Development and Evaluation of National-Scale Operational Hydrological Forecasting Services in Russia

Georgy Ayzel[a] and Aleksei Sorokin[b]

[a] *Russian State Hydrological Institute, 2-ya Liniya V.O. 23, Saint Petersburg, 199053, Russia*
[b] *Computing Center of the Far Eastern Branch of the Russian Academy of Sciences, 65 Kim Yu Chena Ulitsa, Khabarovsk, 680000, Russia*

Abstract

The presented study briefly describes the development and evaluation of two recently developed national-scale operational hydrological forecasting services in Russia — OpenForecast, and OpenLevels. The developed services are based only on open data and software that ensures their transparency and reproducibility. Launched in March 2020, both services provide reliable short-term hydrological forecasts for hundreds of river basins across Russia for more than one year. The forecasts of river discharge and water level have been openly disseminated on the internet, allowing any interested party to convert them into value by guiding decision-making and early warning of extreme events. Further development of OpenForecast and OpenLevels services requires a transition to modern computational techniques (e.g., deep learning) with high computing power and reliability requirements. Thus, this contribution also seeks collaboration with computer scientists and experts in high-performance computing to boost the value of introduced forecasting services.

Keywords 1

Russia, Hydrology, Forecasting, Services, Modelling

1. Introduction

Operational hydrologic forecasting systems aim to provide up-to-date information about the current state of water-related variables, such as river discharge or water level, and their reliable forecast. These systems have particular relevance and high importance for many parties, supporting water management decisions and ensuring timely early warning of extreme events. The development of operational hydrologic forecasting systems has coincided with the emerging cutting-edge computational resources and could be dated back to the early 1980s. Since then, hydrologists have worked back-to-back with computer scientists to establish new and support existing forecasting systems [1, 2].

Many existing hydrological forecasting systems have been operationalized at different spatial scales: from a basin- to global [3, 4]. These systems differ from each other by the type of underlying hydrological model, meteorological forcing, utilized weather forecast products, and the way of forecast dissemination [5]. However, one thing remains common for every system — the value they aim to provide for the end-user. That turned out to be more critical during the covid-19 pandemic when people moved from cities to suburban areas where the proximity to rivers is much closer. Therefore there is a higher vulnerability to floods.

A long-lasting and still ongoing crisis in Russian hydrometeorology hinders the development of operational forecasting systems in Russia. It puts the end-users — the country's citizens — out of focus for decades. Fortunately, available data, technologies, the passion of young researchers, and the

VI International Conference Information Technologies and High-Performance Computing (ITHPC-2021), September 14–16, 2021, Khabarovsk, Russia
EMAIL: hydrogo@yandex.ru (A. 1); alsor@febras.net (A. 2);
ORCID: 0000-0001-5608-9110 (A. 1); 0000-0003-3334-3440 (A. 2)

support of the Russian Fund for Basic Research (RFBR) have recently reached the critical mass that helped shift the situation from the dead point towards the development of the first national-scale forecasting services.

In the presented study, we want to introduce two national-scale hydrological forecasting systems — OpenForecast [6] and OpenLevels [7] —, which are in operational use for almost two years, providing timely and reliable discharge and water level forecasts for up to seven days for hundreds of hydrological gauging stations in Russia.

The paper is organized as follows. In Sect. 2, we describe the components of both developed systems. Section 3 describes the obtained results for the first year of systems‘ operational use. We provide an outlook for further directions of research and development in Section 4.

2. Operational services

In the present section, we describe two operational hydrological forecasting services in Russia – OpenForecast [6] and OpenLevels [7]. These services are conceptually different in both underlying methods and forecasted variables. OpenForecast utilizes meteorological forecasts and hydrological models to provide runoff forecasts up to seven days at daily temporal resolution. Instead, OpenLevels uses a machine learning model fed by water level observations to provide water level forecasts up to seven days at daily temporal resolution. Both services are in operational use since March 2020.

2.1. OpenForecast

The first version of the OpenForecast system was launched in operation in summer 2018 [8]. Since then, many improvements have been incorporated, and the second version has been released in March 2020 [9]. A detailed description of the second version's operational setup is provided in [9].

OpenForecast disseminates a seven-day-ahead runoff forecast for 834 hydrological stations on Russian rivers in different geographical conditions across Russia. The system is entirely based on open data [10, 11, 12] and open-source software [13, 14] that ensures transparency and reproducibility. There are two major computational blocks – offline and online (Figure 1).

Offline routines aim to obtain optimal parameters for hydrological models GR4J [15] and HBV [16], which underpin runoff calculation. Online routines run daily to update and further disseminate runoff forecasts.

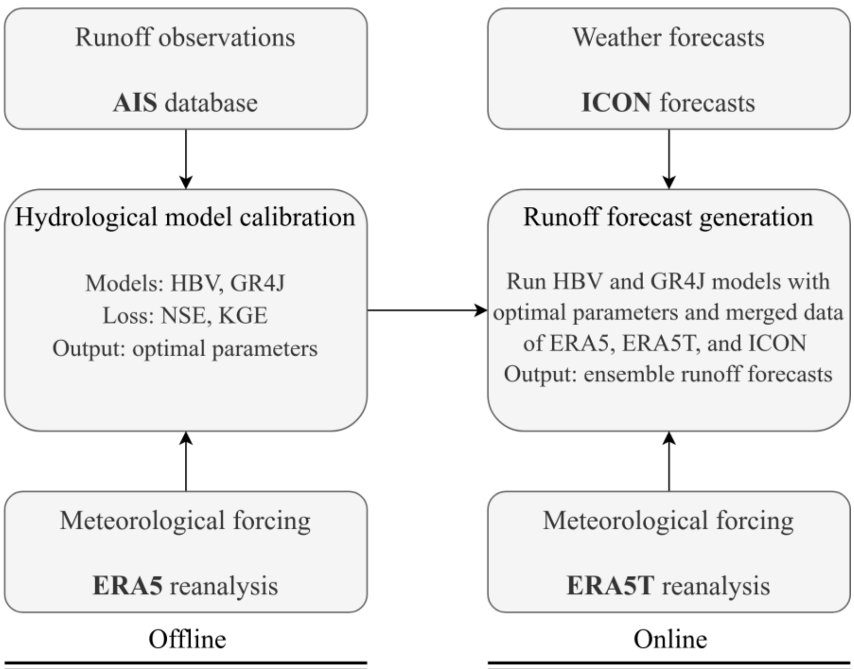

Figure 1: Illustration of the OpenForecast computational workflow

OpenForecast is hosted on a standard cloud server with 4Gb of RAM and two CPU threads. The average computational time for the entire system's run from meteorological data update to publication on the website is 75 min, from which 60 min are used for runoff calculation by hydrological models.

2.2. OpenLevels

OpenLevels is an entirely data-driven system that utilizes operational water level observations at gauging stations and uses a machine-learning model to extrapolate those observations in the imminent future based on the found trends (Figure 2). Similar to OpenForecast, OpenLevels also uses only open data and software. The Unified State System of Information website regarding the situation in the World Ocean (ESIMO) [17] is the provider of operational water level observations. There are more than 1000 gauges in Russia, observations from which are operationally delivered using ESIMO. We use the fbprophet model developed by Facebook [18] as an underlying model for time series forecasting. fbprophet is based on an additive regression model with main components for fitting trend, different seasonality components, and change points [18]. In this way, fbprophet could be considered as a simple yet powerful tool for time-series forecasting.

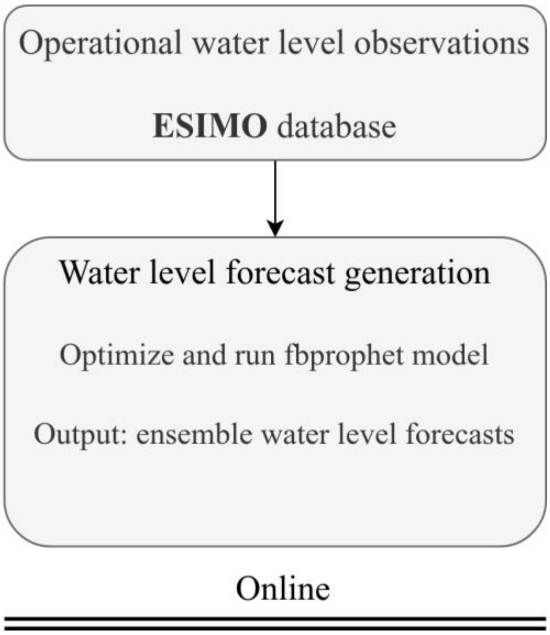

Figure 2: Illustration of the OpenLevels computational workflow

The standard run of OpenLevels starts from the update of water level observations in the ESIMO database. Then, for each station, the individual instance of the fbprophet model is fitted based on the entire time series of obtained observations. After that, individual model instances run in the forecasting model to provide expected predictions in the imminent future for up to seven days and their uncertainty estimates.

OpenLevels is hosted on the same server as OpenForecast. However, their computational routines are separated in time so as not to overlap each other. The average computational time for the entire run of the system is 135 min, from which 120 min are used for model fitting.

3. Results

The evaluation of OpenForecast and OpenLevels predictive efficiencies has been performed using ESIMO data of observed water levels [17]. The evaluation period is from March 2020 to May 2021 and covers all phases of the water regime for investigated river basins (gauges). It should be mentioned that ESIMO data do not undergo quality control and experience long periods during which the data does not enter the system. In this way, we have performed evaluation exercise only for basins where more than a full year of data (at least 365 values) is presented. Also, water level observations have not been available for many basins presented in the OpenForecast system. In this way, the number of basins used for evaluation is 394 for OpenForecast (Figure 2) and 1017 for OpenLevels (Figure 3). We used the Pearson correlation coefficient as a significant evaluation metric. For the sake of brevity, we have calculated the correlation coefficient between forecast and observed values of runoff (or water level) only for the longest lead time of 7 days. That informs us about the lower envelope of the prediction efficiency as efficiency decreases with lead time.

Figure 3 shows the spatial distribution of the correlation coefficient calculated for OpenForecast's runoff forecast and water level observations for a lead time of 7 days. The median efficiency in terms of the correlation coefficient is 0.83 with the interquartile range of 0.21 (0.69 and 0.9 for the 25th and 75th quantiles, respectively). In this way, OpenForecast shows a solid performance for the majority of the analyzed basins in Russia. The most pronounceable deterioration of OpenForecast efficiency is related to the lower reaches of the Kama river. However, it could be considered an outlier because we have abnormally high rates of river runoff for five gauges located there. We need to conduct an additional investigation to reveal possible causes of such behavior.

During the last year, OpenForecast service has been actively used to inform Russian citizens about the current and expected short-term changes in runoff. Furthermore, local hydrometeorological agencies and municipal authorities used OpenForecast results to guide early warning and decision making. In this way, OpenForecast provided a total value for society.

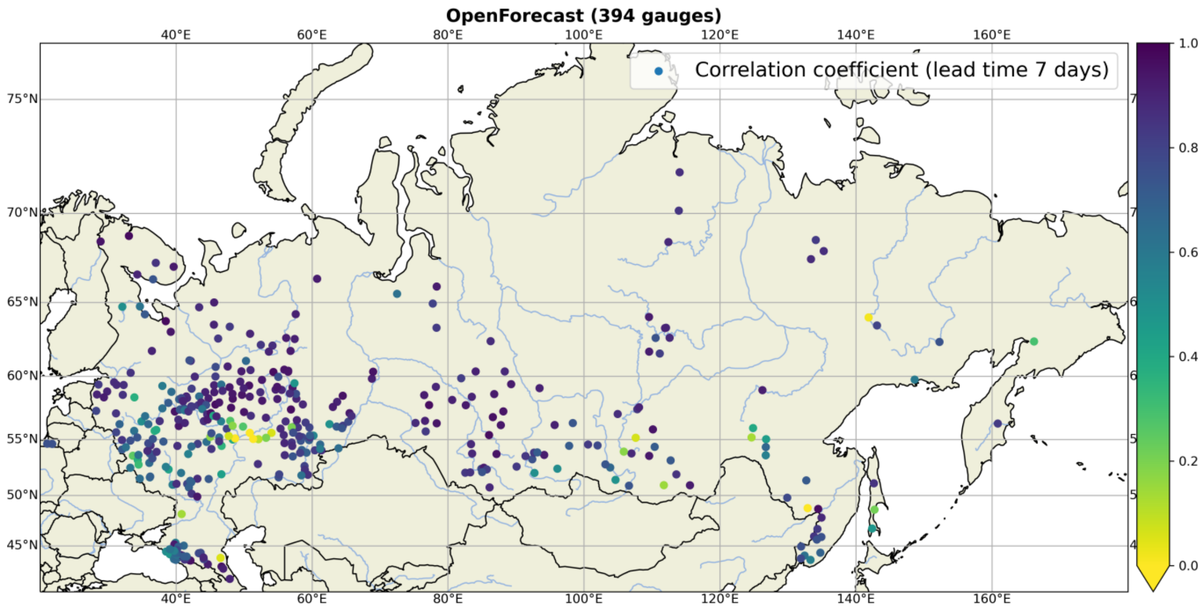

Figure 3: The efficiency of OpenForecast in terms of the correlation coefficient for the lead time of seven days

Figure 4 shows the spatial distribution of the correlation coefficient calculated for OpenLevel's runoff forecast and water level observations for a lead time of 7 days. The median efficiency in terms of the correlation coefficient is 0.87 with the interquartile range of 0.18 (0.76 and 0.94 for the 25th and 75th quantiles, respectively). Similar to OpenForecast, OpenLevels demonstrated a solid performance for the entire territory of Russia.

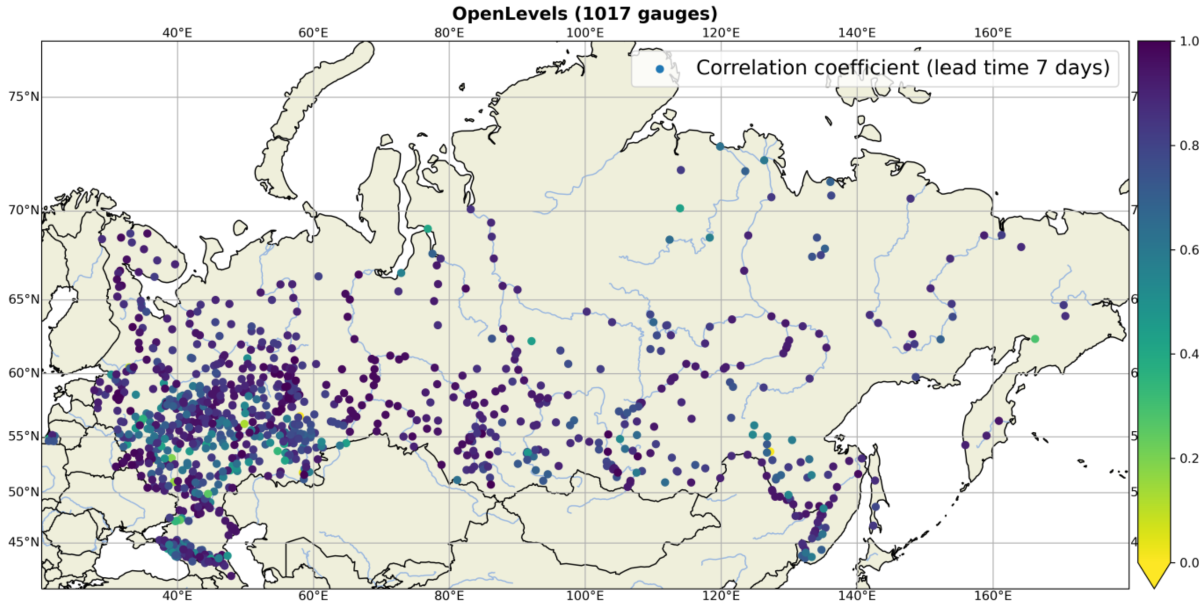

Figure 4: The efficiency of OpenLevels in terms of the correlation coefficient for the lead time of seven days

In contrast to OpenForecast, OpenLevels service has not been extensively promoted in social media due to its premature, proof-of-concept status. However, the results of the preliminary evaluation of its predictive efficiency (Figure 4) show that the machine learning-based fbprophet model provides reliable data-driven water level forecasts.

The noticeable difference between OpenForecast and OpenLevels performances has an obvious explanation. OpenForecast does not use any operationally obtained data for utilization in its computational workflow. Instead, OpenLevels always assimilates the most recent observations from the hydrological gauging network to fit up-to-date models. In this respect, OpenLevels substantially benefits from that kind of data assimilation. However, there are many strategies for incorporating data assimilation routines into OpenForecast's workflow that may lead to a vivid boost of predictive efficiency.

4. Conclusions and Outlook

The presented study aims to provide the preliminary results of the one-year-long evaluation of national-scale operational forecasting services in Russia — OpenForecast, and OpenLevels. Results show that both services provide efficient and reliable short-term forecasts of runoff (Figure 3) and water level (Figure 4). Furthermore, the introduced services are entirely based on open data and software (Figures 1, 2) that ensure transparency and reproducibility. Both OpenForecast and OpenLevels services are freely available on the Internet and could be accessed 24/7. Our services have been actively used to inform every interested person or organization about the current and expected short-term changes on Russian rivers. Local hydrometeorological agencies and municipal authorities have used forecasts to guide decision-making and early warning alerts. We hope to increase further the value of our services for any interested party.

It should be mentioned that the development of these first national-scale hydrological forecasting services in Russia would not happen without hydrological data that has been carefully obtained, processed, stored, and openly distributed by the Russian Federal Service for Hydrometeorology and Environmental Monitoring.

Experience that has been accumulated during the continuous operational run of the introduced forecasting services provides a solid set of possible improvements that have a high potential to boost services' efficiency, reliability, and value. Among them are operational quality-control of observations, sub-daily predictions, utilization of emerging deep learning models for runoff prediction and forecasting [19], data assimilation, utilization of commercial yet open for research data from regional providers [20]. All these would require a critical build-up of computing resources and increasing their reliability.

We are looking forward to further fostering the collaboration with experts in the field of computer sciences and high-performance computing to ensure the robust development of OpenForecast and OpenLevels services.

5. Acknowledgments

The present study was funded by the Russian Foundation for Basic Research (RFBR), project no. 19-05-00087 A.

The studies were carried out using the resources of the Center for Shared Use of Scientific Equipment "Center for Processing and Storage of Scientific Data of the Far Eastern Branch of the Russian Academy of Sciences", funded by the Russian Federation represented by the Ministry of Science and Higher Education of the Russian Federation under project No. 075-15-2021-663.

6. References

[1] R. Emerton, E. Zsoter, L. Arnal, H. L. Cloke, D. Muraro, C. Prudhomme, E. M. Stephens, P. Salamon, F. Pappenberger, Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0, Geosci. Model Dev. 11 (2018) 3327–3346, doi: 10.5194/gmd-11-3327-2018.

[2] S. Harrigan, E. Zoster, H. Cloke, P. Salamon, C. Prudhomme, Daily ensemble river discharge reforecasts and real-time forecasts from the operational Global Flood Awareness System, Hydrol. Earth Syst. Sci. Discuss. [preprint] (2020), doi: 10.5194/hess-2020-532.

[3] R. Emerton, E. Stephens, F. Pappenberger, T. Pagano, A. Weerts, A. Wood, P. Salamon, J. Brown, N. Hjerdt, C. Donnelly, C. Baugh, H. Cloke, Continental and global scale flood forecasting systems, WIREs Water 3 (2016) 391-418, doi: 10.1002/wat2.1137.

[4] W. Wu, R. Emerton, Q. Duan, A. Wood, F. Wetterhall, D. Robertson, Ensemble flood forecasting: Current status and future opportunities, WIREs Water. 7 (2020), doi: 10.1002/wat2.1432.

[5] H. Cloke, F. Pappenberger, Ensemble flood forecasting: A review, Journal of Hydrology 375 (2009) 613-626, doi: 10.1016/j.jhydrol.2009.06.005.

[6] OpenForecast, Runoff forecasts for Russian rivers, 2021. URL: http://www.openforecast.github.io.

[7] OpenLevels, Water level forecasts for Russian rivers, 2021. URL: http://www.openlevels.github.io.

[8] G. Ayzel, N. Varentsova, O. Erina, D. Sokolov, L. Kurochkina, V. Moreydo, OpenForecast: The First Open-Source Operational Runoff Forecasting System in Russia, Water 11(8) (2019) 1546, doi: 10.3390/w11081546.

[9] G. Ayzel, OpenForecast v2: Development and Benchmarking of the First National-Scale Operational Runoff Forecasting System in Russia, Hydrology 8(1) (2021) 3, doi: 10.3390/hydrology8010003

[10] AIS, the Automated Information System for State Monitoring of Water Bodies, 2021. URL: https://gmvo.skniivh.ru/.

[11] ICON, Deterministic forecast of the ICON model by the German Weather Service, 2021. URL: https://opendata.dwd.de/weather/nwp/icon/.

[12] ERA5 hourly data on single levels from 1979 to present, 2021. URL:https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels.

[13] LHMP, the Lumped Hydrological Models Playgroud, 2016. URL: https://github.com/hydrogo/LHMP.

[14] G. Ayzel, Runoff Predictions in Ungauged Arctic Basins Using Conceptual Models Forced by Reanalysis Data, Water Resour 45 (2018) 1–7, doi: 10.1134/S0097807818060180.

[15] C. Perrin, C. Michel, V. Andréassian, Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology 279 (2003) 275-289, doi: 10.1016/S0022-1694(03)00225-7.

[16] G. Lindström, A Simple Automatic Calibration Routine for the HBV Model, Hydrology Research 28(3) (1997) 153–168, doi: 10.2166/nh.1997.0009.
[17] ESIMO, the Unified State System of Information website regarding the Situation in the World Ocean, 2021. URL: http://esimo.ru.
[18] Prophet: Automatic Forecasting Procedure, 2021. URL: https://github.com/facebook/prophet.
[19] G. Ayzel, Does deep learning advance hourly runoff predictions?, in: Proc. of V Intl. Conf. ITHPC-2019 Khabarovsk, Russia, September 16-19, 2019, pp. 62–70, CEUR-WS.org, online CEUR-WS.org/Vol-2426/paper9.pdf.
[20] Emercit, The system of flood monitoring in Krasnodarskiy Kray, Russia, 2021. URL: http://emercit.ru/map/.