The Value of Hydroclimatic Teleconnections for Snow-based Seasonal Streamflow Forecasting in Central Asia

Atabek Umirbekov[1,4], Mayra Daniela Peña-Guerrero[1,4], Iulii Didovets[2], Heiko Apel[3], Abror Gafurov[3] and Daniel Müller[1, 4, 5]

[1] Leibniz Institute of Agricultural Development in Transition Economies (IAMO), Theodor-Lieser-Str. 2, 06120 Halle (Saale), Germany
[2] Potsdam Institute for Climate Impact Research (PIK), Telegrafenberg A 31, 14473 Potsdam, Germany
[3] GFZ Helmholtz Centre for Geoscience, Telegrafenberg, 14473 Potsdam, Germany
[4] Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
[5] Integrative Research Institute on Transformations of Human-Environment Systems (IRI THESys), Humboldt Universität-zu-Berlin, Unter den Linden 6, 10099 Berlin, Germany

Correspondence to: Atabek Umirbekov (umirbekov@iamo.de)

Abstract. Due to the long memory of snow processes, statistical seasonal streamflow predictions in snow-dominated catchments typically rely on snowpack estimates. Using mountainous catchments in Central Asia as a case study, we demonstrate how seasonal hydrological forecasts benefit from incorporating large-scale climate oscillations (COs). First, we examine the teleconnections between the major COs and peak precipitation season in eight catchments across the Pamir and Tian-Shan mountains from February to June. We then employ a machine learning framework that incorporates snow water equivalent (SWE) and dominant COs indices as predictors for mean discharge from April to September. Our workflow leverages an ensemble technique that uses multiple SWE estimates from near-time global data sources and diverse types of explainable machine-learning models. We find that the winter states of the El Niño-Southern Oscillation and the North Atlantic Oscillation enhance SWE-based forecasts of seasonal discharge in the study catchments. We identify three instances in which the inclusion of COs as additional predictors could be instrumental for snowpack-based seasonal streamflow forecasting: 1) when forecasts are issued at extended lead times and accumulated SWE is not yet representative of seasonal terrestrial water storage; 2) when climate variability during the forecasted season plays a larger role in shaping seasonal discharge; and 3) SWE estimates for a catchment are subject to larger uncertainty. Our approach provides a ~~novel~~ useful way to reduce uncertainties in seasonal discharge predictions in data-scarce snowmelt-dominated catchments.

1. Introduction

Snowmelt-driven streamflow is a vital source of water supply for downstream regions around the globe, ~~where it~~ sustainings ecosystems, agriculture, hydropower, and numerous other human activities (Immerzeel et al., 2020; Viviroli et al., 2007). ~~It is estimated that a~~Around two billion people ~~of the world`s population~~ lives in snow-sensitive basins (Mankin et al., 2015). Projections suggest ~~and it is projected~~ that around a quarter of the world`s lowland population will be critically dependent on snow- and glacier-melt runoff from mountains by the middle of the century (Viviroli et al., 2020). Accurate water ~~availability~~ supply forecasts are essential for the sustainability and resilience of water-dependent human and ecological systems in these regions.

Seasonal streamflow forecasts are usually generated using either ~~dynamic~~ process-based or ~~statistical~~ data-driven approaches. ~~Dynamic~~Process-based, dynamical forecasts~~, aka dynamical forecasts,~~ encompass a hydrological or land-surface model to estimate ~~current~~ initial hydrologic conditions, typically with the assimilation of observational data, followed ~~and~~by ~~use of~~ climate forecasts to project future conditions~~update and correct the resulting model state variables~~ (Troin et al., 2021). One major advantage of ~~dynamical~~process-based approaches is the continuous production of future streamflow states (Modi et al., 2022). On the other hand, ~~one of the main~~ a limitations of dynamical forecasts is their ~~their high computational demands and~~ dependence on spatially distributed meteorological variables obtained from numerical climate models, which ~~might be~~are prone to uncertainties. In addition, process-based approaches typically ~~exhibit~~ have higher computational demands. ~~Statistical~~ Meanwhile, ~~D~~data-driven~~, statistical~~ approaches ~~(aka statistical based relyapproaches)~~ rely on the empirical relationship between one or multiple predictor variables and seasonal streamflow. In this respect, ~~statistical~~ data-driven hydrological forecasts in ~~the context of~~ snow-dominated catchments offer advantages in terms of lower computational complexity and reliance primarily on initial hydrological conditions.

Because accumulated snowpack is a ~~primary~~ main source of predictability of river streamflow in snowmelt-dominated basins (Pechlivanidis et al., 2020), statistical forecasts of seasonal streamflow often primarily rely ~~solely~~ on accumulated snowpack~~.~~, with use of additional predictors that contribute to estimation of initial hydrological conditions. ~~Emerging evidence suggests that statistical-based forecast techniques that leverage initial hydrological conditions and large-scale climate indices could improve skill compared to current forecast approaches (Mendoza et al., 2017; Lehner et al., 2018).~~ North America has the longest history of systematically developing seasonal streamflow forecasts, also known in this domain as *water supply forecasts*, using empirical relationships between accumulated snow and spring-summer runoff. While snowpack explains most seasonal streamflow variability in western US basins, climate variability after the forecast issuance date has been identified as the main source of forecast error (Church, 1935, as cited in Pagano & Garen, 2010~~-~~; Schaake & Peck, 1985). Early attempts to use climate oscillation indices in water supply forecasts began in the 1970s, ~~with their integration~~integrating them into operational water supply forecasting at some agencies ~~occurring by~~at the beginning of the 2000s, though ~~,~~widespread adoption did not follow immediately (Pagano and Garen, 2010).

~~Apart from cases when catchments are influenced by strong teleconnections with large-scale climate oscillations (Mendoza et al., 2017), it remains unclear under what other conditions combining snow with climate indices offers benefits.~~

Previous research has shown that integrating climate indices generally improves seasonal streamflow forecasts. Studies from the U.S. suggest that the improvement may be more evident in long-lead forecasts, as climate indices tend to account for future climatic conditions after the forecast issuance dates (Grantz, Rajagopalan, Clark, & Zagona, 2005; Hamid & Matthew, 2010; Kalra, Ahmad, & Nayak, 2013; Kennedy, Garen, & Koch, 2009; Regonda, Rajagopalan, Clark, & Zagona, 2006). Similarly, evidence from parts of High Mountain Asia suggests that climate indices may be better predictors than snowpack for streamflow forecasting in snowmelt-dominated catchments at the beginning of winter (Charles et al., 2018; Umar et al., 2023). While confirming that snowpack is one of the main predictors, a multi-ensemble study on seasonal streamflow forecasting in the Andes also highlights the utility of large-scale ocean-atmospheric factors as additional predictors (Mendoza et al., 2014). However, the improvement from combining climate indices into snowpack-based streamflow forecasts depends on the strength of the teleconnections with large-scale climate oscillations (Mendoza et al., 2017; Opitz-Stapleton et al., 2007). In turn, relationships between large-scale climate oscillations and hydrometeorological variability maybe nonlinear and non-monotonic, making them challenging to capture with linear approaches (Fleming and Dahlke, 2014).

From a methodological perspective, data-driven seasonal streamflow forecasting has undergone two major transformations over the past decades. Historically dominated by the use of linear regression and its extensions, such as Principal Component Regression, the field increasingly adopts machine learning (ML) techniques. ML-based data-driven approaches ~~can~~ generally excel at leveraging diverse datasets, capturing non-linear relationships, and achieving ~~-~~higher predictive accuracy in seasonal streamflow forecasting (Sean W Fleming, Rittger, Oaida Taglialatela, & Graczyk, 2024; Kalra et al., 2013; Korsic et al., 2023). Another notable trend is the growing use of ensemble approaches because they generally offer higher prediction accuracy and better quantify prediction uncertainty compared to single-model methods (Murray, 2018; Zounemat-Kermani et al., 2021). A notable example that integrates these two trends for forecasting in snowmelt-dominated catchments is the multi-model machine-learning metasystem ("M4") in the western US, which uses an ensemble approach with multiple ML-based forecast models and pool outputs to generate a consensus prediction (Fleming et al., 2021). Furthermore, ~~(~~Najafi and Moradkhani~~,~~ (2016) explored techniques for combining outputs from multiple data-driven seasonal forecast models, providing valuable insights into best practices for ensemble forecasting. In addition, ensemble methods, including those based on ML, can be also more effective in addressing challenges associated with small datasets (Alzubaidi et al., 2023; Safonova et al., 2023; Dietterich, 2000). In this context, it is worth noting that observational data gaps -are common in mountainous regions of the Global South (Hock, R. et al., 2019).

Water is inextricably intertwined with the development challenges of Central Asia, yet its ~~timely~~ availability during the ~~vegetation~~ growing season remains erratic. The hydrological discharge in Central Asian rivers is subject to large seasonal temperature and precipitation cycles; the latter falls as snow in winter~~,~~ and its melting contributes to spring and summer runoff. The high variability of precipitation during the cold season ~~thus eventually determines~~ results in high interannual volatility of river streamflow in ~~Central Asian~~ the ~~many~~ endorheic rivers of Central Asia~~n~~ since most discharge originates from snowmelt in the Pamir and Tian Shan mountains (Viviroli & Weingartner, 2004). This high hydroclimatic variability ~~subsequently~~ underscores the need for improved water availability forecasting during the irrigation season (Xenarios et al., 2019).

Research on seasonal river discharge forecasting in Central Asia can be classified into two mainstream approaches. One approach explored the predictability of mean discharge ~~during~~ from April to -September (~~from now on~~hereinafter referred to as '~~vegetation~~ growing season") by using estimates of terrestrial water storage that accumulates in mountain catchments throughout the prec~~ee~~ding November to March (~~from now on~~hereinafter referred to as "cold season"). Terrestrial water storage in Central Asia is dominated by large annual cycles, with most precipitation during the extended cold season ~~that lasts~~ falling ~~from~~ from autumn to spring and accumulating ~~es~~ as snowpack in the mountain catchments. In the absence of in-situ snow water equivalent (SWE) data, several studies explored the use of proxies such as cumulative precipitation over the cold season (Dixon and Wilby, 2016; Schär et al., 2004), ~~or~~ satellite-derived snow cover ~~derived from satellite~~, antecedent discharge, and other predictors (Apel et al., 2019; Gafurov et al., 2016).

Another approach uses climate indices of global climate oscillations as predictors, some of which are known to have a noticeable impact on hydroclimate variability in Central Asia. It was found that El Niño-Southern Oscillation (ENSO) during its warm phase (aka El-Niño) increases precipitation intensity in Central Asia, most pronounced from autumn to summer (Mariotti, 2007; Chen et al., 2018). In contrast, the cold phase of ENSO ~~in its cold phase~~ (i.e. La-Niña) ~~is associated~~ contributes to ~~with~~ below-average precipitation in the region. The Pacific Decadal Oscillation (PDO) can intensify ENSO's effects: during ~~the ENSO's~~ La Niña phase, when the PDO is in its negative phase, Central Asia is more susceptible to severe droughts (Wang et al., 2014). The North Atlantic Oscillation (NAO), Scandinavian pattern (SCAN), and East Atlantic/Western Russia pattern (EAWR), ~~all of~~ which all are ~~refer to~~ periodic fluctuations in atmospheric pressure between specific regions of the Atlantic Ocean and Eurasia, ~~are also known to~~ also affect hydroclimatic variability in Central Asia (Syed et al., 2010). Several studies ~~previously~~ showed that indices of these climate oscillations can be used for forecasting seasonal precipitation (Gerlitz et al., 2019; Umirbekov et al., 2022) and streamflow in the region (Barlow and Tippett, 2008; Dixon and Wilby, 2019).

~~Both approaches have strengths and weaknesses. Using terrestrial water storage estimates as predictors produces accurate seasonal runoff predictions, though their accuracy gradually degrades with extending lead times (Apel et al., 2018). In contrast, climate indices offer seasonal hydrological outlooks well before the start of the vegetation season, though at the cost of higher uncertainties. Combining the strengths of both approaches into a two-tiered approach was suggested for forecasting seasonal~~

~~runoff (Gerlitz et al., 2020). Accordingly, early seasonal outlooks should employ large global oscillations to predict cold season precipitation anomalies. By the start of the vegetation season, seasonal hydrological projections should rely more on the abovementioned proxies of seasonal water storage. However, a common approach to seasonal hydrological forecasting typically builds on two main elements: initial hydrological conditions and future climate variability (WMO, 2021). From this perspective, combining the main predictors the two approaches rely on could be more appropriate: so that the accumulated catchment snowpack represents the initial hydrological conditions, and climate oscillation indices serve as precursors to climate variability during the targeted season.~~

Another challenge hampering the development of advanced forecasting techniques in the region is a scarcity of in-situ meteorological and hydrological observations, particularly for snow mass measurements. In the past, local hydrometeorological agencies ~~used to~~ conducted snow depth measurements across the region's main catchments. This practice was discontinued mainly due to the underfinancing of the relevant agencies that persisted for the past three decades (Xenarios et al., 2019). Satellite or reanalysis datasets available in near-real time can be an alternative source for estimating SWE. Still, they might be prone to inherent uncertainties and insufficient spatial resolution to capture variations of accumulated SWE (Mortimer et al., 2020), with larger errors in mountainous regions (Mortimer et al., 2024). Combining multiple satellite-derived or reanalysis estimates may improve snowpack estimation, thereby reducing streamflow prediction uncertainty (Oğulcan Doğan et al., 2023; Mortimer et al., 2020). ~~in complex terrain. A promising option to reduce uncertainty in modelled products is to apply an ensemble technique instead of relying on a single model estimate (Murray, 2018; Zounemat-Kermani et al., 2021).~~

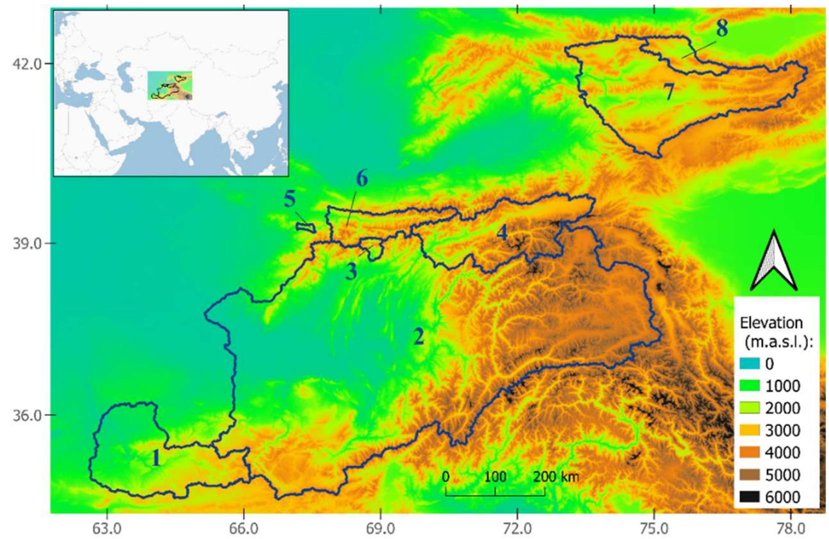
This paper ~~aims to~~ ~~introduces and~~ tests a new framework for seasonal streamflow ~~statistical~~ forecasting~~s of~~ ~~seasonal river discharge~~ in ~~snow-dominated catchments~~ Central Asia, which ~~by coupling~~combines catchment SWE estimates ~~of snow water equivalent~~ with climate oscillation indices that condition hydroclimatic variability in the upcoming season. Given the region's observational data gaps, this study also evaluates the utility of ~~Our approach relies on an ensemble of diverse estimates of catchment-averaged~~ SWE derived ~~or simulated using multiple~~from global reanalysis and satellite products for hydrological forecasting in high-elevated catchments. ~~supplemented with climate indices that affect hydroclimatic variability across Central Asia.~~ We used ~~generalized~~ generalised linear regression and machine learning techniques, such as Random Forest, Gaussian Process, and Support Vector Regression, which produce a range of individual forecasts. Finally, we employ an ensemble stacking approach, a type of ensemble learning that uses the forecasts from individual models as inputs to a model that produces a more reliable final prediction.

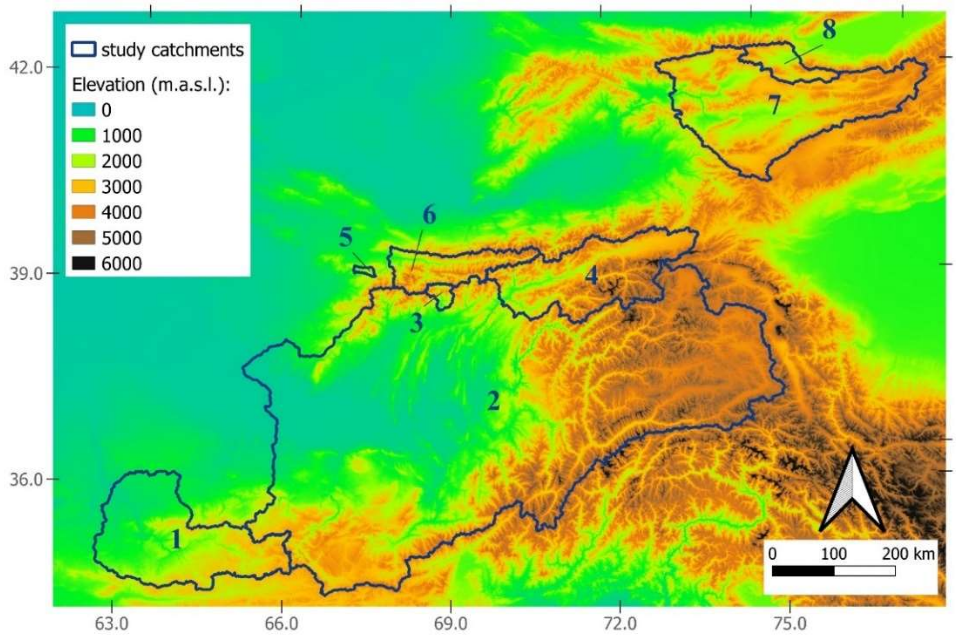
2. Study Area

The study area encompasses eight diverse snowmelt-dominated catchments in the Pamir, Hindukush, and Tian-Shan mountains (Figure 1, Table 1). The size of the selected catchments varies from 343 to 296,000 km^2, and the mean catchment altitude ranges from 1,700 to 3,500 meters. The catchments include the largest rivers in the region, the Amudarya and Naryn (the main tributary of the Syrdarya), which embed several smaller tributary sub-catchments. For reference, the Figure 1 also depicts the annual precipitation cycles in the study catchments, which in the absence of in-situ measurements were estimated using Climate Terra data. (Abatzoglou et al., 2018). It should be noted that while gridded precipitation products consistently capture annual climatology, they may exhibit significant bias in mountainous areas (Hu et al., 2018; Peña-Guerrero et al., 2022).

Table 1. Major geographical and ~~climatic~~ hydrological characteristics of the study catchments.

#	Catchment	Gauging station name	Station location (lat, long)	Catchment area (km^2)	Catchment mean altitude (m.a.s.l)	Mean Apr-Sep discharge during 2000-20018 (m^3/sec)	Coefficient of variation of discharge during 2000-2018
1	Murghab	Takhta Bazar	35.96, 62.91	35,582	1,710	41	0.49
2	Amudarya	Kerki	37.84, 65.23	296,300	2,550	1,876	0.29
3	Varzob	Dagana	38.70, 68.79	1,279	2,700	79	0.23
4	Vaksh	Komsomolabad	38.86, 69.94	28,908	3,530	996	0.1
5	Kashkadarya	Varganza	40.81, 73.26	343	2,663	18	0.35
6	Zarafshan	Dupuli	39.49, 67.80	10,310	3,125	243	0.17
7	Naryn	Toktogul	41.77, 73.29	46,667	2,940	561	0.22
8	Chu	Kochkor	42.25, 75.83	5,305	2,934	35	0.31


42.0
39.0
36.0
63.0
66.0
69.0
72.0
75.0
78.0
1
2
3
4
5
6
7
8
Elevation (m.a.s.l.):
0
1000
2000
3000
4000
5000
6000
0
100
200 km


study catchments
Elevation (m.a.s.l.):
0
1000
2000
3000
4000
5000
6000
42.0
39.0
36.0
63.0
66.0
69.0
72.0
75.0
1
2
3
4
5
6
7
8
0
100
200 km

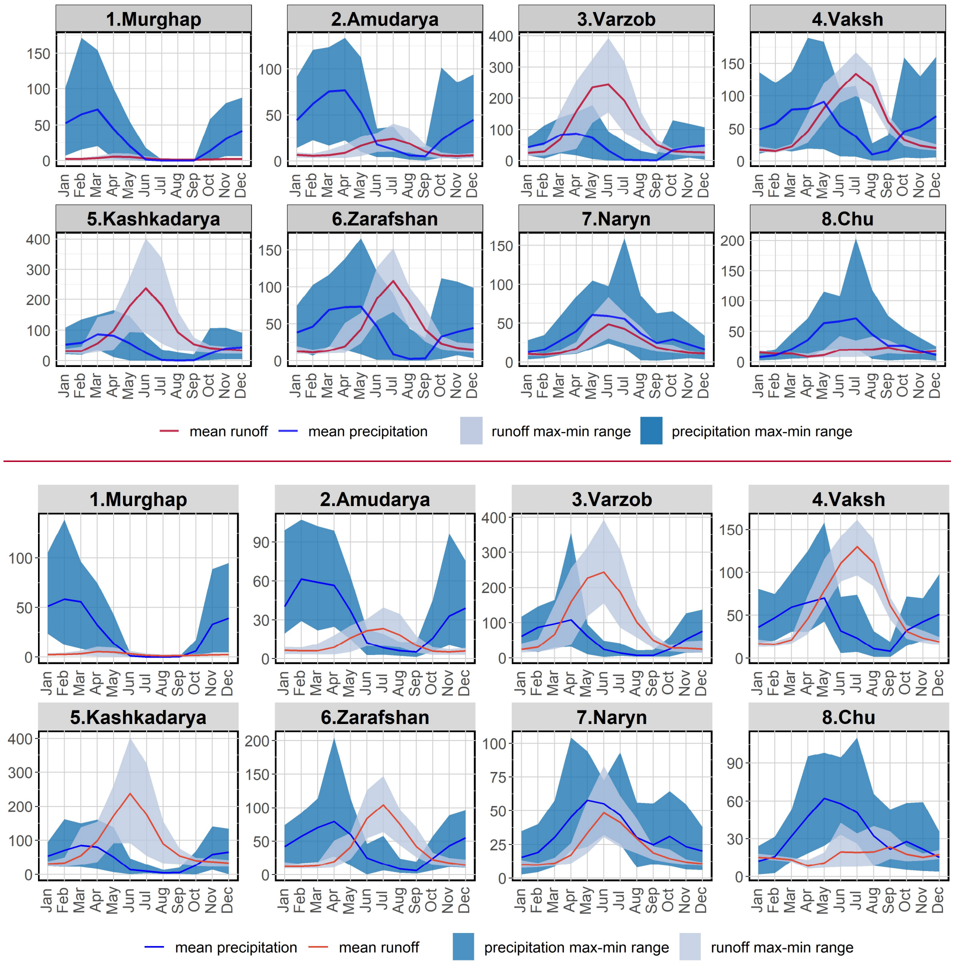


Figure 1. Location of the study catchments (upper map), monthly means and ranges for precipitation and runoff for 2000-2018 in mm (bottom graphs).

3. Data

The predictand~~t~~ variable represents seasonal discharge, calculated as the mean discharge from April to September. We obtained monthly discharge data for the study catchments from 2000 to 2018 from hydrometeorological agencies in Central Asian countries. We aggregated these into mean discharge from April to September, resulting in approximately eighteen observations of seasonal discharge for each catchment.

Operational seasonal streamflow forecasting in snowmelt-dominated catchments primarily relies on snow measurements~~,~~ but usually also incorporates additional variables that reflect initial hydrological conditions, such as accumulated precipitation, antecedent flow, and basin soil moisture. Due to the limited availability of streamflow observations and the overall objective of assessing the added value of teleconnections compared to snowpack-only predictions, we restricted the predictors to two groups: SWE and teleconnections.

As the primary predictor variable, we use four basin-averaged SWE estimates that can be derived from ~~near-real time~~ global climate datasets available in near real~~-~~time (Table 2~~Table 2~~). These include two SWE estimates from global and regional reanalysis datasets, i.e. ERA5-Land (Muñoz-Sabater et al., 2021) and Land Data Assimilation System Central Asia (McNally et al., 2022). In addition, we ~~simulated~~ obtained two SWE estimates using the GEMS snow model (Umirbekov et al., 2023) forced by global precipitation and temperature data available in near-real time. One simulated SWE time series is obtained by forcing the snow model with the Multi-Source Weather dataset, ~~which is~~ generated by bias-correcting and downscaling ERA5 (Beck et al., 2021). The fourth SWE estimate is simulated using precipitation estimates from the Integrated Multi-satellite Retrievals GPM IMERG v6 (Huffman et al., 2019) and temperature estimates from MSWX. We used a 'Late Run' version of GPM IMERG precipitation estimates, accessible in near-real time albeit lacking adjustments using ground precipitation data as in the 'Final' product, which becomes available two months later.

Candidates for additional predictors include the monthly indices of the El Niño–Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO), the North Atlantic Oscillation (NAO) and the Scandinavian Pattern (SCAN).

Table 2. Snow water equivalent estimates and climate oscillation indices that were used as predictors in this study

Type	Predictor (abbreviation)	Description	Source
Snow Water Equivalent estimates	ERA5-L	Retrieved from the ERA5-Land reanalysis dataset	Muñoz-Sabater et al., 2021
	FLDAS	Retrieved from the Land Data Assimilation System Central Asia	McNally et al., 2022
	MSWX	Simulated using GEMS model forced by precipitation and temperature estimates from MSWX dataset	Beck et al., 2021
	GPM	Simulated using GEMS model forced by GPM IMERG precipitation~~,~~ and MSWX temperature	Huffman et al., 2019
Climate Oscillation Indices	SOI	Southern Oscillation Index	Ropelewski and Jones 1987
	PDO	Pacific Decadal Oscillation	Mantua et al 1997
	EAWR	East Atlantic/West Russia pattern (EAWR)	Barnston and Livezey 1987
	NAO	North Atlantic Oscillation (NAO)	Barnston and Livezey 1987
	SCAN	Scandinavian pattern	Barnston and Livezey 1987

The hydrological dataset used in this study comprises 18 seasonal discharge observations per study catchment, spanning 2000 to 2018. This highlights the data availability limitations in the region. To address the challenges posed by this small dataset, we employed the aforementioned approach, which integrates multiple diverse estimates of SWE~~,~~ and utilises an ensemble methodology described below in section 4. "Methods". It is worth noting that historical data for most catchments also spans from the 1970s to the 1990s, with relatively more complete records available for the largest rivers extending up to 2000. However, incorporating these earlier records is challenging, as the datasets used to derive some SWE estimates (FLDAS and GPM) are only available from 2000 onward. Extending the observations back in time would restrict the ensemble to only ERA5-L and MSWX datasets, thereby reducing its diversity and compromising its robustness. In addition, using older discharge records in our framework may be problematic due to the non-stationarity of climate and hydrological systems (Pagano and Garen, 2005; Livneh and Badger, 2020), as well runoff alterations in some large basins induced by land-use changes over the past century (Hou et al., 2023).

4. Methods

4.1 Determining associations between climate oscillations and hydroclimatic variability across study catchments

To determine linkages between the selected climate oscillations and hydroclimatic variability across the catchments, we calculated Spearman's rank correlations with precipitation during months with higher magnitude and interannual variability. We used the ~~CHELSA-W5E5~~global TerraClimate precipitation dataset (~~(~~Abatzoglou et al., 2018~~)Karger et al., 2022~~) to construct ~~a~~ catchment-averaged precipitation time series from 1979 to 2020~~16~~. The annual precipitation cycle in the studied catchments exhibits two distinct sub-regional patterns (see Figure 1~~Figure~~). Catchments in the Pamir and western Tian-Shan experience increasing precipitation during winter, peaking in the spring, and decreasing during summer. In contrast, the Naryn and Chu catchments, located in the interior and northern Tian-Shan, receive most precipitation from late spring to early summer and less precipitation in winter. Across all catchments, the interannual variability is greatest during the months with the highest precipitation totals. We have defined the ~~common~~ standard peak precipitation season for the region as February to June~~,~~ since this period covers the months with the highest precipitation ~~levels~~ amounts and the greatest interannual variability across all the studied catchments. We then calculated Spearman's rank correlation coefficient between the catchment averaged precipitation for February-July (referred to here as 'peak precipitation season') and each climate oscillation index at varied lead-lag times. To identify when oscillations show the strongest association with the precipitation season, we calculated the correlation for each oscillation index from August of the preceding year to July, the final month of the peak precipitation season. In addition, we ~~calculated~~ computed correlations between the climate indices and mean discharge during the ~~vegetation season~~growing season using the same procedure.

4.2 Stacked ensemble-based prediction of seasonal discharge

Our ~~forecast~~ modelling framework employs an ensemble stacking approach. This machine learning technique combines predictions from multiple base models and uses them as inputs to a higher-level meta-learner or stacking model. This approach has seen increasing application in the hydrological field in recent years (Zounemat-Kermani et al., 2021), including long-term streamflow forecasting, drought monitoring, and real-time flood forecasting (Mallick et al., 2022; Granata and Di Nunno, 2024; Li et al., 2019; Xu et al., 2024).

The ensemble stacking workflow consists of four main steps depicted in ~~(~~Figure 2~~Figure)~~. In the first stage, we combine each basin-averaged SWE estimate with climate oscillation indices at months when they exhibit higher association with in-season precipitation peaks. Four different SWE products result in four datasets with varying SWE estimates but the same set of selected climate oscillation indices for each catchment. Any set of predictors for each basin includes a maximum of three variables: one SWE estimate and up to two climate oscillation indices.

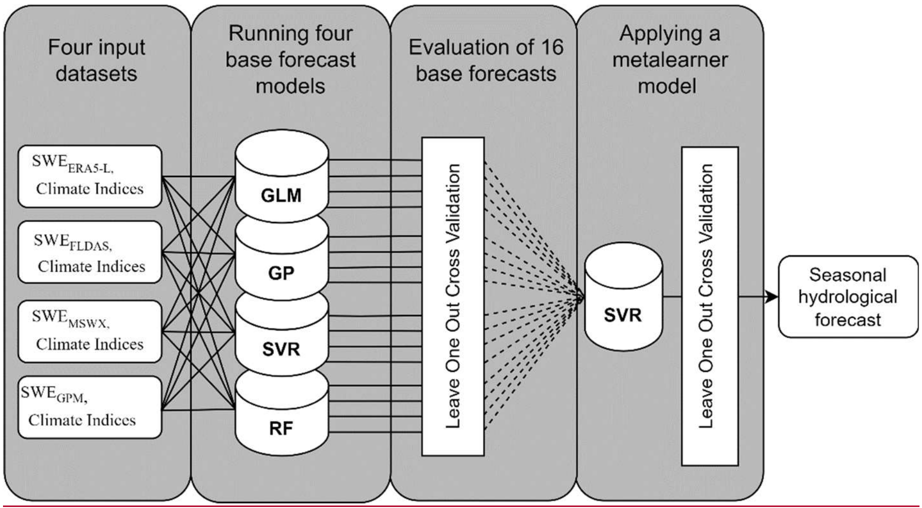


Figure 2. Workflow of the ensemble-based forecast approach

We then use four different forecast models (from now on referred to as "*base models*"), each forced with the four input datasets to produce a range of 16 seasonal forecasts. The four base models are comprised of the generalised linear model (GLM) with Gaussian identity link, gaussian process regression with the linear kernel (GP), support vector regression (SVR) with the linear kernel (SVR) and random forest (RF). The latter two model algorithms have parameters that control internal model complexity. For example, the "*cost"* parameter in SVR limits training errors against maximising the margin of the decision function, and "*mtry"* in RF determines the number of predictors that can be taken into account at each split point of a single tree. We ~~set~~ confine these parameters to relatively lower levels (cost=0.3 in the case of SVR and mtry=1~~2~~ in the case of RF), which helps to avoid overfitting and facilitates a higher degree of generalizability (Najafi and Moradkhani, 2016; Safonova et al., 2023).

In the next step, we evaluate each of the 16 base model predictions using leave-one-out cross-validation (LOOCV), which is well-suited for the small dataset context. It is worth noting that LOOCV is a standard practice in developing and evaluating water supply forecasting models in the western U.S. (Fleming and Garen, 2022). Rather than using all 16 base model forecasts in subsequent steps, we apply a threshold that filters out weaker predictions. This threshold requires a leave-one-out cross-validated R-squared coefficient of base model performance to be greater than 0.2 to be considered for further analysis. This threshold was optimal during LOOCV in terms of predictive performance for the stacking ensemble.

In the final step, those base model predictions that pass the LOOCV test become inputs for a final forecast model (from now on called "*ensemble model*"). Since all selected base model predictions would exhibit some degree of correlation among themselves, we employ the SVR algorithm as a meta-learner model, which is known to be less sensitive to multicollinearity (Farrell et al., 2019). The final prediction of the meta-learner model is again assessed using LOOCV. ~~Finally, we validate the resulting model using a few observations of seasonal discharge, which were held out during the training of the models.~~

We apply the procedures described above for each standard forecast issue time adopted by hydrological agencies in Central Asia, starting from January 1st, that is a three-month lead time concerning the April-September season, and ending with the final forecast issued just before the start of the season, i.e., on April 1st. Each forecast uses inputs that are accessible by its issue date. For example, three-month lead forecasts can only use estimates of catchment SWE by January 1st and state of climate oscillations in previous months. To attain parsimonious forecast models, rather than incorporating all studied climate oscillation indices into a set of predictors, we followed a stepwise approach: each of the climate indices was added one at a time to the predictors set, which was then evaluated. This approach led to a final predictor set with the minimal combination necessary to produce plausible predictions for each catchment and each forecast issue date.

4.3 Determination of supplementary importance of incorporating climate oscillations as additional predictors

We implement two track evaluation analyses to determine the value of adding COs as additional predictors into snow-based forecasts. First, we elaborate forecast models that ~~assimilate~~use only SWE estimates as predictors, using the same approach described in the previous section, and compare their performance with those that ~~assimilate~~ use both SWE and COs. Second, we determine the relative importance of COs using the *feature importance ranking measure* method (Greenwell et al., 2018), which quantifies how much each input variable influences the predictions made by the model. The method assesses the impact of each input variable by estimating partial dependence plots (Friedman, 2001) and assigning higher (lower) importance rank to features that exhibit a steeper (flatter) partial dependence effect.

~~4.~~5. Results

5.1 Evaluation of SWE estimates

~~**Figure 3.** Pearson`s correlation coefficients between the SWE estimates and mean seasonal discharge between April and September at different forecast lead months. The red line is the median across all snow products.~~ Figure 3~~Figure 3~~ ~~summarizes~~ summarises the correlation coefficients between catchment-averaged SWE at different forecast issue dates and mean discharge during the ~~vegetation season~~growing season. The SWE estimates obtained from global reanalysis and satellite data exhibit varied degrees of connection with the seasonal discharge. For all catchments, the correlation in general tends to increase with shorter lead times, i.e., with SWE estimates for January 1st having the lowest correlation and those for April 1st having the highest. ~~There is no discernible best-performing product across the four SWE estimates overall, with some products better performing in one catchment but underperforming in another. Nevertheless,~~ While SWE estimates based on ERA5-L and MSWX generally show ~~-~~a higher correlation with seasonal discharge across most catchments, ~~though~~ in the absence of in-situ snow measurements, it is impossible to assert which of the four SWE estimates is relatively more consistent.~~.~~

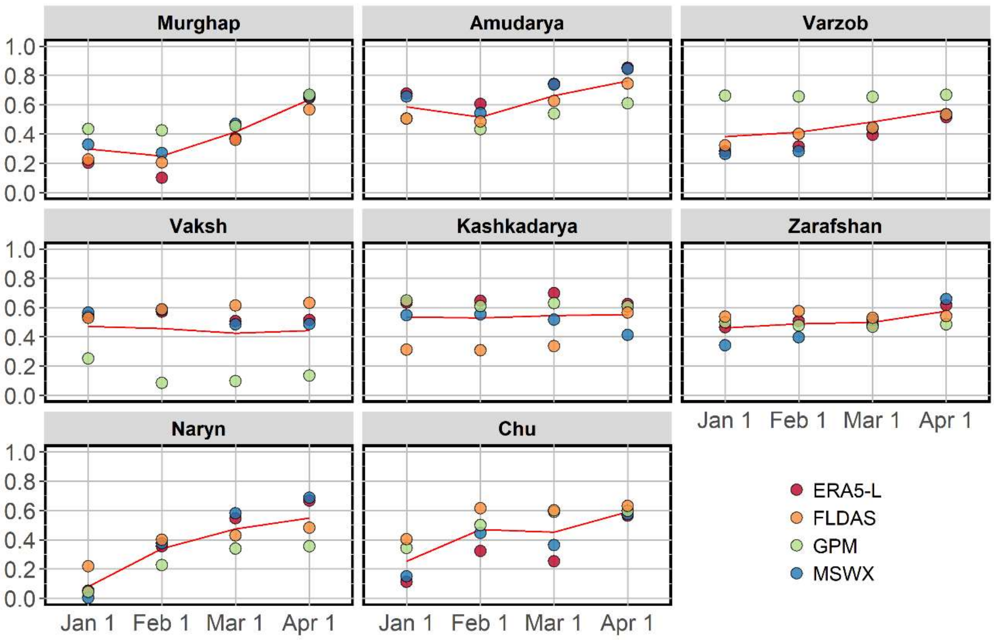


Figure 3. Pearson`s correlation coefficients between the SWE estimates and mean seasonal discharge between April and September at different forecast lead months. The red line is the median across all snow products.

5.2 Association between climate oscillations and hydroclimatic variability across the study catchments

Evaluation of the climate oscillation indices revealed diverse associations with peak season precipitation and mean river discharge during the ~~vegetation season~~growing season across the catchments (Figure 4~~Figure 4~~, upper graph). In all catchments, the February-July precipitation exhibits a robust and persistent association with ENSO, represented by the Southern Oscillation Index (SOI) and the Pacific Decadal Oscillation (PDO), over an extended timeframe compared to other oscillations. ~~There is a significant negative correlation~~ A significant negative correlation exists between peak precipitation season and SOI in all catchments, evident three months before the season's commencement. This relationship persists for a longer duration compared to any other climate oscillation. On the other hand, PDO exhibits a positive link with seasonal precipitation, becoming noticeable as early as four months before the season's onset and reaching its most substantial level in November. The selected lead months of SOI and PDO exhibit ~~higher correlation between each other, possibly because the latter also mirrors the ENSO phenomena~~ a higher correlation, possibly because the latter also mirrors the ENSO phenomenon.

Like ENSO and PDO, the East Atlantic/West Russia pattern (EAWR) consistently demonstrates a stronger correlation across most catchments before the peak precipitation season. Notably, the October state of EAWR shows a substantial positive correlation with peak precipitation across all catchments; however, it becomes more variable as the season progresses.

On the other hand, the North Atlantic Oscillation (NAO) and the Scandinavian Pattern (SCAN) show a relatively less pronounced association with the peak precipitation season, with correlations that vary depending on the lead time. From December to March, the NAO shows a weak but persistent negative relationship with the peak precipitation season in most catchments. In January, at the beginning of the peak precipitation season, a considerable portion of the catchments demonstrates a negative correlation with the state of SCAN. However, as the season progresses and reaches March, there is a noticeable shift, with all catchments showing a stronger and more positive correlation with the state of SCAN.

The correlation between the climate indices and mean river discharge during April-September exhibits almost the same pattern (~~Figure 4~~Figure S1 in the Supplement~~, bottom graph~~)~~, which~~. This implies that interannual discharge variability is predominantly driven by precipitation ~~that falls during~~ between February and July.

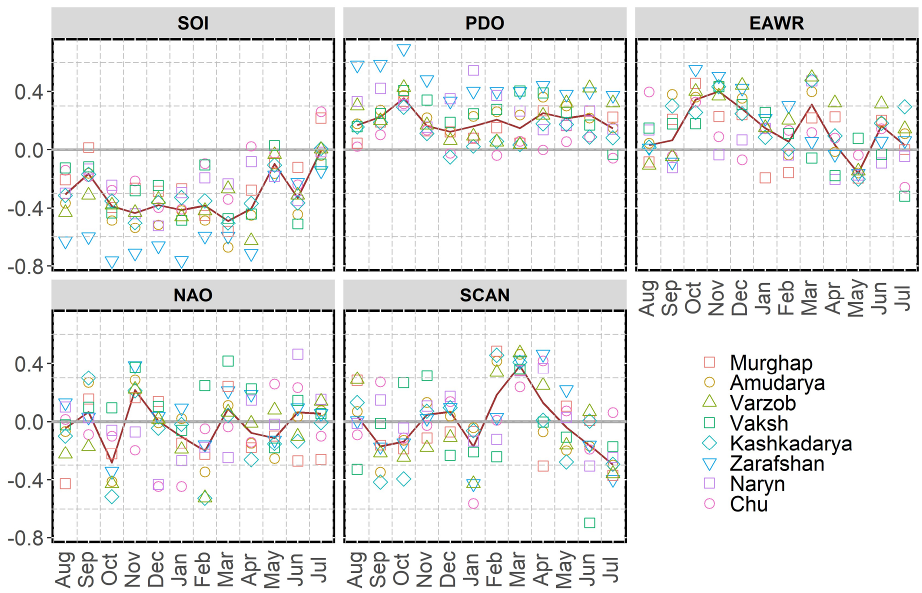


Figure 4. Spearman's rank correlation coefficients between the climate oscillation indices and precipitation from February to July precipitation ~~(top) and mean seasonal discharge from April to September (bottom)~~. The X-axis denotes months of a climate index. The red line represents a median for correlation coefficients across all catchments in each month.

5.3 Performance of seasonal discharge forecasts

Figure 5~~Figure 5~~ below ~~summarizes~~ summarises a set of final predictors per studied catchment, obtained after screening COs associations with peak precipitation and mean discharge during the ~~vegetation season~~growing season and following a stepwise selection procedure using the ensemble-stacking forecast approach described in section 4.2.

While the input dataset for the base models included SWE estimates, the combination of climate oscillations they rely on varies depending on a catchment location and elevation. In most catchments, ~~the base models~~there is a higher –correlation between~~incorporated~~ the late autumn state of PDO ~~or~~ and the ~~Decembe~~winter state of SOI. To avoid redundancy and potential issues with multicollinearity, we did not include both indices as predictors in the models for the same basin. Instead, ~~either~~ PDO or SOI was selected for each basin's model based on which index exhibited a stronger predictive relationship. ~~While~~As a ~~SOI~~result, SOI mostly appears ~~equally crucial in most catchments~~as ~~predictor in Tian-Shan catchments,~~ a predictor in Tian-Shan catchments, and PDO generally ~~persisted~~ persists as a predictor in catchments located in Pamir ~~and west Tian-Shan~~. The winter state of NAO and SCAN are another source of predictability in many of the catchments but have variable temporal signatures. In the case of the Murghap, where workable base models were obtained only for the April 1st forecast, they rely solely on SWE estimates and ~~do not include any of the climate oscillations~~SCAN as predictors.

Selected ~~NAO or SCAN predictors~~climate indices tend to have the same temporal lags for neighbouring catchments. For instance, the Naryn and Chu catchments in the Tian-Shan, which have similar seasonal precipitation patterns, use the NAO condition in January~~December~~ as one of their predictors. The Varzob and Zarafshan~~Vaksh~~ rivers, ~~which are~~ both high-elevation tributaries of the Amudarya, use the January state of SCAN~~, while the latter becomes a more robust forecast for the larger Amudarya watershed only one month later, in February~~.

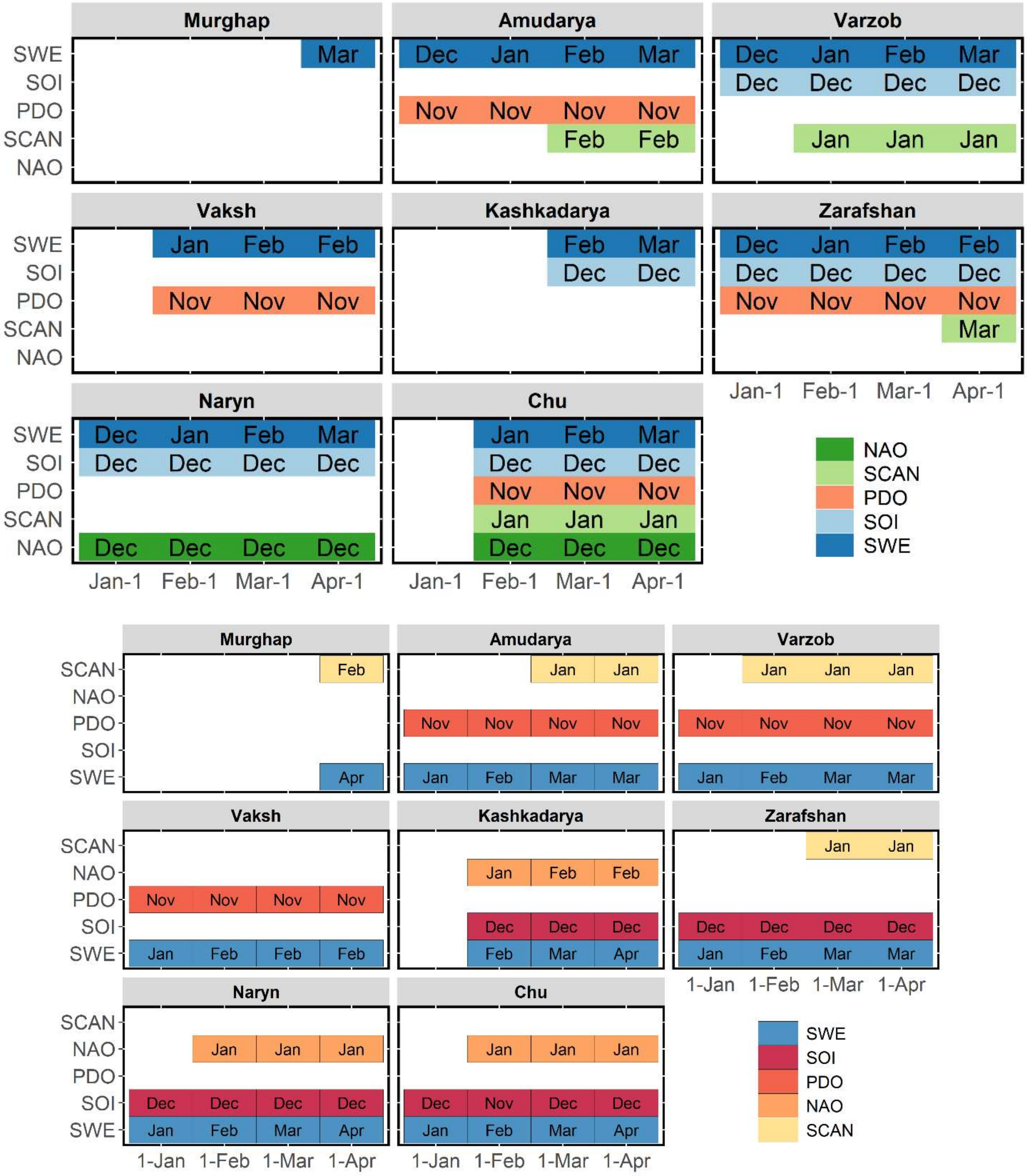


Figure 5. Predictors at different forecast issue times. Abbreviations within boxes indicate the month of the respective climate oscillation index or catchment-averaged SWE used as predictors. For example, the April 1st forecast models for the Amudarya ~~river~~ use as predictors the SWE estimate as of the beginning~~end~~ of March, the state of the PDO index in November, and the SCAN index in February.

The ensemble-based forecasting framework plausibly simulated seasonal discharge across all catchments, albeit with varying temporal performance based on lead time (Figure 6~~Figure 6~~). The ensemble model's LOOCV R-squared coefficient ranges between 0.2~~3~~ and 0.5 for the extended lead time forecast (1 January). It ~~but~~ gradually increases with decreasing lead time, surpassing~~reaching~~ ~~0.8 and~~ 0.9 for the April 1 forecast for most catchments except Murghap and Varzob. The ~~performance~~ accuracy of the meta-learner model forecasts depends on the number and diversity of the resultant individual base models~~and~~, which is ~~typically~~ superior to those of the latter. This underscores the strength of ensemble approaches, which ~~generally tend to~~ outperforms single-model approaches, as demonstrated in similar studies (Hagedorn et al., 2005; Najafi and Moradkhani, 2016; Fleming et al., 2021).

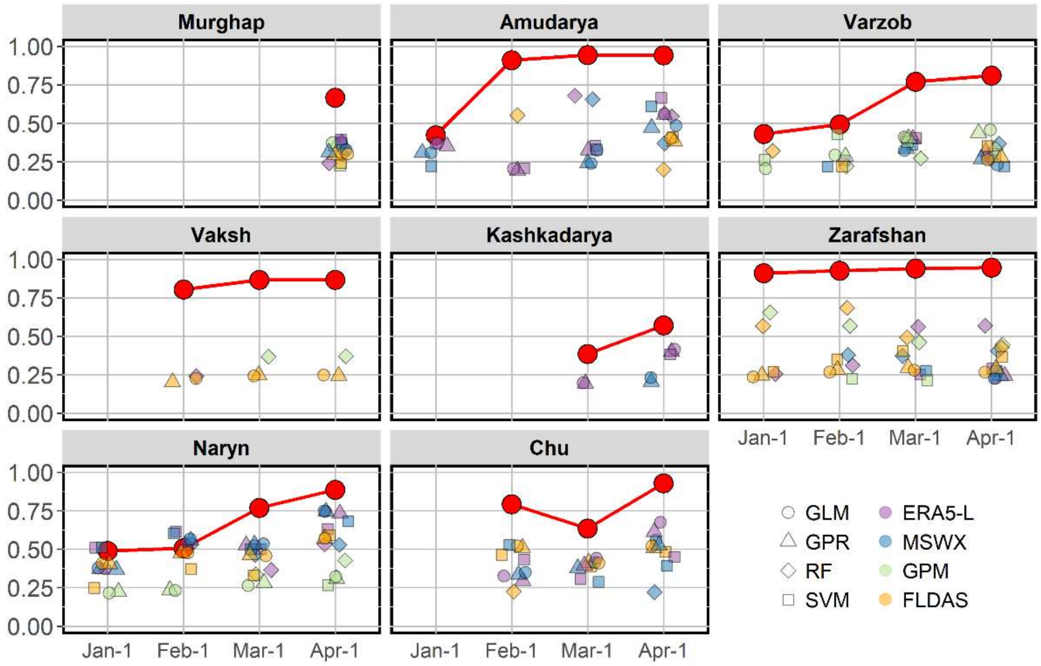
Due to the threshold criterion (R2>0.2) for a base model to be included in the final ensemble prediction, the resulting stacked ensembles typically consist of fewer than 16 base models. We observe two trends in this regard: (1) the later the issue date, the greater the number of base models included in the ensemble, and (2) larger catchments tend to incorporate more base models~~. For~~ for certain rivers, such as the Varzob, -Kashkadarya and Chu, this results in fewer base models being used for ensemble stacking.

~~Fewer base models appear capable of predicting seasonal discharge at longer lead times, resulting in a relatively lower performance of the meta-learner model on the January 1^{st} forecast.~~ For the Murgha~~pb~~ and ~~,~~Kashkadarya~~, and Chu~~ catchments, no feasible base models were obtained for the January 1^{st} forecast. Furthermore, workable base models and the derived meta-learner model for Murghap are only obtainable for the April 1^{st} forecast.

~~There are no discernible winners~~ ~~model types that are consistently superior~~ ~~in terms of~~ No model types are consistently superior in ~~performance~~ accuracy ~~of the base model types~~ across all lead times, especially for the final (April 1^{st}) forecast. However, the base models' performance has some distinct spatial heterogeneity, depending on which SWE product they ~~assimilate~~use. For example, all base models for the ~~Kashkadarya~~ Vaksh retained after cross-validation rely mostly~~only~~ on SWE_{ERA5-L} or SWE_{MSWX} as inputs. In contrast, ~~all~~ forecasts~~base models~~ ~~obtained~~ for the ~~Vaksh~~ Varzob catchment ~~rely only~~have ~~higher number of base models using~~ ~~on~~ more base models using SWE_{FLDAS} and SWE_{GPM}. The seasonal discharge in the largest catchments, such as Amudarya and Naryn, is also better explained by base models that use SWE_{ERA5-L} or SWE_{MSWX}.

The results suggest that models incorporating GPM IMERG have higher uncertainty, reflected in overall lower cross-validation performance, except in the highly elevated Varzob and Zarafshan catchments. This is likely due to the lower accuracy of the GPM IMERG`s Late Run product, which includes only climatological adjustment. In contrast, its final product ("Final Run") comes with adjustments using gauge data. However, the latter is only available at a three-month latency~~ time~~, precluding its operational forecasting use.

We tested several other ML techniques as base models, including using the same models with non-linear kernels. In most cases, the presented combination of models yielded a better accuracy ~~in terms of~~regarding MAE and R-squared coefficients during LOOCV. ~~In some instances~~Sometimes~~, depending on the basin or issue date, certain non-linear models produced slightly better predictions~~, certain non-linear models produced slightly better predictions depending on the basin or issue date. However, ~~when generalizing across all basins and issue dates, the existing structure still showed superior performance in terms of accuracy~~the existing structure still showed superior accuracy when generalising across all basins and issue dates. We assume this may be due to two major and non-exclusive factors: (1) ~~a relatively smaller~~ fewer ~~number of~~ observations and predictors, which makes non-linear machine learning models less efficient and prone to overfitting, and (2) the selection of predictors based on a linear metric (Pearson's correlation) may have inherently ~~favored~~favoured linear models.



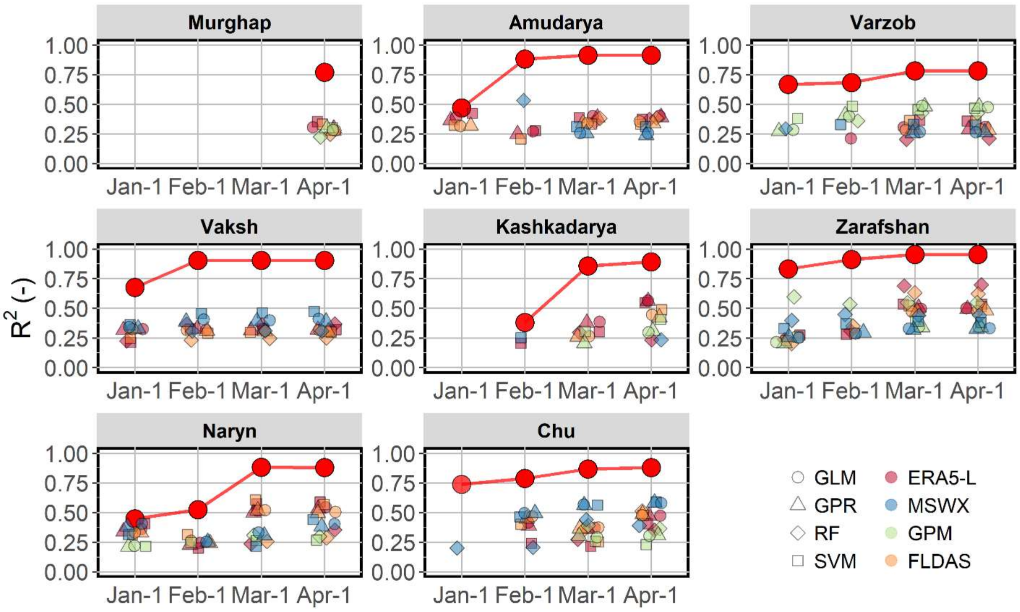


Figure 6. ~~Resulted~~ LOOCV R-squared coefficients of individual base models at different lead months and the LOOCV R-squared of the meta-learner model (red line).

~~The performance of the meta-learner model also yielded comparable results for both LOOCV validation and hold-out test set in terms of normaliszed Mean Absolute Error (Figure 7, upper graph). The normaliszed MAEs follow the same temporal pattern as the R-squared coefficient, with relatively larger errors for early forecasts that gradually decrease with decreasing lead times. Due to a limited number of discharge observations, the hold-out included only three last observations for seasonal discharge (which even became two in the case of the Naryn and Zarafshan catchments). In many instances, the hold-out set across the catchments contained a diverse combination of high and low observed seasonal discharge, which were plausibly predicted by the final meta-learner model (p). For the Kashkadarya and Chu rivers, the model correctly predicted the lowest observed seasonal discharges over the entire observation period, which were in the hold-out set. Similarly, the final meta-learner model coherently reproduced lower discharges from hold-out sets in the Amudarya andNaryn rivers below the 25th quantile of all observation years.~~

~~Figure 7 (bottom graph) also compares the performance of the forecast models based only on SWE estimates. SWE-only forecasts have larger uncertainties, with higher MAE errors for the LOOCV validation~~~~With higher MAE errors for the LOOCV validation, SWE-only forecasts have larger uncertainties, especially the hold-out test set. The mean absolute errors also tend to be larger at shorter lead times, which may imply higher uncertainties in SWE estimates during the snow-melt phase. In addition, we could not develop effective meta-learner models for extended lead times in some catchments in certain catchments, we could not develop effective meta-learner models for extended lead times.~~

The inclusion of climate indices generally enhances forecast accuracy across most catchments, as reflected in the generally lower normalized MAEs for models that combine SWE and climate indices compared to those that use only SWE (Figure 7). However, there are exceptions, such as in the February forecast for Kashkadarya and the January forecast for Vaksh, where SWE-only models exhibit lower errors. The improvement from including climate indices is particularly evident in catchments situated in the Tian Shan mountains, such as Naryn and Chu, where SWE-only forecasts result in substantially higher errors. Moreover, the difference in MAEs between the two model types becomes more pronounced with reduced lead times in these catchments. A similar pattern is observed in the high-elevation Zarafshan catchment. In contrast, in Amudarya and Murghap, large and relatively low-elevation catchments located in the Pamir region, the incremental differences between the two model types for the April 1 forecast are minor or absent, suggesting that the inclusion of climate indices provides limited added value in these cases.

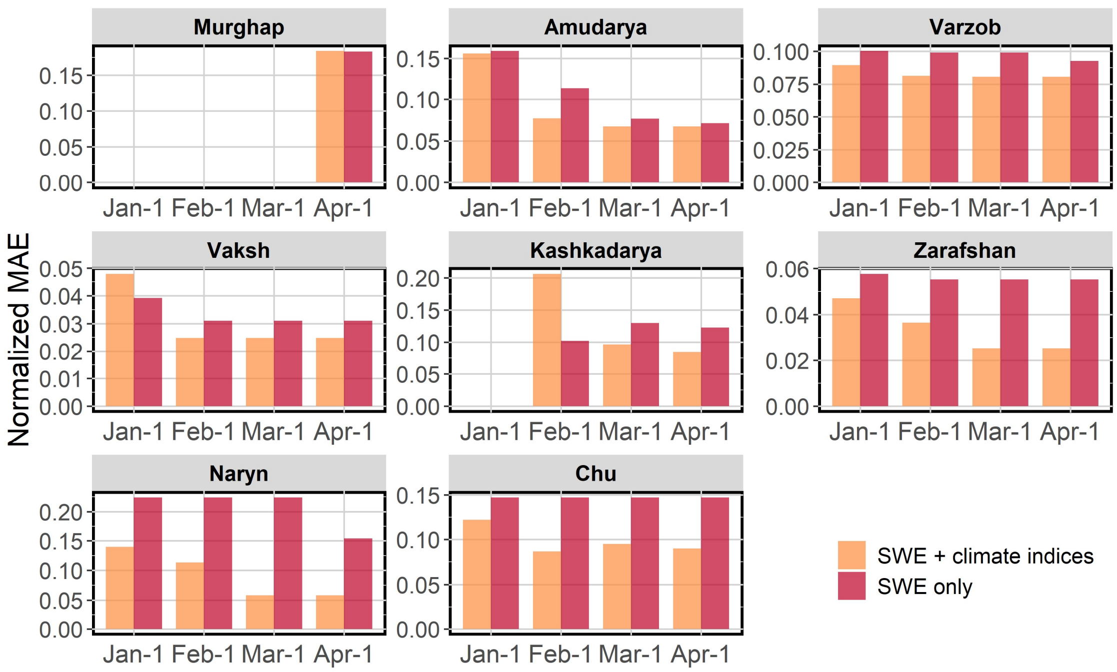


Figure 7. ~~**Normalized**~~ ~~**Normalised**~~**Normalized MAE of the forecasts** ~~**simulated seasonal discharge by ensemble models for training and hold-out sets**~~ **at different forecast lead months**~~**, including SWE and climate oscillation indices (upper) and only SWE (bottom) as predictors.**~~

5.4 Predictive uncertainty

Comparison of observed and predicted seasonal discharge across basins and forecast issue dates (**Figure 8**) showcases the performance and limitations of the modelling framework. The alignment of forecasts with observed discharge in most basins indicates reasonable predictive skill, though deviations are ~~evident in some cases~~sometimes evident. Predictions ~~initialized~~initialised earlier, such as on January 1st, tend to show a larger scatter, highlighting higher uncertainty ~~compared~~

~~to~~than forecasts ~~initialized~~initialised closer to the target season (e.g., March 1^{st} or April 1^{st}), which display tighter clustering around the 1:1 line. Basins such as Murghap and Kashkadarya exhibit greater scatter in predictions across all issue dates, particularly at the extremes, reflecting challenges in low-elevated basins. For the largest basins like Naryn and Amudarya, systematic biases are also evident in early forecasts, with overestimation at lower ~~quantiles~~ and underestimation at higher quantiles. These biases likely stem from the uncertainties of snowmelt and hydrological dynamics in these large basins, where significant spatial variability in snow accumulation complicates predictions. In contrast, ~~basins such as~~ the Zarafshan, Varzob, and Chu basins ~~demonstrate more consistent~~ align~~ment~~ better between forecasts and observations across all ~~initialization~~initialisation dates, suggesting more predictable hydrological responses.

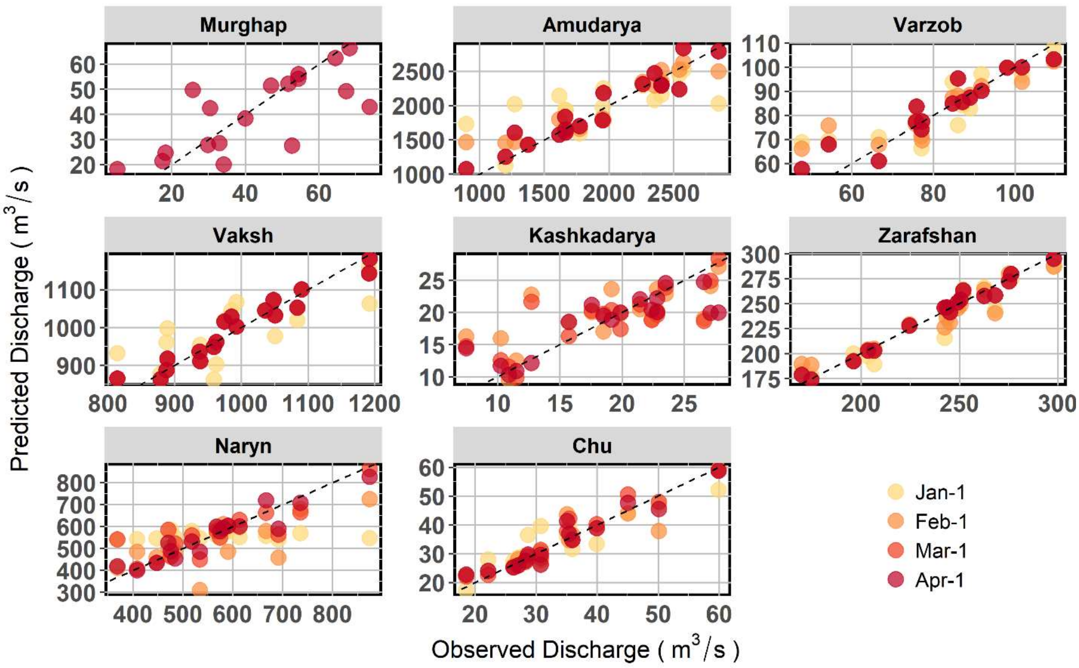
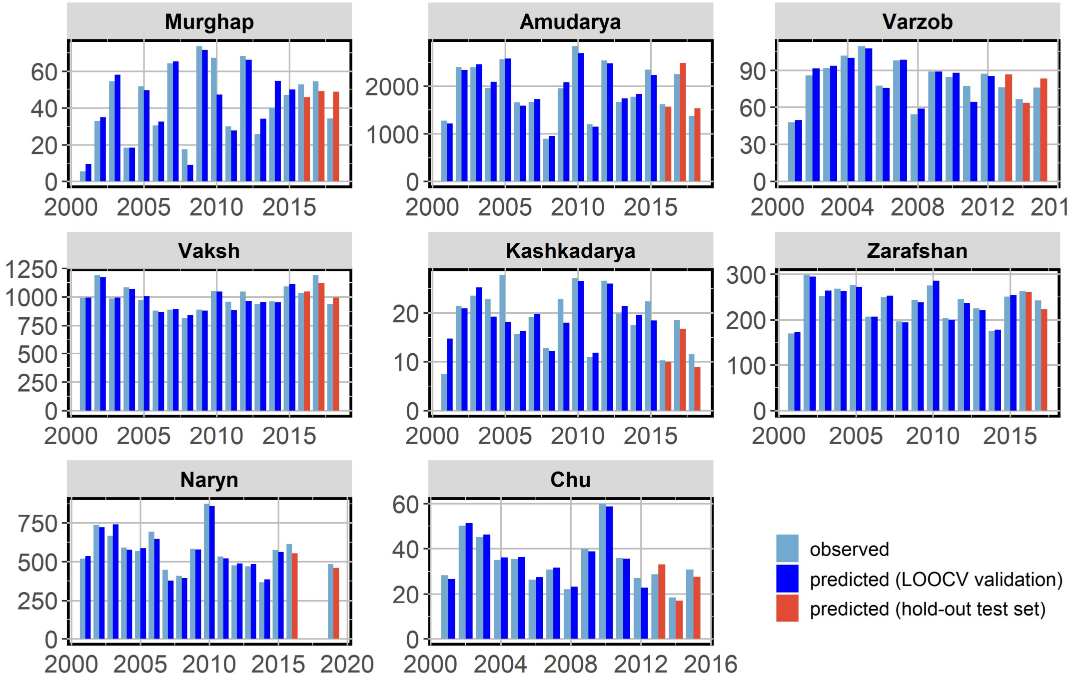


Figure 8. Observed and predicted seasonal discharge at different lead months

~~To assess prediction uncertainty, we have employed a bootstrapping~~ We have employed bootstrapping to assess prediction uncertainty by resampling the data and retraining the SVM meta-learner on each bootstrapped sample. ~~In this way w~~We estimated 80% prediction intervals based on the variability across bootstrapped forecasts. Accordingly (Figure 9), seasonal discharge variability was generally well captured, with predictions closely aligning with observations. However, the width of uncertainty bounds var~~y~~ies across basins, reflecting basin-specific characteristics that influence predictive reliability. Forecasts initiali~~z~~sed earlier, such as January 1st, tend to have broader uncertainty bounds and greater variability~~, whereas as~~. In contrast, as forecast issue dates approach the target season, they become more consistent, with narrower uncertainty bounds and better alignment with observations. Forecasts for basins with smaller catchment areas or/and located at lower elevations, such as Kashkadarya, tend to have wider uncertainty bounds. Similarly, Murghap, a basin with lower seasonal discharge, shows significant deviations in predictions. For these two basins, higher predictive uncertainties may be also be attributed to the comparably higher interannual variability of the seasonal discharge (Table 1). In larger basins like Naryn and Amudarya, variability in uncertainty bounds is more pronounced for early issue dates. However, forecasts for these basins become more consistent with later ~~initialization~~initialisation dates as the models incorporate updated snowpack data, narrowing the uncertainty bounds. In contrast, Zarafshan, Vaksh and Varzob, high-altitude basins, demonstrate stable predictions and tight uncertainty bounds across all ~~initialization~~initialisation dates.



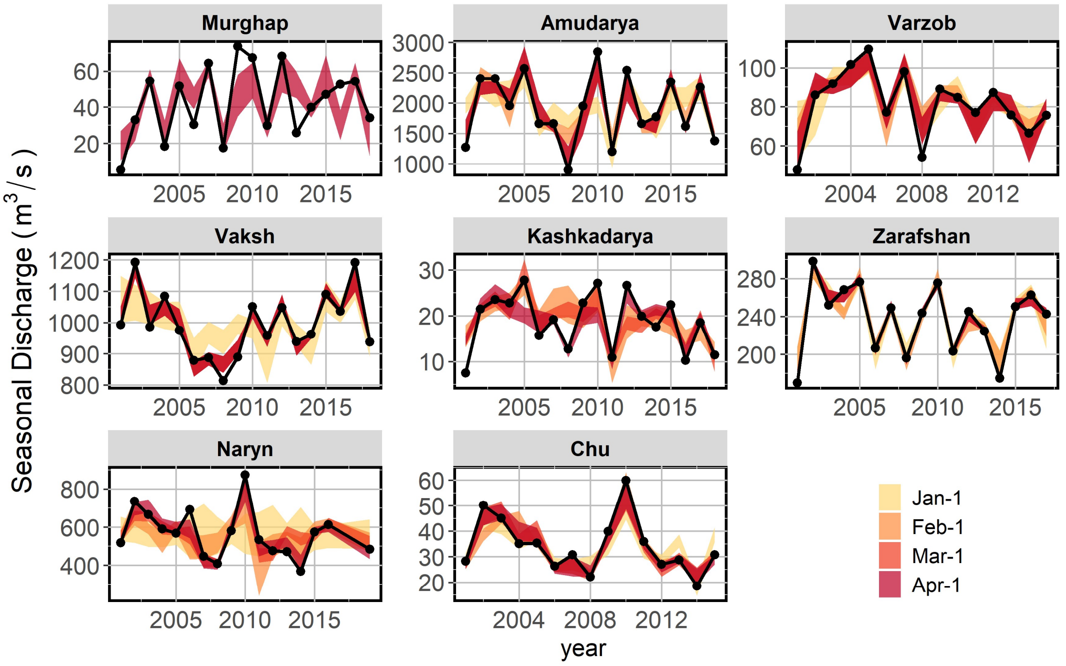
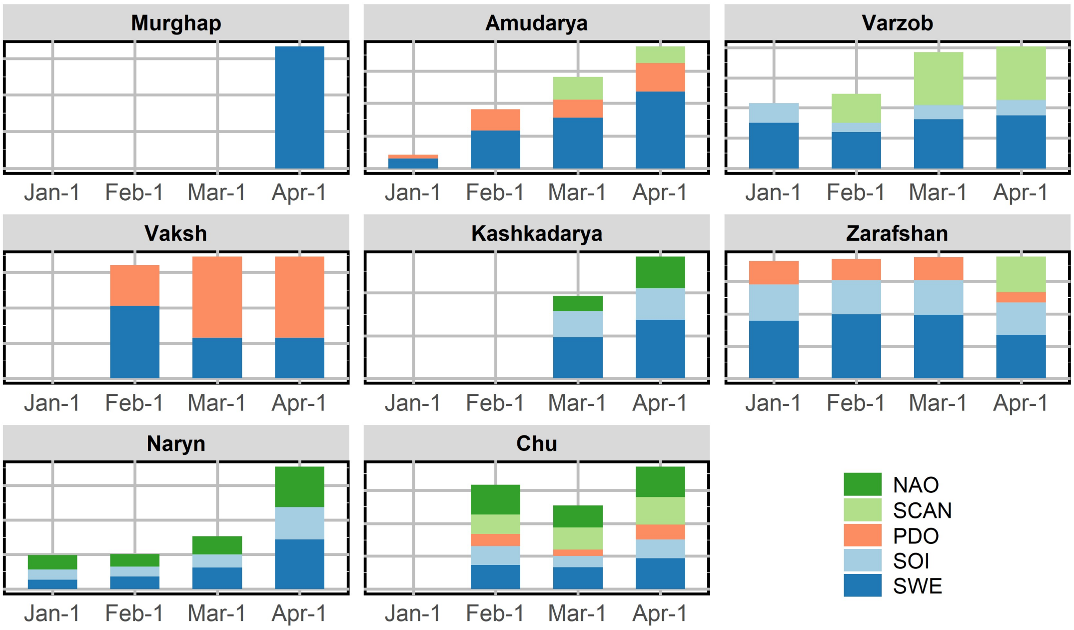


Figure 9. The 80% predictive uncertainty bands of the forecasts at different lead months

~~Figure 8. Observed vs simulated seasonal discharge using Apr 1st forecast ensemble model~~

5.1 Importance of climate oscillation indices as predictors

The importance of predictors varies depending on the catchment location and the forecast issue date (Figure 10~~Figure 9~~). Regardless of the forecast issue date, SWE is a major predictor in most catchments located in Pamir, and its significance generally arises with decreasing lead times. Its incremental value is evident in the basins in the western part of the study area, the Pamir Mountains. Nevertheless, the supplementary predictive value of COs is visible in all basins regardless of their location except for Murghap, where integration of COs does not noticeably improve forecast accuracy compared to ~~forecast~~ models relying only on SWE~~estimates~~. The predictive power of COs is highest for the two catchments located in the inner and northern Tian-Shan, Naryn and Chu. Especially in Chu, the COs contribute to more than half of the predictive power of the forecast models across all forecast issue dates. In addition, a higher reliance on COs is also evident in the high-elevation catchments of Zarafshan and Varzob, with their importance surprisingly increasing at later forecast issue dates.



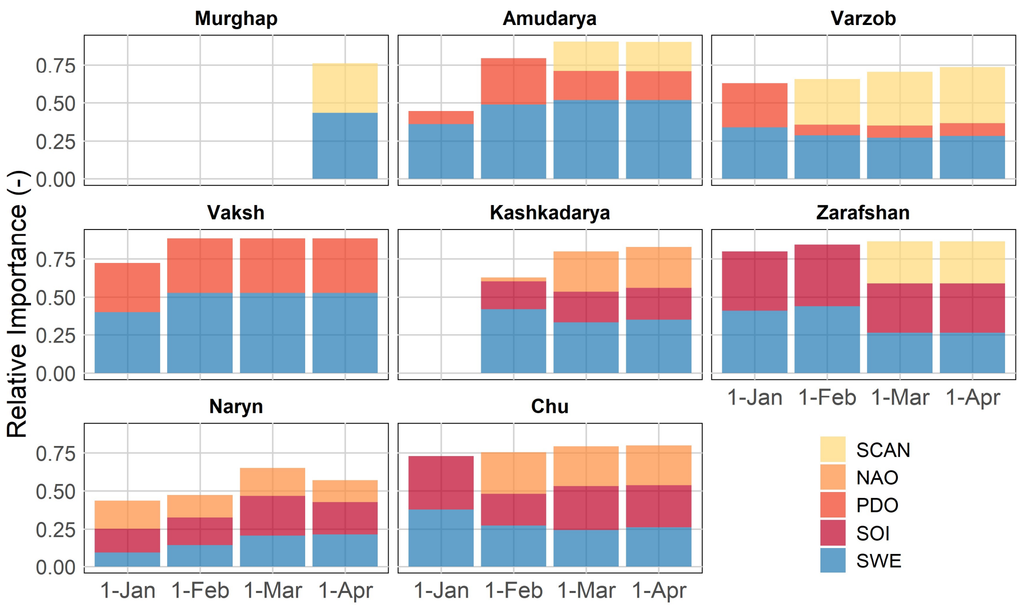


Figure 10~~9~~. The relative importance of predictors at different forecast issue dates

~~5.~~6. Summary and Discussion

Our findings suggest that ~~near real-time~~~~useful~~valid SWE estimates, suitable for operational seasonal river discharge forecasting, can be effectively derived from global reanalysis or satellite data. Still, they are subject to spatial bias and uncertainty, which may be due to uncertainties in underlying precipitation and temperature inputs. The uncertainties in the SWE estimates may propagate across time and ~~enlarge~~ become more pronounced during the snow ablation phase by the end of the cold season. ~~Assimilating~~ Combining SWE data from multiple global sources helps mitigate these biases, and predictions that pass cross-validation filters reflect the accuracy of SWE products specific to catchment locations. Nevertheless, although catchment-averaged SWE estimates improve with the assimilation of multiple snow products, they may still tend to contain spatial uncertainties that increase during the ablation phase.

Multiple global ocean-atmospheric oscillations modulate the seasonal hydroclimatic patterns in the Pamir and Tian-Shan mountains, each with different temporal effects. The findings suggest that the magnitude of both seasonal precipitation and discharge is associated with the late autumn to winter state of ENSO (approximated in our study with SOI). PDO is known to mimic ENSO-like variability on monthly to annual scales and has a pronounced impact on the interdecadal scale (Zhang et al., 1997). This could explain the similarity in dominant lead times observed in our analysis with SOI. Late winter to spring states of NAO and SCAN contribute to hydroclimatic predictability in many studied catchments, mainly showing higher significance in the Tian-Shan domain. All these spatial and temporal patterns are broadly consistent with several earlier findings (e.g., Mariotti 2007, Wang et al. 2014, Dixon and Wilby 2019, Gerlitz et al. 2019).

The associations between the climate indices with ~~both~~ precipitation and discharge exhibit an almost identical pattern, implying that ~~river discharge's~~ interannual volatility of streamflow during the growing season ~~from April to September~~ is substantially driven by the peak precipitation period, which we determine as February to July. This implies that SWE accumulated by the middle of winter is a weak precursor of hydrologic variability in the upcoming season, which our findings assert. On the other hand, this serves as an argument for using climate oscillation indices beside the catchment snowpack in discharge forecasts at extended lead times. Following the traditional approach towards seasonal hydrological predictions, SWE estimates initial hydrological conditions and climate oscillation indices as a proxy of climate variability during the target season.

Our experiment~~, which combines snow and climate indices as predictors,~~ confirms this by demonstrating the complementary roles of SWE and climate oscillation indices in improving discharge forecasts at extended lead times. ~~assumption.~~ The resulting forecast models generate credible simulations of seasonal discharge across all studied catchments, albeit with performance variations depending on lead time. The forecast models incorporating both SWE and COs perform better than the SWE-only models, evidenced by lower forecast ~~uncertainties and minor~~ errors~~ in LOOCV validation and hold-out test sets~~. ~~Due to a limited number of observations, our evaluation only included a small set of the most recent discharge data for some rivers. Despite this limitation, the meta-learner models accurately predicted high and low seasonal discharge values in the hold-out set for some catchments, showcasing its effectiveness in predicting hydrological extremes on seasonal scales.~~

The resulting forecast models underscore the significance of SWE as one of the primary predictors in most catchments in the Pamir region, with its importance becoming more pronounced during the peak SWE period, typically occurring in mid-spring. Nevertheless, the forecast models ~~also~~ gain valuable predictive power from climate oscillation indices during extended and shorter lead times, but t~~. T~~he importance of specific climate oscillations as predictors varied across catchments. In most catchments, the SOI, PDO, or both were utilis~~z~~ed, indicating the dominant influence of ENSO~~ phenomena~~. Moreover, the results suggest that the NAO and SCAN exhibit a relatively higher predictive power for catchments in the Tian-Shan region.

The predictive importance of climate oscillations equalled or exceeded that of SWE in the Naryn and Chu catchments located in Tian-Shan, as well as in high-elevated catchments in Pamir, such as Zarafshan, Varzob, and Vaksh. The former might be explained by a distinctive precipitation cycle across Tian-Shan, which peaks during summer, i.e., considerably later than the final forecast issue date (April 1st). Consequently, SWE estimates have comparably smaller power to capture upcoming hydroclimatic variability than other catchments where precipitation peaking occurs during spring months and thus is embedded in SWE estimates by April 1st. This ~~is better exemplified~~contrasts with~~ by~~ the forecast model for the Murghap catchment, which doesn't integrate any climate oscillations, ~~likely because~~as ~~due to the majority of~~ ~~most seasonal~~ precipitation in this catchment peaks ~~occurr~~es~~ing~~ before spring. The higher predictive power of the oscillations for high-elevated catchments may be attributed to the poorer performance of the satellites and reanalyses of precipitation estimates over high elevations in the region (Peña-Guerrero et al., 2022), which subsequently propagate as uncertainties in the SWE estimates. In this regard, the higher predictive performance of climate oscillation indices across those catchments is ~~assumingly~~ likely ~~reasoned by their~~because they compensate~~ion for~~ for ~~uncertainties~~ errors in SWE estimates.

Based on the abovementioned, we identify three specific cases when the incorporation of COs as additional predictors helps to improve seasonal discharge forecasts in snow-dominated catchments:

1) *Extended lead time forecasts with early seasonal SWE*: When seasonal discharge forecasts are made well in advance, but SWE is not a reliable representation of seasonal terrestrial water storage, climate oscillations may provide additional insights into anticipated hydroclimatic conditions.

2) *Dominant climate variability regime during the target season:* When the seasonal discharge is more influenced by in-season climate variability than by accumulated SWE before the season, climate oscillations can serve as adequate proxies for this variability.

3) *Uncertainties in catchment SWE estimates:* High uncertainties in SWE estimates for a particular catchment result in higher errors in discharge predictions. These uncertainties can be partially compensated by leveraging the forecasts with climate oscillations, leading to more reliable seasonal discharge predictions.

In-situ observations of essential climate variables, such as snowpack properties, are ~~scarce in~~ especially scarce in mountainous regions of the Global South~~, especially mountainous regions~~, impeding hydrological forecasting. Previous research has demonstrated how, in the absence of in-situ snow observations, satellite-derived snow cover, precipitation and temperature can serve as proxies of terrestrial water storage and improve seasonal discharge forecasting in Central Asia (Apel et al., 2018; Gafurov et al., 2016). Additionally, other studies have investigated how climate indices characterize hydroclimatic variability in the region over longer lead times (Dixon and Wilby, 2019). By combining the strengths of these two approaches, our modelling framework offers a new way to make hydrological predictions in the region. It leverages an ensemble technique that uses multiple estimates from global data, a diverse set of more straightforward types of machine learning methods with loose tuning parameters. These elements allow us to achieve ~~reliable~~ plausible forecast models even when in-situ discharge observations are short.

Code and Data Availability

R script to reproduce results in this paper is available at Zenodo (Umirbekov et al., 2024), under the Creative Commons Attribution CC BY 4.0 International license. The same record contains related data files, including~~:~~ mean seasonal discharge of the studied catchment, gridded daily SWE data for the study area, and~~,~~ monthly indices of the climate oscillations.

Author contributions

AU and DM designed the study. All authors evaluated the findings and contributed to the interpretation of results. All authors contributed to writing and reviewing the manuscript. DM supervised the project.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgements

This research has been supported by the Volkswagen Foundation (grant no. 96 264). We extend our sincere appreciation to the open-source developer community and individuals behind numerous R packages, including but not limited to caret (Kuhn, 2008), terra (Hijmans, 2023), vip (Greenwell and Boehmke, 2020), and dplyr (Wickham et al., 2023).

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., and Hegewisch, K. C.: TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015, Sci. Data, 5, 170191, https://doi.org/10.1038/sdata.2017.191, 2018.

Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A. S., Al-dabbagh, B. S. N., Fadhel, M. A., Manoufali, M., Zhang, J., Al-Timemy, A. H., Duan, Y., Abdullah, A., Farhan, L., Lu, Y., Gupta, A., Albu, F., Abbosh, A., and Gu, Y.: A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications, J. Big Data, 10, 46, https://doi.org/10.1186/s40537-023-00727-2, 2023.

Apel, H., Abdykerimova, Z., Agalhanova, M., Baimaganbetov, A., Gavrilenko, N., Gerlitz, L., Kalashnikova, O., Unger-Shayesteh, K., Vorogushyn, S., and Gafurov, A.: Statistical forecast of seasonal discharge in Central Asia using observational records: development of a generic linear modelling tool for operational water resource management, Hydrol. Earth Syst. Sci., 22, 2225–2254, https://doi.org/10.5194/hess-22-2225-2018, 2018.

Aralova, D., Toderich, K., Jarihani, B., Gafurov, D., and Gismatulina, L.: Monitoring of vegetation condition using the NDVI/ENSO anomalies in Central Asia and their relationships with ONI (very strong) phases, in: Earth Resources and Environmental Remote Sensing/GIS Applications VII, 1000512, https://doi.org/10.1117/12.2242164, 2016.

Barlow, M. A. and Tippett, M. K.: Variability and Predictability of Central Asia River Flows: Antecedent Winter Precipitation and Large-Scale Teleconnections, J. Hydrometeorol., 9, 1334–1349, https://doi.org/10.1175/2008jhm976.1, 2008.

Barnston, A. G. and Livezey, R. E.: Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns, Mon. Weather Rev., 115, 1083–1126, https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2, 1987.

Beck, H. E., van Dijk, A. I. J. M., Larraondo, P. R., McVicar, T. R., Pan, M., Dutra, E., and Miralles, D. G.: MSWX: global 3-hourly 0.1° bias-corrected meteorological data including near real-time updates and forecast ensembles, Bull. Am. Meteorol. Soc., 1–55, https://doi.org/10.1175/bams-d-21-0145.1, 2021.

Charles, S. P., Wang, Q. J., Ahmad, M.-D., Hashmi, D., Schepen, A., Podger, G., and Robertson, D. E.: Seasonal streamflow forecasting in the upper Indus Basin of Pakistan: an assessment of methods, Hydrol. Earth Syst. Sci., 22, 3533–3549, https://doi.org/10.5194/hess-22-3533-2018, 2018.

Chen, X., Wang, S., Hu, Z., Zhou, Q., and Hu, Q.: Spatiotemporal characteristics of seasonal precipitation and their relationships with ENSO in Central Asia during 1901–2013, J. Geogr. Sci., 28, 1341–1368, https://doi.org/10.1007/s11442-018-1529-2, 2018.

Church, J. E.: Principles of snow surveying as applied to forecasting stream flow, J. Agric. Res., 51, 97–130, 1935.

Dietterich, T. G.: Ensemble Methods in Machine Learning, in: Multiple Classifier Systems, 1–15, 2000.

Dixon, S. G. and Wilby, R. L.: Forecasting reservoir inflows using remotely sensed precipitation estimates: a pilot study for the River Naryn, Kyrgyzstan, Hydrol. Sci. J., 61, 107–122, https://doi.org/10.1080/02626667.2015.1006227, 2016.

Dixon, S. G. and Wilby, R. L.: A seasonal forecasting procedure for reservoir inflows in Central Asia, River Res. Appl., 35, 1141–1154, https://doi.org/10.1002/rra.3506, 2019.

Farrell, A., Wang, G., Rush, S. A., Martin, J. A., Belant, J. L., Butler, A. B., and Godwin, D.: Machine learning of large-scale spatial distributions of wild turkeys with high-dimensional environmental data, Ecol. Evol., 9, 5938–5949, https://doi.org/10.1002/ece3.5177, 2019.

Fleming, S. W. and Dahlke, H. E.: Parabolic northern-hemisphere river flow teleconnections to El Niño-Southern Oscillation and the Arctic Oscillation, Environ. Res. Lett., 9, 104007, https://doi.org/10.1088/1748-9326/9/10/104007, 2014.

Fleming, S. W. and Garen, D. C.: Simplified Cross-Validation in Principal Component Regression (PCR) and PCR-Like Machine Learning for Water Supply Forecasting, JAWRA J. Am. Water Resour. Assoc., 58, 517–524, https://doi.org/https://doi.org/10.1111/1752-1688.13007, 2022.

Fleming, S. W., Garen, D. C., Goodbody, A. G., McCarthy, C. S., and Landers, L. C.: Assessing the new Natural Resources Conservation Service water supply forecast model for the American West: A challenging test of explainable, automated, ensemble artificial intelligence, J. Hydrol., 602, 126782, https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.126782, 2021.

Fleming, S. W., Rittger, K., Oaida Taglialatela, C. M., and Graczyk, I.: Leveraging Next-Generation Satellite Remote Sensing-Based Snow Data to Improve Seasonal Water Supply Predictions in a Practical Machine Learning-Driven River Forecast System, Water Resour. Res., 60, e2023WR035785, https://doi.org/https://doi.org/10.1029/2023WR035785, 2024.

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, Environ. Earth Sci., 75, 1–15, https://doi.org/10.1007/s12665-016-5869-x, 2016.

Gerlitz, L., Steirou, E., Schneider, C., Moron, V., Vorogushyn, S., and Merz, B.: Variability of the cold season climate in Central Asia. Part II: Hydroclimatic predictability, J. Clim., 32, 6015–6033, https://doi.org/10.1175/JCLI-D-18-0892.1, 2019.

Granata, F. and Di Nunno, F.: Forecasting short- and medium-term streamflow using stacked ensemble models and different meta-learners, Stoch. Environ. Res. Risk Assess., 38, 3481–3499, https://doi.org/10.1007/s00477-024-02760-w, 2024.

Grantz, K., Rajagopalan, B., Clark, M., and Zagona, E.: A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts, Water Resour. Res., 41, https://doi.org/https://doi.org/10.1029/2004WR003467, 2005.

Greenwell, B. M., Boehmke, B. C., and McCarthy, A. J.: A Simple and Effective Model-Based Variable Importance Measure, 1–27, 2018.

Hagedorn, R., Doblas-Reyes, F. J., and Palmer, T. N.: The rationale behind the success of multi-model ensembles in seasonal forecasting — I. Basic concept, Tellus A Dyn. Meteorol. Oceanogr., 57, 219–233, https://doi.org/10.3402/tellusa.v57i3.14657, 2005.

Hamid, M. and Matthew, M.: Long-Lead Water Supply Forecast Using Large-Scale Climate Predictors and Independent Component Analysis, J. Hydrol. Eng., 15, 744–762, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000246, 2010.

Hock, R., G., Rasul, C., Adler, B., Cáceres, S., Gruber, Y., Hirabayashi, M., Jackson, A., Kääb, S., Kang, S., Kutuzov, A., Milner, U., Molau, S., Morin, B., Orlove, B., and Steltzer H: High Mountain Areas, in: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, 2019.

Hou, M., Cuo, L., Murodov, A., Ding, J., Luo, Y., Liu, T., and Chen, X.: Streamflow Composition and the Contradicting Impacts of Anthropogenic Activities and Climatic Change on Streamflow in the Amu Darya Basin, Central Asia, J. Hydrometeorol., 24, 185–201, https://doi.org/10.1175/JHM-D-22-0040.1, 2023.

Hu, Z., Zhou, Q., Chen, X., Li, J., Li, Q., Chen, D., Liu, W., and Yin, G.: Evaluation of three global gridded precipitation data sets in central Asia based on rain gauge observations, Int. J. Climatol., 38, 3475–3493, https://doi.org/https://doi.org/10.1002/joc.5510, 2018.

Huffman, G. J., Stocker, E. F., Bolvin, D. T., Nelkin, E. J., and Tan, J.: GPM IMERG Late Precipitation L3 1 day 0.1 degree x 0.1 degree V06, https://doi.org/https://doi.org/10.5067/GPM/IMERGDL/DAY/06, 2019.

Immerzeel, W. W., Lutz, A. F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B. J., Elmore, A. C., Emmer, A., Feng, M., Fernández, A., Haritashya, U., Kargel, J. S., Koppes, M., Kraaijenbrink, P. D. A., Kulkarni, A. V., Mayewski, P. A., Nepal, S., Pacheco, P., Painter, T. H., Pellicciotti, F., Rajaram, H., Rupper, S., Sinisalo, A., Shrestha, A.

B., Viviroli, D., Wada, Y., Xiao, C., Yao, T., and Baillie, J. E. M.: Importance and vulnerability of the world's water towers, Nature, 577, 364–369, https://doi.org/10.1038/s41586-019-1822-y, 2020.

Kalra, A., Ahmad, S., and Nayak, A.: Increasing streamflow forecast lead time for snowmelt-driven catchment based on large-scale climate patterns, Adv. Water Resour., 53, 150–162, https://doi.org/https://doi.org/10.1016/j.advwatres.2012.11.003, 2013.

Kennedy, A. M., Garen, D. C., and Koch, R. W.: The association between climate teleconnection indices and Upper Klamath seasonal streamflow: Trans-Niño Index, Hydrol. Process., 23, 973–984, https://doi.org/https://doi.org/10.1002/hyp.7200, 2009.

Korsic, S. A. T., Notarnicola, C., Quirno, M. U., and Cara, L.: Assessing a data-driven approach for monthly runoff prediction in a mountain basin of the Central Andes of Argentina, Environ. Challenges, 10, 100680, https://doi.org/https://doi.org/10.1016/j.envc.2023.100680, 2023.

Kuhn, M.: Building Predictive Models in R Using the caret Package, J. Stat. Softw., 28, 1–26, https://doi.org/10.18637/jss.v028.i05, 2008.

Li, Y., Liang, Z., Hu, Y., Li, B., Xu, B., and Wang, D.: A multi-model integration method for monthly streamflow prediction: modified stacking ensemble strategy, J. Hydroinformatics, 22, 310–326, https://doi.org/10.2166/hydro.2019.066, 2019.

Livneh, B. and Badger, A. M.: Drought less predictable under declining future snowpack, Nat. Clim. Chang., 10, 452–458, https://doi.org/10.1038/s41558-020-0754-8, 2020.

Mallick, J., Talukdar, S., and Ahmed, M.: Combining high resolution input and stacking ensemble machine learning algorithms for developing robust groundwater potentiality models in Bisha watershed, Saudi Arabia, Appl. Water Sci., 12, 77, https://doi.org/10.1007/s13201-022-01599-2, 2022.

Mankin, J. S., Viviroli, D., Singh, D., Hoekstra, A. Y., and Diffenbaugh, N. S.: The potential for snow to supply human water demand in the present and future, Environ. Res. Lett., 10, 114016, https://doi.org/10.1088/1748-9326/10/11/114016, 2015.

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., and Francis, R. C.: A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production, Bull. Am. Meteorol. Soc., 78, 1069–1079, https://doi.org/10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2, 1997.

Mariotti, A.: How ENSO impacts precipitation in southwest central Asia, Geophys. Res. Lett., 34, 2–6, https://doi.org/10.1029/2007GL030078, 2007.

McNally, A., Jacob, J., Arsenault, K., Slinski, K., Sarmiento, D. P., Hoell, A., Pervez, S., Rowland, J., Budde, M., Kumar, S., Peters-Lidard, C., and Verdin, J. P.: A Central Asia hydrologic monitoring dataset for food and water security applications in Afghanistan, Earth Syst. Sci. Data, 14, 3115–3135, https://doi.org/10.5194/essd-14-3115-2022, 2022.

Mendoza, P. A., Rajagopalan, B., Clark, M. P., Cortés, G., and McPhee, J.: A robust multimodel framework for ensemble seasonal hydroclimatic forecasts, Water Resour. Res., 50, 6030–6052, https://doi.org/https://doi.org/10.1002/2014WR015426, 2014.

Mendoza, P. A., Wood, A. W., Clark, E., Rothwell, E., Clark, M. P., Nijssen, B., Brekke, L. D., and Arnold, J. R.: An intercomparison of approaches for improving operational seasonal streamflow forecasts, Hydrol. Earth Syst. Sci., 21, 3915–3935, https://doi.org/10.5194/hess-21-3915-2017, 2017.

Modi, P. A., Small, E. E., Kasprzyk, J., and Livneh, B.: Investigating the Role of Snow Water Equivalent on Streamflow Predictability during Drought, J. Hydrometeorol., 23, 1607–1625, https://doi.org/https://doi.org/10.1175/JHM-D-21-0229.1, 2022.

Mortimer, C., Mudryk, L., Derksen, C., Luojus, K., Brown, R., Kelly, R., and Tedesco, M.: Evaluation of long-term Northern Hemisphere snow water equivalent products, Cryosph., 14, 1579–1594, https://doi.org/10.5194/tc-14-1579-2020, 2020.

Mortimer, C., Mudryk, L., Cho, E., Derksen, C., Brady, M., and Vuyovich, C.: Use of multiple reference data sources to cross-validate gridded snow water equivalent products over North America, Cryosph., 18, 5619–5639, https://doi.org/10.5194/tc-

18-5619-2024, 2024.

Muñoz-Sabater, J., Dutra, E., Agust\'\i-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodr\'\iguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Murray, S. A.: The Importance of Ensemble Techniques for Operational Space Weather Forecasting, Sp. Weather, 16, 777–783, https://doi.org/https://doi.org/10.1029/2018SW001861, 2018.

Najafi, R. M. and Moradkhani, H.: Ensemble Combination of Seasonal Streamflow Forecasts, J. Hydrol. Eng., 21, 4015043, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001250, 2016.

Oğulcan Doğan, Y., Arda Şorman, A., and Şensoy, A.: Multi-criteria evaluation for parameter uncertainty assessment and ensemble runoff forecasting in a snow-dominated basin, J. Hydrol. Hydromechanics, 71, 231–247, https://doi.org/doi:10.2478/johh-2023-0003, 2023.

Opitz-Stapleton, S., Gangopadhyay, S., and Rajagopalan, B.: Generating streamflow forecasts for the Yakima River Basin using large-scale climate predictors, J. Hydrol., 341, 131–143, https://doi.org/https://doi.org/10.1016/j.jhydrol.2007.03.024, 2007.

Pagano, T. and Garen, D.: A Recent Increase in Western U.S. Streamflow Variability and Persistence, J. Hydrometeorol., 6, 173–179, https://doi.org/10.1175/JHM410.1, 2005.

Pagano, T. and Garen, D.: Integration of Climate Information and Forecasts into Western US Water Supply Forecasts, Clim. Var. Clim. Chang. Water Resour. Eng., https://doi.org/10.1061/9780784408247.ch06, 2010.

Pechlivanidis, I. G., Crochemore, L., Rosberg, J., and Bosshard, T.: What Are the Key Drivers Controlling the Quality of Seasonal Streamflow Forecasts?, Water Resour. Res., 56, e2019WR026987, https://doi.org/https://doi.org/10.1029/2019WR026987, 2020.

Peña-Guerrero, M. D., Umirbekov, A., Tarasova, L., and Müller, D.: Comparing the performance of high-resolution global precipitation products across topographic and climatic gradients of Central Asia, Int. J. Climatol., 0–1, https://doi.org/https://doi.org/10.1002/joc.7548, 2022.

Regonda, S. K., Rajagopalan, B., Clark, M., and Zagona, E.: A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin, Water Resour. Res., 42, https://doi.org/https://doi.org/10.1029/2005WR004653, 2006.

Ropelewski, C. F. and Jones, P. D.: An extension of the Tahiti-Darwin Southern Oscillation Index, Mon. Weather Rev., 115, 1987.

Safonova, A., Ghazaryan, G., Stiller, S., Main-Knorn, M., Nendel, C., and Ryo, M.: Ten deep learning techniques to address small data problems with remote sensing, Int. J. Appl. Earth Obs. Geoinf., 125, 103569, https://doi.org/https://doi.org/10.1016/j.jag.2023.103569, 2023.

Schaake, J. C. J. and Peck, E. L.: Analysis of water supply forecast accuracy, in: 53rd Annual Western Snow Conference, 1985.

Schär, C., Vasilina, L., Pertziger, F., and Dirren, S.: Seasonal runoff forecasting using precipitation from meteorological data assimilation systems, J. Hydrometeorol., 5, 959–973, https://doi.org/10.1175/1525-7541(2004)005<0959:SRFUPF>2.0.CO;2, 2004.

Syed, F. S., Giorgi, F., Pal, J. S., and Keay, K.: Regional climate model simulation of winter climate over central-southwest Asia, with emphasis on NAO and ENSO effects, Int. J. Climatol., 30, 220–235, https://doi.org/10.1002/joc.1887, 2010.

Troin, M., Arsenault, R., Wood, A. W., Brissette, F., and Martel, J.-L.: Generating Ensemble Streamflow Forecasts: A Review of Methods and Approaches Over the Past 40 Years, Water Resour. Res., 57, e2020WR028392, https://doi.org/https://doi.org/10.1029/2020WR028392, 2021.

Umar, M., Yousaf, W., Ahmad, S., Irteza, S., Abbas, S., and Javaid, A.: Identification of Suitable Climate Indices to Forecast Seasonal Streamflow in the Upper Indus Basin at Tarbela Dam, Pakistan, [Preprint], https://doi.org/10.2139/ssrn.4558997, 2023.

Umirbekov, A., Peña-Guerrero, M. D., and Müller, D.: Regionalization of climate teleconnections across Central Asian mountains improves the predictability of seasonal precipitation, Environ. Res. Lett., 17, 55002, https://doi.org/10.1088/1748-9326/ac6229, 2022.

Umirbekov, A., Essery, R., and Müller, D.: GEMS v1.0: Generalizable empirical model of snow accumulation and melt based on daily snow mass changes in response to climate and topographic drivers, Rev., 2023.

Umirbekov, A., Peña-Guerrero, M. D., Didovets, I., Apel, H., Gafurov, A., and Müller, D.: R script and data for the manuscript "The Value of Hydroclimatic Teleconnections for Snowpack-based Seasonal Streamflow Forecasting," https://doi.org/https://doi.org/10.5281/zenodo.11308066, May 2024.

Viviroli, D. and Weingartner, R.: The hydrological significance of mountains: from regional to global scale, Hydrol. Earth Syst. Sci., 8, 1017–1030, https://doi.org/10.5194/hess-8-1017-2004, 2004.

Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M., and Weingartner, R.: Mountains of the world, water towers for humanity: Typology, mapping, and global significance, Water Resour. Res., 43, https://doi.org/https://doi.org/10.1029/2006WR005653, 2007.

Viviroli, D., Kummu, M., Meybeck, M., Kallio, M., and Wada, Y.: Increasing dependence of lowland populations on mountain water resources, Nat. Sustain., 3, 917–928, https://doi.org/10.1038/s41893-020-0559-9, 2020.

Wang, S., Huang, J., He, Y., and Guan, Y.: Combined effects of the Pacific Decadal Oscillation and El Niño-Southern Oscillation on Global Land Dry-Wet Changes, Sci. Rep., 4, 1–8, https://doi.org/10.1038/srep06651, 2014.

Xenarios, S., Gafurov, A., Schmidt-Vogt, D., Sehring, J., Manandhar, S., Hergarten, C., Shigaeva, J., and Foggin, M.: Climate change and adaptation of mountain societies in Central Asia: uncertainties, knowledge gaps, and data constraints, Reg. Environ. Chang., 19, 1339–1352, https://doi.org/10.1007/s10113-018-1384-9, 2019.

Xu, X., Chen, F., Wang, B., Harrison, M. T., Chen, Y., Liu, K., Zhang, C., Zhang, M., Zhang, X., Feng, P., and Hu, K.: Unleashing the power of machine learning and remote sensing for robust seasonal drought monitoring: A stacking ensemble approach, J. Hydrol., 634, 131102, https://doi.org/https://doi.org/10.1016/j.jhydrol.2024.131102, 2024.

Zhang, Y., Wallace, J. M., and Battisti, D. S.: ENSO-like interdecadal variability: 1900-93, J. Clim., 10, 1004–1020, https://doi.org/10.1175/1520-0442(1997)010<1004:eliv>2.0.co;2, 1997.

Zounemat-Kermani, M., Batelaan, O., Fadaee, M., and Hinkelmann, R.: Ensemble machine learning paradigms in hydrology: A review, J. Hydrol., 598, 126266, https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.126266, 2021.