A Technology for Predictive Estimation of Meteorological Parameters on the Basis of the Global Climate Model CFSv2

V. A. Petrukhina*, N. V. Abasov

Melentiev Energy Systems Institute of Siberian Branch of Russian Academy of Sciences, Irkutsk, Russia

***Abstract* — The paper presents a technology for predictive estimation of meteorological indicators based on the global climate model CFSv2. This technology provides continuous monitoring of current and prognostic indicators of the state of the atmosphere (temperature, pressure, humidity, precipitation, etc.) in the catchment basin of Lake Baikal. The monitoring and data analysis tasks are briefly described as well as the operation algorithms for the main software components intended to obtain predictive distributions of weather indicators for the average values of a given time period, predictive scenarios of the dynamics of their changes for a selected point or a separate basin. The technology involves adjusting the weights of individual ensembles of predictive data of the global CFSv2 system, which provides more reliable predictive estimates.**

***Index Terms*: prognostic estimates, monitoring and data analysis, climate forecast system, climate maps, ensemble approach.**

I. Introduction

Long-term prognostic estimates for 3 months or more are important to manage the operation of hydroelectric power plants (HPP). This affects the generation of electricity in a region with a high share of HPP (for example, for the Irkutsk region it is about 80%). Melentiev Energy Systems Institute of Siberian Branch of the Russian Academy of Sciences (ESI SB RAS) has been developing long-term forecasting of nature-conditioned energy factors for a long time (for more than 60 years) [1–3]. The founders of this area of research are Academician I.P. Druzhinin and Professor A.P. Reznikov. The GIPSAR system, which was developed at the Institute and included various approximative and probabilistic methods [4, 5], made it possible to obtain sufficiently reliable prognostic estimates. Unfortunately, global climate change has disrupted many previously found patterns, which significantly reduced the reliability of prognostic results.

Modern global climate models make it possible to model and predict the climatic situation for various forecast horizons by considering many factors of the state of the atmosphere, ocean, and land. They allow producing long-term prognostic estimates based on the ensemble approach [6]. This approach suggests the technology for processing predictive ensembles and generates the most probable climatic conditions based on processing the data of the global climate model CFSv2 [7, 8].

II. Global Model CFSv2

Data processing of the global climate model *Climate Forecast System* (CFS) is implemented by separate components of the GeoGIPSAR information and climate system, which is a GIPSAR system extension, which includes the matrices of global data grid. This system was developed at the NCES Environmental Modeling Center [9]. The model is fully coupled and shows the interactions among the Earth's atmosphere, oceans, land, and sea ice. The main advantage of CFS is the openness of the data it provides.

The main purpose of the data analysis in CFSv2 is to create long-term global grid representations of atmospheric states generated by the model and the data assimilation system. The use of operational data has made progress in climate research by eliminating fictitious trends caused by model changes and data assimilation by second.

Data sets from the first version of CFS were collected and converted into the form required for the second version, which was a difficult and time-consuming task. This format was brought in line with international standards for the storage and exchange of observational data. Thousands of graphics editors of the model are generated automatically at the end of each reanalyzed month according to the first version and are displayed on the Climate Forecast System Reanalysis (CFSR) website in real time. In contrast to the first CFS version, CFSv2 has the following new features:

* Corresponding author.
E-mail: petruhina@isem.irk.ru

http://dx.doi.org/10.38028/esr.2022.04.0002
Received November 08, 2022. Revised November 11, 2022.
Accepted November 23, 2022. Available online January 30, 2023.

- Analysing ocean parameters every 6 hours.
- Building an interactive ice model.
- Processing satellite signals received over 24 hours.

The main indicators of CFSv2 predictive ensembles, which are used in the GeoGYPSAR system, are precipitation intensity, surface air temperature, pressure above sea level, geopotential indicators on isobaric surfaces of 500 and 850 gpa, and atmospheric circulation rates [10].

Other models can also be used for forecasting, but most of them have commercial security.

III. The Main Components of Geoclimatic Data Processing

The GeoGIPSAR system is implemented in the form of various interconnected components that perform certain tasks. It includes the tools for adding new components and developing the existing ones. The main component aims to obtain the most probable distributions of atmospheric processes for an arbitrary time period [11, 12].

The system components are divided into 2 groups.

The first group includes:

1. Component for internet monitoring of new data of predictive ensembles in the CFSv2 Global Model Data Center [13].
2. Component for converting, aggregating, and writing to a data warehouse.

The second group of components includes:

1. Component for analysis and generation of prognostic indicators (averages for the period of climatic spatial distributions of indicators or their deviations from the norm).
2. Component for building prognostic scenarios of changes in meteorological indicators for a selected point or basin in comparison with the range of changes in actual data.
3. Component for visual representation of probabilistic distributions of the studied indicator for the selected point according to various sets (or weights) of individual ensembles.
4. Component for verification of prognostic indicators based on their comparison with actual data and refinement of the weighting coefficients of various ensembles.

After a reliable prognostic scenario of climatic parameters for the time period under study is obtained [14, 15], an algorithm for obtaining the inflow through the determination of the closest analogous years is applied. Such years are determined based on the spatial distribution over the past similar periods to minimize the proximity of comparative indicators in the following form:

$$\sigma(e,y,P)=\frac{\sum_{(p)} c_p\cdot\left(\sum_{(i,j)} d_{ij}(i,j)\cdot(p_{ij}^e-p_{ij}^y)^2\right)^{0.5}}{\sum_{(p)} c_p\cdot\sum_{(i,j)} d_{ij}(i,j)}, \quad (1)$$

$$c_p\ge 0,\ d_{ij}(i,j)\ge 0,\quad y=\overline{y_1,y_2},$$

where:

e, y denotes the studied periods of the season and those similar for other years (in a range of years);

$p \in P$ defines the type of parameter with a weighting factor c_p from a given set P;

$d_{ij}(i,j)$ is a specified weight function depending on the coordinates of the region i, j, with the maximum values in the area of the studied catchment basin;

p_{ij}^e, p_{ij}^y are aggregated indicators of the season period e and y years for each cell of the coordinate grid of the region.

An operatively formed set of the closest years with indicators $\sigma(e, y, P) \le \sigma_0$ allows obtaining estimates of the dynamics of changes in water content for the nearest period.

IV. Description of the Technology for Predictive Estimation

The technology for predictive estimation of weather indicators based on the global CFSv2 model relies on an integrated approach, which allows considering weather indicators from different angles and to quickly make estimates of water content within the specified limits of the catchment basin. This approach makes it possible to increase the reliability of predictive estimates of water inflows into hydroelectric power plant (HPP) reservoirs. This is necessary for the effective management of HPP operating modes. The software components of the GeoGIPSAR technology perform the following tasks:

1. Internet monitoring allows downloading the necessary data of predictive ensembles of various weather indicators in a binary GRIB format, which has a complex structure with information compression.
2. Data conversion, aggregation, and recording in the GeoGYPSAR data warehouse in the form of specialized format files, which include daily averages.
3. Construction of prognostic maps based on absolute and relative weather indicators.
4. Generation of prognostic probable scenarios of a meteorological indicator for a point or a selected basin.
5. Estimation of the probabilistic distribution of the weather indicator component for the selected period with the calculation of the main statistical characteristics.
6. Determination of the weight coefficients of individual ensembles.
7. Verification based on comparison of prognostic and actual data.
8. Visualization of predictive maps and graphs.

Figure 1 shows the interaction of software components in the tasks of Internet monitoring, data conversion and recording in the of CFS ensemble storage. Daily automatic Internet monitoring provides new predictive ensembles periodically downloaded from the CFSv2 Data Center. Next, the downloaded GRIB format files with a complex binary structure are converted to a lighter and more flexible CFS format, which is necessary for further efficient

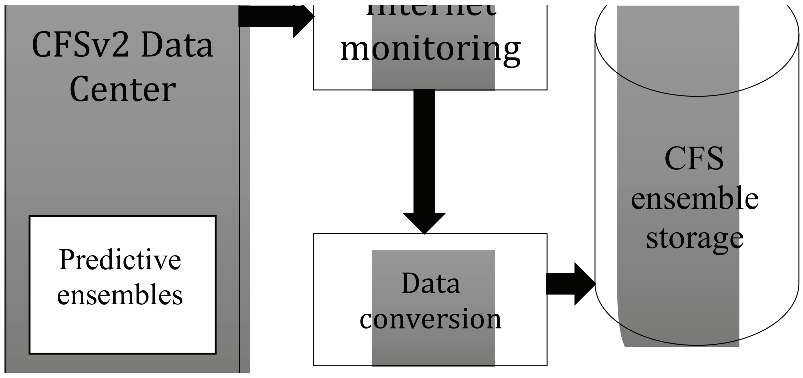

Fig. 1. Interaction of software components of the technology of forming predictive estimates.

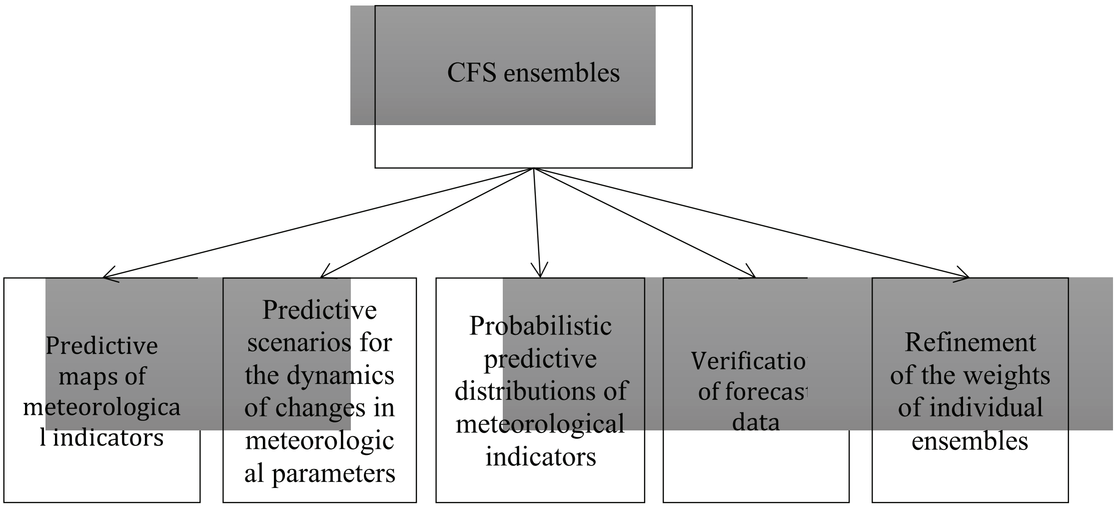

Fig. 2. Application of CFS ensembles within the technology for predictive estimation of meteorological indicators.

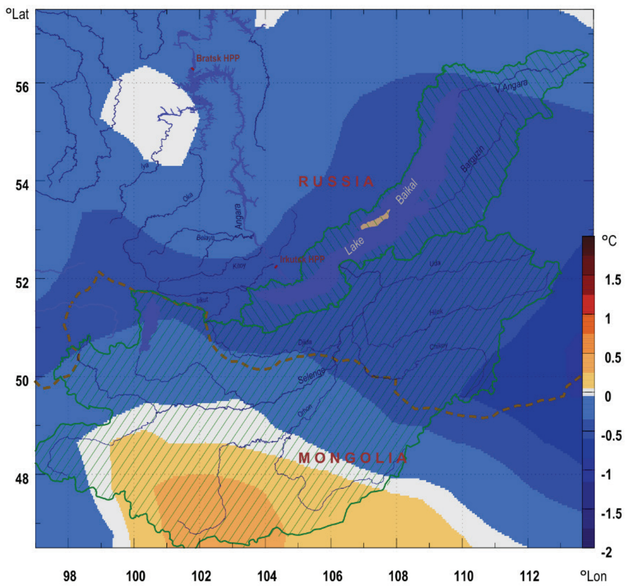

Fig. 3. The air temperature anomalies for June 2022.

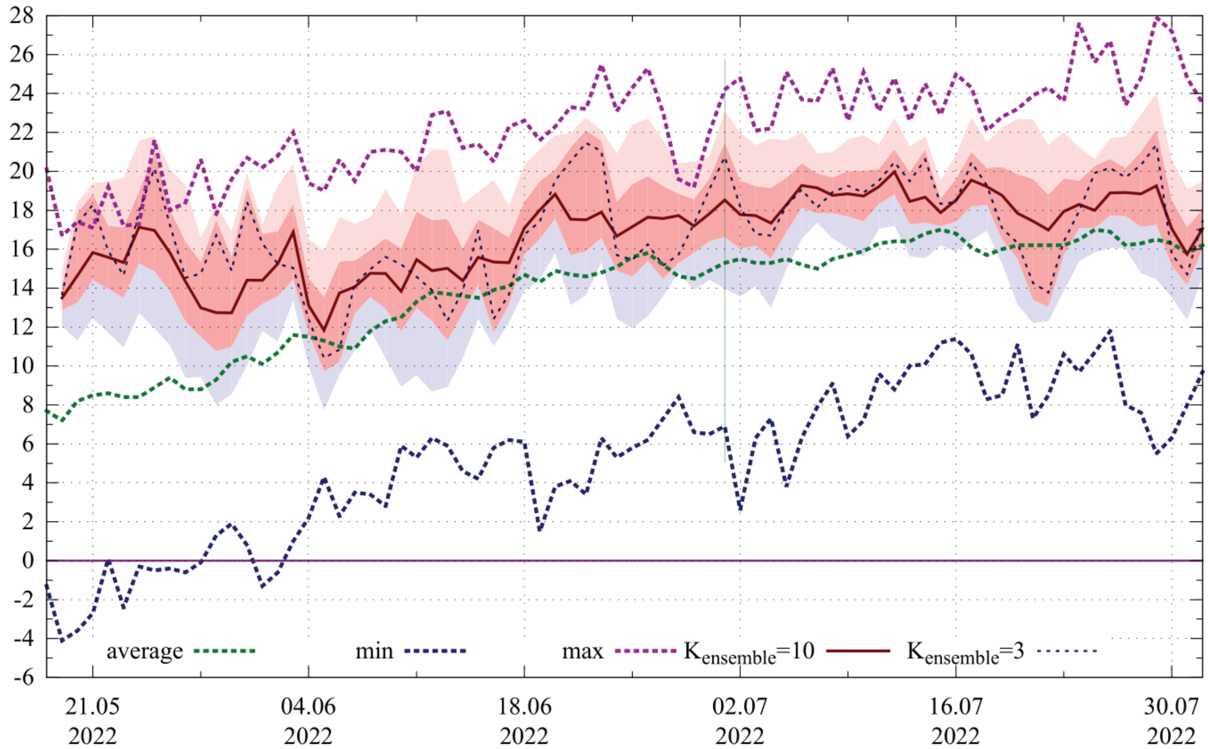

Fig. 4. A predictive scenario of the dynamics of temperature changes for the summer period for the village of Kyakhta.

processing. The converted files are accumulated in the CFS ensemble storage.

The second group of the software components focuses on the use of CFS ensembles for various purposes (Fig. 2). The predictive maps of weather indicators are constructed using pluggable forecasting parameters (a selection of ensembles used, dates of the forecast period, the number of ensembles, connected GIS data, terrain coordinates, the predicted weather indicator, etc.).

The predictive scenario of the dynamics of changes in meteorological indicators is based on the forecast ensembles of CFS, given the range of its daily fluctuations according to the actual data of reanalysis [16].

Probabilistic prognostic distributions show the range and probabilities of the values of the studied meteorological indicator for a given period in comparison with the actual data for the given period.

V. Approach to Processing Predictive Ensembles

The components of processing the predictive ensembles allow constructing predictive maps of meteorological indicators (precipitation, temperature, pressure, etc.) for the studied catchment basins for a given period (up to 9 months); build predictive scenarios of changes in meteorological indicators for river basins; and generate probabilistic distributions of prognostic indicators for the studied period with operational calculation of the main statistical characteristics.

The generation of predictive indicators for a specific meteorological parameter for a certain coordinate grid (selected region) employs absolute and relative indicators, the general form of which is represented by formula:

$$P(k,t)=\left\{\overline{p_{ij}}(k,t),\ i=1,\ldots,N_x,\ j=1,\ldots,N_y\right\}. \quad (2)$$

To obtain predictive indicators with equal weights of ensembles, the formula (3) is used:

$$\overline{p_{ij}}(k,t)=\frac{1}{T}\sum_{t=1}^{T}\overline{p_{ij}}(k,t), \quad (3)$$

Formula (4) is used to determine predictive indicators given the weighting coefficients of ensembles c_k:

$$\overline{P_{ij}}(k,T)=\frac{\sum_{k=1}^{K}c_k\times p_{ij}(k,T)}{\sum_{k=1}^{K}c_k},\ 0\le c_k\le 1, \quad (4)$$

where:
p_{ij} – geoclimatic parameter;
t – time indicator;
N_x – latitude coordinate;
N_y – longitude coordinate;
c_k – weight of the ensemble;
K – number of predictive ensembles;
k – ensemble for a specific parameter;
i, j – spatial distribution indices (latitude, longitude).

The technology also involves the selection of the weighting coefficients of the influence of individual ensembles to obtain final indicators, and has procedures for verification based on actual data. The accuracy of forecasting weather indicators can be increased by

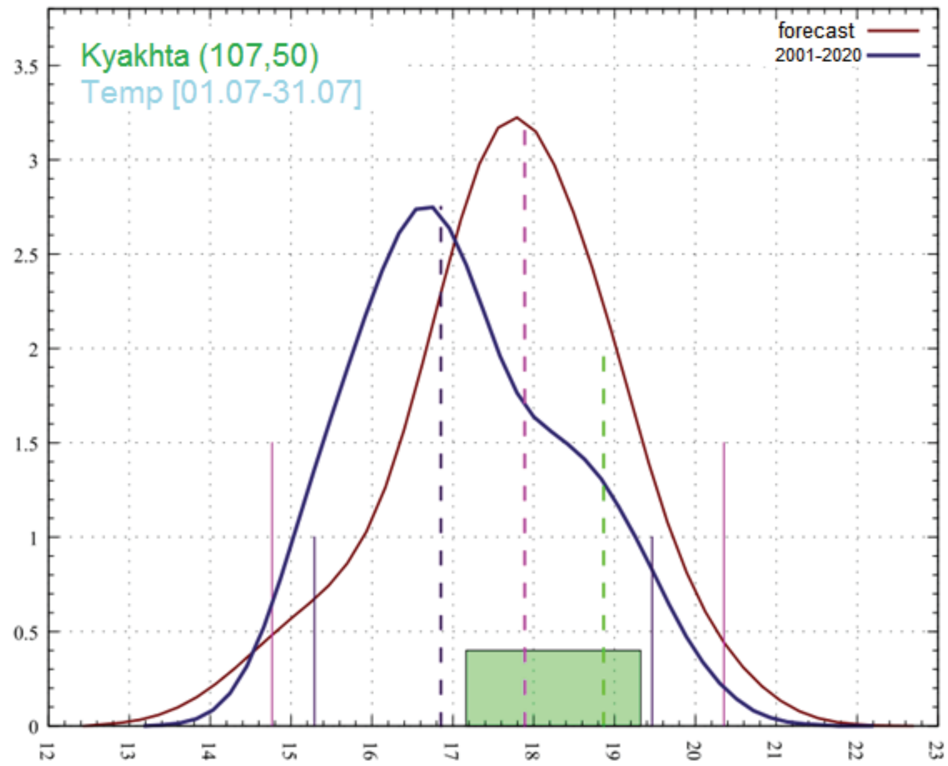

Fig. 5. Probabilistic prognostic and actual (for 2001-2020) distributions of the average temperature of July 2022 in the village of Kyakhta according to the data of prognostic ensembles for the period (01.03-30.04), 2022.

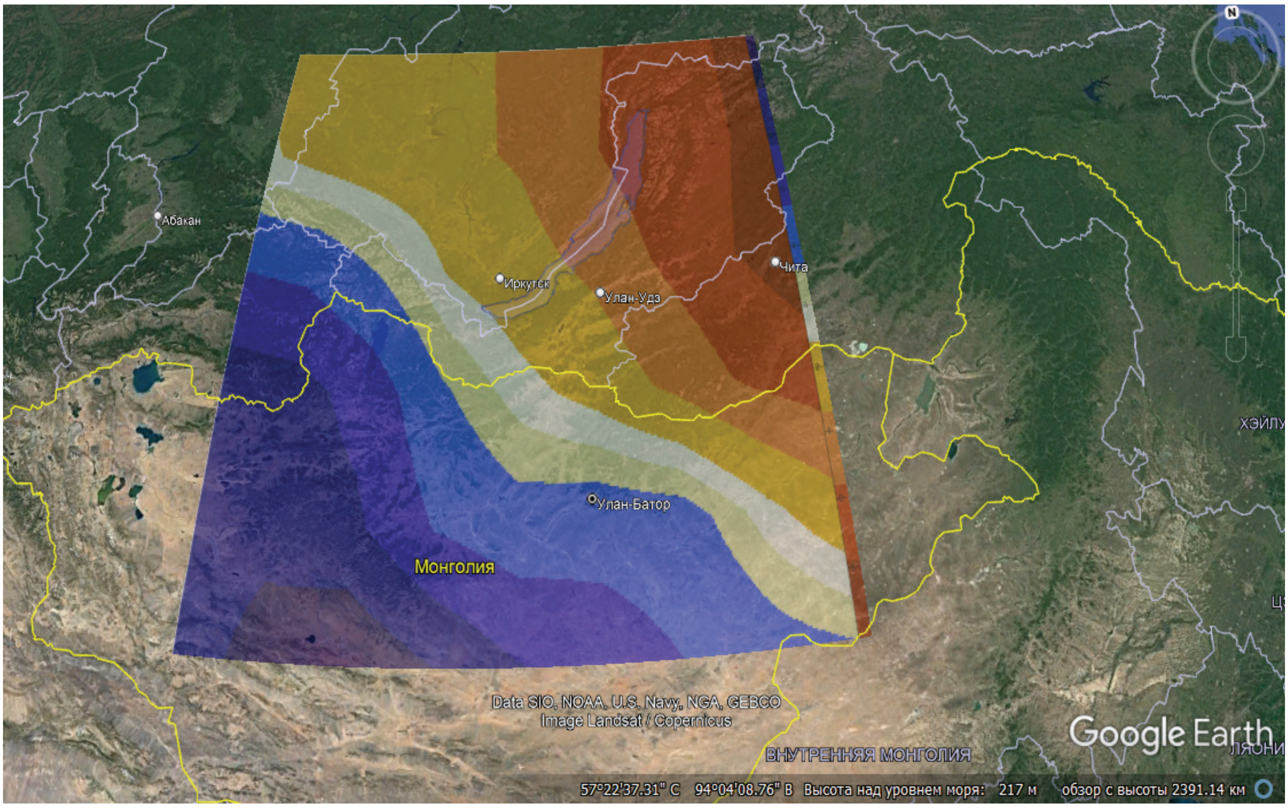

Fig. 6. An example of using a predictive map as a layer in Google Earth.

assigning a weight to an individual data ensemble and varying it depending on the importance. To assign a weight to the ensemble, it is necessary to analyze the degree of influence on the predicted time period.

VI. Examples of Using the technology to Predict Climate Indicators

The performance of the technology can be seen in Fig. 3. It shows a geoclimatic map for temperature anomalies for June 2022. The map is built by processing the data ensembles for the months of January–February. For more accurate results, it is necessary to periodically update the data of the forecast period and adjust the weighting coefficients of the influence of individual ensembles.

A set of accumulated ensembles can be used to build a predictive scenario of the dynamics of changes in the meteorological indicator for a given period. For example, Figure 4 shows the dynamics of temperature changes for the summer period for the village of Kyakhta. The red area highlights the most likely range, in which the air temperature indicator will be located. The solid red line shows the averaged most probable temperature based on processing 10 data ensembles. The blue and pink dashed lines show the lowest and highest possible average daily temperatures.

Figure 5 shows a graph of the probabilistic predictive and actual distribution of the average temperature of July 2022 in the village of Kyakhta according to the data of prognostic ensembles for the months of March and April. The green area is a refinement of the last 3 ensembles, the dashed lines on the graph are the medians of the actual (blue line) and predictive (pink line) values. At the same time, the data beyond 5–95% can be discarded. The actual data are taken into account for the period from 2001 to 2020, which is associated with global climate changes compared to the data of the 20th century.

The predictive indicators obtained are represented by geoclimatic maps, which can have various formats. The technology enables the transformation of maps to overlay a predictive indicator in the form of a layer in standard GIS systems in various services. For example, in Figure 6, a distribution in the form of a kml file is added as a separate layer to the standard GIS system Google Earth.

VII. Conclusion

The technology for predictive estimation of meteorological indicators based on the global climate model CFSv2 provides operational assessments of the state of the atmosphere in the studied catchment basin or at a desired point on the map by coordinates. The estimation at issue in combination with other methods employed in GeoGIPSAR allows experts to build the most probable prognostic picture. This approach, however, does not provide unambiguous prognostic estimates in the case of a long lead time and requires constant refinement (at least once a month) with the possibility of changing prognostic distributions. These data need to be coordinated with the data obtained using other approaches, such as neural network models, to adjust and refine the weighting coefficients of the influence of individual initial ensembles on the final result. The development of this technology in this direction will improve the accuracy of predictive estimates of meteorological parameters.

The proposed software components have a flexible structure and wide possibilities for their development. The technology also allows adding new methods of analysis and forecasting of energy-significant meteorological indicators.

Acknowledgments

The research was carried out under State Assignment Project (no. FWEU-2021-0003, no. FWEU-2021-0008) of the Fundamental Research Program of the Russian Federation 2021–2030 using the resources of the High-Temperature Circuit Multi-Access Research Center (Ministry of Science and Higher Education of the Russian Federation, project no. 13.CKP.21.0038).

References

[1] N. V. Abasov, “The system of long-term forecasting and analysis of environmental factors of energy GeoGIPSAR,” in *Proc. Int. Meeting APN (Enviromis-2012) “Extreme manifestations of global climate change in North Asia”*, Irkutsk, Russia, 2012, pp. 63–66. (In Russian)

[2] N. V. Abasov, T. V. Berezhnykh, A. P. Reznikov, “Long-term forecasting of environmental factors in the energy sector,” in *Systematic studies of energy problems*, N. I. Voropai, Ed. Novosibirsk, Russia: Nauka, 2000, pp. 415–429. (In Russian)

[3] N. V. Abasov, T. V. Berezhnykh, V. M. Nikitin, E. N. Osipchuk, “Methodology for the formation of long-term prognostic scenarios of water content and temperature regimes,” in *Conf. abstract book “Energy of the 21st century: Sustainable development and intelligent management”*, Irkutsk, Russia, 2020, pp. 100–101. (In Russian)

[4] N. V. Abasov, “On the development of the predictive system of GIPSAR based on Lua components,” in *Proc. XVI Baikal All-Russian Conference “Information and mathematical technologies in science and management,”* Irkutsk, Russia, 2011, pp. 212–217. (In Russian)

[5] N. V. Abasov, T. V. Berezhnykh, A. P. Reznikov, “Long-term forecast of nature-conditioned energy factors in the GIPSAR information and prognostic system,” *Izvestiya RAS. Energy*, no. 6, pp. 22–30, 2000. (In Russian)

[6] N. V. Abasov, T. V. Berezhnykh, V. V. Vetrova, O. Yu. Marchenko, E. N. Osipchuk, “Analysis and forecasting of the Baikal region hydropower potential under the conditions of varied climate,” in *Risks and Opportunities of the Energy Sector in East Siberia and the Russian Far East: For Better Risk Management and Sustainable Energy Development*, Sangtu Ko and

Kyong Wan Lee, Eds. Berlin, Germany: Lit Verlang Dr., 2012, pp. 173–186.

[7] A. Rai, S. K. Saha, "Evaluation of energy fluxes in the NCEP climate forecast system version 2.0 (CFSv2)," *Clim Dyn.*, vol. 50, pp. 101–114, 2018.

[8] Xing Yuan, Eric F Wood, Lifeng Luo, Ming Pan, "A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction," *Geophysical Research Letters*, vol. 38, no. 13, Jul. 2011.

[9] A. Rai, S. K. Saha, "Evaluation of energy fluxes in the NCEP climate forecast system version 2.0 (CFSv2)," *Clim Dyn.*, vol. 50, pp. 101–114, 2018.

[10] K. Schamm, M. Ziese, A. Becker, P. Finger, A. Meyer-Christoffer, U. Schneider, M. Schröder, P. Stender, "Global gridded precipitation over land: a description of the new GPCC First Guess Daily product," *Earth System Science Data*, vol. 6, pp. 49–60, 2014. DOI: 10.5194/essd-6-49-2014.

[11] S. E. Hampton, L. R. Izmest'eva, M. V. Moore, S. L. Katz, B. Dennis, E. A. Silow, "Sixty years of environmental change in the world's largest freshwater lake – Lake Baikal, Siberia," *Glob Chang Biol.*, vol. 14, no. 8, pp. 1947–1958, 2008.

[12] N. V. Abasov, V. M. Nikitin, T. V. Berezhnykh, E. N. Osipchuk, "Monitoring and Predictive Estimations of Atmospheric Parameters in the Catchment Area of Lake Baikal," *Atmosphere*, vol. 13, no. 1, pp. 49, 2022. DOI: 10.3390/atmos13010049.

[13] Reanalysis data NOAA [Online]. Available: ftp://ftp.cdc.noaa.gov/Datasets/ncep.reanalysis.dailyavgs. Accessed on Apr. 25, 2022.

[14] T. V. Berezhnykh, O. U. Marchenko, N. V. Abasov, "Changes in the summertime atmospheric circulation over East Asia and formation of long-lasting low-water periods within the Selenga river basin," *Geogr. Nat. Resour.*, vol. 33, no. 3, pp. 223–229, 2012. (In Russian)

[15] R. Törnqvist, J. Jarsjö, J. Pietroń, "Evolution of the hydro-climate system in the Lake Baikal basin," *Journal of Hydrology*, vol. 519, pp. 1953–1962, 2014.

[16] The center's portal GPCC [Online]. Available: https://www.dwd.de/EN/ourservices/gpcc/gpcc.html. Accessed on Apr. 20, 2022.

Valeria Petrukhina graduated from Irkutsk State Transport University with a degree in Information Systems and Technologies. She is currently a research engineer at the Melentiev Energy Systems Institute SB RAS. Her research interests include the design of information systems and the construction of predictive models to improve the efficiency of hydroelectric power plants in the context of a changing climate.

Nikolai Abasov graduated from Moscow Physical Technical Institute in 1978. Currently he is a senior engineer at the Melentiev Energy Systems Institute SB RAS. His research interests include the development of new techniques and methodologies to improve the efficiency of hydroelectric system in the context of a changing climate.