Long-term ensemble forecast of snowmelt inflow into the Cheboksary reservoir under the two different weather scenarios

Alexander Gelfan[1,2], Vsevolod Moreydo[1], Yury Motovilov[1], Dimitri Solomatine[1,3,4]

[1]Water Problems Institute of Russian Academy of Sciences, Watershed Hydrology Lab., Moscow, Russia
[2]Moscow State University, Geographical Department, Moscow, Russia
[3]IHE Delft Institute for Water Education, Chair of Hydroinformatics, Delft, the Netherlands
[4]Delft University of Technology, Water Resources Section, Delft, the Netherlands

Correspondence to: Alexander Gelfan (hydrowpi@iwp.ru)

Abstract. A long-term forecasting ensemble methodology, applied to water inflows into the Cheboksary reservoir, is presented. The methodology is based on a combination of semi-distributed hydrological model ECOMAG that allows for calculation of an ensemble of inflow hydrographs using two different sets of weather ensembles for the lead-time period: the observed weather data constructed on the basis of the Ensemble Streamflow Prediction methodology (ESP-based forecast), and the synthetic weather data simulated by a multi-site weather generator (WG-based forecast). We have studied: (1) whether there is any advantage of the developed ensemble forecasts in comparison with the currently issued operational forecasts of water inflow into the Cheboksary reservoir, and (2) whether there is any noticeable improvement in a probabilistic forecasts when using the WG-simulated ensemble compared to the ESP-based ensemble? We have found that for 35-year period beginning from the reservoir filling in 1982, both continuous and binary model-based ensemble forecasts (issued in deterministic form) overperform the actually used operational forecasts of the April-June inflow volume and, additionally, provided acceptable forecasts of water regime characteristics other then inflow volume. We have also demonstrated that the model performance measures (in verification period) obtained from the WG-based probabilistic forecasts which are based on a large number of possible weather scenarios, appeared to be more statistically reliable than the corresponding measures calculated from the EPS-based forecasts based on observed weather scenarios.

1 Introduction

Spring freshet is a hydrological phenomenon, which magnitude is highly dependent on the amount of water accumulated in surface and subsurface storages of the river basin during several months prior to the snowmelt. This dependency serves as a physical basis to predictability of spring runoff (Li et al., 2009). As stated by Lettenmaier and Waddle (1978, p.1) “snowmelt runoff is one of the few natural phenomena for which relatively accurate long-term forecast can be made”.

Implementation of this opportunity is crucial for the water reservoirs of the Volga-Kama Reservoir Cascade (VKRC) – one of the world’s largest multi-purpose water management systems. The VKRC is located within the largest European Volga

River basin (area of 1 350 000 km^2) and consists of 11 reservoirs that hold from 1 to 58 km^3 of water and conduct seasonal and multiyear flow regulation. The VKRC was designed to redistribute the highly uneven runoff of the Volga River, with 2/3 of the annual runoff volume occurring during the two to four months of the spring-summer freshet. This task, aimed at optimizing reservoir management for power production, navigation and flood protection, is even more complex due to the requirement of annual spring water release to Lower Volga aimed at allowing for sturgeon spawning. Such several-weeks-long regulated release with predefined amount and temperature of water during the spring freshet is a unique task for a water management system (Avakyan, 1998). Hence, a reliable and forehand forecast of snowmelt inflow into the VKRC reservoirs is crucial for the decision makers.

By the mid-1960s, the specific methods were developed, which underlie the contemporary operational forecast for VKRC management (water supply forecast). For different reservoirs, the produced forecasts are based on two primary techniques: the index methods, and the so-called physical-statistical methods (Gelfan and Motovilov, 2009; Borsch and Simonov, 2016). Both methods produce deterministic (despite the term "physical-statistical"), purely data-driven forecasts and relate the predictors (such as initial snow water equivalent, soil freezing and soil moisture indices, precipitation amount for the forecast period) to the main predictand - the spring inflow into a reservoir. The initial basin characteristics are derived from observations, yet the precipitation amount is typically set to the climatic mean. The operational water supply forecasts methodology is used in real practice by water managers and remains unchanged over the past half-century.

While the utility of data-driven flow forecasts (which currently may be based on advanced statistical and machine learning techniques) has been demonstrated on various examples (see, e.g. Abrahart et al., 2012), their skill and reliability depend on the amount and stationarity of available data and are not always adequate. The forecast improvement could not be expected within the framework of the purely data-driven approach because of reduction of the observational network in the Volga basin (estimated at 30% in (Borsch and Simonov, 2016)), non-homogeneity of the observations caused by changes in the measurement techniques, changes in climate, land use and so on.

An opportunity to improve the operational water supply forecasts of water inflow into VKRC lies in shifting from the traditional exclusively data-driven forecasts towards hydrological model-based forecasts, and from deterministic methodology to the one using ensembles with a possibility to characterise forecast uncertainty. During the last 20-30 years there have been a general understanding of the necessity of such a shift to ensemble streamflow prediction (ESP) systems (e.g., Day, 1985), and a considerable research effort in this direction (Franz et al., 2003; Wood and Lettenmaier, 2006; Li et al., 2009; Shukla and Lettenmaier, 2011; Yossef et al., 2013; Najafi and Moradkhani 2015; Demirel et al. 2015; Beckers et al. 2016; Arnal et al., 2017; Mendoza et al. 2017). Such systems are currently used more and more in operational mode by national weather services in the Untied States (e.g. McEnery et al., 2005), Canada (Druce, 2001) and other countries (Pappenberger et al., 2016).

In its original form, an ESP is based on an assumption that historical time series of observed meteorological variables are representative for a local climate. These series are used as an ensemble of meteorological inputs into a hydrological model to simulate corresponding ensembles of streamflow forecasts. This allows for considering uncertainty in weather conditions during the forecast horizon and provide an opportunity to quantify the corresponding uncertainty (and hence, risk) in the forecast-based decision support systems for reservoir management. In addition, utilizing the process-oriented hydrological models results in increasing the physical adequacy of forecasts and, potentially, in improving forecast accuracy in comparison with the methods currently used in operational practice. However, such quantitative comparisons is not a common place; to our knowledge the only example is the comprehensive experiment presented by Mendoza et al. (2017) and comparing the ESP model-based forecasts with the operational data-driven forecasts for a multi-year historical period.

The observed weather scenarios that are used within the ESP framework do not encompass all of the possible weather conditions for the forecast period. It is desirable to account not only for the observed weather, but for possible weather condition that might lead to freshet events of rare occurrence. Assessing the magnitude of such an event might be crucial for decision making. Moreover, hence the ensemble size is limited to the number of the historical years, statistical problems can appear stemming from large sample errors. For instance, Buizza and Palmer (1998) demonstrate improvement of the weather forecast skill as the ensemble size increases, wherein degree of improvement depends on the verification measure used. Particularly, the ranked probability skill score is strongly dependent on ensemble size and negatively biased (see also Müller et al. 2005, Weigel et al., 2007). Different aspects of the ensemble size effect on statistical properties of the ensemble weather forecast and verification scores are studied by Richardson (2001), Ferro et al. (2008), Najafi et al. (2012). The problem can be solved by incorporating synthetic, stochastically generated time series of weather variables instead of the historical data used within the ESP framework. As a result, hydrological system response to a large variety of possible weather conditions can be reproduced and large ensemble of forecasts can be predicted.

There are not too many attempts to use stochastic weather generator (WG) within framework of long-term ensemble forecasting. Hanes et al. (1977) were probably the first who used Monte-Carlo simulated sequences of daily precipitation to drive the conceptual US Geological Survey hydrological model and provide ensemble seasonal forecast of snowmelt runoff volume. A physically-based distributed hydrological model was used in combination with a weather generator to create a long-term probabilistic forecast of spring runoff of rivers in Central Russia in Kuchment and Gelfan (2007), Gelfan et al. (2015). Caraway et al. (2014) incorporated a stochastic weather generator into the ESP to make a probabilistic seasonal climate forecasts and applied the modified methodology to the San Juan River snowmelt dominated basin. Beckers et al. (2016) used ENSO-conditioned weather generator to compensate for the reduction of ensemble size in the post-processing ensemble forecast scheme presented for the Columbia River basin.

The studies and examples mentioned above serve as the background and the main motivation for this study. The objective of this study is to contribute to the EPS-related studies, with the focus on the comparative analysis between the data-driven techniques used in operational forecasts, and the ensemble forecasts of sreamflow, using two different weather scenarios: a)

based on the historical data, and b) employing the WG-based forecasts. The case study is the Cheboksary reservoir of the VKRC cascade for which the operational forecasts are available since 1982.

Thus, this study is an attempt to answer the following two research questions: (1) Does the model-based ensemble methodology allow one to improve reliability and skill of the operational forecast of spring inflow into the Cheboksary reservoir, and to what extent? (2) Does the enlarged ensemble size lead to any noticeable advantage when using the WG-simulated ensemble compared to the ESP-based ensemble?

The remaining part of this paper is organized as follows. The case study is described in the next section. The operational forecast methodology, as well as the proposed forecasting approach including modelling tools (hydrological model and stochastic weather generator), forecasting schemes, experimental design and forecast verification measures are described in Section 3. Results and discussion are presented in Section 4. The overall conclusions and recommendations are given in Section 5.

2. Case study basin

The Cheboksary reservoir is located on the Volga River in central part of the European Russia. It was constructed in 1982 to become the 11th member of Volga-Kama reservoir cascade with Nizhegorodskoe reservoir upstream and Kujbysevskoe reservoir downstream of it. Total unregulated basin area of the Cheboksary reservoir is 373 800 km^2 (Fig. 1). Its main tributaries - Oka, Sura and Vetluga Rivers - account for 80 to 90% of annual inflow into the reservoir.

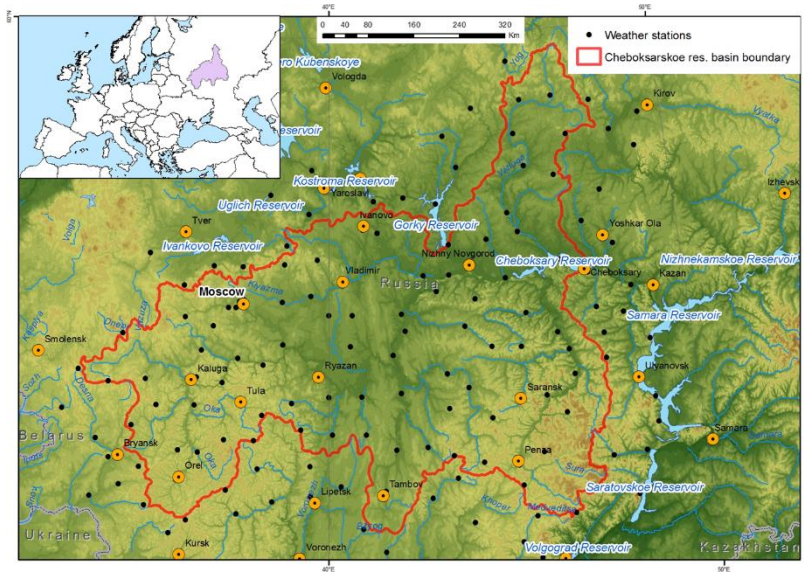

Figure 1: Cheboksary reservoir basin: topography, river network, weather stations

Local climate conditions can be described as moderately continental, with cool snow-abundant winter and relatively hot summer. Mean annual temperature ranges from 1.4°C in the northern part of the basin to 4.8°C in the southern part. During

wintertime air temperature may fall as low as -35 - -40°C. Annual precipitation amount ranges between 650 and 750 mm throughout the territory. Around 60% of the precipitation occurs as rain. Most of winter precipitation is stored as snow cover, emerging in mid-December and lasting until mid-April. Snow water equivalent ranges from 50 mm in the South-Western part up to 100 – 120 mm in the North. Springtime snowmelt contributes to high-flow freshet – the dominating hydrological season accounting for around 65% of the total annual inflow into the reservoir (51.3 km^3). Typically, freshet commences around mid-April and lasts until June. Mean volume of inflow for the period of reservoir operation (1982-2016) is 33.4 km^3, mean maximum inflow discharge is 9355 m^3/s.

3. Method

3.1 Operational data-driven forecast of spring inflow into the Cheboksary reservoir: a current practice

Methodology for forecasting the spring inflow into the Cheboksary reservoir was developed by E.P. Chemerenko (1992) on the basis of so-called physical-statistical approach originally proposed for the reservoirs of Middle Volga in mid-1960s (Zmieva, 1964; Gelfan and Motovilov, 2009). This approach is currently in use by the Russian Hydrometeorological service (Roshydromet) for inflow forecasting into all reservoirs located on the Middle Volga River.

Water inflow volume Y into the Cheboksary reservoir is forecasted according to the equation:

$$Y = \sum_{i=1}^{5} \alpha_i y_i + \beta \quad (1)$$

where y_i is the runoff forecast at streamflow gauge i, located on the reservoir tributaries: i=1 – Oka River, Polovskoe gauge (drainage area F=99000 km^2), i=2 – Klyazma River, Kovrov gauge (F=24900 km^2), i=3 – Vetluga River, Vetluzhsky gauge (F=27400 km^2), i=4 – Sura River, Poretsky gauge (F=50100 km^2), i=5 – Tsna River, Knyazhevo gauge (F=13600 km^2), α_i, β – regression coefficients estimated from the streamflow data observed at the corresponding gauge.

The runoff volume y_i at ith gauge is forecasted by a unified procedure. The predictors are: basin averaged snow water equivalent (S, mm), soil freezing depth, (FD, cm), soil moisture index (W, dimensionless) on a forecast issue date, and total precipitation (x, mm) during the forecast horizon.

Runoff volume at each gauge is calculated as follows:

$$y = (1-f)\left\{(S_1 + x) - P_0\left[1 - \exp\frac{(S_1 + x)}{P_0}\right]\right\} + f\eta(S_2 + x) \quad (2)$$

$$P_o = a\exp\left[-b(\theta - \theta_{\min})^c\right] \quad (3)$$

where S_1, S_2 are the snow water equivalent at the forecast issue date within the deep frozen (FD≥60cm) and non-deep frozen (FD<60 cm) parts of the river basin, respectively, derived from snow observations; x is the total precipitation for the forecast

horizon, assigned as climatic mean; *f* is the fraction of the basin area covered by deep-frozen soil, derived from soil freezing observations; θ is the soil moisture index, calculated from precipitation amount during the preceding autumn period; η is the runoff coefficient from the basin fraction with non-deep frozen soil calculated as a function of θ; *a*, *b*, *c*, $\theta_{\min}$ are the parameters derived from hindcasts for the 30-year period before the reservoir filling in 1982.

The operational forecast of water inflow volume into the Cheboksary reservoir for April-June period is issued just before the beginning of this period (March 27) and then updated 2-3 times during April-May. In this paper, the operational deterministic forecast (not updated, i.e. issued before the beginning of April) is compared with the deterministic forecast derived from the model-based ensemble forecasts described below.

3.2 Model-based ensemble forecast technique and verification measures

3.2.1 Modelling tools

Hydrological model

The ECOMAG (ECOlogical Model for Applied Geophysics) is a semi-distributed process-based hydrological model describing snow accumulation and melt, soil freezing and thawing, water infiltration into unfrozen and frozen soil, evapotranspiration, thermal and water regime of soil, overland, subsurface and channel flow on a daily time step (Motovilov et al., 1999). The model accounts for measurable watershed characteristics such as surface elevation, slope, aspect, land cover and land use, soil and vegetation properties. The parameters are spatially distributed by partitioning the watershed into sub-basins (elementary basins). Parameterization of the sub-grid processes is described by Motovilov (2016). The model is driven by time series of daily air temperature, air humidity and precipitation intensity.

The model was earlier applied for hydrological simulations in many river basins with highly varying sizes and characteristics - from small-to-middle size European basins (Gottschalk et al., 2001) to the large Volga, Lena, Mackenzie with watershed areas exceeding a million km^2 (Motovilov 2016; Gelfan et al., 2017).

In this study, 1x1 km spatial resolution DEM was used for the basin discretization and river network construction. A total of 1045 elementary basins were delineated, with average area of 340 sq. km. The model forcing data for each elementary basin were interpolated from the 157 weather station data (see Fig. 1) employing the inverse-distance method. Most parameters are physically meaningful and were derived through available measurements of the basin characteristics (topography, soil and vegetation properties).

The model was calibrated and validated against the Cheboksary reservoir daily water inflow observations beginning from January, 1, 1982 (the 1st year after the reservoir filling to capacity) to December 31, 2016: calibration covered the period of 2000-2010, the rest of the data were used for the model evaluation. The ECOMAG calibration procedure is described in detail by Gelfan et al. (2015). Here, we emphasize the two issues concerning this procedure. First, the values of several key

parameters pre-assigned from literature or from the available measurements are considered as the initial approximations of the optimal values and the latter are sought within the neighborhood of the initial, pre-assigned values. Second, during the calibration process, the ratios between the initial values of the distributed parameter corresponding to different soils, landscapes and vegetation are preserved. The Nash and Sutcliffe (1970) efficiency criterion NSE is adopted to represent the goodness of fit of the simulated and measured variables.

Multi-site weather generator (MSFR_WG)

The Multi-Site FRagment-based stochastic Weather Generator (MSFR_WG) is the stochastic model using Monte-Carlo simulation to generate time-series of daily weather variables (precipitation, air temperature and air humidity deficit), retaining statistical properties, both spatial and temporal, of the corresponding observed variables. This modeling procedure is based on the so-called "spatial fragments' (SFR) resampling method" initially presented by Gelfan et al (2015). The SFR-method is a modification of the temporal fragments' (TFR) method proposed by Svanidze (1980) for stochastic simulation of highly autocorrelated time-series.

The SFR resampling method includes the following steps:

1. Computation of *N* normalized fields ("spatial fragments", SFR) of weather variables on the basis of the available meteorological data. SFRs are computed for each of *N* years of observations by dividing each daily value of the specific variable by the corresponding spatially averaged annual value.

2. Monte-Carlo simulation of synthetic time-series of *M* spatially averaged annual weather variables reproducing temporal statistical features of the corresponding annual variables derived from observation data. Cross-correlation between annual values of the simulated weather variables is taken into account through the Cholesky's decomposition method (see e.g. Press et al., 2007).

3. Calculation of the synthetic daily fields of weather variables by multiplying the computed SFRs (see step 1) on the Monte-Carlo simulated spatially averaged annual value of the corresponding variables (see step 2). SRFs are randomly chosen from the available set by the Latin Hypercube method (McKay et al., 1979).

The advantage of MSFR_WG is that it has a small number of free parameters in comparison with the widely used milti-site weather generators (see, e.g. Khalili et al. (2011) and references therewith), and does not require complex estimation procedures. Such features typically indicate that the model has high robustness.

3.2.2 Ensemble forecasting technique

The proposed ensemble forecasting procedure utilized in this study was verified by producing hindcasts of water inflow into the Cheboksary reservoir from the 1st of April for 3 months ahead (up to 30th of June). The hindcasts cover a 35-year period

between 1982 and 2016. (Hereafter, we use the term "forecasts" for these hindcasts). For each *i*-th year of the verification period (*i*=1, 2,…, 35), the procedure includes the following steps:

(1) Spin-up ECOMAG-based simulations ("warm start") using meteorological observations data prior to the forecast issue date (March, 31) in order to calculate the initial watershed hydrological state (soil, snow and channel water contents, groundwater level, soil freezing depth, etc.) that initializes the forecast. The simulations start from the end of the previous freshet, i.e. 8-9 months before the forecast issue date;

(2) Selection of a weather scenario[1] from the N_{EPS}-member ensemble of the observed weather or from the N_{WG}-member ensemble of the generated weather for the forecast horizon;

(3) Simulating daily inflow hydrograph by the ECOMAG model driven by the selected scenario;

(4) Repetitive selection of the next weather scenario (step 2) from th ensemble and calculation of the corresponding inflow hydrograph (step 3). The corresponding ensemble of *N* inflow hydrographs is formed ($N=N_{EPS}$ or $N=N_{WG}$);

(5) From each of the modelled hydrographs, the following inflow characteristics are derived: (1) inflow volume (hereafter referred to as *W*), (2) maximum inflow discharge (Q_{max}), (3) number of days with the inflow discharge above the mean observed discharge for the forecast horizon (N_q), (4) number of days with the inflow discharge above the mean maximum observed discharge for the forecast horizon (N_{qMax}).

(6) Deterministic (ensemble mean) and probabilistic forecast are derived and verified for each of the inflow characteristics.

3.2.3 Verification measures

To verify deterministic and probabilistic model-based forecasts, as well as to compare them with each other and with the operational data-driven forecast of water inflow into the Cheboksary reservoir, we used the following, quite traditional, measures of the forecasts' efficiency and skill.

For the deterministic forecast verification, we used the mean error (ME), relative bias (BIAS), root-mean-squared error (RMSE), and Pearson's correlation coefficient R. In addition, for presentation, we used the Taylor diagram (Taylor, 2001) which combines three forecast characteristics in one chart, namely the forecast standard deviation, RMSE and the correlation coefficient between the observations and the forecasted values.

[1] Hereafter, by "weather scenario" we mean an array of weather time-series (daily precipitation amount, air temperature and humidity deficit) that are used to drive the hydrological model for the forecast horizon

For categorical forecast verification, we used measures that can be calculated from a contingency table (Ferro and Stephenson, 2011), such as probability of detection (POD, shows the correct forecast fraction of the observed events), false alarm ratio (FAR, shows the fraction of forecasts that did not occur), frequency BIAS (shows correspondence of the observed and the forecasted events), Heidke skill score (HSS, shows the advantage of the forecast as compared to a random forecast), Hanssen and Kuipers score (KSS, can detect if the forecast is hedging) and Symmetric Extremal Dependency Index (SEDI, evaluates the performance of the forecast of rare binary events).

The probabilistic ensemble forecasts performance was assessed by several verification measures. The ability of forecasts to correctly predict the category of the event occurred within several categories was measured by Ranked Probability Score (RPS) (Wilks, 1995) that can also be treated as the mean squared error of the probabilistic forecast. The probability forecast efficiency relating to streamflow climatology was measured by the Ranked Probability Skill Score (Wilks, 1995). To visualize the specifics of probabilistic forecasts, three diagrams were employed. The predictive Q-Q plot (Laio and Tamea, 2007) was used to assess the degree of correspondence between tthe cumulative distribution function of predictions and the observed values. The reliability diagram (Hartmann et al., 2002) was used to plot the forecast probability against the relative frequency of the observations in the corresponding forecast probability bin. Finally, the discrimination diagram (Wilks, 1995) was used to show the frequency of each forecast probability for events and non-events.

Full list of the aforementioned verification measures, their formulations, units, and value ranges are presented in Table 1S (Suppl. Materials)

4. Results and discussion

4.1 Calibration and evaluation of the hydrological model

The hydrological model was calibrated and evaluated against the daily time-series of water inflow into the Cheboksary reservoir for the periods of 2000-2010 and 1982-1999, 2011-2016, respectively. The observed inflow data do not account for inflow from the upstream Nizhegorodskoe reservoir. Fig. 2 compares hydrographs of the observed and the simulated daily inflow discharges. Nash-Sutcliffe efficiency for daily inflow discharge is rather high (NSE=0.80) and ranges from 0.79 for the evaluation period to 0.83 for the calibration one. One can see that the model demonstrates good performance with respect to this criterion. Additionally, a small difference between the criteria estimated for the calibration and evaluation periods confirms the model robustness (Gelfan et al., 2015).

The model performance was tested also by comparison of the observed and simulated inflow characteristics, which were used then for the forecast verification and listed in sub-section 3.2.2. Figure 3 shows scatterplots of the observed and simulated characteristics of the inflow into the Cheboksary reservoir in April – June. In general, the inflow volume is well simulated, yet slightly underestimated for the high flows (above 50 km^3, see Fig. 3a). Maximum inflow discharge is a highly uncertain characteristic, but is still well simulated by the model (Fig. 3c). The number of days above a certain inflow

discharge threshold is a highly important characteristic for various users, e.g. waterways navigation and water supply. For the number of days above long-term (1982-2016) mean inflow discharge during the period between April and June, the model shows less days than the observed ones (Fig. 3b) – 31 against 36 days, on average for the whole period. For the number of days above long-term mean maximum inflow discharge the model also shows less days (Fig. 3d) – 13 against 17 days, on average.

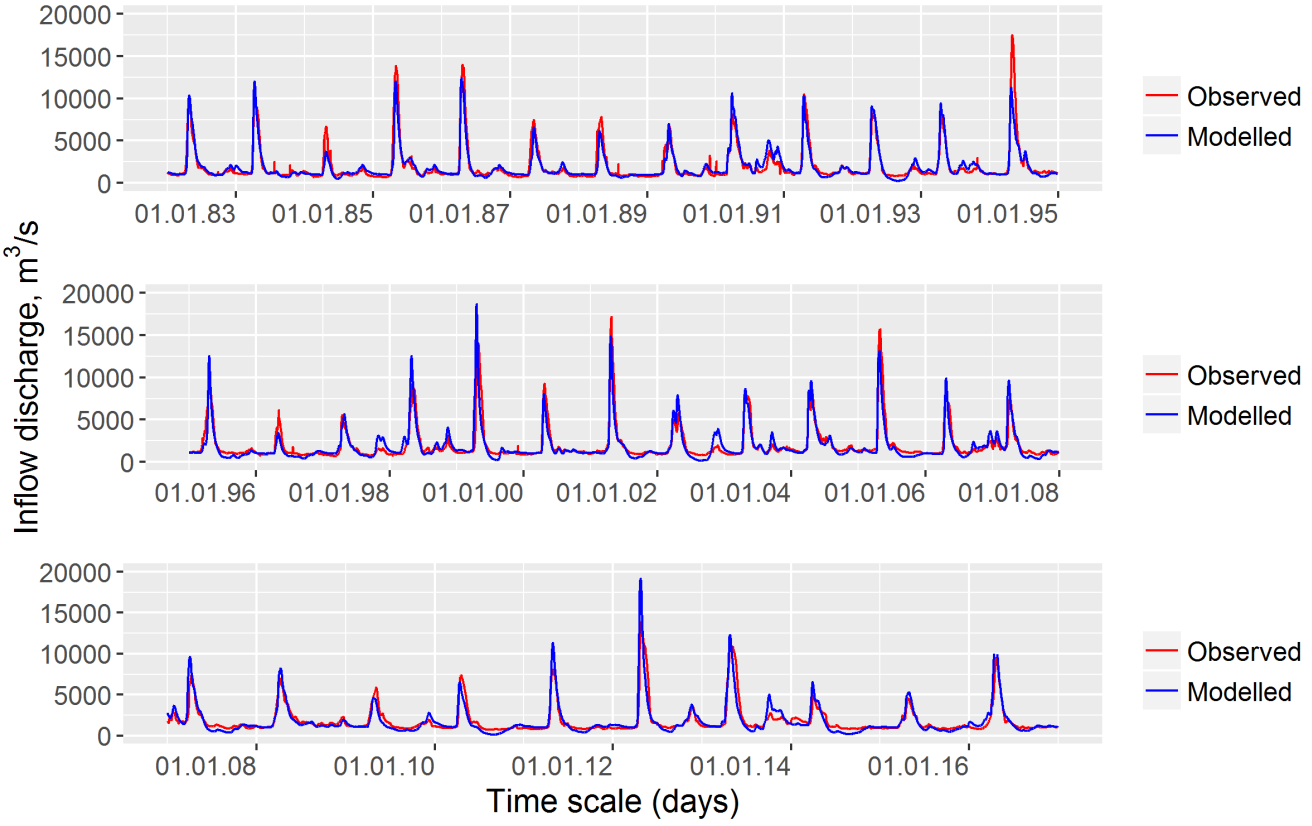

Figure 2: Observed and simulated daily discharge of inflow into the Cheboksary reservoir

Relative bias (BIAS) of the inflow volume in April – June for the whole period 1982 – 2016 is BIAS=-3%; RMSE of the inflow volume (5.23 km^3) consists 55% of the observed data standard deviation (σ_W=9.41 km^3). Relative bias of the maximum inflow discharge BIAS=5% m^3/s; RMSE=2321 m^3/s, that is 30% lower than the standard deviation of the observed maximum inflow discharge (3385 m^3/s).

The obtained results allow us to conclude that the developed model can be considered as a suitable tool for the long-term hydrological forecast of spring water inflow into the Cheboksary reservoir.

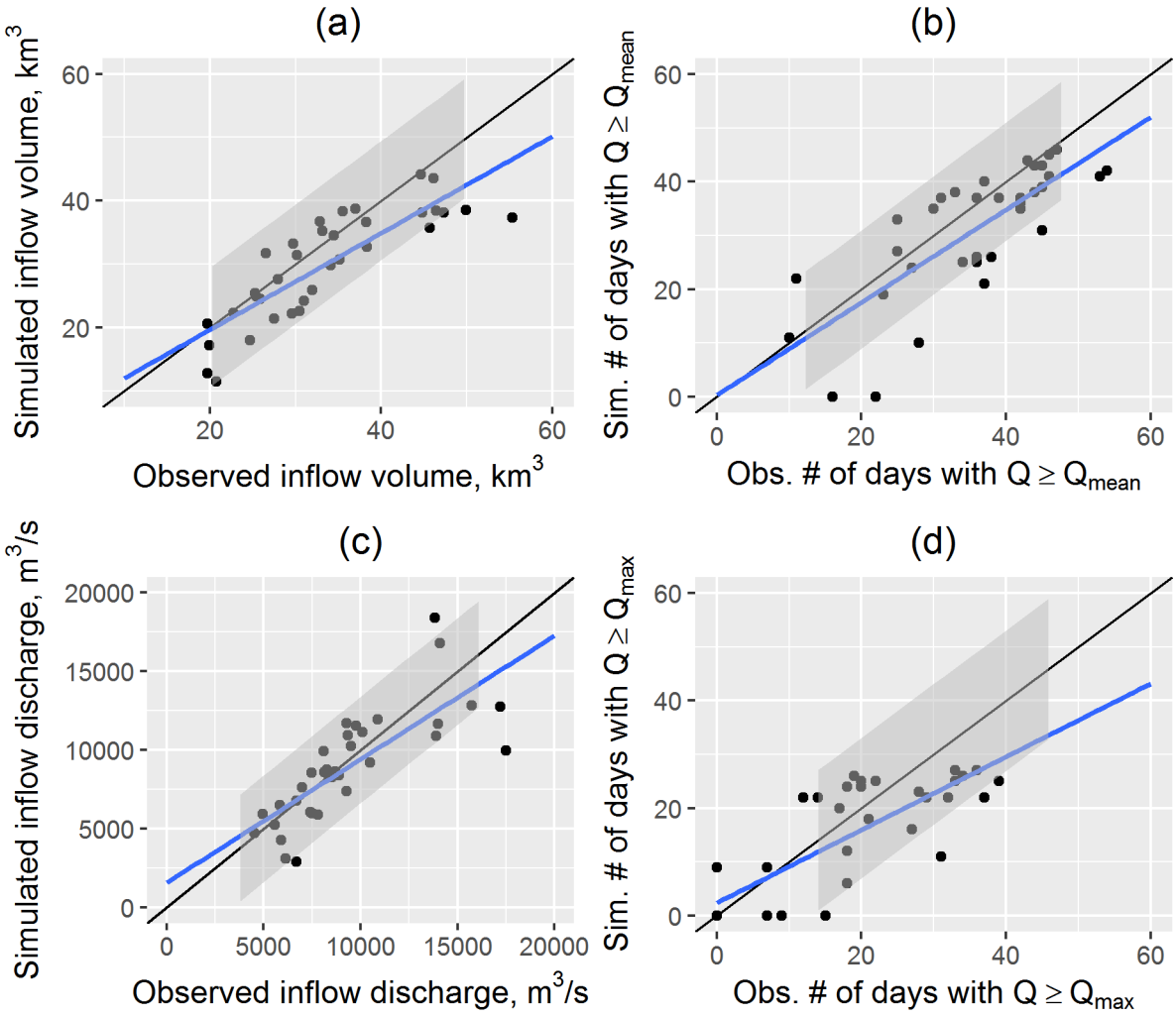

Figure 3: Scatterplots for the observed and simulated characteristics of inflow into the Cheboksary reservoir during April – June: volume W (a), maximum discharge Q_{max} (c), number of days above mean inflow discharge N_q (b), number of days above mean maximum inflow discharge N_{qMax} (d) Blue line represents linear fit. Grey shaded area denotes variance band of ±1 standard deviation of the respective observed values

4.2 MSFR_WG: parameter estimation and model testing

Time series of daily precipitation, air temperature and humidity deficit observed at the meteorological stations located at the Cheboksary reservoir basin for 51 years (1966-2016) are used to estimate the nine parameters of the developed stochastic model. The parameters estimated by the method of moments are shown in Table 2S in the Supplementary Materials. The stochastic models were comprehensively tested through their ability to reproduce the main statistical characteristics of meteorological processes at the Cheboksary reservoir basin. For testing, we compared only those characteristics of the observed and simulated time-series, which are neither the parameters of the model nor a single-valued function of the

parameters as suggested in Gelfan (2010). Statistics of the 1000-member Monte-Carlo generated ensemble of daily meteorological variables were compared with the following corresponding statistics derived from observations: mean and variation of annual and monthly values, autocorrelation functions of daily and monthly values of the specific variables. Results demonstrating comparison between statistical properties of the observed and simulated series are shown in Supplementary Materials for spring months and for several selected stations (Figs. 1S-8S).

4.3 Forecast verification

4.3.1 Ensemble (model-based) and operational (data-driven) deterministic forecasts

We verified the two types of the ensemble forecasts (ESP-based and WG-based), compared them with each other and with the operational forecasts of water inflow into the Cheboksary reservoir for April-June 1982-2016. To make a deterministic forecast, the forecasted inflow characteristics were averaged over the corresponding ensembles (51-member in the case of the ESP-based forecast and 1000-member for the WG-based forecast) to produce a single-value forecast of the desired characteristic: W, Q_{max}, N_q, N_{qMax}. Operational forecasts of inflow volume for the same April-June periods of 1982-2016 were obtained from the official Roshydromet forecast bulletins (reports). All forecasts were analyzed to assess the forecast performance measures: mean absolute error, BIAS, and RMSE. The results are presented in Table 1.

Table 1 – Statistics of the operational (op.) and the ensemble (ESP and WG) deterministic forecasts of inflow into the Cheboksary reservoir for April-June in 1982-2016

Inflow	Obs.	Mean			Mean error			BIAS*			RMSE		
characteristics		Op.	ESP	WG	Op.	ESP	WG	Op.	ESP	WG	Op.	ESP	WG
W, km^3	33.4	32.9	32.3	33.5	-0.5	-1.1	0.1	-1%	-3%	0%	6.52	5.06	5.19
Q_{max}, m^3/s	9355	N/A	9463	9958	N/A	108	603	N/A	1%	6%	N/A	1970	2244
N_q, days	35.9	N/A	35.9	36.1	N/A	0	0.2	N/A	0%	1%	N/A	8.0	8.8
N_{qMax}, days	17.0	N/A	16.2	17.1	N/A	-0.8	0.1	N/A	-5%	1%	N/A	7.4	8.2

* The measure abbreviations are defined in Table 1S

First, the deterministic forecasts of the inflow in April-June were compared to the operational forecasts for 1982-2016. As shown in Table 1, the mean error of the operational forecasts appeared to be quite low (around 1%) and close to those of the ESP and WG ensemble average values. However, the operational forecasts RMSE values are significantly higher than those of the ESP and WG forecasts and accounts for almost 70% of the observed inflow volume variability σ_W=9.61 km^3. For the ESP and WG-based forecasts these values are around half of σ_W.

Fig. 4 compares the inflow volume forecast errors of the operational forecasts (Fig. 4a) with the ESP-based forecast errors (Fig. 4b). The shaded area in the figures represents the area of the acceptable error [$-0.674\sigma_W$; $0.674\sigma_W$] = [-6.48 km^3; 6.48

km^3]. In the Russian operational forecasting practice, a forecast is considered acceptable if its error falls into this area, and the forecast acceptability is calculated as the ratio of the acceptable forecasts to the whole number of forecasts. According to the assumption of the Gaussian distribution of the forecast errors, 50% of the forecasts by climatology should fall into this interval. It can be seen from Fig. 4 that five of 35 ESP-based forecasts (in 1985, 1994, 2002, 2005 and 2011) and every third (12 of 35) operational forecasts were not acceptable, i.e. the ESP-based forecast acceptability is 89% and that of the operational forecast is 66%. Note that the inacceptable forecasts in both cases occurred in the years with spring precipitation amount significantly differing from the corresponding climatic mean.

To compare the ESP-based and the WG-based forecasts, we present them in the form of the Taylor diagram (Fig. 5; Taylor, 2001), which combines three forecast characteristics in one chart, namely, the forecast standard deviation, RMSE and the correlation coefficient between the observed and the forecasted values of the inflow characteristics. The values of all characteristics are normalized by dividing the RMSE by the standard deviation of the observations. This normalization provides a demonstration of the forecast efficiency expressed in fractions of the observed standard deviation. As long as the forecast RMSE is less than the standard deviation of the observations, the forecast can be considered efficient against climatology.

It can be seen from Fig. 5 that the ESP-based forecasts of W, Q_{max} and N_{qMax} are slightly better correlated with the observations than the WG-based forecasts. Pearson's R values of the ESP-based forecasts are over 0.8 for all characteristics, except for N_q. Forecasts of N_q are less correlated with the observations, with values of R for ESP-based and WG-based forecasts equal 0.73 and 0.63, respecrively. Forecasts of Q_{max} and N_{qMax} show normalized RMSE values around 58-67% of the standard deviation of the corresponding observed characteristics.

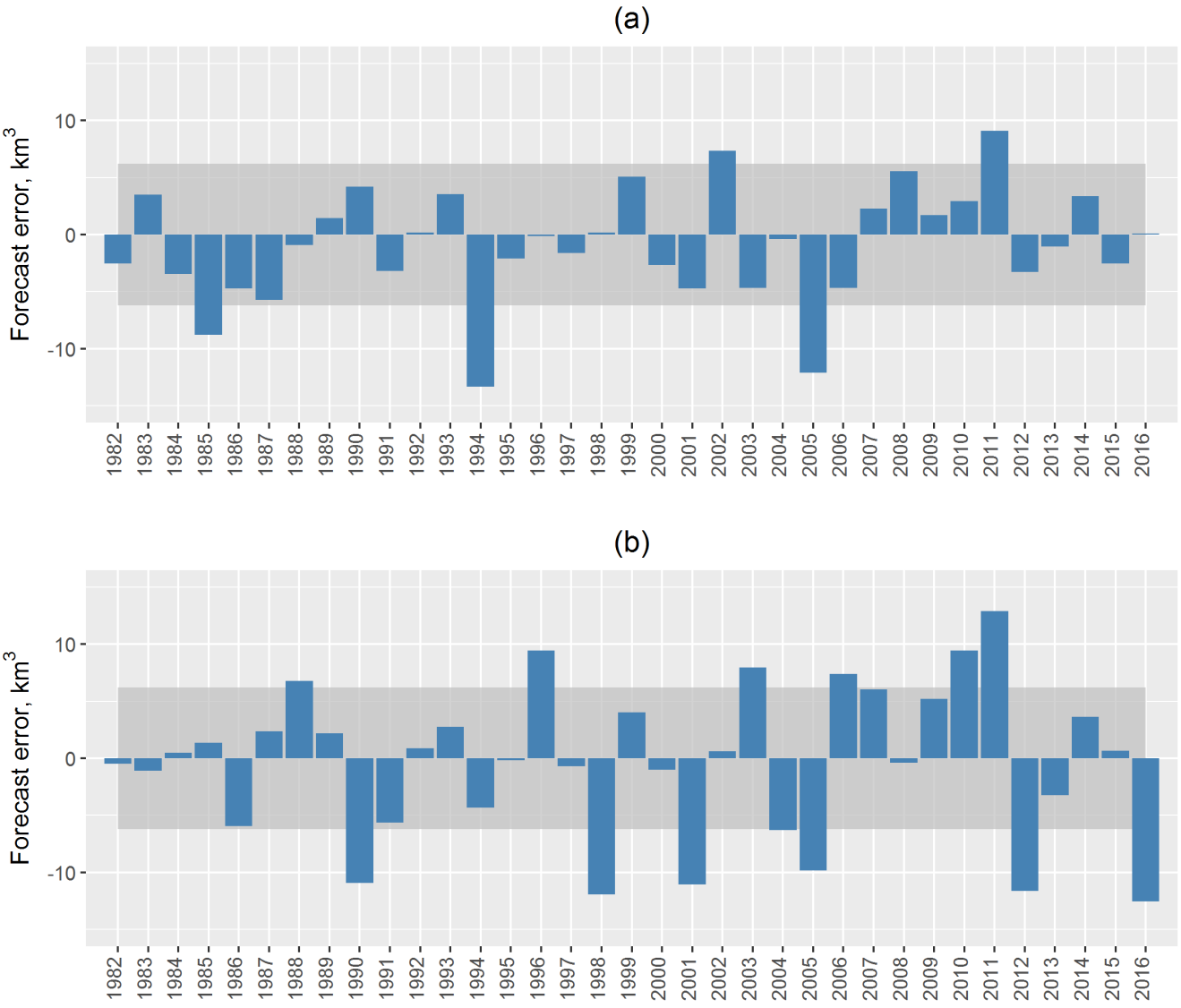

Figure 4: Errors of the ESP-based (a) and operational (b) forecasts of the April-June volume of water inflow into the Cheboksary reservoir. (Solid lines present boundaries of the acceptable error equaled $\pm 0.674\sigma_W$, where σ_W=9.61 km^3 is the standard deviation of the observed inflow volume)

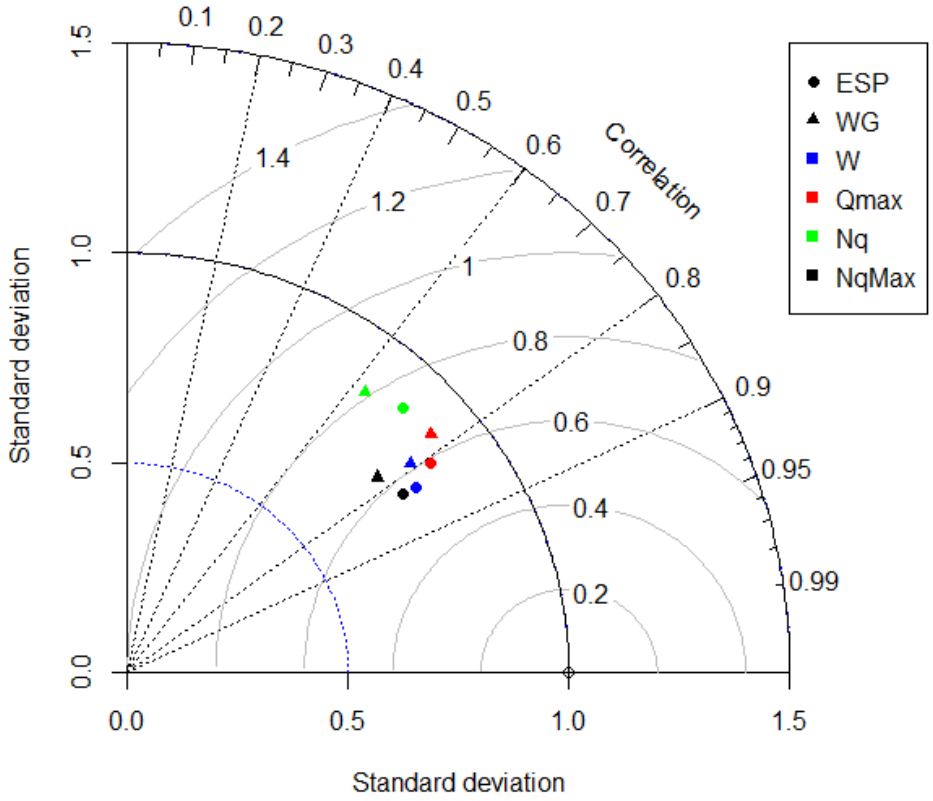

Figure 5: Taylor diagram of ESP (circle) and WG-based (triangle) forecasts of inflow volume *W* (in blue), maximum inflow discharge Q_{max} (red), number of days with inflow discharge above mean N_q (green), number of days with inflow discharge above maximum N_{qMax} (black).

For the purpose of reservoir management, it is often crucial to determine whether the expected inflow characteristic will exceed the corresponding mean value. To verify the methodology capability of predict this exceedance, the observations and forecasts were converted into binary vectors with "zero" value representing the event of non-exceedance of the mean annual value and "one" for the event occurrence. For example, for W, the event occurs with the exceedance of mean inflow volume during April – June. The forecast binary measures assessed with the use of the contingency tables and described in Table 1S are shown in Table 2.

The forecasts show good detection estimates (even perfect for Q_{max}) for both model-based methodologies. However, as the frequency Bias is high, this might be the result of overprediction, so with the high values of False Alarm Ratio and Hansen-Kuipers score. For *W* and Q_{max}, the forecast accuracy with the Heidke Skill Score (HSS) of around 60% is better than the accuracy of random chance; this means that the forecast is capable of detecting the occurrence of rare extreme events, which is shown by high values of the Symmetric Extremal Dependency Index (SEDI). Overall, the presented binary verification measures demonstrate a slight advantage of the ESP-based forecasts over the WG-based forecasts, though the differences are not significant.

Table 2 – Verifications measures for the binary forecasts (Cheboksary reservoir, April-June of 1982-2016)

Inflow characteristics	POD* R** ∈ [0;1]; PFM*** =1			FAR R∈ [0;1]; PFM=0			Freq. BIAS R∈ (-∞;+∞) PFM=0			HSS R∈ (-∞;1]; PFM=1			KSS R∈ (-1;1) PFM=1			SEDI R∈ (-1;1) PFM=1		
	Op.	**ESP**	**WG**	**Op.**	**ESP**	**WG**	**Op.**	**ESP**	**WG**	**Op.**	**ESP**	**WG**	**Op.**	**ESP**	**WG**	**Op.**	**ESP**	**WG**
W, km^3	0.69	0.87	0.87	0.31	0.24	0.29	1.31	1.13	1.27	0.42	0.66	0.55	0.42	0.67	0.57	0.57	0.82	0.73
Q_{max}, m^3/s	N/A	1.00	1.00	N/A	0.33	0.37	N/A	1.50	1.58	N/A	0.66	0.61	N/A	0.74	0.70	N/A	0.93	0.92
N_q, days	N/A	0.77	0.75	N/A	0.23	0.20	N/A	1.00	0.91	N/A	0.39	0.41	N/A	0.39	0.42	N/A	0.53	0.57
N_{qMax}, days	N/A	0.79	0.76	N/A	0.17	0.20	N/A	0.95	0.91	N/A	0.60	0.57	N/A	0.60	0.62	N/A	0.76	0.67

* The measure abbreviations are defined in Table 1S

** R is the Range of the measure value

*** PFM is the Perfect Forecast Measure value

Binary measures of the operational forecasts of inflow volume are worse than those of the model-based forecasts. For instance, only 69% of the observed events (exceedance of mean inflow volume) are correctly forecasted by the current operational methodology and its accuracy relative to that of random chance is less then 50%. An ability of a user to detect rare events on the basis of the operational forecast is also much lower than with the help of ensemble forecasts.

Thus, both continuous (Table 1) and binary (Table 2) model-based forecasts of inflow volume appeared to be more preferable, in general, than the corresponding operational forecasts. However, as one can see from Fig. 4, there were several years, when the operational forecasts were more accurate (in terms of the absolute error) than the ensemble ones. We found that most often the operational forecast outperforms the model-based forecast in those years when the modelled initial snow water equivalent (SWE) on the forecast issue date significantly differed from the observed SWE. Since the latter is the main factor affecting the freshet volume, more accurate (observed) initial snow conditions used in formula (2) resulted in more accurate forecast than the one initiated from the simulated SWE.

4.3.2 Freshet of 2017: testing the ensemble methodology

In the beginning of spring 2017, the basin's pre-melt conditions were close to climatology: snow water storage was 10 to 15% above the long-term mean value, soil water content and freezing depth were close to the corresponding mean values. However, the weather conditions during the spring freshet formation appeared to be significantly different from climatology. Anomalous warm and sunny weather that settled over the basin in the first half of March has led to commencement of snowmelt and river stage ascent at least half-month earlier than the mean dates. The last decade of March was, on the contrary, cold and damp, and the precipitation amount was twice above normal for this period. As a result, by the end of March the inflow volume (5.13 km^3) into the reservoir exceeded mean March inflow by 32%. Periods of intense snowmelt, interchanged with cold spells and large amount of precipitation, including snowfalls during April and May of 2017 (a number of stations have registered snowfall even in June). Such diversity in weather conditions during the snowmelt period and their difference from the climatology resulted in rather untypical regime of inflow into the Cheboksary reservoir.

The ESP-based forecasting technique was tested in operational mode during the freshet period of 2017. The forecasts were issued on March 1^{st}, 15^{th} and 27^{th} for the period from April, 1^{th} till June, 30^{th}. Fig. 6 shows daily forecast ensembles for this period compared to the observed inflow data.

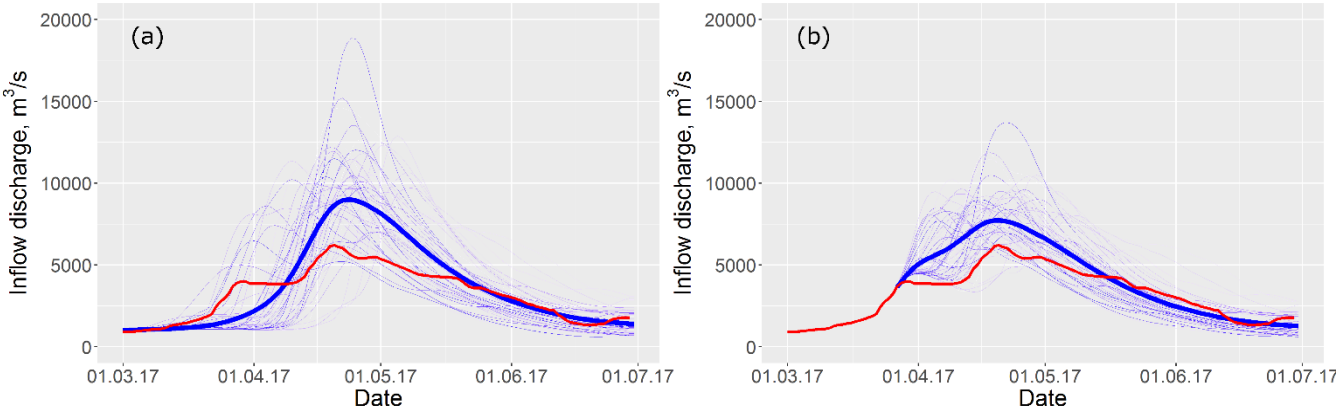

Figure 6: The ESP-based forecast of daily inflow discharge for the period from April, 1th till June, 30th issued on March 1st (a) and March 27th (b). Thin blue lines – ensemble of the forecasted hydrographs; bold blue line – mean ensemble hydrograph, red line – observed hydrograph of inflow into the Cheboksary reservoir.

Box plots of the ESP-based forecasts of different inflow characteristics are presented in Fig 7. All forecasts of inflow volume showed low errors (Fig. 7a), unlike the maximum discharge forecasts (Fig. 7b), however, the Q_{max} forecast range envelopes the observed maximum inflow discharge. Both forecasts of number of days over thresholds showed low errors (Fig. 7b, d), e.g. just before the beginning of April we correctly forecasted low freshet with the absence of days when inflow discharge exceeds the mean maximum discharge for the period of observations.

In 2017, Roshydromet also issued forecast of spring water inflow into the Cheboksary reservoir on the basis of the methodology presented above (Fig. 8). In contrast with the results presented in Table 1 and Fig. 4, which demonstrate general advantage of the ESP-based forecasts over the operational forecasts for 1982-2016, in 2017, operational forecast of inflow volume appeared to be more preferable. Possible explanation is again in errors in simulated pre-melt SWE used as initial conditions for the ESP-based forecast.

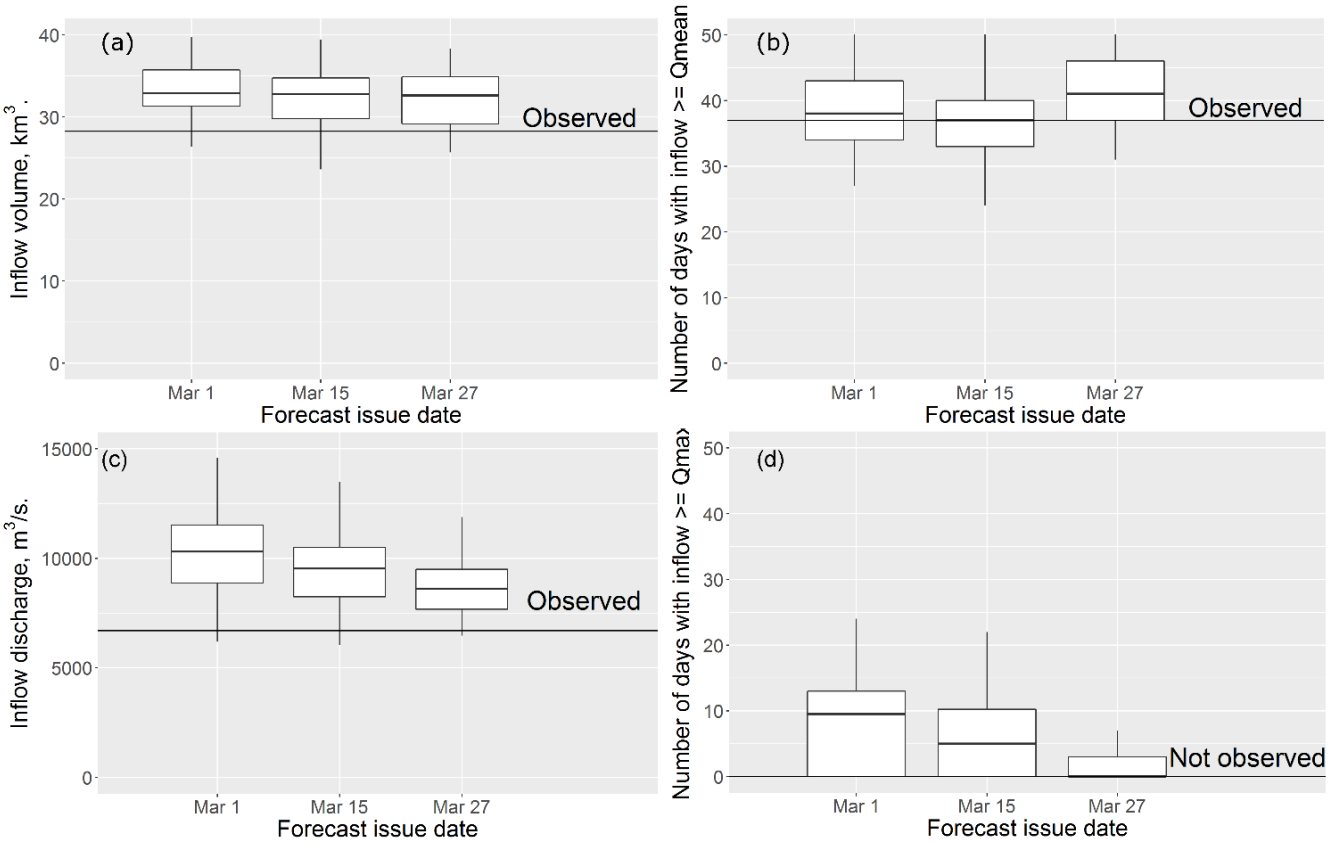

Figure 7: Box plots of the ESP-based forecast of inflow volume (a), maximum inflow discharge (c), number of days with inflow discharge above mean observed discharge (b), and number of days with inflow discharge above mean maximum observed discharge (d) for the period from April, 1th till June, 30th of 2017

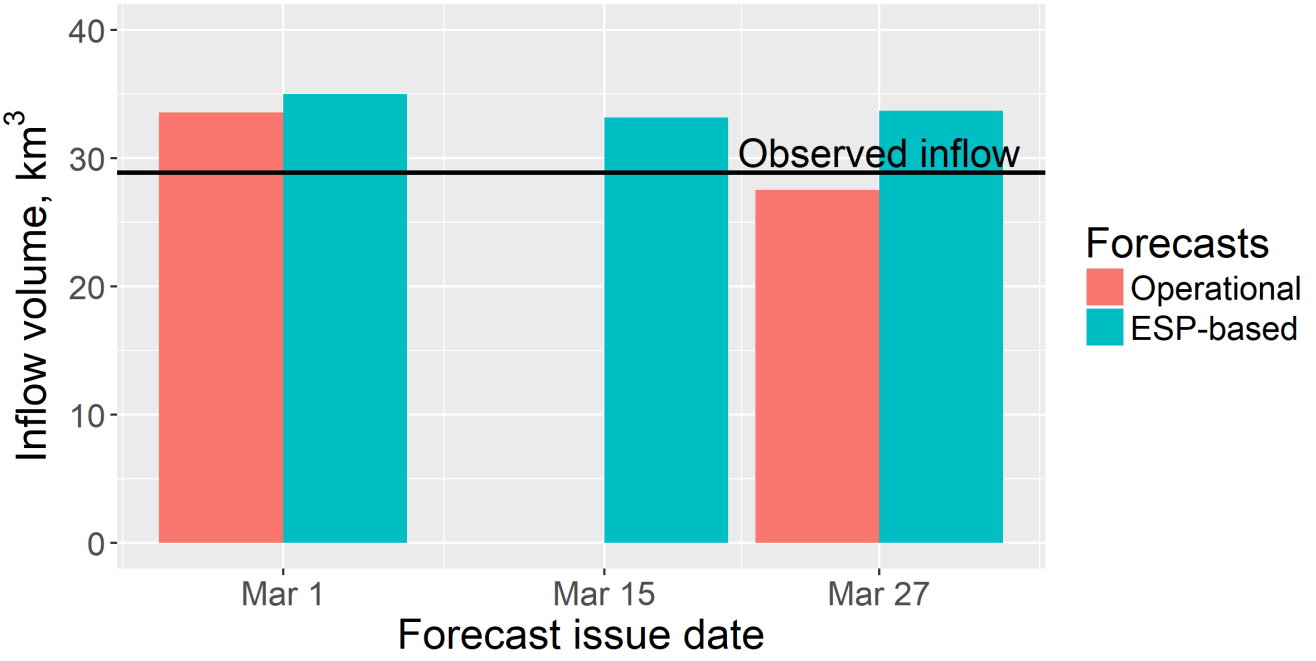

Figure 8: The ESP-based and operational forecasts of volume of water inflow into the Cheboksary reservoir for the period of 01/04/2017-30/06/2017 (line presents the observed value)

4.3.3 Probabilistic forecast

In this section, the operational forecast, which is issued in deterministic form only, is not discussed.

One of the main advantage of the ensemble forecasting is the ability to assess the uncertainty that is nested in the future possible behavior of the hydrological system. The resulting ensemble is used to create cumulative distribution function (CDFs) of the desired characteristic in j^{th} forecast as

$$F_m(j) = \sum_{i=1}^{m} f_i(j),\ m=1,\ldots,M;\ j=1,\ldots,N, \tag{4}$$

where M are the forecast probability bins on the interval [0;1]; N – total number of forecasts; f_i – probability of forecast in m^{th} bin.

CDFs of the forecasted inflow volume W for the period from April, 1 to June, 30 of 35 years (1982-2016) are shown in Fig. 9. Three CDFs are combined in each plot: two CDFs of forecasts calculated under ESP-based and WG-based weather scenarios, and CDF of the observed inflow volume in the specific year (CDFs of observations can be represented as the Heaviside step function). One can see from Fig. 9 that for the most of the years the inflow is not far from the most probable one, in other words, CDF of the forecasts crosses CDF of observations around 50% probability. For almost all years

observed inflow lies within the range of the ensemble. Exceptions are 1994, 2002, 2005 and 2011, i.e. once per 8-9 years, on average, ensemble forecast range does not cover observed inflow because of large forecast errors.

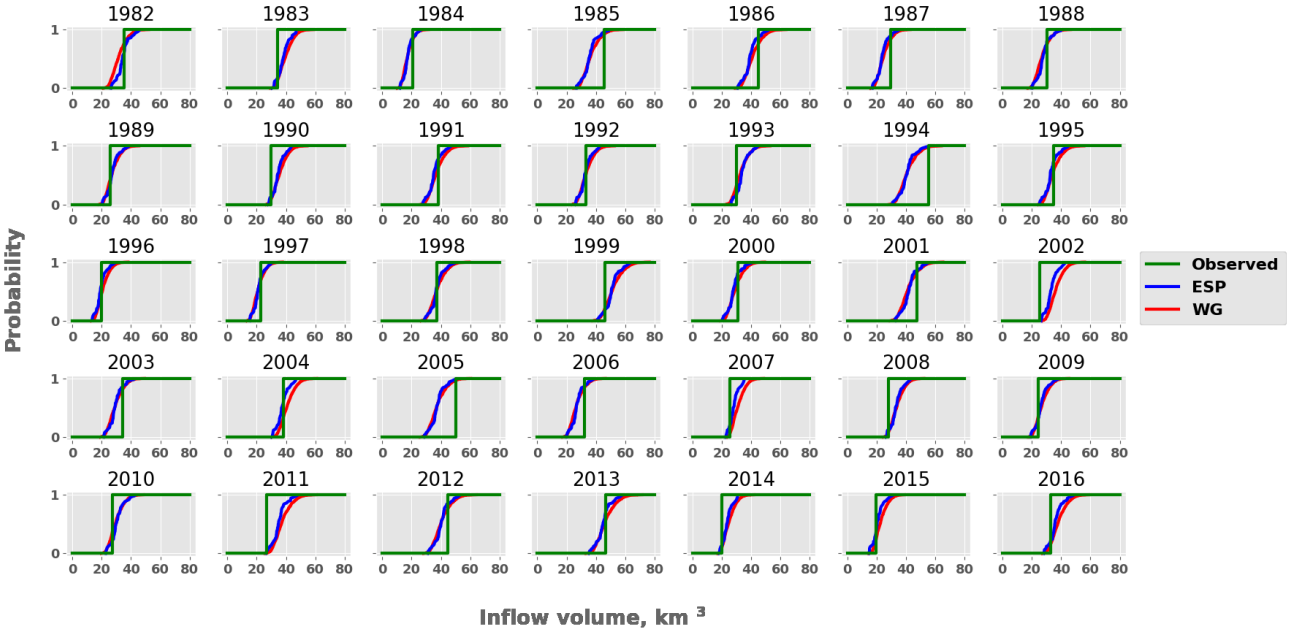

Figure 9: Cumulative probability distribution functions for *W* in April – June for all years between 1982 and 2016. Green line – observed inflow, orange – ESP-based forecast, blue – WG-based forecast

To quantify an ability of forecasts to predict probability of the event occurred within the pre-assigned inflow categories, we used the RPS measure. The forecast efficiency was measured by the RPSS criterion relating the verified forecast to streamflow climatology (both RPS and RPSS formulations are presented in Table 1S)

Both forecasts demonstrate moderate improvement over climatology: according to the RPSS value, around 30% on average both for *W* and for Q_{max} (Table 3). Probably, accounting for a seasonal weather forecast and conditioning historical weather patterns on this forecast could result in a higher improvement over the streamflow climatology, however a reliable seasonal weather forecast for the study region is not available.

In addition, we compared the ESP-based and the WG-based forecasts by setting the former one as a reference forecast. The modified RPSS is formulated in this case as:

$$RPSS_{\text{modified}} = 1 - \frac{RPS_{WG}}{RPS_{ESP}} \quad (5)$$

Table 3. Ranked Probability Score and Skill score for the forecasts

Inflow characteristics	RPS_{ESP}	RPSS (ESP vs. climatology)	RPS_{WG}	RPSS (WG vs. climatology)	$RPSS_{modified}$ (WG vs. ESP)
W	0.33	0.38	0.38	0.28	-0.16
Q_{max}	0.43	0.28	0.49	0.20	-0.13

As one might expect from a comparison of the unmodified RPSS measures, the modified one showed that the WG forecasts are less skillful than the ESP with modified RPSS values of -0.16 for W and -0.13 for Q_{max}.

To compare quantiles of the forecasted characteristics with the quantiles of the corresponding observations, we used the predictive Q-Q plot (Laio and Tamea, 2007). As one can see from Fig. 10a, the predictive Q-Q plot of the inflow volume forecasts demonstrates good agreement with the distribution of the observations. This is fairly consistent for both methodologies and for all quantiles, but for rare events there is an underestimation of the predictive uncertainty, expressed as an offset of 1:1 line in the upper right corner of the plot. For the maximum inflow discharge (Fig. 10b), one can see overprediction in both methodologies. However, the behavior of ESP-based and WG-based forecasts of rare events is different in terms of predictive uncertainty. Particularly, the WG-based forecasts of the events of low exceedance probability appeared to be closer to the 1:1 line.

Additional comparative features of the ensemble forecasts of both types, one can detect from the reliability and discrimination diagrams presented in Figs. 9S-12S of the Supplementary Materials.

Overall, all presented measures of the probabilistic forecast performance are slightly better for the ESP-based forecasts than for the WG-based forecasts, though the differences are not significant and hardly interpretable. At the same time, verification measures obtained from the large ensemble of the WG-based forecasts should be more statistically reliable, that is demonstrated in the next section for two measures: CDF and RPSS.

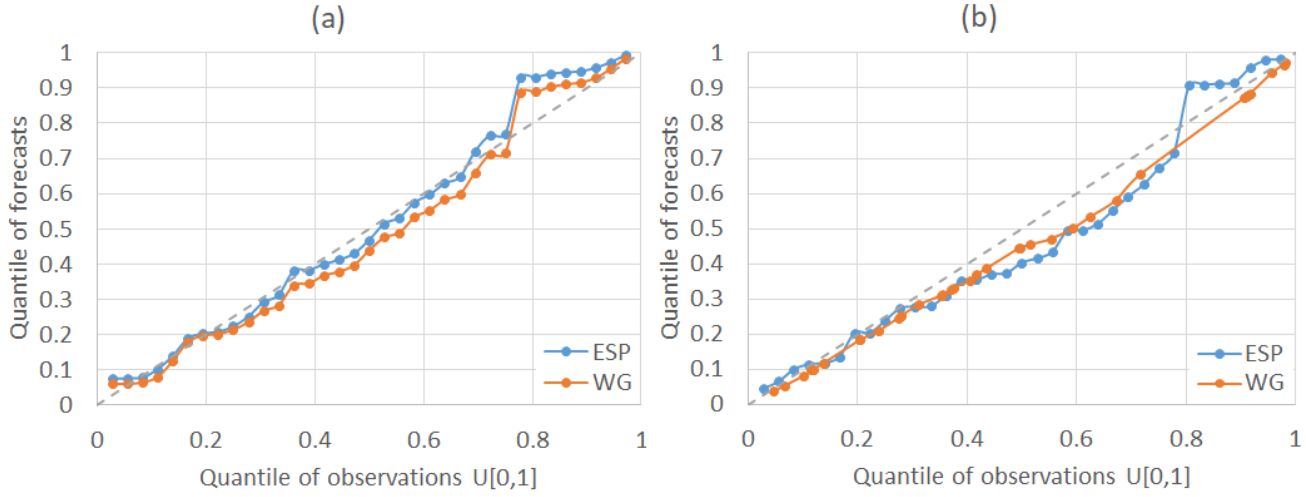

Figure 10: Predictive quantile-quantile plots for inflow volume (a) and inflow discharge (b) forecasts

4.3.4 Ensemble size effect on the verification measures: two examples

It can be seen from Fig. 9 that the CDFs appeared to be close to each other for both used ensemble methodology. However, the sample variance of the CDF is significantly different due to different amount of scenarios in the ensembles: 51 in the ESP-based ensemble and 1000 in the WG-based ensemble. To illustrate such difference, we assessed confidence bands for CDFs derived from both forecasting approaches. A two sided confidence bands were expressed from Dvoretzky-Kiefer-Wolfowitz inequality as (e.g. Massart, 1990):

$$\mathrm{P}\left(\sup_{x\in R}\left|\hat{F}_n(x)-F(x)\right|\geq\varepsilon\right)\leq 2\exp\left(-2n\varepsilon^2\right) \quad (6)$$

where $F(x)$ and $\hat{F}_n(x)$ are the CDF's ordinate and its empirical estimation from a sample of size *n*, respectively; ε is the constant depending on the significance level α as

$$\varepsilon=\sqrt{\frac{1}{2n}\ln\left(\frac{2}{1-\alpha}\right)} \quad (7).$$

For the pre-assigned confidence probability $p=(1-\alpha)$, the upper ($U(x)$) and the lower ($L(x)$) confidence bands of the empirical CDF $\hat{F}_n(x)$ are defined from (5) and (6) as:

$$U(x)=\min\left[\hat{F}(x)+\varepsilon,1\right] \quad (8)$$

$$L(x)=\max\left[\hat{F}(x)-\varepsilon,0\right] \quad (9)$$

Fig. 11 demonstrates difference between 95%-confidence intervals of the ESP-based inflow volume forecast as compared to the corresponding intervals of the WG-based forecast. We believe that this difference should be taken into account by the ensemble forecast developers when they use statistical verification measures for assessment of forecast performance, as well as by the users when they interpret the forecast.

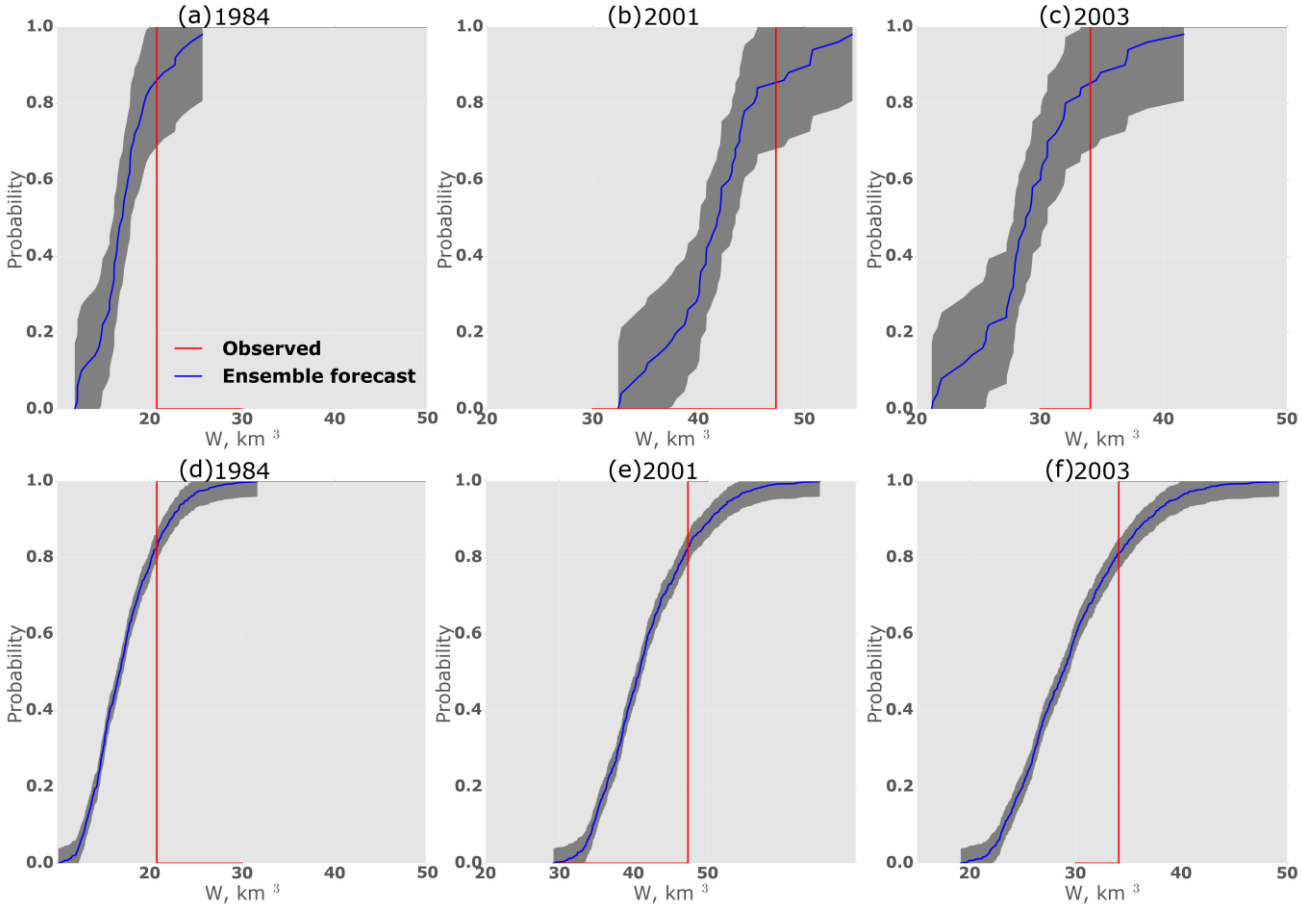

Figure 11: Cumulative probability distribution functions for W in April – June for selected years between 1982 and 2016 for ESP-based forecast (a – c) and WG-based forecast (d – f). Shaded area presents confidence interval of 95% confidence probability.

One can conclude from Table 3 that RPSS criterion demonstrates advantage of the ESP-based probabilistic forecast over the WG-based one as compared to climatology. However, it is important to take into account that the RPSS measure is strongly dependent on ensemble size and negatively biased (see, for instance, Müller et al. 2005, Weigel et al., 2007). A debiased estimate of RPSS can be formulated as (Weigel et al., 2007):

$$RPSS_{\mathrm{D}} = 1 - \frac{RPS}{RPS_{ref} + D} \tag{10}$$

where D is the correction term depending on the ensemble size, the climatological probabilities, and the number of categories. For very large ensemble size, the correction term D converges toward zero and the RPSSD toward the RPSS.

Fig. 12 demonstrates the dependence of the RPSS bias on sample size built with the use of approximation of D presented in (Weigel et al., 2007).

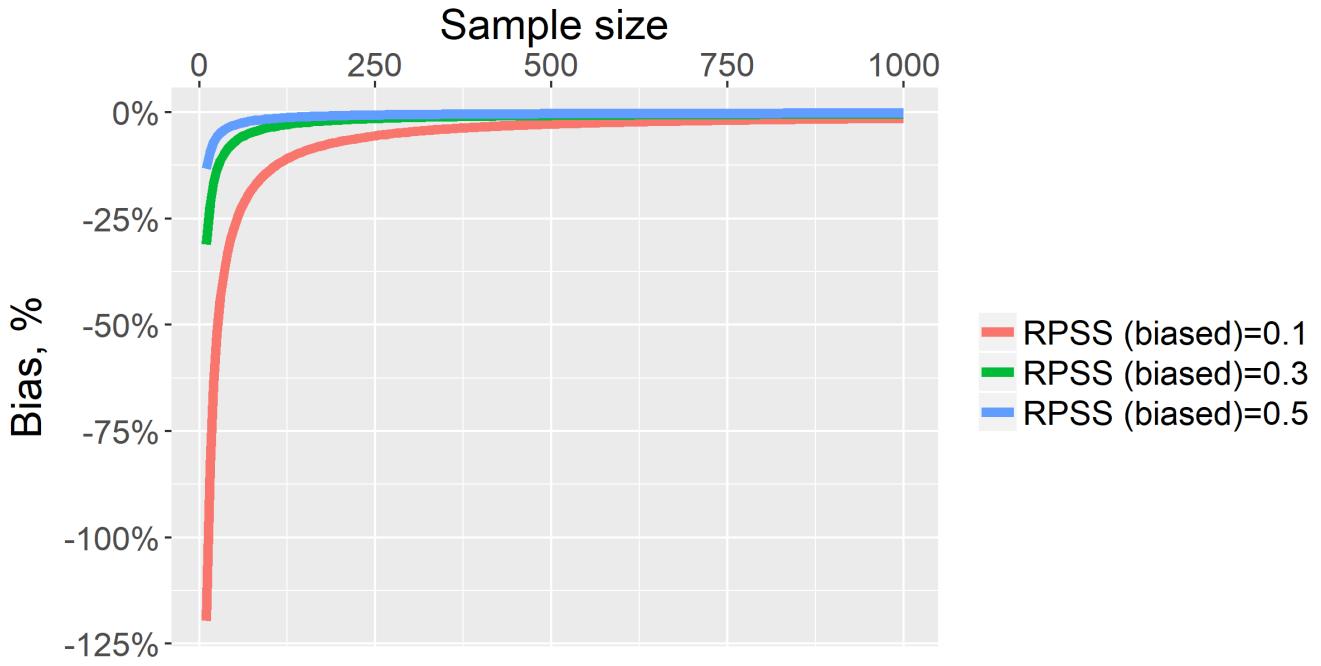

Figure 12: Negative bias of the RPSS-estimate in dependence on the ensemble size and the RPSS value

One can see from this Fig. 12 that under the used 51-member ensemble (i.e. the ESP-based ensemble) the bias can reach tens of percent depending on the RPSS estimate. Under the used 1000-member ensemble, the bias is close to zero.

5 Conclusions

The paper describes the development and preliminary assessment results of the long-term forecasting methodology for the water inflow into the Cheboksary reservoir – one of the eleven major river reservoirs of the Volga-Kama Reservoir Cascade. The methodology is based on a combination of semi-distributed hydrological model ECOMAG that allows for ensemble calculation of inflow hydrographs with the two different sets of weather ensembles for the lead-time period: observed weather data constructed on the basis of the ESP methodology, and synthetic weather data simulated by a weather generator. As it is mentioned in the Introduction section, we studied: (1) whether there is any advantage of the developed ensemble forecasts in comparison with the currently issued operational forecasts of water inflow into the Cheboksary reservoir, and (2) whether there is any noticeable improvement in the probabilistic forecasts when using the WG-simulated ensemble compared to the ESP-based ensemble.

Our findings can be summarised as follows:

1. For the 35-year period starting from the reservoir filling in 1982, both continuous and binary model-based ensemble forecasts (issued in deterministic form), overperformed the operational forecasts (currently used in practice) of the

April-June inflow volume. However, for several years (including 2017), the operational forecasts were more accurate in terms of the absolute errors. We found that the larger errors of the ensemble forecasts in these years were resulting from the errors in the modelled initial snow water equivalent on the forecast issue date comparing with the observed SWE. The prospects for improving the ensemble forecasts are in assimilation of the observation data (accounting for their reliability) on the forecast issue date.

2. Model-based ensemble approach extended a list of the forecasted inflow characteristics in comparison with the operational forecast. In addition to the inflow volume for the period of April-June, both ESP-based and WG-based methodology provided acceptable forecasts of maximum inflow discharge, number of days with inflow discharge above the mean observed discharge, and number of days with inflow discharge above the mean maximum observed discharge for this period. Thus, the ensemble methodology enhanced the informational content of the forecast in comparison with the operational one.
3. Overall, all the presented measures of the deterministic and probabilistic forecast performance are slightly better for the ESP-based forecasts than for the WG-based forecasts, though the differences are not significant and hardly interpretable. At the same time, verification measures obtained from the large ensemble of the WG-based forecasts appear to be more statistically reliable than the measures obtained from the ensemble size limited to the number of the historical years.

Currently we are in the process of updating and fine-tuning the presented forecast methodology for its practical tests during the freshet of 2018.

Our further research will be aimed at the following:

1. Refinement of the initial (on the forecast issue date) basin conditions through assimilation of the available observation data, first of all snow observations, into the hydrological model. Ensemble Kalman filtering is seen as as a promising procedure for that.
2. Use of the medium-range and seasonal weather forecasts for pre-processing the weather scenarios (both the ESP-based and the WG-based) for the hydrological forecast lead-time time.

Acknowledgements

The authors are very grateful to three anonymous Reviewers for criticisms and constructive comments. Also, we would like to thank to Ilias Pechlivanidis (handling editor) for his valuable suggestions.

The research related to developing methods of ensemble forecast and forecast verification technique was financially supported by the Russian Foundation for Basic Research (grants No. 16-05-00679 and 16-05-00599). The other research components, including those related to comparison of the ensemble forecast with the operational one, were financially supported by the Russian Science Foundation (grant No. 17-77-30006).

The present work was carried out within the framework of the Panta Rhei Research Initiative of the International Association of Hydrological Sciences (IAHS).

References

Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., Shamseldin, A. Y., Solomatine, D. P., Toth, E. and Wilby, R. L.: Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting, Prog. Phys. Geogr., doi:10.1177/0309133312444943, 2012.

Arnal, L., Wood, A. W., Stephens, E., Cloke, H. L. and Pappenberger, F.: An Efficient Approach for Estimating Streamflow Forecast Skill Elasticity, J. Hydrometeorol., doi:10.1175/JHM-D-16-0259.1, 2017.

Avakyan A.B.: Volga-Kama cascade reservoirs and their optimal use. Lakes and Reservoirs Research and Management, 3:113 – 121, 1998

Beckers, J. V. L., Weerts, A. H., Tijdeman, E. and Welles, E.: ENSO-conditioned weather resampling method for seasonal ensemble streamflow prediction, Hydrol. Earth Syst. Sci., doi:10.5194/hess-20-3277-2016, 2016.

Borsch, S. and Simonov, Y.: Operational Hydrologic Forecast System in Russia, in Flood Forecasting: A Global Perspective., 2016.

Buizza, R. and Palmer, T. N.: Impact of Ensemble Size on Ensemble Prediction, Mon. Weather Rev., doi:10.1175/1520-0493(1998)126<2503:IOESOE>2.0.CO;2, 1998.

Caraway, N. M., McCreight, J. L. and Rajagopalan, B.: Multisite stochastic weather generation using cluster analysis and k-nearest neighbor time series resampling, J. Hydrol., doi:10.1016/j.jhydrol.2013.10.054, 2014.

Chemerenko, E.P.: Long-term forecasting of spring inflow into the Cheboksary reservoir. Proceedings of the Hydrometeorological centre of Russia (in Russian). Vol. 324. P. 16-21, 1992

Day, G. N.: Extended Streamflow Forecasting Using NWSRFS, J. Water Resour. Plan. Manag., doi:10.1061/(ASCE)0733-9496(1985)111:2(157), 1985.

Demirel, M. C., Booij, M. J. and Hoekstra, A. Y.: The skill of seasonal ensemble low-flow forecasts in the Moselle River for three different hydrological models, Hydrol. Earth Syst. Sci., doi:10.5194/hess-19-275-2015, 2015.

Druce, D. J.: Insights from a history of seasonal inflow forecasting with a conceptual hydrologic model, J. Hydrol., doi:10.1016/S0022-1694(01)00415-2, 2001.

Ferro, C. A. T. and Stephenson, D. B.: Extremal Dependence Indices: Improved Verification Measures for Deterministic Forecasts of Rare Binary Events, Weather Forecast., doi:10.1175/WAF-D-10-05030.1, 2011.

Franz, K. J. and Hogue, T. S.: Evaluating uncertainty estimates in hydrologic models: Borrowing measures from the forecast verification community, Hydrol. Earth Syst. Sci., doi:10.5194/hess-15-3367-2011, 2011.

Gelfan, A. N. and Motovilov, Y. G.: Long-term hydrological forecasting in cold regions: Retrospect, current status and prospect, Geogr. Compass, doi:10.1111/j.1749-8198.2009.00256.x, 2009.

Gelfan, A. N., Motovilov, Y. G. and Moreido, V. M.: Ensemble seasonal forecast of extreme water inflow into a large reservoir, in IAHS-AISH Proceedings and Reports., 2015.

Gelfan, A., Gustafsson, D., Motovilov, Y., Arheimer, B., Kalugin, A., Krylenko, I. and Lavrenov, A.: Climate change impact on the water regime of two great Arctic rivers: modeling and uncertainty issues, Clim. Change, doi:10.1007/s10584-016-1710-5, 2017.

Gelfan, A.: Extreme snowmelt floods: Frequency assessment and analysis of genesis on the basis of the dynamic-stochastic approach, J. Hydrol., doi:10.1016/j.jhydrol.2010.04.031, 2010.

Gottschalk, L., Beldring, S., Engeland, K., Tallaksen, L., Sælthun, N. R., Kolberg, S. and Motovilov, Y.: Regional/macroscale hydrological modelling: a Scandinavian experience, Hydrol. Sci. J., doi:10.1080/02626660109492889, 2001.

Hanes, W. T., Fogel, M. M. and Duckstein, L.: Forecasting Snowmelt Runoff: Probabilistic Model, J. Irrig. Drain. Div., 1977.

Hartmann, H. C., Pagano, T. C., Sorooshian, S. and Bales, R.: Confidence builders: Evaluating seasonal climate forecasts from user perspectives, Bull. Am. Meteorol. Soc., doi:10.1175/1520-0477(2002)083<0683:CBESCF>2.3.CO;2, 2002.

Kuchment, L. S. and Gelfan, A. N.: Long-term probabilistic forecasting of snowmelt flood characteristics and the forecast uncertainty. In: Boegh, E., Kunstmann, H., Wagener, T., Hall, A., Bastidas, L., Franks, S., Gupta, H., Rosbjerg, D., and Schaake, J. (eds) Quantification and reduction of predictive uncertainty for sustainable water resources management, Vol. 313. Wallingford, UK: IAHS Publishers, p. 213–221, 2007

Laio, F. and Tamea, S.: Verification tools for probabilistic forecasts of continuous hydrological variables, Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hessd-3-2145-2006, 2006.

Lettenmaier, D.P. and Waddle, T.J.: Forecasting Seasonal Snowmelt Runoff: A Sum-mary of Experience with Two Models Applied to Three Cascade Mountain, Washington Drainages, WRS 59. – 1978. – 97 P., 1978

Li, H., Luo, L., Wood, E. F. and Schaake, J.: The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting, J. Geophys. Res. Atmos., doi:10.1029/2008JD010969, 2009.

Massart, P.: The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality, Ann. Probab., doi:10.1214/aop/1176990746, 1990.

McEnery, J., Ingram, J., Duan, Q., Adams, T. and Anderson, L.: NOAA's advanced hydrologic prediction service: Building pathways for better science in water forecasting, Bull. Am. Meteorol. Soc., doi:10.1175/BAMS-86-3-375, 2005.

McKay, M. D., Beckman, R. J. and Conover, W. J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, doi:10.1080/00401706.2000.10485979, 2000.

Mendoza, P. A., Wood, A. W., Clark, E., Rothwell, E., Clark, M. P., Nijssen, B., Brekke, L. D. and Arnold, J. R.: An intercomparison of approaches for improving operational seasonal streamflow forecasts, Hydrol. Earth Syst. Sci., doi:10.5194/hess-21-3915-2017, 2017.

Motovilov, Y. G.: Hydrological simulation of river basins at different spatial scales: 1. Generalization and averaging algorithms, Water Resour., 43(3), 429–437, doi:10.1134/S0097807816030118, 2016.

Motovilov, Yu., Gottschalk, L., Engeland, K., and Belokurov, A.: ECOMAG – regional model of hydrological cycle. Application to the NOPEX region. Department of Geophysics, University of Oslo, Institute Report Series no. 105, 88 p., 1999.

Müller, W. A., Appenzeller, C., Doblas-Reyes, F. J. and Liniger, M. A.: A debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes, J. Clim., doi:10.1175/JCLI3361.1, 2005.

Najafi, M. R. and Moradkhani, H.: Ensemble Combination of Seasonal Streamflow Forecasts, J. Hydrol. Eng., doi:10.1061/(ASCE)HE.1943-5584.0001250, 2016.

Najafi, M. R., Moradkhani, H. and Piechota, T. C.: Ensemble Streamflow Prediction: Climate signal weighting methods vs. Climate Forecast System Reanalysis, J. Hydrol., doi:10.1016/j.jhydrol.2012.04.003, 2012.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I - A discussion of principles, J. Hydrol., doi:10.1016/0022-1694(70)90255-6, 1970.

Pappenberger, F. et al.: Hydrological Ensemble Prediction Systems Around the Globe, in Handbook of Hydrometeorological Ensemble Forecasting, edited by Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H. L. Cloke, and J. C. Schaake, pp. 1–35, Springer Berlin Heidelberg, Berlin, Heidelberg., 2016.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P.: Numerical Recipes 3rd Edition: The Art of Scientific Computing., 2007.

Richardson, D. S.: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size, Q. J. R. Meteorol. Soc., doi:10.1256/smsqj.57714, 2001.

Shukla, S. and Lettenmaier, D. P.: Seasonal hydrologic prediction in the United States: Understanding the role of initial hydrologic conditions and seasonal climate forecast skill, Hydrol. Earth Syst. Sci., doi:10.5194/hess-15-3529-2011, 2011.

Svanidze, G. G.: Mathematical Modeling of Hydrologic Series. Water Resources Publications, Fort Collins, Colorado, USA, 324 p., 1980

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res. Atmos., doi:10.1029/2000JD900719, 2001.

Weigel, A. P., Liniger, M. A. and Appenzeller, C.: Generalization of the Discrete Brier and Ranked Probability Skill Scores for Weighted Multimodel Ensemble Forecasts, Mon. Weather Rev., doi:10.1175/MWR3428.1, 2007.

Wilks, D. S.: Statistical methods in the atmospheric sciences, Int. Geophys. Ser., doi:10.1007/s13398-014-0173-7.2, 1995.

Wood, A. W. and Lettenmaier, D. P.: A test bed for new seasonal hydrologic forecasting approaches in the western United States, Bull. Am. Meteorol. Soc., doi:10.1175/BAMS-87-12-1699, 2006.

Yossef, N. C., Winsemius, H., Weerts, A., Van Beek, R. and Bierkens, M. F. P.: Skill of a global seasonal streamflow forecasting system, relative roles of initial conditions and meteorological forcing, Water Resour. Res., doi:10.1002/wrcr.20350, 2013.

Zmieva E.S.: Forecasts of water inflow into the Kuibyshevskoe and Volgogradskoe reservoirs (in Russian). Gidrometizdat. Moscow., 255 p., 1964