DKUCNRS WORKING PAPER (2022) 53-72

https://doi.org/10.29258/DKUCRSWP/2022/53-72.eng

Working Papers Series

Analysis of sets of factors affecting the variable flow of the Amu Darya River to create a seasonal prognostic model

G. Shadurdyyev

Kazakh-German University, Pushkin St., 111, Almaty, Kazakhstan, 050011.

ABSTRACT

The Amu Darya River is a transboundary river whose flow of the river in high-water years reaches up to 108 km^3 and in low-water years up to 47 km^3 and these are huge fluctuations in the water flow of the river for Tajikistan, Kyrgyzstan, Uzbekistan, Turkmenistan, and Afghanistan, that share water among themselves. The point to consider is that the downstream countries Turkmenistan and Uzbekistan (and possibly Afghanistan in the future) use a lot of water for irrigation, and therefore these countries are the ones most in need of an accurate forecast of the volume of water for the upcoming season. An accurate forecast of the volume of water on the seasonal scale is necessary for better planning of the structure of crops, and subsequently water use in the irrigation of crops. An acceptable solution to this challenge is the construction of an empirical time series model that will be used to predict the seasonal flows of the Amu Darya River to improve the planning and management of water resources in downstream countries.

This article considers three important discharge time series in the larger Amu Darya Basin. These include the Kerki Gauge on the Amu Darya, Darband Gauge on Vaksh River and Khorog Gauge on Gunt River. Long-term time series from these stations are available for the study of the development and implementation of time-series based models for the prediction of discharge in the basin. At this stage, we attempt to demonstrate a proof-of-concept which can in a second step convince stakeholders to share such type of discharge data operationally for more effective water allocation between sectors and countries. All our work was carried out with the quantitative tools R/RStudio and QGIS. It can serve as a stepping stone for more complex forecasting models in the future.

ARTICLE HISTORY

Submitted: July 19, 2022
Accepted: November 18, 2022
Published: December 21, 2022

KEYWORDS

QGIS, R/RStudio, Amu Darya River, hydrological characterization, discharge prediction.

CONTACT **G. Shadurdyyev** govsutsadurdyev@gmail.com, Kazakh-German University, Pushkin St., 111, Almaty, Kazakhstan, 050011.

Published by Kazakh German University, Almaty/Kazakhstan

Introduction

1.1 Background

The Amu Darya River, the largest and, therefore, the most important freshwater river water in Central Asia and Afghanistan is used by Tajikistan, Kyrgyzstan, Afghanistan, Uzbekistan, and Turkmenistan. The water management of the Amu Darya River is often divided into two periods: before and after the collapse of the USSR. In Soviet times, the Scientific and Technical Council of the USSR's Ministry of Water Resources in 1987 made from Protocol 566 determined the quotas of Tajikistan, Uzbekistan, Turkmenistan, and the Kyrgyz Republic in the Amu Darya River Basin (Kamil, 2021). Water resources were distributed as shown in Table 1. Uzbekistan and Turkmenistan were and still are the most significant water users, because these downstream countries had advantageous geographical locations and fertile soils at that time since, and in the USSR, a large amount of water was directed to irrigate the agricultural fields (the crops were mainly cotton and wheat). After the collapse of the USSR, an agreement in 1992 was made between the states of Central Asia to continue the sharing of the waters of the Amu Darya River and Syr Darya River basins following past practice and volumes (Agreement between the Republic of Kazakhstan, the Kyrgyz Republic, the Republic of Tajikistan, Turkmenistan and the Republic of Uzbekistan on Cooperation in the Field of Joint Management on Utilization and Protection of Water Resources from Interstate Sources, n.d.). This meant that the limits on the division of water resources remained like those of the USSR. So, Turkmenistan and Uzbekistan together consume the bulk of the annually renewable water resources of the Amu Darya River for irrigation.

Table I. Water allocation quotas as per Protocol 566 (Kamil, 2021)

Country	BCM per year	Percentage
Kyrgyzstan	0.40	0.6
Tajikistan	9.50	15.4
Turkmenistan	22.0	35.8
Uzbekistan	29.60	48.3
Total	61.50	100

The water flow of the Amu Darya River fluctuates significantly between years. "In connection with this, in ancient times, the river was called Jeyhun - this, translated from the Semitic language, meant " mad ", " unbridled ", which very accurately characterizes the behavior of this river" (Muradov, 2013). The problem to note is the volume of water in the flow of the Amu Darya River changes every year, then, accordingly, countries that use water for irrigation necessary to change the area under crops every year based on the volume of water flow. Therefore, forecasting the volume of water runoff is the most crucial tool in the irrigation of agricultural land, and this prompted us to create a digital model to predict the volume of water runoff for the next season.

Since the Amu Darya River is a transboundary river, the considered best way to manage water resources is through a regional approach. Based on this, the creation of this predictive model is done so that it is used at the international level and to benefit all interested organizations and researchers. But the focus of the study is on the issues of water security in Turkmenistan because of the relevance of the river for the country. “The Amu Darya River is the most crucial source of fresh water in Turkmenistan. The total volume of water resources of Turkmenistan is 25 km^3 per average annual water content. Amu Darya River shares 22 km^3 or 88% of the total volume of surface water resources; the remained part – 1.631 km^3 (6.5%) belongs to Murgap River, 0.869 (3.5) – Tejen River, 0.354 (1.4) – Etrek River, Sumbar River, and Chandir River, 0.15 km^3 (0.6%) – small rivers” (About Turkmenistan, n.d.).

A further important agreement to note shown for this work is that in 1996, between Turkmenistan and Uzbekistan, there was an agreement signed on the division of the waters of the Amu Darya River at the Kerki gauge station where a 50% - 50% division was agreed on (Agreement between Uzbekistan and Turkmenistan on Cooperation on Water Management Questions, n.d.). Figure 1 plate f below shows a map with the location of the Kerki gauge station (17011).

Why this agreement did consider, and why is this essential for this work? The answer to this question is the favorable location of the Kerki gauge station makes it possible not to build a predictive model for the entire territory of the Amu Darya River, to know the volume of water supplied from the Amu Darya River for the countries of the lower reaches of the river, it will be enough to predict the volume of water only at the Kerki gauge station. Since this station shares the water resources of the two countries of Turkmenistan and Uzbekistan. Those having received a forecast on the volume of water for the next season at the Kerki gauge station can obtain data on the entire volume of water for Turkmenistan and Uzbekistan received from the Amu Darya River.

It should be noted that the frequently used types of modeling for forecasting used by water organizations require large amounts of data from gauging stations. In our case, we do not have sufficient data from gauging stations. Therefore, the research question will be: “Is it possible, meeting the required percentage accuracy for water management in irrigation, to create a predictive model for the water discharge of the Amu Darya River at the Kerki gauge station, without having complete data at the station itself, but supplementing it with data from two gauging stations located upstream of the river and remote sensing data?”

To answer this question, a suitable set of factors for the creation of a seasonal predictive model for the Kerki gauge station in the Amu Darya River have to be identified. Based on this, the purpose of the present work is: “To collect data on the Amu Darya River basin, as well as data on the flow of the river, and make an analysis of these data, to further build a hydrological model that will predict the seasonal scale flow of the river at Kerki gauge station.”

1.2. Motivation

This work is motivated by the idea of combining relevant time series data, and as a result, the creation of an empirical forecasting model for predicting the amount of water in the Amu Darya River for the next season, which will be able to update the information every year, i.e. operationally. In the hope that this work will be helpful for better planning and integrated water resources management at the regional and national levels of individual countries. Therefore, the aim was faced with of collecting data useful for predicting the flow of the Amu Darya River for the next season. To achieve the aim, objectives were set, which consist of the following points:

- Create a digital map by collecting data from ground stations and information from satellites for remote sensing and doing a basin characterization.
- Data preparation for further use in time series modeling and forecasting of Amu Darya River discharge
- The necessary data getting from various sources, as well as compared with each other for greater accuracy. These are data from satellite remote sensing and in-situ ground, data as well as data from previous similar scientific works.
- Using QGIS and R/RStudio for implementation
- Provide an overview of modern techniques for forecast discharge of Amu Darya River at seasonal scales (6 months into the future).

More specifically, GIS technology helped in this work to take all the information about the basin itself. And get an answer about the characteristics of the basin itself and anthropogenic impacts. The main factor of this basin is that the river is formed because of the melting of snow and glaciers in the highlands of this basin, namely in the Pamir / Hindukush and Gissar-Alay mountains during the summer periods. And winter periods fall because heavy precipitation in the mountains replenishes their reserves of snow and ice. Therefore, for hydrological modeling, we must pay great attention to these processes.

Currently, the use of GIS technologies is an integral part of the work of most engineers worldwide. GIS technology is improving constantly and becoming a more productive tool in the hands of engineers. For hydrologists, too, these technologies provide tremendous help in solving problems in many directions. But if we are talking about measuring the flow of a river, it should be borne in mind that data taken from remote sensing (remote sensing of the Earth) is to a certain extent an alternative to ground hydrological gauging stations. That is, data taken from gauging stations should be much more accurate than data from satellites. But there are problems in solving our problem. The problem is that each country with gauging stations along the river's discharge in its part of the country is reluctant to share data on river flow. Also, many of the most essential gauging stations went out of order after the collapse of the USSR, which added to the problems.

Despite all the difficulties, we must not forget that the Amu Darya River needs to be managed at the regional level, and for this, we need to consolidate all information throughout the entire basin. Since remote sensing data are an open source of information, this work would help to collect all the available information into to create a time-series hydrological model. The creation of GIS data, for this model, worked, is in parallel with the data obtained from the gauging stations. Namely, work for empirical modeling using time series implemented in the R programming environment.

2. Data and Methods

2.1. Data

For this work, it is essential to collect information about the most important factors influencing the creation of flow and, of course, information about past flow data to understand the characteristics of the river. Speaking about the data of our work, we can divide it into two parts. This is information about the Amu Darya River basin and information about the river flow itself. Information about the basin is developed using remote sensing data and GIS technologies. To date, we have built a digital map of the Amu Darya River basin, consisting of different layers, which reflect essential aspects of the bay.

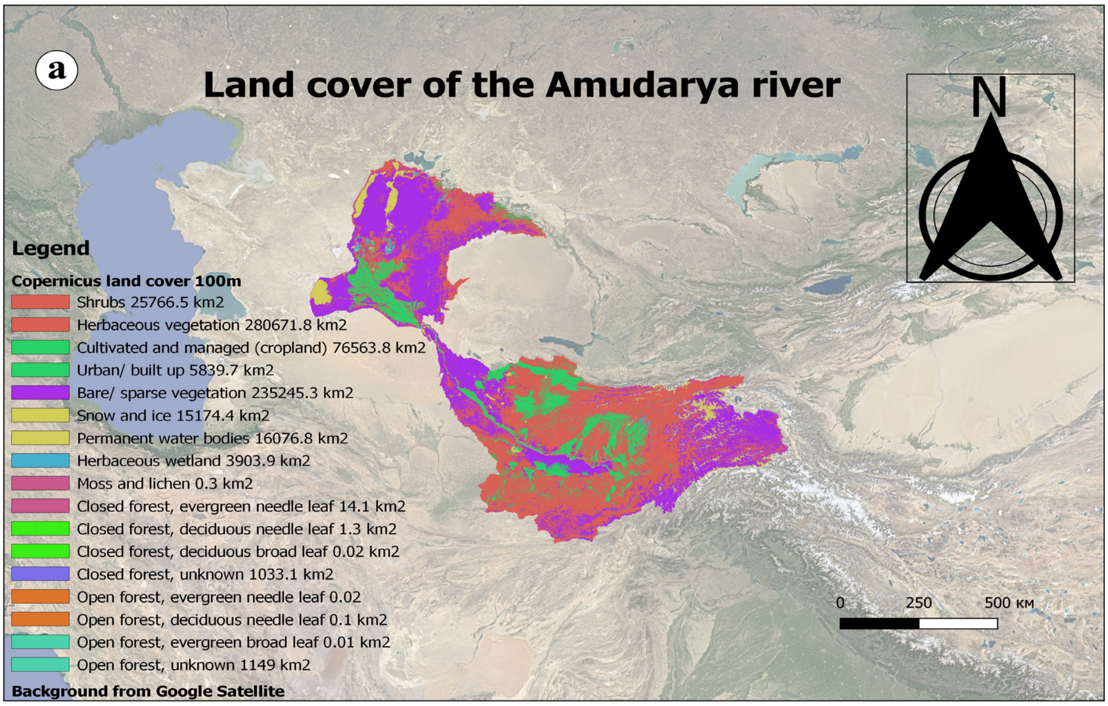
Let's start listing the data available in our digital map with land cover, taken from Land Cover sources (Figure 1, plate a). This layer of the map shows the land cover in the Amu Darya River basin (not including irrigated areas of canals from these waters, in order not to deviate far from the current topic). The map shows and labels the varieties of land cover, divided into corresponding colors, and you can visually see which areas are occupied by one or another cover. But it is also possible to calculate the exact sizes of the occupied territories using the QGIS program function, which we have also shown in the legend of our map for better understanding.

The main part of the water - in the Amu Darya River is formed due to the snowmelt on glaciers in the Pamir mountains, and then flows downstream. Therefore, in this work, it is essential to consider the topography of the basin, as well as the area and volume of glaciers on which the main volume of snow accumulates. Therefore, we have collected and analyzed DEM (Digital Elevation Model) data and glacier data which are shown in (Figure 1, plate b).

Climatic factors also significantly affect renewable water resources in the Amu Darya River basin. Consequently, we have included temperature climatology of the Amu Darya River information in our digital map from 1979 – 2011. In Figure 1, plate c, averaged climatological information is shown over these years. Information from temperature climatology of the Amu Darya River will help us in the future to calculate the partitioning of precipitation into evaporation, runoff or it is becoming intermittently stored.

To calculate the amount volume of renewable water resources, we need to know information about precipitation in the Amu Darya River basin, and this is also important. We have included information on precipitation climatology of the Amu Darya River information in our digital map from 1979 - 2011, and we, showed figure averaged information during these years (Figure 1, plate d).

Typically, runoff in the Central Asian River basin is ~80% of snowmelt, and of course, this is one of the most important pieces of information that one must know. Based on this, we included in our digital map information that shows the fraction of total precipitation falling as snow over the years 1981 – 2010, and we show long-term averaged data during these years (Figure 1, plate e). The fraction of total precipitation falling as snow helps a lot in the analysis of our region and makes it possible to accurately determine the effect of snowmelt on water runoff by comparing information from gauging stations and snowmelt volumes.



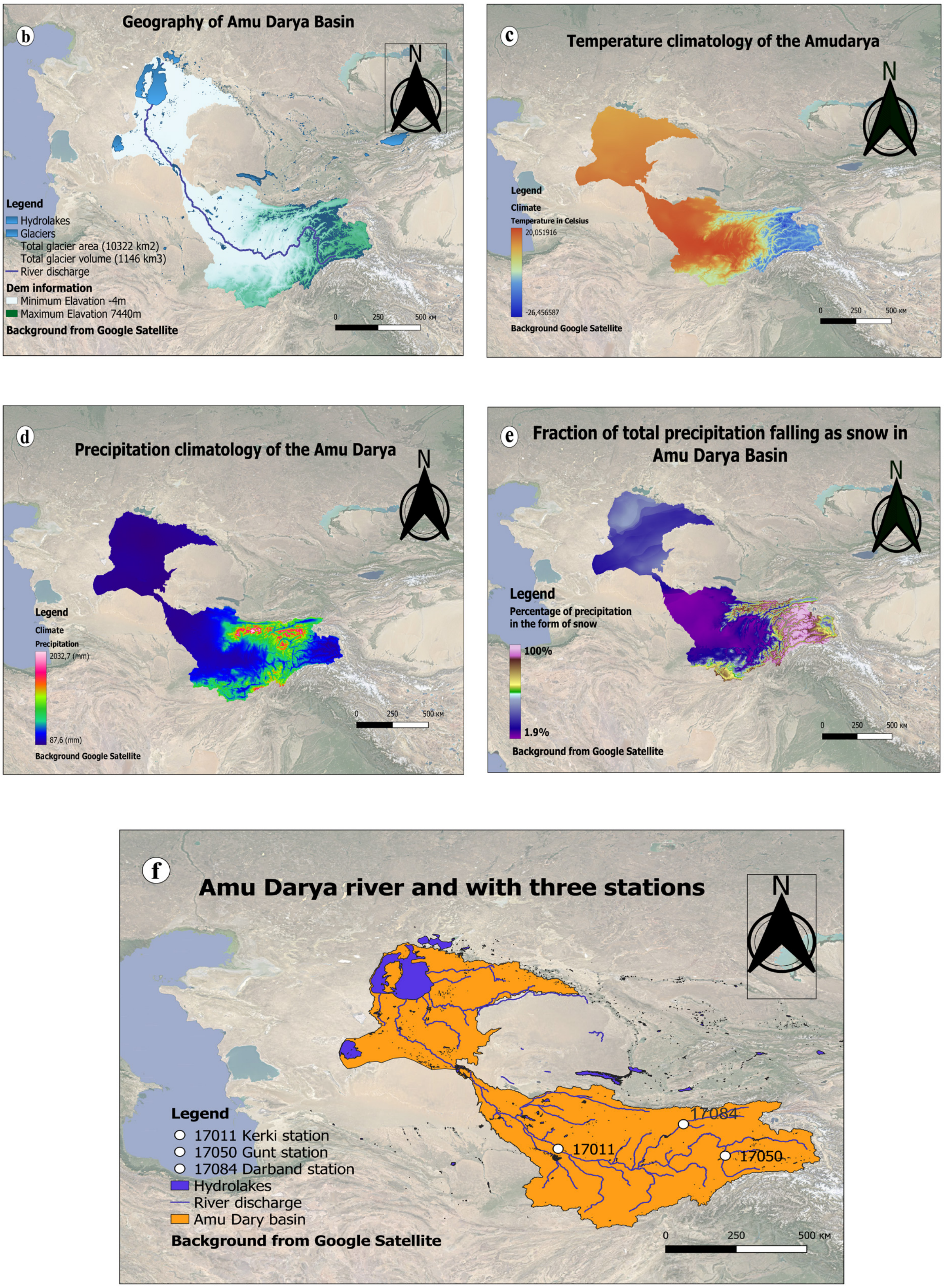


Figure 1. Overview characteristics of the Amu Darya River basin. Plate a) shows Copernicus landcover (Buchhorn et al., 2020), plate b) SRTM digital elevation model (Farr et al., 2007), plate c) is the CHELSA V21 temperature climatology from 1979 – 2011(Karger et al., 2020), plate d) is the CHELSA V21 precipitation climatology from 1979 – 2011 (Karger et al., 2020), plate e) is the fraction of precipitation falling as snow (derived from CHELSA V21 data) and plate f) important gauging stations in the river basin.

As it has already become apparent, this model is hybrid and combines all the essential data for the region we have chosen, and this is a novelty today. In other words, we can look at existing problems from a new point of view. Today, in addition to data from GIS technologies, we also have data from gauging stations, which will significantly help better modeling. Taken about the runoff data from possible gauging stations today, these are Kerki gauge station (17011), Darband gauge station (17084), and Gunt gauge station (17050). A better, understanding of the locations of the stations are shown in (Figure 1, plate f). The data taken from the stations is used to calibrate the model. I would like to note that the more data we can get, the more accurate our hydrological model will be.

After receiving the GIS information, we simultaneously started working with the data obtained from the gauging stations. Namely, we started work on the empirical modeling of time series using the R/RStudio program. We have had information about the flow of water since 1911, and this is a station in Kerki gauge station. To better understand the information, we collected all available information from all three stations and built a graph using R/RStudio. The time series of the three stations are shown in (Figure 2, plate a).

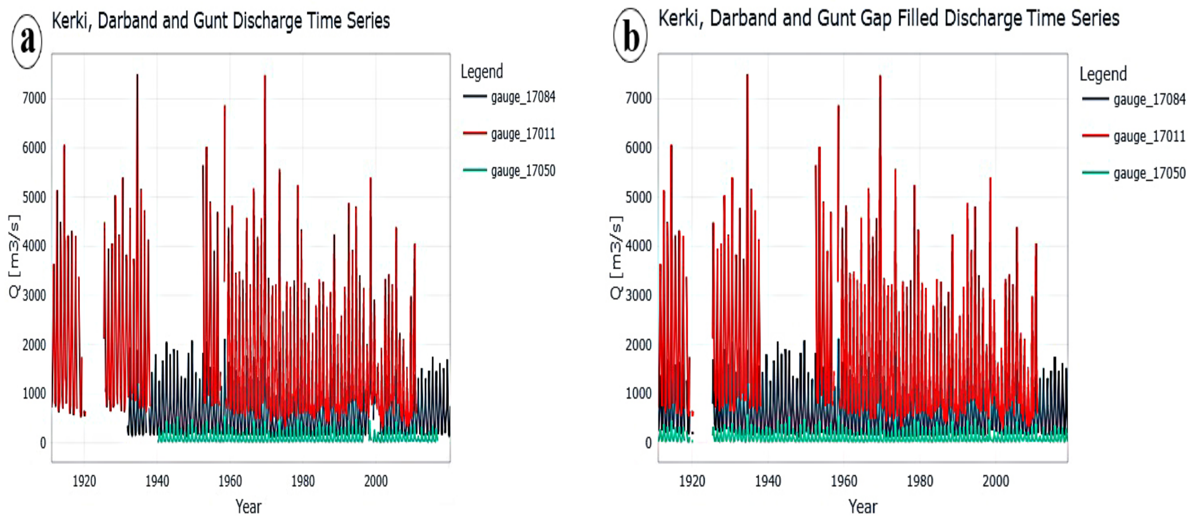


Figure 2. Plate a) The time series of the data from the three gauging stations are shown (Uzbekhydromet and Tajikhydromet, Annual Hydrological Yearbooks). 17011 is Kerki gauge station data, 17050 is Gunt/Khorog gauge station data, and 17084 is Vaksh Darband gauge station data. Plate b) Gap filled time series of the same stations

Often when working with data for forecasting, one may encounter the problems of data gaps, which is what we have in our case. If the missing data is not large and there is a lot of available data, the lost data can fill in by modeling based on previously available results. In our case, we applied this technique and used a command in the RStudio program that helps to automatically fill in the appropriate small gap based on data from a similar ten-year norm. This helped to fill in the data at 17050 is Gunt/Khorog gauge station data which had missing data in 1997 and was filled with missing data in 1998, 1999, and 2000 at 17084 is Vaksh Darband gauge station using a linear model. Now our graph looks as shown in (Figure 2, plate b).

And, when forecasting river discharge at seasonal scales, it is interesting to account for atmospheric teleconnections to see if remote Ocean conditions can improve forecasting. We collected information on several relevant ocean indices from 1950 until 2020. A comprehensive list of ocean indices is provided here: https://psl.noaa.gov/data/climateindices/list/. For our work, we collected the following data using the rsoi R package (https://github.com/boshek/rsoi/):

- Antarctic Oscillation (AAO)
- Arctic Oscillation (AO)
- Dipole Mode Index (DMI)
- Oceanic Nino Index (ONI)
- Southern Oscillation Index (SOI)
- North Pacific Gyre Oscillation (NPGO)
- Multivariate ENSO Index Version 2 (MEI)
- North Atlantic Oscillation (NAO)
- Pacific Decadal Oscillation (PDO)

Figure 3 shows a comprehensive plot of the standardized ocean indices time series.

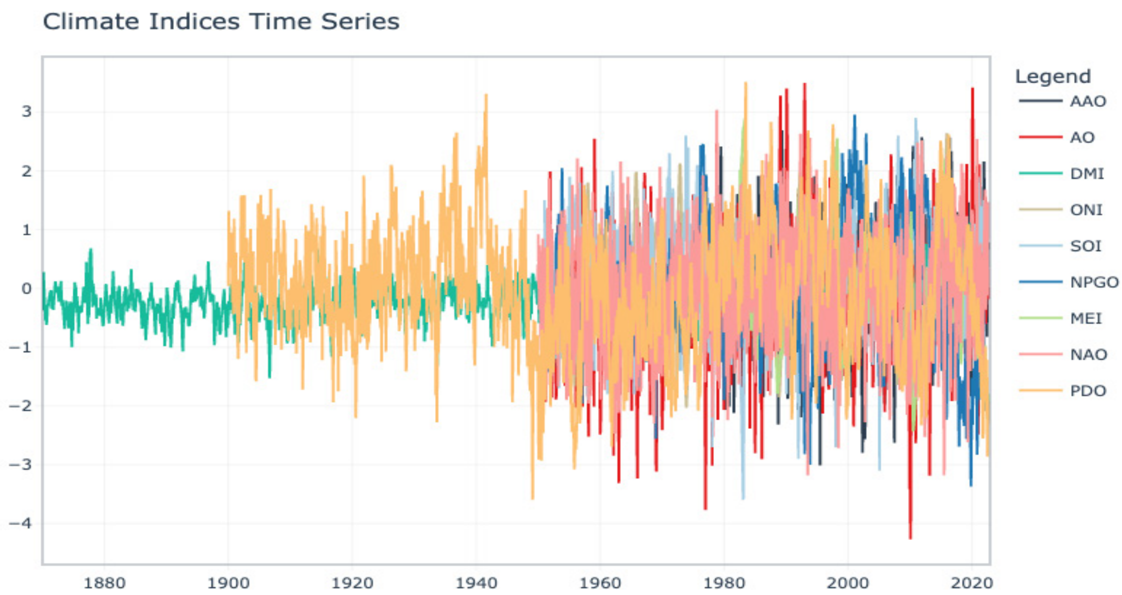


Figure 3. Available set of time series for forecasting. The ocean indices ((AAO), (AO), (DMI), (ONI), (SOI), (NPGO), (MEI), (NAO), (PDO) (Reynolds et al., 2002), (Reynolds & Smith, 1994)).

To see the relationship between water runoff data and ocean indices, must center and scale the runoff data 17011 Kerki gauge station, 17050 Gunt/Khorog gauge station data, and 17084 Vaksh Darband gauge station, so it could be compared to the index data. Of course, we did it, you will see the normalized data to test teleconnection about the water flow (see Figure 4).

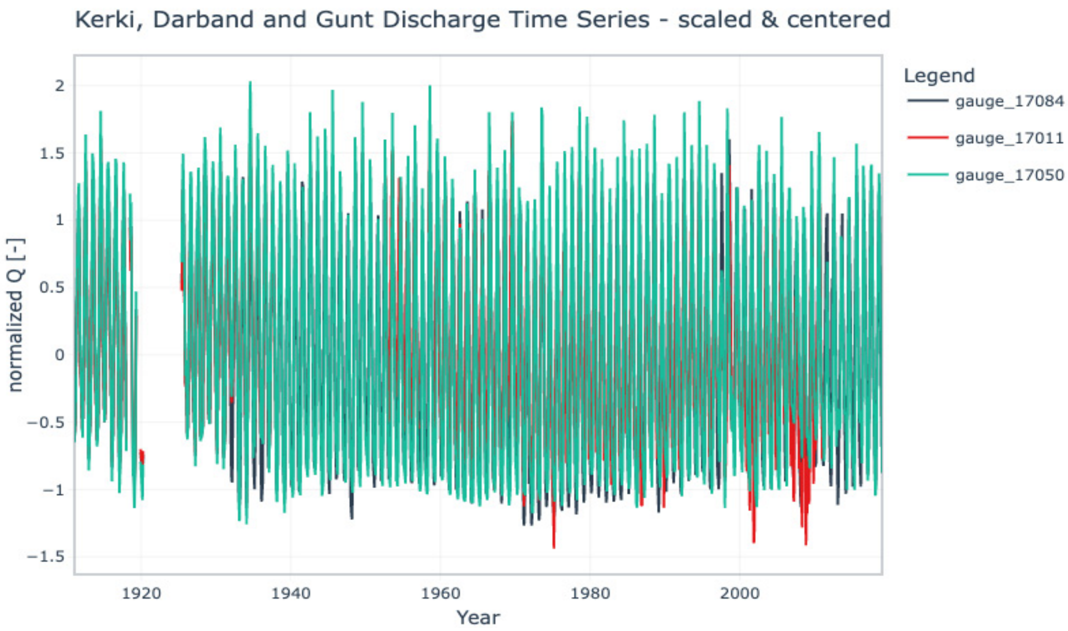


Figure 4. Scaled and centered discharge data from 17011 Kerki gauge station, 17050 Gunt/Khorog gauge station data, and 17084 Vaksh Darband gauge station.

Speaking about the characteristics of the basin, based on the obtained data, we can confirm that the geography of the basin is quite complex, and the water discharge shows high interannual fluctuations. It is also worth noting that the data taken from the gauging stations have gaps, which makes our work difficult. For now, this is all the information we have, and we hope that this data will be enough to build a good forecasting model.

2.2 Literature Review & Methods

From our side, we studied a lot of articles before starting work. Let's dwell on some of them. For example, in this article (Annina Sorg 1,2*, Tobias Bolch 3,4, Markus Stoffel 1,2, Olga Solomina 5 and Martin Beniston 1, 2012), the authors tried, to use the available data from the stations and update the information taken from the GIS systems, to model individual sections of the Tien Shan mountains to be able to predict future changes in the volume of glaciers and, accordingly, the flow of water knowing the metrological data for coming years in the selected region. To measure environmental, social, and economic impacts.

Next (Kaya et al., 2019) authors in this article wanted to show “Assessment and Modeling with GIS and RS Data of the Land Use Effects on Water Quality of Mamasın Dam” (page 1522)

First, they took information from satellites. Next, samples were taken from soils at eight points on land, and data were taken from them in laboratories. Further, all information from satellites and the laboratory was processed in the R/RStudio program, having downloaded ready-made free function packages in advance.

Then we want to discuss the work by (Kostianoy et al., 2013). The authors of this article “demonstrated modern capabilities of satellite remote sensing technologies in environmental monitoring and showed examples of the use of satellite data and imagery for the analysis of morphometric characteristics, sea/lake level, temperature, water quality, wind speed, wave height in the main water bodies in Turkmenistan.” (page 228).

In this paper, the main water bodies of Turkmenistan are selected for measurement. Caspian Sea, Kara-Bogaz-Gol Bay, Sarykamysh and Altyn Asyr Lakes, and the Amu Darya River. Measurements were taken to calculate the level of water bodies, water temperature, wind, and consequently, the level of waves. Information was taken from satellites from information banks, then calibrated and corrected using GIS programs.

And in (White et al., 2014) this study, the authors set a goal to identify the impact of climate change in the long term in the Amu Darya River basin, as well of possible consequences on economic and environmental issues, during 2070-2099.

There are three scenarios were taken that may be in the future between 2070-2099. This is a scenario, of moderate and severe effects of climate change. Also, these data were compared with data obtained on water runoff from 1961 to 1990 to obtain future scenarios for changes in water runoff.

The authors in this paper took information from different sources: (based on available data) in most cases this was information taken from Soviet archives, from aero photographs taken during the Soviet era, and if possible, from public meteorological stations, and took data from current GIS technology and computing software.

Then we studied article (Apel et al., 2018) where authors set themselves the goal of building a model that will work on almost all relatively 13 rivers in Central Asia, when predicting the volume of water in the irrigation season for 6 months from April to September. These authors would identify and choose the best predictors for each month and for a particular river. Data from gauging stations were taken as measurements to verify the data obtained from the model, as well as to correct the results obtained. Also, to include predictors affecting the river flow, data were taken from the MODIS program, such as precipitation, snow, temperature, etc.

The authors did data analysis automatically using the RStudio program, by combining individual factors and in their totality, as a result, a large number of checks came out that the program itself carried out and found the best combinations.

This model will potentially work in all river basins in Central Asia. But of course, in the case of regulated basins, prediction becomes difficult since human flow alterations can only be predicted of the operating regime of a dam is known. This is foe example not the case for Nurek dam discharge which will cause problems for forecasting. However, since our model is learning from past, it factors in past operating regimes and, in this sense, can at least on average anticipate to what extent anthropogenic factors influence discharge. Also, not all models met the required accuracy at the beginning of April, and only May or June improved. This also creates problems because this program is to use data to calculate irrigation areas, and they should be ready by April.

Hydrologists have learned how to predict the volume of water runoff for a long time. And of course, because of this, there are many ways and types of forecasting. It is impossible to choose one of them as an ideal predictive tool since each method is suitable in certain conditions and for each forecasting intention. It affects the forecasting period, whether it is short-term, medium-term, or long-term forecasting. And, of course, the choice of prediction is influenced by the available information.

Let's look at what we want and what we have for this. We have discharge data from 3 gauging stations also have SRTM DEM data that was used to delineate basins and convey height ranges, climate data for this, the relevant literature has been studied (Karger, n.d.), (Karger, Wilson, et al., 2021), (Karger et al., 2020), (Karger, Lange, et al., 2021), (Karger et al., 2017), (Karger, Dirk Nikolaus et al., 2020), (Beck et al., 2020), (Brun et al., 2022) (daily precipitation and temperature values, aggregated to monthly values, are taken from CHELSA climate data and used to drive the model (ISIMIP Repository, n.d.)), land cover information (Copernicus Global Land Service: Land Cover 100m: Collection 3: Epoch 2019: Globe.), snow and glacier information within the basin (Randolph Glacier Inventory – a Dataset of Global Glacier Outlines: Version 6.0: Technical Report). And we, using these data, would like to predict in the future volume of river flow for six months, which is called seasonal forecasting.

We found the ideal option that was previously used (T. Siegfried & B. Marti, 2022). They do empirical simulations. Time-series-based hydrological models use past runoff observations and auxiliary variables such as watershed snow cover and, or precipitation to predict the future based on past learned patterns. Also, the QGIS program was used for information on GIS technologies, and the R / RStudio program was used to work with runoff information and runoff modeling. Today it is one of the most modern modeling methods, that was a crucial choice factor.

Another important source was the article by (Dixon & Wilby, 2019). Here, the articles had two goals. First, they assessed the spatial and temporal variation in correlations between selected climate regimes and precipitation, temperature, and reservoir inflows for three sizeable hydroelectric power plants in CA. Second, they developed a simple procedure for generating early forecasts of summer inflow at one of these sites (Nurek Reservoir, Tajikistan) using the previous Niño 3.4 winter. So, to speak, measuring the temperature of the Pacific Ocean. This idea prompted us to use in our work the concept of measuring significant sources of climate change. Scientists have already done the calculation of the factors of the Earth and left it in open sources.

Speaking about the methodology in this work, data from earth remote sensing satellites, data from previous similar scientific works, and time series data from three hydrological gauging stations are used. For geospatial analysis, the QGIS program is used, and for empirical time series modeling, the R/RStudio program is used. Different existing approaches for time series forecasting have been tested (Dixon & Wilby, 2019), (Apel et al., 2017), (Gerlitz, 2020).

Now we need to collect all the data that we will use for empirical modeling and identify the relationship between the factors that affect the river flow. To do this, we began work with the correlation of the volume of the river runoff, with the data of the ocean indices we obtained. Nowadays, our aim of the work is to understand better and analyze the characteristics of the basin to use these data further to predict the river flow in seasonal periods. We started building our model with ocean index teleconnections data and gauging station data. A lag diagnostic was carried out as shown in Figure 5.

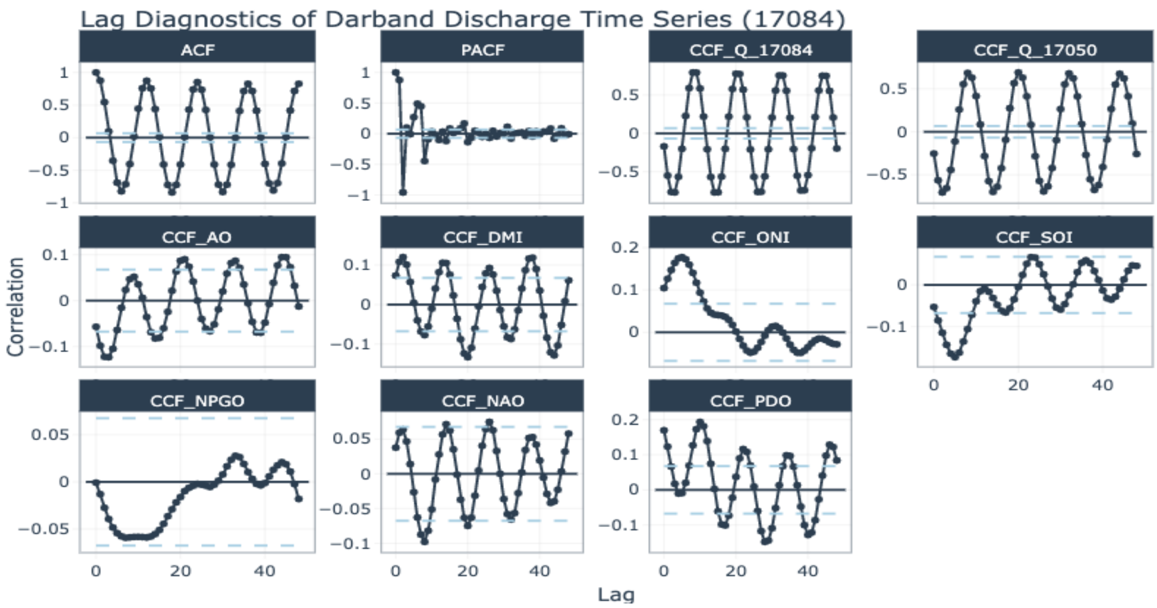


Figure 5. Analyses autocorrelation (ACF) and partial autocorrelation (PCAF) for discharge at Darband station and cross-correlations with ocean indices.

For forecasting, it is standard to differentiate between model calibration and model validation. For this purpose, we define the 1950 to 2000 period as training period and the period from 2001 – 2020 as the period over which we assess model performance.

Since we are interested in seasonal forecasting, i.e., to forecast mean discharge over the next coming 6 months. Instead of multi-step forecasting, we use a simple one-step forecast approach by constructing a target time series that consists of mean discharge data averaged from t+1 and t+6.

3. Results

We have used linear time series modeling to investigate the potential of different predictors to help forecast mean monthly discharge averaged over 6 months at Darband Station (Index 17084). The philosophy here is to iteratively test different types of predictors by including them in a forecasting model for testing their predictive power.

In a first step, we have added time-based predictor features that help to improve modeling. These features include indices of the year, half-year, the semester, the quarter, and month. A linear model that uses these predictors alone can already very well explain seasonal variability as well as long-term trends (through the inclusion of the year as a predictor). We obtain an adjusted R-square value of 0.7534 as model performance.

In a second step, we have added Fourier-based frequency features to our linear regression. Based on the diagnostics of lags (see Figure 5), we have included period 8 and 19 features and tested a maximum of 2 Fourier orders. By inclusion of these predictors, we obtained and adjusted R-squared value of 0.7517. There was thus not much information gain from these predictors. The reason is that we have already added the seasonality to our model with the time-based features in a first step.

In a third and final step, we have included different lagged features in our simple linear forecasting model. Detailed results are provided in the regression table in Appendix 7.1 below. The importance of lagged predictors in time-series based modeling is apparent. With an adjusted R-squared value of 0.9905, we could identify highly relevant predictors for the forecasting task.

These are of course encouraging news for us Figure 6 shows model results. With these preliminary investigations, we can easily show important points. These are

Seasonal discharge at Darband station can be forecasted.Time-based and lagged features are highly relevant for forecasting. Climate indices can help to improve model forecasting skills.

For comparison, We would like to note that (Apel et al., 2018) made studies to predict water resources for all the large rivers of Central Asia, namely, there are 15 of them. In this study for the Amu Darya separately until April, the results showed R-squared best 0.878 mean 0.839. As a result, the researchers' data improves when they forecast the April-September data in June, including the real April and May data already received, it turns out Best 0.982 Mean 0.977. But it must be considered that for countries using water for irrigation, data received before 1-st April is more important.

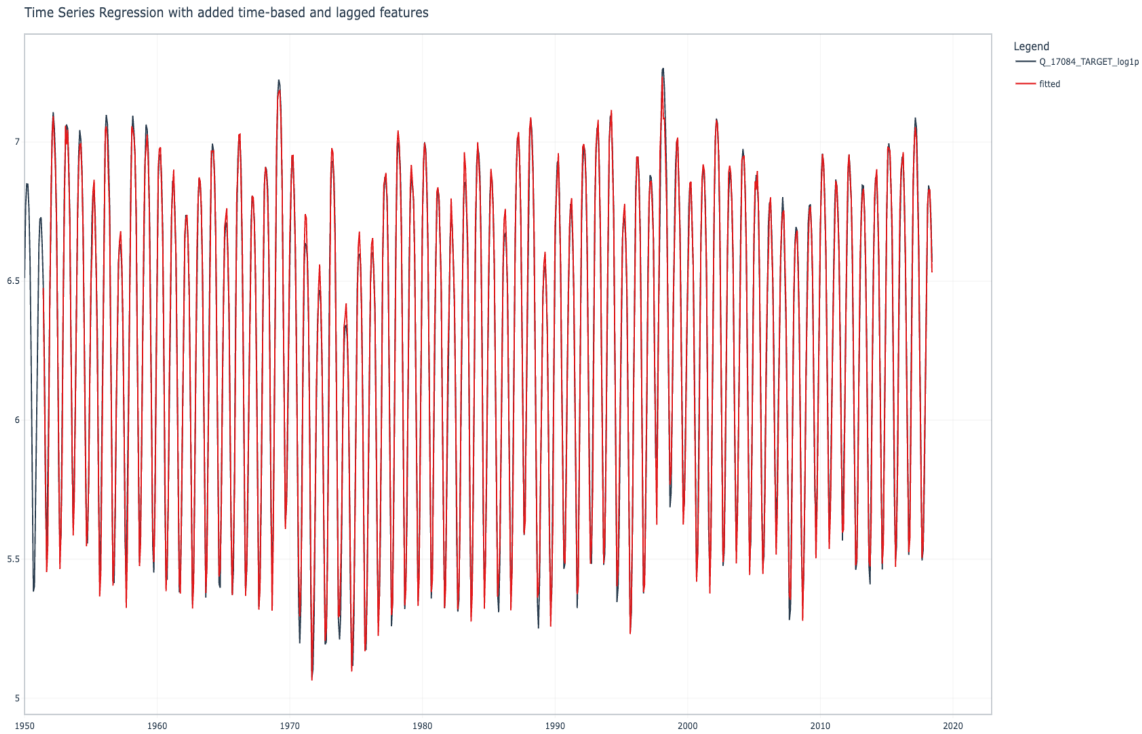


Figure 6. Modeling results using a simple linear model to predict seasonal discharge

This simple model yields satisfactory results in terms of its ability to predict seasonality and peak and minimum flows. However, while the overall R squared value is excellent, the model does not yet a sufficiently good job to predict high and low flows properly.

4. Conclusions and Outlook

In this paper, we have collected relevant information about the Amu Darya River basin using data from three gauging stations. We have cleaned these data, filled gaps, and prepared them for forecasting. Together with this, we have prepared relevant GIS information. Finally, monthly data from climate indices was retrieved and equally prepared for forecasting

Also, we have seen that at 3 different stations of the Amu Darya River, although the volumes of water are different, the amplitudes of the increase in the volume of water and the decrease in the volume of water are still the same. This led us to realize that the frequency of the change in water volume changes in the same way, albeit with different volumes, and we decided to average and rescale the data to get station data that show almost the same amplitude of increase and decrease. In our study, we included ocean indices that which are relevant with regard to climate teleconnections, and in the future, we will analyze how much the ocean indices affect our data and which oceans most affect our river flow in gauging stations and whether they affect at all? And we will reveal with what delay they have influenced. That is if changes occur in the oceans, how long will it take to affect our data gauging stations? We have developed a simple linear model 1-step seasonal forecasting method and tested this method by gradually increasing the number of relevant predictors. Our results are very promising even though we have so far only used a very simple model. In the future, we will take these findings to the next level by using advanced feature engineering and machine learning models to help to work towards operational seasonal forecasting of discharge time series in Central Asia. Also, we want to include upstream precipitation and snow cover information in our predictive model to account for water resources stored as snow in the high mountain regions.

After calculating these data, we intend to implement the data obtained from remote sensing satellites as factors influencing the flow of the river. Specifically, satellite-derived snow cover information will also be investigated.

The received data have digital designations in the attribute's column in the QGIS program, they will be transferred to the R / RStudio program, where the model itself will be built and we will already compare it with the data taken from the stations. When the model suits us with its accuracy and we choose the right factors to include in our model, we will incorporate future climate data into our model to get a forecast for the next season.

Analyzing the data, it can easily be seen that the amount of water has been decreasing in recent years. Our study does not intentionally reveal the reasons for the decrease in flow since the survey is only aimed at measurements and model construction. But considering that the source of the river is snow and ice in the highlands of this region, we can assume that we are losing glacier volumes and, consequently, the amount of water. Despite this, I would like to note that today the downstream countries intend to increase their areas for irrigation, and this will require more water than is used today. But is there enough water for this? To do this, we need basin modeling, as well as subsequent river flow prediction. The accuracy of the finished model will be proved by comparing the obtained results of the model with the available data from gauging stations in the same period. Percentage accuracy will show whether this model will be a good tool for planning further agricultural decisions.

Summing up, it should be emphasized that today's technologies for building a hydrological model can significantly help in the management of water resources both at the national and regional levels. But the accuracy of the hydrological model depends on the availability of accurate data. Operational runoff forecasting will become possible if data on river discharge are received online and on-time. We hope that having seen our good intentions of using data for regional management, the countries of Central Asia and Afghanistan will begin to share their information from gauging stations more easily. This is currently one of the big challenges one must consider when trying to model future runoff.

5. Acknowledgement

Being today a student of the master's program "Integrated Water Resources Management" at the Kazakh-German University, I would like to express my gratitude to all the management, all teachers, and all sponsors, thanks to which our students have the opportunity to gain knowledge and work in a scientific direction.

I am also grateful to the USAID Regional Water and Vulnerable Environment Activity project for sponsoring this work.

And of course, I want to thank my supervisor Dr. Tobias Siegfried, for his great work, patience, and thoughts which he kindly presented to me. This work would not have been possible without his help.

6. References

About Turkmenistan. (n.d.). Retrieved July 2, 2022, from http://www.dntours.com/index.php?option=com_content&view=category&id=39&Itemid=69&lang=en

Agreement between the Republic of Kazakhstan, the Kyrgyz Republic, the Republic of Tajikistan, Turkmenistan and the Republic of Uzbekistan on Cooperation in the Field of Joint Management on Utilization and Protection of Water Resources from Interstate Sources. (n.d.). Retrieved July 2, 2022, from http://www.icwc-aral.uz/statute1.htm

Agreement between Uzbekistan and Turkmenistan on cooperation on water management questions. (n.d.). Retrieved July 2, 2022, from https://cis-legislation.com/document.fwx?rgn=25361

Annina Sorg1,2*, Tobias Bolch3,4, Markus Stoffel1,2, Olga Solomina5 and Martin Beniston1. (2012). Climate change impacts on glaciers and runoff in Tien Shan (Central Asia). *NATURE CLIMATE CHANGE*. https://www.nature.com/articles/nclimate1592

Apel, H., Abdykerimova, Z., Agalhanova, M., Baimaganbetov, A., Gavrilenko, N., Gerlitz, L., Kalashnikova, O., Unger-Shayesteh, K., Vorogushyn, S., & Gafurov, A. (2018). Statistical forecast of seasonal discharge in Central Asia using observational records: Development of a generic linear modelling tool for operational water resource management. *Hydrology and Earth System Sciences*, 22(4), 2225–2254. https://doi.org/10.5194/hess-22-2225-2018

Apel, H., Abdykerimova, Z., Agalhanova, M., Baimaganbetov, A., Gerlitz, L., Kalashnikova, O., Unger-Shayesteh, K., & Vorogushyn, S. (2017). *Statistical forecast of seasonal discharge in Central Asia for water resources management: Development of a generic linear modelling tool for operational use*. 31.

Beck, H. E., Wood, E. F., McVicar, T. R., Zambrano-Bigiarini, M., Alvarez-Garreton, C., Baez-Villanueva, O. M., Sheffield, J., & Karger, D. N. (2020). Bias Correction of Global High-Resolution Precipitation Climatologies Using Streamflow Observations from 9372 Catchments. *Journal of Climate*, 33(4), 1299–1315. https://doi.org/10.1175/JCLI-D-19-0332.1

Brun, P., Zimmermann, N. E., Hari, C., Pellissier, L., & Karger, D. N. (2022). Global climate-related predictors at kilometre resolution for the past and future. *Earth System Science Data Discussions*, 1–44. https://doi.org/10.5194/essd-2022-212

Buchhorn, M., Smets, B., Bertels, L., De Roo, B., Lesiv, M., Tsendbazar, N.-E., Herold, M., & Fritz, S. (2020). *Copernicus Global Land Service: Land Cover 100m: collection 3*: epoch 2019: Globe. https://doi.org/10.5281/zenodo.3939050

Dixon, S. G., & Wilby, R. L. (2019). A seasonal forecasting procedure for reservoir inflows in Central Asia. *River Research and Applications*, rra.3506. https://doi.org/10.1002/rra.3506

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., & Alsdorf, D. (2007). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, 45(2), RG2004. https://doi.org/10.1029/2005RG000183

Gerlitz, L. (2020). Climate informed seasonal forecast of water availability in Central Asia_ State-of-the-art and decision making context. *Water Security*, 12.

ISIMIP Repository. (n.d.). Retrieved April 28, 2022, from https://data.isimip.org/10.48364/data.isimip.org

Kamil, I. (2021). Afghanistan, the Amu Darya Basin and Regional Treaties. *Chinese Journal of Environmental Law*, 5(1), 37–62. https://doi.org/10.1163/24686042-12340063

Karger, D. N. (n.d.). High-resolution monthly precipitation and temperature time series from 2006 to 2100. 10.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). *Climatologies at high resolution for the earth's land surface areas*. Scientific Data, 4(1), Article 1. https://doi.org/10.1038/sdata.2017.122

Karger, D. N., Lange, S., Hari, C., Reyer, C. P. O., & Zimmermann, N. E. (2021). *CHELSA-W5E5* v1.0: W5E5 v1.0 downscaled with CHELSA v2.0. https://doi.org/10.48364/ISIMIP.836809

Karger, D. N., Schmatz, D. R., Dettling, G., & Zimmermann, N. E. (2020). *High-resolution monthly precipitation and temperature time series from 2006 to 2100. Scientific Data*, 7(1), Article 1. https://doi.org/10.1038/s41597-020-00587-y

Karger, D. N., Wilson, A. M., Mahony, C., Zimmermann, N. E., & Jetz, W. (2021). *Global daily 1 km land surface precipitation based on cloud cover-informed downscaling. Scientific Data*, 8(1), 307. https://doi.org/10.1038/s41597-021-01084-6

Karger, Dirk Nikolaus, Dabaghchian, Babek, Lange, Stefan, Thuiller, Wilfried, Zimmermann, Niklaus E., & Graham, Catherine H. (2020). *High resolution climate data for Europe* (1.0, p. 10 KB, KB) [NetCDF,PDF]. EnviDat. https://doi.org/10.16904/ENVIDAT.150

Kaya, E., Agca, M., Adiguzel, F., & Cetin, M. (2019). Spatial data analysis with R programming for environment. *Human and Ecological Risk Assessment: An International Journal*, 25(6), 1521–1530. https://doi.org/10.1080/10807039.2018.1470896

Kostianoy, A. G., Lebedev, S. A., & Solovyov, D. M. (2013). Satellite Monitoring of the Caspian Sea, Kara-Bogaz-Gol Bay, Sarykamysh and Altyn Asyr Lakes, and Amu Darya River. In I. S. Zonn & A. G. Kostianoy (Eds.), The Turkmen Lake Altyn Asyr and Water Resources in Turkmenistan (Vol. 28, pp. 197–231). *Springer Berlin Heidelberg*. https://doi.org/10.1007/698_2013_237

Muradov R. (2013, November 22). Along the banks of the Jayhoon: The innkeepers and settlements on the ancient trade route. [Vdol' beregov Dzhejhuna: Postoyalye dvory i poseleniya na drevnem torgovom puti.] https://turkmenportal.com/blog/2182

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., & Wang, W. (2002). An Improved In Situ and Satellite SST Analysis for Climate. J*ournal of Climate*, 15(13), 1609–1625. https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2

Reynolds, R. W., & Smith, T. M. (1994). Improved Global Sea Surface Temperature Analyses Using Optimum Interpolation. *Journal of Climate*, 7(6), 929–948. https://doi.org/10.1175/1520-0442(1994)007<0929:IGSSTA>2.0.CO;2

T. Siegfried & B. Marti. (2022, March 4). Modeling of Hydrological Systems in Semi-Arid Central Asia.

https://hydrosolutions.github.io/caham_book/

White, C. J., Tanton, T. W., & Rycroft, D. W. (2014). The Impact of Climate Change on the Water Resources of the Amu Darya Basin in Central Asia. *Water Resources Management*, 28(15), 5267-5281. https://doi.org/10.1007/s11269-014-0716-x

7. Appendix

7.1 Regression results of the final predictive model in R/RStudio program

```
Call:
stats::lm(formula = .formula, data = df)

Residuals:
      Min        1Q    Median        3Q       Max
-0.202944 -0.031835 -0.000599  0.029747  0.247848

Coefficients: (1 not defined because of singularities)
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                71.0714530 24.1159430   2.947 0.003304 **
date                       -0.0120865  0.0041490  -2.913 0.003682 **
AO                          0.0001982  0.0026415   0.075 0.940205
DMI                         0.0079191  0.0074115   1.068 0.285637
ONI                         0.0078764  0.0041240   1.910 0.056515 .
SOI                        -0.0006602  0.0030155  -0.219 0.826753
NPGO                        0.0026160  0.0020000   1.308 0.191256
NAO                         0.0012534  0.0024835   0.505 0.613933
PDO                         0.0025887  0.0036210   0.715 0.474888
half                       -0.2514253  0.0162641 -15.459  < 2e-16 ***
quarter                     0.1168411  0.0091658  12.748  < 2e-16 ***
month                       0.3620530  0.1262032   2.869 0.004232 **
index.num.zs                       NA         NA      NA       NA
Q_17084_log1p              -0.2238967  0.0136427 -16.412  < 2e-16 ***
Q_17050_log1p              -0.0259440  0.0069232  -3.747 0.000192 ***
year_zs                    93.0820849 31.9504093   2.913 0.003679 **
Q_17084_TARGET_log1p_lag1   1.3469631  0.0241355  55.808  < 2e-16 ***
Q_17084_TARGET_log1p_lag2  -0.3907060  0.0271512 -14.390  < 2e-16 ***
Q_17084_TARGET_log1p_lag6   0.2443534  0.0210934  11.584  < 2e-16 ***
Q_17084_log1p_lag2         -0.0145662  0.0137308  -1.061 0.289094
Q_17084_log1p_lag3         -0.0299532  0.0135525  -2.210 0.027386 *
Q_17084_log1p_lag8         -0.0016282  0.0119806  -0.136 0.891932
Q_17084_log1p_lag9          0.0033455  0.0120777   0.277 0.781854
Q_17050_log1p_lag2         -0.0414540  0.0068524  -6.050 2.26e-09 ***
```

```
Q_17050_log1p_lag8        -0.0096700  0.0075376  -1.283 0.199912
AO_lag2                   -0.0029536  0.0020138  -1.467 0.142874
DMI_lag2                  -0.0080300  0.0074886  -1.072 0.283921
ONI_lag5                   0.0029166  0.0039501   0.738 0.460517
SOI_lag5                  -0.0032836  0.0030763  -1.067 0.286122
NAO_lag8                  -0.0034114  0.0019211  -1.776 0.076177 .
PDO_lag1                  -0.0009719  0.0035321  -0.275 0.783271
PDO_lag10                 -0.0018284  0.0021481  -0.851 0.394939
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05383 on 774 degrees of freedom
  (71 observations deleted due to missingness)
Multiple R-squared:  0.9908,	Adjusted R-squared:  0.9905 
F-statistic: 2793 on 30 and 774 DF,  p-value: < 2.2e-16
```