Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-340
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 21 June 2017

Hydrology and Earth System Sciences
Discussions

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 – for the period 2000-2015 provided skilful forecasts for

most catchments already in January. The skill of the prediction increased every month, with R^2 values often in the range 0.8 – 0.9 in April just before the prediction 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.

1 Introduction

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) . Tien Shan and Pamir mountains exhibit particularly high relative water yield compared to the lowland parts of these catchments (Viviroli et al., 2007). Related to the economic water demands in the lowland plains primarily for irrigated agriculture, Tien Shan and Pamir mountains are among the most important contributors of stream water worldwide (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, yield estimation and management of reservoir capacities in the upper parts of the catchments. Seasonal forecasts are one of the major responsibilities of the state hydrometeorological (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, questionnaire survey). 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 apply the hydrological forecast model AISHF (Agaltseva et al., 1997) developed at the Uzbek Hydrometeorological 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) 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 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 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), 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

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-340
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 21 June 2017

meteorological gauging stations used for the seasonal flow forecast. 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 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.

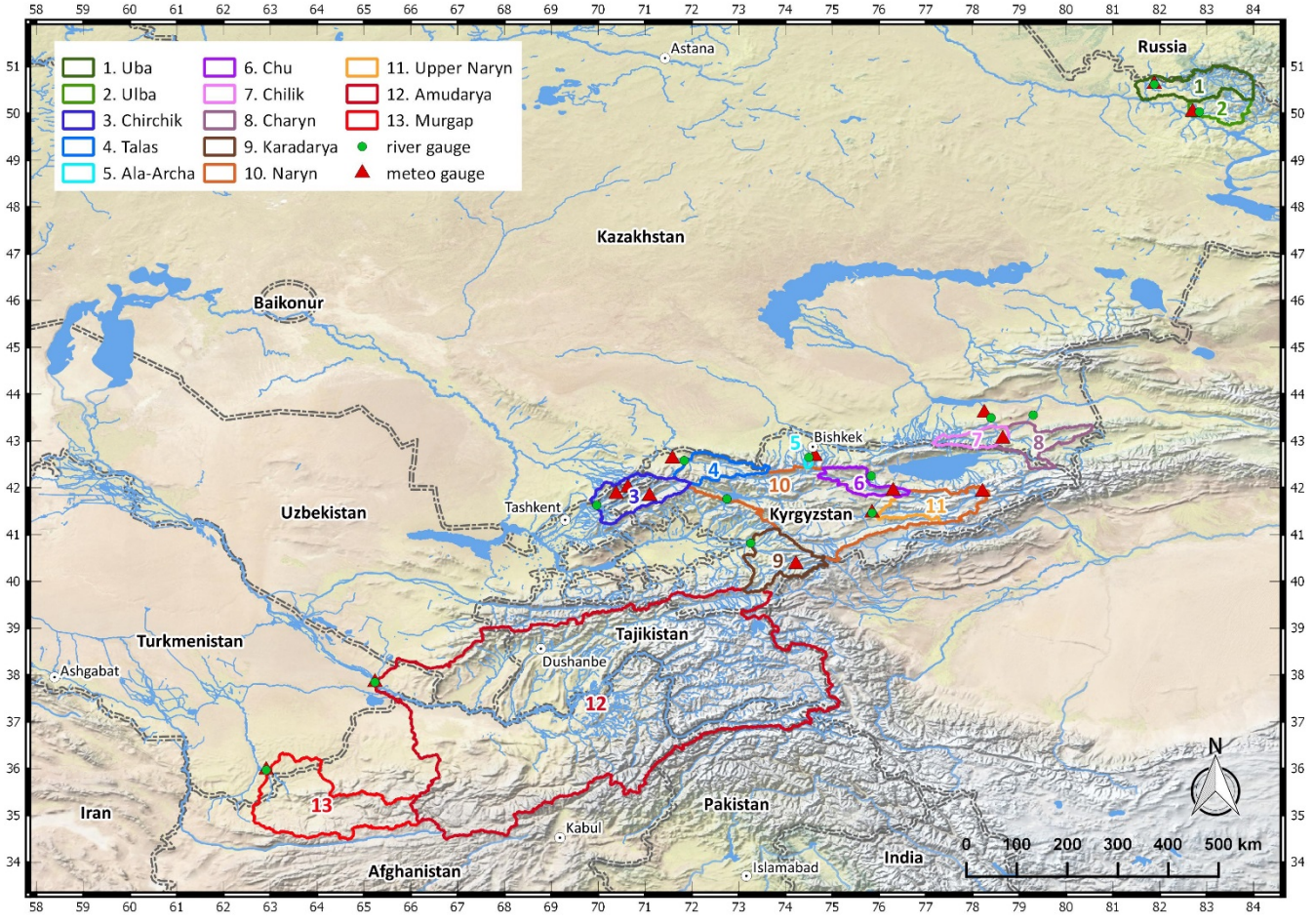

Figure 1: Overview of the catchments for which prediction models were established, with locations of discharge and meteorological gauging stations used (coordinates in latitude/longitude).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-340
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 21 June 2017

Hydrology and Earth System Sciences
Discussions

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.

	catchment	discharge station	Q deg. lat	Q long	meteo station	meteo deg. lat	meteo deg. long	meteo altitude [m]	catchment area [km^2]	Q mean seasonal [m^3/s]	mean altitude [m]
1	Uba	Shemonaikha	50.620	81.880	Shemonaikha	50.620	81.880	300	9324	269.2	740
2	Ulba	Perevalochnaya	50.033	82.843	Oskemen	50.030	82.700	375	5080	151.4	950
3	Chirchik	Charvak	41.626	69.969	Chatkal	41.822	71.097	2300	10903	346.21	2575
					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
5	Ala-Archa	Kashka-Suu	42.650	74.500	Baytik	42.670	74.630	1579	239	8.83	3288
6	Chu	Kochkor	42.250	75.833	Kara Kuzhur	41.930	76.300	855	4961	34.53	2934
7	Chilik	Malybai	43.494	78.392	Shelek	43.597	78.249	600	3964	70.67	2603
8	Charyn	Sarytogai	43.553	79.293	Zhalanash	43.043	78.642	1690	7921	59.06	2260
9	Karadarya	Andijan	40.814	73.257	Ak-Terek	40.365	74.222	1190	11670	186.21	2663
10	Naryn	Toktogul	41.760	72.750	Naryn city	41.460	75.850	2040	51926	653.13	2850
11	Upper Naryn	Naryn city	41.460	75.85	Tien Shan	41.910	78.210	3614	10343	168.64	3546
12	Amudarya	Kerki	37.842	65.23	Kerki	37.842	65.230	237	287714	2551.02	2578
13	Murgab	Takhta Bazar	35.966	62.907	Takhta Bazar	35.966	62.907	354	35767	40.13	1707

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. 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 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, 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 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 different inter-annual variability patterns of the different catchments. Therefore cross-correlations of the seasonal discharges are calculated and hierarchically clustered (Figure 3). 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. If skilful forecasts are obtained for all catchments, it can be argued that the approach delivers robust forecasts and can be transferred to other regions with similar streamflow generation characteristics.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-340
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 21 June 2017

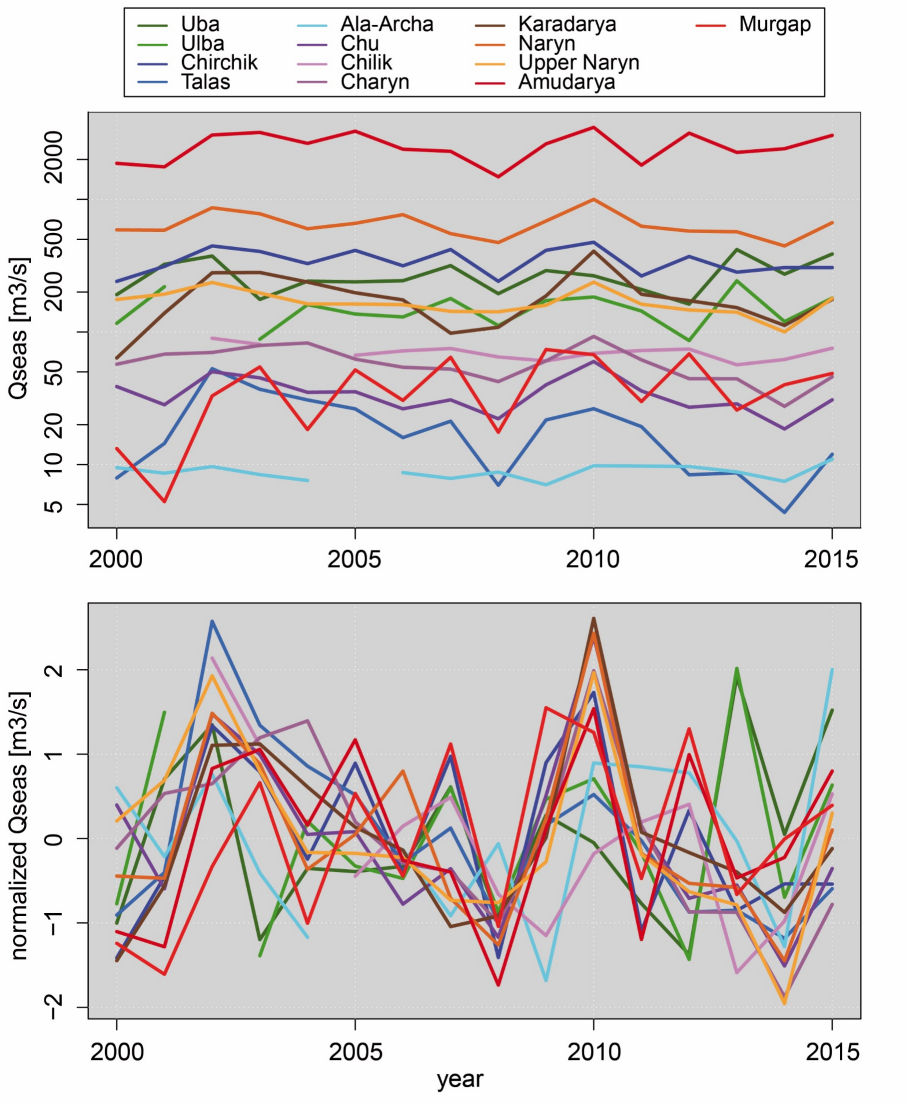

Figure 2: Seasonal discharge (mean monthly discharge for the period April – September) for the catchments under study. The lower panel shows the seasonal discharge normalized to zero mean and standard deviation of 1.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-340
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 21 June 2017

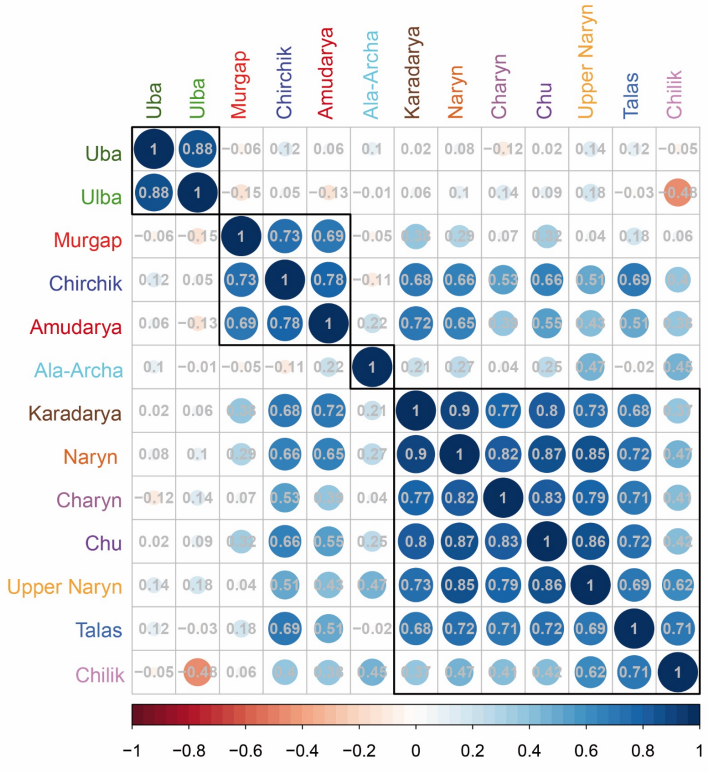

Figure 3: Correlation matrix of the seasonal discharges of the catchment under study. The catchments are hierarchically clustered using the Ward algorithm. The colour and size of the circles indicate the direction and strength of the correlations, with blue colours indicating positive, and red colours 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. 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.

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 a series of different prediction dates were defined. The first prediction of the seasonal mean discharge (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. 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 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 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 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.

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. The best models were selected based on the lowest Predicted Residual Error Sum of Squares (PRESS) value obtained by a Leave-One-Out Cross Validation (LOOCV). In the LOOCV one data point of the time series of seasonal discharge 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 value is the sum of squared errors of all seasonal discharges left out and the associated predicted LOOCV values. 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 the presented study not only the single-best model according to PRESS of the LOOCV-MLR was selected as prediction model, but rather the best 20 models. 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 set of specific models of the best models can be selected according to their performance measures and temporal dynamics by experts knowledgeable of the individual catchments.

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 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. 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 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 all catchments under study, the MLR fitting with LOOCV described in 3.2 was applied for different forecast dates ranging from January 1st to June 1st. The best 20 models according to the PRESS resulting from the LOOCV were retained for prediction. In general the performance of the linear models increased from January to June, with the best models reaching R^2 values in the range of 0.86 – 0.96 in April and 0.88 – 0.99 in June. For most of the catchments high R^2 values in the range of 0.58 – 0.88 were already obtained in January. Only for Amudarya and Chirchik the performance is unsatisfyingly low in January, but increases to > 0.7 already one month later in February. Table 2 lists the R^2 values of the best LOOCV models for all catchments and forecast months. Note that the 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. 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 picture of the model performance, Figure 4 shows the temporal evolution of the R^2 evaluated for the complete time period of the single best LOOCV model, the minimum R^2 of the best 20 models, the mean 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), and the PRESS value of the best model. Note that the highest R^2 value is not necessarily the R^2 of the single best model, because the best model is selected according to the lowest PRESS in the LOOCV, and not the best R^2 evaluated using the whole time series. Therefore the mean R^2 in January is occasionally higher than the R^2 of the best LOOCV model, i.e. the most robust model. In general, Figure 4 shows that the different R^2, RMSE and PRESS values are similar in their evolution in time, i.e. increase (R^2), resp. decrease (RMSE, PRESS) with later forecast months. This indicates that for all best 20 models the performance is improving with later forecast months.

Furthermore, the difference between min R^2 and mean R^2 to the 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 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 is below 20% in January already. For the important April forecast the normalized RMSE is 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 values of the best models and the development with the forecast months. As for the R^2 values, the PRESS values generally decrease (i.e. improve) with prediction month. However, occasionally increases can be observed for later forecast months. This can be also seen in the 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 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 also means that the models become more robust with later prediction months. To illustrate this more clearly, Figure 4 also shows the relation between the mean R^2 of the LOOCV for all 20 models to the mean R^2 of the full model fit. The mean R^2 of the LOOCV is calculated from the LOOCV residuals used to calculate the PRESS. 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 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: 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.

		January		February		March		April		May		June	
		best	*mean*	best	*mean*	best	*mean*	best	*mean*	best	*mean*	best	*mean*
1	**Uba**	0.747	*0.705*	0.874	*0.793*	0.887	*0.816*	0.865	*0.849*	0.858	*0.853*	0.971	*0.965*
2	**Ulba**	0.740	*0.582*	0.802	*0.626*	0.848	*0.773*	0.909	*0.860*	0.961	*0.953*	0.988	*0.982*
3	**Chirchik**	0.306	*0.424*	0.710	*0.677*	0.683	*0.703*	0.922	*0.918*	0.952	*0.953*	0.979	*0.974*
4	**Talas**	0.720	*0.592*	0.853	*0.796*	0.863	*0.803*	0.873	*0.848*	0.918	*0.894*	0.972	*0.966*

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-340
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 21 June 2017

Hydrology and Earth System Sciences
Open Access
Discussions

5	**Ala-Archa**	0.580	*0.569*	0.731	*0.609*	0.833	*0.683*	0.745	*0.714*	0.799	*0.804*	0.887	*0.817*
6	**Chu**	0.643	*0.645*	0.773	*0.703*	0.811	*0.758*	0.912	*0.791*	0.831	*0.663*	0.881	*0.837*
7	**Chilik***	0.919	*0.883*	0.793	*0.837*	0.946	*0.910*	0.880	*0.850*	0.920	*0.872*	0.955	*0.922*
8	**Charyn**	0.745	*0.612*	0.882	*0.822*	0.864	*0.842*	0.910	*0.868*	0.957	*0.959*	0.990	*0.982*
9	**Karadarya**	0.695	*0.651*	0.744	*0.655*	0.917	*0.872*	0.983	*0.977*	0.984	*0.985*	0.986	*0.981*
10	**Naryn**	0.738	*0.759*	0.773	*0.751*	0.876	*0.886*	0.884	*0.874*	0.903	*0.920*	0.896	*0.929*
11	**Upper Naryn**	0.880	*0.867*	0.929	*0.901*	0.940	*0.925*	0.962	*0.939*	0.899	*0.893*	0.965	*0.961*
12	**Amudarya**	0.278	*0.348*	0.894	*0.822*	0.893	*0.897*	0.919	*0.871*	0.926	*0.908*	0.988	*0.981*
13	**Murgap**	0.622	*0.605*	0.827	*0.699*	0.858	*0.738*	0.862	*0.818*	0.979	*0.971*	0.998	*0.997*

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

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-340
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 21 June 2017

Hydrology and Earth System Sciences
Discussions

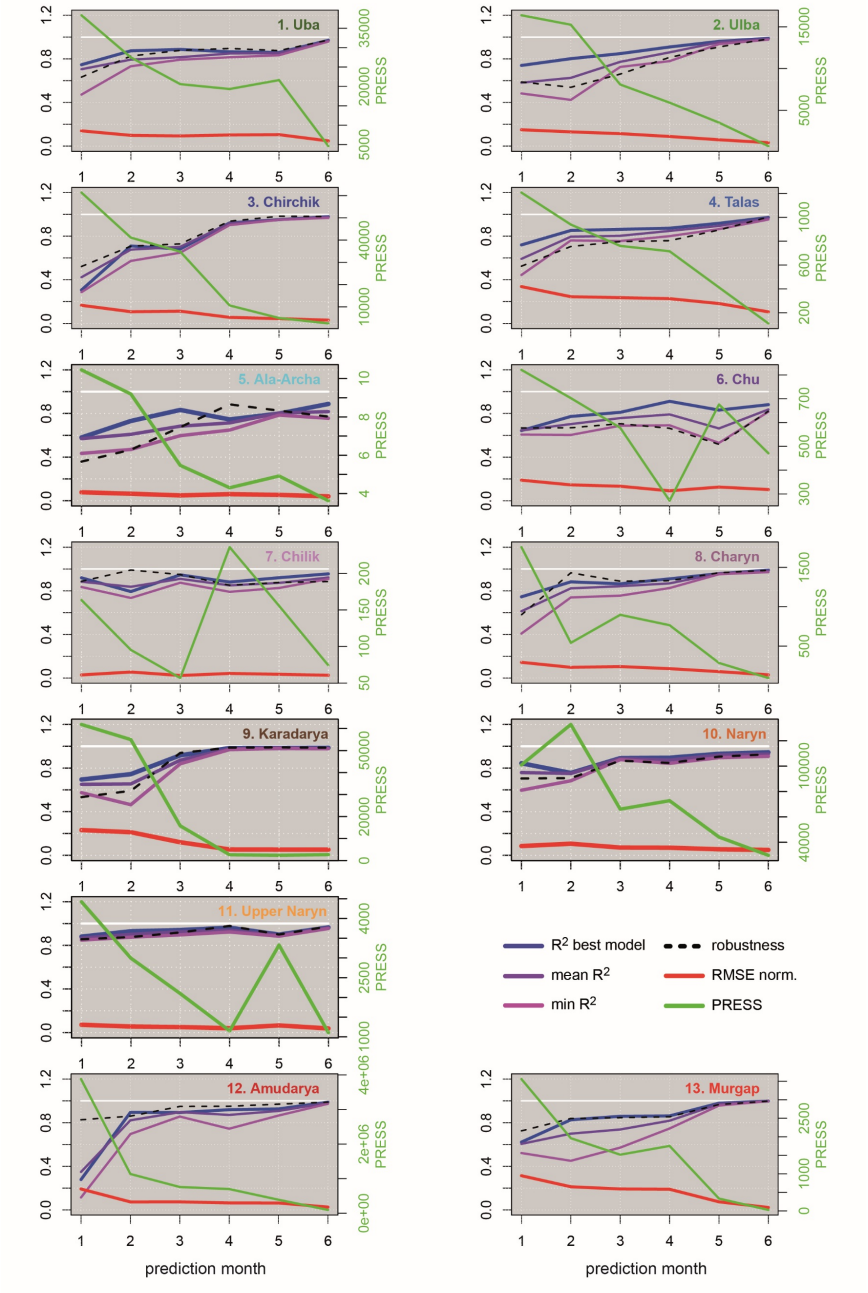

Figure 4: Performance of the prediction models for the different catchments and prediction months. R^2 best model is R^2 of the single best LOOCV model, mean R^2 is mean R^2 of the best 20 LOOCV models, min R^2 is minimum R^2 of the best 20 LOOCV models, robustness is mean LOOCV-R^2 of the best 20 models divided by the mean R^2, RMSE norm. is the root mean squared error of the single best model normalized to mean multi-annual seasonal discharge, PRESS is predictive residual sum of squares of the single best model.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-340
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 21 June 2017

Hydrology and Earth System Sciences
Open Access
Discussions

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.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-340
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 21 June 2017

Hydrology and Earth System Sciences
Discussions

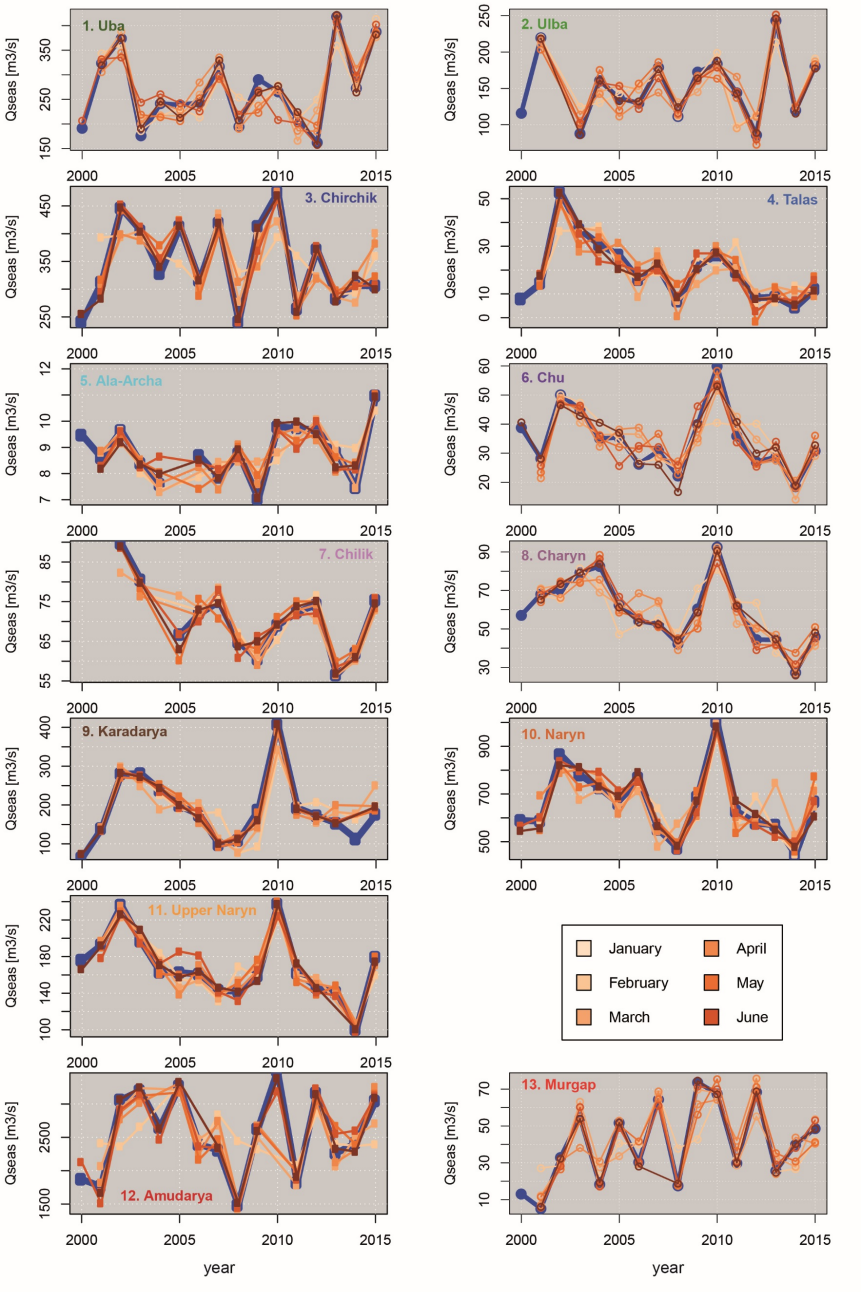

Figure 5: Forecasts of the seasonal discharge by the single best model selected by the LOOCV for the individual catchments and all prediction months. The blue lines show the observed seasonal discharges. Note that some models do not provide forecasts for every year due to missing predictor data.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-340
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 21 June 2017

Hydrology and Earth System Sciences
Open Access
Discussions

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

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

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 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% of the years (14 out of 16 years) for most of the catchments. For the smallest and the largest catchment (Ala-Archa and Amudarya respectively) the numbers were lower, but still as high as 73% and 81%. For all catchments the percentages increase further for the later forecast months. These findings are also valid for all 20 selected best 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 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 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*	88	*86*	88	*86*	88	*83*	88	*91*	94	*96*
2	**Ulba**	80	*63*	87	*70*	87	*79*	93	*87*	93	*91*	93	*95*
3	**Chirchik**	50	*53*	75	*72*	69	*74*	88	*93*	94	*96*	100	*99*
4	**Talas**	75	*68*	94	*81*	88	*81*	88	*88*	94	*92*	94	*94*
5	**Ala-Archa**	67	*63*	73	*65*	80	*71*	73	*74*	80	*79*	87	*80*
6	**Chu**	75	*75*	75	*71*	88	*78*	94	*80*	88	*73*	100	*91*
7	**Chilik**	85	*82*	85	*85*	85	*95*	92	*88*	100	*93*	100	*100*
8	**Charyn**	75	*67*	88	*82*	81	*83*	94	*90*	94	*93*	88	*91*
9	**Karadarya**	75	*72*	75	*72*	88	*83*	88	*88*	88	*89*	94	*93*
10	**Naryn**	88	*78*	75	*79*	88	*88*	88	*87*	100	*96*	100	*99*
11	**Upper Naryn**	88	*85*	88	*87*	88	*89*	94	*92*	94	*86*	94	*91*
12	**Amudarya**	44	*47*	81	*74*	75	*78*	81	*80*	100	*95*	88	*87*
13	**Murgap**	75	*72*	88	*80*	88	*78*	88	*88*	94	*94*	88	*92*

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-340
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 21 June 2017

4.1 Predictive uncertainty

In order to quantify the predictive uncertainty the empirical 10% and 90% percentiles of the residuals of the best 20 models were calculated for every prediction month. 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 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). This indicates that the uncertainty estimation derived from the regression residuals provide a reliable uncertainty information for decisions based on the forecasts given by the MLR models. 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 time series used for the "real" variability of the seasonal discharge in Central Asia.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-340
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 21 June 2017

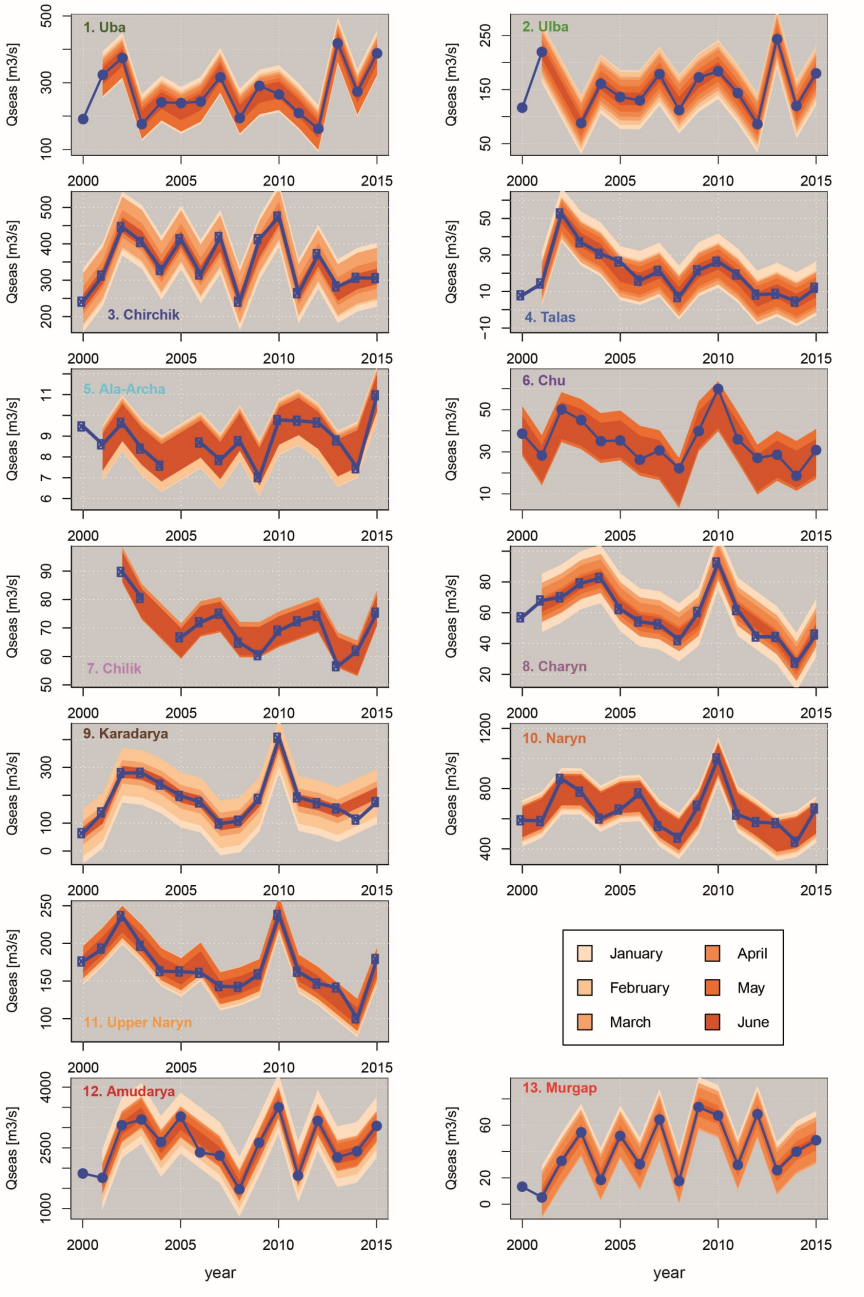

Figure 6: 80% predictive uncertainty bands for all catchments and forecasts months. The blue lines indicate the observed seasonal discharges.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-340
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 21 June 2017

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

Figure 7 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 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 is quite stable, indicating that multicollinearity of the predictors does not impede the predictor selection. Moreover, this can be interpreted that the predictor selection is not random, but rather and actually has 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:

- Typically there is no single factor dominating the explained variance, with the exception of 9. Andijan, 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 small catchment size and thus the quick response of discharge to precipitation events and faster 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) snow coverage as an individual factor is of less importance compared to the other regions. This is due to different snow cover characteristics of 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) 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. Thus, snowmelt is not important in these catchments for seasonal 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 Tien Shan catchments (Nr. 8. to 11.), where snow coverage alone explains up to 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 is thus already a good indicator of the expected seasonal discharge.

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 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.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-340
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 21 June 2017

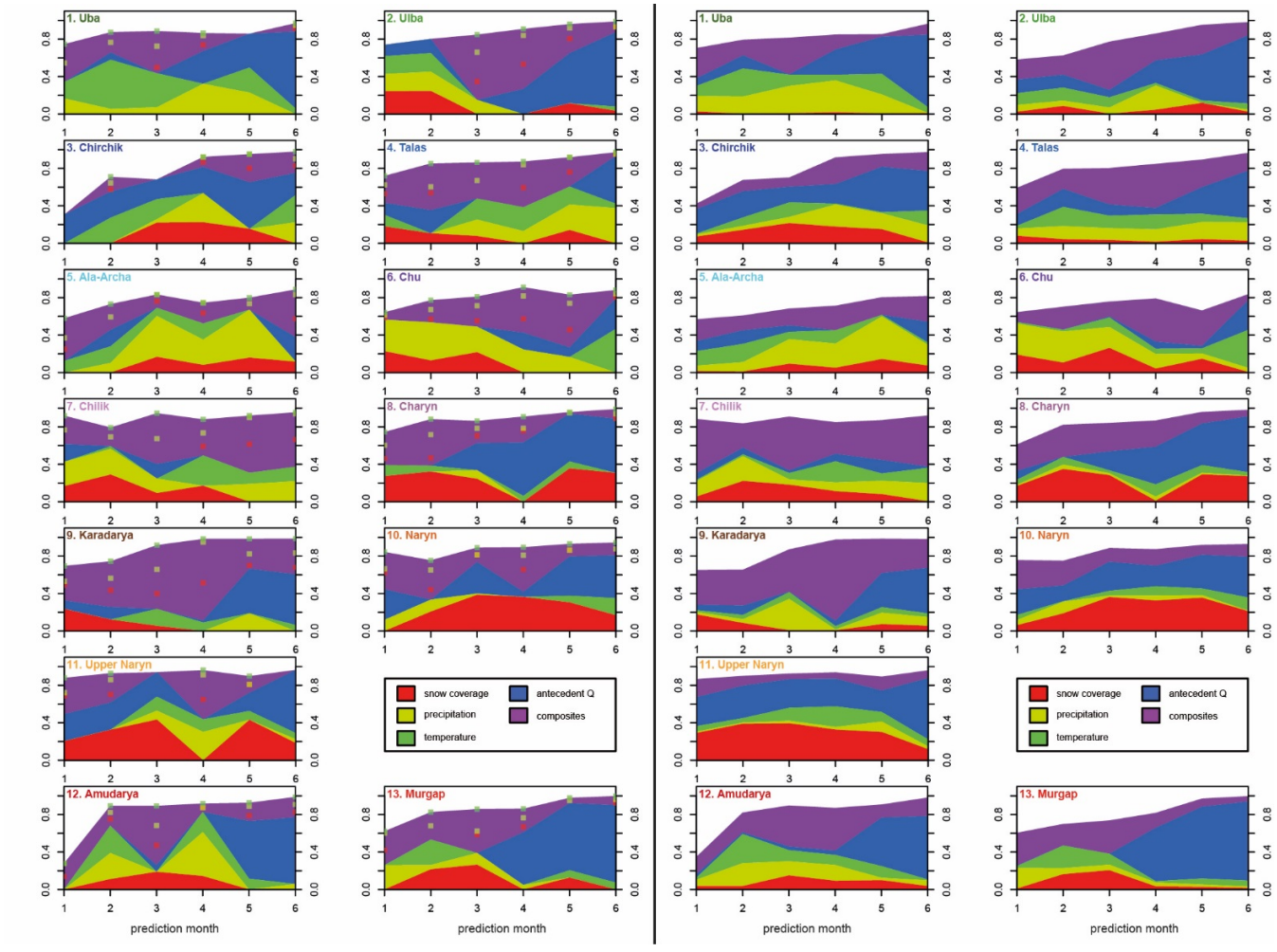

Figure 7: Importance of the predictors in the linear models as absolute contribution to the explained variance (R^2) for all catchments and prediction months. Left: of the best LOOCV model; Right: on average for the best 20 LOOCV models. Squares in the left panel figures indicate the presence of the 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 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

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 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 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 the hydromet services of the region. 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 of Squares (PRESS) 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 (R^2) is already high in the range of 0.6 – 0.88 for 11 catchments. For the following prediction on February 1st the explained variance is above 0.7 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, R^2 values are in the range 0.86 – 0.96, 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 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 can be used by the responsible 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. 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 fall this year.

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 (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.

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.

Annex

Annex 1: Predictors used for the different prediction dates

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