

Automated short-term forecast system based on open-source hydrological models for the Tikhvinka river (Leningrad region of Russia)

Eduard Kazakov^{1*}, *Kirill Shemanaev*², *Sergey Zhuravlev*¹, *Mikhail Sarafanov*^{1,3}, *Yulia Borisova*^{1,4}, and *Lyubov Kurochkina*¹

¹State Hydrological Institute, 23 2-nd line, Vasilievsky Island, 199004, St. Petersburg, Russia

²LLC "LGMI", 22a Zvenigorodskaya str, 191119, St. Petersburg, Russia

³ITMO University, 49 Kronversky prospekt, 197101, St. Petersburg, Russia

⁴Saint Petersburg State University, 7/9 Universitetskaya nab, 199034, St. Petersburg, Russia

Abstract. In recent decades there has been a trend towards an increase in the number of dangerous hydrological events, especially floods. In order to protect citizens and solve economic problems, it is important to develop and actively introduce into operational practice methods of hydrological forecasting, as well as to build more modern and convenient interfaces of interaction between hydrometeorological services, municipal authorities and citizens. This work discusses a compact automated short-term hydrological forecasting system that uses open-source conceptual models HBV, SimHYD and GR4J as its core. The system is connected to data streams on the observed temperatures and precipitation in the watershed basin, as well as the predicted values of these parameters (in a current implementation, the WRF model with a forecast for 84 hours is used). Also, for operational calibration in daily mode, the system can assimilate (if available) data on observed water levels. Testing of the system is carried out on the example of Tikhvin city (the Tikhvinka river), which in recent years has been characterized by frequent flooding.

1 Introduction

Dangerous hydrological events associated with short-term increases in water levels remain one of the most common destructive natural phenomena that annually cause human casualties and billions of dollars in damage [1]. Potential flood zones have been one of the significant factors in assessing real estate insurance rates. These zones are included in the master plans of rural and urban settlements. However, a large number of people still live in flooded areas. To reduce economic costs and human casualties, methods of short-term hydrological forecasting are used. It allows warning in advance of a few days about the high probability of water level rise above a critical level. In hydrological practice the models of different nature are used, in particular simple statistical (empirical), parametric conceptual, as well as more complex physically-based and spatially distributed [2]. Models

* Corresponding author: ee.kazakov@gmail.com

based on machine learning methods are also gaining popularity. Conceptual models are a compromise option – they do not impose too large requirements on the quantity and quality of the initial information (unlike most spatially distributed), but are more efficient and predictable than statistical ones, and also lend themselves to a more natural interpretation than models based on machine learning methods.

The objective of the presented work is to create an automated hydrological short-term forecasting system based on conceptual models with open source code, with the requirements of maximum compactness and ease of obtaining source data. Such systems could subsequently be quickly created for various hydrological gauges with acceptable efficiency but without the need for scrupulous and qualified collection of high-quality basic information about landscapes, soils, geological structure and so on. Two gauges were selected as the objects for testing: the Gorelukha and Tikhvin gauges (both at the Tikhvinka river, Leningrad region of Russia). For Tikhvin town, the presence of such a forecasting system is relevant: with a frequency of approximately every 3 years in the spring, floods occur in the city. And the lowest parts of Tikhvin are flooded almost every year.

2 Data

Most of the conceptual models for hydrological forecasting produce surface runoff rather than the level. Therefore, for their adjustment and operation at the target gauge, historical data on water consumption should be available, as well as the possibility of obtaining operative data on costs for calibration. In the presented work, the following information is used as initial one:

1. Historical data on water discharge at the gauge Tikhvinka – Gorelukha for the period from 1965 to 2019 (for the model calibrations);
2. Historical data on water levels at the gauge Tikhvinka – Gorelukha and Tikhvinka – Tikhvin for the period from 2012 to 2019 (for recalculating water levels at one gauge to another, for estimation of discharge from water level data and backwards at the hydrological gauge Tikhvinka – Gorelukha);
3. Historical data on air temperature and precipitation for the period from 1965 to 2019 (for the models calibration);
4. Operative information about water levels at the gauge Tikhvinka – Gorelukha (for calibration in real time);
5. Operative information on meteorological parameters at Tikhvin meteorological station (for calibration in real time);
6. Daily forecast of meteorological parameters according to the WRF model [3] with a lead time of 84 hours (directly for hydrological forecast).

Historical data on water levels and discharges, operative data on water levels, historical data and daily forecast of meteorological parameters according to the WRF model were provided by the Federal State budgetary institution «Northern Administration for Hydrometeorology and Environmental Monitoring» (NAHEM).

3 Hydrological models

Three conceptual models with open source code were selected for inclusion in the system: HBV [4], SimHYD [5] and GR4J [6]. All of them are rainfall-runoff models, linking precipitation, air temperature and surface runoff. These models are widely known and are often used in different regions of the planet. An automated approach similar to that proposed in [7] was used to select the optimal parameters for each model: at first, 100000 iterations of random selection of parameters were performed with the quality estimation

according to the Nash-Sutcliffe criterion, based on historical meteorological and hydrological data for the period from 1965 to 2019. Then the best combination was selected, and a new iterative selection was performed for it – each parameter was increased or decreased slightly, the gradient of the Nash-Sutcliffe test was estimated, the change was fixed or cancelled. After the best local value of the criterion was reached for the parameter, the transition to the next parameter was carried out. 50 cycles of such calibration were performed for all parameters. The quality of the models according to the Nash-Sutcliffe criterion was estimated at 0.75-0.80, which indicates a sufficient model efficiency. Final values of calibrated parameters for HBV, GR4J and SimHYD models are shown in Tables 1-3.

Table 1. HBV parameters calibrated for the Tikhvinka River.

BETA	CET	FC	K0	K1	K2	LP	MAXBAS
1.76	0.11	60	0.396	0.134	0.025	1.137	2
PERC	UZL	PCORR	TT	CFMAX	SFCF	CFR	CWH
01.07	68	0.848	1.49	1.58	0.552	0.008	0.171

Identifiers for HBV parameters: *BETA* – parameter that determines the relative contribution to runoff from rain or snowmelt; *CET*– evaporation correction factor; *FC*– maximum soil moisture storage; *K0* – recession coefficient for surface soil box (upper part of SUZ – storage in upper zone); *K1* – recession coefficient for upper groundwater box (main part of SUZ); *K2* – recession coefficient for lower groundwater box (whole SLZ – storage in lower zone); *LP* – threshold for reduction of evaporation (SM/FC); *MAXBAS* – routing parameter, order of Butterworth filter; *PERC* – percolation from soil to upper groundwater box; *UZL* – threshold parameter for groundwater boxes runoff, mm; *PCORR* – precipitation (input sum) correction factor; *TT* – temperature which defines the separation of rain and snow fraction of precipitation; *CFMAX* – snow melting rate, mm/day per Celsius degree; *SFCF* – Snowfall Correction Factor; *CFR* – refreezing coefficient; *CWH* – fraction (portion) of meltwater and rainfall which retain in snowpack (water holding capacity).

Table 2. GR4J parameters calibrated for the Tikhvinka River.

X1	X2	X3	X4	X5	X6
582	-0.29	48.5	2.6	0.93	2.6

Identifiers for GR4J parameters are the following: *X1* – production store capacity, mm; *X2* – intercatchment exchange coefficient, mm/day; *X3* – routing store capacity, mm; *X4* – time constant of unit hydrograph, day; *X5* – dimensionless weighting coefficient of the snow pack thermal state; *X6* – day-degree rate of melting, mm/(day*°C).

Table 3. SimHYD parameters calibrated for the Tikhvinka River.

INSC	COEFF	SQ	SMSC	SUB	CRAK	K	etmul	DELAY	X_m	X5	X6
5	310	3.5	570	0.89	0.99	0.09	1.1	3.7	0.02	0.93	2.5

Identifiers for SimHYD parameters are the following: *INSC* – interception storage capacity, mm; *COEFF* – maximum infiltration loss; *SQ* – infiltration loss exponent; *SMSC*

– soil moisture storage capacity; *SUB*– constant of proportionality in interflow equation; *CRAK* – constant of proportionality in groundwater recharge equation; *K* – baseflow linear recession parameter; *etmul* – added parameter to convert max T to potential evapotranspiration; *DELAY* – runoff delay; *X_m* – transformation parameter; *X5* – dimensionless weighting coefficient of the snow pack thermal state; *X6*– day-degree rate of melting, mm/(day*°C).

4 Automated system

To ensure automatic daily production of hydrological forecasts with a lead time of 3 days and their delivery to consumers, an information system based on the principles of client-server architecture was created. The application has been deployed on a server that is accessible via the Internet. It collects and organizes the data from external sources, performs calculations using hydrological models, and provides results via HTTP for third-party servers and for publishing on a web page. The server application uses only open-source technical tools: the GNU / Linux Ubuntu operating system and its scheduled launch tools, compact single-file SQLite database management system, Python programming language with libraries: Flask, SQLAlchemy, SciPy, GDAL, Dash. Models HBV, SimHYD, GR4J are also implemented in Python. The general diagram of system is shown in Fig. 1.

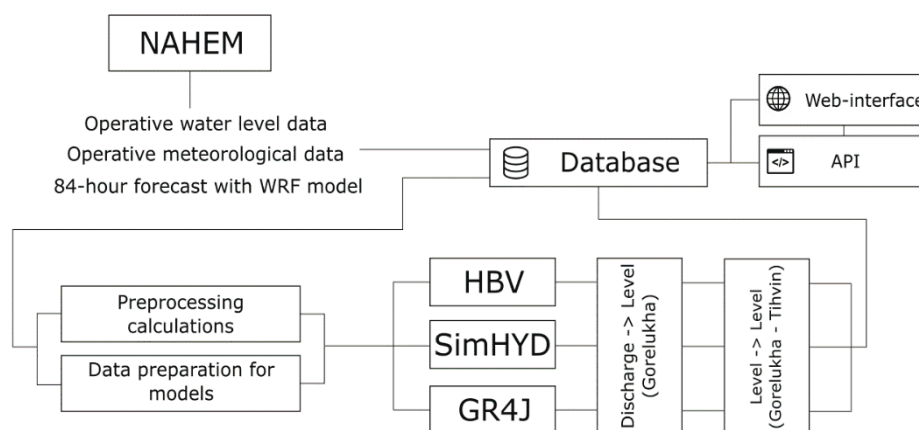


Fig. 1. General diagram of the information system.

To demonstrate the principles of the system's operation, a daily sequence of actions is presented (automatic start on a schedule every day at 22.00):

1. The following data is received from the servers of the NAHEM: the level at the Gorelukha gauge for 8 am (today); temperature and precipitation at Tikhvin meteorological station for today; actual forecast for the WRF model (for 84 hours).
2. The WRF forecast is processed, the forecasts dataset is cut off along the boundaries of the watershed basin of the gauge Tikhvinka – Gorelukha, the temperatures for each day are averaged, precipitation are summarized.
3. The levels on the gauge Tikhvinka – Gorelukha are recounted into the levels on the gauge Tikhvinka – Tikhvin; the levels at the gauge in Gorelukha are recalculated to discharges. Recalculations are based on machine learning methods, in particular, the support vector machines approach is used.
4. All received data is stored in the DBMS.

5. Information about all available meteorological observations (since 1965) is extracted from the database. Data is being prepared in the form required at the input by hydrological models together with three-days WRF forecast records.
6. According to this synthesized meteorological dataset, the discharges at the gauge in Gorelukha are calculated with models HBV, SimHYD, GR4J.
7. Forecasted discharges on the gauge in Gorelukha are recalculated into forecasted levels.
8. Forecast levels on the gauge in Gorelukha are recalculated into forecast levels at the gauge in Tikhvin.
9. Based on the last day for which their level observations are available, an error is considered for each of the models. The error coefficient is calculated, it is applied to all three predictive values of the level.
10. Finally, the database contains all the forecasts and their errors.

The currently functioning web interface offers interactive charts with a history of observed water levels and a forecast (three days in advance) (Fig. 2). Also, access to the forecast data can be carried out using the program interface (API) via the HTTP protocol.

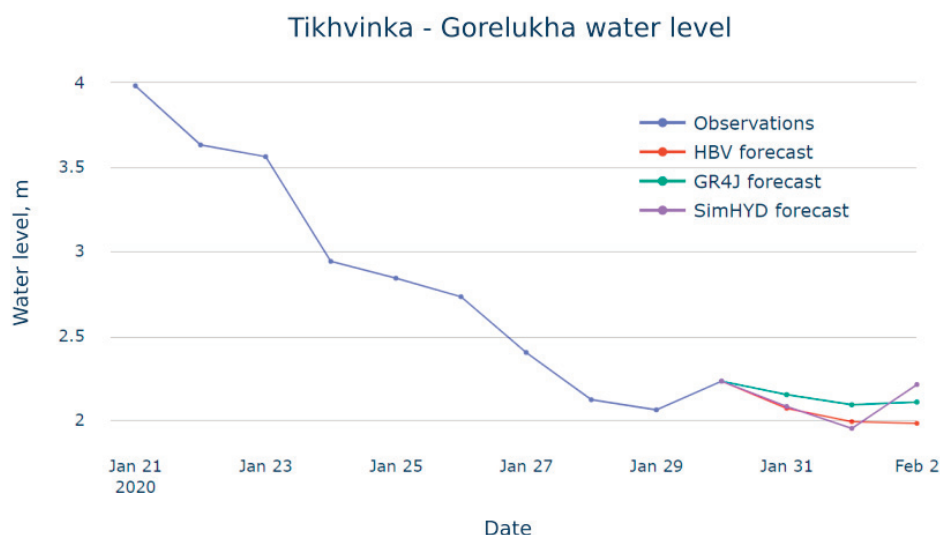


Fig. 2. A fragment of the web interface of the information system. The plot of the observed water levels and forecasts according to three models for the Gorelukha gauge is shown. A similar plot is available for the Tikhvin gauge.

5 Conclusion

The automated information system which is created in prototype mode shows a rather high prognostic efficiency, while being compact, based only on open source software and capable of being quickly calibrated and deployed for new objects. The availability of historical meteorological and hydrological data to select the optimal parameter values and access to operative information on water level observations for calibration during operation is required to run the application. Despite their simplicity, the HVB, GR4J, and SimHYD models demonstrate good quality indicators and can be implemented in different regions as an affordable predictive tool.

References

1. M. N. Istomina, A. G. Kocharyan, I. P. Lebedeva, WARE, **32**(4), 389-398 (2005)
2. G. K. Devia, B. P. Ganasri, G. S. Dwarakish, Aquat. Procedia, **4**, 1001 – 1007 (2015)
3. J. Michalakes, S. Chen, J. Dudhia, L. Hart, J. Klemp, J. Middlecoff, W. Skamarock, *Proceedings of Ninth ECMWF Workshop on the Use of High Performance Computing in Meteorology*, 269-276 (2001)
4. G. Lindström, B. Johansson, M. Persson, M. Gardelin, S. Bergström, *J. Hydrol.*, **201**, 272–288 (1997)
5. F. H. Chiew, M. C. Peel, A.W. Western, Mathematical models of small watershed hydrology and applications (2002)
6. C. Perrin, C. Michel, V. Andreassian, *J. Hydrol.*, **279**, 275–289 (2003)
7. G. Lindström, *Hydrol Res*, **28**, 153–168 (1997)