extreme values of the Bowen relation differed eightfold. These data indicated a considerable variability of the heat exchange conditions inside the moss cover and on its boundary with the atmosphere depending on the momentum and energy fluxes on the surface; this aspect was not considered in earlier papers [9, 22].

In [38], a three-dimensional temperature field that formed inside the moss cover under the influence of evaporation from the surface at different insolation was measured in laboratory conditions. A considerable nonuniformity of this field (differences up to 5°C) was revealed at horizontal inside the moss layer, and this magnitude weakly depended on the plant species. In some experiments, the Rayleigh number values inside the cover exceeded critical values, but the vertical heat flux that formed was not measured directly. Using the empirical relations of the Nusselt number (Nu) with the Grashoff number known from the literature, the authors estimated Nu ≈ 2.2−5.6.

In [39], the problem of convective instability of the submerged part of the moss layer was considered based on the analysis of the linearized problem and numerical solution of the Darcy equations. The Rayleigh number critical values were found, and it was shown

that the simulation results agreed with the data of specially designed experiments. The results of the paper were limited by the assumption of stationarity of the equation of motion. The authors also employed the free convection approximation assuming that the water level was inside the moss cover, so that the momentum flow on its surface was not large. For a moss layer above the water level, similar analysis should be performed taking into consideration the friction stress on the upper boundary [37].

The theory of convection in porous media can be taken as the basis for parametrization of the circulating heat and mass exchange in the moss cover [40]. In the case of free convection, the theoretical and semiempirical relations between the Nusselt and Rayleigh numbers are found. The case of shear convection is more general; this scenario is thoroughly investigated both numerically and experimentally for the Poiseuille type currents (caused by a horizontal pressure gradient), while the Couette type current in porous medium (with the velocity or momentum flux imposed on the upper boundary) is a more adequate model of flow for the moss cover. Apparently, the semiempirical heat exchange laws for this current can be obtained using the results of the direct numerical simulation; this methodology is successfully applied to the Couette type flow in continuous medium (see, for example, [41]).

3. HEAT AND MOMENTUM FLUXES ON THE BOUNDARY WITH ATMOSPHERE

The Monin−Obukhov similarity theory [42] is conventionally applied to calculate turbulent fluxes of momentum (τ) and sensible heat (H) in the surface layer of the atmosphere:

$$\tau \equiv \rho u_*^2, \qquad (5)$$

$$H \equiv \rho c_p u_* T_*, \qquad (6)$$

where the velocity u_* scale is the dynamic or friction velocity and T_* is the temperature scale. Integrating the equation of the similarity theory for the mean wind velocity u

$$\frac{du}{dz} = \frac{u_*}{\kappa z} \phi\left(\frac{z}{L}\right)$$

provides the following relation connecting the wind velocity at level z with the friction velocity:

$$u = \frac{u_*}{\kappa}\left[\ln\left(\frac{z}{z_{0u}}\right) - \Psi_u\left(\frac{z}{L}\right) + \Psi_u\left(\frac{z_{0u}}{L}\right)\right], \qquad (7)$$

where κ is the von Karman constant (≈ 0.4) and $L = -\dfrac{u_*^3}{\kappa \beta H}$ is the Obukhov scale ($\beta = \dfrac{g}{T}$ is the buoyancy factor, g is the gravity acceleration). The summand $\ln\left(\dfrac{z}{z_{0u}}\right)$ determines the wind-velocity logarith-

mic profile under conditions of neutral stratification, function $\Psi_u(\varsigma)$ is the integral universal function for momentum describing the effects of the surface layer stratification, and $\varsigma = z/L$ is the dimensionless height (stability parameter).

The integration constant z_{0u} sets the length scale; the vertical size of the surface roughness elements, is proportional to this length scale, which is known as the parameter of aerodynamic roughness. The surface roughness is the crucial factor of development of turbulent conditions in the surface layer of the atmosphere: it determines the momentum, heat, moisture, and air constituents exchange between the surface and the atmosphere [43−45].

Parameter z_{0u} is defined as the height at which the mean wind velocity turns zero. However, the mean velocity is actually zero exactly on the surface and, at heights of about z_{0u}, the similarity theory (in particular, Eq. (7)) is not valid. The standard physical interpretation of the parameter z_{0u} is that this is an only parameter in Eq. (7) which describes the integral effect of the mean wind velocity profile.

The equation for the temperature gradient, i.e. acting similar to deriving Eq. (7) from the similarity theory, we get the thermal roughness coefficient z_{0T} involved:

$$T - T_s = \frac{T_*}{\kappa_T}\left[\ln\left(\frac{z}{z_{0T}}\right) - \Psi_T\left(\frac{z}{L}\right) + \Psi_T\left(\frac{z_{0T}}{L}\right)\right], \qquad (8)$$

where T_s is the surface temperature and $\Psi_T(\zeta)$ is the integral universal function for temperature. A similar equation can be readily derived for air humidity.

The important anatomical and morphological feature of mosses when compared with vascular plants is the absence or poor development of conducting elements to move up moisture from the upper soil layer and the absence of stomata in gametophytes (green plants) of mosses. In this regard, there is almost no transpiration observed in mosses. This means that the vapor transfer in the viscous sublayer is controlled by the molecular diffusion not involving biological processes[1], so it is possible to assume the closeness of roughness coefficients for the temperature and the water vapor.

Equations (7) and (8) form the basis of the algorithms of calculation of the heat and moisture exchange and friction between the atmosphere and the surface applied all modern weather forecasting models and the climate theory. The roughness coefficients z_{0u} and z_{0T} are not measured directly. These parameters are expressed in units of length; they depend on geometrical properties of the surface, friction velocity, and the coefficients of molecular viscosity and tem-

---

[1] Transpiration is included in the schemes of calculating total evaporation within the land models via the use of "stomatal resistance" concept that can be reformulated in terms of the moisture roughness coefficient.

perature conductivity. Introducing these parameters makes it possible to avoid a detailed description of wind velocity and temperature profiles in the immediate proximity to the underlying surface. Equations (7) and (8) can be considered the definitions of z_{0u} and z_{0T} providing the way to calculate these parameters. Traditionally, the parameter z_{0u} is considered stipulated only by the form of the surface. If this form is fixed, this parameter can be calculated at neutral stratification, when the universal function $\Psi_u(\zeta) = 0$ and the uncertainty associated with the setting of the form of this function is absent. The friction rate u_* is expressed via the momentum flux, which is measured by the eddy covariance method [46], so parameter z_{0u} is calculated by measurements of fluctuations of three components of the wind velocity at the same height. An alternative way is to calculate the aerodynamic roughness coefficient by the data on the mean velocity at two heights [47].

Similarly, the coefficient of thermal roughness can be found for the case of neutral stratification using the logarithmic law, by the data on the temperature and heat flux at one level, or by the data on the air temperature at two levels (assuming that the surface temperature is known). In the general case of a stratified surface layer, Eqs. (7) and (8) can be also used to calculate z_{0u} and z_{0T}. The values of z_{0u} obtained in this way and from a logarithmic profile built using the data on $u(z)$ at two levels differ insignificantly, even at greatly stable and greatly unstable stratifications, if the measurements are made at levels up to ~5 m above the surface [48]. Corrections for stratification employing functions $\Psi_u(\zeta)$ and $\Psi_T(\zeta)$ are also small at strong winds when $L \to \pm\infty$ and $\Psi_u(\zeta), \Psi_T(\zeta) \to 0$.

Upon the interaction of the surface layer with the surface covered with large elements of roughness (forest and urban area), the atmospheric flow is displaced from the underlying surface by the value of D (displacement height). In this case, height z in formulas (7) and (8) is replaced by $(z - D)$, so that, in order to find the aerodynamic roughness parameter, it is necessary to measure at least three quantities: for example, u_* at one height and $u(z)$ at two heights [49]. In the case, when the vertical size of roughness elements and the typical distance between them are comparable with L or exceed it, the most energy-carrying turbulent vortices penetrate between the roughness elements and the aerodynamic resistance of the surface increases. This suggests the z_{0u} dependence on D/L [50]; however, for dwarf vegetation (in particular, mosses and lichens), this effect, as well as the value of D, should be small.

The calculation of roughness parameters from formulas (7) and (8) in real conditions can lead to errors, because the actual profile of the wind velocity does not obey the logarithmic law even at neutral stratification (due to inhomogeneous surface and nonstationary

weather environment), but the analysis of corresponding special events is beyond the scope of this review.

Roughness parameters for the momentum and temperature (moisture) are not identical, since the momentum transfer and heat and mass exchange transfer through the rough surface are driven by different mechanisms [51]. The aerodynamic roughness parameter in all turbulent fluxes studied exceeds the thermal-roughness parameter. The momentum flux τ through the underlying surface is composed of τ_p, which is the flux caused by the difference of pressures on the windward and leeward sides of the roughness elements (so-called form resistance), and τ_v, which is the tangential stress on the surface linked to the molecular viscosity. The value of τ_p dominates over the value of τ_v, so the momentum transfer from the surface layer to the underlying surface depends primarily on the form and sizes of the streamlined irregularities and depends weakly on molecular viscosity. On the contrary, the heat or mass flux in the immediate proximity to the surface is completely controlled by the molecular thermal conductivity and diffusion, respectively, independently of the surface smoothness or roughness. Due to form resistance effect, the momentum transfer through the surface is more efficient than the transfer of scalar characteristics, which determines the inequality $z_{0u} > z_{0T}$. The difference between the molecular diffusion and viscosity coefficients plays a secondary role in the difference between z_{0u} and z_{0T}.

While theoretical methods of estimating the roughness aerodynamic parameter received wide attention in the literature, the progress in the development of similar methods for the thermal roughness parameter is much more modest. There are different approaches to describing the dependence of the temperature-roughness parameter on roughness Reynolds number $\mathrm{Re}_{z_{0u}}$, determined by the friction velocity u_*, molecular viscosity coefficient ν, and one of length scales (z_{0u} or h, which is the typical height of roughness elements). In [51, 52], it was shown that the natural logarithm of the ratio of the aerodynamic roughness to the temperature roughness should be a function of $\mathrm{Re}_{z_{0u}}^{1/2}$. On the other hand, in [44, 53, 54], it was theoretically found that a similar dependence should contain $\mathrm{Re}_{z_{0u}}^{1/4}$. This important distinction in scaling is found in the most widely used parametrizations of the temperature roughness [55–60]. Apart from these two dependences, the linear dependence of $\ln\left(z_{0u}/z_{0T}\right)$ on $\mathrm{Re}_{z_{0u}}$ [61, 62] and the square-law dependence on $\ln(\mathrm{Re}_{z_{0u}})$ [63, 64] was proposed at $\mathrm{Re} > 2.5$ for the rough and smooth sea ice. D. Li et al. showed [65] that different powers of $\mathrm{Re}_{z_{0u}}$ in the formula for $\ln\left(z_{0u}/z_{0T}\right)$ could be obtained by simplifying the advection-diffusion equation of a scalar quantity (temperature, moisture) in a viscous sublayer. Apparently, the dimensionless empirical con-
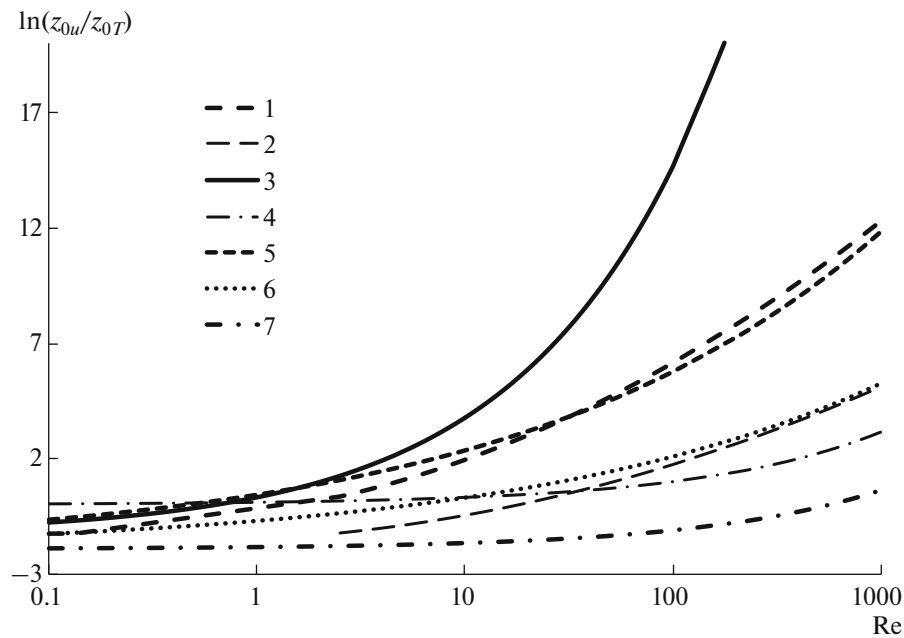
Fig. 2. Dependence of the logarithm of the ratio between the dynamic and thermal roughness coefficients on the roughness Reynolds number. The line numbers correspond to the parametrization numbers in Table 1.

stants included in the particular forms of dependence $\ln\left(z_{0u}/z_{0T}\right) = f\left(\mathrm{Re}_{z_{0u}}\right)$ depend on the surface type, which is confirmed, for example, by a considerable difference between the constants in the formula $f\left(\mathrm{Re}_{z_{0u}}\right) = a\,\mathrm{Re}_{z_{0u}}^{1/2} + b$ for the water surface [51] and high grass [66]. In Fig. 2 and Table 1, the main parametrizations of $\ln\left(z_{0u}/z_{0T}\right)$ dependence on $\mathrm{Re}_{z_{0u}}$, obtained by theoretical models and from experimental data over different types of surfaces, are presented.

Figure 2 shows the dependence of the roughness-temperature parameter on the surface type, especially at large roughness Reynolds numbers. The good agreement between parametrizations (2) and (6) (rough ice and urban area) and (1) and (5) (snow-covered smooth ice and even surface with low grass cover) confirms that the height of roughness elements is an important factor for shaping the ratio z_{0u}/z_{0T}. In the case of sea surface (3) at strong winds, the wave breaking and the layer of foam and splashes influence the exchange processes, which is most likely the reason for the strong difference of the parametrization developed for these conditions from parametrizations for other surfaces. The fact that the aerodynamic parameter of the sea surface roughness increases with the increase in the wind velocity is also significant.

The results of eddy covariance measurements presented in [67] are of interest: the thermal roughness parameters of five types of vegetation were investigated, including croplands, meadows, bushes, and forests. It was found that parameter z_{0T} considerably depended on the type of vegetation, but, in the case of dense vegeta-

tion (forest, agricultural acreages), it poorly depended on the friction velocity u_*, i.e. on $\mathrm{Re}_{z_{0u}}$. Concerning other types of vegetation, none of the power dependences on $\mathrm{Re}_{z_{0u}}$ provided satisfactory result.

Data on the roughness parameter of moss cover are scarce in the literature. In [68], the values of the moss aerodynamic roughness in the range from 0.025 to 0.04 m were presented. Large variations of the aerodynamic roughness in time were mostly explained by the seasonal change of phenophases of herbaceous plants that were quite frequent on the moorland site. Thus, the value of z_{0u} can be used as an implicit indicator of area of the herbage cover above the moss.

Considering the close relationship between the roughness parameters and the form of the underlying surface, the data on the "roughness" of the moss surface that is defined as a measure of spatial inhomogeneity of the surface height can be useful. In [69], the "roughness," interpreted as the root-mean-square deviation or the mean module of height deviation of the sphagnum moss surface, was measured in laboratory conditions by images in visible and infrared spectra. Nine moss samples with different water contents were analyzed. The influence of the moss moisture on its roughness was shown. The relation of the roughness parameter of the moss with its porosity and moisture was also mentioned in [70], where the authors proposed using the roughness as a measure of moss-cover productivity. The importance of the roughness parameter in estimating the photosynthetic activity of the moss cover and its role in biogeochemical processes were also noted in [68].

Table 1. Parametrization of the $\ln\left(z_{0u}/z_{0T}\right)$ dependence on $\mathrm{Re} = \mathrm{Re}_{z_{0u}}$

No.	Reference	Formula	Surface type
1	[64]	$\ln\left(\dfrac{z_{0u}}{z_{0T}}\right) = \begin{cases} -1.25 & \mathrm{Re} \le 0.135 \\ -0.145 + 0.55(\ln \mathrm{Re}) & 0.135 < \mathrm{Re} < 2.5 \\ -0.317 + 0.565(\ln \mathrm{Re}) + 0.183(\ln \mathrm{Re})^2 & \mathrm{Re} > 2.5 \end{cases}$	Snow covered level ice. $\mathrm{Re} > 2.5$.
2	[63]	$\ln\left(\dfrac{z_{0u}}{z_{0T}}\right) = -1.5 + 0.2(\ln \mathrm{Re}) + 0.11(\ln \mathrm{Re})^2 \; \mathrm{Re} > 2.5.$	Rough ice
3	[51]	$\dfrac{1}{\kappa}\ln\left(\dfrac{z_{0u}}{z_{0T}}\right) = \begin{cases} -2 & \mathrm{Re} < 0.1 \\ 4\,\mathrm{Re}^{1/2} - 3.2 & \mathrm{Re} \ge 0.1 \end{cases}$	Sea surface
4	[52]	$\ln\left(\dfrac{z_{0u}}{z_{0T}}\right) = 0.1\,\mathrm{Re}^{1/2}$	Theoretical calculations, even land surface
5	[44]	$\ln\left(\dfrac{z_{0u}}{z_{0T}}\right) = 2.46\,\mathrm{Re}^{1/4} - 2$	Low vegetation cover
6	[54]	$\ln\left(\dfrac{z_{0u}}{z_{0T}}\right) = 1.29\,\mathrm{Re}^{1/4} - 2$	Urban area
7	[66]	$\ln\left(\dfrac{z_{0u}}{z_{0T}}\right) = 0.08\,\mathrm{Re}^{1/4} - 1.9$	Grass cover

In [71], the results of the studies in a wind tunnel of the z_{0u} dependence on the height and sparseness of the vegetation cover were presented. It was shown that the lower the cover sparseness was, the greater the role of the displacement height in generating aerodynamic roughness was. It was found that with the increase in the cover density the roughness increased, while, with the increase in the wind velocity, the roughness decreased due to the flexibility of the vegetation cover. Experimental data on estimations of the thermal-roughness parameter for the moss are extremely scarce, but it can be expected that this parameter also depends on the form of the moss surface and, in particular, on its moistening.

CONCLUSIONS

In conclusion, it is necessary to note that, despite the well-known thermal insulation properties of the moss and lichen layer, the approaches to the heat-transfer parametrization in this layer are still underdeveloped. The experimental data make it possible to reliably assert that the moss thermal-conductivity coefficient strongly (and usually linearly) depends on the contents of liquid moisture; the fact of convection is proved for some moss species at cooling the upper boundary and the convection causes a severalfold growth of the effective thermal conductivity coefficient and greatly influences the Bowen ratio. The expressions developed for soil with changed constants are used for the thermal-conductivity coefficient when

describing the heat and moisture transfer in mosses within the Earth system models. Apparently, the convection effects are not considered here. A considerable uncertainty remains in setting the parameters of aerodynamic and thermal roughness of the moss surface. Presumably, there are no papers devoted to studies of the thermal roughness of ecosystems with a predominance of bryophytes.

In the opinion of the authors of this review, the following directions of research can be recommended for developing a physically sound mathematical description of heat transfer in a moss and lichen cover which would ensure the high accuracy of calculations of the thermal regime:

(i) field estimates of the moss thermal-conductivity coefficient based on measurements by thermometers, thermal balance plates, and conductometers;

(ii) definition of parameters of the aerodynamic and thermal roughness of the moss surface according to the eddy covariance and profile measurements in the surface layer;

(iii) development of the bryophyte classification according to porous space geometry; selecting the groups with the most probable development of the convective circulation;

(iv) application of the theory of convection in porous media for the parametrization of convective heat exchange in the moss cover; carrying out additional laboratory experiments to adjust the parametrization constants.

The discussion of some important issues concerning the implementation of physically justified bryophytes parametrization within the Earth system model was left outside the framework of this review. Thus, the heat exchange in active layer is closely connected with the moisture exchange, which in itself deserves a special review. In addition, the variety of thickness and geometry of the interior structure of different moss communities means the expediency of using digital maps of the distribution of these communities in weather-forecasting models and climate theory.

FUNDING

REFERENCES

1. A. J. Shaw, C. J. Cox, and B. Goffinet, "Global patterns of moss diversity: taxonomic and molecular inferences," Taxon **54**, 337−352 (2005).

2. T. S. Romdal, M. B. Araujo, and C. Rahbek, "Life on a tropical planet: niche conservatism and the global diversity gradient," Global Ecol. Biogeogr. **22,** 344−350 (2012).

3. R. G. Mateo, O. Broennimann, S. Normand, B. Petitpierre, M. B. Araujo, J. -C. Svenning, A. Baselga, F. Fernandez-Gonzalez, V. Gomez-Rubio, J. Munoz, G. M. Suarez, M. Luoto, A. Guisan, and A. Vanderpoorten, "The mossy North: an inverse latitudinal diversity gradient in European bryophytes," Sci. Rep. **6,** 25546 (2016).

4. R. E. Longton, *Biology of Polar Bryophytes and Lichens* (Cambridge University Press, Cambridge, 1988).

5. M. C. F. Proctor, "Physiological ecology," in *Bryophyte Biology*, Ed. by A. J. Shaw (Cambridge University Press, Cambridge, 2009), pp. 237−268.

6. H. Rydin, "Population and community ecology of bryophytes," in *Bryophyte Biology*, Ed. by A. J. Shaw (Cambridge University Press, Cambridge, 2009), pp. 393−444.

7. E. E. Machul'skaya and V. N. Lykosov, "Modeling of the thermodynamic reaction of permafrost to seasonal and interannual variations of atmospheric parameters," Izv. Ross. Akad. Nauk, Fiz. Atmos. Okeana **38** (1), 20−33 (2002).

8. M. Jorgenson, V. Romanovsky, J. Harden, Y. Shur, J. O' Donnell, E. Schuur, M. Kanevskiy, and S. Marchenko, "Resilience and vulnerability of permafrost to climate change," Can. J. For. Res. **40**, 1219−1236 (2010).

9. N. A. Soudzilovskaia, P. M. Bodegom, and J.H.C. van Cornelissen, "Dominant bryophyte control over high-latitude soil temperature fluctuations predicted by heat transfer traits, field moisture regime and laws of thermal insulation," Funct. Ecol. **27** (6), 1442−1454 (2013).

10. P. Porada, A. Ekici, and C. Beer, "Effects of bryophyte and lichen cover on permafrost soil temperature at large scale," Cryosphere **10** (5), 2291−2315 (2016).

11. J. S. Price, T. W. D. Edwards, Y. Yi, and P. N. Whittington, "Physical and isotopic characterization of evaporation from Sphagnum moss," J. Hydrol. (Amsterdam, Neth.) **369** (1−2), 175−182 (2009).

12. B. R. Voortman, R. P. Bartholomeus, P. M. Bodegom, H. Gooren, S. E. A. T. M. van der Zee, and J.-P. M. Witte, "Unsaturated hydraulic properties of xerophilous mosses: towards implementation of moss covered soils in hydrological models," Hydrol. Processes **28** (26), 6251−6264 (2014).

13. J. Beringer, A. H. Lynch, F. S. Chapin III, M. Mack, and G. B. Bonan, "The representation of arctic soils in the land surface model: The importance of mosses," J. Clim **14** (15), 3324−3335 (2001).

14. S. Chadburn, E. Burke, R. Essery, J. Boike, M. Langer, M. Heikenfeld, P. Cox, and P. Friedlingstein, "An improved representation of physical permafrost dynamics in the JULES land-surface model," Geosci. Model Dev. **8,** 1493−1508 (2015).

15. Y. Wu, D. L. Verseghy, and J. R. Melton, "Integrating peatlands into the coupled Canadian Land Surface Scheme (CLASS) v3.6 and the Canadian Terrestrial Ecosystem Model (CTEM) v2.0," Geosci. Model Dev. **9** (8), 2639−2663 (2016).

16. J. A. O' Donnell, V. E. Romanovsky, J. W. Harden, and A. D. McGuire, "The effect of moisture content on the thermal conductivity of moss and organic soil horizons from black spruce ecosystems in interior Alaska," Soil Sci. **174** (12), 646−651 (2009).

17. A. Druel, P. Peylin, G. Krinner, P. Ciais, N. Viovy, A. Peregon, V. Bastrikov, N. Kosykh, and N. Mironycheva-Tokareva, "Towards a more detailed representation of high-latitude vegetation in the global land surface model ORCHIDEE (ORC-HL-VEGv1.0)," Geosci. Model Dev. **10** (12), 4693−4722 (2017).

18. K. Mägdefrau, "Life-forms of bryophytes," in *Bryophyte Ecology,* Ed. by A. J. Smith (Chapman and Hall, London,1982), pp. 45−58.

19. J. M. Glime, "Adaptive strategies: growth and life forms," in *Bryophyte ecology. Vol. 1. Physiological Ecology* (Michigan Tech. Univ., Houghton, 2017), p. 37

20. V. E. Fedosov, "Oceanic and continental elements in the moss flora of the North Holarctic," in *Mosses: Ecology, Life Cycle and Significance* (Nova Science, N.-Y., 2018), pp. 107−151.

21. H. Park, S. Launiainen, P. Konstantinov, Y. Iijima, and A. Fedorov, "Modelling the effect of moss cover on soil temperature and carbon fluxes at a tundra site in Northeastern Siberia," J. Geophys. Res.: Biogeosci. **123** (9), 3028−3044 (2018).

22. F. J. W. Parmentier, T. C. Maximov, and F. Berendse, "The cooling capacity of mosses: controls on water and energy fluxes in a Siberian tundra site," Ecosystems **14** (7), 1055−1065 (2011).

23. J. L. Gornall, I. S. Jónsdóttir, S. J. Woodin, and R. van der Wal, "Arctic mosses govern below-ground environ-

ment and ecosystem processes," Gecologia **153** (4), 931−941 (2007).

24. A. V. Sosnovskii and N. I. Osokin, "The effect of moss and snow covers on permafrost stability on Western Spitsbergen during climate change," Vestn. Kol'sk. Nauch. Tsentra RAN. No. 3 (10), 178−184 (2018).

25. A. A. Tishkov, N. I. Osokin, and A. V. Sosnovskii, "The impact of moss synusia on the active layer of arctic soil and subsoil," Izv. Ross. Akad. Nauk. Ser. Geogr., No. 3, 39−46 (2015).

26. V. R. Sobol, P. N. Goman, I. V. Dedyulia, A. G. Brovka, and O. N. Mazurenko, "Thermal properties of the forest soil cover material at a characteristic moisture content," J. Eng. Phys. Thermophys. **84** (5), 1165−1174 (2011).

27. N. Zhang and Z. Wang, "Review of soil thermal conductivity and predictive models," Int. J. Therm. Sci. **117,** 172−183 (2017).

28. O. Wiener, "Die Theorie des Mischkörpers für das Feld der stationären Strömung. Erste Abhandlung. Die Mittelswertsätze für Kraft, Polarisation und Energie," Abh. d. Math. Klasse der Konigl. Sachs. Ges. Wiss. **32,** 509−604 (1912).

29. O. Johansen, *Thermal Conductivity of Soils. No. CRREL-TL-637* (Cold Regions Research and Engineering Lab Hanover NH, 1977).

30. S. X. Chen, "Thermal conductivity of sands," Heat Mass Transfer **44** (10), 1241−1246 (2008).

31. J. Cote and J. -M. Konrad, "Thermal conductivity of base-course materials," Can. Geotech. J. **42** (1), 61−78 (2005).

32. F. Donazzi, E. Occhini, and A. Seppi, "Soil thermal and hydrological characteristics in designing underground cables," Proc. Inst. Electr. Eng. **126** (6), 506 (1979).

33. S. Lu, T. Ren, Y. Gong, and R. Horton, "An improved model for predicting soil thermal conductivity from water content at room temperature," Soil Sci. Soc. Am. J. **71** (1), 8 (2007).

34. C. D. Peters-Lidard, E. Blackburn, X. Liang, and E. F. Wood, "The effect of soil thermal conductivity parameterization on surface energy fluxes and temperatures," J. Atmos. Sci. **55** (7), 1209−1224 (1998).

35. V. Balland and P. A. Arp, "Modelling soil thermal conductivities over a wide range of conditions," J. Environ. Eng. Sci. **4** (6), 549−558 (2005).

36. P. A. Moore, A. G. Smolarz, C. E. Markle, and J. M. Waddington, "Hydrological and thermal properties of moss and lichen species on rock barrens: Implications for turtle nesting habitat," Ecohydrology **12** (2), 2057 (2019).

37. A. Aisar, Y. Katoh, H. Katsurayama, M. Koganei, and M. Mizunuma, "Effects of convection heat transfer on Sunagoke moss green roof: A laboratory study," Energy Build. **158,** 1417−1428 (2018).

38. S. K. Rice, T. A. Gagliardi, and R. A. Krasa, "Canopy structure affects temperature distributions and free convection in moss shoot systems," Am. J. Bot. **105** (9), 1499−1511 (2018).

39. C. Rappoldt, G. -J. J. M. Pieters, E. B. Adema, G. J. Baaijens, A. P. Grootjans, and C. J. Van Duijn, "Buoyancy-driven flow in a peat moss layer as a mecha-nism for solute transport," Proc. Natl. Acad. Sci. U. S. A. **100** (25), 14937−14942 (2003).

40. D. A. Nield and A. Bejan, "Mechanics of fluid flow through a porous medium," in *Convection in Porous Media* (Springer New York, New York, 2013), p. 1.

41. E. V. Mortikov, A. V. Glazunov, and V. N. Lykosov, "Numerical study of plane Couette flow: turbulence statistics and the structure of pressure−strain correlations," Russ. J. Numer. Anal. Math. Model. **34** (2), 119−132 (2019).

42. A. S. Monin and A. M. Obukhov, "Basic laws of turbulent mixing in the atmosphere near the ground)," Trudy Geofiz. Inst. AN SSSR, No. **24** (151), 163−187 (1954).

43. M. R. Raupach, "Drag and drag partition on rough surfaces," Boundary-Layer Meteorol. **60,** 375−395 (1992).

44. W. Brutsaert, *Evaporation into the Atmosphere: Theory, History and Applications* (Springer Science & Business Media, Dordrecht, 2013).

45. J. R. Garratt, *The Atmospheric Boundary Layer* (Cambridge University Press, Cambridge, UK, 1997).

46. X. Lee, "Principle of eddy covariance," in *Fundamentals of Boundary-Layer Meteorology* (Springer, Cham, 2018), p. 149−173.

47. O. N. Laban, C. M. Maghanga, and K. Joash, "Determination of the surface roughness parameter and wind shear exponent of Kisii Region from the on-site measurement of wind profiles," J. Energy **2019,** art. 8264061 (2019).

48. M. P. Langleben, "A study of the roughness parameters of sea ice from wind profiles," J. Geophys. Res. **77** (30), 5935−5944 (1972).

49. M. Bottema, "Landscape roughness parameters for Sherwood forest − Experimental results," Boundary-Layer Meteorol. **89** (2), 285−316 (1998).

50. S. S. Zilitinkevich, I. Mammarella, A. A. Baklanov, and S. M. Joffre, "The effect of stratification on the aerodynamic roughness length and displacement height," Boundary-Layer Meteorol. **129,** 179−190 (2008).

51. S. S. Zilitinkevich, A. A. Grachev, and C. W. Fairall, "Scaling reasoning and field data on the sea-surface roughness lengths for scalars," J. Atmos. Sci. **58,** 320−325 (2001).

52. S. Zilitinkevich, "Non-local turbulent transport pollution dispersion aspects of coherent structure of convective flows," in *Air Pollution III, Air Pollution Theory and Simulation,* Ed. by H Power, N. Moussiopoulos, and C. A. Brebbia (Computational Mechanics Publ., Southampton, Boston, 1995), pp. 53−60.

53. W. Brutsaert, "The roughness length for water vapor sensible heat, and other scalars," J. Atmos. Sci. **32** (10), 2028−2031 (1975).

54. M. Kanda, M. Kanega, T. Kawai, R. Moriwaki, and H. Sugawara, "Roughness lengths for momentum and heat derived from outdoor urban scale models," J. Appl. Meteorol. Climatol **46,** 1067−1079 (2007).

55. K. Yang, T. Koike, H. Ishikawa, J. Kim, X. Li, H. Liu, S. Liu, Y. Ma, and J. Wang, "Turbulent flux transfer over bare-soil surfaces: Characteristics and parameterization," J. Appl. Meteorol. Climatol **47** (1), 276−290 (2008).

56. F. Chen and Y. Zhang, "On the coupling strength between the land surface and the atmosphere: From viewpoint of surface exchange coefficients," Geophys. Res. Lett. **36,** L10404 (2009).

57. Y. Chen, K. Yang, D. Zhou, J. Qin, and X. Guo, "Improving the Noah land surface model in arid regions with an appropriate parameterization of the thermal roughness length," J. Hydrometeorol. **11** (4), 995−1006 (2010).

58. Y. Chen, K. Yang, J. He, J. Qin, J. Shi, J. Du, and Q. He, "Improving land surface temperature modeling for dry land of China," J. Geophys. Res. **116,** D20104 (2011).

59. W. Zheng, H. Wei, Z. Wang, X. Zeng, J. Meng, M. Ek, K. Mitchell, and J. Derber, "Improvement of daytime land surface skin temperature over arid regions in the NCEP GFS model and its impact on satellite data assimilation," J. Geophys. Res. **117,** D06117 (2012).

60. E. Haghighi and D. Or, "Evaporation from porous surfaces into turbulent airflows: Coupling eddy characteristics with pore scale vapor diffusion," Water Resour. Res. **49,** 8432−8442 (2013).

61. E. Haghighi, E. Shahraeeni, P. Lehmann, and D. Or, "Evaporation rates across a convective air boundary layer are dominated by diffusion," Water Resour. Res. **49,** 1602−1610 (2013).

62. X. Zeng and R. E. Dickinson, "Effect of surface sublayer on surface skin temperature and fluxes," J. Clim **11** (4), 537−550 (1998).

63. C. J. P. P. Smeets and M. R. V. D. Broeke, "The parameterisation of scalar transfer over rough ice," Boundary-Layer Meteorol. **128** (3), 339−355 (2008).

64. E. L. Andreas, "Theory for scalar roughness and scalar transfer coefficients over snow and sea ice," Boundary-Layer Meteorol. **38,** 159−184 (1987).

65. D. Li, A. Rigden, G. Salvucci, and H. Liu, "Reconciling the Reynolds number dependence of scalar roughness length and laminar resistance," Geophys. Res. Lett. **44** (7), 3193−3200 (2017).

66. M. Mölder, A. Lindroth, and A. Grelle, "Experimental determination of the roughness length for temperature over a field of tall grass in central Sweden," Geogr. Annaler: Ser. A, Phys. Geogr. **81** (1), 87−100 (1999).

67. A. Rigden, D. Li, and G. Salvucci, "Dependence of thermal roughness length on friction velocity across land cover types: A synthesis analysis using AmeriFlux data," Agric. For. Meteorol. **249,** 512−519 (2018).

68. P. K. Alekseychik, A. Korrensalo, I. Mammarella, T. Vesala, and E. S. Tuittila, "Relationship between aerodynamic roughness length and bulk sedge leaf area index in a mixed-species boreal mire complex," Geophys. Res. Lett. **44** (11), 5836−5843 (2017).

69. S. Ondimu and H. Murase, "Image processing and roughness analysis as a tool for quantification of physiological well-being in plants: results for Sunagoke moss," in *Proc. 17th World Congress Int. Fed. Automatic Control (Seoul, Korea, July 6−11, 2008),* p. 641.

70. S. K. Rice and N. Schneider, "Cushion size, surface roughness, and the control of water balance and carbon flux in the cushion moss *Leucobryum glaucum* (Leucobryaceae)," Am. J. Bot. **91** (8), 1164−1172 (2004).

71. L. Kang, J. Zhang, X. Zou, H. Cheng, C. Zhang, and Z. Yang, "Experimental investigation of the aerodynamic roughness length for flexible plants," Boundary-Layer Meteorol. **172** (3), 397−416 (2019).

Translated by N. Semenova

On the Mechanism of Arctic Climate Oscillation with a Period of About 15 Years According to Data of the INM RAS Climate Model

E. M. Volodin*

Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, 119333 Russia
e-mail: volodinev@gmail.com

Abstract—According to data from a preindustrial experiment, a mechanism of Arctic climate oscillation with a period of about 15 years in the INM-CM5 climate model is considered. For this purpose, a technique for computing the contribution of the different components to the generation of oscillation energy and phase variation is used. It is shown that, in the positive phase of an oscillation (the warm Arctic), a negative anomaly of the temperature and salinity occurs at middle latitudes of the North Atlantic. These anomalies are accompanied by a positive North Atlantic Oscillation index. A quarter of the period after Arctic warming, warming and salinization occur at middle latitudes of the North Atlantic. In the Arctic Ocean, the temperature anomalies are generated due to the intensifying advection of the Atlantic water. The evolution of the oscillation phase is accounted for by the heat transport from the Atlantic and the heat flux across the surface. Anomalies of the currents transporting the heat from the Atlantic to the Arctic Ocean are generated mainly by surface wind stress.

Keywords: climate, oscillation, mechanism, generation, phase, temperature, salinity, currents, anomaly

INTRODUCTION

It is known from the observation data [1] that the climate of the Arctic and the North Atlantic is subject to oscillations with a typical period of about 60 years. Oscillations with a typical period of 10−20 years also occur. However, it is extremely difficult to analyze the mechanism of such oscillations based on the observation data. The available time series of observations in the ocean is too short, and the accuracy and coverage are insufficient. It is shown in [2] that the Arctic warming at the beginning of the 21st century can be related to the increase in the heat transport to high latitudes in the atmosphere, but it refers only to the correlation of the warming observed and the vortex activity index in the atmosphere. Work [3] demonstrates the relationships between the Arctic warming in the middle of the 20th century and the decadal oscillation indices, but its data are not sufficient either to conclude if they cause or result from Arctic warming. The natural temperature oscillations in the Arctic and the North Atlantic also manifest themselves in the climate models, which provides additional possibilities for the diagnosis of these oscillations; e.g., it was suggested in [4] that the small amount of ice in the Barents Sea can cause the low pressure above it and the winds that facilitate the penetration of surface currents from the Atlantic to the Arctic, thus contributing to further warming. By now, several mechanisms of natural climate oscillations in the Arctic and the North Atlantic that occur in the climate models have been proposed,

e.g., [5] and [6]. However, it is hardly provable that this is the mechanism realized in this climate model.

Arctic climate oscillations with a typical period of about a few decades also occur in the INM-CM5 climate model. The first empirical orthogonal function (EOF) of the surface temperature is localized in the Arctic, and the spectrum of its Fourier coefficient has a markedly pronounced maxima at periods of about 60 years and about 15 years. The mechanism of 60-year oscillations was considered in [7]. For this purpose, the technique for computing the contribution of different components to the generation of oscillation energy and its phase evolution was used. In studying a 15-year oscillation, we also use this technique and emphasize that the mechanism of a 15-year oscillation differs from the mechanism of a 60-year oscillation.

MODEL, NUMERICAL EXPERIMENT, AND DIAGNOSTIC METHOD OF A 15-YEAR OSCILLATION

The data from a numerical experiment using the INM-CM5 climate model are considered. The ocean model was described in [8] and has blocks of atmospheric and ocean dynamics and an aerosol block. In the atmospheric block, the resolution is $2° × 1.5°$ in latitude and longitude and 73 levels vertical to an altitude of approximately 60 km. In the ocean block, the resolution is $0.5° × 0.25°$ and 40 levels. The numerical experiment was conducted in the preindustrial climate regime, when all impacts on the climate system were assigned at the

level of the year 1850, which makes it possible to identify eigenoscillations in the climate system. The experiment was performed within the framework of comparing the CMIP6 models (Coupled Model Intercomparison Project, phase 6 [9]). The experiment lasted 1200 years. For the latter 200 years of the experiment, the monthly average values of all summands in the prediction equations were preserved for the temperature, salinity, and velocity of currents, in addition to the typical issue of an ocean block. For the analysis of the 15-year oscillations, harmonics with periods shorter than 10 years and longer than 20 years were removed from the time series of all variables considered. As an oscillation index I, we took the temperature of the ocean surface in the region of $0°-80°$ E, $70°-82°$ N, where the first EOF of the temperature has the maximum. If there is ice on the surface, we take the temperature of the ocean surface under the ice instead of the ice temperature. Harmonics with periods shorter than 10 years and longer than 20 years were filtered from this time series like from all the others. We calculated the C composites of different atmospheric and ocean fields for the time offsets Δn from -7 to $+7$ years between the field under consideration and the oscillation index

$$C_{\Delta n} = \sum_{n=1}^{N} F_{n+\Delta n} I_n \bigg/ \sqrt{N \sum_{n=1}^{N} I_n^2}.$$

Here, n is the number of the time point (in this case, the year number) and N is the number of time points. The value of $C_{\Delta n}$ calculated this way represents an average-statistical anomaly of the F field in a year lagging behind the warming maximum in the Arctic by Δn years.

To diagnose an oscillation mechanism, the contributions of the summands to the generation of oscillation energy and evolution of its phase were calculated. This technique was described in detail in [10]. It considers the mechanism of quasi-biennial oscillation of the wind velocity in the equatorial stratosphere and clearly shows the impacts and extent of the change in this oscillation phase. Principally, this technique can be used in studying an oscillation of any nature. Work [7] presents its use in studying the mechanism of the 60-year oscillation of the Arctic climate based on the data of the INM-CM5 model. This work is just a brief summary of the main ideas.

Let the evolution equation be solved for certain Ψ value

$$\frac{\partial \Psi}{\partial t} = \sum_{j=1}^{J} F_j,$$

where t is the time and J is the number of summands in the right part. If we expand the left and right part of this equality in a Fourier series with respect to time and consider that oscillation energy E_k for the kth harmonic is expressed as $E_k = \Psi_k^2/2$, we may calculate the generation of energy for the kth harmonics by the Jth summand of $Gj(k)$, as well as the contribution of $Pj(k)$ to the evolution of the oscillation phase. The sum of energy generation by all summands should be equal to zero, and the contribution to the phase change can be normalized so that the sum of all contributions from the different terms was equal to 1.

We consider the prognostic variables of the ocean models as Ψ: temperature T, salinity S, and horizontal components of the current velocity U and V. In the right side of the equation, the following summands were preserved for T during the calculation of the model: FT_{ADV} is the contribution of advection by ocean currents to the temperature change; FT_{VD} is the contribution of vertical diffusion, including the fluxes on the surface: a sensible heat flux, a latent heat flux, and a long-wave radiation balance; FT_{IS} is the contribution of isopycnic diffusion; and FT_{RAD} is the contribution of solar radiation. There are no other summands in the equation for the temperature in the model. Solar radiation is considered here separately from thermal radiation, since solar radiation partly penetrates water and thermal radiation is fully absorbed by the surface. When the ocean surface is ice-covered, the heat fluxes across the surface imply the fluxes at the ice–water interface, and these fluxes also include effects related to the ice formation and melting. Similarly, for salinity we have summands describing the contributions of advection, vertical diffusion and the flux across the surface, and isopycnic diffusion FS_{ADV}; FS_{VD}; FS_{IS}. The fluxes across the surface in the summand FS_{VD} comprise sediments, evaporation, and the influence exerted by the ice formation and melting on salinity. For the U component of the current velocity, we have the summands FU_{ADV}, FU_{VD} (this summand also includes the flux across the surface, i.e., in this case, the surface wind stress), FU_{HD} (the contribution of horizontal diffusion), and FU_{PC} (the contribution of the pressure gradient and the Coriolis force). The V component of the velocity has similar summands. We calculate the values of G and P by the fields of composites with a time offset for each node of the model network and then average them over altitude, longitude, and latitude with respect to the cell volume and for the case of P with respect to the oscillation energy, so that the total value of G is equal theoretically to zero and the sum of all values of P is equal to 1. In fact, however, the sums can noticeably differ from the theoretical values, first and foremost, due to the finite precision of machine arithmetic, especially where the total tendency is calculated in the model as a small difference of large values. Another reason is the finiteness of the time series in use for identifying composites and time harmonics.

DIAGNOSIS OF THE ARCTIC OSCILLATION

We consider the composites of the fields characterizing a 15-year oscillation of the pressure at the sea level, the surface temperature, the ocean salinity at the surface, and the velocities of currents at the ocean surface during the different oscillation phases. Figure 1
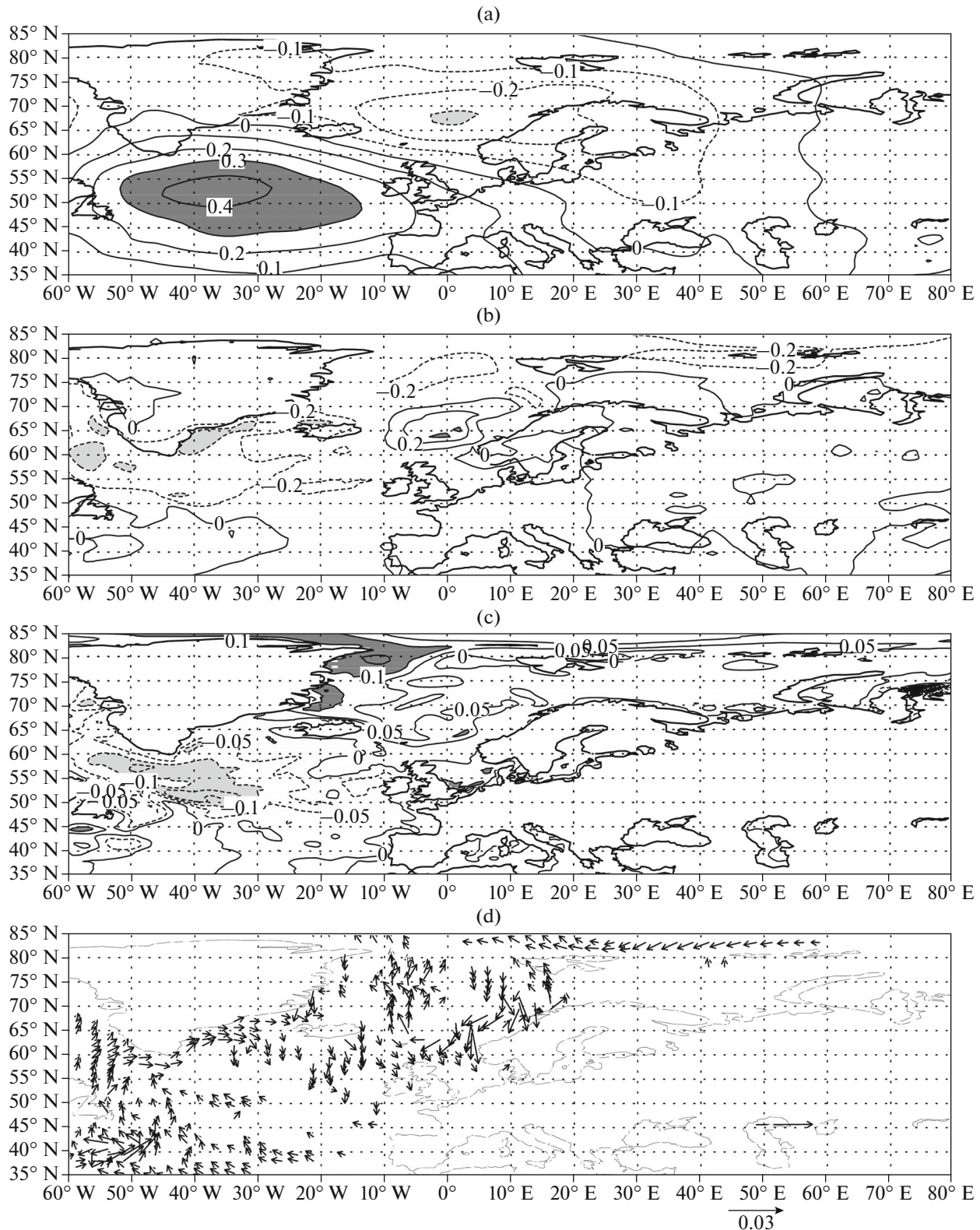
Fig. 1. Composites for anomalies of (a) pressure at the sea level, hPa; (b) surface temperature, K; (c) salinity at the surface, per-mille; and (d) surface current velocity, m/s, four years before the maximum temperature in the Arctic.

presents the composites of these fields over 4 years before the maximum of the oscillation index, Fig. 2 demonstrates the composites over 2 years before the maximum, Fig. 3 depicts them during the maximum, and Fig. 4 depicts them 2 years after the maximum.

The surface temperature anomalies are negative over 4 years (a little longer than a quarter of the period) to the maximum temperature in the Arctic almost everywhere in the North Atlantic and the Atlantic sector of the Arctic, except for a small positive region in the Norwegian Sea. The pressure anomalies resemble the North Atlantic Oscillation (NAO), but the minimum of pressure is displaced to the east and the maximum, to the northwest of their typical positions at a high NAO index. The salinity anomalies are primarily negative in the North Atlantic and positive in the Atlantic sector of the Arctic. Though the anomalies of the currents are differently directed, the reduced flux of the Atlantic water to the Arctic Ocean along the northern coast of Europe is still dominant.

Figure 2 clearly shows the positive temperature anomalies of ~1° in the Barents and Kara Seas, as well as in the north of the Greenland−Norwegian Sea 2 years before the maximum of the temperature in the Arctic. The pressure anomaly almost corresponds to the positive NAO index, with the western transport intensifying near 60° N. The anomalies of the near-surface currents mostly correspond to the anomaly of the geostrophic wind with respect to the Ekman turn to the right. The Atlantic water flux to the north and east in the Norwegian and Barents seas is greater than an ordinary one. The anomalies of water salinity at the surface are negative in the North Atlantic and positive in the Atlantic sector of the Arctic Ocean.

At the maximum temperature in the Arctic (Fig. 3), the distribution of the anomalies is similar to that which takes place 2 years before that; the temperature anomalies reach 1.6°−1.8°; and the currents, directed to the north and east in the Norwegian, Barents, and Kara seas, are even more noticeable.

Two years after the maximum temperature, the ocean surface-temperature anomalies still reach 1° in the Greenland−Norwegian, Barents, and Kara seas; however, in the North Atlantic, negative anomalies change to positive and the pressure anomalies are small and do not correspond to the NAO; the anomalies of the currents are small everywhere except for the Greenland−Norwegian Sea.

We emphasize that Figs. 1−4 present composites of the average annual anomalies. If we construct the same composites, but with respect to the annual course, i.e., separately for Januaries, Februaries, etc., it will turn out for the pressure and temperature of the surface that, in the winter months, the anomalies are similar to the average monthly ones, but greater than them by a factor of 2−3, and in the summer months the anomalies are usually slight. Thus, the composites of the pressure and temperature of the surface are determined primarily by the winter months. At the same time, the composites of salinity and velocity of currents for the different months of a year are quite close to average annual ones. Since we studied the process with a period of about 15 years and the calculation of the composites is a linear operation, the composites of the fields, e.g., 6 years after the maximum warming of the Arctic, are close to the composites that lag behind half of the period (7.5 years), i.e., 1.5 years before the maximum with a minus sign. Thus, the maximum negative NAO phase is found at a time offset with respect to the maximum warming in the Arctic equal to 6 years, and the minimum temperature in the Arctic will take place in another 1.5 years.

The composites for the 15-year oscillation are similar to the composites for the 60-year oscillation in that the positive NAO index and the Atlantic water flux in the Arctic Ocean, which is more intense than usually, correspond to the warm Arctic. In addition, the warm Arctic is preceded by the cold and fresh North Atlantic for the both oscillations.

The main differences from the composites for the 60-year oscillation are that the Arctic warming is not preceded by a negative anomaly of salinity at depths from 200 to 1000 m along the continental slope in the entire Arctic Ocean, which was recorded in [7]. For the 15-year oscillation, there is no increase in the Atlantic water flux in the Arctic Ocean over a quarter of the period to the maximum temperature, which was also recorded in [7] and is attributed to the salinity anomaly at depths. Therefore, the mechanism of energy generation and change in the phase of the 15-year oscillation should be similar to and different from the mechanism of the 60-year oscillation in some ways. In order to understand these similarities and differences, we consider the contribution of the different summands to the generation of the energy of the 15-year oscillation and the change in its phase (Table 1). All calculations were averaged over the region 70°−82° N, 0°−80° E, 0−100 m, where the oscillation has the maximum amplitude in the temperature.

The transport by the ocean currents makes the greatest contribution to the generation of the temperature anomalies. The contribution of solar radiation is also positive and is approximately twice as small as the contribution made by the transport. This can be interpreted as the fact that the temperature anomalies are primarily generated by the advection anomalies and the solar radiation plays the role of an additional positive feedback: during Arctic warming, the ice area shrinks, which leads to the greater absorption of solar radiation by the surface. The contribution of vertical diffusion and surface fluxes to the generation of the temperature anomalies is negative, and this summand primarily compensates the positive summands. This means that, during Arctic warming, the ocean releases more heat to the atmosphere. The contribution of the isopycnic diffusion is negative. All conclusions on the
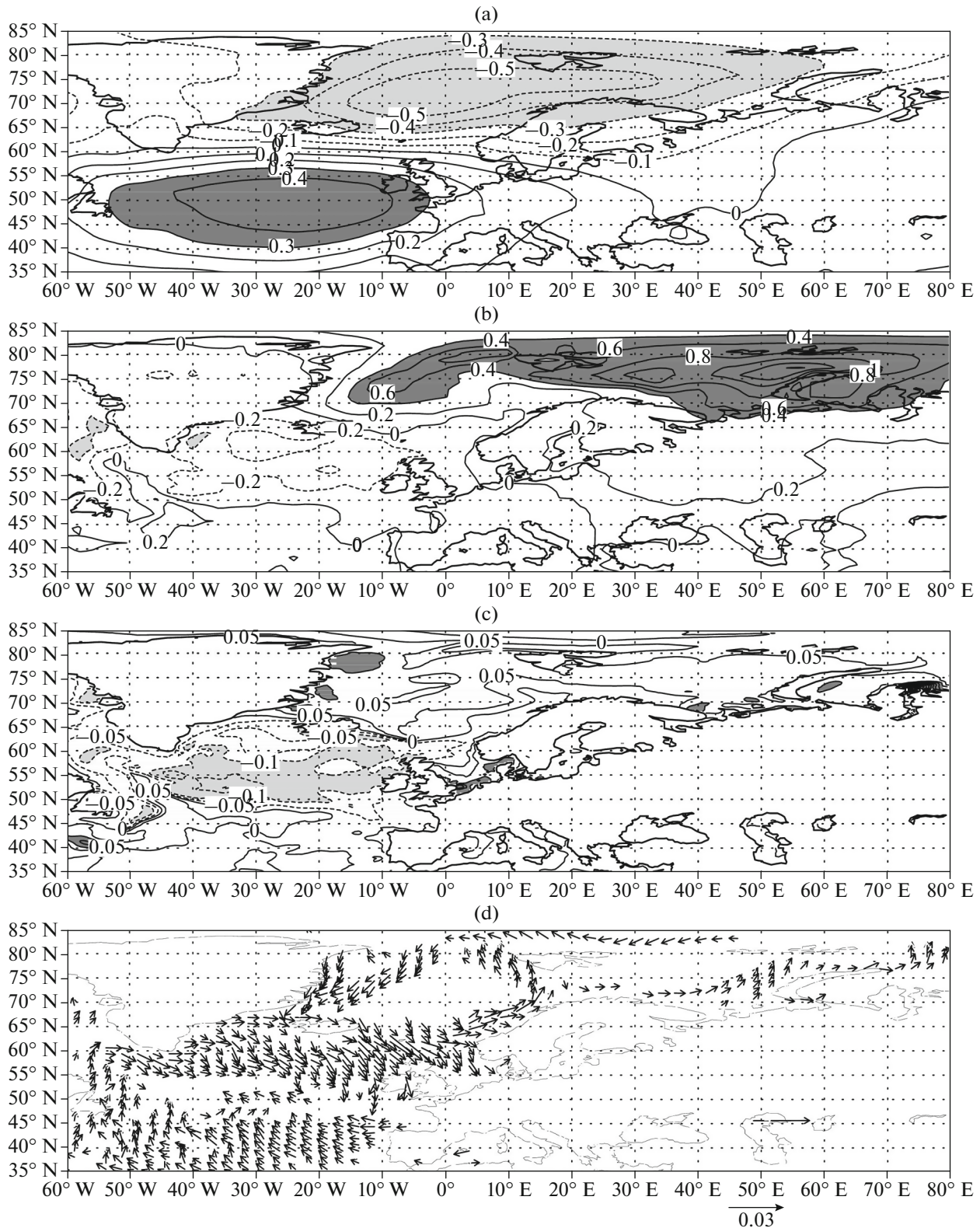
Fig. 2. Composites for anomalies of (a) pressure at the sea level, hPa; (b) surface temperature, K; (c) salinity at the surface, per-mille; and (d) surface current velocity, m/s, two years before the maximum temperature in the Arctic.
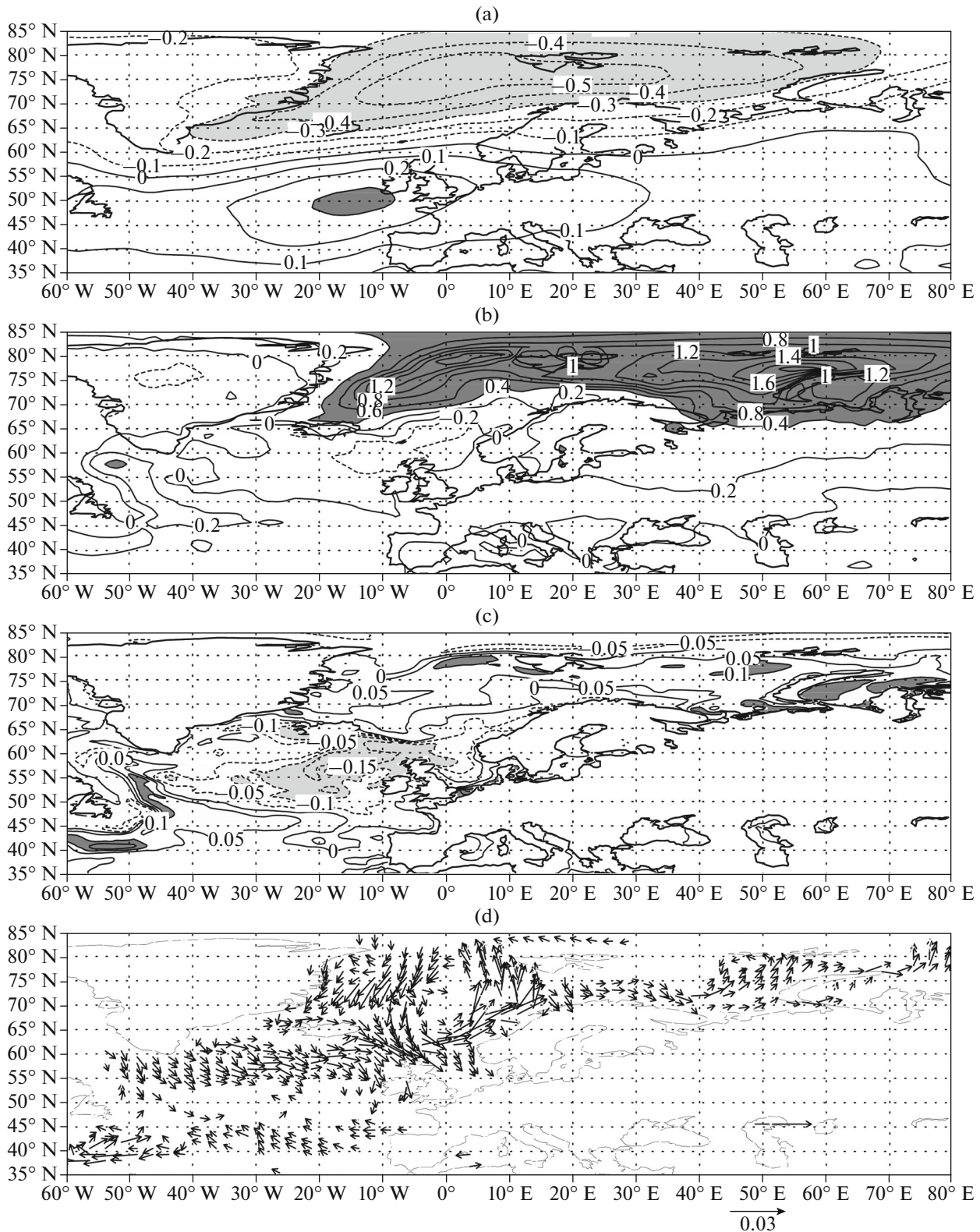
Fig. 3. Composites for anomalies of (a) pressure at the sea level, hPa; (b) surface temperature, K; (c) salinity at the surface, per-mille; and (d) surface current velocity, m/s, at the maximum temperature in the Arctic.
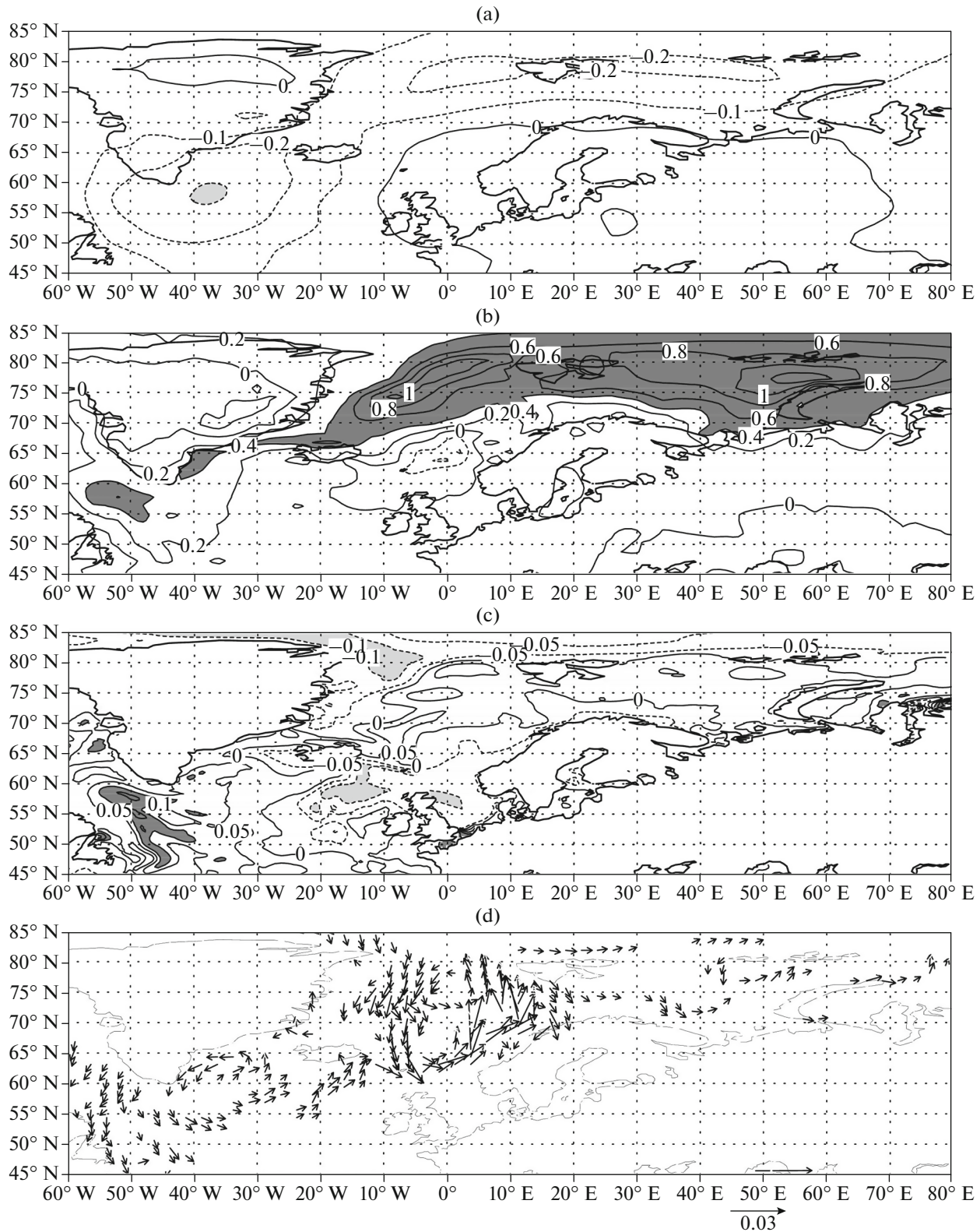
Fig. 4. Composites for anomalies of (a) pressure at the sea level, hPa; (b) surface temperature, K; (c) salinity at the surface, permille; and (d) surface current velocity, m/s, two years after the maximum temperature in the Arctic.

Table 1. Contribution of different summands to the generation of energy G and the change in phase P of oscillation of temperature T, salinity S and current velocity U, V in the Arctic in the region $70°-82°$ N, $0°-80°$ E, $0-100$ m. The designations of the summands are given in section "Model, Numerical Experiment, and Method of Diagnostics of a 15-Year Oscillation"

		FT_{ADV}	FT_{VD}	FT_{IS}	FT_{RAD}
T	G, 10^{-10} K^2/s	4.48	−5.29	−1.21	2.27
	P	0.44	0.46	0.01	0.01
		FS_{ADV}	FS_{VD}	FS_{IS}	
S	G, 10^{-12}‰ ^{2}/s	6.1	−5.9	0.1	
	P	0.77	0.17	0.00	
		FU_{ADV}, FV_{ADV}	FU_{VD}, FV_{VD}	FU_{HD}, FV_{HD}	FU_{PC}, FV_{PC}
U, V	G, 10^{-10} m^2/s^3	−0.8	12.6	−10.8	−1.1

generation of the temperature anomalies are qualitatively the same as for the 60-year oscillation.

The phase of the temperature oscillation changes under the impact of summands of the transport and the surface fluxes. The phase of the 60-year oscillation changes only under the influence of the transport, and the contribution of surface fluxes is negative. For the 15-year oscillation during the transition from the cold to the warm phase, the atmosphere gets warmer a little earlier and the heat fluxes to the ocean from the atmosphere contribute to the change in the oscillation phase in the ocean.

For salinity, like for the case of the 60-year oscillation, the transport makes a major contribution to anomaly generation and phase evolution.

The analysis of the contributions to the generation of horizontal currents shows that the only positive summand is the contribution of the surface wind stress and vertical diffusion; i.e., the current anomalies that generate the temperature oscillations are merely wind ones in the upper 100-m layer. For the 60-year oscillation, the surface wind stress also makes a major contribution to the generation of the currents, but the contribution of a pressure gradient is also significant.

The contribution of the main mechanisms to maintaining the 15-year oscillation of the temperature in the Arctic is illustrated in Fig. 5, which shows the composites of the surface temperature, the temperature transport summand, the sensible heat flux from the atmosphere to the ocean, as well as the NAO index. To use the same scale, we need to have all data normalized to the root-mean-square deviation. The temperature has the maximum at time point 0. The summand of the transport leads in phase the temperature by approximately 1.5 years. It is almost maximum at the maximum temperature, i.e., makes a positive contribution to the generation, and leading in phase contributes to the phase evolution. The sensible heat flux is almost in antiphase with the temperature; when the temperature is maximum, it is nearly the lowest, which means a negative contribution to the genera-tion. At the same time, the minimum flux occurs 1 year later than the maximum temperature, which ensures the positive contribution of this summand to the phase evolution. Figure 5 also displays the time course of the NAO index, which is calculated as a pressure difference in the region $40°$ W–$40°$ E, $35-45°$ N–$40°$ W–$60°$ E, and $65-80°$ N. The NAO index leads in phase the temperature in the Arctic by $1-2$ years and is found almost in the phase with the anomaly of the heat transport by the currents. This may mean that the surface wind stress associated with the NAO anomalies rapidly generates the wind currents that transport the heat to the Arctic. The positive NAO index also provides the transport of the warmer air in the atmosphere during the transition from the cold to the warm Arctic, which leads to the offset between the maximum temperature and the minimum sensible heat flux to the ocean.

DIAGNOSIS
OF THE NORTH ATLANTIC OSCILLATION

We also present the results for the generation of the anomalies of temperature, salinity, and velocities in the near-surface layer down to 100 m in the North Atlantic ($60°-0°$ W, $45°-62°$ N, Table 2). The major summand that contributes to the generation of the temperature anomalies and the evolution of the oscillation phase is the transport by the currents. This result differs from similar data for the 60-year oscillation, where the comparable contributions to the generation and the change in the temperature phase in the Atlantic are made by several summands.

For salinity, the main contribution to the generation of anomalies is made by the flux across the surface. This is when the water is colder than on the average; it evaporates less and exactly this mechanism maintains the negative anomaly of salinity in the Atlantic at $50°-62°$ N, depicted in Figs. $1-3$. At the same time, southward of $50°$ N at a zero-time offset, the positive temperature anomaly leads to greater evaporation, which maintains the positive salinity
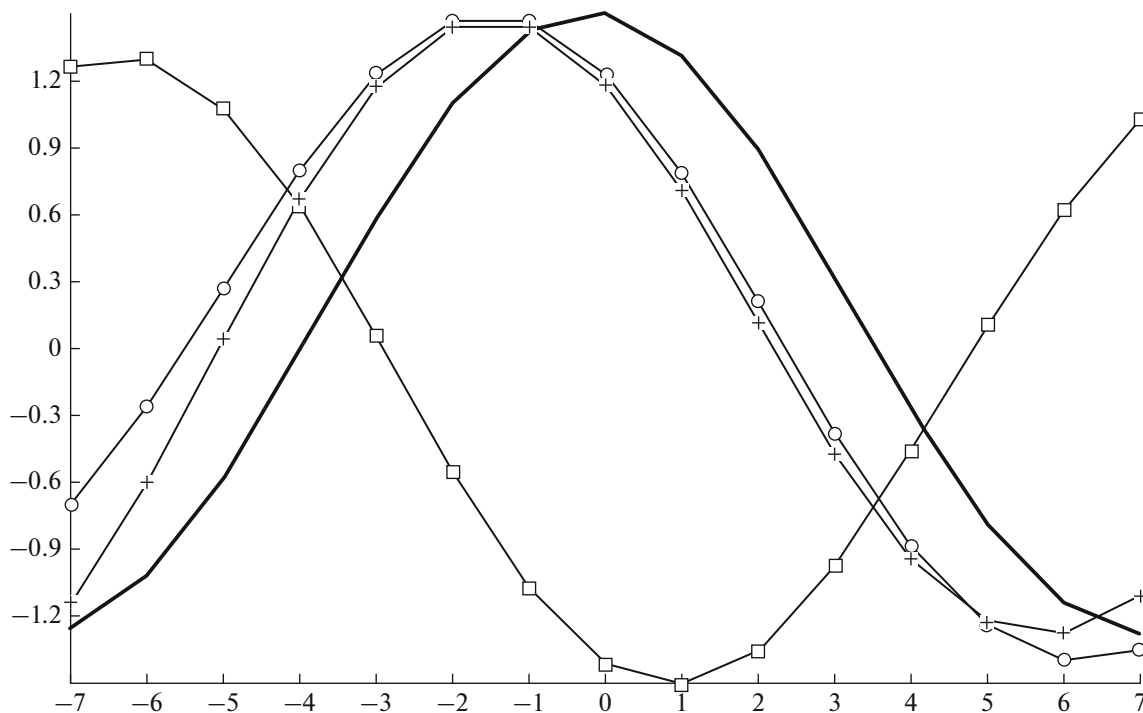
Fig. 5. Composites for the surface temperature (solid line without marks); water heating in the layer of 0−100 m due to the sensible heat flux from the atmosphere (line with squares); water heating in the layer of 0−100 m due to temperature advection by currents (line with circles) in the region 70°−82° N, 0°−80° E; as well as the NAO index (line with crosses) for time offsets from −7 to 7 years with respect to index *I*. All values are normalized to root-mean-square deviation.

anomaly that also appears northward at a time offset +2 years. The change in the salinity oscillation phase occurs due to the transport. The main summand that provides the generation of the near-surface currents in the North Atlantic are surface wind stress; i.e., the currents are purely wind ones. This is very consistent with the data in Figs. 1−4, where, as a rule, the North Atlantic currents turn to the right at an angle of about 30° relative to the geostrophic wind.

Unlike the 60-year oscillation, the noticeable anomalies of meridional circulation in the North Atlantic are

attributed to the 15-year oscillation. The anomalies of the meridional streamfunction are maximum in value 1−2 years before the maximum temperature in the Arctic (Fig. 6) and correspond to the maximum NAO index. The maximum streamfunction of about 0.5 Sv is located at a depth of ~1500 m at 35° N; the decrease occurs at 40°−50° N, and the increase occurs southward of 30° N. Such a structure is close to the first EOF of the meridional streamfunction obtained for most of the climate models in [11]. It is shown in [12] that this structure is excited if, in the numerical experiment,

Table 2. Contribution of different summands to the generation of energy *G* and the change in phase *P* of oscillation of temperature *T*; salinity *S*; and current velocity *U, V* in the North Atlantic in the region 45°−62° N, 60°−0° W, 0−100 m

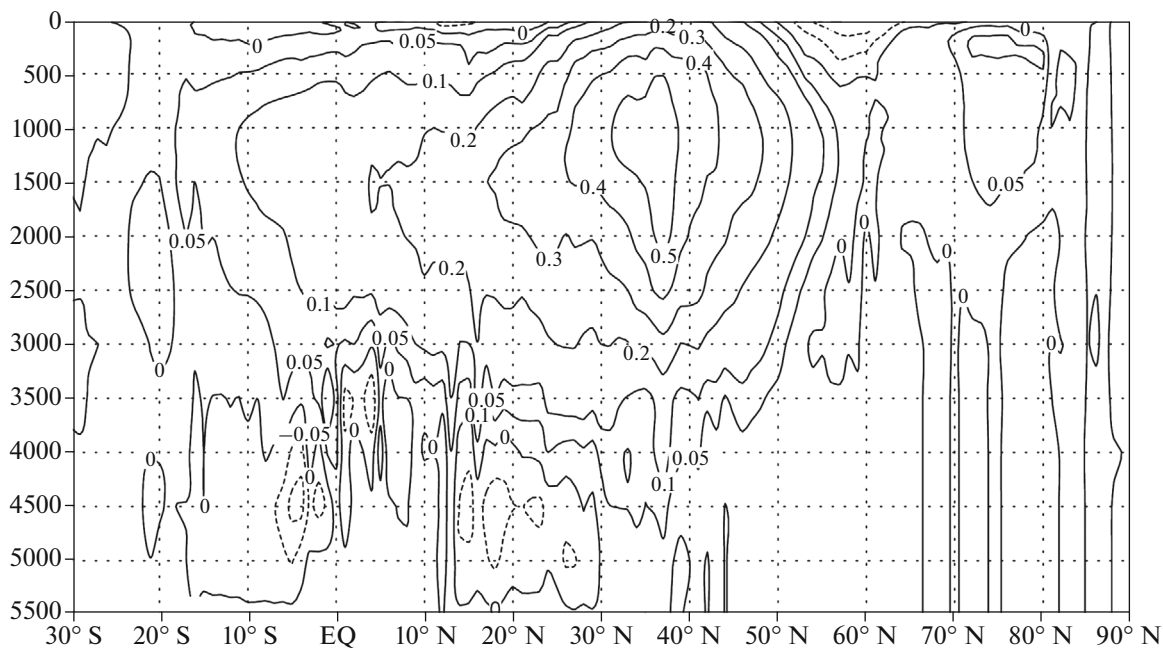|		FT_{ADV}	FT_{VD}	FT_{IS}	FT_{RAD}
T	*G*, 10^{-10} K^2/s	2.50	−2.16	−0.54	0.12
	P	0.85	−0.02	0.04	0.01
		FS_{ADV}	FS_{VD}	FS_{IS}	
S	*G*, 10^{-12}‰ ^{2}/s	−0.76	4.68	−2.52	
	P	0.86	−0.02	0.05	
		FU_{ADV}, FV_{ADV}	FU_{VD}, FV_{VD}	FU_{HD}, FV_{HD}	FU_{PC}, FV_{PC}
U, V	*G*, 10^{-10} m^2/s^3	2.6	56.7	−15.4	−43.7

Fig. 6. Composite of the anomaly of the meridional streamfunction in the Atlantic (Sv) two years before the temperature maximum in the Arctic.

the surface wind stress perturbation is artificial, corresponding to the NAO in the form of an oscillation with a period of 10 years. In this case, the increase in the meridional circulation of about 35° N also corresponds to the positive NAO index. Work [13] shows that the Ekman vertical velocities generated by the NAO anomalies are close to the total vertical velocities; therefore, the anomalies of the meridional circulation related to the NAO index are primarily generated by the surface wind stress and divergence of wind currents. Thus, in our case, the anomaly of the meridional streamfunction is also likely to be determined by the anomaly of the surface wind stress, the direct evidence of which requires additional diagnostics.

CONCLUSIONS

According to the data of a preindustrial numerical experiment with the use of the INM-CM5 climate model, the Arctic climate oscillation with a period of about 15 years was studied. It was shown that, in the Arctic, the generation of the temperature anomalies is ensured by the heat transport by the ocean currents and, to a lesser extent, by the absorption of solar radiation. The oscillation phase changes almost equally due to the heat transport by the ocean currents and due to the heat flux from the atmosphere to the ocean, primarily the whole sensible heat flux. In turn, the anomalies of the currents are generated by the surface wind stress. The NAO index reaches the maximum over 1−2 year to the maximum temperature in the Arctic. The related surface wind stress generates the

currents that bring more Atlantic water to the Arctic Ocean and lead to the maximum temperature.

In the Northern Atlantic, the anomalies of the currents related to the 15-year oscillation are also wind anomalies. The generation of the temperature anomalies and the evolution of the temperature oscillation phase take place there due to the anomaly of the heat transport by the ocean currents. The maximum of the meridional streamfunction in the Atlantic near 35° N also occurs 2 years before the maximum temperature in the Arctic.

The question that remains unclear is whether the pressure anomalies corresponding to the different oscillation phases are the atmospheric response to the respective anomalies of the ocean surface temperature. It is evident that this question can be answered only after additional numerical experiments are carried out.

REFERENCES

1. G. W. Moore, J. Halfar, H. Majeed, W. Adey, and A. Kronz, "Amplification of the Atlantic multidecadal

oscillation associated with the onset of the industrial era warming," Sci. Rep. **7,** 40861 (2017).

2. G. V. Alekseev, "Development and amplification of global warming in the Arctic," Fund. Prikl. Klimatol., No. 1, 6−21 (2015).

3. V. N. Malinin and P. A. Vainovskii, "On the causes of the first Arctic warming in the 20th century," Uch. Zap. RGGMU. Meteorol., No. 53, 34−55 (2018).

4. V. A. Semenov and M. Latif, "The early twentieth century warming and winter Arctic sea ice," Cryosphere **6,** 1231−1237 (2012).

5. T. Delworth, S. Manabe, and R. J. Stouffer, "Inter-decadal variations of the thermohaline circulation in a coupled atmosphere-ocean model," J. Clim. **6,** 1993−2011 (1993).

6. L. M. Frankcombe and H. A. Dijkstra, "The role of Atlantic − Arctic exchange in North Atlantic multidecadal climate variability," Geophys. Res. Lett. **38,** 16603/1−16603/5 (2011).

7. E. M. Volodin, "The nature of 60-year oscillations of the Arctic climate according to the data of INM RAS climate model," Russ. J. Numer. Anal. Math. Modelling **33** (6), 359−366 (2018).

8. E. M. Volodin, E. V. Mortikov, S. V. Kostrykin, V. Ya. Galin, V. N. Lykosov, A. S. Gritsun, N. A. Diansky, A. V. Gusev, and N. G. Yakovlev, "Simulation of modern climate with the new version of the INM RAS climate model," Izv., Atmos. Ocean. Phys. **53** (2), 142−155 (2017).

9. V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization," Geosci. Model Dev. **9,** 1937−1958 (2016).

10. E. M. Volodin, "Estimation of the contribution of different mechanisms to the phase evolution of quasi-biennial oscillation using the results of climate simulation," Izv., Atmos. Ocean. Phys. **55,** 32−37 (2019).

11. A. S. Gritsun, "Low frequency variability and sensitivity of the Atlantic meridional overturning circulation in selected IPCC climate models," Russ. J. Numer. Anal. Math. Modelling **33** (6), 341−350 (2018).

12. A. Bellucci and K. J. Richards, "Effect of NAO variability on the North Atlantic Ocean circulation," Geophys. Res. Lett. **33,** L02612 (2006).

13. B. Huang, Z. Hu, E. Schneider, Z. Wu, Y. Xue, B. Klinger, "Influences of tropical-extratropical interaction on the multidecadal AMOC variability in the NCEP climate forecast system," Clim. Dyn. **39,** 531−555 (2012).

Translated by L. Mukhortova