Hydrologic forecasts verification and comparison of forecasting methods

Sergei Borsch[1], Andrei Khristoforov[1], Vladimir Krovotynzev[2], Ekaterina Leontieva[1], Yuri Simonov[1], and Victoria Deriugina[2]

[1] Hydrometcentre of Russia; orgp@mecom.ru
[2] State Research Centre "Planeta"; asmus@planet.iitp.ru
* Correspondence: yuri.simonov@mail.ru; z-victoria@yandex.ru; Tel.: +7-916-828-64-04

Abstract: This paper presents the methods of estimating the mean square error of hydrological forecasts, allowing for assessment of their practical applicability. Depending upon the amount and composition of available hydrometeorological data, an appropriate method for forecast error estimation is chosen. A system of statistical tests for comparison of different forecasting methods for the same hydrologic characteristic with the same lead time is presented. These tests allow for choosing an optimal and most accurate forecasting method. Hydrological forecasting method efficiency estimation is based on comparing the forecast error with climatology or inertial (persistence) forecast error using presented tests.

Keywords: hydrologic forecast verification, mean squared forecast error, methods of forecast error estimation, comparison of hydrologic forecasting methods, forecast applicability assessment.

1. Introduction

The purpose of hydrologic forecasting is to predict the expected future water regime characteristics for rivers, channels, lakes, reservoirs, and other water bodies. Hydrologic forecasts are used for water resources management and water-related hazard management planning and operational activities. The quality of hydrologic forecasts is defined in turn by their accuracy and lead time. Therefore, the development of modern objective methods for evaluating the quality of operational river flow forecasting methods is of great scientific and practical importance [11, 12].

To estimate deterministic hydrologic forecast error and to assess the forecast applicability, the system of statistical methods is recommended. The use of those methods allows for obtaining objective quality assessment of hydrologic forecasting algorithms. Using these methods allows for selecting optimal scheme and developing scheme enhancement strategies taking into account the specific features of forecasting scheme, as well as the amount of hydrologic and meteorological data, their contents and quality investigated during scheme development, testing and operational use.

Forecast verification implies statistical analysis of the relation between actual values of hydrologic regime elements and their forecast values using this scheme. Verification procedures are the final necessary step to complete the process of forecast scheme development and implementation. Hydrological forecast error determination and analysis are an essential step in the process of scheme development and operation. First of all, for a given forecast lead time, forecast error is the key index of practical value of that forecast. Moreover, forecast error analysis allows for determining scheme weaknesses and outlining strategies for scheme enhancement. Analysis results for a set of forecast of different elements of hydrologic regime within a whole region may determine and prove the strategies for improving the observation network and data acquisition and processing system.

In case when different methods may be used for forecasting the same hydrologic characteristic with the same lead time, there is a need to choose the most accurate method. If an advanced modification

of some previously used scheme is presented, its advantage must be substantiated. In all such cases, statistically significant difference between the estimated errors of hydrologic forecast methods must be determined.

Hydrologic regime elements forecasting method are based on hydrological and meteorological information about variability factors of these elements available on the date of forecast issue. An alternative forecast is based on long-term data statistical analysis of predicted element only. Climatology forecasting method ("climatology") using mean long-term value of predicted element could be considered such unconditional alternative for long-range hydrological forecasts as well as for some medium-range forecasts. Inertial, or persistence forecast ("persistence") using the known value of forecast element for the date of forecast issue could be considered an unconditional alternative for short-range and some medium-range hydrological forecasts. The practical use of a hydrological forecast method is reasonable if its accuracy is higher than that of alternative forecast [2, 6,12].

2. General principles of forecast error estimation

Hydrological forecast general error V is defined as mean squared difference between its actual value Y and value $\widetilde{Y}$ predicted using considered scheme:

$$V = M[(Y - \widetilde{Y})^2].\qquad(1)$$

Before presenting the measure V estimation methods, considering forecast error estimation using dependent sample would be useful. Consider a hydrological forecasting method based on n-year joint hydro- and meteorological observation data. If average number of annually issued forecasts is l, the number N of hydrological observations of predicted value taken into account when developing that method is nl, as well as that of hydrometeorological observations of its predictors. Those hydrometeorological observations are used for generating the series of N forecasts. As a result, a test forecast error series $Y_1 - \widetilde{Y}_1,...,Y_N - \widetilde{Y}_N$ is generated. Forecast error is characterized by the following index:

$$S^2 = \frac{1}{(N-k)}\sum_{i=1}^{N}(Y_i - \widetilde{Y}_i^*)^2,\qquad(2)$$

where k is the number of forecasting model or formula parameters estimated using the same data of N hydrometeorological observations. In case of using graphs representing the relationship between predictors and predictand, k is defined using the following rule: $k = 2$ for a linear relationship between one predictor and predictand; $k = 3$ for a parabolic, exponential or logarithmic relationship, etc. [1, 7, 10, 16].

Very important aspect of considered problem is that the test forecast error series used for producing the estimate S^2 was generated based on the data used for developing the evaluated forecasting method. Choosing predictors, developing the forecasting formula, estimating its parameters or plotting relationship graphs was performed in such a way that the discrepancy between observed and predicted values Y_i and $\widetilde{Y}_i$ of hydrologic characteristic for $i = 1, ..., N$ was minimal.

Due to the fact that dependent sample–based estimate S^2 is characteristic of only residual variance $\widetilde{\sigma}^2$ of the relationship between predictors and predictand Y, the true value of forecast error is therefore considerably underestimated.

Forecast error is defined by forecasting method robustness against the used observation data as well. Ceteris paribus, the larger the number of predictors and parameters (or the more flexible the set of relationship graphs), the lower the robustness of that predictors–predictand relationship and the larger the difference between residual variance $\widetilde{\sigma}^2$ and forecast error V [6, 8, 9, 15, 19].

To recognize the discrepancy between $\widetilde{\sigma}^2$ and V, consider a case when stochastic relationship between characteristic Y and its predictors is described by a linear regression model for each date of forecast issue over the observation period and over projected period of operational use. Model parameters are estimated using the least squares method basing on N joint hydro- and meteorological

2

observation series. The relationship between residual variance $\widetilde{\sigma}^2$ and forecast error V is described by the following equation:

$$V = \widetilde{\sigma}^2 \, (1 + \frac{k - 2/N}{N - k - 1}) \approx \widetilde{\sigma}^2 \, (\frac{N-1}{N-k-1}) . \tag{3}$$

The formula (3) shows that forecast error is rapidly increasing with increasing number k of parameters, as residual variance $\widetilde{\sigma}^2$ is not that rapidly decreasing due to more comprehensive and adequate description of predicted event [2, 9]. The above results demonstrate a well-known fact that in case of verification using independent sample, simple but robust forecasting methods quite often can be more accurate than those using comprehensive and adequate models with many parameters to be estimated [2, 6, 9, 12, 18, 20].

3. Forecast error estimation methods

Depending on the volume and composition of available observation data used for developing and testing a forecast, one of the following methods is recommended for estimating deterministic hydrological forecast error [2]. The absence of autocorrelation in series of forecast errors is supposed.

Method 1.

If $\widetilde{N}$ forecasts of a hydrological variable Y were issued using developed and implemented hydrological forecasting method, a series of forecast errors $Y_{N+1} - \widetilde{Y}_{N+1}$,..., $Y_{N+\widetilde{N}} - \widetilde{Y}_{N+\widetilde{N}}$ is generated, representing the results of testing using independent sample. Forecast error estimate computed from that series is defined by the formula:

$$V_I^* = \frac{1}{\widetilde{N}} \sum_{j=1}^{\widetilde{N}} (Y_{N+j} - \widetilde{Y}_{N+j})^2 . \tag{4}$$

The estimate V_I^* is unbiased, i.e., is free of systematic error. For quite large number $\widetilde{N}$ of independent test forecasts, V_I^* is close to actual forecast error V [1, 8, 10, 11, 16].

For proper use of estimates V_I^* or $\sqrt{V_I^*}$, their statistical error of determination must be computed. For V_I^* and $\sqrt{V_I^*}$, their root mean squared errors (square root of variance) are approximately calculated using the formulas:

$$\sigma(V_I^*) = V_I^* \sqrt{\frac{2}{\widetilde{N}}} , \tag{5}$$

$$\sigma(\sqrt{V_I^*}) = \sqrt{V_I^*} \sqrt{\frac{1}{2\widetilde{N}}} . \tag{6}$$

As more observation data is available, a hydrological forecast may be corrected (updated) by re-estimating model parameters or refining the relationship graphs, which results in changing that forecasting method, in fact, in developing a new one which has to be tested separately.

Method 2.

In case of quite long n-year hydro- and meteorological observation data series, a "leave-p-out" cross-validation method can be used [2, 4, 8, 11, 13, 16, 19]: 'truncated' n_0-year random sample of original n-year data is used as training set for developing a forecasting method and the remaining $\hat{n} = n - n_0$ observations are used as an independent validation set. On average, $\hat{l}$ forecasts are issued during each of $\hat{n}$ years; forecast error series length $\hat{N}$ is therefore equal to $\hat{n}\hat{l}$. That test data subset should be used for estimating the forecast error using a formula similar to (2.9).

Test dataset–based forecasting scheme remains the same, though its parameters or graphs may change. Thus, it is reasonable to change the indications and to denote such forecast by $\widetilde{Y}_{0,j}$ for $j = 1,..., \hat{N}$.

Using less observation data leads to lower accuracy of model parameters estimation. As a result,

3

'truncated' forecast error is somewhat larger than actual forecast error V derived from full data set of N observations. A correction coefficient resulting from formula (2.5) may be used to eliminate that fault [2]. If k parameters of forecasting method are to be estimated, the forecast error estimate computed from all N observations should be defined by the following formula:

$$V_{II}^* = \frac{1}{\hat{N}} \sum_{j=1}^{\hat{N}} (Y_j - \widetilde{Y}_{0,j})^2 \left(\frac{N_0 - k - 1}{N_0 - 1}\right)\left(\frac{N - 1}{N - k - 1}\right). \tag{7}$$

Accuracy of that method may be considerably improved using the following approach: the above procedure should be repeated m times in such a way that the resulting test set consists of non-duplicate data for each of all available n years. That number m of iterations should satisfy the condition that the sum of all test sample lengths $\hat{N}_1 + \ldots + \hat{N}_m$ is equal to total number N of observations. To ensure satisfying that condition, the numbers $\hat{n}_1, \ldots, \hat{n}_m$ may differ at each iteration.

Resulting forecast error estimate is computed from the series of forecast error estimates $V_{II,1}^*$, ..., $V_{II,m}^*$ for m iterations:

$$\overline{V}_{II}^* = \frac{1}{m} \sum_{i=1}^{m} V_{II,i}^* . \tag{8}$$

Root mean squared errors of estimates V_{II}^* and $\sqrt{V_{II}^*}$ are approximately calculated using the formulas (5) and (6).

Method 3.

Consider a less favorable situation when original hydro- and meteorological observation data sample is quite short, not allowing for obtaining a validation set long enough, whereas no operational verification has been performed yet or there is still not enough data available for error estimation using the above method 1. In this case, a "leave-one-out" method, based on J.W. Tukey's "jack-knife" method, is optimal [2, 9]. In case of small n, this is an optimal modification of method 2 [2, 5, 9, 13, 17, 18].

Leaving out the i-th year from n-year observation period, the remaining $n-1$ observations are used for developing an i-th variant of forecasting method modification with new relationship parameters or graphs. Denote the number of forecasts issued during the i-th year by l_i. The series of forecast errors $Y_{i,j} - \widetilde{Y}_{i,j}$ for $j = 1, \ldots, l_i$ can be used as independent validation sample, as the values $Y_{i,j}$, as well as their predictors, were not taken into account when developing the i-th variant of forecasting method.

For $i = 1, \ldots, n$, the described procedure is repeated for each i-th year, each time returning the year left out on previous step to the training sample. Thus, method 3 is a modification of the above cross-validation method 2 in case of $n_0 = n-1$, $\hat{n} = 1$, and $m = n$. As a result, for $j = 1, \ldots, l_i$ and $i = 1, \ldots, n$, a series of test forecast errors $Y_{i,j} - \widetilde{Y}_{i,j}$ is generated, characterizing the forecast error derived from $(n-1)$-year observation data, whereas the verified scheme is based on n-year observation data. However, this may be considered negligible, as the correction coefficient following from formula (2.5) is close to 1. The "leave-one-out"-estimated forecast error is defined by the formula:

$$V_{III}^* = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{l_i} (Y_{i,j} - \widetilde{Y}_{i,j})^2 . \tag{9}$$

In respect of theoretical results presented in [9], approximate root mean squared errors of estimates V_{III}^* and $\sqrt{V_{III}^*}$ are defined by formulas (5) and (6).

Method 4.

The presented verification method is the easiest, allowing for validating the forecast using a dependent sample (the same data used for scheme development). This method is based on a hypothesis that stochastic relationship between hydrologic characteristic and its predictors is described by a linear regression model and its k parameters are estimated using least squares method. Using the estimates

S^2 , respectively, forecast error can be estimated as follows:

$$V_{IV}^* = S^2 \left(\frac{N-1}{N-k} \right).$$ (10)

In respect of theoretical results presented in [2, 9], approximate root mean squared errors of estimates V_{IV}^* and $\sqrt{V_{IV}^*}$ are defined by formulas (5) and (6) too.

The following guidelines are recommended for deterministic hydrological forecast verification:

1. If there are enough operational forecasts issued to form an independent validation data set, it is reasonable to use method 1. Otherwise, one of the remaining three methods should be used.

2. If there is enough joint hydro- and meteorological data available for developing and validating a forecasting scheme, it is reasonable to use method 2.

3. Otherwise, it is reasonable to use method 3.

4. For any data volume available, it is reasonable to use method 4, if the forecasting equation is linear in its parameters.

5. We recommend that several verification methods be used if possible and analysis of results be performed.

4. Hydrological forecast verification example

Consider using the above methods for short-range daily hydrologic forecast verification (case study of the Sochi River at Sochi, forecast lead time is 1 day). Conceptual snowmelt- and rainfall–runoff model was developed for forecasting. To feed that model, daily hydrometric and weather data observed at Sochi streamgauge and Sochi weather station are used, as well as weighted mean temperature and precipitation 1-day ahead forecasts derived from four available operational meteorological model outputs (COSMO-Ru 7, NCEP, REGION, and UKMO). For each month, using least squares method, 13 parameters of hydrological forecasting model were estimated from 18-year hydrometeorological observation data series over the period from 1984 to 2005 (accounting for observation gaps). Forecasts are issued once a day on daily basis; annual number of issued forecasts is therefore equal to 365-366. In respect of the number of days in a month, total monthly number N of daily observations varies from 508 to 558. Mean long-term total annual number of daily observations is 6506 [3].

As the above forecasting model has been operationally implemented fairly recently, there is still not enough data available for using verification method 1 (see section 2.2). Therefore, methods 2, 3 and 4 were used for estimating forecast error.

Example of using method 2.

Randomly choosing $\hat{n} = 5$ years 1984, 1990, 1998, 2000, and 2004 as 'independent data', initial 18-year observation data were cut to a 5-year validation set and 13-year training set. For each month, model parameters were estimated using that training sample and 5-year forecast error series was derived from test sample.

Figure 1 shows daily actual vs. predicted flow (discharge) plots, which coincide fairly closely.
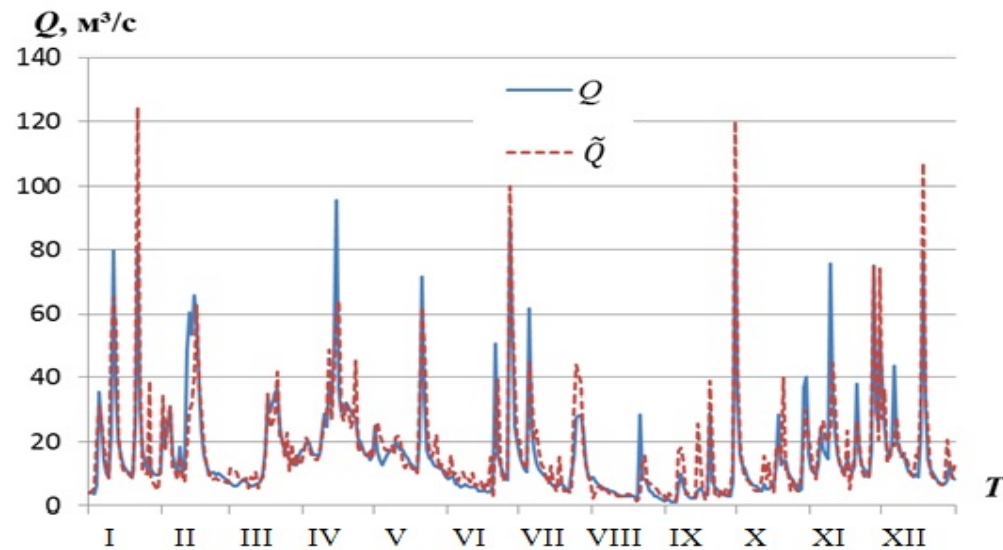
Figure 1. Daily discharge of the Sochi River at Sochi in the year 1984: actual (blue solid line) vs. forecast (red dashed line)

For each month, forecast error was estimated using equation (7), where $\hat{N}$ = 141 - 155. For each month and the whole year, the estimates $\sqrt{V_{II}^*}$ m^3/s are given in table 1. According to (6), relative root mean squared error of estimating $\sqrt{V_{II}^*}$ is approximately equal to 6.4% for each month.

Example of using method 3

For i =1,…, n (n = 18), consequently leaving out every i -th year, forecasting model parameters were estimated from remaining n –1 = 17 years. The data of that i -th year were used for generating an independent set of forecast errors. For each month and the whole year, forecast error was estimated using equation (9). Estimates $\sqrt{V_{III}^*}$ m^3/s are presented in table 1.

Example of using method 4.

As the conditions for using method 4 are satisfied, equations (2) and (10) were used for estimating forecast error $\sqrt{V_{IV}^*}$ m^3/s for each month and for the whole year is given in table 1..

Table 1

Daily stream flow forecast error estimates, cubic meters per second (case study of the Sochi River at Sochi)

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	year
$\sqrt{V_{II}^*}$	12.3	10.2	10.4	9.6	16.3	11.9	10.2	12.5	9.6	16.4	15.7	17.1	12.7
$\sqrt{V_{III}^*}$	11.4	8.2	11.1	8.8	17.2	12.5	9.7	10.4	9.8	14.5	16.1	15.7	12.1
$\sqrt{V_{IV}^*}$	10.4	8.5	10.7	8.6	14.8	12.1	9.2	11.2	8.8	15.3	14.5	16.2	12.0

As table 1 shows, for each month, identical forecast error estimates are derived using different verification methods. The features of hydrological regime, in particular, intramonthly flow variability, are well represented by monthly error distribution [2, 3].

5. Comparison of hydrologic forecasting methods

6

Consider using two concurrent methods of forecasting the same characteristic Y with the same lead time. The differences in forecast methods may be related with forecast models, meteorological data assimilation schemes, the set of predictors taken into account, the set of parameters, parameter estimation methods, or forecast relation graphs. Denote the forecasts of characteristic Y constructed using methods "1" and "2" by $\widetilde{Y}_1$ and $\widetilde{Y}_2$, respectively, and mean squared forecast errors defined by formula (1), i. e., true error values of compared methodologies, by V_1 and V_2, respectively. The methods presented before must be used for computing the estimates V_1^* and V_2^* based on test forecast series using dependent or independent samples , series lengths are N_1 and N_2, respectively. There may be some number $N_{1,2}$ of test forecasts issued on the same time using compared schemes.

Considering test forecast series common for both schemes "1" and "2", denote the forecasts of value Y_i for i-th forecast time by $\widetilde{Y}_{1,i}$ and $\widetilde{Y}_{2,i}$. There may be strong correlation between forecast errors $Y_i - \widetilde{Y}_{1,i}$ and $Y_i - \widetilde{Y}_{2,i}$ due to the common observed value Y_i and to the same (or strongly correlated) predictors. The factors not taken into account, for instance, weather conditions over hydrological forecast lead time period [2], contribute considerably to forecast error correlation.

Standard estimate of forecast error $Y_i - \widetilde{Y}_{1,i}$ and $Y_i - \widetilde{Y}_{2,i}$ correlation coefficient for $i = 1,\ldots, N_{1,2}$ is defined by formula:

$$r = \frac{\sum_{i=1}^{N_{1,2}} (Y_i - \widetilde{Y}_{1,i})(Y_i - \widetilde{Y}_{2,i})}{\sqrt{\sum_{i=1}^{N_{1,2}} (Y_i - \widetilde{Y}_{1,i})^2}\sqrt{\sum_{i=1}^{N_{1,2}} (Y_i - \widetilde{Y}_{2,i})^2}}. \tag{11}$$

Forecasting experience demonstrates that positive correlation coefficient estimate r is virtually always statistically significant. Error correlation of forecasts using different methods must be taken into account when comparing the estimates V_1^* and V_2^*, as that correlation results in increase in probability of advantage of one, i.e., in power increase of statistical tests used for comparing those methods [2, 9].

Suppose method "1" seems to be more accurate, i.e., $V_1^* < V_2^*$ likely due to the fact that method "1" is actually more accurate than method "2", i.e., $V_1 < V_2$. However, method "1" may only seem to have an advantage due to statistical error of compared estimates. Submitting or rejecting the advantage of one method over another actually comes to testing a statistical hypothesis that $(V_1 = V_2)$ against $(V_1 < V_2)$.

Depending on the properties of each test forecast error series, we recommend two statistical tests be used for assessing statistical significance of method "1" advantage. When using these tests, the following fact must be taken account: if scheme errors were estimated using dependent sample, the number of estimated parameters (k_1 and k_2 for methodologies "1" and "2", respectively) must be taken into account. Otherwise, if forecasts errors were estimated using independent sample, the values k_1 and k_2 must be replaced with zero for further calculations.

Test 1.

This test, given in [9], is a modification of asymptotically most powerful Wald likelihood-ratio test [2]. It can be used if the following conditions are satisfied:

1 Test forecasts of the same forecast time are used for testing both forecasting methodologies, i.e., $N_1 = N_2 = N_{1,2} = N$.

2 For each forecasting method, test forecast error series is a random sample of normal probability distribution.

3 Forecast error estimates V_1^* and V_2^* were computed using methods 1, 2, or 3 described before.

For significance level α, the inequality $V_1^* < V_2^*$ should be considered statistically significant and scheme "1", obviously more accurate, if the following condition is satisfied:

$$B = N \ln[1 + \frac{(V_2^* - V_1^*)^2}{4 V_1^* V_2^* (1 - r^2)}] > \chi_1^2(\alpha),\qquad(12)$$

where $\chi_1^2(\alpha)$ is the quintile of central chi-squared distribution with one degree of freedom corresponding to probability of exceedance α. For $\alpha = 5\%$, $\chi_1^2(\alpha) = 3.84$ [2, 9].

One should pay attention to the fact that the left part of inequality (12) increases with increasing value of coefficient r. This is indicative of the fact that in case of strongly correlated forecast errors of compared methods, even a slight advantage of one of forecasting methodology over another becomes statistically significant.

Test 2.

The presented statistical test is recommended for the use in the most general case of no limitations of test forecast error series properties. In this case, we recommend the multidimensional statistical analysis Mahalonobis distance based test given in [2]. The errors $\sigma(V_1^*)$ and $\sigma(V_2^*)$ of estimates V_1^* and V_2^*, defined by formula (5). The only condition of using that test is sufficiently large number of test forecasts, N_1 and N_2 for methods "1" and "2", respectively.

For significance level α, the inequality $V_1^* < V_2^*$ should be considered statistically significant and method "1", obviously more accurate, if the following condition is satisfied:

$$M = \frac{V_2^* - V_1^*}{\sqrt{\sigma^2(V_1^*) + \sigma^2(V_2^*) - 2\frac{N_{1,2}}{\sqrt{N_1 N_2}} r^2 \sigma(V_1^*)\sigma(V_2^*)}} > t(\alpha),\qquad(13)$$

where $t(\alpha)$ is the normal probability distribution quintile corresponding to probability of exceedance α. In particular, for $\alpha = 5\%$, $t(\alpha) = 1.64$.

6. Hydrologic forecasting schemes comparison example

Compare two methods of short-range stream flow forecasting for the Sochi River at Sochi (lead time is 1 day). Both were developed using the same hydrometeorological observation data for the period from 1984 to 2005 accounting for observation gaps (series length $n = 18$ years). Total monthly number of daily observations varies from 508 to 558. For the whole year, mean long-term total number of daily observations is equal to 6506.

As for forecasting algorithm "1" presented in section 2.4, its forecasting formula has 13 parameters estimated for each month using least squares method.

As for forecasting algorithm "2" based on the same conceptual snowmelt- and rainfall-runoff model and using the same hydrometeorological data, its forecasting formula is simplified and has only $k_2 = 7$ parameters estimated for each month.

Concerning specific features of the forecasting algorithms, with V_1^*, V_2^* and r estimates taken into account, the above statistical tests are used for comparing both forecasting methods.

Using test 1.

Correlation coefficient r of forecast "1" and forecast "2" errors for the same days was estimated using formula (11) for each month and for the whole year (see table 2). The values of test measure B defined by formula (12) are also given in table 2. For significance level $\alpha = 5\%$, critical value of B is $\chi_1^2(5\%) = 3.84$. According to the data given in table 2, for the whole year and for all months except for June, July, and September, the inequality (12) is true and there is statistically significant accuracy advantage of scheme "1" over scheme "2".

Using test 2.

Root mean squared errors $\sigma(V_1^*)$ and $\sigma(V_2^*)$ of estimates V_1^* and V_2^* respectively, defined by formula (5), are presented in table 2. The values of test measure M, defined by formula (13) for $N_1 = N_2 = N_{1,2} = N$, are given in table 2. For significance level $\alpha = 5\%$, critical value of M is $t(5\%) =$ 1.64. According to the data given in table 3, for the whole year and for all months except for June, July, and September, the inequality (13) is true and there is statistically significant accuracy advantage of forecasting algorithm "1" over method "2". For the Sochi River, flood frequency is quite low in the period from July to September following the period of relatively low spring flood. Therefore, the impact of using simplified forecasting formula "2" is minimal [3].

Table 2

Stream flow forecasting schemes comparison results using test 1 and 2 for Sochi River flow forecasting

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	year
N	554	508	526	540	558	540	558	526	540	558	540	558	6506
V_1^*	108	72	114	74	219	146	85	125	77	234	210	262	144
$\sigma(V_1^*)$	6.6	4.6	6.2	5.1	18.9	17.3	11.1	16.7	6.5	17.2	14.1	19.1	3.8
V_2^*	161	88	164	83	279	149	86	135	81	266	253	331	172
$\sigma(V_2^*)$	15.6	6.7	10.6	6.4	30.6	18.0	11.7	16.8	7.3	20.1	18.6	24.9	5.0
r	0.81	0.87	0.84	0.92	0.85	0.96	0.98	0.97	0.93	0.90	0.90	0.89	0.90
B	61.5	20.7	7.9	3.3	3.9	0.8	0.4	2.1	1.4	5.9	4.6	5.7	12.7
M	4.9	5.1	7.9	3.3	3.9	0.8	0.4	2.1	1.4	5.9	4.6	5.7	12.7

Thus, the results of using the above both tests allow to make the same conclusion that for the whole year and for almost all months, there is statistically significant accuracy advantage of forecasting algorithm "1" over "2" [2, 3].

7. Forecast applicability assessment

Climatology forecasting method ("climatology"), as well as inertial (persistence) forecast, may be used as unconditional alternative of some hydrological forecasting method. If the latter is obviously more accurate than unconditional alternative forecast, the studied methodology has proved to be used in hydrological forecasting practice [2].

Climatology method is used as unconditional alternative of long-range and some medium-range hydrological forecasting techniques, as well as of occurrence dates forecasting of events characterizing the hydrological regime of a water body. The forecast of some hydrologic characteristic Y produced using the climatology method is characterized by its long-term value $\overline{Y}$ averaged over the n-year period of observations ($Y_1,..., Y_n$). Such forecast error is usually characterized by Y variance estimate defined by the following formula:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \ . \tag{14}$$

The "climatology" forecast error is compared to the error of forecast produced using the evaluated method. That error is usually characterized by an index S^2 defined by formula (2). That measure is an approximate estimate of residual variance related to the variance of forecasted value and correlation ratio R [18]. Squared R estimate is defined by the following formula:

$$R^2 = 1 - \frac{S^2}{\sigma^2} \ . \tag{15}$$

9

The index R^2 is a correction of Nash-Sutcliffe efficiency index widely used in a number of countries for assessing the hydrological forecast efficiency [2, 14, 19].

Inertial forecast is the unconditional alternative of short-range and some medium-range hydrological forecasting methods. For Δt-day forecast lead time, inertial forecast $\widetilde{Y}_I(t)$ of some flow characteristic for day t is based on that characteristic's value $Y(t - \Delta t)$ known on the date of forecast issue and is defined by the formula:

$$\widetilde{Y}_I(t) = Y(t - \Delta t) + \overline{\Delta}, \tag{16}$$

where $\overline{\Delta}$, providing the absence of systematic error, is mean variation of forecasted hydrologic characteristic over the lead time period. That value is the arithmetic mean of the series Δ_i for $i = 1,...,$ N, where N is the total number of forecasted runoff characteristic variation values observed over a hydrological forecast lead time period. Inertial forecast error estimate σ_Δ is defined by the formula:

$$\sigma_\Delta = \sqrt{\frac{1}{N-1} \sum_{i=1}^n (\Delta_i - \overline{\Delta})^2}. \tag{17}$$

In some cases of medium-range hydrological forecasting, a problem of choosing an unconditional forecasting alternative may arise. The choice of "climatology" or inertial forecast is driven by the σ and σ_Δ ratio. For Δt-day forecast lead time, that ratio is determined by the angle of incline (a) of the line $Y(t) = f(Y(t - \Delta t))$, where $Y(t)$ is the forecasted hydrological variable and $Y(t - \Delta t)$ is its actual value observed Δt days before day t. The criterion of choice is as follows:

if $a \geq 1/2$, then $\sigma > \sigma_\Delta$ and inertial forecast is therefore recommended as unconditional alternative of evaluated forecasting scheme;

if $a < 1/2$, $\sigma < \sigma_\Delta$ and "climatology" is therefore recommended as unconditional alternative of evaluated forecasting scheme [2, 7].

Applicability assessment of a hydrological forecast is based on comparing its error with that of unconditional alternative forecast. One of the methods described before should be used to get the error estimate V^* of forecast depending on the volume and quality of available data.

In case of choosing "climatology" as alternative, we need to compare V^* with σ^2 defined by formula (14). If $V^* < \sigma^2$, we recommend the statistical tests presented before be used for testing the statistical significance of accuracy advantage of evaluated method over "climatology".

In case of choosing inertial forecast as alternative, we need to compare V^* with σ_Δ^2 defined by formula (17). If $V^* < \sigma_\Delta^2$, we recommend the above mentioned tests be used for testing the statistical significance of accuracy advantage of evaluated method over inertial forecast.

8. Forecast applicability assessment example

Applicability evaluation of one-day lead time stream flow forecasting scheme for the Sochi River at Sochi is considered as the second example of forecasting method applicability assessment. Error estimates defined by method 4 are considered in this example. One-day lead time inertial forecast defined by formula (17) for $\Delta t = 1$ is used as unconditional forecast alternative.

For each month and for the whole year, table 3 presents the following data: the ratio of evaluated forecast errors to inertial forecast errors $\sqrt{V^*}/\sigma_\Delta$, and the coefficient r of correlation between same-day evaluated forecast errors and inertial forecast errors. There is considerable difference between the evaluated forecast error probability distribution and normal probability distribution, as well as between the inertial forecast probability distribution and normal probability distribution. In respect of the above, the most generalized test 2 was used for assessing the efficiency of considered forecasting scheme. The values of that test index M, defined by formula (13), are presented in table 3. For each month and for the whole year, those values are large enough for the inequality (13) to be true for any feasible significance level α.

Table 3

Applicability characteristics of short-range daily stream flow forecasting scheme for the Sochi River at Sochi

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	year
S/σ_Δ	0.62	0.60	0.58	0.61	0.65	0.72	0.63	0.66	0.64	0.58	0.57	0.63	0.62
r	0.36	0.50	0.33	0.44	0.29	0.49	0.40	0.45	0.30	0.41	0.41	0.50	0.41
M	10.9	10.0	11.2	9.9	11.2	7.0	7.7	7.0	9.3	9.7	11.1	10.0	32.7

Thus, the considered method of 1-day lead time daily stream flow forecasting for the Sochi River at Sochi may be considered efficient for each month and for the whole year [5, 6].

8. Forecast applicability assessment example

Special web application had been developed for operational visualization and analysis of hydrological forecasting results for the Kuban River basin. The web app enables users to compare streamflow and flood category forecasts with actual hydrological situation in this river basin. The web app provides access to all hydrological data via Internet to remote users. For this web application development, some GIS Amur technologies were used [1, 2, 3].

The web application is a user interface managing and supporting several web services, including web map service, actual hydrological data service, forecast hydrological data service, and satellite web service. Web map service includes a set of topographic and administrative maps (Rosreestr, OpenStreetMap, Esri topo maps, etc.) and digital elevation models (Esri and USGS web services). Actual hydrological web service allows users to display observed river stages at Roshydromet and EMERCOM (Emercit) streamgauges, whereas forecast hydrological web service provides the results of hydrological modeling and forecasting. Satellite web service provides Russian and foreign satellite data of high- and medium spatial resolution provided by the State Research Center "Planeta".

Two Roshydromet organizations, the Hydrometcenter of Russia and the State Research Center "Planeta" perform the technical support of the web application. The Hydrometcenter has a web server hosting the web app (the management component) and GIS servers hosting the hydrological web services, organized into a cluster in order to increase the web service performance and reliability. The database providing information for the web services had been deployed on a separate server.

The State Research Center "Planeta" has a GIS server hosting an automatically updated satellite database and satellite web services.

ArcGIS Enterprise was used for building and managing the web app and web services. Microsoft SQL Server 2014 Enterprise for Windows is used for managing the hydrological and satellite databases.

To make working with hydrological data more convenient, a number of functionalities had been implemented in the "Kuban" web app. In particular, the most recent data is displayed first time the users launch the app. The app allows users to work with both operational and archival data. Red color is used to display locations (streamgauges) where water stage exceeds a dangerous threshold. Moreover, the web app allows users to integrate satellite and ground data, to use tables and graphs for their data visualization. Using the app, for any date of river forecast issue a forecast hydrograph can be plotted for a selected location together with the hydrograph for the past few days, allowing users to analyze streamflow changes over some time for 1 to 5 days ahead. Moreover, plotting back predicted data against an actual hydrograph for a given time period is a useful tool to analyze forecast accuracy over that time

period. Other useful options are: searching, filtering, and downloading data using specific criteria; multiple time animation of actual and forecast data; etc.

Figure 1 presents the forecast of flood situation in the Kuban River basin in terms of flood category (from no flooding to major flooding using green-yellow-red color scheme, respectively) and streamflow changes (flow rise is displayed with red figures; flow fall, with blue figures).
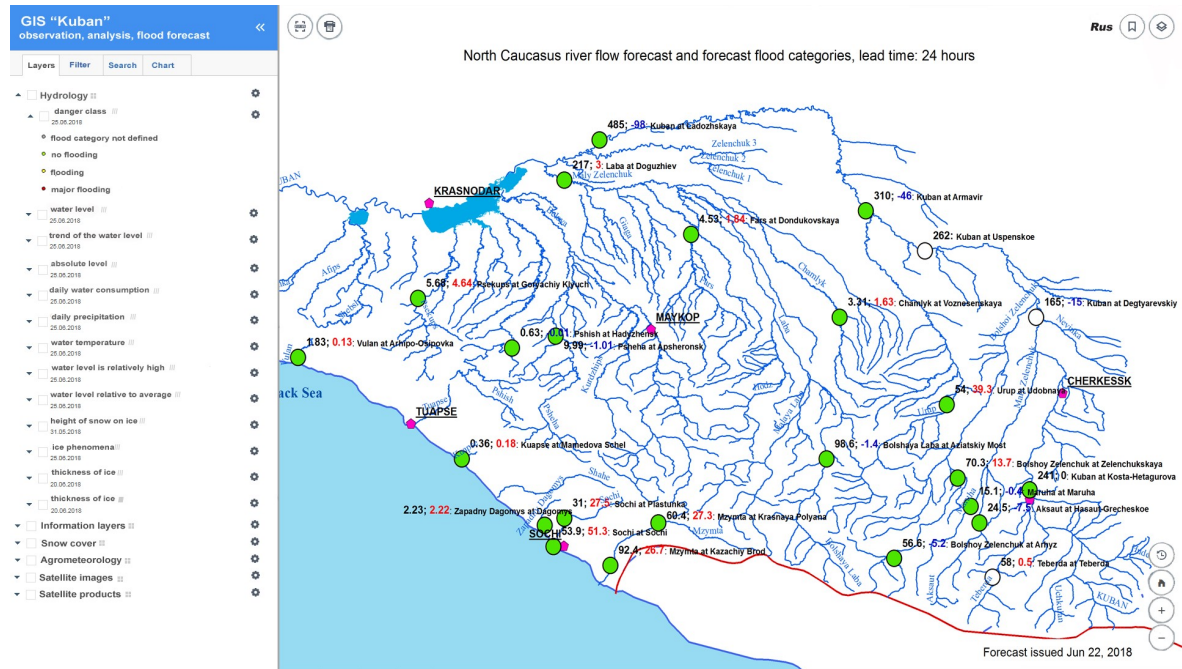

Figure 1. Hydrological forecasts, visualized in the WEB-GIS application

9. Conclusion

Hydrological forecast verification rules taking into account the amount and composition of available hydrometeorological data are presented in this paper, as well as their mean square error estimation methods. To choose an optimal forecasting method, statistical tests for comparison of different methods of forecasting the same hydrologic characteristic with the same lead time are offered. This paper presents hydrological forecasting method efficiency estimation procedures based on comparing the forecast error with climatology or inertial forecast error. Verification of short-range flood forecasts of the Sochi River (Krasnodar region, Russian Federation) is an example of using the above recommended verification rules.

References

[1] American Society of Civil Engineers,1996: Hydrology Handbook, Second Edition, ASCE Manual and Reports on Engineering Practice No. 28, New York, 784.

[2] Borsch S.V., Khristoforov A.V. Hydrologic flow forecast verification. Moscow, Proceedings of Hydrometcentre of Russia, Special Issue 355, 2015, 198.

[3] Borsch S.V., Simonov Y.A., Khristoforov A.V. Flood forecasting and early warning system for rivers of the Black Sea shore of Caucasian region and the Kuban River basin. Moscow, Proceedings of Hydrometcentre of Russia, Special Issue 356, 2015, 247.

[4] Burman P., Chow E., Nolan D. A cross-validatory method for dependent data. Biometrika, 1994, № 81, 351-358.

12

[5] Efron B. The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, 1982, 92.

[6] Guide to Hydrological Practices. Volume II. Management of Water Resources and Application of Hydrological Practices // World Meteorological Organization - No. 168, 2009, 302.

[7] Instruction to forecasting service. Section 3. Part 1. Forecasting of land hydrological regime. Leningrad, Gidrometeoizdat, 1962, 193.

[8] Jolliffe I.T., Stephenson, D.B. Forecast Verification. – Wiley, 2003, 240.

[9] Khristoforov A.V. Reliability of river flow calculations. Moscow, Moscow State University, 1993, 166.

[10] Management Overview of Flood forecasting Systems (MOFFS): Version 3. Geneva: HWR, 1995, 26.

[11] Manual on Flood Forecasting and Warning. // World Meteorological Organization - No. 1072, 2011, 138.

[12] Lambert A.O. Development and Use of the Management Overview of Flood Forecasting Systems (MOFFS) / Technical Reports in Hydrology and Water Resources No. 55. Geneva: HWR, 1994, 23.

[13] Langford J. Quantitatively Tight Sample Complexity Bounds. - Carnegie Mellon Thesis, 2002, 124.

[14] Nash J. E., Sutcliffe J. V. River flow forecasting through conceptual models. Part 1 – A discussion of principles // Journal of Hydrology. Vol. 10, 1970, 282–290.

[15] National Research Council, 1988: Estimating probabilities of extreme floods, National Academy Press, Washington, D.C., 141.

[16] National Weather Service River Forecast Verification Plan. Report of the Hydrologic Verification System Requirements Team. U.S. Department of Commerce. National Oceanic and Atmospheric Administration, 2006, 44.

[17] Tukey J.W. Exploratory Data Analysis. Reading, Mass, Addison-Wesley, 1977, 688.

[18] Wang Juemou. Methods of Verification of Hydrological Forecasts. / Technical Reports in Hydrology and Water Resources No. 44. Geneva: HWR, 1994, 15.

[19] Welles E. Verification of river stage forecasts: Dissertation. University of Arizona, 2005, 157.

[20] World Meteorological Organization. Manual on Flood Forecasting and Warning // WMO, No. 1072, Geneva, 2011, 138.