Dear Dr. Pechlivanidis,

Thank you very much for the kind editor decision and letter. We have now thoroughly revised the manuscript, taking up all the suggestions from the reviewers. As requested, we explained the suitability of the MLR for the forecast in more detail based on hydrological processes. Additionally we included the formal tests of the MLR assumptions in the manuscript, partly in the main text, partly in the Annex. We also included the reliability analysis as suggested by reviewer 2 in the manuscript. Due to the additions the manuscript is now longer compared to the original submission, but we believe still concise enough for a scientific article.

I hope that we have met your and the reviewers requirements, and looking forward to the review results of the revised manuscript.

Kind regards,
Heiko Apel
On behalf of all co-authors.

Reply to reviewer comment hess-2017-340-RC1

Heiko Apel[1], Zharkinay Abdykerimova[2], Marina Agalhanova[3], Azamat Baimaganbetov[4], Nadejda Gavrilenko[5], Lars Gerlitz[1], Olga Kalashnikova[6], Katy Unger-Shayesteh[1], Sergiy Vorogushyn[1], Abror Gafurov[1]

[1]GFZ German Research Centre for Geoscience, Section 5.4 Hydrology, Potsdam, Germany

[2]Hydro-Meteorological Service of Kyrgyzstan, Bishkek, Kyrgyzstan

[3]Hydro-Meteorological Service of Turkmenistan, Ashgabat, Turkmenistan

[4]Hydro-Meteorological Service of Kazakhstan, Almaty, Kazakhstan

[5]Hydro-Meteorological Service of Uzbekistan, Tashkent, Uzbekistan

[6]CAIAG Central Asian Institute for Applied Geoscience, Bishkek, Kyrgyzstan

Correspondence to: Heiko Apel (heiko.apel@gfz-potsdam.de)

General referee comment:

There is an urgent need to improve the safety and operation of impoundments in Central Asia, yet hydrometeorological data to support inflow forecasting and water management are in short supply. This manuscript seeks to address these needs by developing a standard multiple linear regression model of melt season (April-September) discharge in 13 catchments. Forecasts are based on suites of predictors (precipitation, temperature, snow cover and composite variables) in January to June, and tested using a cross-validation technique applied to 16 years of monthly data. Following an exhaustive evaluation of all possible permutations of monthly and averaged predictors, best-performing models, and 20 near-optimal models are retained. The attendant mix of predictors and uncertainty bounds are then examined for months leading up to and at the start of the forecast season. Variations in forecast skill are qualitatively linked to catchment characteristics.

The overall approach to model development is necessarily pragmatic given the data and technical constraints of the region. Despite the simplicity of the approach, high explained variance (R2) is reported and the authors have bounded forecasts using envelopes of predictor suite uncertainty. However, it is unclear whether the underpinning data comply with the assumptions of the MLR model (i.e. linearity of relationships, homoscedasticity, no outliers, normally distributed and uncorrelated residuals). Furthermore, given the small number of cases (16) and relatively large number of independent variables (4) it is essential that significance levels and adjusted R2 values are reported for all retained MLR models. Significance of the model coefficients should also be tested and any insignificant variables removed. In some models, the predictor variable (e.g. May discharge) is not fully independent of the forecast variable (April to September discharge).

On this basis, publication is recommended subject to the following major revisions, minor corrections and clarifications.

We thank the referee for the critical and constructive comments. We provide detailed answers and justifications below, were the main comments are listed.

Main comments

[Abstract] Please incorporate more headline results, such as the range of forecast skill for forecasts issued before the onset of the main melt season, as well as typical forecast biases.

The suggestion has been taken up and the abstract reads now as follows:

The semi-arid regions of Central Asia crucially depend on the water resources supplied by the mountainous areas of the Tien Shan, Pamir and Altai mountains. During the summer months the snow and glacier melt dominated river discharge originating in the mountains provides the main water resource available for agricultural production, but also for storage in reservoirs for energy generation during the winter months. Thus a reliable seasonal forecast of the water resources is crucial for a sustainable management and planning of water resources. In fact, seasonal forecasts are mandatory tasks of all national hydro-meteorological services in the region. In order to support the operational seasonal forecast procedures of hydro-meteorological services, this study aims at the development of a generic tool for deriving statistical forecast models of seasonal river discharge. The generic model is kept as simple as possible in order to be driven by available meteorological and hydrological data, and be applicable for all catchments in the region. As snowmelt dominates summer runoff, the main meteorological predictors for the forecast models are monthly values of winter precipitation and temperature, satellite based snow cover data and antecedent discharge. This basic predictor set was further extended by multi-monthly means of the individual predictors, as well as composites of the predictors. Forecast models are derived based on these predictors as linear combinations of up to 3 or 4 predictors. A user selectable number of best models is extracted automatically by the developed model fitting algorithm, which includes a test for robustness by a leave-one-out cross validation. Based on the cross validation the predictive uncertainty was quantified for every prediction model. Forecasts of the mean seasonal discharge of the period April to September are derived every month starting from January until June. The application of the model for several catchments in Central Asia - ranging from small to the largest rivers (240 km^2 to 290,000 km^2 catchment area)– for the period 2000-2015 provided skilful forecasts for most catchments already in January with adjusted R^2 values of the best model in the range of 0.3 – 0.8. The skill of the prediction increased every following month, i.e. with reduced lead time, with adjusted R^2 values usually in the range 0.8 – 0.9 for the best and 0.7 – 0.8 for the ensemble mean in April just before the prediction period. The later forecasts in May and June improve further due to the high predictive power of the discharge in the first 2 months of the snow melt period. The improved skill of the model ensemble with decreasing lead time resulted in very narrow predictive uncertainty bands at the beginning of the snow melt period. In summary, the proposed generic automatic forecast model development tool provides robust predictions for seasonal water availability in Central Asia, which will be tested against the official forecasts in the upcoming years, with the vision of operational implementation.

[Table 1] Add additional information on the mean annual precipitation, temperature and winter snow cover area in each catchment.

Thanks for the suggestion. We will extend Table 1 as shown below.

Table 1: List of the catchments for which prediction models are derived with discharge (Q) and meteorological gauging stations used for the prediction. Note that Charvak, Andijan and Toktogul are reservoir inflows summing several tributary inflows. For the Charvak reservoir the mean temperature and precipitation data of three meteo stations located in the catchment was used. Latitude and longitudes are in decimal degrees (WGS84). Q mean seasonal is multiannual mean seasonal discharge from April to September for the period 2000-2015. Mean annual P ist the mean annual precipitation sum of the meteo station for the period 2000-2015. Mean annual T is the mean annual mean temperature of the meteo station for the period 2000-2015. Mean winter SC is the mean of the mean daily snow coverage of January to February for the period 2000-2015.

	catchment	discharge station	Q deg. lat	Q deg. long	meteo station	meteo deg. lat	meteo deg. long	meteo altitude [m]	catchment area [km^2]	Q mean seas. [m^3/s]	mean altitude [m]	mean ann. P [mm]	mean ann. T [°C]	mean winter SC [%]
1	Uba	Shemonaikha	50.620	81.880	Shemonaikha	50.620	81.880	300	9324	269.2	740	460	3.6	69.2
2	Ulba	Perevalochnaya	50.033	82.843	Oskemen	50.030	82.700	375	5080	151.4	950	483	3.8	87.7
3	Chirchik	Charvak	41.626	69.969	Chatkal	41.822	71.097	2300	10903	346.21	2575	708	5.5	97.3
					Oygaing	42.000	70.633	1620	10903					
					Pskem	41.861	70.384	2220	10903					
4	Talas	Kluchevka	42.581	71.836	Kyzyl-Adyr	42.616	71.586	1764	6663	19.62	2424	327	9.0	72.1
5	Ala-Archa	Kashka-Suu	42.650	74.500	Baytik	42.670	74.630	1579	239	8.83	3288	559	3.2	79.6
6	Chu	Kochkor	42.250	75.833	Kara Kuzhur	41.930	76.300	855	4961	34.53	2934	253	1.1	59.4
7	Chilik	Malybai	43.494	78.392	Shelek	43.597	78.249	600	3964	70.67	2603	274	11.0	74.5
8	Charyn	Sarytogai	43.553	79.293	Zhalanash	43.043	78.642	1690	7921	59.06	2260	507	6.1	82.4
9	Karadarya	Andijan	40.814	73.257	Ak-Terek	40.365	74.222	1190	11670	186.21	2663	913	9.5	82.4
10	Naryn	Toktogul	41.760	72.750	Naryn city	41.460	75.850	2040	51926	653.13	2850	374	-5.8	88.0
11	Upper Naryn	Naryn city	41.460	75.85	Tien Shan	41.910	78.210	3614	10343	168.64	3546	345	-5.8	91.0
12	Amudarya	Kerki	37.842	65.23	Kerki	37.842	65.230	237	287714	2551.02	2578	173	17.9	56.7
13	Murgab	Takhta Bazar	35.966	62.907	Takhta Bazar	35.966	62.907	354	35767	40.13	1707	217	18.2	37.5

[Section 2.1] Explain the method and purpose of the hierarchical clustering. What metrics were used to compare catchments and to establish cluster membership? The three clusters should be linked much more explicitly to subsequent discussions of predictor sets (in section 4.2).

We want to show that the different catchments show some differences in the inter-annual variability of the seasonal discharge. This is important, because if all catchment would have the same inter-annual variability, the discharge could theoretically be equally well forecasted with meteorological variables from other catchments with the same variability. This would mean in turn, that the presented ability of the approach to predict the seasonal discharge for the selection of different catchments would provide no additional evidence for the suitability of the approach as a single test case. Cluster memberships were established based on the dissimilarities of the correlation between the seasonal discharge time series of the different catchments, i.e.

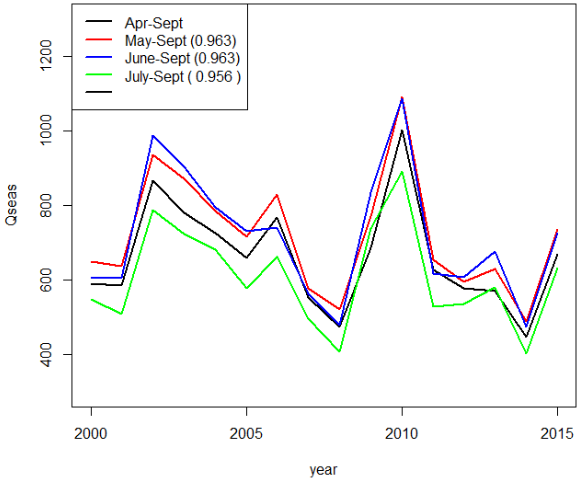
basically on the similarity/dissimilarity of the variability of the seasonal discharge as shown in Figure 2. The cluster algorithm starts by assigning a single cluster for each catchments, and starts to reduce the number of clusters by joining the most similar clusters. For the construction of the clusters the Ward algorithm was chosen, which minimizes the variability within the clusters and maximizes the variability between the clusters. This is a standard procedure. Details on this can be found in any statistical textbook.

[Section 3] How significant are evaporative losses from the catchments and how might this component of the water balance be represented within MLR models?

The catchments presented are all mountainous catchment with a cold climate and fast flowing rivers. The evaporative losses from the rivers are thus expected to be low, and do not substantially influence the seasonal discharge to be predicted. Evaporative losses from reservoirs are more likely, but all catchments in the study are without reservoirs (except the Nurek dam in Amudarya, whos influence is negligible in this large catchment. See also reply to short comment SC1), or represent inflows into reservoirs. And in general, evaporative losses are difficult to observe directly and thus to include in the MLR models. Moreover, we do not believe past evaporative losses would have a high predictive power for future discharge. Evaporation is strongly related to radiation/temperature and past temperature is already included as a potential predictor into the MLR models.

[Section 3.1] To maintain independence of the predictors, only variables up to March should be used to build models of April-September discharge. After March, the forecast period should be progressively reduced. For instance, predictors between January and April could be used to build models of May-September discharge, January to May variables for June-September discharge, and so forth. Results from models with overlapping predictors and forecast variable should be removed.

We agree, in order to guarantee independence of the predictors from the predictand this would be the appropriate procedure enabling a fair comparison of the skill of the forecasts before and during the vegetation period. But this is not the purpose of the presented study. We rather aim at providing the best possible forecasts with the given data at hand. As shown in the results and discussion, the observed discharge values (antecedent discharge predictors) from the start of the vegetation period have a high predictive power for the whole vegetation period. Therefore these should be used for the prediction, particularly when a possible application in operational forecast is considered. Besides this, the results would very likely not change much, because the seasonal discharge for April to September is highly correlated to the seasonal discharges for shorter periods. The following figure shows this exemplarily for the Naryn basin. The shorter seasonal mean discharges are very similar to the whole vegetation period April to September, and are highly correlated. The numbers in the legend show the linear correlation coefficient. All correlations are highly significant (p-values < 10^{-8}). This means that the performance of models predicting only the discharge ahead is pretty much identical to the presented performance.



Moreover, the presented approach is in line with the official forecast procedures in the Central Asian hydromet services. In order to obtain acceptance of the proposed method in the services and their use in the official forecast procedures it is advisable to follow the prescribed procedures. It is required from the Hydromet Services to issue updated (corrected) forecasts, which include the entire vegetation period (April-September), The water regulation procedures and e.g. agricultural yield estimation are traditionally based on bulk numbers for the entire period. If these procedures are not followed, the obtained results, which are better than the forecasts issued with the existing procedures, might not be implemented and come into practise, and thus a chance would be missed to bring research results into application.

[Section 3.2] More rigour is needed in testing for violations of MLR assumptions (i.e. linearity of relationships, homoscedasticity, no outliers, normally distributed and uncorrelated residuals). This could be captured in tabular format with a matrix showing which assumptions (if any) are violated in each catchment.

The general answer to this comment is: no, the discharge generation in the catchments is not linear, particular if all relevant processes are considered. This has been shown in many hydrological studies. However, this does not mean that linear models cannot be applied. In fact, runoff generation can be approximated by linear models. This has been proven by the many hydrological modelling studies based on linear concepts, e.g. linear storage models. Moreover, hydrological processes can be even better approximated on longer time scales, or on larger spatial scales. This is the basis for the still wide spread use of linear regression in (seasonal) forecast studies (Seibert et al., 2017;Delbart et al., 2015;Dixon and Wilby, 2015). Furthermore, if the processes to be described show significant non-linear features, using linear models will result in low(er) performance. Predictions and model performance cannot be improved by linear models if processes are non-linear. We thus argue, that the use of linear regression for seasonal forecasts as presented is justifiable by these general considerations, and is actually supported by the good results obtained.

However, we also tested the MLR assumptions as suggested by the reviewer in order to show that our general argument holds, i.e. that the seasonal runoff generation in Central Asia can be approximated with linear models.

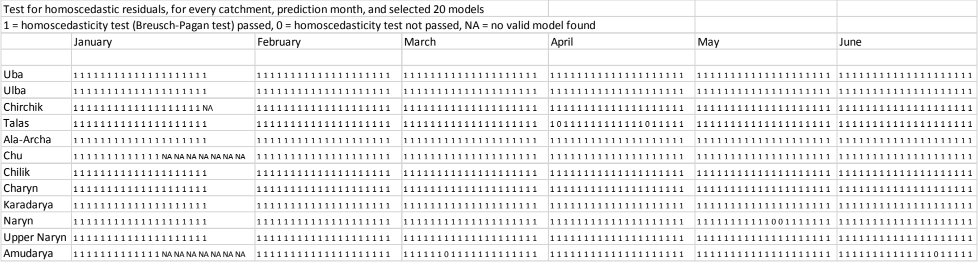
First we tested if the residuals of the models are normally distributed with the Shapiro-Wilk test for normality. Doing so, one has to bear in mind that this test is based on a sample size of maximal 16 values for each model only, so the test may not provide meaningful results. The table below shows the test result for every model, catchment, and forecast month. Note that the test was performed for a new set of models, where models with insignificant predictors were removed (cf. comment below). A “1” indicates a normal distributed residuals, “0” not normal distributed residuals. “NA” indicates that no more models with significant predictors could be found. For every forecast month up to 20 indices are given. The table shows that for most of the models (89.5%) the test was positive, i.e. the residuals are normally distributed, even for this rather low and possibly not representative sample size.

Test for normal distributed residuals, for every catchment, prediction month, and selected 20 models						
1 = normal distributed, 0 = not normal distributed, NA = no valid model found						
	January	February	March	April	May	June
Uba	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 0 1 0 1 0 1 1 0 1 1 0 1 0 0 1 0 1 1 0	1 0 0 0 0 1 1 1 0 1 1 1 0 0 0 0 1 0 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Ulba	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 0 1 1 1 0 1 1 1 1 1 0 1 1 1 1 0	1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 0 0 0 0	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Chirchik	1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1 1 1 NA	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 0	0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 1	0 0 1 1 1 0 0 1 1 1 0 1 1 1 1 1 1 1 1 1
Talas	1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 0 1 0 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 1
Ala-Archa	1 1 1 1 1 0 1 1 1 1 1 1 0 0 1 1 1 1 1 1	1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1	1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 0 1 1 0 1 0 1 1 1 1 1 1 1 0 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Chu	1 1 1 1 0 1 1 1 1 1 1 1 1 NA NA NA NA NA NA NA	1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1	1 1 1 1 0 0 0 1 1 0 1 1 1 1 1 1 0 0 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Chilik	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 0 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1	1 1 1 0 1 1 1 1 1 0 1 1 1 1 0 1 1 1 1 1	1 1 1 0 1 0 1 0 1 1 1 0 1 1 1 0 1 1 0 0	1 1 1 1 1 1 1 1 1 0 1 0 0 1 1 1 0 1 0 1
Charyn	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Karadarya	1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1 0	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 0 0 0 1 1 1 0 1 1 1 0 1 1 0	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 0 1 1 1 0 0 1 1 0 0 1 1 1 1 1 0 0
Naryn	1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 0 0 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1
Upper Naryn	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 0 1 1 1 0 0 0 1 1 1 1 1 1 1 1	0 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0	1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1	1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Amudarya	1 1 1 1 1 1 1 1 1 1 1 1 1 NA NA NA NA NA NA NA	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1 1 0	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Murgap	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1	1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 0 1 1 1 0 0 1 1 1 0 1	1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1 1	1 1 1 1 1 0 1 0 0 1 1 1 1 1 1 0 1 1 0 1

Next we tested if the residuals are independent applying a test for autocorrelation with lag 1 at significance level p = 0.05. In the table below a “0” indicates independence, a “1” dependence. It shows that 95.8% of the models have independent residuals.

Test for autocorrelated (independent) residuals, for every catchment, prediction month, and selected 20 models, lag = 1						
1 = correlated, 0 = not correlated, NA = no valid model found						
	January	February	March	April	May	June
Uba	0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 1 0	1 0 0 0 1 0 0 0 0 0 0 0 1 0 1 1 1 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Ulba	1 0 0 1 0 1 0 1 0 0 0 0 0 0 1 0 1 0 0 1	0 1 0 0 0 0 1 1 0 0 0 1 0 0 0 0 0 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	1 1 1 1 0 1 0 1 1 1 0 0 1 0 0 0 0 0 0 0
Chirchik	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 NA	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Talas	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Ala-Archa	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Chu	0 0 0 0 0 0 0 0 0 0 0 0 0 NA NA NA NA NA NA NA	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0
Chilik	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Charyn	0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
Karadarya	0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Naryn	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
Upper Naryn	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0	1 1 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 0 1 0
Amudarya	0 0 0 0 0 0 0 0 0 0 0 0 0 NA NA NA NA NA NA NA	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
Murgap	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0	1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

Furthermore we applied the Breusch-Pagan test for heteroscedasticity. This test shows that 99.5% of the models have homoscedastic residuals.



Test for homoscedastic residuals, for every catchment, prediction month, and selected 20 models

1 = homoscedasticity test (Breusch-Pagan test) passed, 0 = homoscedasticity test not passed, NA = no valid model found

	January	February	March	April	May	June
Uba	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Ulba	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Chirchik	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 NA	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Talas	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Ala-Archa	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Chu	1 1 1 1 1 1 1 1 1 1 1 1 1 NA NA NA NA NA NA NA	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Chilik	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Charyn	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Karadarya	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Naryn	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Upper Naryn	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Amudarya	1 1 1 1 1 1 1 1 1 1 1 1 1 NA NA NA NA NA NA NA	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1

In summary, we believe that the provided arguments and tests provide sufficient reason and arguments for the use of MLR models for the seasonal forecasts in Central Asia. The tables shown above can be included in an appendix to the manuscript.

[Section 3.3] Equations for the lmg algorithm should be provided, and the method of selecting predictors should be described more clearly. Significance of all model coefficients should be formally tested and any insignificant variables removed. All reported R2 values should be adjusted for sample size, and accompanied by a statement of significance. Then, only models that pass the specified level(s) of significance should be retained.

This comment refers to several section of the manuscript, not only section 3.3. The whole process of predictor selection and model fitting and model selection is described in sections 3.1 and 3.2. Section 3.3 describes the procedure of calculating the predictor importance, which is not relevant for model and predictor selection, but is rather a help for interpreting the selected models and their predictors. We answer to the different points referring to the different sections.

Section 3.1: The selection of the predictors used in the MLR models is described in this section. Additionally tables providing the selected predictors for each forecast month are given in the appendix. We actually think that this is clearly described, and do not see how to improve the description further. The reviewer comment does not provide guidance for this, while the second reviewer seems to be satisfied with our explanations. Therefor we will leave this section as it is, unless more information about what is unclear is provided.

Section 3.2: First, we did not report the significant levels in section 4 and Table 2, as we thought that it is actually obvious that models with such a high explained variance are highly significant, even for this limited sample size. We indicate the significance levels for the best models in Table 2 below, as well as the lowest significance of the selected 20 models for the mean performances.

However, we did not check the significance of the individual predictors in the models. We thank the reviewer for stressing this point. The model selection process has been modified in a way that only models with all predictors significant at $p = 0.1$ are retained. The selection of the models to be retained is still based on the PRESS value from the LOOCV. However, we weighed the PRESS by the number of years for which forecasts are available in order to reduce possible biases due to missing predictor values (i.e. reduced number of samples). This resembles a Predictive Residual Error Mean Squares. Additionally the performance of the models is now reported in terms of adjusted R^2 values, as suggested. This lead to lower performance values mainly for the early forecasts, while the high performance of the late forecasts remain very high. Additionally we added the

Mean Absolute Error MAE (relative to the mean seasonal discharge, just as the RMSE) to the performance plots in Figure 4, as requested by the second reviewer. Figure 4 is updated to the figure below, and Table 2 also reports now the adjusted R^2 values of the best LOOCV model and the mean of the selected models, where all predictors are significant at p = 0.1. In Figure 4 the green lines for the PRESS values is replaced by a brown line in order avoid red-green blindness problems, as suggested.

Table 2: Adjusted R^2-values of the best performing prediction models from the LOOCV for all catchments and prediction months. "best" indicates the single best model according to the LOOCV, "mean" indicates the mean percentage over the best 20 models according to the LOOCV. The adjusted R^2 values are associated with indicators for significance levels.

		January		**February**		**March**		**April**		**May**		**June**	
		best	*mean*	best	*mean*	best	*mean*	best	*mean*	best	*mean*	best	*mean*
1	**Uba**	0.678 ++	0.511 ++	0.824 +++	0.714 +++	0.842 +++	0.743 +++	0.811 +++	0.790 +++	0.823 +++	0.804 +++	0.959 +++	0.951 +++
2	**Ulba**	0.624 o	0.429 +	0.714 +++	0.444 +	0.781 +++	0.672 ++	0.869 +++	0.811 +++	0.943 +++	0.932 +++	0.983 +++	0.975 +++
3	**Chirchik**	0.253 ++	0.278 --	0.594 +++	0.556 ++	0.650 +++	0.593 ++	0.891 +++	0.884 +++	0.945 +++	0.941 +++	0.971 +++	0.964 +++
4	**Talas**	0.669 +++	0.408 +	0.794 +++	0.703 +++	0.808 +++	0.728 +++	0.823 +++	0.787 +++	0.886 +++	0.852 +++	0.961 +++	0.954 +++
5	**Ala-Archa**	0.393 +	0.353 o	0.597 ++	0.431 o	0.758 +++	0.524 +	0.761 +++	0.623 ++	0.739 +++	0.624 ++	0.837 +++	0.738
6	**Chu**	0.274 +	0.260 --	0.709 +++	0.440 o	0.903 +++	0.729 +++	0.680 +++	0.569 ++	0.800 +++	0.740 +++	0.887 +++	0.862 +++
7	**Chilik***	0.865 +++	0.818 ++	0.856 +++	0.787 ++	0.910 +++	0.873 +++	0.757 +++	0.770 ++	0.880 +++	0.805 +++	0.933 +++	0.821 +++
8	**Charyn**	0.643 +++	0.503 +	0.844 +++	0.786 +++	0.792 +++	0.765 +++	0.873 +++	0.810 +++	0.949 +++	0.944 +++	0.985 +++	0.975 +++
9	**Karadarya**	0.573 ++	0.449 +	0.589 +++	0.411 ++	0.880 +++	0.845 +++	0.976 +++	0.968 +++	0.977 +++	0.979 +++	0.981 +++	0.973 +++
10	**Naryn**	0.782 +++	0.679 +++	0.657 +++	0.657 +++	0.844 +++	0.800 +++	0.853 +++	0.819 +++	0.906 +++	0.887 +++	0.924 +++	0.899 +++
11	**Upper Naryn**	0.832 +++	0.810 +++	0.898 +++	0.850 +++	0.916 +++	0.897 +++	0.947 +++	0.923 +++	0.858 +++	0.847 +++	0.950 +++	0.947 +++
12	**Amudarya**	0.213 +	0.304 +	0.841 +++	0.691 +++	0.857 +++	0.840 +++	0.878 +++	0.839 +++	0.897 +++	0.876 +++	0.983 +++	0.972 +++
13	**Murgap**	0.465 ++	0.367 o	0.757 +++	0.551 +	0.802 +++	0.642 ++	0.807 +++	0.700 ++	0.970 +++	0.960 +++	0.997 +++	0.993 +++

* the performance of Chilik is not representative and comparable to the other catchments due to too many missing discharge and predictor data.

Significance p: +++ = 0.01, ++ = 0.05, + = 0.1, o = 0.2, -- = >0.2; for mean the lowest significance of the model ensemble

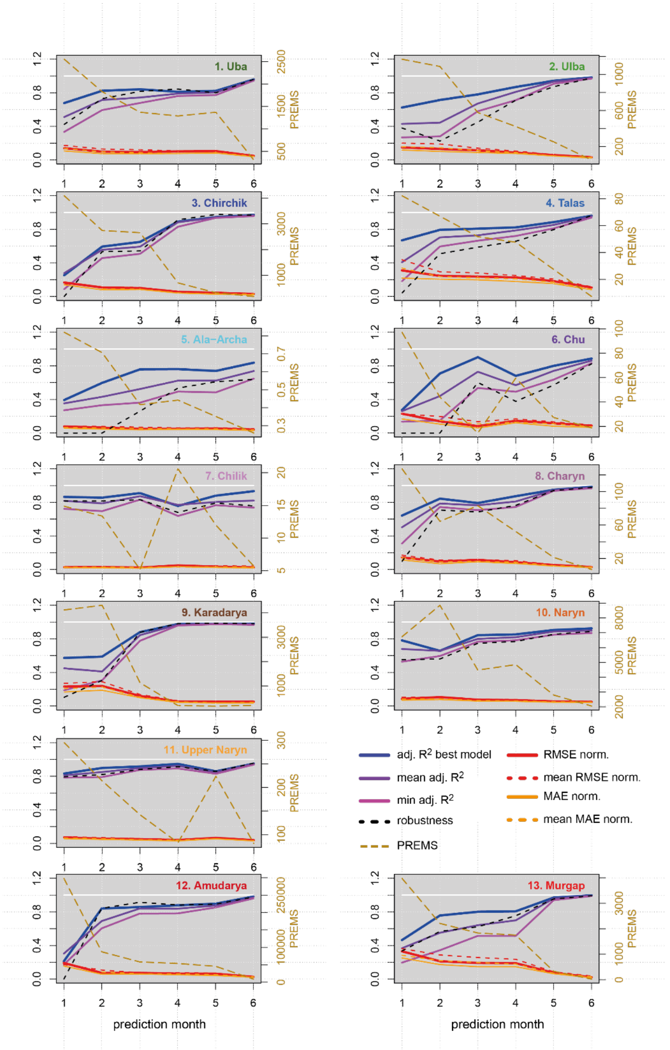


Figure 4: Performance of the prediction models for the different catchments and prediction months. Adj. R^2 best model is the adjusted R^2 of the single best LOOCV model, mean adj. R^2 is the mean adj. R^2 of the best 20 LOOCV models,

min adj. R^2 is minimum adj. R^2 of the best 20 LOOCV models, robustness is mean LOOCV-adj. R^2 of the best 20 models divided by the mean adj. R^2, RMSE/MAE norm. is the root mean squared error/mean absolute error of the single best model normalized to mean multi-annual seasonal discharge, mean RMSE/MAE norm is the mean root mean square error/mean absolute error of the best 20 LOOCV models normalized to the multi-annual seasonal discharge; PREMS is the predictive residual sum of squares (PRESS) of the single best model, divided by the number of prediction months.

Section 3.3: The analysis of predictor importance is performed after the best models are selected, i.e. it has no influence on the predictor and model selection. The *lmg* algorithm calculates how much of the overall explained variance is explained by the individual predictors of the selected models. This is principally performed by re-running the model with a single of the selected predictors and calculating the explained variance. Then the other predictors are added and the gain in explained variance is determined (->sequential R^2s). Then the importance of a predictor is given either as percentage of the overall explained variance, or as absolute fraction of explained variance. However, in this procedure the sequence of predictors influences the explained variance. In other words, it matters with which predictor the importance analysis starts. The lmg algorithm tests all predictor orderings and calculates the mean importance of every ordering in order to overcome the problem of predictor ordering in sequential R^2s. More details are given in the reference provided. We will add some sentences as above for explanation in the revised manuscript.

[Table 2] Report only adjusted R2 values for the overall best, and 20 best models. Results for forecasts issued in April, May and June should only cover the periods May-September, June-September and July-September respectively. The legend should be updated accordingly.

Adjusted R2 values are now reported in Table 2 and Figure 4 (see above). However, as already explained in an earlier answer, we would keep the current procedure, because of the high correlation of the seasonal and sub-seasonal mean discharge and for better transfer into operational practice in Central Asia.

[Section 4.1] The discussion of predictive uncertainty should acknowledge other components, including from data quality, choice of model type/ structure, choice of objective unction(s), model parameters. As noted, the uncertainty bands associated with the 20 best models reflect the number of models retained. When more stringent tests of model skill are applied (see comments on section 3.3 above), fewer models may pass. In any event, the criteria for model inclusion within the ensemble used for uncertainty estimation should be stated explicitly.

The criteria for model inclusion in the ensemble is as stated the best model performance in the cross validation, i.e. the lowest PREMS (=PRESS divided by the number of years for which forecasts can be made by the individual models) value. However, the number of models for the ensemble is set subjectively to 20. This selection is aiming at obtaining a sufficient number of models for an ensemble evaluation of the forecasts. With the newly set restriction on model selection (only models with significant predictors), a few ensembles, particularly for the January prediction have less than 20 models, because not enough models fulfilling the new selection criteria could be identified. There is actually no rule for the number of ensembles members applied. We left sufficient amount of freedom for this, in order to enable an expert selection of models by the forecasters of the Central Asian hydromet services. The forecasters have a lot of experience with their catchments, and can decide better which forecasts are valuable for them. The forecasters check every model retained for their performances (qantitatively and qualitatively), and select the models accordingly. This means that in practice less models that the 20 presented in the manuscript might be selected, or even more. Another possible rule for ensemble model selection could be to set a threshold in explained variance for the model. However, due to the high explained variances, the threshold must be very high in order to reduce the ensemble members. A fixed R^2 threshold would more likely increase the ensemble members in most cases.

We will explicitly discuss other uncertainty sources. Other uncertainty sources are, as mentioned, model structure, which is rather low given the high explained variances; data sources, which is not quantifiable, but might be high, particularly the discharge data; and performance criteria for selecting the best models. This last aspect has actually been tested, but is not included in the manuscript in order to keep the manuscript concise. Using other performance criteria as PRESS for model selection usually results in slight different selection of best models, and often in a different order of the best models. The best PRESS model is not necessarily the best cross validated R^2 model. However, as this mainly affects the ordering of the best models, the results in terms of ensemble predictions, if unweighed as presented, will remain the same. In order to illustrate this, we will some sentences in the manuscript. We could also provide tables with further performance criteria (R2, adj., R2, SSQ, MAE, central Asian performance criteria, all for cross validation and full models) for every selected model, forecast month and catchment in the appendix. This will, however, result in 13 x 6 wide tables, i.e. in quite a long appendix. We would abstain from including this bulk of data into the manuscript, unless the editor explicitly requests this.

[Section 4.3] Add a paragraph on the specific operational decisions that are already, or could be, supported by seasonal discharge forecasts in Central Asia.

A lot of management and strategic decisions are based on seasonal forecasts of water availability in CA. The main consumer of water resources in the Aral Sea basin is the agricultural sector with has one of the world's largest irrigation systems (Dukhovny and de Schutter, 2011). Very important decisions based on water availability forecasts are the planning of agricultural production crop types and water allocation through the irrigation network. Also the estimation of agricultural yield is related to water availability and is needed for country income planning that heavily depends on agricultural export in some countries.

[Conclusions] Note that seasonal forecasting of precipitation could provide useful in-formation in catchments and years with relatively little winter snowpack accumulation. Seasonal and sub-seasonal forecasts of extreme rainfall could also be important for hazard management (floods, landslides) and dam safety. Note also that the winter precipitation, summer melt situation applies in the Western U.S. too. Add a paragraph on further research opportunities.

Thanks for the suggestions. We will discuss the possibility of the forecasts for hazard management on more detail.

However, seasonal forecasts of precipitation in Central Asia are difficult and very uncertain. We have studied this in another publication (Gerlitz et al., 2016). We showed, that winter precipitation amounts are highly related to tropical and extratropical circulation modes (such as ENSO and NAO) and thus exhibit a certain degree of predictability. In contrast, summer precipitation in Central Asia is usually convective, i.e. is triggered by surface heating and associated atmospheric instability. Summer precipitation sums are composed of few single events (occasionally of high intensity) which, however, are rather randomly distributed and non-predictable.

Minor corrections and clarifications

[P1, L19] Note that seasonal forecasts can also contribute to improved dam safety.

Thanks for the hint. We will add this in the introduction.

[P1, L31] State the range of river catchment areas.

Will be included.

[P2, L7] Typo “The Central Asian region. . .”

Will be corrected.

[P2, L25] Omit “actually”.

Will be omitted.

[P3, L4] Provide full publication details for the Hydromet Services questionnaire.

The questionnaire was project internal and not published. This was a prerequisite for participation in the questionnaire. The replies to the questionnaires were rather heterogeneous and were further elaborated and specified in the dialogue with the Central Asian forecasting specialists during a workshop. It would, however, go far beyond the scope of this manuscript to compile and publish the detailed answers of the questionnaires and interviews. The content of the questionnaire would also neither improve the quality of the manuscript, nor change the focus, but rather distract the focus.

[P4, L27] Typo “catchmentss”.

Will be corrected.

[P4, L27] Note that some of the catchments are nested (i.e. not independent) such as the Upper Naryn and Naryn, so the actual sample size is smaller than 13.

This comment applies for the Naryn basin only. The catchments were nested in order to analyse if the method also works for high alpine catchment such as Upper Naryn with a high degree of glacierization. We will state this in the revised manuscript.

[P7, L12] Non sequitur – please clarify why the need for cross-validation and hierarchical clustering follows from the observation that the discharge regimes vary between catchments.

Thanks for spotting this. The sentences will be changed in "This plot indicates similar but also different inter-annual variability patterns of the different catchments. In order to distinguish between similar and different inter-annual variabilities cross-correlations of the seasonal discharges are calculated and hierarchically clustered (Figure 3).".

[P10, L20] Presumably all variables used in composites (e.g. temperature and precipitation) are normalized by their mean and variance such that they have equal weight in the MLR model?

No, the variables in the composites were simply multiplied.

[P11, L4] Provide the equation for the Predicted Residual Error Sum of Squares (PRESS). Note also that had a different objective function been selected, different sets of predictors might have emerged.

We will include the following description of PRESS:

"The PRESS residuals are defined as $e_{(i)} = |y_i - \hat{y}_{(i)}|$ where $\hat{y}_{(i)}$ is the regression estimate of y_i based on a regression equation computed leaving out the ith observation. The process is repeated for all n observations resulting in:

$PRESS = \sum_{i=1}^{n} e_{(i)}{}^2$ "

And yes, a different objective criteria might result in a different set of models, but in most cases usually in a different order of the best models. We commented on this earlier. We argue that PRESS is the most appropriate selection criteria, because a desirable forecast would reduce the residuals to a minimum in cross validation.

Moreover, we now normalized the PRESS by the number of years for which forecasts could be made by the individual models in order to avoid biases caused by missing predictor values. This resembles a Predictive Residual Error Mean of Squares, which we term PREMS.

[P11, L14] Please clarify "a set of specific models of the best models".

This refers to the option of selecting individual models by experts of the catchments, i.e. the responsible forecasters in CA. some models might have acceptable performance criteria, but the temporal dynamics might be not acceptable. Or, some models might show high performance, but have too many missing predictors resulting in a spurious good performance. Such models can be excluded from the ensemble by the forecasters.

[Figure 4] Improve legibility by removing the grey background from each panel. Avoid use of red with green lines as these will be indistinguishable for some readers.

The green line for PRESS (now PREMS) has been replaced be a beige/light brown dashed line. We would actually keep the grey background, because we believe that this supports the legibility and enhances the graphical appearance. Because the other reviewer did not comment on this, we would ask the editor to decide whether the background should be changed or not.

[Table 3] Explain how the number of "good" forecasts can be higher for the mean than for the best model in some catchments (e.g. Uba, January).

As mentioned above, the best models are selected according to PRESS. In PRESS the residuals are squared, resulting in a different order and occasionally selection of the set of best models compared to a sorting based on absolute residuals, as the Mean Absolute Error MAE and the CA performance criteria defined in equation (1). This means that the best model according to PRESS is not necessarily the best model according to the CA performance criteria. This can result in more "good" forecasts according to the CA criteria on average for the selected ensemble model compared to the best model.

[P19, L22] Please clarify "possible lack of representativeness of the time series used for the "real" variability of the seasonal discharge in Central Asia".

This refers to the limited length of the time series used in this study, which might show a different variability compared to longer time series.

[P21, L7] Please clarify the sentence "This indicates that the predictor selection. . ."

We want to express that a) for the best models similar predictors are selected, i.e. that the predictor selection is not random and follows hydrological principles of runoff generation, and b) that the procedure of predictor selection for the models avoids the selection of correlated predictors from the same group, which could be a problem if the restriction in predictor selection were not set.

We will clarify this in the revised manuscript.

[Figure 7] Ideally the presentation and discussion of the predictors would be organized by the three clusters described in section 2.1.

This is a welcome suggestion. We will include references to the clusters in the revised manuscript. Also note that Figure 7 has slightly changed because if the updated model selection (only significant predictors), and because the importance is now quantified as absolute contribution to adjusted R^2 values:

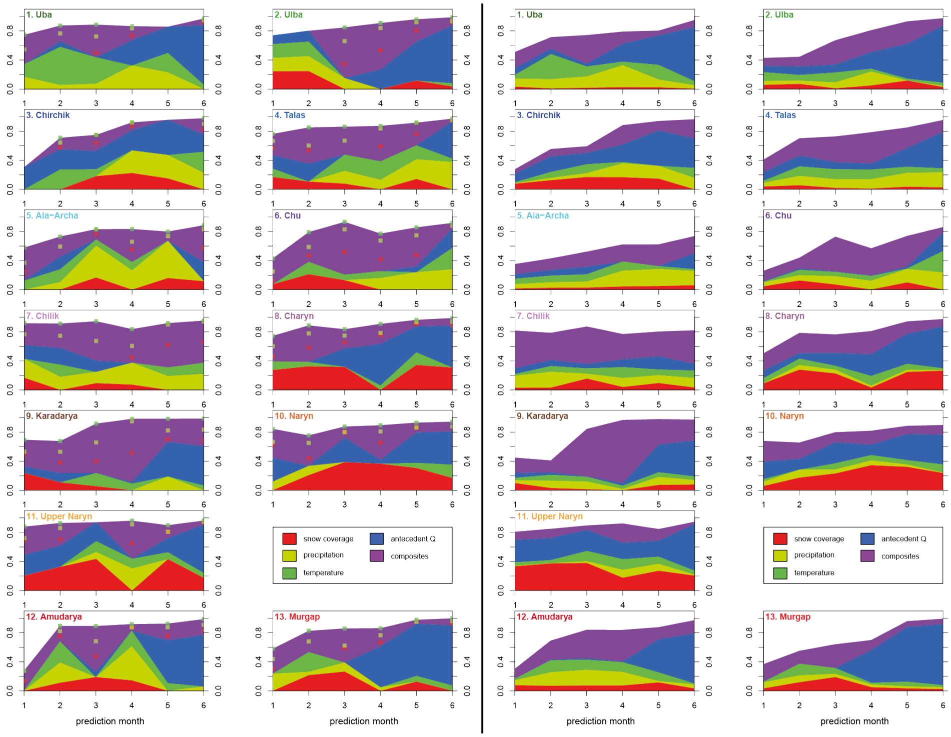


Figure 7: Importance of the predictors in the linear models as absolute contribution to the explained variance expressed as adjusted R^2 for all catchments and prediction months. Left: of the best LOOCV model; Right: on average for the best 20 LOOCV models. Squares in the left panel figures indicate the presence of the different predictors used in the composites: snow cover, precipitation and temperature, using the same colour codes as for the individual predictors.

[P24, L3] Typo “precipiutation”.

Will be corrected.

[P24, L25] Report only adjusted R2 values with accompanying significance level(s).

Will be done.

References:

Delbart, N., Dunesme, S., Lavie, E., Madelin, M., Régis, and Goma: Remote sensing of Andean mountain snow cover to forecast water discharge of Cuyo rivers Journal of Alpine Research | Revue de géographie alpine, 103, DOI : 10.4000/rga.2903 2015.

Dixon, S. G., and Wilby, R. L.: Forecasting reservoir inflows using remotely sensed precipitation estimates: a pilot study for the River Naryn, Kyrgyzstan, Hydrological Sciences Journal, 61, 1-16, 10.1080/02626667.2015.1006227, 2015.

Dukhovny, V. A., and de Schutter, J. L. G.: Water in Central Asia: Past, Present and Future, CRC Press/Balkema,Taylor & Francis Group: London, UK, 2011.

Gerlitz, L., Vorogushyn, S., Apel, H., Gafurov, A., Unger-Shayesteh, K., and Merz, B.: A statistically based seasonal precipitation forecast model with automatic predictor selection and its application to central and south Asia, Hydrol. Earth Syst. Sci., 20, 4605-4623, 10.5194/hess-20-4605-2016, 2016.

Seibert, M., Merz, B., and Apel, H.: Seasonal forecasting of hydrological drought in the Limpopo Basin: a comparison of statistical methods, Hydrol. Earth Syst. Sci., 21, 1611-1629, 10.5194/hess-21-1611-2017, 2017.

Reply to reviewer comment hess-2017-340-RC2

Heiko Apel[1], Zharkinay Abdykerimova[2], Marina Agalhanova[3], Azamat Baimaganbetov[4], Nadejda Gavrilenko[5], Lars Gerlitz[1], Olga Kalashnikova[6], Katy Unger-Shayesteh[1], Sergiy Vorogushyn[1], Abror Gafurov[1]

[1]GFZ German Research Centre for Geoscience, Section 5.4 Hydrology, Potsdam, Germany

[2]Hydro-Meteorological Service of Kyrgyzstan, Bishkek, Kyrgyzstan

[3]Hydro-Meteorological Service of Turkmenistan, Ashgabat, Turkmenistan

[4]Hydro-Meteorological Service of Kazakhstan, Almaty, Kazakhstan

[5]Hydro-Meteorological Service of Uzbekistan, Tashkent, Uzbekistan

[6]CAIAG Central Asian Institute for Applied Geoscience, Bishkek, Kyrgyzstan

Correspondence to: Heiko Apel (heiko.apel@gfz-potsdam.de)

General referee comment:

This paper proposes to use standard multiple linear regression (MLR) to predict season streamflow for 13 catchments in Central Asia. The predictors are antecedent precipitation, streamflow, temperature, and snow depth. The different combinations of predictors are tested using MLR under the framework of leave-one-out cross validation (LOOCV) and using the metric of predicted residual error sum of squares (PRESS). At the end, “the best 20 forecast models” are picked out for the prediction of future streamflow. In general, the paper is well-written and the results are clearly presented. In the meantime, there are comments for further improvements of the paper:

First of all, it is widely known that the predictability of seasonal streamflow is generally from two sources, i.e., catchment storage and future climate [Hamlet and Lettenmaier, 1999; Chiew and MacMahon, 2002; Wood et al., 2002; Schepen et al., 2012; Crochemore et al., 2017]. However, in this paper, the predictors of future climate, which can be atmospheric circulation indices and GCM/RCM outputs, are not considered at all. That is to say, this paper only accounts for the predictability from catchment storage. As a result, the forecasts as are presented in this paper are not deemed “best” and they can be further improved. The authors are encouraged to consider circulation indices in seasonal streamflow forecasting. It is noted that NOAA provides a collection of more than 30 climatic indices (https://www.esrl.noaa.gov/psd/data/climateindices/list/).

We thank the reviewer for the constructive comments. We fully agree that the predictability of seasonal streamflow depends on the information about catchment storage and future climate, particularly rainfall. However, in Central Asia much of the discharge stems from snow melt, i.e. the winter accumulation, resp. the precipitation in winter. In the Altai catchments and along the Northern rim of the Tien Shan some additional precipitation occurs during spring and early summer (March-July). This precipitation is eventually considered as observations in the late forecasts presented here. However, reliable information about the spring precipitation in advance could possibly improve the early forecasts. We actually studied the seasonal predictability of precipitation in Central Asia using NAO, ENSO and EA indices as well as automatically selected seas surface temperature regions as predictors in a preceding paper (Gerlitz et al., 2016). Although some skillful models were obtained for winter precipitation, the variability of the seasonal precipitation in Central Asia was strongly underestimated. Furthermore, summer precipitation in Central Asia is usually convective, i.e. is triggered by surface heating and associated atmospheric instability. Summer precipitation sums are composed of few single events (occasionally of high intensity) which, however, are rather randomly distributed and non-predictable. Therefor we did not include the seasonal forecasts of precipitation in the presented linear models, because no additional gain in performance can be expected. Another (practical) reason for this decision was the envisaged operational use by the CA hydromet services. Using station data as presented is fairly easy for them to include in the operational routines, while using climate indices could pose a mental as well as technical barrier for the staff in the services.

Regarding the term “best” we of course refer to the best forecast obtained with the presented approach and predictors. We do by no means imply that the presented models are the best forecast obtainable in general and will make this clear in the revised manuscript.

Second, the analysis of predictive uncertainty is too simple to be informative in this paper. It is pointed out that for ensemble and probabilistic forecasts, the attributes of reliability and skill are of key importance [Murphy, 1993, What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting]. Reliability can be diagnosed using the PIT reliability diagram or PIT histogram [e.g., Wang et al., 2009; Crochemore et al., 2017]. Meanwhile, Skill can be measured using the continuous ranked probability score (CRPS), which is for both deterministic and ensemble forecasts and is equivalent to the mean absolute error (MAE) for deterministic forecasts [Hersbach, 2000]. In addition to the illustrative plots of predictive uncertainty, the authors are encouraged to perform a comprehensive examination of forecast reliability and skill.

Many thanks for the suggestion. Because we are using deterministic models, we have now evaluated the skill in terms of the MAE and plotted it in Figure 4. The MAE is normalized to the mean seasonal discharge for each basin, just as already shown for the RMSE. The MAE skill is very similar to the RMSE, being in the range of 10%-20% for the January forecasts, and below 10% for the most important April forecast:

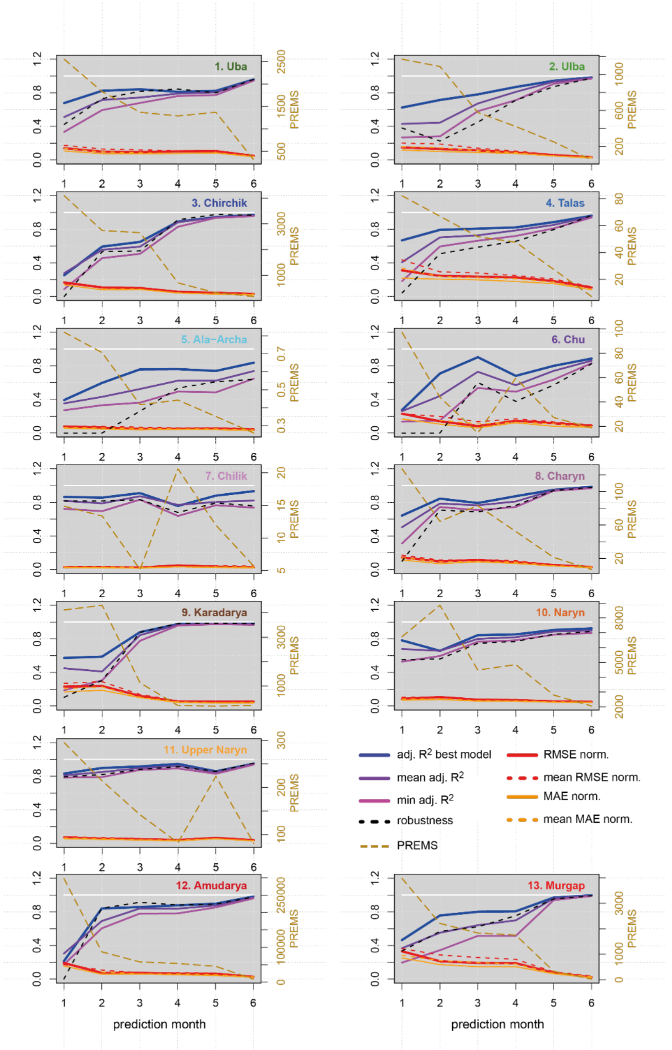


Figure 4: Performance of the prediction models for the different catchments and prediction months. Adj. R^2 best model is the adjusted R^2 of the single best LOOCV model, mean adj. R^2 is the mean adj. R^2 of the best 20 LOOCV models,

min adj. R^2 is minimum adj. R^2 of the best 20 LOOCV models, robustness is mean LOOCV-adj. R^2 of the best 20 models divided by the mean adj. R^2, RMSE/MAE norm. is the root mean squared error/mean absolute error of the single best model normalized to mean multi-annual seasonal discharge, mean RMSE/MAE norm is the mean root mean square error/mean absolute error of the best 20 LOOCV models normalized to the multi-annual seasonal discharge; PREMS is the predictive residual sum of squares (PRESS) of the single best model, divided by the number of prediction months.

We also evaluated the reliability by means of PIT diagrams, as suggested. The plot below shows the PIT diagrams for every catchment and all forecast months using the prediction of the selected ensemble models. The PIT diagrams show that the model ensemble predictions are in most cases close to the 1:1 line, i.e. provide reliable forecasts. However, in some cases the predictive uncertainty is under-estimated, i.e. the predictive uncertainty bands presented in Figure 6 are too narrow. We further calculated a PIT score as the area between the PIT curve and the 1:1 line as a summarizing indicator for the reliability. The theoretically least reliable model has a score of 0.5, a perfect model a score of 0. The highest score, i.e. the lowest reliability, of all models is 0.2, with the majority of the models being in the range of 0.07-0.15. Interpreting the scores with the curves in the PIT diagram it can be stated that the reliability of the models is good for PIT scores <= 0.1. For higher scores the predictive uncertainty is likely to be underestimated. We will include this analysis in the revised manuscript as suggested by the reviewer, and provide the PIT scores as guidelines for the interpretation of the predictive uncertainty bounds.

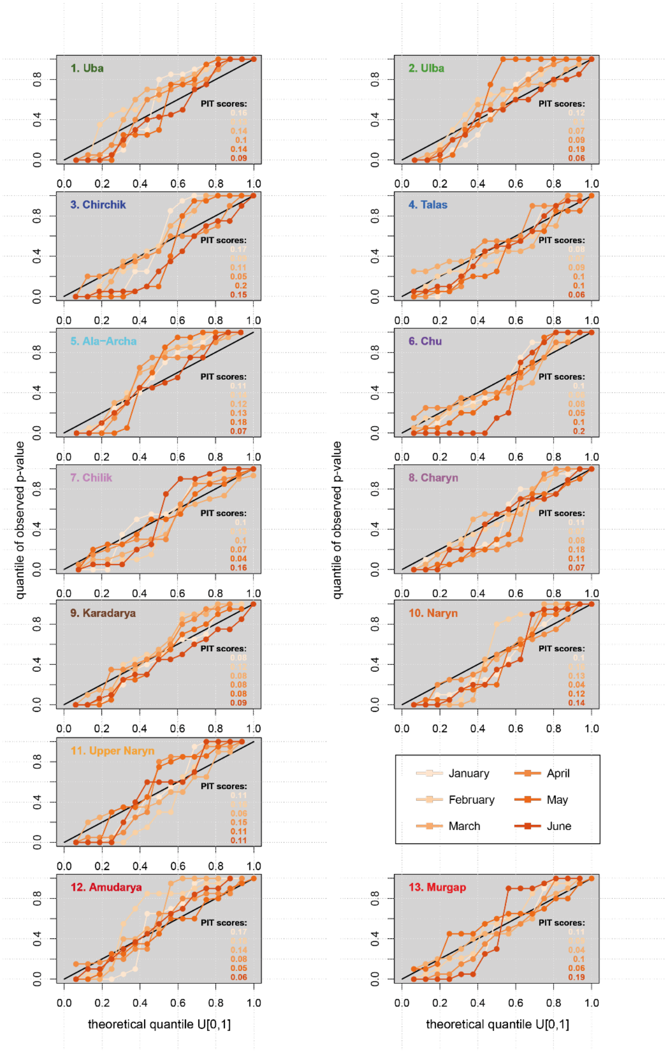
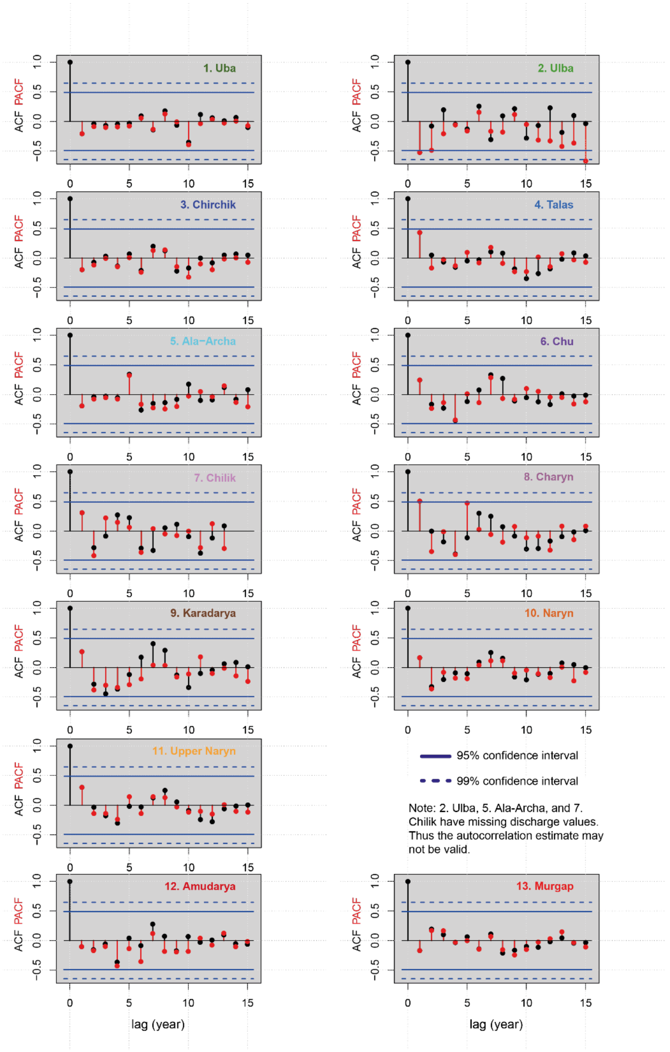

1. Uba
PIT scores:
0.16
0.13
0.14
0.1
0.14
0.09
2. Ulba
PIT scores:
0.12
0.1
0.07
0.09
0.19
0.06
3. Chirchik
PIT scores:
0.17
0.09
0.11
0.05
0.2
0.15
4. Talas
PIT scores:
0.08
0.07
0.09
0.1
0.1
0.06
5. Ala−Archa
PIT scores:
0.11
0.14
0.12
0.13
0.18
0.07
6. Chu
PIT scores:
0.1
0.06
0.08
0.05
0.1
0.2
7. Chilik
PIT scores:
0.1
0.13
0.1
0.07
0.04
0.16
8. Charyn
PIT scores:
0.11
0.07
0.08
0.18
0.11
0.07
9. Karadarya
PIT scores:
0.08
0.12
0.08
0.08
0.08
0.09
10. Naryn
PIT scores:
0.1
0.16
0.13
0.04
0.12
0.14
11. Upper Naryn
PIT scores:
0.11
0.16
0.06
0.15
0.11
0.11
12. Amudarya
PIT scores:
0.17
0.18
0.14
0.08
0.05
0.06
13. Murgap
PIT scores:
0.11
0.09
0.04
0.1
0.06
0.19
January
February
March
April
May
June
quantile of observed p-value
quantile of observed p-value
0.0
0.4
0.8
0.0
0.2
0.4
0.6
0.8
1.0
theoretical quantile U[0,1]
theoretical quantile U[0,1]

Figure: PIT reliability diagrams for every catchment and forecast month. The PIT score is calculated as the area between the reliability plots and the 1:1 line as suggested in Renard et al. (2010). The lower the PIT score, the higher the reliability. The least reliability score is 0.5, the best 0.

There are also some minor comments:

1. As for LOOCV, it can lead to artificial over-estimation of forecast skill if the streamflow series exhibit strong auto-correlation. It is worthwhile to check the serial autocorrelation of streamflow. Or, a more rigorous leave-five-years-out cross validation (L5OCV) ought to be applied.

We checked the autocorrelation and partial autocorrelation of the streamflow time series and plotted it in the figure below. Hardly any autocorrelation at p = 0.05 could be detected. Only for 2. Ulba the partial autocorrelation shows some autocorrelation for lag 1 and 2 just above p = 0.05. But in summary for all catchments, it can be stated that autocorrelation does not exist in the discharge time series, and thus the proposed LOOCV is an appropriate validation method. We propose to include the figure below in the appendix of the revised manuscript and include the statement above in the text.



Note: 2. Ulba, 5. Ala-Archa, and 7. Chilik have missing discharge values. Thus the autocorrelation estimate may not be valid.

2. In terms of predictors of catchment storage, the use of multi-monthly means as the predictor values is sensible.

Thanks for the supporting comment.

3. The paper suggests to use the “the best 20 forecast models”. This setting is empirical and it is rare in peer studies. Please clarify why.

We commented on this already in the reply to reviewer 1, thus we quote the reply here:

The number of models for the ensemble is set subjectively to 20. This selection is aiming at obtaining a sufficient number of models for an ensemble evaluation of the forecasts. With the newly set restriction on model selection (only models with significant predictors), a few ensembles, particularly for the January prediction have less than 20 models, because not enough models fulfilling the new selection criteria could be identified. There is actually no rule for the number of ensembles members applied. We left sufficient amount of freedom for this, in order to enable an expert selection of models by the forecasters of the Central Asian hydromet services. The forecasters have a lot of experience with their catchments, and can decide better which forecasts are valuable for them. The forecasters check every model retained for their performances (quantitatively and qualitatively), and select the models accordingly. This means that in practice fewer models than the 20 presented in the manuscript might be selected, or even more.

Another possible rule for ensemble model selection could be a defined threshold of R^2 for the model. However, due to the high explained variances, the threshold must be very high in order to reduce the number of ensemble members. A fixed R^2 threshold would more likely increase the ensemble members in most cases. The selection of the threshold level would also be subjective.

References:

Gerlitz, L., Vorogushyn, S., Apel, H., Gafurov, A., Unger-Shayesteh, K., and Merz, B.: A statistically based seasonal precipitation forecast model with automatic predictor selection and its application to central and south Asia, Hydrol. Earth Syst. Sci., 20, 4605-4623, 10.5194/hess-20-4605-2016, 2016.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resources Research, 46, 10.1029/2009WR008328, 2010.

Reply to short comment hess-2017-340-SC1 by Samuel G. Dixon

Heiko Apel[1], Zharkinay Abdykerimova[2], Marina Agalhanova[3], Azamat Baimaganbetov[4], Nadejda Gavrilenko[5], Lars Gerlitz[1], Olga Kalashnikova[6], Katy Unger-Shayesteh[1], Sergiy Vorogushyn[1], Abror Gafurov[1]

[1]GFZ German Research Centre for Geoscience, Section 5.4 Hydrology, Potsdam, Germany

[2]Hydro-Meteorological Service of Kyrgyzstan, Bishkek, Kyrgyzstan

[3]Hydro-Meteorological Service of Turkmenistan, Ashgabat, Turkmenistan

[4]Hydro-Meteorological Service of Kazakhstan, Almaty, Kazakhstan

[5]Hydro-Meteorological Service of Uzbekistan, Tashkent, Uzbekistan

[6]CAIAG Central Asian Institute for Applied Geoscience, Bishkek, Kyrgyzstan

Correspondence to: Heiko Apel (heiko.apel@gfz-potsdam.de)

Short comment:

Figure 1: Some gauges are located downstream of impoundments (e.g. catchment 12, Amu Darya). Are the data used corrected for management of upstream reservoirs or does management impact the flow record? A figure showing the annual regime could help to depict whether flows are natural or managed.

We used the discharge data provided by the hydromet services, which are not corrected for reservoir management. However, except the large Amudarya catchment all the gauges are located upstream of reservoirs, thus the discharge is not regulated. Within the Amudarya catchment the Nurek dam exists in the Vakhsh river, which is the right headwater tributary forming the Amudarya at the conjunction with the Panj river. The catchment area of the Vakhsh river is 31,415 km² at the outlet. This is about 11% of the whole Amudarya catchment at gauge Kerky. Because the Nurek dam is located upstream of the conjunction, the Amudarya catchment area affected by the dam is less than 10%. Assuming further that the reservoir can manage only a fraction of the total discharge of the Vakhsh river, and that the dynamics of the water retention are further buffered by the seasonal mean discharge spanning six months, it can be assumed that the regulating effect of the Nurek dam on the overall seasonal discharge is rather low. Additionally, the dam is operational since 1980, therefore a discontinuity in the time series 2000-2015 can be ruled out. We thus argue that the anthropogenic influence of the seasonal discharge time series of the Amudarya is negligible for the presented study. We will point this out in the revised manuscript.

Table 2: Adjusted R2 values may be more suitable to report due to small sample sizes

As mentioned in the reply to the comments of reviewer 1, we report adjusted R^2 values throughout the revised manuscript. The adjusted R^2 values decrease (compared to the R^2 values) for the early forecasts, while they remain high for the late forecasts.

Table 2: Model performance could be benchmarked against the long term average or persistence forecast to quantify additional skill provided by MLR models.

This is implicitly done in the R^2 values. The coefficient of determination R^2 (which is synonymous to explained variance and Nash-Sutcliffe efficiency) benchmark the squared residuals, i.e. the observations minus simulated or forecasted discharges, against the squared differences between observations and the mean observed discharge. In other words, R^2 values benchmark the model against the most simple model, which is the mean of the observed time series. Any R^2 value above 0 thus indicates an improvement compared to using the mean as predictor. Schaefli and Gupta (2007) nicely illustrate this in their paper about the value of Nash-Sutcliffe efficiency. Therefor we don't see any gain in reporting the mean discharge as reference.

General: Winter hydropower production is also a key use of water in the region as well as irrigation provision. Comment might be made as to whether these models could be useful for hydropower planning as well as summer irrigation demands.

The models are valuable for any planning concerning the use of water resources. This also includes the hydropower generation. Reliable forecasts support the reservoir management by planning the release, resp. the storage of water in the reservoirs for the winter season. However, for reservoir management also international treaties between the riparian states need to be considered. The demands of the upstream and downstream countries are often quite opposite, which is the core problem of water management in Central Asia.

General: The inclusion of local stakeholders in the authorship adds significant insight into the paper. This could be enhanced via the authors commenting on how the forecasts presented here facilitate improved water management in the region, possibly providing examples of better decisions made possible by the forecasts. Furthermore, insight could be provided regarding if the forecasts produced here fulfil the requirements of hydromet agencies, or if there are any specific areas in which the models do not perform satisfactorily requiring further research.

This is an interesting suggestion. The matter is, however, very complex. A detailed analysis would certainly by beyond the scope of this manuscript. Additionally there is the time problem. An encompassing assessment of the model performance, advantages and deficits can only be done after the forecasted vegetation season. i.e. after September. Considering the time required for collecting data, evaluating the forecasts, and the experiences and acceptance in the different hydromet services would likely take several months, certainly exceeding the time schedule for this manuscript. However, this suggestion is very welcome and we will consider to collect and summarize the experience with the models in operational forecasting and publish them additionally, maybe as a short comment.

Minor corrections

P6, L11: Typo - capitalised while

P7, L3: States "continuous time series for all data and stations were available" when later it is stated that there is some missing data (e.g. Figure 2)

P11, L27: “Figures presented in 4.3” – should this be 4.2?

P19, L19 and P21, L12: Catchment 9 is referred to as Andijan rather than Karadarya.

Figure 7: Possibly label x-axis as Jan, Feb, etc. rather than 1-6 to ease interpretation

General: Inconsistent spelling of Murgab/Murgap, e.g. Table 1 and Figure 1

Thanks for spotting these errors. We will correct them in the revised manuscript. However, we would keep the x-axis labels as 1-6, in order to avoid overloading the figures. Printed on A4 paper they are already very dense.

References:

Schaefli, B., and Gupta, H. V.: Do Nash values have value?, Hydrological Processes, 21, 2075-2080, 2007.

Revised manuscript with changes marked

Statistical forecast of seasonal discharge in Central Asia for water resources management: development of a generic linear modelling tool for operational use

Heiko Apel[1], Zharkinay Abdykerimova[2], Marina Agalhanova[3], Azamat Baimaganbetov[4], Nadejda Gavrilenko[5], Lars Gerlitz[1], Olga Kalashnikova[6], Katy Unger-Shayesteh[1], Sergiy Vorogushyn[1], Abror Gafurov[1]

[1]GFZ German Research Centre for Geoscience, Section 5.4 Hydrology, Potsdam, Germany

[2]Hydro-Meteorological Service of Kyrgyzstan, Bishkek, Kyrgyzstan

[3]Hydro-Meteorological Service of Turkmenistan, Ashgabat, Turkmenistan

[4]Hydro-Meteorological Service of Kazakhstan, Almaty, Kazakhstan

[5]Hydro-Meteorological Service of Uzbekistan, Tashkent, Uzbekistan

[6]CAIAG Central Asian Institute for Applied Geoscience, Bishkek, Kyrgyzstan

Correspondence to: Heiko Apel (heiko.apel@gfz-potsdam.de)

Abstract. The semi-arid regions of Central Asia crucially depend on the water resources supplied by the mountainous areas of the Tien Shan, Pamir and Altai mountains. During the summer months the snow and glacier melt dominated river discharge originating in the mountains provides the main water resource available for agricultural production, but also for storage in reservoirs for energy generation during the winter months. Thus a reliable seasonal forecast of the water resources is crucial for a sustainable management and planning of water resources. In fact, seasonal forecasts are mandatory tasks of all national hydro-meteorological services in the region. In order to support the operational seasonal forecast procedures of hydro-meteorological services, this study aims at the development of a generic tool for deriving statistical forecast models of seasonal river discharge. The generic model is kept as simple as possible in order to be driven by available meteorological and hydrological data, and be applicable for all catchments in the region. As snowmelt dominates summer runoff, the main meteorological predictors for the forecast models are monthly values of winter precipitation and temperature, satellite based snow cover data and antecedent discharge. This basic predictor set was further extended by multi-monthly means of the individual predictors, as well as composites of the predictors. Forecast models are derived based on these predictors as linear combinations of up to 3 or 4 predictors. A user selectable number of best models is extracted automatically by the developed model fitting algorithm, which includes a test for robustness by a leave-one-out cross validation. Based on the cross validation the predictive uncertainty was quantified for every prediction model. Forecasts of the mean seasonal discharge of the period April to September are derived every month starting from January until June. The application of the model for several catchments in Central Asia - ranging from small to the largest rivers (240 km^2 to 290,000 km^2 catchment area) – for the period 2000-2015 provided skilful forecasts for most catchments already in January~~.~~ with adjusted R^2 values of the best model in the range of 0.3 – 0.8. The skill of the prediction increased every following month, i.e. with reduced lead time, with adjusted R^2 values ~~often~~usually in the range 0.8 – 0.9 for the best and 0.7 – 0.8 for the ensemble mean in April just before the prediction period. The later forecasts in May and June improve further due to the high predictive power of the discharge in the first 2 months of the snow melt period. The improved skill of the model ensemble with decreasing lead time resulted in very narrow predictive uncertainty bands at the beginning of the snow melt period. In summary, the proposed generic automatic forecast model development tool provides robust predictions for seasonal water availability in Central Asia, which will be tested against the official forecasts in the upcoming years, with the vision of operational implementation.

Formatiert: Tabstopps: Nicht an 11.5 cm

1 Introduction

The Central Asian region encompassing the five countries Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan and Uzbekistan as well as northern parts of Afghanistan and north-western regions of China is characterized by the presence of two major mountain systems Tien Shan and Pamir drained by a number of endorheic river systems such as Amudarya, Syrdarya, Ili, Tarim and a few smaller ones. The Central Asian river basins are characterized by the semi-arid climate with strong seasonal variation of precipitation. Most precipitation falls as snow during winter and spring months in Western and Northern Tien Shan (Aizen et al., 1995, 1996;Sorg et al., 2012). In contrast, Central and Eastern Tien Shan receive their largest precipitation input during the summer months. The Pamir mountains receive the highest portion of precipitation during winter and spring months with minimum in summer (Schiemann et al., 2008;Sorg et al., 2012).

Precipitation also exhibits a high spatial variation, with e.g. less than 50 mm/year in the desert areas of Tarim and around 100 mm/year on leeward slopes of Central Pamir to more than 1000 mm/year in the mountain regions exposed to the westerly air flows being a major moisture source in the region (Aizen et al., 1996;Bothe et al., 2012;Hagg et al., 2013;Schiemann et al., 2008). This makes Tien Shan and Pamir Mountains the regional 'water towers', with snow melt to be the dominant water source during spring and early summer months. During summer, glacier melt and liquid precipitation gain importance depending on the basin location and degree of glacierisation ~~(Aizen et al., 1996) .~~(Aizen et al., 1996). The Tien Shan and Pamir mountains exhibit particularly high relative water yield compared to the lowland parts of these catchments ~~(Viviroli et al., 2007).~~(Viviroli et al., 2007). Related to the economic water demands in the lowland plains primarily for irrigated agriculture, the Tien Shan and Pamir mountains are among the most important contributors of stream water worldwide ~~(Viviroli et al., 2007).).~~(Viviroli et al., 2007). They are also among those river basins with the highest share of glacier melt water in summer, particularly in drought years (Pritchard, 2017). Within the Aral Sea basin, to which the Amudarya and Syrdarya rivers drain, the ~~actually~~ irrigated area amounts to approximately 8.2-8.4 million ha (Conrad et al., 2016;FAO, 2013) Additionally, considerable irrigation areas are located in the Aksu/Tarim basin, where agricultural land doubled in the period 1989-2011 and land use for cotton production increased even 6-fold ~~(Feike et al., 2015)~~. Irrigated agriculture in Central Asia (CA) is mainly fed by the stream water diversion with only small portion of groundwater withdrawal (FAO, 2013;Siebert et al., 2010). Hence, reliable prediction of seasonal runoff during vegetation period (April – September) is crucial for agricultural planning~~,~~ and yield estimation ~~and~~ in the low lying countries in the Aral Sea basin, as well as for the management of reservoir capacities including dam safety operations in the upper parts of the catchments. Seasonal forecasts are one of the major responsibilities of the state ~~hydrometeorological~~hydro-meteorological (hydromet) services of the Central Asian countries and are regularly released starting from January till June with the primary forecast issued end of March – beginning of April for the upcoming 6-months period. In some post-soviet countries, these forecasts are typically developed based on the empirical relationships for individual basins relating precipitation, temperature and snow depth/SWE records to seasonal discharge, partly available only in analogue form as look-up tables or graphs (Hydromet Services, unpublished questionnaire survey undertaken within the CAWa project). Particularly, point measurements of snow depth and/or snow water equivalent (SWE), which have been carried out by helicopter flights or footpath surveys in mountain regions in the past decades, are costly or not feasible due to access problems nowadays. Other ~~hydromet services~~Hydromet Services apply the hydrological forecast model AISHF ~~(Agaltseva et al., 1997)~~ developed at the Uzbek ~~Hydrometeorological~~Hydro-meteorological Service (Uzhydromet), which computes discharge hydrographs by considering temperature, snow accumulation and melt. Snow pack is accumulated in winter and temperature and precipitation are taken from an analogous year to drive the model in the forecast mode. Hydrometeorological services rely on the available meteorological and hydrological data acquired by the network of climate and discharge stations, which, however, strongly diminished during the 1990s (Unger-Shayesteh et al., 2013) and slowly recovers nowadays, partly with substantial international support (e.g. ~~Schöne et al. (2013)~~; CAHMP Programme by World Bank; previous programmes by SDC and USAID). Hence, to fulfill their task, hydrometeorological services need the timely to near real-time data and simple methodologies capable of utilizing available information.

~~Schär et al. (2004)~~ showed the potential of the ERA-15 precipitation data from December-April period to explain about 85% of the seasonal runoff variability in May-September in the large-scale Syrdarya river basin. The explained variance for the Amudarya River amounted, however, to only about 25%, presumably due to poor precipitation modelling in the ERA dataset, strong influence of glacier melt and water abstraction for irrigation purposes. ~~Similarly, Barlow and Tippett (2008)~~Similarly, Barlow and Tippett (2008) explored the predictive power of NCEP-NCAR cold-season (November-March) precipitation for warm-season (April-August) discharge forecast using canonical correlation analysis. Though for some of the 24 Central Asian gauges, no skillful prediction could be achieved, for a few catchments 20 to 50% explained variance could be attained. ~~Archer and Fowler (2008)~~ utilized temperature and discharge records additionally to precipitation for spring and summer seasonal flow forecast on the southern slopes of Himalaya in northern Pakistan using multiple linear regression models. Despite good predictions of spring and early summer flows, late summer discharges were poorly forecasted due to the strong influence of summer monsoon. Recently, Dixon and Wilby (2015) demonstrated the skill of a linear regression model for the Naryn basin, Kyrgyzstan, based on TRMM precipitation from October-March to explain 65% of the seasonal flow variance in the vegetation period. The authors selected specific TRMM pixels in the catchments showing the highest correlation to seasonal discharge. They also explored the predictive skills of multiple linear regression models additionally including temperature and antecedent discharge and testing different lead times from one to three months. They showed that forecasts based on multiple linear regression models are always superior to zero order forecasts, i.e. the mean flow.

The fact that substantial snow accumulation in Central Asian mountain regions during the winter and spring months significantly governs runoff in the vegetation period can be effectively utilized for seasonal forecasts. For a similar climatic setting, ~~Pal et al. (2013)~~ included the measurements of snow water equivalent at point locations into multiple linear regression models along with precipitation, antecedent discharge and temperature-based predictors. Linear models with multiple predictor combinations achieved skilful forecasts of the spring (March-June/April-June) seasonal flow in northern India on the southern Himalaya slopes. Point snow measurements are, however, rarely available and remotely sensed snow cover extent can provide a viable alternative. Based on the monitored snow cover extent, e.g. using optical satellite imagery, and additionally considering temperature and precipitation to implicitly approximate snow water equivalent (SWE) a solid basis for seasonal discharge forecast can be formed. The MODIS snow cover product was shown to deliver high accuracy for the Central Asian region ~~(Gafurov et al., 2013)~~. Methodologies to remove cloud obstruction of optical imagery have matured over the past decade (Gafurov and Bárdossy, 2009;Gafurov et al., 2016) and tools for the automated image acquisition and processing reached the operational level (Gafurov et al., 2016). MODIS snow cover data was used for runoff forecast in the Argentinian Andes in the high-water season (September-April), though no cloud elimination algorithms were applied (Delbart et al., 2015). Snow cover in September-October could explain about 60% of the high-water season discharge variance. However, no skilful forecast with lead times greater than zero were possible. ~~Rosenberg et al. (2011)~~ proposed a hybrid (statistical – hydrological model) framework for seasonal flow prediction in Californian catchments using accumulated precipitation in antecedent period and SWE modelled by a distributed hydrological model. These two predictors were linked to seasonal discharge by principal component and Z-score regression ~~(Rosenberg et al., 2011)~~. The hybrid approach was found comparable and in some cases superior to a purely statistical approach, however, at the cost of effort for hydrological simulation of the SWE dynamics.

Based on the finding of the studies listed above, we propose a simple methodology for the operational forecast of seasonal runoff for the vegetation period (April-September) for all Central Asian catchments, which areas range over three orders of magnitude. The method is based on multiple linear regression models with automatic predictor selection, whereas the predictors are based on the readily available precipitation, temperature and discharge gauge records and additionally leverage by the operationally processed cloud-free MODIS snow cover product. It is argued, that in linear modelling the use of meteorological data from a single gauging station for a large catchment is justified, as long as the variability of the station records are representative for the whole catchment. We demonstrate the model predictive skill and robustness in a cross-validation and discuss the relative ~~significance~~importance of the automatically selected predictors depending on the prediction lead time.

2 Study sites and data

For the testing of the forecast models 13 ~~catchmentss~~catchments were selected. They cover a wide range of geographical regions, ranging from catchments along the western slopes of the Altai mountains in Eastern Kazakhstan (Uba, Ulba), over catchments at the western and northern rim of the Tien-Shan (Chirchik, Talas, Ala-Archa, Chu, Chilik, Charyn) and central Tien-Shan (Karadarya~~ (Andijan), Naryn) mountains (cf. Aizen et al., 2007)~~, Naryn) mountains (cf. Aizen et al., 2007), to the northern and central Pamir (Amudarya) and the northern Hindukush (Murgap). The size of the catchments varies over three orders of magnitude from 239 km^2 to 288,000 km^2. Figure 1 provides an overview of the location and size of the catchments, while Table 1 additionally lists the discharge and meteorological gauging stations used for the seasonal flow forecast. Note that the Naryn catchments are nested. The Upper Naryn represents a high alpine catchment and is the headwater catchment of the larger Naryn catchment draining into the Toktogul reservoir. This separation was undertaken in order to test the proposed method also for a high alpine catchment with a comparably high degree of glacialisation. The wide range of catchment locations, climatic conditions and sizes enable a testing of the proposed forecast models under different boundary conditions, and thus provides an indication of the applicability, robustness and transferability of the approach.

The catchment boundaries are derived to map the catchment area draining to the selected discharge stations. For the meteorological data (temperature and precipitation) meteorological stations run by the individual ~~hydrometeorological services~~Hydromet Services were selected. Ideally those are located in the catchment area and have sufficient data coverage of at least 16 years (starting in 2000 in order to be consistent with the MODIS temporal coverage). However, in some catchments meteorological stations fulfilling these criteria were not available. For those catchments stations nearby were selected for the prediction.

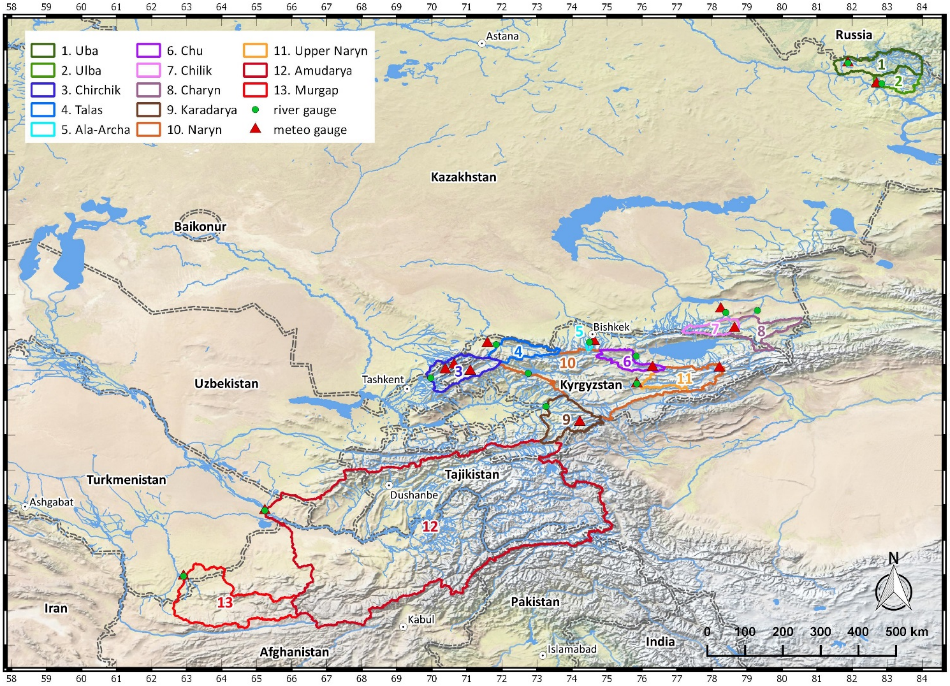


Figure 1: Overview of the catchments for which prediction models were established, with locations of discharge and meteorological gauging stations used (coordinates in latitude/longitude).

Table 1: List of the catchments for which prediction models are derived with discharge (Q) and meteorological gauging stations used for the prediction. Note that Charvak, Andijan and Toktogul are reservoir inflows summing several tributary inflows. For the Charvak reservoir the mean temperature and precipitation data of three meteo stations located in the catchment was used. Latitude and longitudes are in decimal degrees (WGS84). Q mean seasonal is multiannual mean seasonal discharge from April to September for the period 2000-2015. Mean annual P ist the mean annual precipitation sum of the meteo station for the period 2000-2015. Mean annual T is the mean annual mean temperature of the meteo station for the period 2000-2015. Mean winter snow cover (SC) is the mean of the mean daily snow coverage from January to February for the period 2000-2015.

	catchment	discharge station	Q deg. lat	Q deg. long	meteo station	meteo deg. lat	meteo deg. long	meteo altitude [m]	catchment area [km^2]	Q mean ~~seasonal~~seas. [m^3/s]	mean altitude [m]	mean ann. P [mm]	mean ann. T [°C]	mean winter SC [%]
1	Uba	Shemonaikha	50.620	81.880	Shemonaikha	50.620	81.880	300	9324	269.2	740	460	3.6	69.2

Eingefügte Zellen

Eingefügte Zellen

Eingefügte Zellen

Formatiert: Schriftart: 8 Pt.

Formatierte Tabelle

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

2	Ulba	Perevalochnaya	50.033	82.843	Oskemen	50.030	82.700	375	5080	151.4	950	483	3.8	87.7
3	Chirchik	Charvak	41.626	69.969	Chatkal	41.822	71.097	2300	10903	346.21	2575	708	5.5	97.3
					Oygaing	42.000	70.633	1620	10903					
					Pskem	41.861	70.384	2220	10903					
4	Talas	Kluchevka	42.581	71.836	Kyzyl-Adyr	42.616	71.586	1764	6663	19.62	2424	327	9.0	72.1
5	Ala-Archa	Kashka-Suu	42.650	74.500	Baytik	42.670	74.630	1579	239	8.83	3288	559	3.2	79.6
6	Chu	Kochkor	42.250	75.833	Kara Kuzhur	41.930	76.300	855	4961	34.53	2934	253	1.1	59.4
7	Chilik	Malybai	43.494	78.392	Shelek	43.597	78.249	600	3964	70.67	2603	274	11.0	74.5
8	Charyn	Sarytogai	43.553	79.293	Zhalanash	43.043	78.642	1690	7921	59.06	2260	507	6.1	82.4
9	Karadarya	Andijan	40.814	73.257	Ak-Terek	40.365	74.222	1190	11670	186.21	2663	913	9.5	82.4
10	Naryn	Toktogul	41.760	72.750	Naryn city	41.460	75.850	2040	51926	653.13	2850	374	4.4	88.0
11	Upper Naryn	Naryn city	41.460	75.85	Tien Shan	41.910	78.210	3614	10343	168.64	3546	345	-5.8	91.0
12	Amudarya	Kerki	37.842	65.23	Kerki	37.842	65.230	237	287714	2551.02	2578	173	17.9	56.7
13	~~Murgab~~Murgap	Takhta Bazar	35.966	62.907	Takhta Bazar	35.966	62.907	354	35767	40.13	1707	217	18.2	37.5

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

Formatiert: Schriftart: 8 Pt.

For both discharge and meteorological data monthly values were obtained for the stations listed in Table 1, i.e. monthly mean discharges, monthly mean temperatures and monthly precipitation sums. For the presented study meteorological station data was used, because of the operational availability to the CA ~~hydromet services.~~Hydromet Services. Gridded re-analysis products like ERA-Interim typically have a latency of weeks to months, and thus cannot be used for operational forecasts to fulfil the mandatory regulations ~~While~~while station temperature and precipitation data are likely not representative for basin average values, it is assumed that the variability of the catchment averages and the station data is similar. This, in turn, enables the use of the station data in the statistical forecast using multiple linear regressions.

In addition to the station data, mean monthly snow coverages for the individual catchments were calculated using daily snow cover data derived by the MODSNOW-Tool (Gafurov and Bárdossy, 2009;Gafurov et al., 2016). MODSNOW uses the MODIS satellite snow cover product and applies a sophisticated cloud elimination algorithm (Gafurov and Bárdossy, 2009;Gafurov et al., 2016) to obtain cloud free daily snow cover images. The MODSNOW-Tool runs operationally in most of the CA ~~hydromet services~~Hydromet Services, thus enabling the use of snow cover information for operational forecasts.

Due to the use of MODIS snow cover, which is available since March 2000, the time series of the data used for the construction of the forecast models had to be limited to post-2000. The time period for the model development and testing was thus set to 2000 – 2015, for which mostly complete continuous time series for all data and stations were available.

The seasonal discharge, i.e. the predictand of the forecasts, is calculated as the mean monthly discharge for the period April to September.

2.1 Seasonal discharge variability

Figure 2 shows the seasonal discharges for all catchment considered in this study. The top panel highlights the differences in the magnitude of the seasonal discharge, spanning almost three orders of magnitude (cf. also Table 1). Discontinuous lines indicate data gaps. In order to illustrate differences in the inter-annual variability of the seasonal discharge the lower panel of Figure 2 plots the seasonal discharges normalized to zero mean and standard deviation of 1. This plot indicates similar but also different inter-annual variability patterns of the different catchments. ~~Therefore~~In order to distinguish between similar and different inter-annual variabilities cross-correlations of the seasonal discharges are calculated and hierarchically clustered (Figure 3). Cluster memberships were established using the Ward algorithm clustering the catchments based on the dissimilarities of the correlation between the seasonal discharge time series of the different catchments. The correlation matrix in Figure 3 shows that the seasonal discharges mainly cluster according to their geographical location. The variability of the seasonal discharge of the two catchments in the Altai region (Uba, Ulba) is distinctively different to all the others. Also the two most southern catchments (Amudarya and Murgap) form a distinct cluster that is joined by the most western catchment of the northern Tien Shan, Chirchik. However, Chirchik is also well correlated to the largest group, the catchments in the Tien Shan, which all show similar inter-annual variability of the seasonal discharge. An exception to this is the smallest catchment in the study, Ala-Archa, which is not correlated to any of the other catchments, presumably due to the strong influence of local meteorology and glacier-melt dominated discharge formation in the summer months.

The analysis of the inter-annual variability thus maps the geographical and climatic differences of the catchments considered in this study to a large extent. These differences in variability, but also in the magnitude of the discharges and catchment size imply that the forecast methods can be tested against a wide range of boundary conditions~~.~~ and seasonal variabilities. If skilful forecasts are obtained for all catchments, it can be argued that the approach delivers robust forecasts ~~and~~that are not obtained by chance or due to similar variabilities in all catchments. If successful, it could also be inferred that the approach can be transferred to other regions with similar streamflow generation characteristics.

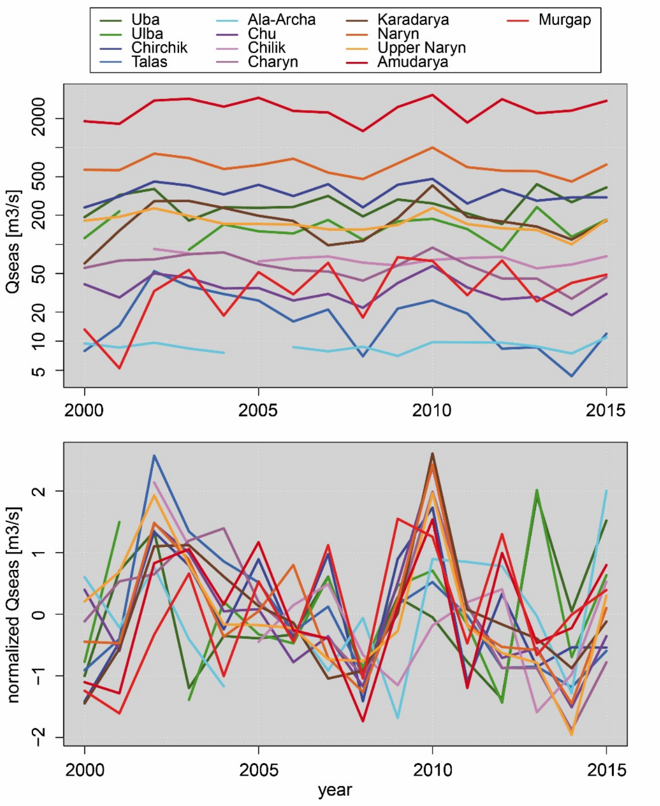


Figure 2: Seasonal discharge (mean monthly discharge for the period April – September) for the catchments under study~~.~~ (upper panel). The lower panel shows the seasonal discharge normalized to zero mean and standard deviation of 1.

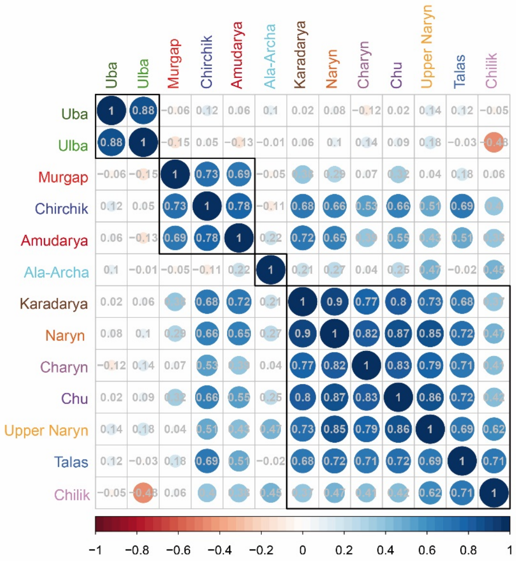


Figure 3: Correlation matrix of the seasonal discharges of the ~~catchment~~catchments under study. The catchments are hierarchically clustered using the Ward algorithm. The colour and size of the circles indicate the direction and strength of the correlations, with blue colours indicating positive, and red colours indicating negative correlations. The numbers provide the actual linear correlation coefficient. The coloured circles indicate significant correlation at a significance level of p = 0.05.

3. Method

As mentioned in the introduction, the seasonal discharge during the vegetation period of April to September in CA is dominated by snow melt in the ~~mountain regions.~~mountains. Therefore a good estimation of the snow accumulation and snow water equivalent in the catchments during the winter months may provide reliable forecasts of the discharge during the vegetation period. However, data about the depth and snow water equivalent are not regularly acquired except for some dedicated research sites. Thus alternative data containing proxy information about the snow depth and water equivalent must be used. Therefore predictors for the forecast models were derived from mean monthly temperature records, monthly sums of precipitation and monthly mean snow coverage of the catchments. It is argued that the combination of these factors is able to serve as proxy data for snow depth and water equivalent. While the precipitation directly contains information about the snow fall amount and thus accumulation, temperature may contain information on the wetness of the snow pack. In combination with snow coverage, temperature and precipitation may thus provide information about the snow volume and water content. In addition to the climate data monthly antecedent discharge can serve as an indicator about the magnitude of the snow melt process and groundwater storage state and release, and is used as predictor, too.

For some regions, particular the Altai catchments, early summer (May – July) precipitation plays a larger role for the seasonal discharge. This precipitation is partly considered as observations in the late forecasts presented here. However, reliable information about the spring precipitation in advance could possibly improve the early forecasts. But due to the low predictability of the typically convection type summer precipitation (Gerlitz et al., 2016) this is not considered in the predictor set.

Evaporative losses in the presented mountain catchments are considered low due to the low summer temperatures and fast catchment response and high water flow velocities in the rivers. Higher losses can occur in reservoir lakes, but with the exception of the large Amudarya basin there are no reservoirs present in the selected catchments. In the Amudary catchment the Nurek reservoir lake at the Vakhsh river exist. However, evaporative losses from the lake surface area of 98 km^2 can be considered negligible in comparison to the large catchment size. Therefore evaporation is not directly considered as predictor for the forecasts.

Moreover, the catchment area of the Vaksh river at the conjunction with the Panj river amounts to 31,415 km², equivalent to about 11% of the Amudarya catchment at Kerky considered here. Assuming further that the reservoir can manage only a fraction of the total discharge of the Vakhsh river, and that the effects of the water retention are further buffered by the seasonal mean discharge spanning six months, it can be assumed that the regulating effect of the Nurek dam on the overall seasonal discharge of the Amudarya at Kerky is rather low. Additionally, the dam is operational since 1980, therefore a discontinuity in the time series 2000-2015 can be ruled out. We thus argue that the anthropogenic influence of the seasonal discharge time series of the Amudarya is negligible for the presented study.

3.1 Generation of the predictor set

The core set of predictors consists of the monthly values preceding the prediction date. According to the operational forecast schemes of the CA ~~hydromet services~~Hydromet Services a series of different prediction dates were defined. The first prediction of the seasonal mean discharge for the vegetation period (April to September) is issued on January 1st, followed by predictions on February 1st, March 1st, April 1st, May 1st, and June 1st. The predictions January to March are preliminary forecasts, while the prediction on April 1st is the most important for the water resource planning in the CA states. The following forecasts serve as corrections of the April forecast. They are actually partial hindcasts, as the predictors already cover a part of the prediction season. This means that the predictors for the late forecasts are not fully independent from the predictand. This decision was on the first hand made due to the foreseen implementation of the method in the Central Asian Hydromet Services. The presented procedure is in line with the official forecast procedures in the Central Asian Hydromet Services. In order to obtain acceptance of the proposed method in the services and their use in the official forecast procedures it is advisable to follow the prescribed procedures. It is required from the Hydromet Services to issue updated (corrected) forecasts, which include the entire vegetation period (April-September). The water regulation procedures and e.g. agricultural yield estimation are traditionally based on bulk numbers for the entire period. If these procedures are not followed, the obtained results, which are better than the forecasts issued with the existing procedures, might not be implemented and come into practise, and thus a chance would be missed to bring research results into application. Furthermore, shortened seasonal discharge time series, where the predictors and predictand are independent, are highly correlated (cf. Annex 1). This means that results obtained with seasonal discharges calculated only for months following the prediction data will obtain very similar results to the ones using the full discharge time series.

For the prediction up to the 1st of April the monthly values over the whole winter period, i.e. from October onwards are used. For later predictions this was limited to data of the prediction year, i.e. from January onwards, in order to keep the number of predictor combinations in reasonable limits. The monthly predictor values were accompanied by multi-monthly means, spanning over two and three months prior to the prediction date, and mean values for the whole predictor period defined above, i.e. either from October to the prediction month, or from January to the prediction month, respectively.

Furthermore, composites were calculated from the climatological data in order to extend the predictor set. They are introduced in order to explore their potential to ~~map~~reflect snow wetness~~ better and thus to improve the prediction~~. It is argued that composites can improve the prediction by linear models, as some non-linear interactions might be ~~mapped~~reflected better by composites compared to the raw data ~~(as shown in e.g. Hall et al., 2017)~~. Analogously to the original data, monthly and multi-monthly composites were derived. For the composites, products of “temperature and precipitation”, “temperature and snow coverage”, “precipitation, snow coverage and temperature”, “precipitation and snow coverage” were used. Antecedent discharge was not included in the composites, because this should not influence the snow cover characteristic.

3.2 Statistical modelling

For the development of the statistical forecast models standard multiple linear regression (MLR) was applied. ~~All~~It is argued that the discharge generation from snow melt over whole catchments and on a seasonal time scale can be approximated by linear models. In fact, this was shown by a large number of studies using hydrological models based on linear concepts like linear storage, e.g. Duethmann et al. (2014) and Duethmann et al. (2015) in Central Asia. Additionally a number of studies have shown that linear regression is a valid approach for seasonal forecasts (e.g. Delbart et al., 2015;Dixon and Wilby, 2015;Seibert et al., 2017). A linear modelling approach is thus seen as a valid approach for seasonal forecasts in the study region from a general point of view. However, in order to statistically support the assumption that the runoff generating processes can be approximated by linear models, the formal assumptions of MLR were also tested: the assumption of normal distribution of the residuals was tested by the Shapiro-Wilk test, the independence of the residuals was tested by calculating the autocorrelation with lag 1, and the heteroscedasticity of the residuals was tested by the Breusch-Pagan test.

In the model selection procedure all possible predictor combinations, which are different for every prediction month as described in 3.1, are used in the MLR for the construction of forecast models. However, some restrictions were put on the predictor combinations in order to avoid overfitting and thus spurious regression results:

1. The predictors are grouped into 8 groups: snow cover, temperature, precipitation, antecedent discharge, and the four composite types.
2. The maximum number of predictors in a regression is limited to four.
3. Only one predictor from each group of predictors can be used in an individual regression model.

Formatiert: Einzug: Links: 0.63 cm, Hängend: 0.63 cm, Abstand Vor: 12 Pt., Nach: 12 Pt., Zeilenabstand: Doppelt

This resulted in 7,728 predictor combinations, i.e. multiple linear models to be tested in January, and increased to 155,690 possible models in April. A complete list of the predictors for the different prediction months is provided in the Annex. The coefficients for all these linear models were automatically fitted during the MLR by the least squares method. Only models with all predictors statistically significant at p = 0.1 and with an overall model significance of at least p = 0.1 were retained. The best models were selected based on the lowest Predicted Residual Error ~~Sum~~Mean of Squares (~~PRESS~~PREMS) value obtained by a Leave-One-Out Cross Validation (LOOCV). In the LOOCV one ~~data point~~year of the ~~time series of~~ seasonal discharge time series is removed from the data set for fitting the MLR. The missing data point is then estimated by the model fitted to the remaining data. The ~~PRESS~~PREMS value is the ~~sum~~mean of squared errors of all seasonal discharges left out and the associated predicted LOOCV values. PREMS is thus defined as:

$$PREMS = \frac{1}{n}\sum_{i=1}^{n} e_{(i)}{}^{2}$$

with $e_{(i)}$ being the residuals of the LOOCV:

$$e_{(i)} = |y_i - \hat{y}_{(i)}|$$

where $\hat{y}_{(i)}$ is the regression estimate of y_i based on a regression equation computed leaving out the ith observation of the overall number of *n* observations. The PREMS was used in this study instead of the usual PRESS (Predictive Residual Error Sum of Squares) in order to avoid biases possibly introduced by missing predictor or predictand data. Using the sum of squares could favour models with missing data compared to models providing predictions for all 16 years. Using the mean of the squares can avoid this to a large extent.

The LOOCV is testing the MLR for robustness and can avoid overfitting and incidental good MLR results valid for the whole data set only. In order to avoid an over-estimation of the forecast skill the seasonal discharge time series were tested for autocorrelation, which could lead to spurious estimation of model robustness.

Model skill was evaluated in a number of measure: adjusted R2, root mean square error RMSE, and mean absolute error MAE. The robustness of the model ensemble was quantified as the ration of the adjusted R2 based on the LOOCV residuals to the adjusted R2 of the complete model residuals. The reliability of the model was analysed by PIT diagrams (e.g. Crochemore et al., 2017) and quantified as PIT-scores (Renard et al., 2010).

In the presented study not only the single-best model according to ~~PRESS~~PREMS of the LOOCV-MLR was selected as prediction model, but rather the best 20 models, if more than 20 models pass the significance tests. This selection aims at the analysis of the differences between the best models in terms of performance and predictors, but also serves as a model ensemble for the forecast of the seasonal discharge. The distribution of the residuals of the best 20 forecast models was evaluated to provide 80% predictive uncertainty intervals for every forecast. However, it has to be noted that the choice to use the best 20 models is subjective, and this number can be increased or reduced. Moreover, a different set of specific models ~~of~~from the best models can be selected according to their performance measures and temporal dynamics by experts knowledgeable of the individual catchments. Sufficient amount of freedom was left for the selection of the number of best models to be retained, in order to enable an expert selection of models by the forecasters of the Central Asian Hydromet Services. The forecasters check every model retained for their performances (quantitatively and qualitatively), and select the models accordingly.

3.3 Predictor importance

The predictors of the selected best models were analysed for their importance, i.e. their share of the overall explained variance (R^2) of the individual models. This was achieved by the *lmg* algorithm implemented in the R-package *relaimpo* ~~(Grömping, 2006)~~. *lmg* is based on sequential R^2s, but explicitly eliminating the dependence on predictor orderings by averaging over orderings using simple unweighted averages. ~~This procedure yields information about the importance of the individual predictors at different~~In sequential R^2 calculations, the model is re-run with a single predictor only and the explained variance is calculated. Then the next predictor is added and the gain in explained variance is calculated. By this procedure the variance explained by individual predictors can be quantified. However, in this procedure the sequence of predictors added influences the share of explained variance associated to the individual predictors. Therefore the *lmg* algorithm tests all possible predictor sequences and calculates the mean importance of every sequence in order to overcome the problem of predictor ordering in sequential R^2s. The predictor importance calculation yields information about the importance of the individual predictors at different forecast points in time for the catchments under study, which in turn can be used for a discussion of the factors responsible for the winter snow accumulation and snow water content in the catchments.

However, such a discussion is complicated by the use of the composite predictors. Therefore the importance of composite predictors is divided into equal proportions to the components of the composites. If more than one composite is used in a model, the proportions associated to the component factors (snow cover, precipitation, temperature) are summed up and displayed as parts of the composite importance in the figures presented in 4.~~3~~2. This analysis is not meant to provide a quantitative estimation for the component importance of the composite predictors, but rather to enhance the discussion and interpretation of the predictors of the selected forecast models.

In addition to the importance ~~of~~for an individual model (here the best LOOCV model), the mean importance over the best 20 LOOCV models is calculated. This is achieved by calculating the fractions of the sum of importance of an individual predictor for all 20 models to the sum of the R^2 values of all 20 models for each catchment and month. These fractions are then multiplied by the mean R^2 values of the best 20 models. This mean predictor importance can be compared to the predictor importance of the best model in order to analyse the stability of the predictor selection within the best 20 LOOCV models.

4. Results

~~For~~ In order to test the suitability of the LOOCV for the seasonal streamflow forecast the autocorrelation and partial autocorrelation of the streamflow time series was calculated and plotted (Annex 2). Any autocorrelation in the discharge time series could lead to artificial over-estimation of the forecasts skill by the LOOCV. Hardly any autocorrelation at α = 0.05 could be detected. Only for the Ulba some significant autocorrelation for lag 1 and 2 is shown just above α = 0.05, but by the partial autocorrelation only. No autocorrelation was found at α = 0.01. Therefore it can be stated that autocorrelation does not exist in the discharge time series of all catchments~~ under study, the ~~, and thus the proposed LOOCV is an appropriate validation method.

The MLR fitting with LOOCV ~~described in~~(cf. 3.2) was applied for different forecast dates ranging from January 1st to June 1st~~.~~ for all catchments. The best 20 models according to the ~~PRESS~~PREMS resulting from the LOOCV were retained for the forecasts. The tests for possible violations of the formal MLR assumptions showed, that 89.5% of all models for all catchments and prediction~~. ~~ months fulfilled the criteria of normal distributes residuals, 95.8% of the models passed the test for independence of the residuals, and 99.5% of the models have homoscedastic residuals (cf. Annex 3). In summary the formal requirements of MLRs are fulfilled by almost all models, and the use of linear models for seasonal discharge forecast is justified also from a formal point of view.

In general the performance of the linear models ~~increased~~increases from January to June, with the best models reaching ~~R^2~~adjusted R^2 (adj. R^2) values in the range of 0.~~86~~76 – 0.~~96~~89 in April and 0.~~88~~84 – 0.99 in June. For most of the catchments high adj. R^2 values in the range of 0.~~58~~57 – 0.~~88~~78 were already obtained in January. Only for Amudarya, Chu and Chirchik the performance is unsatisfyingly low in January, but increases to adj. R^2 > 0.~~7~~59 already one month later in February. Table 2 lists the adj. R^2 values of the best LOOCV models for all catchments and forecast months. Note that the adj. R^2 in the table are calculated using the coefficients of the linear models fitted to the whole data set, i.e. they are not cross validated.

While for most of the catchments, the performance of the models gradually increases with decreasing lead time, the performance for ~~the catchment no. 7, ~~Chilik, shows significant decreases and increases. This is caused by a comparatively large number of missing discharge and predictor data. The automatic fitting algorithm takes advantage of this by finding models able to explain the fewer data points better compared to the full time series~~.~~ despite the use of PREMS instead of PRESS. However, these models can already represent an overfitting and are thus less reliable or stable in time compared to models fitted to longer time periods.

In order to get a more ~~encompassing~~comprehensive picture of the model performance, Figure 4 shows the temporal evolution of the adj. R^2 evaluated for the complete time period of the single best LOOCV model, the minimum adj. R^2 of the best 20 models, the mean adj. R^2 of the best 20 models, the root mean square error (RMSE) of the single best LOOCV model calculated for the full data set normalized to the mean seasonal discharge (cf. Table 1), the normalized mean absolute error MAE, and the ~~PRESS~~PREMS value of the best model. Note that the highest adj. R^2 value is not necessarily the adj. R^2 of the single best model, because the best model is selected according to the lowest ~~PRESS~~PREMS in the LOOCV, and not the best adj. R^2

evaluated using the whole time series. Therefore the mean adj. R^2 in January is occasionally higher than the adj. R^2 of the best LOOCV model, i.e. the most robust model. In general, Figure 4 shows that the different adj. R^2, RMSE, MAE and ~~PRESS~~PREMS values are similar in their evolution in time, i.e. increase (adj. R^2), resp. decrease (RMSE, ~~PRESS~~MAE, PREMS) with later forecast months. This indicates that for all best 20 models the performance is improving with later ~~forecast months~~forecasts.

Furthermore, the difference between min adj. R^2 and mean adj. R^2 to the adj. R^2 of the single best LOOCV model is typically larger in the early prediction months. This indicates a wider spread of model performance within the selected 20 models for the predictions with longer lead times. This difference decreases with shorter lead times, meaning that more models with similar high performance can be found, and thus uncertainty of the model ensemble is reduced. To a certain extent this is likely caused by the larger number of possible predictors for later prediction months, but it is also well justified to assume that the later predictors have more predictive power: data from the late winter months can better describe the snow coverage and water content compared to predictors from the previous autumn. This issue will be discussed further in Section 4.3.

Figure 4 shows that the RMSE as well as the MAE of the best model of the LOOCV is at maximum about 35% of the long term seasonal mean discharge (Talas in January). However, for most catchments the normalized RMSE and MAE is below 20% in January already. For the important April forecast ~~the normalized RMSE is~~they fall generally below 10%, except for Talas and Murgap, where it remains at 20%. These values state the high performance of the linear forecast models in terms of actual discharge, and are thus a useful information for practitioners in order to assess the value of the forecasts.

Figure 4 also shows the ~~PRESS~~PREMS values of the best models and the performance development with the forecast months. ~~As for the R^2 values, the PRESS~~The PREMS values generally decrease (i.e. improve) with ~~prediction month~~decreasing lead time. However, occasionally increases can be observed for later forecast months. This can be also seen in the adj. R^2 values, but less pronounced because of the scale of the left y-axis. This phenomenon is caused by the changing predictor sets from forecast month to forecast month. Particularly multi-monthly predictors change for each prediction date according to the parameter selection outlined in Section 3.1. As this phenomenon of increasing ~~PRESS~~PREMS values usually occurs in April or May, it can be hypothesized that the information of the late winter/early spring months used in the later forecasts does not contain better information about the snow cover as the previous months. With respect to a practical application, the better performing forecasts from the previous months can be used, which is equivalent to an extension of the predictor set by ~~including~~ the predictors of the previous month.

This general reduction of ~~PRESS~~PREMS also means that the models become more robust ~~with~~for later prediction months. To illustrate this more clearly, Figure 4 also shows the relation between the mean adj. R^2 of the LOOCV for all 20 models to the mean adj. R^2 of the full model fit. The mean adj. R^2 of the LOOCV is calculated from the LOOCV residuals used to calculate the ~~PRESS~~PREMS. According to the rationale of the LOOCV, a model is more robust and less prone to overfitting, if the LOOCV-R^2 is very close to the overall R^2. Figure 4 shows that this is generally the case for the catchments with very high adj. R^2 values, and also for later prediction months. This means that the selection of the predictors is likely stable even if additional data is added to the time series in future. However, there are some catchments for which comparably less robust models could be derived even for later prediction months (5. Ala-Archa, 6. Chu). For these catchments it is likely that the predictor selection will change with additional data.

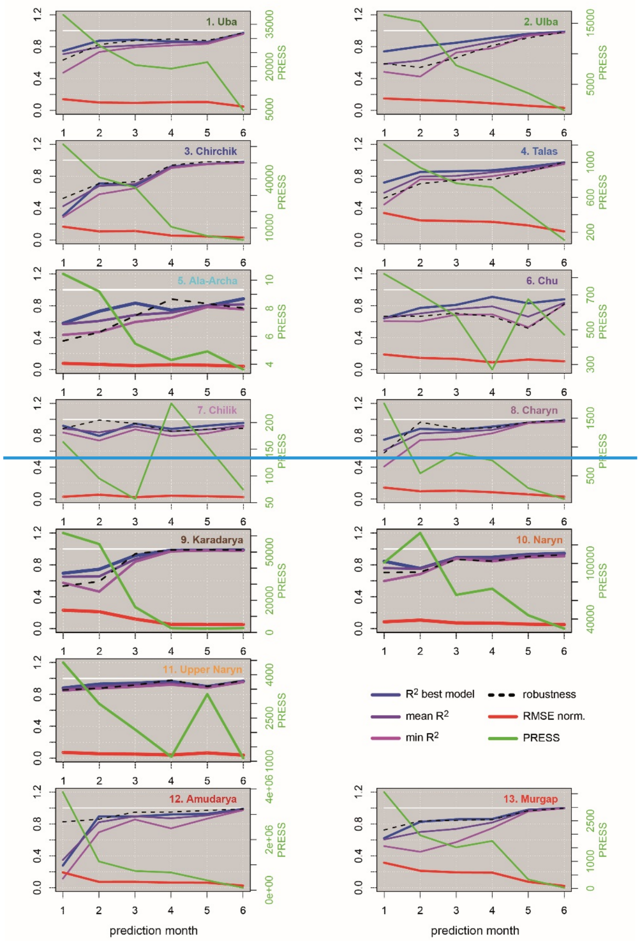
Table 2: Adjusted R²-values of the best performing prediction models from the LOOCV for all catchments and prediction months. "best" indicates the single best model according to the LOOCV, "mean" indicates the mean percentage over the best 20 models according to the LOOCV. The adjusted R² values are associated with indicators for significance levels.

		January		February		March		April		May		June	
		best	*mean*	best	*mean*	best	*mean*	best	*mean*	best	*mean*	best	*mean*
1	**Uba**	0.~~747~~678 ++	0.~~705~~511 ++	0.~~874~~824 +++	0.~~793~~714 +++	0.~~887~~842 +++	0.~~816~~743 +++	0.~~865~~811 +++	0.~~849~~790 +++	0.~~858~~823 +++	0.~~853~~804 +++	0.~~971~~959 +++	0.~~965~~951 +++
2	**Ulba**	0.~~740~~624 o	0.~~582~~429 +	0.~~802~~714 +++	0.~~626~~444 +	0.~~848~~781 +++	0.~~773~~672 ++	0.~~909~~869 +++	0.~~860~~811 +++	0.~~961~~943 +++	0.~~953~~932 +++	0.~~988~~983 +++	0.~~982~~975 +++
3	**Chirchik**	0.~~306~~253 ++	0.~~424~~278 --	0.~~710~~594 +++	0.~~677~~556 ++	0.~~683~~650 +++	0.~~703~~593 ++	0.~~922~~891 +++	0.~~918~~884 +++	0.~~952~~945 +++	0.~~953~~941 +++	0.~~979~~971 +++	0.~~974~~964 +++
4	**Talas**	0.~~720~~669 +++	0.~~592~~408 +	0.~~853~~794 +++	0.~~796~~703 +++	0.~~863~~808 +++	0.~~803~~728 +++	0.~~873~~823 +++	0.~~848~~787 +++	0.~~918~~886 +++	0.~~894~~852 +++	0.~~972~~961 +++	0.~~966~~954 +++
5	**Ala-Archa**	0.~~580~~393 +	0.~~569~~353 o	0.~~731~~597 ++	0.~~609~~431 o	0.~~833~~758 +++	0.~~683~~524 +	0.~~745~~761 +++	0.~~714~~623 ++	0.~~799~~739 +++	0.~~804~~624 ++	0.~~887~~837 +++	0.~~817~~738
6	**Chu**	0.~~643~~274 +	0.~~645~~260 --	0.~~773~~709 +++	0.~~703~~440 o	0.~~811~~903 +++	0.~~758~~729 +++	0.~~912~~680 +++	0.~~791~~569 ++	0.~~831~~800 +++	0.~~663~~740 +++	0.~~881~~887 +++	0.~~837~~862 +++
7	**Chilik***	0.~~919~~865 +++	0.~~883~~818 ++	0.~~793~~856 +++	0.~~837~~787 ++	0.~~946~~910 +++	0.~~910~~873 +++	0.~~880~~757 +++	0.~~850~~770 ++	0.~~920~~880 +++	0.~~872~~805 +++	0.~~955~~933 +++	0.~~922~~821 +++
8	**Charyn**	0.~~745~~643 +++	0.~~612~~503 +	0.~~882~~844 +++	0.~~822~~786 +++	0.~~864~~792 +++	0.~~842~~765 +++	0.~~910~~873 +++	0.~~868~~810 +++	0.~~957~~949 +++	0.~~959~~944 +++	0.~~990~~985 +++	0.~~982~~975 +++
9	**Karadarya**	0.~~695~~573 ++	0.~~651~~449 +	0.~~744~~589 +++	0.~~655~~411 ++	0.~~917~~880 +++	0.~~872~~845 +++	0.~~983~~976 +++	0.~~977~~968 +++	0.~~984~~977 +++	0.~~985~~979 +++	0.~~986~~981 +++	0.~~981~~973 +++
10	**Naryn**	0.~~738~~782 +++	0.~~759~~679 +++	0.~~773~~657 +++	0.~~751~~657 +++	0.~~876~~844 +++	0.~~886~~800 +++	0.~~884~~853 +++	0.~~874~~819 +++	0.~~903~~906 +++	0.~~920~~887 +++	0.~~896~~924 +++	0.~~929~~899 +++
11	**Upper Naryn**	0.~~880~~832 +++	0.~~867~~810 +++	0.~~929~~898 +++	0.~~901~~850 +++	0.~~940~~916 +++	0.~~925~~897 +++	0.~~962~~947 +++	0.~~939~~923 +++	0.~~899~~858 +++	0.~~893~~847 +++	0.~~965~~950 +++	0.~~961~~947 +++
12	**Amudarya**	0.~~278~~213 +	0.~~348~~304 +	0.~~894~~841 +++	0.~~822~~691 +++	0.~~893~~857 +++	0.~~897~~840 +++	0.~~919~~878 +++	0.~~871~~839 +++	0.~~926~~897 +++	0.~~908~~876 +++	0.~~988~~983 +++	0.~~981~~972 +++
13	**Murgap**	0.~~622~~465 ++	0.~~605~~367 o	0.~~827~~757 +++	0.~~699~~551 +	0.~~858~~802 +++	0.~~738~~642 ++	0.~~862~~807 +++	0.~~818~~700 ++	0.~~979~~970 +++	0.~~971~~960 +++	0.~~998~~997 +++	0.~~997~~993 +++

* the performance of Chilik is not representative and comparable to the other catchments due to too many missing discharge and predictor data.

Significance p: +++ = 0.01, ++ = 0.05, + = 0.1, o = 0.2, -- = >0.2; for mean the lowest significance of the model ensemble is used.

Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...


1. Uba
2. Ulba
3. Chirchik
4. Talas
5. Ala-Archa
6. Chu
7. Chilik
8. Charyn
9. Karadarya
10. Naryn
11. Upper Naryn
12. Amudarya
13. Murgap
0.0
0.4
0.8
1.2
1
2
3
4
5
6
PRESS
5000
20000
35000
15000
10000
40000
200
600
1000
4
6
8
10
300
500
700
50
100
150
200
1500
0
50000
40000
100000
1000
2500
4000
0e+00
2e+06
4e+06
R2 best model
mean R2
min R2
robustness
RMSE norm.
PRESS
prediction month
prediction month

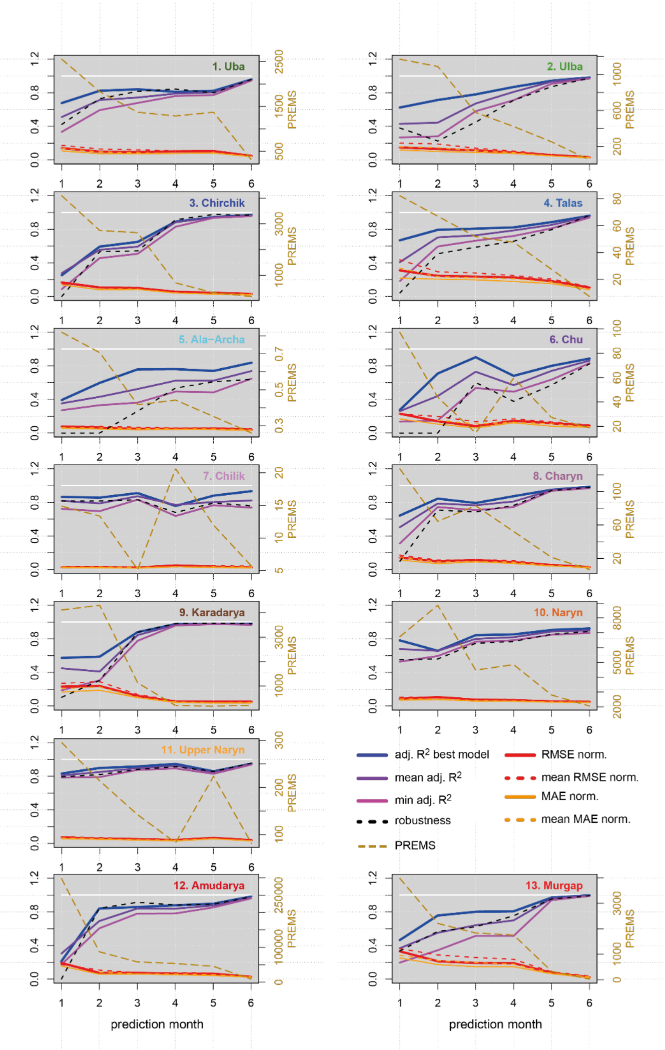


Figure 4: Performance of the prediction models for the different catchments and prediction months. Adj. R² best model is the adjusted R² of the single best LOOCV model, mean adj. R² is the mean adj. R² of the best 20 LOOCV models, min adj. R² is minimum

Formatiert: Schriftart: 9 Pt., Fett

Formatiert: Schriftart: 9 Pt., Fett

Formatiert: Standard

Formatiert: Schriftart: 9 Pt., Fett

Formatiert: Schriftart: 9 Pt., Fett

Formatiert: Schriftart: 9 Pt., Fett

Formatiert: Schriftart: 9 Pt., Fett

Formatiert: Schriftart: 9 Pt., Fett

adj. R^2 of the best 20 LOOCV models, robustness is mean LOOCV-adj. R^2 of the best 20 models divided by the mean adj. R^2, RMSE/MAE norm. is the root mean squared error/mean absolute error of the single best model normalized to mean multi-annual seasonal discharge, ~~PRESS~~mean RMSE/MAE norm is the mean root mean square error/mean absolute error of the best 20 LOOCV models normalized to the multi-annual seasonal discharge; PREMS is the predictive residual sum of squares (PRESS) of the single best model~~.~~, divided by the number of prediction months (i.e. mean of squares).

In addition to the performance metrics Figure 5 plots the temporal dynamics of the best LOOCV models for all six prediction months. It can be seen that the models can map the high variability of the observed seasonal discharges very well, often already in January or February. This graphically corroborates the findings derived from the performance metrics and underlines that the good performance of the models is not a statistical artefact.

Formatiert: Schriftart: 9 Pt., Fett

Formatiert: Schriftart: 9 Pt., Fett

Formatiert: Schriftart: 9 Pt., Fett

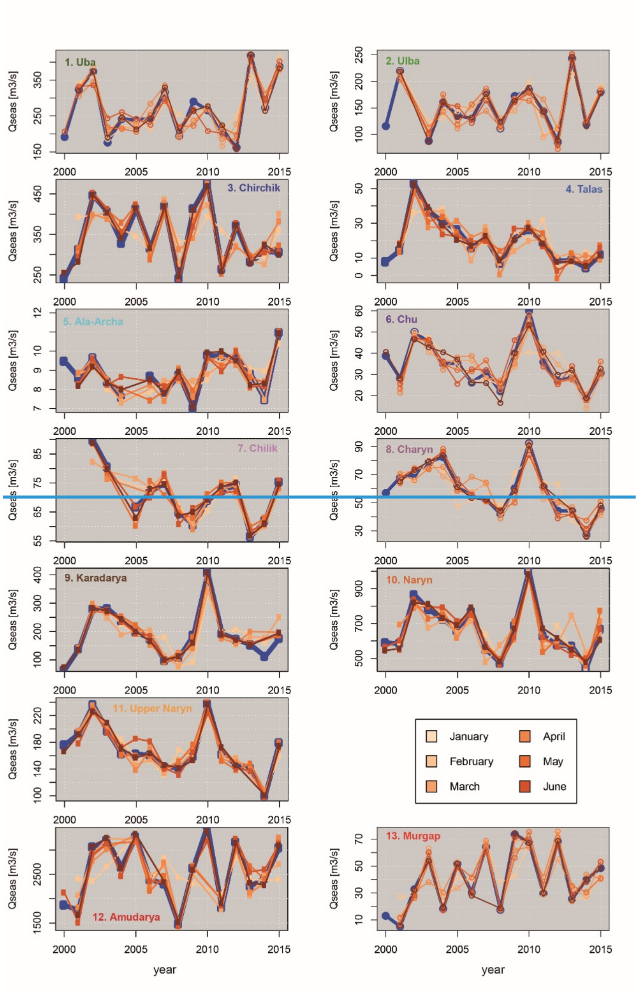
Formatiert: Schriftart: 9 Pt., Fett

Formatiert: Schriftart: 9 Pt., Fett

Formatiert: Schriftart: 9 Pt., Fett

Formatiert: Schriftart: 9 Pt., Fett

Formatiert: Schriftart: 9 Pt., Fett


1. Uba
2. Ulba
3. Chirchik
4. Talas
5. Ala-Archa
6. Chu
7. Chilik
8. Charyn
9. Karadarya
10. Naryn
11. Upper Naryn
12. Amudarya
13. Murgap
Qseas [m3/s]
2000
2005
2010
2015
year
January
February
March
April
May
June

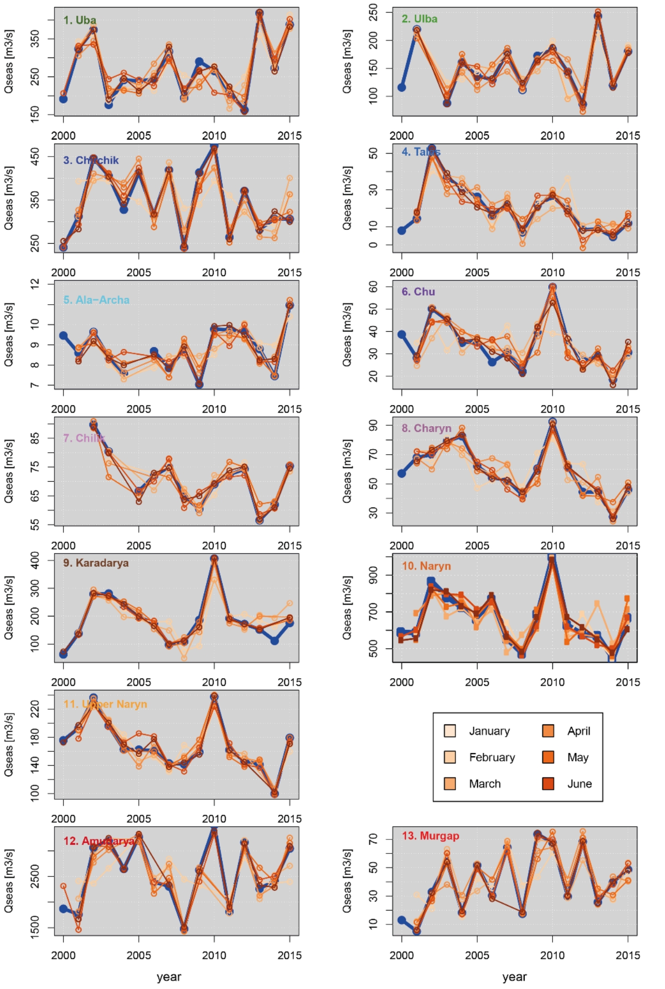


Figure 5: Forecasts of the seasonal discharge by the single best model selected by the LOOCV for the individual catchments and all prediction months. The blue lines show the observed seasonal discharges. Note that some models do not provide forecasts for every year due to missing predictor data.

Formatiert: Block

In order to set the performance of the presented models in the context of the routines and guidelines of the Central Asian ~~hydromet services~~Hydromet Services, the performance of the models was also estimated according to the performance criteria used by the ~~hydromet services.~~Hydromet Services. This is defined by:

$$S_\sigma = \frac{|res|}{\sigma_{Qs}} \qquad (1)$$

With |res| denoting the absolute value of the residual of an individual forecast, and σ_{Qs} the standard deviation of the seasonal discharge (here calculated for the discharge time series used, i.e. for the period 2000-2015). According to the protocols of the ~~hydromet services~~Hydromet Services an acceptable ("good") forecast is defined by $S_\sigma < 0.675$. Table 3 shows how often this criteria was fulfilled during the analysis period 2000-2015 for the best model, and on average by the best 20 models. For the critical forecast month April the criteria was fulfilled for ~~88~~at least 82% of the years (~~14~~13 out of 16 years) by the model ensemble for most of the catchments. For ~~the smallest and the largest catchment (~~Ala-Archa and ~~Amudarya respectively)~~Chu the numbers were lower, but still as high as ~~73~~75% and ~~81~~77%. For all catchments the percentages increase further for the later forecast months. These findings are also valid for ~~all 20 selected~~the ensemble of the best 20 models, as the very similar percentages of the mean of all models compared to the best model indicate. This means that the developed models would provide acceptable forecasts for the ~~hydromet services~~Hydromet Services in the range of 80%-90% for the important forecast month April.

Table 3: Number of times the models yield acceptable prediction according to the criteria of the Central Asian ~~hydromet services~~Hydromet Services for all catchments and prediction months. Numbers indicate percentage of the years of the period 2000-2015 for which the criteria for an acceptable forecast is fulfilled. "best" indicates the best model according to the LOOCV, "mean" indicates the mean percentage over the best 20 models according to the LOOCV.

		January		February		March		April		May		June	
		best	*mean*	best	*mean*	best	*mean*	best	*mean*	best	*mean*	best	*mean*
1	**Uba**	69%	~~74~~70%	88%	~~86~~85%	88%	~~86~~85%	88%	83%	88%	~~91~~92%	94%	~~96~~97%
2	**Ulba**	80%	~~63~~62%	87%	~~70~~71%	87%	~~79~~77%	93%	87%	93%	91%	93%	95%
3	**Chirchik**	50%	~~53~~54%	75%	~~72~~73%	~~69~~75%	~~74~~75%	88%	93%	~~94~~100%	~~96~~98%	100%	~~99~~100%
4	**Talas**	~~75~~81%	~~68~~67%	94%	~~81~~82%	88%	81%	88%	88%	94%	92%	94%	94%
5	**Ala-Archa**	67%	~~63~~59%	73%	~~65~~63%	80%	~~71~~69%	~~73~~87%	~~74~~75%	80%	~~79~~76%	87%	~~80~~81%
6	**Chu**	~~75~~69%	~~75~~55%	~~75~~81%	~~71~~70%	88%	~~78~~81%	~~94~~81%	~~80~~77%	~~88~~81%	~~73~~83%	~~100~~94%	~~91~~89%
7	**Chilik**	85%	~~82~~83%	85%	~~85~~82%	85%	~~95~~93%	92%	~~88~~87%	100%	93%	100%	~~100~~93%
8	**Charyn**	75%	67%	88%	~~82~~84%	~~81~~88%	83%	94%	~~90~~88%	~~94~~88%	93%	88%	~~91~~90%

Formatierte Tabelle ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...
Formatiert ...

9	**Karadarya**	75%	~~72~~70%	~~75~~69%	~~72~~69%	88%	~~83~~84%	88%	88%	88%	89%	94%	~~93~~94%
10	**Naryn**	88%	~~78~~79%	75%	79%	88%	~~88~~84%	88%	87%	100%	~~96~~94%	100%	~~99~~98%
11	**Upper Naryn**	88%	~~85~~86%	88%	87%	88%	~~89~~90%	94%	92%	94%	86%	94%	~~91~~93%
12	**Amudarya**	44%	~~47~~51%	81%	~~74~~70%	75%	~~78~~79%	81%	~~80~~82%	~~100~~94%	~~95~~90%	88%	~~87~~88%
13	**Murgap**	75%	~~72~~66%	88%	~~80~~76%	88%	78%	88%	~~88~~86%	94%	94%	88%	~~92~~95%

4.1 Predictive uncertainty

In order to quantify the predictive uncertainty the empirical 10% and 90% percentiles of the residuals of the forecasts ensembles consisting of the up to best 20 models according to PREMS were calculated for every prediction month. Note that for the early prediction months occasionally less than 20 models passed the significance test. The tables in Annex 3 indicate when this was the case. The quantiles of the residuals were then added to the actual model predictions, thus providing an 80% predictive uncertainty band, i.e. an interval in which the true value of the seasonal discharge should lie with a probability of at least 80%. Figure 6 shows the predictive uncertainty bands for every catchment along with the observed seasonal discharge. The predictive uncertainty for the different prediction months are shown in shades of orange. In general it can be seen, that the predictive uncertainty bands narrow with later prediction months, illustrating the better prediction during later prediction months described above. While this is perfectly visible for most catchments (e.g. 3. Chirchik, 7. Karadarya), it is not the case for some others (5. Ala-Archa, 6. Chu, 10. Naryn). The main reason for this is the larger difference between the predictions and performance of the best 20 models compared to the other catchments, as indicated by the difference between the best and mean adj. R^2 shown and listed in Figure 4 and Table 2, respectively. This causes a wider distribution of the residuals of the best 20 models and thus higher predictive uncertainty. However, if only the best or a smaller selection of the best 20 models are used for a forecast, the predictive uncertainty would also be reduced. This means, that the uncertainty bands derived depend on the subjective choice of the number of models to be kept in the model ensemble. Another reason for wider predictive uncertainty bands for later months is the observed decline in performance during later months in some catchments due to the changed predictor set (e.g. for 6. Chu). This causes again higher predictive uncertainty bands, which overlay the narrower band from the previous month.

From a formal point of view the uncertainty bands correctly include at least 80% of the observed seasonal discharges, even for very narrow bands (e.g. in June for 3. Chirchik or 9. ~~Andijan~~Karadarya). This indicates that the uncertainty estimation derived from the regression residuals provide a reliable estimation of the uncertainty ~~information for~~associated to model selection, and can be used to derive decisions based on the forecasts given by the MLR ~~models~~model ensembles. However, it must be noted that the derived uncertainty bands represent the predictive uncertainty of the MLR models fitted to the available time series. They do not account for any uncertainty stemming from a possible lack of representativeness of the rather short time series used ~~for the "real"~~. Longer discharge time series might show a different variability of ~~the~~ seasonal discharge, which would then not be covered by the derived models. However, as the models can be updated every year in future, this potential problem is expected to decrease with further use of the approach in the Central ~~Asia~~Asian Hydromet Services.

However, it has to be noted that the estimated uncertainty cover only the model selection uncertainty. Other uncertainty sources are:

- model structure, which is assumed to be rather low given the high explained variances;
- data sources, which is not quantifiable, but might be high, particularly the discharge data;
- and the performance criteria for selecting the best models.

The last aspect has been tested. Using other performance criteria as PREMS can result in a slightly different selection of best models, but more often just in a different order of the best models. The best PREMS model is not necessarily the best cross validated R^2 model, or the best MAE or RMSE model. However, as this mainly affects the ordering of the best models, the results in terms of ensemble predictions and predictive uncertainty, if unweighted as presented, would be very similar.

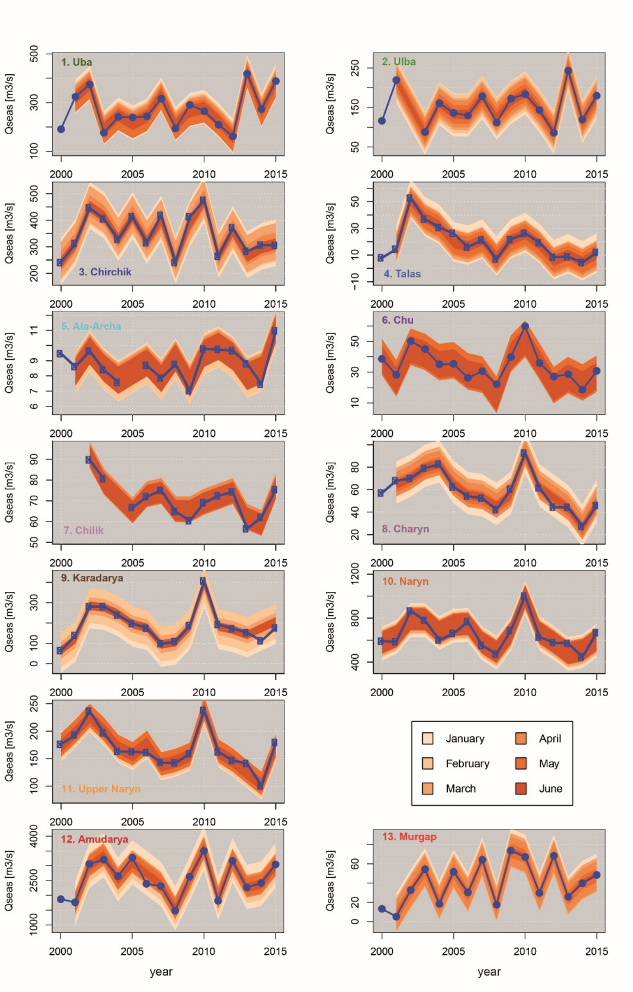


Figure 6: 80% predictive uncertainty bands for all catchments and forecasts months. The blue lines indicate the observed seasonal discharges.

In addition to the predictive uncertainty also the reliability of the forecasts was quantified by PIT diagrams and PIT scores. Figure 7 shows the PIT diagrams for every catchment and all forecast months using the forecasts of the selected ensemble models. The PIT diagrams show that the model ensemble predictions are in most cases close to the 1:1 line, i.e. provide reliable forecasts. However, in some cases the predictive uncertainty is under-estimated (PIT diagram lines with pronounced vertical component around the 50% quantile). This means that some of the predictive uncertainty bands presented in Figure 6 are too narrow to reliably quantify the predictive uncertainty. This is mostly the case for the late forecasts with high skill, where the models in the ensemble often produce very similar forecasts. In addition to the diagrams a PIT score was calculated as the area between the PIT curve and the 1:1 line as a summarizing indicator for the reliability (Renard et al., 2010). The theoretically least reliable model has a score of 0.5, a perfect model a score of 0. The highest score, i.e. the lowest reliability, of all models is 0.2, with the majority of the models being in the range of 0.07-0.15. Interpreting the scores with the curves in the PIT diagram it can be deduced that the reliability of model ensembles with PIT scores ≤ 0.1-0.14 is acceptable. For higher scores the predictive uncertainty derived from the model ensemble is likely to be underestimated.

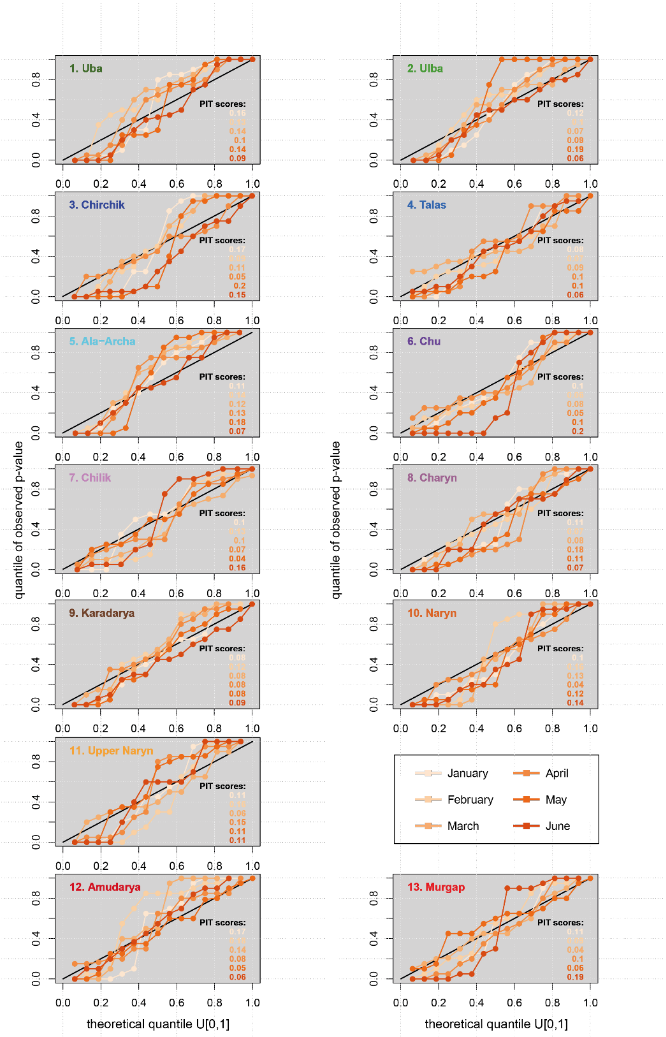


Figure 7: PIT reliability diagrams for every catchment and forecast month. The PIT score is calculated as the area between the reliability plots and the 1:1 line as suggested in Renard et al. (2010). The lower the PIT score, the higher the reliability. The least reliability score is 0.5, the best 0. The colour codes of the PIT scores indicate the forecast month as in the legend.

4.2 Predictor importance (Is there some hydrological process information in ~~the~~ linear ~~model~~models?)

Figure ~~7~~8 illustrates the importance of the predictors of the selected MLR models as absolute fractions of the R^2 values, whereas it is not differentiated between individual predictors, but rather between predictor classes described in 3.1. The left panel of Figure ~~7~~8 shows the importance for the single best LOOCV model, while the right panel shows the average importance of the predictors for the best 20 LOOCV models. A comparison of the left and right panels shows that the predictor selection and importance for the different catchments and prediction months of the best model is quite similar to the mean of the best 20 models. This indicates that the predictor selection for ~~good forecasts~~the models in the ensemble is quite stable, ~~indicating that multicollinearity of the predictors does not impede the predictor selection. Moreover, this can be interpreted~~and hence that the predictor selection is not random, but rather ~~and actually~~ has some hydrological meaning. However, an interpretation of the contributions of the different factors is complicated by the use of the composites, which are almost always selected as one or more predictors in the MLR models. Nevertheless, some general features can be identified from Figure 7:

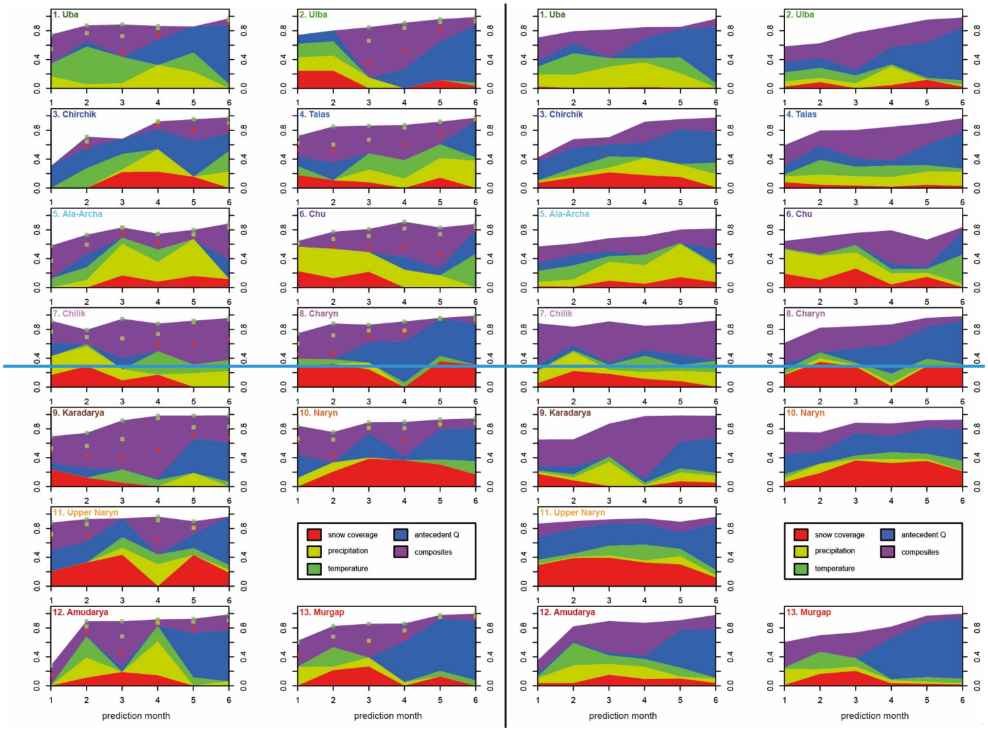
Formatiert: pb_abstract

- Typically there is no single factor dominating the explained variance, with the exception of ~~9. Andijan~~Karadarya, where the composites have an exceptionally large share on the explained variance. But as the composites are comprised of the other predictors (except antecedent discharge), this statement is actually valid for all catchments. This indicates that the winter snow accumulation providing the bulk of the seasonal discharge is best described by a combination of the factors determining the extent and water equivalent of the snow pack in the catchments (precipitation, temperature, snow coverage). Omitting one of these predictors leads in fact to a reduction in model performance.
- There is a general and plausible trend for higher importance of antecedent discharge in the later prediction months. In this period it can be expected that antecedent discharge has higher predictive power of the seasonal discharge compared to the winter months, i.e. during the accumulation phase, because it directly indicates the magnitude of the discharge generation from snow melt. This finding is valid for most catchments except ~~3.~~ Chirchik, ~~5.~~ Ala-Archa and ~~7.~~ Chilik. For Chirchik the importance of antecedent discharge is almost constant throughout the prediction months, both for the best model and on average. Contrary to this, antecedent discharge has very little importance for Ala-Archa and Chilik. For Ala-Archa this observation can be explained by the very small catchment size and thus the quick response of discharge to precipitation events and ~~faster~~lower transit times, but also with the high proportion of glacier melt during the summer months. ~~Thus the lower importance of antecedent discharge matches the catchment characteristics.~~ The high importance of precipitation, which is higher than in any other catchment particularly in the later prediction months, also supports this reasoning. For Chirchik and Chilik, however, no plausible explanation can be derived from the basic catchment characteristics presented here.
- The importance of the snow coverage predictors indicate a regional differentiation of the predictor importance. For the two catchments in the Altai region (1. Uba, 2. Ulba, cluster 1 in Figure 3) snow coverage as an individual factor

is of less importance compared to the other regions. This is due to different snow cover characteristics ~~of~~in these catchments, which have comparatively lower altitudes compared to other catchments in this study. Therefore, snow accumulation in these catchments is comparably low and quickly responds to increasing temperature already in the spring months. Seasonal snow cover variations obtained from MODSNOW-Tool (Gafurov et al, 2016) for these catchments also ~~show (not shown in this manuscript)~~illustrate sudden snow cover depletion in the month of April for both catchments, and for ~~1.~~ Uba with multiple depletions also in winter months until April~~.~~ (analysis not shown in this manuscript). Thus, snowmelt is not important in these catchments for seasonal summer discharge, although it may be of high importance for spring discharge which is beyond the focus of this study. The reverse line of argument can be applied for the relatively high importance of snow coverage for the high altitude central Tien Shan catchments (~~Nr. 8. to~~10. Naryn & 11~~.)~~, Upper Naryn) with mean annual temperatures below zero (cf. Table 1), where snow coverage alone explains up to almost 40% of the explained variance by the MLR models, in addition to the share of snow coverage contained in the composites. For these catchments snow coverage alone is thus already a good indicator of the ~~expected~~ seasonal discharge.

- In terms of predictor importance no obvious differences can be detected on average (right panel in Figure 8) between the Tien Shan and Pamir discharge regimes identified in the cluster analysis (Figure 3), with the exception of the Naryn catchments as stated above. The mean predictor importance figures for those catchments are all very similar. This can indicate that although the variability of seasonal discharges varies with geographical location, the runoff generating processes seem to be similar.

This general interpretation of the predictor importance shows that the selection of the predictors, particularly the change of predictors with prediction months and geographic region, has some hydrological meaning. ~~However, this is on a rather abstract level describing the general runoff generation processes in high mountain catchments.~~ Due to the simplicity of the approach and the simple linear relationship between the predictors, it is unlikely that more hydrological process information and understanding can be extracted from the MLR results. ~~If~~If this can be achieved at all, then on individual catchment basis only and by the interpretation of the exact predictors, i.e. not aggregated by classes as above. This is, however, beyond the scope of this study. But nevertheless, the observation described above indicate that the general runoff generation processes can be described by linear models, and that the presented forecast results are unlikely obtained by pure chance only.



1. Uba
2. Ulba
3. Chirchik
4. Talas
5. Ala-Archa
6. Chu
7. Chilik
8. Charyn
9. Karadarya
10. Naryn
11. Upper Naryn
12. Amudarya
13. Murgap
snow coverage
precipitation
temperature
antecedent Q
composites
0.0
0.4
0.8
1
2
3
4
5
6
prediction month

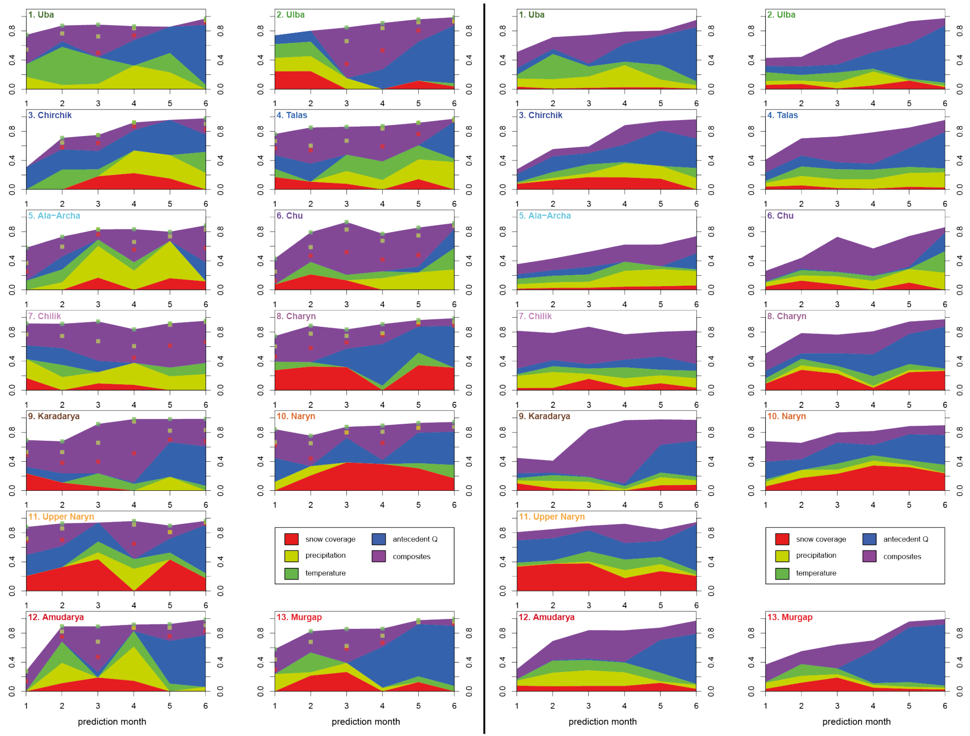


Figure ~~7~~8: Importance of the predictors in the linear models as absolute contribution to the explained variance (R^2) for all catchments and prediction months. Left: of the best LOOCV model; Right: on average for the best 20 LOOCV models. Squares in the left panel figures indicate the presence of the different predictors used in the composites: snow cover, precipitation and temperature, using the same colour codes as for the individual predictors.

4.3 Potential of operational application

~~The presented method for deriving forecast models was designed according to the needs and data availability of the Central Asian hydromet services. It~~A lot of management and strategic decisions are based on seasonal forecasts of water availability in CA. The main consumer of water resources in the Aral Sea basin is the agricultural sector, which is based on one of the world's largest irrigation systems (Dukhovny and de Schutter, 2011). Very important decisions based on water availability forecasts are the planning of agricultural production crop types and water allocation through the irrigation network. Also the estimation of agricultural yield is related to water availability and is needed for country income planning that heavily depends on agricultural export in some countries. Therefore reliable forecasts of seasonal water availability is essential for the economies of Central Asian states.

In order to design a generic and readiliy applicable forecats tool the presented method was designed according to the needs and data availability of the Central Asian Hydromet Services, which are responsible for the seasonal forecasts. The method is based on station data readiliy available to the state agencies, thus fulfilling a core prerequisite for an operational implementation of the method. Moreover, the procedure for deriving forecast models is fairly simple and implemented in the open source software R. Therefore no limitations due to licence issues exist. The model development is automated requiring only some basic definitions as e.g. the formatting and provision of the predictor data as ASCII text files, and the specification of the prediction month. Therefore the code can be applied by the staff of the ~~hydromet services~~Hydromet Services after a short training. However, it has to be noted that the provided predictor data should be as complete as possible in order to avoid spurios model fitting results (overfitting). Due to the automatic model fitting the algorithm may find best performing models fitted to a few years only, if too many predictor data are missing. The chances of overfitting are then greatly increased as the degree of freedom of the linear models, i.e. the ratio of the years used for fitting to the variables in the prediction models, decreases.

The presented model system can also be run with alternative predictor data. For example, it has been tested using gridded ERA-Interim re-analysis data for ~~precipiutation~~precipitation and temperature, averaged monthly over the individual catchment areas. Similar, if not better results as presented were obtained. However, due to the latency of at least two months until the data is released, an operational use of the model system with ERA-Interim data is not feasible at the moment.

~~6~~6 Discussion and Conclusions

The presented study aimed at the development of a flexible and generic forecast model system for the prediction of the seasonal (April-September) discharge in Central Asian river basins, with the final goal of operational use ~~in~~at the ~~hydromet services of the region~~Central Asian Hydromet Services. In order to achieve this the data requirements were kept as low as possible, using only monthly precipitation and temperature data from a single station in the individual catchments, accompanied by operationally processed monthly MODIS snow coverage data and monthly antecedent discharge. Based on this core predictor data set, a variety of monthly, multi-monthly and composite predictors were automatically derived for different prediction dates. The predictors were then used for predicting the seasonal discharge with Multiple Linear Regression models (MLR). In order to avoid overfitting, restrictions were set on the selection and number of predictors in each MLR, and the models were tested for robustness by a Leave-One-Out Cross Validation (LOOCV). An ensemble of prediction models was then selected based on the best Predictive Residual Error ~~Sum~~Mean of Squares (~~PRESS~~PREMS) of the LOOCV.

The prediction model system was tested for the period 2000 – 2015 on a selection of 13 different river basins in different geographic and climatic regions, and with different catchment characteristics. It could be shown that the models provided good to excellent predictions for all catchments and for all defined prediction dates, resp. lead times. For the first prediction on January 1st, i.e. for a lead time of three months, the explained variance (expressed as adjusted R^2) is already high in the range of 0.~~6~~46 – 0.~~88~~86 for ~~11~~9 catchments. For the following prediction on February 1st the explained variance is above 0.~~7~~59 for all catchments, and increases further with the following months. For the important prediction date for the planning of water resources in the region on April 1st just before the high flow season, adj. R^2 values of the best models for each catchment are in the range 0.~~86~~68 – 0.~~96~~97, indicating exceptional high performance for a seasonal forecast.

The automatic selection of the predictors and their importance revealed some geographic or temporal patterns. Geographically the northern Altai catchments differ in the predictor selection of the best LOOCV-MLR models from the other regions as snowmelt in this region has less contribution to seasonal discharge (April – September), with snow cover often reduced to zero already in early spring. For all catchments the importance of antecedent discharge is increasing with progressing prediction dates. This is plausible from a hydrological perspective: While during the winter months the discharge is dominated by groundwater contribution, the discharge in April and later contains information about the snow melt process, and thus predictive power. Moreover, for predictions following April 1st the antecedent discharge represents already part of the predictand, and has thus an even higher predictive power. This means in summary that the selected predictors and their importance have some hydrological meaning, thus supporting the validity of the forecast models derived by the model system. However, it has to be noted that specific features of runoff generation in the catchments cannot be detected and discovered by the rather abstract level of predictors, predictor importance and the very basic catchment characteristics.

Overall, the presented simple forecast system proved to be able to provide robust~~ and~~, very skilful, and reliable forecast models for Central Asia. Moreover, it also provides a generic and flexible tool for the development of seasonal discharge forecast models for Central Asian rivers~~, which~~. This tool can be used by the responsible ~~hydromet services~~Hydromet Services without the need for larger investments in hardware, software, and education and training of staff. In fact, the model system is already tested in four Central Asian national ~~hydro-meteorological services.~~Hydromet Services. The forecasts provided by the MLR models for the summer discharge of 2017 is benchmarked against existing forecast routines and finally the measured discharge in late fall ~~this year~~2017.

The reason for the high performance is surely the separation of the largest share of annual precipitation (snow in winter), and the runoff generation (snow melt in spring and summer). Due to this temporal separation there is no need to perform a seasonal forecast of the precipitation for the summer period, which is ~~typically~~ very difficult and uncertain in Central Asia (Gerlitz et al., 2016). The forecast is rather based on an estimation of the snow pack accumulated in winter and its snow water equivalent, for which the predictors and their combinations provide proxy information. Moreover, the proxy information is not forecasted, but measured, thus providing more reliable information compared to forecasted predictors. As the timely separation of precipitation and runoff is a unifying feature of all Central Asian headwater catchments encompassing high-mountain ranges , the model system is able to perform exceptionally well for all tested catchments, though with different predictor combinations.

It is thus also very likely, that the model system will also work well in the Central Asian catchments not included in this study, with some limitations for very small catchments. Moreover, it can be reasoned that it is likely that the model system will also work well in other regions with similar climatic settings, e.g. the South American dry Andes~~.~~ or the Western U.S. (e.g. the Sierra Nevada). The provided information of seasonal water availability could also be used in dam operation and dam safety procedures, and strategic flood hazard management plans.

Further research using the same method could focus on the use of near-real time satellite based data for the forecasts. As indicated by the successful experiment using ERA-Interim data as predictors, spatially aggregated temperature and precipitation data for whole catchment areas could further improve the forecasts. Additionally the method could be extended to ungauged basins. A potential alternative source for precipitation data could be the near-real time TRMM rainfall estimates. This needs to be accompanied by a satellite based temperature product. MODIS based surface temperature estimates could serve as data source for this. Another completely different data source with prediction potential are the water storage variations derived from gravity records of the GRACE mission. Ongoing research aims at the provision of a near-real time product of GRACE (EGSIEM project, egsiem.eu), which could then be used for operational forecasts. Considering that the gravity based water storage variation should actually map the overall winter accumulation, it can be expected that this data could well serve as predictor for the seasonal discharge. However, due to the low spatial resolution, GRACE data are currently applicable for larger river basins only.

Acknowledgements

This work was undertaken within the frame of the CAWa project (www.cawa-project.net) funded by the German Federal Foreign Office as part of the German Water Initiative for Central Asia (grant number AA7090002).

References

~~Agaltseva, N. A., Borovikova, L. N., and Konovalov, V. G.: Automated system of runoff forecasting for the Amudarya River basin, IAHS-AISH Publication, 193-201, 1997.~~

~~Aizen, V. B., Aizen, E. M., and Melack, J. M.: CLIMATE, SNOW COVER, GLACIERS, AND RUNOFF IN THE TIEN SHAN, CENTRAL ASIA1, JAWRA Journal of the American Water Resources Association, 31, 1113-1129, 10.1111/j.1752-1688.1995.tb03426.x, 1995.~~

~~Aizen, V. B., Aizen, E. M., and Melack, J. M.: Precipitation, melt and runoff in the northern Tien Shan, Journal of Hydrology, 186, 229-251, http://dx.doi.org/10.1016/S0022-1694(96)03022-3, 1996.~~

~~Aizen, V. B., Aizen, E. M., and Kuzmichonok, V. A.: Glaciers and hydrological changes in the Tien Shan: simulation and prediction, Environmental Research Letters, 2, Artn 045019~~
~~10.1088/1748-9326/2/4/045019, 2007.~~

~~Archer, D. R., and Fowler, H. J.: Using meteorological data to forecast seasonal runoff on the River Jhelum, Pakistan, Journal of Hydrology, 361, 10-23, http://dx.doi.org/10.1016/j.jhydrol.2008.07.017, 2008.~~

~~Barlow, M. A., and Tippett, M. K.: Variability and Predictability of Central Asia River Flows: Antecedent Winter Precipitation and Large-Scale Teleconnections, Journal of Hydrometeorology, 9, 1334-1349, 10.1175/2008jhm976.1, 2008.~~

~~Bothe, O., Fraedrich, K., and Zhu, X.: Precipitation climate of Central Asia and the large-scale atmospheric circulation, Theoretical and Applied Climatology, 108, 345-354, 10.1007/s00704-011-0537-2, 2012.~~

~~Conrad, C., Schonbrodt-Stitt, S., Low, F., Sorokin, D., and Paeth, H.: Cropping Intensity in the Aral Sea Basin and Its Dependency from the Runoff Formation 2000-2012, Remote Sensing, 8, ARTN 630~~
~~10.3390/rs8080630, 2016.~~

~~Delbart, N., Dunesme, S., Lavie, E., Madelin, M., Régis, and Goma: Remote sensing of Andean mountain snow cover to forecast water discharge of Cuyo rivers Journal of Alpine Research | Revue de géographie alpine, 103, DOI : 10.4000/rga.2903 2015.~~

~~Dixon, S. G., and Wilby, R. L.: Forecasting reservoir inflows using remotely sensed precipitation estimates: a pilot study for the River Naryn, Kyrgyzstan, Hydrological Sciences Journal, 61, 1-16, 10.1080/02626667.2015.1006227, 2015.~~

~~Irrigation in Central Asia in figures. AQUASTAT Survey-2012: http://www.fao.org/NR/WATER/AQUASTAT/countries_regions/asia_central/index.stm, 2013.~~

~~Feike, T., Mamitimin, Y., Li, L., and Doluschitz, R.: Development of agricultural land and water use and its driving forces along the Aksu and Tarim River, PR China, Environmental Earth Sciences, 73, 517-531, 10.1007/s12665-014-3108-x, 2015.~~

~~Gafurov, A., and Bárdossy, A.: Cloud removal methodology from MODIS snow cover product, Hydrol. Earth Syst. Sci., 13, 1361-1373, 2009.~~

~~Gafurov, A., Kriegel, D., Vorogushyn, S., and Merz, B.: Evaluation of remotely sensed snow cover product in Central Asia, Hydrology Research, 44, 506-522, 10.2166/nh.2012.094, 2013.~~

~~Gafurov, A., Lüdtke, S., Unger-Shayesteh, K., Vorogushyn, S., Schöne, T., Schmidt, S., Kalashnikova, O., and Merz, B.: MODSNOW-Tool: an operational tool for daily snow cover monitoring using MODIS data, Environmental Earth Sciences, 75, 1-15, 10.1007/s12665-016-5869-x, 2016.~~

~~Gerlitz, L., Vorogushyn, S., Apel, H., Gafurov, A., Unger-Shayesteh, K., and Merz, B.: A statistically based seasonal precipitation forecast model with automatic predictor selection and its application to central and south Asia, Hydrol. Earth Syst. Sci., 20, 4605-4623, 10.5194/hess-20-4605-2016, 2016.~~

~~Grömping, U.: Relative importance for linear regression in R: The package relaimpo, Journal of Statistical Software, 17, 2006.~~

~~Hagg, W., Mayer, C., Lambrecht, A., Kriegel, D., and Azizov, E.: Glacier changes in the Big Naryn basin, Central Tian Shan, Global and Planetary Change, 110, 40-50, 10.1016/j.gloplacha.2012.07.010, 2013.~~

~~Hall, R. J., Jones, J. M., Hanna, E., Scaife, A. A., and Erdélyi, R.: Drivers and potential predictability of summer time North Atlantic polar front jet variability, Climate Dynamics, 48, 3869-3887, 10.1007/s00382-016-3307-0, 2017.~~

~~Pal, I., Lall, U., Robertson, A. W., Cane, M. A., and Bansal, R.: Predictability of Western Himalayan river flow: melt seasonal inflow into Bhakra Reservoir in northern India, Hydrol. Earth Syst. Sci., 17, 2131-2146, 10.5194/hess-17-2131-2013, 2013.~~

~~Pritchard, H. D.: Asia's glaciers are a regionally important buffer against drought, Nature, 545, 169-+, 10.1038/nature22062, 2017.~~

~~Rosenberg, E. A., Wood, A. W., and Steinemann, A. C.: Statistical applications of physically based hydrologic models to seasonal streamflow forecasts, Water Resources Research, 47, Artn W00h14~~
~~10.1029/2010wr010101, 2011.~~

~~Schär, C., Vasilina, L., Pertziger, F., and Dirren, S.: Seasonal Runoff Forecasting Using Precipitation from Meteorological Data Assimilation Systems, Journal of Hydrometeorology, 5, 959-973, 10.1175/1525-7541(2004)005<0959:srfupf>2.0.co;2, 2004.~~

~~Schiemann, R., Luthi, D., Vidale, P. L., and Schar, C.: The precipitation climate of Central Asia - intercomparison of observational and numerical data sources in a remote semiarid region, International Journal of Climatology, 28, 295-314, 10.1002/joc.1532, 2008.~~

~~Schöne, T., Zech, C., Unger-Shayesteh, K., Rudenko, V., Thoss, H., Wetzel, H. U., Gafurov, A., Illigner, J., and Zubovich, A.: A new permanent multi-parameter monitoring network in Central Asian high mountains - from measurements to data bases, Geosci Instrum Meth, 2, 97-111, 10.5194/gi-2-97-2013, 2013.~~

~~Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Doll, P., and Portmann, F. T.: Groundwater use for irrigation - a global inventory, Hydrology and Earth System Sciences, 14, 1863-1880, 10.5194/hess-14-1863-2010, 2010.~~

~~Sorg, A., Bolch, T., Stoffel, M., Solomina, O., and Beniston, M.: Climate change impacts on glaciers and runoff in Tien Shan (Central Asia), Nature Climate Change, 2, 725-731, 10.1038/Nclimate1592, 2012.~~

~~Unger-Shayesteh, K., Vorogushyn, S., Farinotti, D., Gafurov, A., Duethmann, D., Mandychev, A., and Merz, B.: What do we know about past changes in the water cycle of Central Asian headwaters? A review, Global and Planetary Change, 110, 4-25, 10.1016/j.gloplacha.2013.02.004, 2013.~~

~~Viviroli, D., Durr, H. H., Messerli, B., Meybeck, M., and Weingartner, R.: Mountains of the world, water towers for humanity: Typology, mapping, and global significance, Water Resources Research, 43, Artn W07447~~
~~10.1029/2006wr005653, 2007.~~

Agaltseva, N. A., Borovikova, L. N., and Konovalov, V. G.: Automated system of runoff forecasting for the Amudarya River basin, IAHS-AISH Publication, 193-201, 1997.

Aizen, V. B., Aizen, E. M., and Melack, J. M.: CLIMATE, SNOW COVER, GLACIERS, AND RUNOFF IN THE TIEN SHAN, CENTRAL ASIA1, JAWRA Journal of the American Water Resources Association, 31, 1113-1129, 10.1111/j.1752-1688.1995.tb03426.x, 1995.

Aizen, V. B., Aizen, E. M., and Melack, J. M.: Precipitation, melt and runoff in the northern Tien Shan, Journal of Hydrology, 186, 229-251, http://dx.doi.org/10.1016/S0022-1694(96)03022-3, 1996.

Aizen, V. B., Aizen, E. M., and Kuzmichonok, V. A.: Glaciers and hydrological changes in the Tien Shan: simulation and prediction, Environmental Research Letters, 2, Artn 045019
10.1088/1748-9326/2/4/045019, 2007.

Archer, D. R., and Fowler, H. J.: Using meteorological data to forecast seasonal runoff on the River Jhelum, Pakistan, Journal of Hydrology, 361, 10-23, http://dx.doi.org/10.1016/j.jhydrol.2008.07.017, 2008.

Barlow, M. A., and Tippett, M. K.: Variability and Predictability of Central Asia River Flows: Antecedent Winter Precipitation and Large-Scale Teleconnections, Journal of Hydrometeorology, 9, 1334-1349, 10.1175/2008jhm976.1, 2008.

Bothe, O., Fraedrich, K., and Zhu, X.: Precipitation climate of Central Asia and the large-scale atmospheric circulation, Theoretical and Applied Climatology, 108, 345-354, 10.1007/s00704-011-0537-2, 2012.

Conrad, C., Schonbrodt-Stitt, S., Low, F., Sorokin, D., and Paeth, H.: Cropping Intensity in the Aral Sea Basin and Its Dependency from the Runoff Formation 2000-2012, Remote Sensing, 8, ARTN 630
10.3390/rs8080630, 2016.

Crochemore, L., Ramos, M. H., Pappenberger, F., and Perrin, C.: Seasonal streamflow forecasting by conditioning climatology with precipitation indices, Hydrol. Earth Syst. Sci., 21, 1573-1591, 10.5194/hess-21-1573-2017, 2017.

Delbart, N., Dunesme, S., Lavie, E., Madelin, M., Régis, and Goma: Remote sensing of Andean mountain snow cover to forecast water discharge of Cuyo rivers Journal of Alpine Research | Revue de géographie alpine, 103, DOI : 10.4000/rga.2903 2015.

Dixon, S. G., and Wilby, R. L.: Forecasting reservoir inflows using remotely sensed precipitation estimates: a pilot study for the River Naryn, Kyrgyzstan, Hydrological Sciences Journal, 61, 1-16, 10.1080/02626667.2015.1006227, 2015.

Duethmann, D., Peters, J., Blume, T., Vorogushyn, S., and Güntner, A.: The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia, Water Resources Research, 50, 2002-2021, 10.1002/2013WR014382, 2014.

Duethmann, D., Bolch, T., Farinotti, D., Kriegel, D., Vorogushyn, S., Merz, B., Pieczonka, T., Jiang, T., Su, B., and Güntner, A.: Attribution of streamflow trends in snow and glacier melt-dominated catchments of the Tarim River, Central Asia, Water Resources Research, 51, 4727-4750, 10.1002/2014WR016716, 2015.

Dukhovny, V. A., and de Schutter, J. L. G.: Water in Central Asia: Past, Present and Future, CRC Press/Balkema,Taylor & Francis Group: London, UK, 2011.

Irrigation in Central Asia in figures. AQUASTAT Survey-2012: http://www.fao.org/NR/WATER/AQUASTAT/countries_regions/asia_central/index.stm, 2013.

Feike, T., Mamitimin, Y., Li, L., and Doluschitz, R.: Development of agricultural land and water use and its driving forces along the Aksu and Tarim River, PR China, Environmental Earth Sciences, 73, 517-531, 10.1007/s12665-014-3108-x, 2015.

Gafurov, A., and Bárdossy, A.: Cloud removal methodology from MODIS snow cover product, Hydrol. Earth Syst. Sci., 13, 1361-1373, 2009.

Gafurov, A., Kriegel, D., Vorogushyn, S., and Merz, B.: Evaluation of remotely sensed snow cover product in Central Asia, Hydrology Research, 44, 506-522, 10.2166/nh.2012.094, 2013.

Gafurov, A., Lüdtke, S., Unger-Shayesteh, K., Vorogushyn, S., Schöne, T., Schmidt, S., Kalashnikova, O., and Merz, B.: MODSNOW-Tool: an operational tool for daily snow cover monitoring using MODIS data, Environmental Earth Sciences, 75, 1-15, 10.1007/s12665-016-5869-x, 2016.

Gerlitz, L., Vorogushyn, S., Apel, H., Gafurov, A., Unger-Shayesteh, K., and Merz, B.: A statistically based seasonal precipitation forecast model with automatic predictor selection and its application to central and south Asia, Hydrol. Earth Syst. Sci., 20, 4605-4623, 10.5194/hess-20-4605-2016, 2016.

Grömping, U.: Relative importance for linear regression in R: The package relaimpo, Journal of Statistical Software, 17, 2006.

Hagg, W., Mayer, C., Lambrecht, A., Kriegel, D., and Azizov, E.: Glacier changes in the Big Naryn basin, Central Tian Shan, Global and Planetary Change, 110, 40-50, 10.1016/j.gloplacha.2012.07.010, 2013.

Hall, R. J., Jones, J. M., Hanna, E., Scaife, A. A., and Erdélyi, R.: Drivers and potential predictability of summer time North Atlantic polar front jet variability, Climate Dynamics, 48, 3869-3887, 10.1007/s00382-016-3307-0, 2017.

Pal, I., Lall, U., Robertson, A. W., Cane, M. A., and Bansal, R.: Predictability of Western Himalayan river flow: melt seasonal inflow into Bhakra Reservoir in northern India, Hydrol. Earth Syst. Sci., 17, 2131-2146, 10.5194/hess-17-2131-2013, 2013.

Pritchard, H. D.: Asia's glaciers are a regionally important buffer against drought, Nature, 545, 169-+, 10.1038/nature22062, 2017.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resources Research, 46, 10.1029/2009WR008328, 2010.

Rosenberg, E. A., Wood, A. W., and Steinemann, A. C.: Statistical applications of physically based hydrologic models to seasonal streamflow forecasts, Water Resources Research, 47, Artn W00h14
10.1029/2010wr010101, 2011.

Schaefli, B., and Gupta, H. V.: Do Nash values have value?, Hydrological Processes, 21, 2075-2080, 2007.

Schär, C., Vasilina, L., Pertziger, F., and Dirren, S.: Seasonal Runoff Forecasting Using Precipitation from Meteorological Data Assimilation Systems, Journal of Hydrometeorology, 5, 959-973, 10.1175/1525-7541(2004)005<0959:srfupf>2.0.co;2, 2004.

Schiemann, R., Luthi, D., Vidale, P. L., and Schar, C.: The precipitation climate of Central Asia - intercomparison of observational and numerical data sources in a remote semiarid region, International Journal of Climatology, 28, 295-314, 10.1002/joc.1532, 2008.

Schöne, T., Zech, C., Unger-Shayesteh, K., Rudenko, V., Thoss, H., Wetzel, H. U., Gafurov, A., Illigner, J., and Zubovich, A.: A new permanent multi-parameter monitoring network in Central Asian high mountains - from measurements to data bases, Geosci Instrum Meth, 2, 97-111, 10.5194/gi-2-97-2013, 2013.

Seibert, M., Merz, B., and Apel, H.: Seasonal forecasting of hydrological drought in the Limpopo Basin: a comparison of statistical methods, Hydrol. Earth Syst. Sci., 21, 1611-1629, 10.5194/hess-21-1611-2017, 2017.

Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Doll, P., and Portmann, F. T.: Groundwater use for irrigation - a global inventory, Hydrology and Earth System Sciences, 14, 1863-1880, 10.5194/hess-14-1863-2010, 2010.

Sorg, A., Bolch, T., Stoffel, M., Solomina, O., and Beniston, M.: Climate change impacts on glaciers and runoff in Tien Shan (Central Asia), Nature Climate Change, 2, 725-731, 10.1038/Nclimate1592, 2012.

Unger-Shayesteh, K., Vorogushyn, S., Farinotti, D., Gafurov, A., Duethmann, D., Mandychev, A., and Merz, B.: What do we know about past changes in the water cycle of Central Asian headwaters? A review, Global and Planetary Change, 110, 4-25, DOI 10.1016/j.gloplacha.2013.02.004, 2013.

Viviroli, D., Durr, H. H., Messerli, B., Meybeck, M., and Weingartner, R.: Mountains of the world, water towers for humanity: Typology, mapping, and global significance, Water Resources Research, 43, Artn W07447
10.1029/2006wr005653, 2007.

Annex

~~Annex 1~~Annex 1: Correlation of seasonal discharge to sub-seasonal discharge

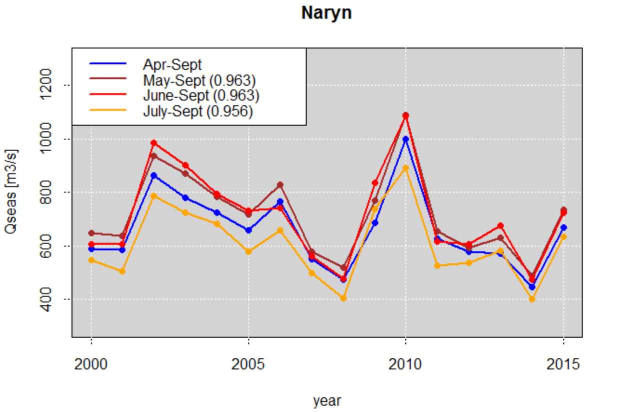


Figure A1: Comparison of seasonal discharge for the whole vegetation period April to September to sub-seasonal discharge time series taking the Naryn basin as example. The sub-seasonal series are highly correlated to the seasonal time series. Numbers in the legend provide the linear correlation coefficient of the sub-seasonal discharges to the seasonal discharge of the whole vegetation period.

Annex 2: Predictors used for the different prediction dates

Formatiert: Schriftart: Fett

The following paragraphs list the predictors created and used for the different forecasts dates, ranging from January 1st to June 1st. The predictors are abbreviated, with *snowcov* and *sc* denoting the snow coverage in the catchment derived by the MODSNOW-tool, *precip* the station records of precipitation, *temp* the station records of temperature, *Q* the discharge recorded at the river gauges. Catchment characteristics and the locations of the gauges are listed in Table 1. The data for all predictors are monthly values (mean for snow coverage, temperature and discharge, sum for precipitation), with *jan* indicating January values, *feb* February values, *mar* March values, *apr* April values, *may* May values and *jun* June values.

Multi-monthly values are mean values of the monthly values spanning over several months, whereas the range of the months included is indicated by the concatenation of the indicators of the months, e.g. *janapr* means multi-monthly means for the period January to April, or *febmar* indicates the mean of the months February and March. The predictor abbreviations are combined with the indicators for the months. *snowcov_apr* thus stands for the mean snow coverage of the catchment in April, or *precip_janmar* for the mean of the monthly precipitation sums for the months January to March.

For the composites the predictors included are listed by their abbreviations, followed by the indicators for the months. For calculating the composites, the monthly values of the predictors denoted by the month indicators are multiplied. E.g. *sc_temp_mar* thus means the product of the mean snow cover in March and the mean temperature in March, or *sc_temp_precip_janmay* denotes the product of the multi-monthly means January to May of snow coverage, temperature and precipitation.

Predictors used for prediction on January 1st

Snow cover:

snowcov_dec snowcov_nov snowcov_oct snowcov_octdec

Precipitation:

precip_dec precip_nov precip_oct precip_novdec precip_octdec

Temperature:

temp_dec temp_nov temp_oct temp_novdec temp_octdec

Composites snow cover x temperature:

sc_temp_octdec

Composites snow cover x precipitation:

sc_precip_octdec

Composites temperature x precipitation:

temp_precip_dec temp_precip_nov temp_precip_oct temp_precip_octdec

Composites snow cover x temperature x precipitation:

sc_temp_precip_octdec

Antecedent discharge:

Q_dec Q_nov Q_oct Q_novdec Q_octdec

Predictors used for prediction on February 1st

Snow cover:

snowcov_jan snowcov_dec snowcov_nov snowcov_oct snowcov_octjan

Precipitation:

precip_jan precip_dec precip_nov precip_oct precip_decjan precip_novjan precip_octjan

Temperature:

temp_jan temp_dec temp_nov temp_oct temp_decjan temp_novjan temp_octjan sc_temp_jan

Composites snow cover x temperature:

sc_temp_jan

Composites snow cover x precipitation:

sc_precip_jan

Composites temperature x precipitation:

temp_precip_jan temp_precip_dec temp_precip_nov temp_precip_oct temp_precip_decjan temp_precip_novjan temp_precip_octjan

Composites snow cover x temperature x precipitation:

sc_temp_precip_octjan

Antecedent discharge:

Q_jan Q_dec Q_nov Q_oct Q_decjan Q_novjan Q_octjan

Predictors used for prediction on March 1st

Snow cover:

snowcov_feb snowcov_jan snowcov_janfeb snowcov_dec snowcov_nov snowcov_oct snowcov_octfeb

Precipitation:

precip_feb precip_jan precip_dec precip_nov precip_oct precip_janfeb precip_decfeb precip_novfeb precip_octfeb

Temperature:

temp_feb temp_jan temp_dec temp_nov temp_oct temp_janfeb temp_decfeb temp_novfeb temp_octfeb

Composites snow cover x temperature:

sc_temp_jan sc_temp_feb sc_temp_janfeb

Composites snow cover x precipitation:

sc_precip_jan sc_precip_feb sc_precip_janfeb

Composites temperature x precipitation:

temp_precip_jan temp_precip_feb temp_precip_dec temp_precip_nov temp_precip_oct temp_precip_janfeb temp_precip_novfeb temp_precip_octfeb

Composites snow cover x temperature x precipitation:

sc_temp_precip_janfeb sc_temp_precip_octfeb

Antecedent discharge:

Q_feb Q_jan Q_dec Q_nov Q_oct Q_janfeb Q_decfeb Q_novfeb Q_octfeb

Predictors used for prediction on April 1st

Snow cover:

snowcov_mar snowcov_feb snowcov_jan snowcov_janmar snowcov_febmar

Precipitation:

precip_mar precip_feb precip_jan precip_dec precip_nov precip_oct precip_febmar precip_janmar precip_decmar precip_novmar precip_octmar

Temperature:

temp_mar temp_feb temp_jan temp_dec temp_nov temp_oct temp_febmar temp_janmar temp_decmar temp_novmar temp_octmar

Composites snow cover x temperature:

sc_temp_mar sc_temp_febmar sc_temp_janmar

Composites snow cover x precipitation:

sc_precip_mar sc_precip_febmar sc_precip_janmar sc_precip_mar_decmar sc_precip_mar_novmar

Composites temperature x precipitation:

temp_precip_jan temp_precip_feb temp_precip_mar temp_precip_febmar temp_precip_janmar temp_precip_decmar temp_precip_novmar

Composites snow cover x temperature x precipitation:

sc_temp_precip_mar sc_temp_precip_febmar sc_temp_precip_janmar

Antecedent discharge:

Q_mar Q_feb Q_jan Q_dec Q_nov Q_oct Q_febmar Q_janmar Q_decmar Q_novmar Q_octmar

Predictors used for prediction on May 1st

Snow cover:

snowcov_apr snowcov_mar snowcov_feb snowcov_janapr snowcov_febapr snowcov_marapr

Precipitation:

precip_apr precip_mar precip_feb precip_jan precip_marapr precip_febapr precip_janapr precip_decapr precip_novapr precip_octapr

Temperature:

temp_apr temp_mar temp_feb temp_jan temp_marapr temp_febapr temp_janapr temp_decapr temp_novapr temp_octapr

Composites snow cover x temperature:

sc_temp_mar sc_temp_apr sc_temp_marapr sc_temp_febapr

Composites snow cover x precipitation:

sc_precip_mar sc_precip_apr sc_precip_marapr sc_precip_febapr

Composites temperature x precipitation:

temp_precip_jan temp_precip_feb temp_precip_mar temp_precip_apr temp_precip_febapr temp_precip_marapr temp_precip_octapr

Composites snow cover x temperature x precipitation:

sc_temp_precip_mar sc_temp_precip_apr sc_temp_precip_marapr sc_temp_precip_janapr

Antecedent discharge:

Q_apr Q_mar Q_feb Q_jan Q_marapr Q_febapr Q_janapr Q_decapr Q_novapr Q_octapr

Predictors used for prediction on June 1st

Snow cover:

snowcov_apr snowcov_mar snowcov_feb snowcov_janapr snowcov_febapr snowcov_marapr

Precipitation:

precip_may precip_apr precip_mar precip_feb precip_jan precip_aprmay precip_marmay precip_febmay precip_janmay precip_octmay

Temperature:

temp_may temp_apr temp_mar temp_feb temp_jan temp_aprmay temp_marmay temp_febmay temp_janmay temp_octmay

Composites snow cover x temperature:

sc_temp_mar sc_temp_apr sc_temp_marmay

Composites snow cover x precipitation:

sc_precip_mar sc_precip_apr sc_precip_marmay

Composites temperature x precipitation:

temp_precip_feb temp_precip_mar temp_precip_apr temp_precip_may temp_precip_marmay temp_precip_octmay

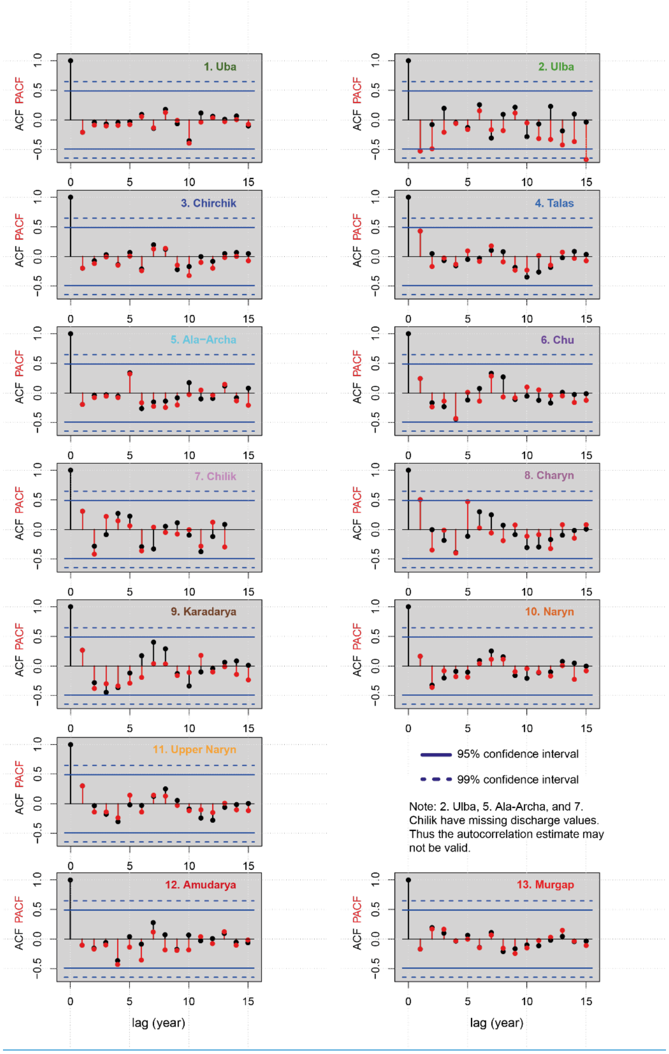
Composites snow cover x temperature x precipitation:

sc_temp_precip_mar sc_temp_precip_apr sc_temp_precip_marmay sc_temp_precip_janmay

Antecedent discharge:

Q_may Q_apr Q_mar Q_feb Q_jan Q_aprmay Q_marmay Q_febmay Q_janmay Q_octmay

Annex 3: Autocorrelation of seasonal discharge time series



Note: 2. Ulba, 5. Ala-Archa, and 7. Chilik have missing discharge values. Thus the autocorrelation estimate may not be valid.

Figure A2: Auto-correlation (black) and Partial auto-correlation (red) of the seasonal discharge tome series for all catchments and possible lags

Annex 4: Formal test for MLR assumptions

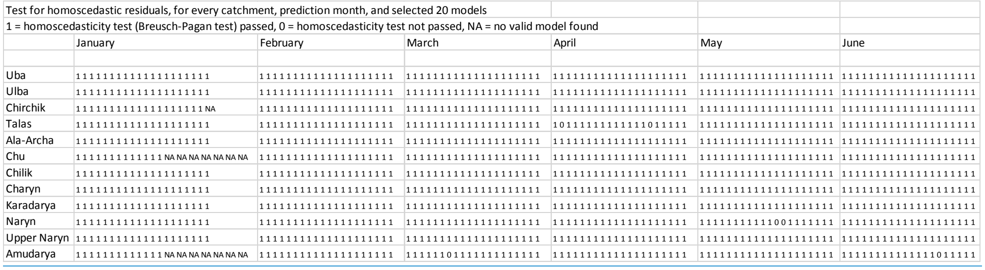
The residuals of the models are tested for normality by the Shapiro-Wilk test for normality. Doing so, one has to bear in mind that this test is based on a sample size of maximal 16 values for each model only, so the test may not provide meaningful results. The table below shows the test result for every model, catchment, and forecast month. A "1" indicates normal distributed residuals, "0" not normal distributed residuals. "NA" indicates that no more models with significant predictors could be found. For every forecast month up to 20 indices are given according to the set of best 20 models to be retained. The table shows that for most of the models (89.5%) the test was positive, i.e. the residuals are normally distributed, even for this rather low and possibly not representative sample size.

Test for normal distributed residuals, for every catchment, prediction month, and selected 20 models						
1 = normal distributed, 0 = not normal distributed, NA = no valid model found						
	January	February	March	April	May	June
Uba	11111111111111111111	10101011011010010110	10000111011100001011	11111111111111101110	11111111111111111111	11111111111111111111
Ulba	11111111111111111111	11110111011111011110	11101111111111111111	11111111111111111111	11111111111000010000	11111111111111111111
Chirchik	1111111110101111111 NA	11111111111111111111	11111111111111111111	11111111111111110100	00000111100000000001	00111001110111111111
Talas	11111111110001111111	11111111111011111111	11111111111111111111	11111111111111111111	11111111111111111111	10101111011111111001
Ala-Archa	11111011111100111111	10110111111111111011	11101111111111001111	11111111111111111111	11101101011111110111	11111111111111111111
Chu	1111011111111 NA NA NA NA NA NA NA	11101111111111111111	11111111111111111101	11110001101111110011	11111111111111011111	11111111111111111111
Chilik	11111111111111111111	11111111111111111111	10111111011011111111	11101111101111011111	11101010111011101100	11111111101001110101
Charyn	11111111111111111111	11111111111111111111	11111111111111111111	11111111111111111111	11111111111111111111	11111111111111111111
Karadarya	11111111111010111110	11111111111111111111	11111100011101110110	11111111111111110111	11111111111111111111	11101110011001111100
Naryn	11111110111111111111	11111111111111111111	11111111111111111111	11111111111111111111	11100111110111111111	11111111110111111111
Upper Naryn	11111111111111111111	11111011100011111111	00101111111111111110	11111111111110111111	11111100001111111111	11111111111111111111
Amudarya	1111111111111 NA NA NA NA NA NA NA	11111111111111111111	11111111111111111111	11111101111111111111	11111111101111011110	11111111111111111111
Murgap	11111111111111111101	11111111111110111111	11111111111101111111	11111111101110011101	11111111110111101111	11111010011111101101

Furthermore it was tested if the residuals are independent applying a test for autocorrelation with lag 1 at significance level p = 0.05. In the table below a "0" indicates independence, a "1" dependence. It shows that 95.8% of the models have independent residuals.

Test for autocorrelated (independent) residuals, for every catchment, prediction month, and selected 20 models, lag = 1						
1 = correlated, 0 = not correlated, NA = no valid model found						
	January	February	March	April	May	June
Uba	00000000001001000010	10001000000010111000	00000000000000000000	00000000000000000000	00000000000000000000	00000000000000000000
Ulba	10010101000000101001	01000011000100000010	00000000000000000000	00000000000000000000	00000000000000000000	11110101110010000000
Chirchik	0000000000000000000 NA	00000000000000000000	00000000000000000000	00000000000000000000	00000000000000000000	00000000000000000000
Talas	00000000000000000000	00000000000000000000	00000000000000000000	00000000000000000000	00000000000000000000	00000000000000000000
Ala-Archa	00000000000000000000	00000000000000001101	00000000000000000000	00000000000000000000	00000000000010000000	00000000000000000000
Chu	0000000000000 NA NA NA NA NA NA NA	00000000000000000000	01100000000000000000	00000000000000000000	00000000000000000000	00000110000000100000
Chilik	00000000000000000000	00000000000000000000	00000000000000000000	00000000000000000000	10001000000000000000	00000000000000000000
Charyn	00000000000100000000	00000000000000000000	00000000000000000000	00000000000000000000	00000000000000000000	00000000000000000100
Karadarya	01000000010000010000	00000000000000000000	00000000000000000000	00000000000000000000	00000000000000000000	00000000000000000000
Naryn	00000000000000000000	00000000000000000000	00000000000000000000	00000000000000000000	00110000000000000000	00000000000000010000
Upper Naryn	00000000000000000000	00000000000000000000	00000100000000000000	00000000000000000000	00000011110000000000	11000000011000001010
Amudarya	0000000000000 NA NA NA NA NA NA NA	00000000000000000000	00000000000000000000	00000000000000000010	00000000000001000000	00000000000010000000
Murgap	00000000000000000000	10001000000000000000	00000000000000000000	00000001000000000000	10000000000000000000	00000000000000000000

Last the Breusch-Pagan test for heteroscedasticity was applied to the residuals. This test shows that 99.5% of the models have homoscedastic residuals. In the table below a "1" indicates homoscedastic residuals, a "0" heteroscedastic residuals according to the test.



Test for homoscedastic residuals, for every catchment, prediction month, and selected 20 models

1 = homoscedasticity test (Breusch-Pagan test) passed, 0 = homoscedasticity test not passed, NA = no valid model found

	January	February	March	April	May	June
Uba	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Ulba	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Chirchik	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 NA	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Talas	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Ala-Archa	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Chu	1 1 1 1 1 1 1 1 1 1 1 1 1 NA NA NA NA NA NA NA	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Chilik	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Charyn	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Karadarya	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Naryn	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Upper Naryn	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Amudarya	1 1 1 1 1 1 1 1 1 1 1 1 1 NA NA NA NA NA NA NA	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1