Modeling the effects of climate change and human activities on the Aral Sea in a reproducible framework

MSc Thesis project

A.B. van der Veen

Delft University of Technology

TU Delft

Modeling the effects of climate change and human activities on the Aral Sea in a reproducible framework

by

A.B. van der Veen

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday, 9 June 2026 at 13:45.

Student number:	4080564	
Project duration:	June, 2025 – June, 2026	
Thesis committee:	Dr. ir. R. Hut,	TU Delft, supervisor
	Prof. dr. ir. M. Hrachowitz,	TU Delft

Cover: By NASA - https://worldview.earthdata.nasa.gov/, Public Domain, https://commons.wikimedia.org/w/index.php?curid=109924994

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

List of Figures

List of Tables

Abstract

Endorheic basins cover roughly 20% of Earth's land surface but account for 50% of water-stressed regions. The Aral Sea is perhaps the most dramatic example of environmental degredation in these systems; since 1960 more than 90% of it's water volume has been lost. Although the expansion of irrigated agriculture, increasing from 47,000 km^2 to 83,000 km^2, is generally considered the primary driver of this decline, the relative contribution of climate variability remains uncertain. This study aims to quantify the relative influence of human water use and climate variability on the Aral Sea water balance using a fully reproducible modelling workflow implemented within the eWaterCycle environment. A reproducible model chain was designed and implemented using the eWaterCycle platform, which promotes FAIR (Findable, Accessible, Interoperable, Reusable) computational hydrology. New workflow components were developed for forcing generation, spatial regridding and downscaling, bias correction, and regional model calibration, and integrated as reusable tools within eWaterCycle. Meteorological forcing was derived from ERA5 reanalysis data (1940-2020) for historical reconstruction and CMIP6 global climate Models for future scenarios (2025-2100). Catchment hydrology was simulated using PCR-GLOBWB2 (PCRaster GLOBal Water Balance model, version 2) following regional calibration against observed discharge data. Simulated discharges were subsequently used as input as inflow to an Aral Volume Balance model developed for this study. Methodologically, the study delivers a modular and reproducible workflow enabling consistent preparation and use of climate forcing data within eWaterCycle. The developed reprocessing tools allowed CMIP6 datasets to be directly applied in PCR-GLOBWB2 simulations, while bias correction ubstantially reduced temperature and precipitation biases. Regional calibration improved discharge simulations from double digit negative KGE to positive values. Hydrologically, the coupled modelling framework successfully reproduced the observed desiccation trend of the Aral Sea, and showed the effect of different models and pathways on the future of the Aral Sea. The study demonstrates that the eWaterCycle platform can be used as a robust, transparent, and reusable workflow environment. The developed tools extend the capabilities of the eWaterCycle ecosystem and enable transparent, reusable hydrological experimentation for endorheic basins and other large-scale water systems.

Acknowledgements

The research presented in this thesis made use of a range of data sources, models, and computational platforms. Simulations and analyses were conducted using the eWaterCycle framework and the SURF Research Cloud infrastructure. River discharge observations were obtained from the Global Runoff Data Centre (GRDC). Climate forcing data were derived from the ERA5 reanalysis and CMIP6 climate model datasets. Earth observation Imagery were obtained from NASA products where relevant. Hydrological simulations were performed using PCR-GLOBWB 2, developed and maintained by researchers at Utrecht University.

The use of generative artificial intelligence tools in this project was in accordance with the guidelines and requirements of "TU Delft OPEN Publishing policies" (n.d.). AI tools were used to support programming, debugging, language editing, and document preparation.

Glossary / Acronyms

BMI	Basic Model Interface
CanESM5	Canadian Earth System Model version 5
CMA-ES	Covariance Matrix Adaptation Evolution Strategy
CMIP6	Coupled Model Intercomparison Project Phase 6
CMIP7	Coupled Model Intercomparison Project Phase 7
DEM	Digital Elevation Model
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA5	ECMWF Reanalysis 5th Generation
GCM	Global Climate Model
GRDC	Global Runoff Data Centre
HBV	Hydrologiska Byråns Vattenbalansavdelning
KGE	Kling–Gupta Efficiency
MIROC6	Model for Interdisciplinary Research on Climate version 6
MPI-ESM1-2-HR	Max Planck Institute Earth System Model version 1.2, High Resolution
MRI-ESM2-0	Meteorological Research Institute Earth System Model version 2.0
NSE	Nash–Sutcliffe Efficiency
PCR-GLOBWB2	PCRaster GLOBal Water Balance model 2
RMSE	Root Mean Square Error
$\mathbf{RMSE}_{\Sigma}$	Root Mean Square Error of cumulative discharge
SSP	Shared Socioeconomic Pathway

Contents

1

Introduction

1.1. Endorheic Basins and the Aral Sea

Endorheic basins are drainage basins that lack a natural surface outflow to the ocean. Water is retained within the basin and lost through either evaporation, infiltration, or human abstraction. As a result, rivers in endorheic systems terminate in inland lakes, wetlands, or salt flats, rather than discharging to the sea (Yapiyev et al., 2017, 2024). Endorheic basins cover roughly 20% of the Earth's land surface and account for 50% of water-stressed regions (Wang et al., 2018) and are home to approximately 400 million people (Varis & Kummu, 2012). Figure 1.1 shows their global distribution.

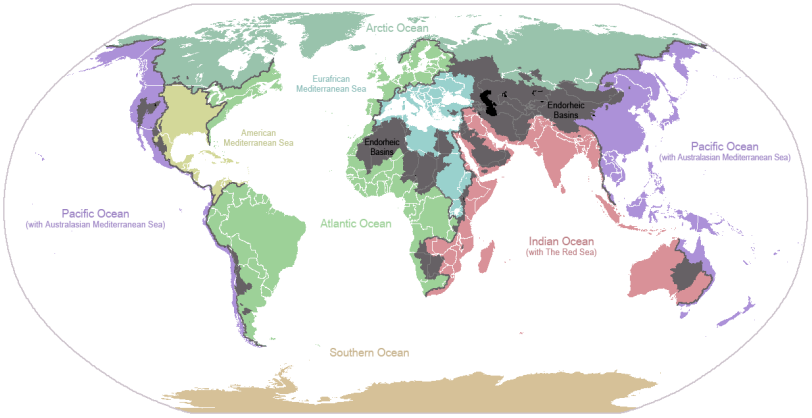


Figure 1.1: Overview worldwide drainage basins with endorheic basins shown in gray. (Wikimedia Commons contributors, 2015)

Endorheic basins are particularly sensitive to climatic variability and change due to their closed hydrological nature. Changes in precipitation, temperature, and evaporation directly affect water storage in terminal lakes and groundwater systems, often leading to pronounced fluctuations in lake levels and surface area (Yapiyev et al., 2017; Zhang et al., 2021). Unlike exorheic basins, where excess water can be transported downstream, endorheic systems accumulate the effects of long-term imbalances in the water budget. Consequently, climate-driven shifts in hydroclimatic conditions can result in rapid salinization, ecosystem degradation, and the loss of open water bodies. On the other hand, long term water accumulation can also lead to flooding (Byrne et al., 2024).

Climate change is already affecting the endorheic basins: basins in the Tibetan plateau are getting 'wetter', basins on the Mongolian plateau getting drier, most of the basin in Central Asia are getting wetter, except for the southern Aral sea. (Zhang et al., 2021). Beyond the local effect, water removed from endorheic water storage contributes directly to sea level rise in a way similar to that of the melting of glaciers (Wang et al., 2018).

Human interventions further influence the hydrology of endorheic basins. Large-scale water withdrawals for irrigation, reservoir construction and land-use change have significantly altered natural flow regimes, often strengthening the impacts of climate variability. The examples of water bodies drying out, such as the Aral Sea, illustrate how anthropogenic water use can dominate the basin-scale water balance (Yapiyev et al., 2017). Endorheic basins are home to approximately 410 million people (Varis and Kummu, 2012) making their water security a matter of socio-economic importance. Because of their fragile and sensitive nature, it is crucial to improve the modeling and understanding of their water balance under combined climatic and anthropogneic pressures.

The Aral Sea Basin is perhaps the most dramatic example of degradation in endorheic systems (T. T. Berdimbetov et al., 2020; Frenken, 2013; Touge et al., 2024). Once the world's fourth-largest lake, the Aral Sea has lost more than 90% of its volume since the late 1950s (Cretaux et al., 2013; He et al., 2022; Micklin et al., 2014) (Figure 1.2). This desiccation is primarily attributed to large-scale anthropogenic water diversions for agricultural irrigation, though recent studies highlight the increasing role of climate-driven changes (Boomer et al., 2009; Tao et al., 2020; Zhang et al., 2021)

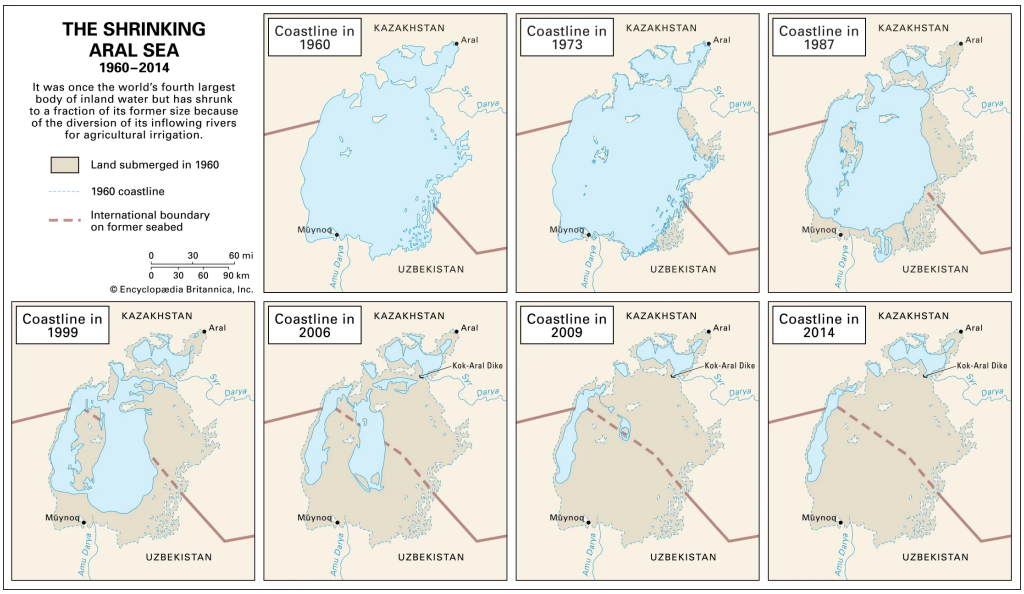


Figure 1.2: Drying of the Aral Sea between 1960 and 2014, showing the decrease in lake area, fragmentation into separate basins, and shoreline retreat. Map graphic published by Encyclopædia Britannica (2026), based on historical reconstructions and data by Philip Micklin. Note the size of the sea in 1960, which extended approximately 400 km from north to south and 300 km from east to west.

1.2. Research Gap

Research into the Aral Sea and Aral Sea basin exist in a wide variety of subjects. Even within hydrology, there has been a lot of research which has varied from large-scale water balance assessments and groundwater flux estimates to detailed historical reconstructions using archaeological evidence (Aus Der Beek et al., 2011; T. Berdimbetov et al., 2024; T. T. Berdimbetov et al., 2020; Boomer et al., 2009; Brite, 2016; Deng & Chen, 2017; Duan et al., 2024; Krapivin et al., 2019; Stulina & Eshchanov, 2013; Tao et al., 2020).

Few studies provide detailed models of the whole Aral Sea basin (Touge et al., 2024) and relatively few use future climate scenarios (e.g. CMIP6) to project the evolution of the Aral Sea system towards the end of the 21st century (He et al., 2022; Pan et al., 2022).

The quantitative attribution to the declining water levels of the Aral Sea due to anthropogenic water use or climate change are also under scientific debate. While it is widely agreed that human influence is the primary driver, estimates on the exact contribution of climate change are unclear, even though investigative studies exist (Aus Der Beek et al., 2011; Cretaux et al., 2013; Duan et al., 2024).

Perhaps the most critical gap lies in the transparency and reproducibility of hydrological research. Computational hydrology is currently grappling with a reproducibility crisis (Hall et al., 2022). Stagge et al. (2019)

found that only 1.6% of tested articles in the field were fully reproducible. Many hydrological models function as "black boxes," relying on closed workflows or incomplete methodological descriptions that prevent independent verification or the reuse of results or methods. Some of the bottlenecks are the lack of clear instructions or the use of hard-coded file paths that only work on the system of the authors (C. Hutton et al., 2016).

To address these limitations, the increasing availability of open scientific infrastructure, such as the *eWaterCycle* platform (Hut et al., 2022), provides an opportunity to promote FAIR (Findable, Accessible, Interoperable, Reusable) and Open Science principles (De Visser et al., 2023; Hall et al., 2022; Hut, 2022).

By using software containers and standardized interfaces such as the Basic Model Interface (BMI) (E. Hutton et al., 2020), the hydrological model can be separated from the specific experiment, ensuring that workflows are transparent and verifiable. In combination with workflow management systems such as Snakemake (Mölder et al., 2025), this enables fully reproducible and automated modelling pipelines. This in turn facilitates reuse and reproducibility by other researchers (Hut & Hall, 2025).

1.3. Research Goal and Questions

Building upon the methodological frameworks and principles of sustainable data analysis introduced in the previous section, this thesis aims to contribute to narrowing the reproducibility gap in hydrology by developing a fully transparent and reusable modelling workflow. By applying this framework to the Aral Sea basin, it enables a robust attribution analysis of the system's historic and future desiccation.

1.3.1. Study Goals

The objectives of this research are twofold:

1. **Methodological Objective:** To develop and implement a reproducible modelling workflow using the *eWaterCycle* platform and *Snakemake*, enabling modular and transparent hydrological simulations.
2. **Research Objective:** To quantify the relative contributions of climate variability and anthropogenic water use to historical changes in the Aral Sea water balance, and to evaluate future system behaviour under climate scenario projections.

1.3.2. Research Questions

To achieve these goals, this study addresses the following central research question: *To what extent can a reproducible modelling framework quantify the relative impact of human water use versus climate change on the Aral Sea water balance from 1940 to 2100?*

This central question is supported by sub-questions:

1. **Historical Validation:** To what extent can the historical (1940–2020) river inflow and lake level dynamics of the Aral Sea be reproduced using a modular modelling chain forced with ERA5 reanalysis data?
2. **Future Projections:** What are the projected river inflows and Aral Sea water level changes for the period 2025–2100 under climate scenario forcing?
3. **Quantitative Attribution:** What is the relative contribution of climate variability and anthropogenic water use to historical and projected changes in the Aral Sea water balance?

1.3.3. Expected Contributions

By answering these questions, this research contributes to hydrology in three ways.

Research contribution: It provides an updated assessment of the relative impacts of climate variability and anthropogenic water use on the Aral Sea water balance, including historical attribution and sensitivity analysis under CMIP6 scenarios.

Methodological contribution: It delivers a "FAIR by design" workflow that lowers the technical barrier for others to build upon or adapt this study to other hydrological applications.

Software contribution: It contributes scripts and workflow components to the *eWaterCycle* ecosystem, improving reproducibility and enabling reuse in future hydrological studies.

1.4. Modelling Workflow

The research workflow consists of three main components and is described in further detail in Chapter 4.

- **Forcing generation and preprocessing:** This step includes data preparation, spatial and temporal harmonization, and bias correction using tools from the *eWaterCycle* framework, complemented by custom scripts and tools developed for this study.
- **Regional hydrological modelling:** PCR-GLOBWB2 is used as the primary distributed hydrological model, which includes an irrigation water demand module (E. H. Sutanudjaja et al., 2018). For this model, a calibration workflow was developed, as well as a BMI-based experiment in which river storage is modified at each timestep to represent processes that are not natively included in PCR-GLOBWB2, within the *eWaterCycle* framework. The HBV model (Lindström et al., 1997) was used for exploratory analysis and workflow testing, running approximately three orders of magnitude faster than PCR-GLOBWB2. Both models are further described in section 3.2.
- **Lake water balance modelling:** Simulated river discharge from the hydrological models is used as input for an Aral Sea volume–area–elevation balance model developed in this study.

Figure 1.3 provides a conceptual overview of the full modelling chain.

1.5. Thesis Outline

This thesis is structured as follows: Chapter 2 provides the geographical and historical context of the Aral Sea Basin. Next, Chapter 3 describes the data and models used, including ERA5/CMIP6 forcing and the PCR-GLOBWB2 and HBV models, while Chapter 4 explains the implementation of the workflow and the Aral Sea volume balance model. The results of the historical reconstruction and future projections are presented in Chapter 5, followed by a methodological and scientific evaluation in Chapter 6. To conclude, Chapter 7 contains the conclusion and recommendations for future research.

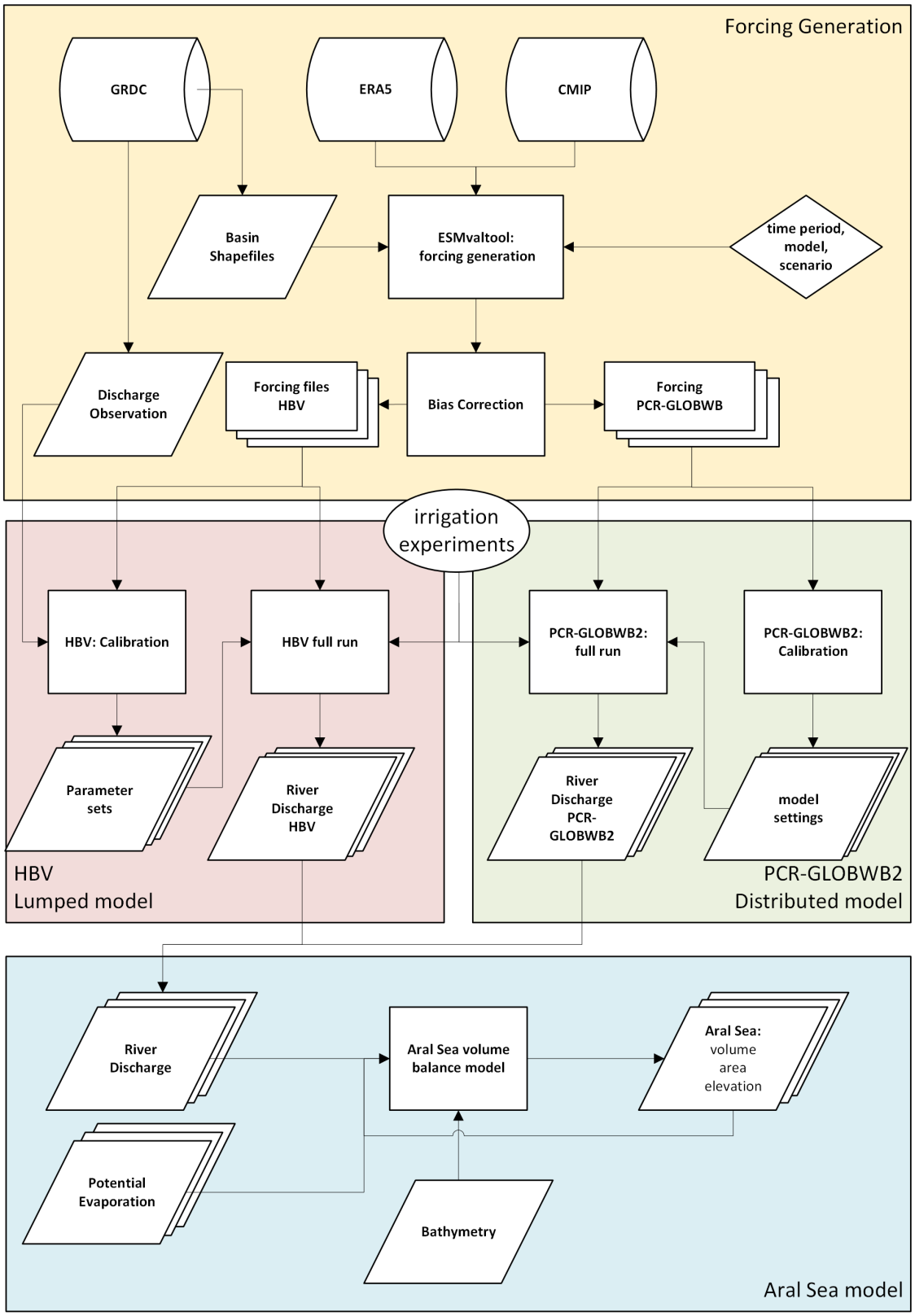


Figure 1.3: Conceptual overview of the modeling workflow.

2

Study Area and Regional Setting

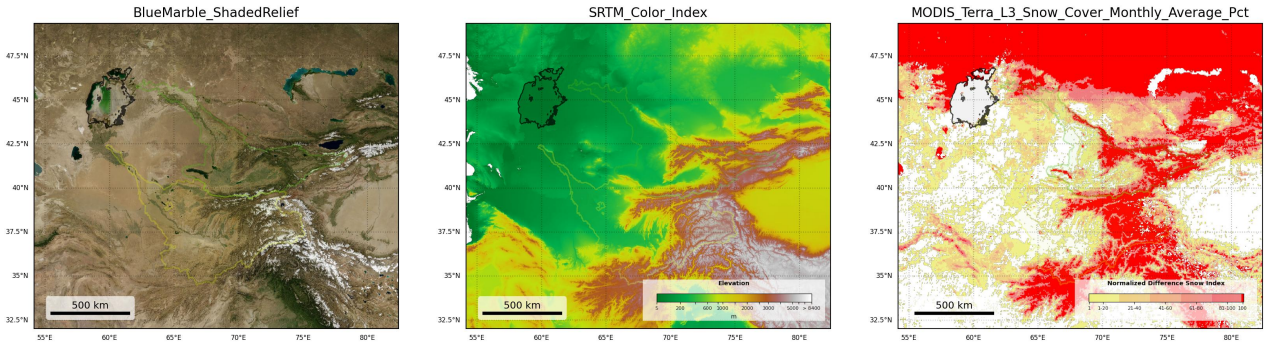
2.1. Location

The Aral Sea Basin is a large endorheic basin located in Central Asia, east of the Caspian Sea. The basin also contains the endorheic lake, The Aral Sea, once the worlds' 4th largest lake. Both are shown in figure 2.1a. The basin covers an area of approximately 1 760 000 km^2 (FAO, 2012; Micklin et al., 2014), making it one of the largest endorheic basins in the world. For comparison, the Rhine drainage basin covers approximately 160 000 km^2, the Danube drainage basin about 800 000 km^2, and the Yangtze drainage basin approximately 1 700 000 km^2 (GRDC: The global runoff data centre, n.d.).

The Aral Sea Basin includes large parts of several Central Asian countries: Kazakhstan, Uzbekistan, Turkmenistan, Tajikistan, and Kyrgyzstan (Frenken, 2013). Exact definitions of the boundaries of what consist of the Aral sea basin differ (FAO, 2012). This study uses the Aral Sea Basin boundaries from Lehner and Grill (2013) focusing on the catchments of the Amu Darya and Syr Darya rivers as these consist of essentially all surface water inflow into the Aral Sea itself (Aus Der Beek et al., 2011; Micklin et al., 2014; Touge et al., 2024).

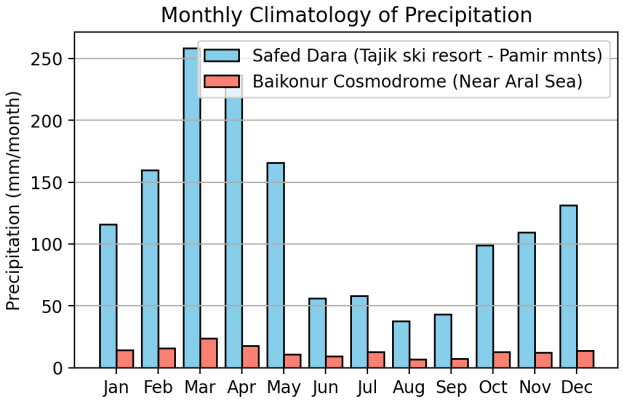
2.2. Climate and Topography

The spatial distribution of precipitation varies strongly across the Aral Sea Basin. Most precipitation occurs in the Tien Shan and Pamir Mountains in the southeast, where snowfall and glacial melt contribute significantly to river discharge (Micklin et al., 2014; Murodov et al., 2023). Figure 2.1c shows the distribution of snow and figure 2.1b shows the Digital Elevation Model (DEM). These mountain ranges reach elevations of up to

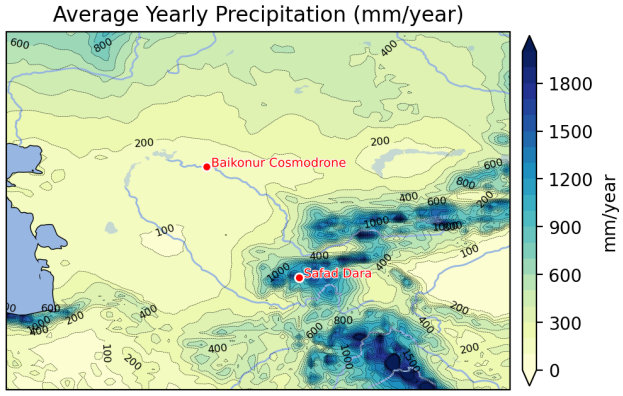


(**a**) NASA Blue Marble Composite (**b**) Elevation (SRTM) (**c**) Snow Cover (MODIS Terra)

Figure 2.1: Study area context: (a) NASA Blue Marble composite (2004), showing land cover, including arid lowlands and surrounding high mountain systems, and a residual water body in the eastern aral sea prior to further desiccation; (b) SRTM elevation, highlighting the contrast between the low-lying Aral Basin and the surrounding Central Asian mountain systems; and (c) MODIS Terra snow cover fraction (February 2025). Catchment boundaries are overlaid on NASA GIBS imagery for geographic context (NASA Earth Science Data and Information System (ESDIS), 2026).



(**a**) Contrast in average monthly precipitation between the mountainous southeastern and arid northwestern parts of the basin.



(**b**) Contour lines of average yearly precipitation.

Figure 2.2: Comparison of precipitation patterns across the Aral Sea basin for the period 1955–1970, based on ERA5 data (Hersbach et al., 2020). Locations are chosen to illustrate spatial variability in precipitation.

approximately 7000 m and are characterised by strong local elevation differences.

In contrast, the western and northern parts of the basin consist mostly of low-lying, semi-arid to arid terrain with low annual precipitation (figure 2.2). This strong spatial contrast is closely linked to the basin's topography and is reflected in the Köppen–Geiger climate classification, with climates ranging from hot and cold deserts in the lowlands to tundra climates in the high mountains (figure 2.3; (Beck et al., 2018)).

2.3. Hydrological context

The two principal rivers of the Aral Sea system are the Amu Darya and the Syr Darya, with catchment areas of approximately 450 000 km^2 and 310 000 km^2, respectively GRDC: The global runoff data centre, n.d. These river systems generate an estimated total discharge of approximately 115–120 km^3/year in the upstream mountain ranges of the Tien Shan and Pamirs (Boomer et al., 2000; Frenken, 2013; Micklin et al., 2014).

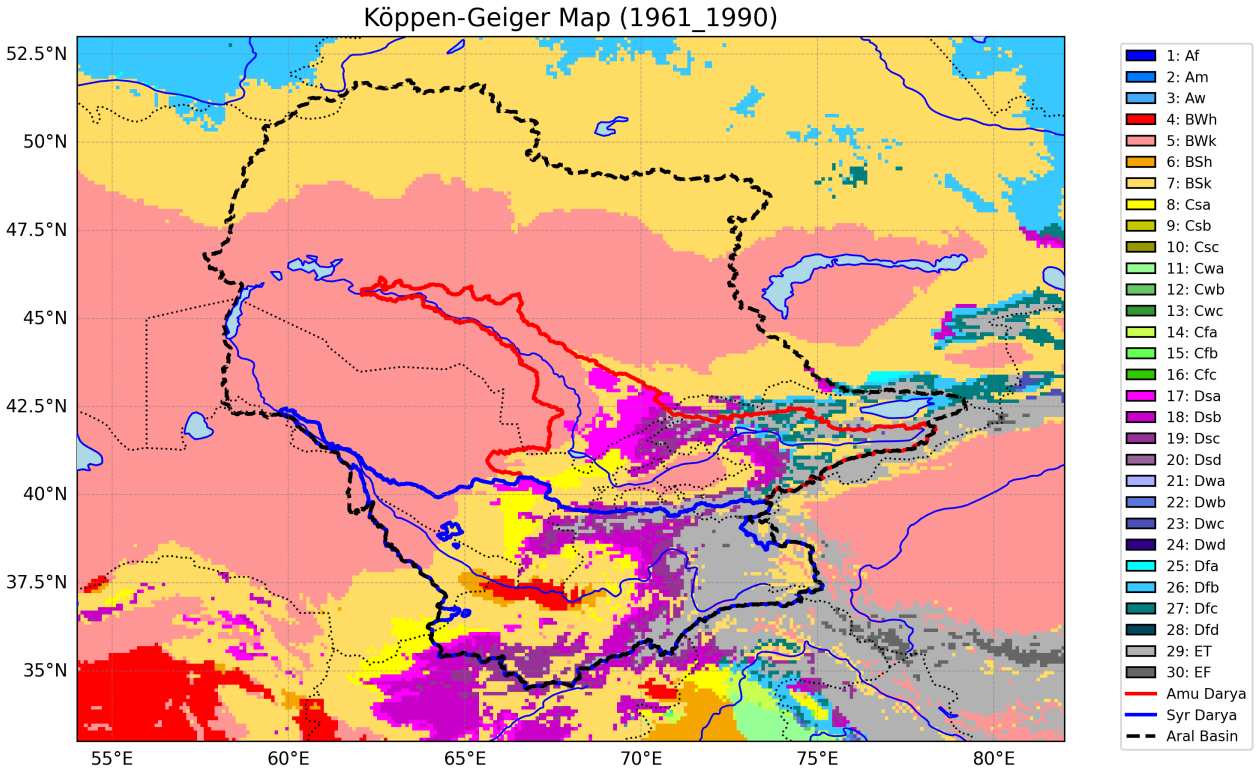
Historically, during the stable period prior to 1960, only about half of this generated volume, averaging approximately 56 km^3/year, actually reached the Aral Sea (Cretaux et al., 2013; Micklin et al., 2014). The remaining 60 km^3/year was lost to natural processes, including evaporation, desert filtration, and transpiration within the vast river deltas, or was diverted for early-scale irrigation (Cretaux et al., 2013; Micklin et al., 2014)

In addition to these main catchments, the Aral Sea Basin contains several smaller endorheic sub-basins (Figure 2.4b). The largest of these are the Shu (or Chu) and Sarysu river systems, which terminate in ephemeral lakes within the Ashchykol Depression and the Moiynkum Desert (Figure 2.4a). These system are hydrologically not connected and do not contribute to the inflow to the Aral Sea (Frenken, 2013).

2.4. Long-Term Basin Evolution

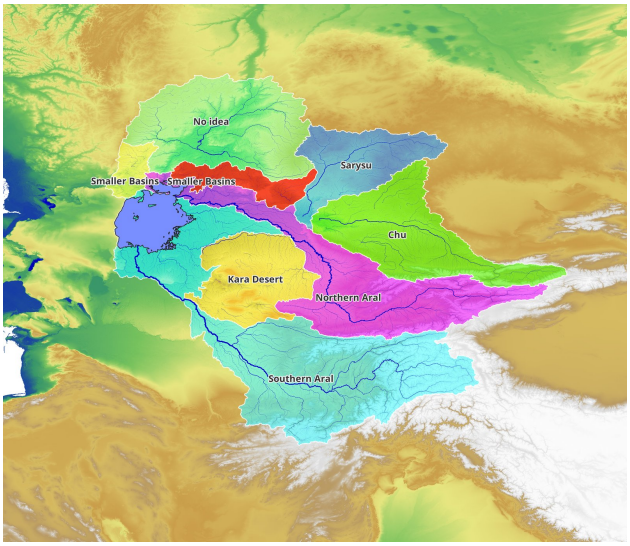
Millions of years before human settlement, the Aral Sea Basin was part of the larger Paratethys Sea, which extended from present-day Western Europe to Eastern Kazakhstan. Around 5 million years ago, tectonic uplift associated with the rising Alps and Carpathian Mountains isolated the basin from the ocean, initiating a long-term drying trend. The Black Sea, the Caspian Sea, and the Aral Sea are remnants of this ancient system (Boomer et al., 2000). Subsequent basin evolution was strongly controlled by changes in the courses of its major rivers, particularly the Amu Darya.

The Amu Darya, historically known as the Oxus (Zon et al., 2009), has shifted its course multipile times due to a combination of natural sedimentation, tectonic activity, and human intervention (Boomer et al., 2009; Brite, 2016; Micklin et al., 2014). Up to approximately 5000 years ago, the river followed a more southwesterly route, south of the Kyzylkum Plateau, discharging directly into the Caspian Sea (Boomer et al., 2000). Later, the river shifted northward, flowing through Lake Sarygamysh and the Uzboy channel before reaching the Caspian Sea (Létolle et al., 2007; Yapiyev et al., 2017). Subsequent course changes redirected flow toward the Aral Sea, Lake Sary Kamysh, or some combination of both (Boomer et al., 2000; Krivonogov, 2009; Krivonogov

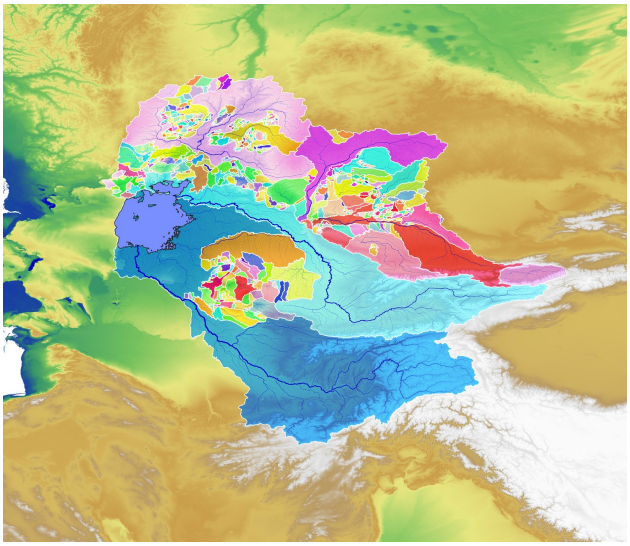


	Climate description	Aral basin (%)	Amu Darya (%)	Syr Darya (%)	Plotted Area (%)
BWk	Cold desert	47.9	11.9	30.5	39.0
BSk	Cold steppe	30.5	33.1	24.3	35.3
ET	Tundra	7.6	19.8	13.3	8.3
Dsb	Snow, dry summer, warm summer	3.7	9.7	8.1	2.9
Dsc	Snow, dry summer, cold summer	3.4	10.5	6.1	1.8
Dsa	Snow, dry summer, hot summer	2.2	3.4	9.2	1.3
Csa	Mediterranean, hot summer	1.8	5.9	3.3	1.3
Dfc	Snow, no dry season, cold summer	1.0	0.0	4.0	1.2
BWh	Hot desert	0.7	3.3	0.0	2.3
BSh	Hot steppe	0.5	2.3	0.0	0.8
Other	Other classes	0.5	0.1	1.2	5.9

Figure 2.3: Köppen–Geiger climate classification of the Aral Basin (dashed black outline) for 1961–1990, based on the high-resolution 1 km dataset by Beck et al. (Beck et al., 2023). The map highlights cold, humid mountain regions in the southeast and arid desert lowlands in the west and north of the basin. The most common climate classes in the basin are BWk (cold desert), BSk (cold steppe), ET (tundra), and Dsb (cold dry-summer continental). The accompanying table presents the 10 most prevalent climate classes in the Aral Basin, showing the relatively high fractions of ET, Dsb, Dsc, and Dsa in the Amu Darya (blue outline) and Syr Darya (red outline) catchments compared to the basin overall.



(**a**) Overview of the various drainage basins based on level 5 hydrobasins (Lehner and Grill, 2013)



(**b**) Overview of the various drainage basins based on level 12 hydrobasins (Lehner and Grill, 2013)

Figure 2.4: Overview of the Aal Sea drainage basins. Of the area considered to be part of the ASB only the area labeled north and south aral drain into the aral sea. The other areas are all smaller endhorheic basins or the flow does not reach the rivers feeding into the Aral Sea, notably the Chu and Sarysu rivers. Based on HydroBasins data (Lehner and Grill, 2013) and DEM data from the HydroSheds project (Lehner et al., 2008)

et al., 2014).

Human activity has interacted with these natural systems for millennia. There has been archeological evidence for irrigation systems dating back to 2000 BCE (Oberhänsli et al., 2007). Some of these irrigation systems supported large agricultural societies (Brite, 2016). These early systems already modified local water distributions and in part affected both river discharge and Aral sea levels.

Shifts in river flow sometimes coincided with destruction or abandonment of irrigation infrastructure, such as during the Arab Conquests and Mongol invasions, though some research suggests that climate variability, rather than military events, may have driven these changes (Krivonogov et al., 2014; Toonen et al., 2020).

Palaeoclimate reconstructions indicate that the Aral Sea Basin experienced significant climatic variability prior to modern times. Periods of increased precipitation occurred between AD 400–900 and AD 1200–1500, when annual precipitation increased from $P < 250$ mm to $P \approx 250$–500 mm (Baker et al., 2015; Sorrel et al., 2007). Such variations influenced river discharge, lake levels, and the extent of irrigated agriculture.

2.5. Modern transformation

Since the 1960, the Aral Sea region has undergone profound anthropogenic change. Agricultural intensification initiated under the Soviet Union and sustained after significantly increased water demand. The total area under cultivation expanded from approx. 47.000 km^2 in 1960 to approx 83.000 km^2 in 2000 (Aus Der Beek et al., 2011), as shown in figure 2.6.

An essential part of this expansion of agricultural land was the development of the Karakum canal. Stretching over 1400 km this canal enabled an increase of 400% of agricultural land in Turkmenistan, withdrawing an estimated 12-18 km³ per year (Aus Der Beek et al., 2011; Frenken, 2013; Micklin et al., 2014; Murodov et al., 2023; Touge et al., 2024).

These massive projects and large-scale diversions of the Amu Darya and Syr Darya rivers lead to irrevocable water losses in the basin (Stulina & Eshchanov, 2013). The subsequent decline in river inflow discharges caused dramatic reductions in the Aral Sea's area and volume (Micklin et al., 2014).

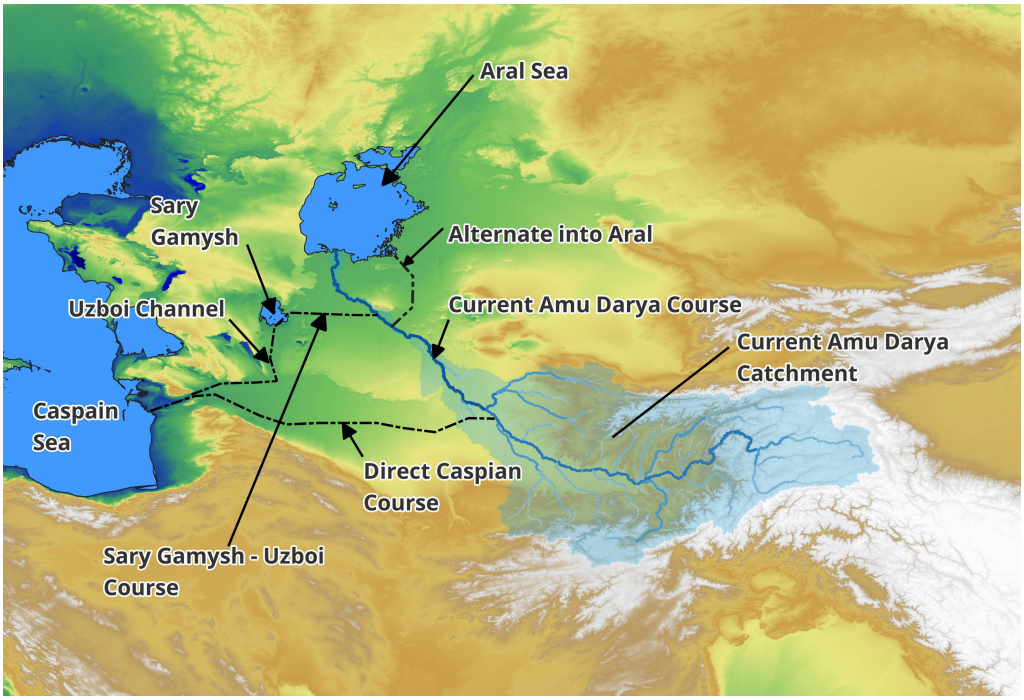
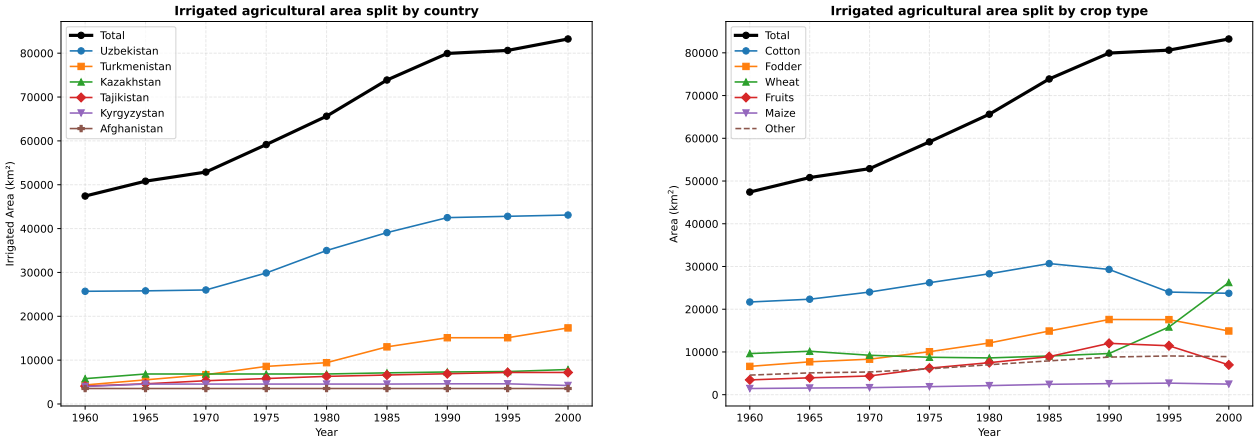


Figure 2.5: Schematic overview of approximate historical courses of the Amu Darya River. Based on DEM from Lehner et al. (2008) and watersheds from Lehner and Grill (2013). Dotted lines an interpretation of some of the historical river courses.



(**a**) irrigated agriculture split by country

(**b**) irrigated agriculture split by crop type.

Figure 2.6: Increase in irrigated agricultural land use, split by country and by crop type. Based on data from Aus Der Beek et al. (2011). Of note is the initial increase in cotton production, while later wheat becomes dominant. These crops have different water demands, explaining in part the decrease in water use.

2.6. Historical Variations in Aral Sea Water Levels

if one would one look at the water levels of the Aral sea since the year 1800 or so, table 2.1, one would get the impression that the water level of the Aral Sea has always been stable at 53-54 meter above sea level (m asl), but this period is not representative of the long term history of the Aral Sea. Throughout the Holocene, the period of the last 12.000 years, the Aral Sea has undergone dramatic transgressions and regressions, periods of high water levels as well as nearly desiccated states (Cretaux et al., 2013; Micklin et al., 2014).

Table 2.1: Historical water levels and key drivers of the Aral Sea. Data primarely based on (Cretaux et al., 2013; Krivonogov et al., 2014; Oberhänsli et al., 2007) and the other papers cited in this section

Period	Level (m a.s.l.)	Key drivers and notes
3000–800 BC	54–58	Amu Darya flows to Aral; first agriculture; canals up to 150 km long.
800 BC–400 AD	10–27	River diverted via Sarykamysh/Uzboy; $\approx 35,000$ km^{2}under irrigation under irrigation; comparable to modern levels.
400–900 AD	52–54	Sea recovers; Amu Darya returns to Aral; irrigation infrastructure destroyed.
900–1200 AD	≈53	Stable; introduction of the *chigir* waterwheel.
1221–1570	29–31	Collapse following Mongol/Timurid destruction of canal infrastructure; river routed via Sarykamysh/Uzboy.
1573–1960	50–53	Amu Darya shifts back to Aral Sea; level stable for nearly 400 years.
1960	53.4	Soviet irrigation expansion begins; Karakum Canal construction starts.
1970	51.4	Continued diversion.
1980	45.8	Accelerating decline.
1989	40	North–South basin split.
2000	33	—
2005	—	Kok-Aral Dam completed; North Aral stabilised at ≈42 m.
2008	29.2	East–West lake split in southern basin.
2011	28	—
2026	23	Southern basins continue to shrink.

A striking example of this variability is the discovery of the Kerderi Mausolea from the 14th century in 2000, located at 29 m asl, some 20 meters below the lake level of 1960, indicating periods when lake levels were substantially lower than in the 19th and 20th century (Krivonogov, 2009; Micklin, 2016; Oberhänsli et al., 2007).

Earlier literature cited a 72–73 m asl as maximum highstand, recent research show that a Holocene level above 64–65 m asl was impossible as it would have triggered a massive overflow into the Caspian sea. The revised maximum lake level is now estimated to be only 54–55 m asl (Boomer et al., 2009; Boomer et al., 2000; Cretaux et al., 2013).

Following a final return of the Amu Darya to the Aral Basin around the end of the 16th century, the sea entered a relatively stable phase. From the late 18th century until 1960, the water level remained relatively stable before entering its current phase of rapid decline (Micklin et al., 2014).

3

Data & models

This chapter describes the datasets and hydrological models used in this study. It contains river discharge observations, Aral Sea observations, land and water use information and forcing data. The hydrological model is also described in detail in this chapter. These form the input basis for the modeling procedure described in Chapter 4.

3.1. eWaterCycle

This research is conducted using the eWaterCycle platform (Hut et al., 2022). eWaterCycle is a framework for hydrological modelling that aims to improve the FAIRness (Findable, Accessible, Interoperable, Reusable) of computational hydrology workflows (Hut, 2022). Its development responds to the observation by C. Hutton et al. (2016) that much computational hydrology research is not reproducible, with some studies citing only 1.6% of of studies can be fully reproduced (Hall et al., 2022; Stagge et al., 2019).

The eWaterCycle framework provides a standardized environment for running hydrological models and conducting experiments. It is designed to improve reproducibility, enhance model reusability, and increase the transparency of computational workflows.

Separation of model and experiment

A key principle of eWaterCycle is the strict separation between model code and experimental setup. This allows hydrological models to be reused across different studies without modifying the model implementation itself. Instead, experimental configurations define forcing data, parameters, and simulation settings. This separation improves reproducibility and facilitates collaboration between researchers.

Software containers

To ensure computational reproducibility, models in eWaterCycle are executed inside Docker containers. These containers encapsulate the full software environment, including libraries, compilers, and dependencies. As a result, simulations are independent of the local computing environment, preventing inconsistencies caused by software version changes or system-specific configurations.

Technical interoperability and runtime control

Technical interoperability within eWaterCycle is enabled by the Basic Model Interface (BMI) 2.0 (E. Hutton et al., 2020). This interface defines a standardized set of functions that allows hydrological models to be controlled from a Python environment, including initialization, time stepping, and data access, regardless of the programming language in which the model is implemented.

Communication between the Python environment and the model container is handled through the grpc4bmi library (van den Oord et al., 2023). This enables remote execution of hydrological models implementing the Basic Model Interface (BMI) and allows runtime interaction with the model state, including querying and modifying variables during simulation.

Supporting tools and data workflows

eWaterCycle integrates additional tools to support reproducible hydrological workflows. Experiment code is typically made available and executed in isolated environments, often through the use Jupyter notebooks, which improves transparency and interpretability of experiments.

To ensure consistency in model inputs, standardized preprocessing of forcing datasets is used. Tools such as ESMValTool (Eyring, Righi, et al., 2016; Righi et al., 2020) are be applied for processing climate reanalysis and model data, including CMIP datasets and ERA5 reanalysis data (see Section 3.4.2 and Section 3.4.1).

3.2. Hydrological Models

As described in section 1.4, the workflow consists of a chain in which generated meteorological forcing data is applied to a hydrological model. The resulting model output is subsequently used in an Aral Sea volume model to compute changes in water levels. The Aral Sea model is made specifically for this work and described in Section 4.4.

The hydrological models used in this study are PCR-GLOBWB2 and HBV. PCR-GLOBWB2 is applied as the primary distributed hydrological model, while HBV is used for testing and benchmarking purposes.

3.2.1. PCR-GLOBWB2 - Distributed Model

PCRaster GLOBal Water Balance model 2 (PCR-GLOBWB2) is a global hydrological model developed at Utrecht University (E. H. Sutanudjaja et al., 2018), and is based on the earlier PCR-GLOBWB model (Van Beek et al., 2011). It simulates the terrestrial water cycle and human water use at a daily time step. The model has been widely applied in hydrological and water resources research (Guan et al., 2026; Hoch et al., 2023; Ruijsch et al., 2021; Wu et al., 2024; Yang et al., 2024).

PCR-GLOBWB2 is a distributed model with a spatial resolution of 5 arcminutes (approximately 10×10 km at the equator). It represents the hydrological system through a set of interacting modules, including:

- meteorological forcing,
- land surface processes,
- groundwater dynamics,
- surface water routing,
- irrigation and water use.

For each grid cell, the model simulates the vertical water balance. A conceptual overview of this process is shown in Figure 3.1. Key processes include precipitation partitioning, infiltration, evapotranspiration, groundwater recharge, and runoff generation.

Irrigation and water use module

A particularly relevant component for this study is the irrigation and water use module. This module dynamically estimates sector-specific water demand (e.g., agriculture, industry, and domestic use). Water abstractions from surface water and groundwater directly affect river discharge and groundwater storage. In addition, return flows from water use are represented and reintroduced into the hydrological system.

Model limitations

Despite its global applicability, PCR-GLOBWB2 has several limitations. All processes are represented at the grid-cell scale, and as such does not include large scale water diversions for large scale agriculture, aqueducts or inter-basin water transfer (Wada et al., 2016). In addition, the spatial averaging can smoothen steep topographic and temperature gradients which can reduce the accuracy of simulations in areas with strong spatial gradients, such as mountainous regions. Another limitation is that glaciers and snow (re)distribution processes are not modelled (Wiersma et al., 2022).

A full technical description of PCR-GLOBWB2 is provided in E. H. Sutanudjaja et al. (2018).

3.2.2. HBV - Conceptual Model

HBV (Hydrologiska Byråns Vattenbalansavdelning) is a conceptual rainfall-runoff model originally developed for Scandinavian catchments (Bergström, 1976). Numerous variants of the model exist and have been

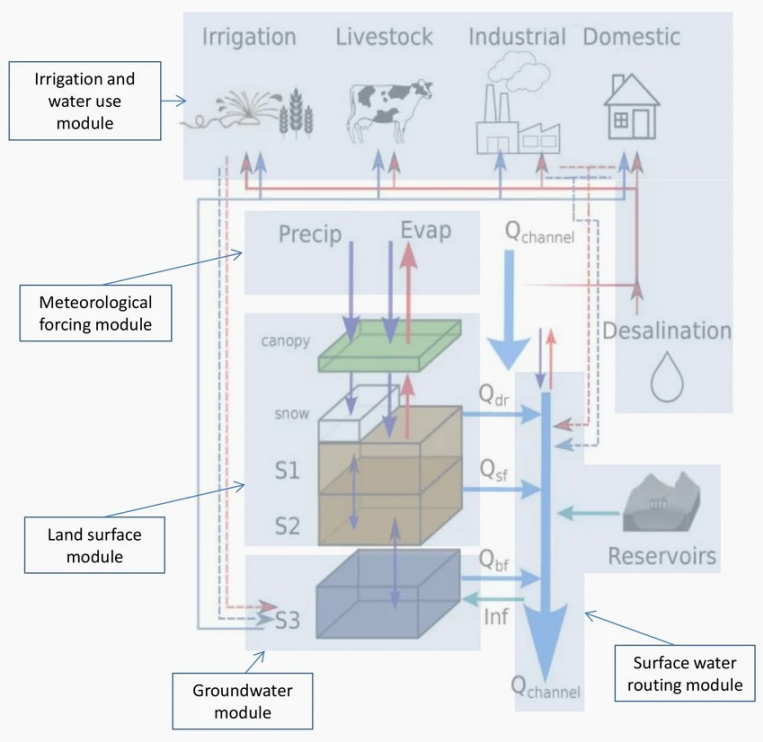


Figure 3.1: Schematic overview of PCR-GLOBWB2 showing its different model modules. Hydrological processes are simulated per grid cell, in contrast to HBV, which represents the entire catchment as a lumped system. Reproduced from E. H. Sutanudjaja et al. (2018).

widely applied in hydrological research (Lindström et al., 1997; Seibert & Bergström, 2022). The implementation used in this study is based on earlier work in Haasnoot (2024) within the eWaterCycle project.

HBV is a model that represents the hydrological system using a set of conceptual reservoirs ("buckets") connected through parameterized fluxes. Rather than explicitly resolving spatial processes, the model captures dominant hydrological dynamics through simplified process representations.

A schematic overview of the model structure is shown in Figure 3.2.

The exchange of water between these reservoirs is governed by a limited number of model parameters. Although strongly simplified compared to distributed models, HBV is capable of reproducing key hydrograph dynamics with relatively low computational cost.

Role as benchmark model

In this study, HBV is primarily used as a computational benchmark and workflow testing tool. Compared to the distributed PCR-GLOBWB2 model, HBV simulations are computationally inexpensive, enabling rapid testing of the modelling workflow, debugging of model coupling, and evaluation of calibration strategies. In addition, the model is used to benchmark calibration performance and optimization speed, for example when applying parameter optimization algorithms such as CMA-ES.

3.3. River Discharge Observations

As an endorheic lake system, the water balance and level of the Aral Sea are fundamentally determined by the river inflow from its two main tributaries, the Amu Darya and Syr Darya. While many studies cited in earlier chapters provide annual estimates or long-term averages of river discharge into the Aral Sea, they generally lack the spatial or temporal resolution required for hydrological modeling. Historical discharge observations are therefore critical for the calibration and validation of the hydrological model-chain. To achieve this, this study relies on daily and monthly streamflow observations from the GRDC for model calibration, while supplementing the analysis with the long-term annual records from the INTAS report to provide historical context.

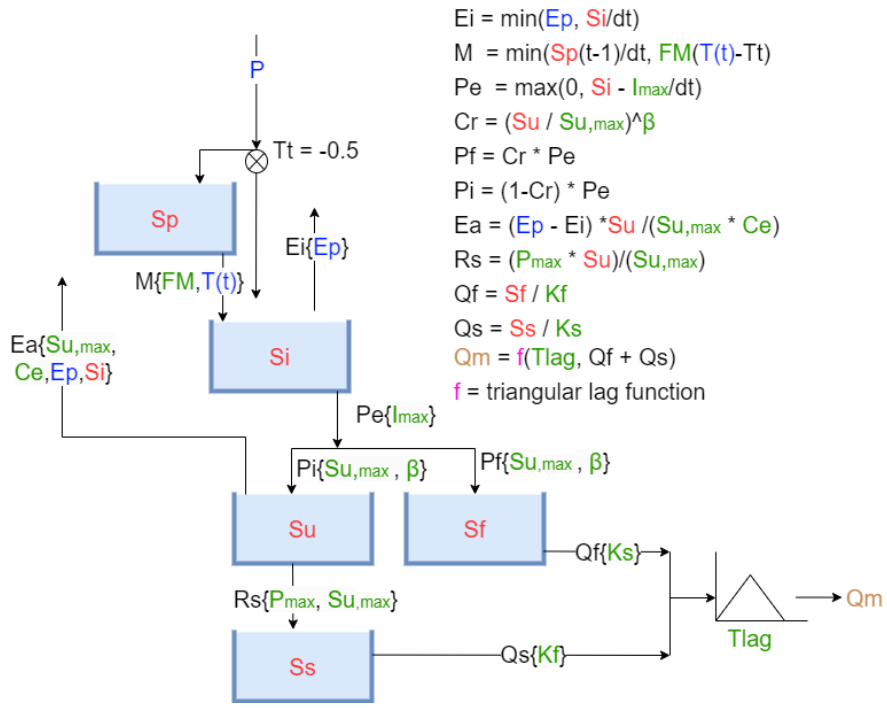


Figure 3.2: Schematic overview of the HBV hydrological model and its governing equations. Forcing variables are shown in blue, state variables in red, and parameters in green. (Haasnoot, 2024)

GRDC - observations

The primary source for historical streamflow is the Global Runoff Data Centre (GRDC), which provides access to hydrological station streamflow data across the world (GRDC: The global runoff data centre, n.d.). For the sub-basins in this study, a total of 85 stations are available (see Figure 3.3). These stations vary in temporal resolution, with some providing monthly data and others daily observations. The record lengths also vary, and several time series contain significant gaps. A complete overview of the available GRDC data is provided in tables A.1 and A.2 in Appendix A.

The main gauging stations on the Amu Darya are Chatly and Kerki. For the Syr Darya, the principal stations are Kazalinsk, Karaozek, and Tyumen-Aryk. The remaining stations primarily cover tributaries within the basin.

As an illustrative example, discharge data from the Kerki (Turkmenistan) and Chatly (Uzbekistan) stations are shown in figure 3.4. Differences in record length and data gaps are clearly visible. An additional feature is that minimum flows at the downstream station (Chatly) are lower than at the upstream station (Kerki). This may reflect upstream water abstractions for irrigation and other human activities, which reduce downstream discharge, particularly during low-flow periods.

INTAS Technical Report

The INTAS-2006 project report (Nachtnebel, 2006) provides a comprehensive historical dataset of yearly discharge volumes into the Aral Sea deltas for the period 1950–2005. This report is essential for reconstructing the "quasi-stationary" period of the water balance (pre-1960) when river inflow and evaporation were nearly in equilibrium. It also serves as a validation for the modelled yearly discharge volumes.

3.4. Meteorological Forcing Data

3.4.1. ERA5

eWaterCycle offers direct access to ECMWF Reanalysis 5th Generation (ERA5) as forcing data. ERA5 is the most recent global reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) (C3S, 2018; Hersbach et al., 2020), assimilating historical observation data into physics based climate models. It provides hourly estimates of a wide range of meteorological variables from 1940 to the

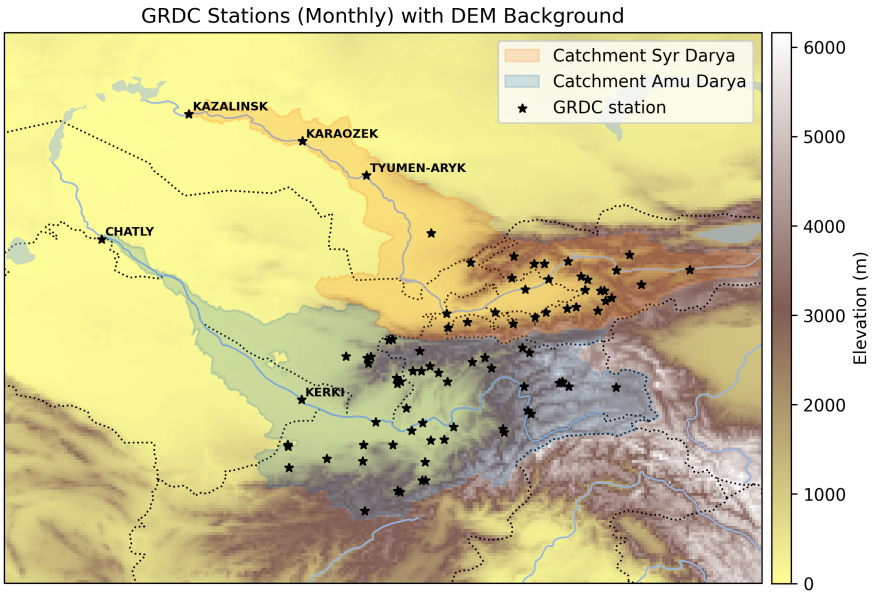


Figure 3.3: Overview of all GRDC stations in the Aral Sea basin, overlaid on a DEM derived from E. H. Sutanudjaja et al., 2018. Selected stations used extensively in this study are highlighted by name.

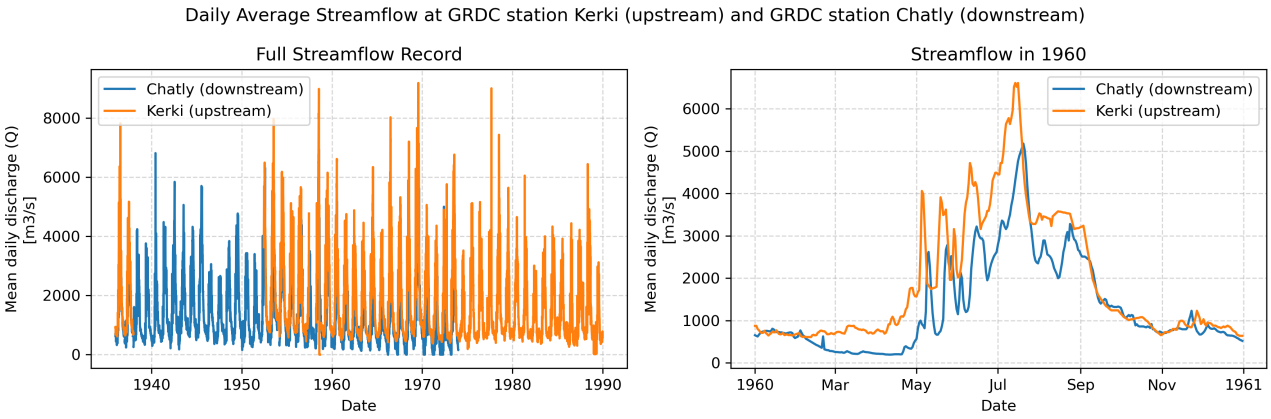


Figure 3.4: Daily discharge at the Kerki (upstream) and Chatly (downstream) GRDC stations (GRDC: The global runoff data centre, n.d.). The left panel shows the full temporal coverage of the observations, while the right panel zooms in on 1960, illustrating the reduction in discharge between the upstream and downstream stations.

present. Through the tools provided by eWaterCycle, ERA5 data can be easily preprocessed and used as forcing input for the hydrological models included in this study.

3.4.2. CMIP

eWaterCycle also provides access to climate projections from the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring, Bony, et al., 2016)). CMIP6 uses Shared Socioeconomic Pathway (SSP) scenarios, which are described as a number: SSPx-y.y where x describes the socioeconomic scenario and y.y the extra radiative forcing in W/m^2 reached by the year 2100 (O'Neill et al., 2016; Riahi et al., 2017; Tebaldi et al., 2021)

Within this study, CMIP6 outputs are used to represent future climate conditions under various scenarios in the Aral Sea basin. The eWaterCycle framework provides tools to make them easily usable as forcing input in the hydrological modelling chain. For consistency only the default ensemble member *r1i1p1f1* is used. Data from the following models was used during this study:

1. Max Planck Institute Earth System Model version 1.2, High Resolution (MPI-ESM1-2-HR)
2. Model for Interdisciplinary Research on Climate version 6 (MIROC6)
3. Canadian Earth System Model version 5 (CanESM5)
4. Meteorological Research Institute Earth System Model version 2.0 (MRI-ESM2-0)

While details of the Coupled Model Intercomparison Project Phase 7 (CMIP7) scenario framework have been released (Van Vuuren et al., 2026), these do not include any meteorological forcing data; therefore only meteorological forcing derived from CMIP6 models is used in this study.

3.5. Aral Sea observations

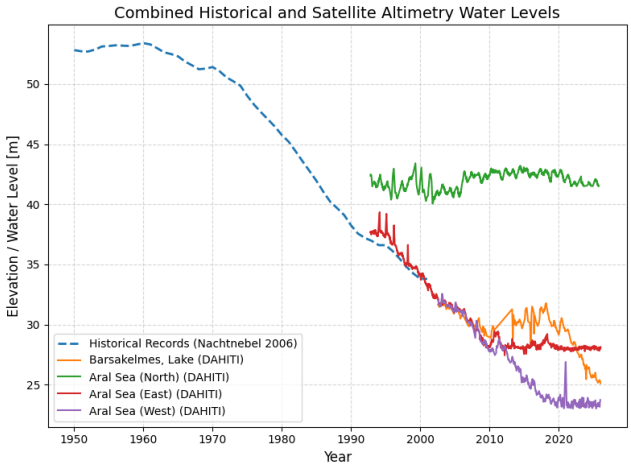


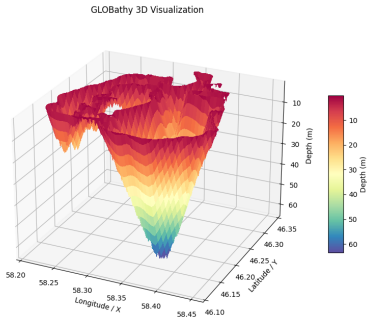
Figure 3.5: Historical and satellite-derived water levels of the Aral Sea. The historical record (1950–2001) is from Nachtnebel (2006) and is shown as a dashed blue line. Satellite altimetry data (1992–2025) from Schwatke et al. (2015) are shown as solid colored lines, with each line corresponding to a specific sub-basin or location indicated by its label. This combined time series illustrates the long-term trends in water level and the impact of human and environmental changes on the Aral Sea.

To accurately model the volumetric changes of the shrinking lake system, reliable geometric relationships are required. The INTAS-2006 technical report (Nachtnebel, 2006) mapped the historical bathymetry of the contiguous Aral Sea seabed (figure 3.6b), providing detailed area-elevation-volume curves over the baseline period of 1950–2005.

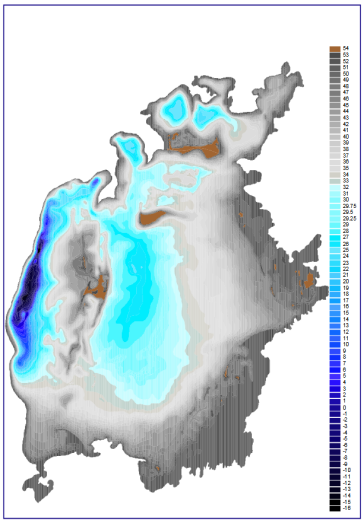
To validate the accuracy of these INTAS curves, they were cross-examined and validated against independent records. Specifically, the area-volume-elevation relationships for the Large (Southern) Aral Sea published in Table 2 of Crétaux et al. (2005). This cross-comparison confirmed a strong agreement between the two datasets, demonstrating that the INTAS curves are robust and suitable for use in the volume balance model.

To verify the continuous performance of the resulting lake water balance model over time, modern high-frequency satellite altimetry data from the DAHITI platform (Schwatke et al., 2015) was also used.

While global bathymetry products like GLOBathy exist (Khazaei et al., 2022), they are intended for data-scarce regions and performed poorly here, showing significant geometric discrepancies (figure 3.5). Given the availability of local cross-validated data, this global approximation was excluded.



(**a**) bathymetry of the southern Aral Sea, based on data from the GLOBathy project (Khazaei et al., 2022).



(**b**) Bathymetry from Nachtnebel (2006). Dark blue indicates deeper areas, grey indicates shallow or dry regions.

Figure 3.6: Comparison of bathymetry datasets for the Aral Sea: global reconstruction (left) and detailed historical survey (right).

3.6. Agriculture: Land and Water Use

A substantial amount of agricultural information is available at regional or national scales within the Aral Sea basin. For example, annual cotton production and irrigated area statistics are reported per country in FAO assessments (Frenken, 2013). Such datasets provide valuable long-term information on agricultural development and water use. Another example is the data provided by Aus Der Beek et al. (2011) which documents the evolution of agriculture split by crop type and country (figure 2.6).

However, high-resolution spatial datasets describing historical agricultural dynamics are largely unavailable. In particular, spatially explicit information on monthly crop water requirements is scarce for the period prior to the satellite era. Many available land-use products only cover recent years and therefore do not capture the rapid agricultural expansion that began in the 1960s. For instance, the ESA WorldCover dataset provides detailed global land-use information but represents only conditions around the year 2020 (Zanaga et al., 2022).

PCR-GLOBWB

Within PCR-GLOBWB2, agricultural land and water use are represented using a combination of global land-use and crop datasets. In the model, pasture and rain-fed crops are considered short natural vegetation and do not interact with the water demand module described in section 3.2. Irrigated agriculture is split between irrigated paddy (rice) and irrigated non-paddy. Considering rice production is not significant (figure 2.6b), the irrigated non-paddy effectively represents all irrigated agriculture in this study. The simulation of crop water consumption is done following the standardized FAO guidelines of Allen et al. (1998).

In the standard model configuration (E. Sutanudjaja et al., 2017) the irrigated water demand is calculated on irrigated area per cell and per crop type. Data for these is from MIRCA200 (Portmann et al., 2010) and the Global Crop Water Model (Siebert & Döll, 2010). For the non-paddy irrigation, water is applied to refill the soil water content up to field capacity whenever it gets below a pre set value. The net irrigation water demand is multiplied with a country specific water loss factor from Rohwer et al. (2007).

The water demand determined in the water use model can come from three sources: groundwater (both fossil and non-fossil), desalinated water and from surface water. Of note for this research is that water distribution networks are not fully modeled. Water availability is therefore pooled over zones of $1° \times 1°$ for surface water and desalinated water, and $0.5° \times 0.5°$ for groundwater, and these zones can be used to meet local crop-water

demand (and other demand) if the local supply is not sufficient. An overview of dimensions is given in Table 3.1. For a full explanation see the model documentation (E. H. Sutanudjaja et al., 2018).

	$(0.083° \times 0.083°)$ PCR-GLOBWB2 grid		$(0.5° \times 0.5°)$ Groundwater		$(1° \times 1°)$ Surface & desalinated Water	
Reference lat.	E–W (km)	N–S (km)	E–W (km)	N–S (km)	E–W (km)	N–S (km)
35°	7.59	9.27	45.54	55.60	91.09	111.19
40°	7.10	9.27	42.59	55.60	85.18	111.19
45°	6.55	9.27	39.31	55.60	78.63	111.19
50°	5.96	9.27	35.74	55.60	71.47	111.19
55°	5.31	9.27	31.89	55.60	63.78	111.19

Table 3.1: Cell dimensions for different resolutions used in water allocation in PCR-GLOBWB2 calculated for different reference latitudes. Surface and desalinated water zones are $1° \times 1°$ while groundwater zones are $0.5° \times 0.5°$. The native spatial resolution in PCR-GLOBWB2 is 5 arc-minutes ($\approx 0.083°$). Cell dimensions are given as east–west (E–W) and north–south (N–S) extents in km.

Karakum Canal

Construction of the Karakum Canal began in the 1950s. Reported withdrawal volumes increased from approximately 4 km^3/year in the 1960s to about 18 km^3/year around 1990, while recent reconstructions estimate current withdrawals at approximately 12 km^3/year (Aus Der Beek et al., 2011; Frenken, 2013; Touge et al., 2024). The canal remains one of the largest anthropogenic diversions from the Amu Darya and is considered a major contributor to reduced inflow to the Aral Sea (Micklin et al., 2014).

4

Methodology

4.1. Software and Workflow

For this research, various software functions and tools within the eWaterCycle environment were developed. Some of these tools aim to reduce the amount of manual adjustments required in notebooks, while others introduce new functionalities. Some of the main points are:

1. Update PCR-GLOBWB in a docker container image to the latest available version
2. Simplification and standardization of forcing generation
3. Streamlined model execution
4. parameter calibration workflow
5. Various analysis scripts
6. Improved reporting and plotting options

All experiments can be run in Jupyter notebooks, but most workflows were converted to Python scripts (`.py`) to reduce manual intervention and improve reproducibility. The full code is publicly available (van der Veen, 2026).

Reproducability

To ensure the process can be fully replicated, the study follows a structured configuration and documentation approach based on practices recommended in De Visser et al. (2023), Hall et al. (2022), and Hut and Hall (2025):

- The workflow is automated using Snakemake (Mölder et al., 2025), which ensures a reproducable execation of all experiment steps.
- all experiment parameters and settings are stored in a central configuration `.yaml` file.
- PCR-GLOB model is containerized
- datasets are accessed via standardized ewaterycle protocols.
- `.ipynb` (Jupyter Notebooks) are provided as interactive examples to demonstrate the core logic and workflow steps
- README.md files where necessary.

A further overview of the adherence to recommended FAIR practices can be found in appendix F

4.2. Forcing Generation & Pre-processing

Two separate forcing generation workflows were applied in this study using the eWaterCycle infrastructure, which enables consistent generation and handling of both gridded and basin-scale forcing data.

For the PCR-GLOBWB2 simulations, ERA5 and CMIP6 datasets were used as gridded fields. CMIP6 datasets were regridded and downscaled to a common latitude–longitude grid with ERA5 and subsequently bias-corrected at the spatial grid level prior to hydrological simulation, as described in Sections 4.2.1 and 4.2.2.

For the Aral Sea volume model, one-dimensional meteorological time series representing local conditions over the Aral Sea region were generated to use as forcing. These were generated using the same eWaterCycle infrastructure. Bias correction was applied using the QDM methodology described in Section 4.2.2. Because these forcing data are strictly one-dimensional, no spatial regridding or interpolation was required for this component of the workflow.

4.2.1. Harmonization of PCR-GLOBWB2 meteorological inputs

To enable a consistent comparison between simulations forced by reanalysis data and climate model projections, all forcing datasets were harmonized prior to model application.

The ERA5 dataset described in Section 3.4.1 was directly compatible with the PCR-GLOBWB forcing requirements within the eWaterCycle framework (Hut et al., 2022). In contrast, the CMIP datasets (Section 3.4.2) differ in spatial resolution and grid definition and calendar. Therefore, all CMIP forcing data were transformed to the ERA5 spatial grid. By enforcing a common latitude–longitude grid, hydrological differences between simulations can be attributed to differences in climate signals instead of numerical artefacts related to grid resolution.

For use in PCR-GLOBWB2 the variables precipitation and near-surface temperature were spatially interpolated to the ERA5 grid using xESMF regridding framework (Zhuang et al., 2025). Temporal chunking (1-year blocks) was applied during this step to enable efficient processing of the long time series, while spatial interpolation was performed using a bilinear scheme. This harmonization ensures consistent representation of basin-scale precipitation patterns, temperature gradients, and derived hydrological processes such as evapotranspiration and runoff generation.

Additionally the calendar systems of CMIP models were harmonized to account for leap year inconsistencies. For models using 365-day calendars the value for 28 February was copied to 29 February. This ensured temporal consistency between datasets and was required to meet the input requirements of PCR-GLOBWB2 The forcing preparation workflow was applied consistently to all CMIP simulations to guarantee comparability across scenarios.

4.2.2. Bias correction of climate forcing

To account for systematic biases in the CMIP6 climate simulations relative to ERA5, a bias correction was applied. Following Maraun and Widmann (2018), bias is defined as the systematic difference between modeled and reference variables. This bias is considered to be constant over the evaluated distribution. Applying bias correction improves the consistency between CMIP6 model output and ERA5 reanalysis data, making the corrected model data more comparable to the reference data and suitable as input to the hydrological model for subsequent simulations and analyses. A more detailed explanation of the use of ERA5 and CMIP6 can be found in Section 3.4.

Several methods exist to correct for such biases. To limit the scope of this research a bias correction method from the bias correction python package from Schwertfeger (2025) was used.

Quantile Delta Mapping (QDM)

For this project, Quantile Delta Mapping (QDM) was applied. The method was originally developed by Cannon et al. (2015). In this study, the practical implementation was carried out using the Schwertfeger (2025) python library, which builds on the QDM formulation of Tong et al. (2021).

Instead of applying a single correction to the entire dataset, QDM divides the distribution into segments (quantiles) and treats each segment separately. This means that each segment (from light drizzles to extreme precipitation events) receives it's own correction procedure and that the whole distribution is taken into account.

Within the QDM framework, the following steps are performed in quantile space:

1. *Quantile transformation*. Modelled values are expressed in terms of their corresponding quantiles in the historical model distribution.

2. *Bias correction via quantile mapping*. These quantiles are mapped onto the reference (ERA5) distribution using the inverse cumulative distribution function, thereby correcting systematic biases across the full distribution.
3. *Preservation of the change signal*. The relative change between historical and future model quantiles is retained and transferred to the bias-corrected distribution, ensuring that the projected climate signal is preserved in quantile space.

For the purposes of this thesis, ERA5 reanalysis data are treated as observational reference data, representing a pseudo-reality used for bias correction.

The datasets are split cleanly into two consecutive, non-overlapping eras: a historical period (1950–2014) and a future projection period (2015–2100). For the historical period, QDM effectively reduces to standard quantile mapping, as no projected change signal is applied. In this step, systematic biases between CMIP6 simulations and ERA5 are corrected across the full distribution. For the future period, the bias correction derived from the historical period is applied to the corresponding CMIP6 projections. This ensures consistency with the characteristics of each Global Climate Model (GCM) while preserving inter-model differences. For example, the historical period of the `mpi-esm1-2-hr` model is used to derive bias corrections, which are then applied to its future projections, and analogously for all other GCMs.

Due to memory constraints associated with the large spatial and temporal extent of the PCR-GLOBWB2 compatible gridded CMIP6 and ERA5 datasets, spatial chunking was applied during processing. The time dimension was kept intact to ensure correct computation of quantile-based bias correction, while all datasets (reference, historical, and projections) were chunked consistently along the spatial dimensions (latitude and longitude). The one-dimensional forcing data used for the Aral Sea volume model were not chunked.

Table 4.1: Overview of preprocessing and bias-correction methods applied to meteorological forcing variables.

Variable	Symbol	Application	Processing steps	Bias correction
Temperature	tas	PCR-GLOBWB	Regridding, downscaling, bias correction	Additive
Precipitation	pr	PCR-GLOBWB	Regridding, downscaling, bias correction	Multiplicative
Precipitation	pr	Aral Sea model	Bias correction only	Multiplicative
Potential evaporation	evspsblpot	Aral Sea model	Bias correction only	Multiplicative

4.3. Catchment modelling with PCR-GLOBWB2

4.3.1. PCR-GLOBWB updates

This section describes the updates to the PCRaster GLOBal Water Balance model 2 (PCR-GLOBWB2) model that were implemented and used in this thesis. These updates were required to enable model calibration and support model state manipulation to enable simulating large-scale irrigation interventions.

Calibration parameters

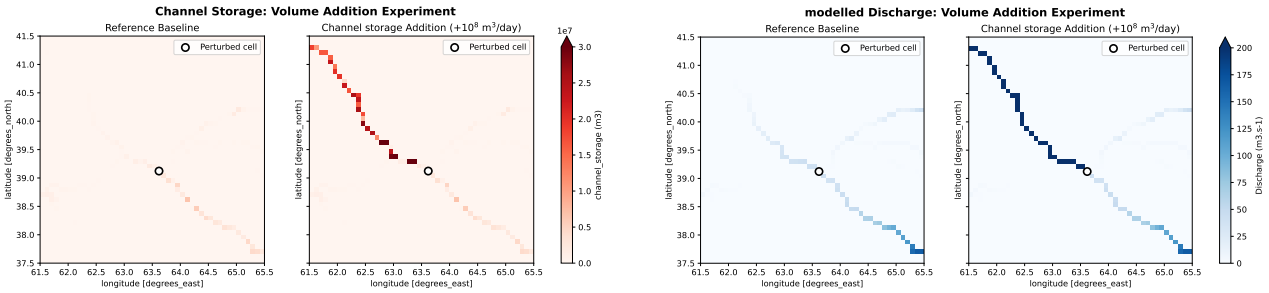
The calibration implementation in this study was made possible by introducing additional calibration parameters within the PCR-GLOBWB2 model configuration (.ini) files. Within this setup, spatially distributed parameter fields are not modified directly. Instead, multiplicative scaling factors defined in the configuration files are applied to these fields during model initialization and runtime.

This approach allows the original PCR-GLOBWB2 source code to remain unchanged while enabling systematic adjustment of key hydrological processes through external configuration. The implementation is fully compatible with the eWaterCycle workflow.

update of BMI - model state manipulation

To support the experimental design, a modification was implemented in Basic Model Interface (BMI) (E. Hutton et al., 2020) implementation of PCR-GLOBWB2 to enable controlled manipulation of internal model states. This update allows for dynamic perturbation of selected state storage variables during experiment runtime in for example Jupyter notebooks, without modifying the core model structure.

Figure 4.1 shows an example of this procedure by comparison of channel storage and simulated discharge under perturbed and reference conditions. The results illustrate how modifications in internal storage states propagate (figure 4.1a through the PCR-GLOBWB2 river routing scheme and affect downstream discharge behavior (figure 4.1b).



(**a**) Comparison of channel storage.

(**b**) Comparison of simulated river discharge.

Figure 4.1: Model results for the channel storage perturbation experiment, where changing the model river channel storage state also leads to changes in river discharges downstream.

4.3.2. Regional Calibration of PCR-GLOBWB 2

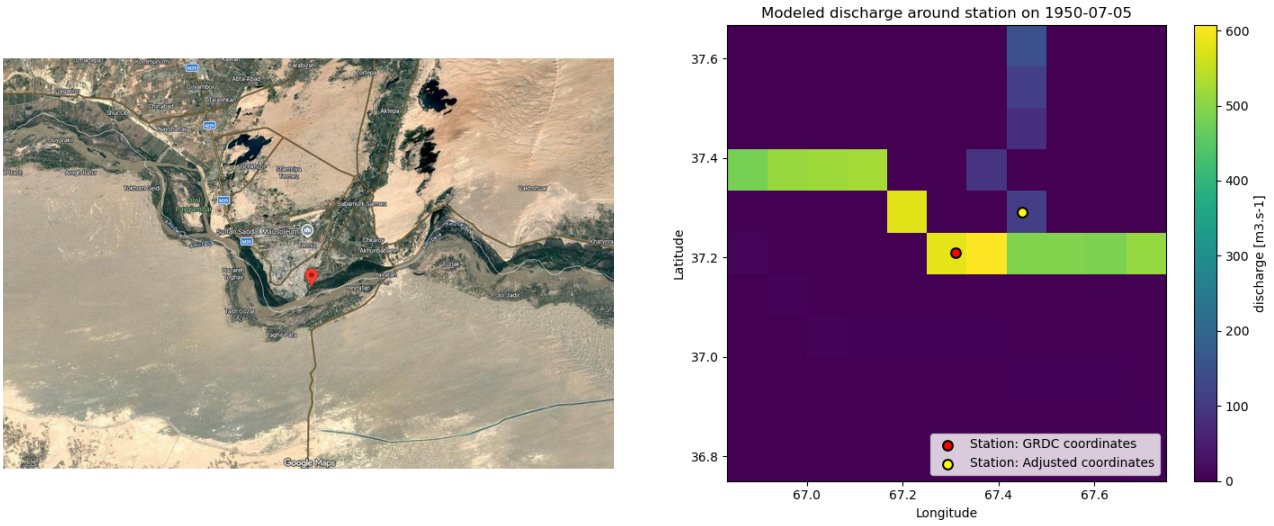
In order to increase performance of the PCR-GLOBWB2 model, described in section 3.2.1, it was calibrated. Instead of using the provided standard model global parameter values, the model was calibrated against observed discharge to increase regional performance.

Calibration data and forcing

Meteorological forcing was derived from the ERA5 reanalysis product (Hersbach et al., 2020). The model was calibrated using 49 monthly discharge stations from GRDC: The global runoff data centre (n.d.). Only stations with at least 95%data coverage in the calibration period were included in the calibration dataset.

The reported GRDC station coordinates required correction to ensure consistency with the river network represented in the model. Two main issues motivated these adjustments. First, the spatial resolution of the model grid can smooth local topographic gradients, causing small tributaries in mountainous regions to be poorly represented or effectively averaged out within a grid cell. Second, stations located near confluences may be assigned to the main river stem when using the original coordinates, while the observations correspond to discharge measured in a tributary. Although automated approaches are available (van der Ven, 2021), the small number of stations made manual matching more practical.

The correction procedure is illustrated in Figure 4.2. The exact coordinates used for calibration are listed in Appendix A.2.



(**a**) GRDC station location based on provided coordinates shown using Google Maps

(**b**) Original and adjusted GRDC station location within the generated PCR-GLOBWB2 river network.

Figure 4.2: Example of the GRDC station location correction procedure. Using the original GRDC coordinates places the station on the main stem of the Amu Darya, whereas the discharge observations correspond to a tributary inflow. Station locations were therefore manually adjusted to match the actual measurement location.

Spin-up and calibration period

To minimise the influence of initial conditions, a spin-up period from 1947 to 1949 was used to initialise soil moisture, groundwater storage, and routing components. The calibration period covered 1950 to 1954, during which model performance was evaluated against observed discharge. This period was chosen because that represents the period before large scale expansion of irrigated agriculture.

Objective function

Model performance was evaluated using the Nash–Sutcliffe Efficiency (NSE) by Nash and Sutcliffe (1970) and the Kling–Gupta Efficiency (KGE) by Gupta et al. (2009).

The Nash–Sutcliffe Efficiency (NSE) quantifies the relative magnitude of the residual variance compared to the variance of the observations:

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{n}(Q_t^{obs} - Q_t^{sim})^2}{\sum_{t=1}^{n}(Q_t^{obs} - \overline{Q}^{obs})^2}. \tag{4.1}$$

NSE values range from $-\infty$ to 1, where a value of 1 indicates perfect agreement between simulated and observed discharge, while values below zero indicate that the observational mean provides a better estimate than the model simulation.

Because NSE is unbounded for poor model performance, large negative values may dominate aggregated performance metrics or multi-objective evaluation criteria. To reduce this sensitivity, NSE was transformed following

$$\mathrm{NSE}^* = \frac{\mathrm{NSE}}{2 - \mathrm{NSE}}, \tag{4.2}$$

which rescales the metric to the bounded interval $[-1, 1]$. This transformation preserves the relative ranking of model performance for positive NSE values while limiting the influence of extremely poor simulations, thereby preventing the NSE component from disproportionately dominating the overall evaluation metric.

The KGE decomposes model performance into correlation, bias, and variability components:

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}, \tag{4.3}$$

where r is the linear correlation coefficient between observations and simulations, α is the ratio of simulated to observed standard deviation, and β is the bias ratio. KGE was computed separately for each station based on the corresponding time series.

To combine complementary aspects of model performance, a composite station-level score was defined as a weighted sum of both metrics:

$$P_s = w_{\mathrm{NSE}}\,\mathrm{NSE}_s^* + w_{\mathrm{KGE}}\,\mathrm{KGE}_s, \tag{4.4}$$

where $w_{\mathrm{NSE}} + w_{\mathrm{KGE}} = 1$. This formulation was used for exploratory calibration to ensure that both error magnitude (NSE^*) and distributional agreement (KGE) contributed to model evaluation.

For optimisation purposes, the objective function was defined as:

$$J_s = 1 - P_s. \tag{4.5}$$

Although a single composite objective was used during calibration, all station-level performance metrics were retained for each trial. This enables post-hoc analysis of spatial variability and calibration robustness across the model domain. The composite metric is therefore intended as a diagnostic optimisation tool rather than a final performance indicator.

To account for differences in hydrological representativeness, station contributions were weighted according to the square root of the upstream contributing catchment area:

$$w_s = \sqrt{A_s}, \tag{4.6}$$

where A_s is the upstream contributing area associated with station s, as provided by GRDC. This weighting scheme represents a compromise between equal station weighting and linear area-proportional weighting. Larger downstream stations, which integrate hydrological processes over substantial portions of the basin and are particularly relevant for downstream water balance analysis, are thereby given greater influence while still preserving the contribution of upstream tributary stations.

The global calibration objective was defined as the weighted mean loss:

$$J = \frac{\sum_{s=1}^{N} w_s \, J_s}{\sum_{s=1}^{N} w_s}. \tag{4.7}$$

Calibration aimed at minimising J.

Parameter selection

A total of nine parameters were selected for calibration, representing key hydrological processes across meteorology, land surface, groundwater, routing, and water use components of the PCR-GLOBWB model, as described in Section 3.2.1. The selected parameters were based on previous studies applying calibration methods to regionalize the PCR-GLOBWB2 model (Guan et al., 2026; López López et al., 2017; Ruijsch et al., 2021).

Parameter scaling was implemented either as linear scaling multiplier or logarithmic multipliers applied to existing model parameters. An overview of the calibrated parameters is provided in Table 4.2.

Linear scaling was defined as:

$$X_{\text{new}} = k_f \cdot X_{\text{original}}, \tag{4.8}$$

where k_f is a dimensionless scaling factor.

While logarithmic scaling was applied as:

$$\log_{10}(X_{\text{new}}) = k_f + \log_{10}(X_{\text{original}}), \tag{4.9}$$

which is equivalent to scaling by 10^{k_f} in linear space. Scalar parameters were directly constrained within predefined bounds.

Parameter	Model component	Transformation	Range	Initial
potential evapotranspiration	Meteorological forcing	linear multiplier	$[0.5,\ 2.0]$	1.0
precipitation	Meteorological forcing	linear multiplier	$[0.4,\ 1.5]$	1.0
temperature	Meteorological forcing	linear multiplier	$[0.96,\ 1.04]$	1.0
Manning's N	Surface water routing	scalar	$[0.005,\ 0.1]$	0.04
Storage Capacity	Groundwater	linear multiplier	$[0.5,\ 5.0]$	3.0
Recession Coefficient	Groundwater	$\log_{10}$ multiplier	$[0.5,\ 5.0]$	1.0
Saturated Hydraulic Conductivity	Land Surface	$\log_{10}$ multiplier	$[0.5,\ 5.0]$	3.0
Crop coeffecient	Irrigation & water use	linear multiplier	$[0.5,\ 5.0]$	1.0
minimum soil depth fraction	Land Surface	linear multiplier	$[0.5,\ 5.0]$	1.0

Table 4.2: Calibration parameters used in the PC-GLOBWB calibration. Parameters are either applied directly (scalar) or as multiplicative factors in linear or $\log_{10}$ space of spatially distributed parameter fields. Bounds and Initial value guesses also provided.

Computational constraints

PCR-GLOBWB has a substantial computational cost, with an average runtime of approximately 4.5 seconds per simulated day (approximately 30 minutes per simulated year) in a single-run configuration. When simulations are executed in parallel on the computational infrastructure used in this thesis (see Appendix D), runtime per evaluation can increase to approximately 10 seconds per simulated day due to I/O and CPU contention effects. In contrast to simpler rainfall-runoff models such as HBV (Appendix E), this limits the number of feasible model evaluations during calibration. As a result, the total number of function evaluations was constrained by available computational resources, making efficient sampling strategies essential.

Optimization method and configuration

The calibration was performed using the Tree-structured Parzen Estimator (TPE) sampler (Bergstra et al., 2011; Watanabe, 2025) within the Optuna hyperparameter optimization framework (Akiba et al., 2019). TPE is a Bayesian optimization method that models the probability distributions of well-performing and poorly performing regions in the parameter space. New samples are selected by maximizing the expected improvement acquisition function. In this implementation, the sampler initially performs 20 random trials before constructing the probabilistic density models used by TPE. For highly non-linear models, such as the hydrological model used in this study, global exploration of the parameter space is preferred to avoid premature convergence to local optima (Saltelli, 2008).

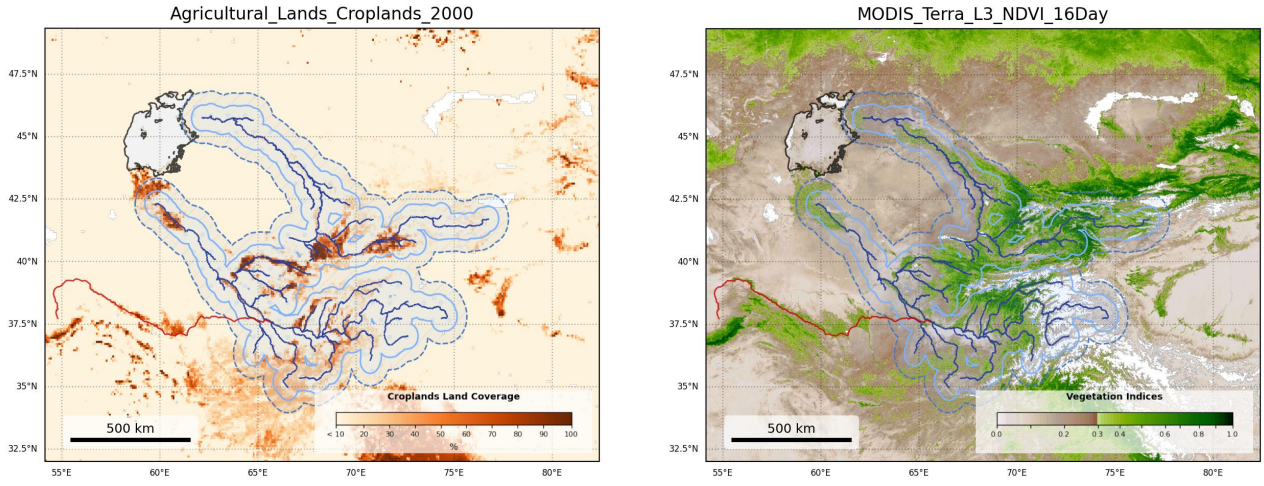
Due to the computational cost of PCR-GLOBWB simulations, the calibration was constrained by a fixed evaluation budget. The optimization was conducted using two parallel runs on 7 computational cores, with 17 evaluations per core, resulting in a total of 119 model evaluations per run. This setup was chosen to improve coverage of the parameter space while remaining within available computational resources. The objective function defined in Equation 4.7 was minimized in all runs. All simulated discharge time series were stored to enable post-optimization evaluation.

Validation setup

Model validation was performed using the same set of GRDC discharge stations, but for a period independent of the calibration window where data availability allowed. Model performance during validation was assessed using the same objective metrics (NSE and KGE_{2009}) to ensure consistency between calibration and evaluation. This allowed for a direct assessment of model transferability beyond the calibration period.

4.3.3. Modeling water transfer schemes - Karakum Canal

As detailed in section 3.6, PCR-GLOBWB2 models rain-fed agricultural practices with grid-cell-based water balances while irrigated agriculture is simulated using larger zones from which irrigation water demand can be supplied. However, the 1,400 km scale diversion dynamics of the Karakum Canal (Figure 4.3, section 3.6) (FAO, 2012; Frenken, 2013; Nachtnebel, 2006) fall completely outside these model mechanisms. To resolve this limitation, the massive water abstraction by the canal was introduced through a direct channel-storage state manipulation.



(**a**) croplands circa 2000

(**b**) NDVI-16Day

Figure 4.3: River networks are shown together with 50 km and 100 km zones, which are used in PCR-GLOBWB to represent the spatial extent of irrigation water demand. The figure indicates that the majority of irrigated areas fall within the 100 km buffer. Cropland extent (circa 2000) and buffered NDVI-16Day are included for context. The Karakum Canal is highlighted in red (OpenStreetMap contributors, 2026). Rivers are shown for reference and are not representative of the PCR-GLOBWB river routing scheme (Lehner & Grill, 2013).

At each daily timestep (t), the implemented abstraction scheme determines the volume of water stored within the specific Amu Darya river channel reach (S_{reach}) through the grpc4bmi interface within the eWaterCycle framework. The fraction of available storage required to fulfill the targeted daily extraction demand (Q_{Karakum}) is calculated and subtracted from the state between model update steps to account for the permanent removal

of the diverted water mass. This subtraction was limited to a maximum of 80% of available channel storage to ensure numerical stability and to avoid full depletion of the river reach within each single timestep.

To evaluate the downstream impacts of temporal demand variations, the extraction was simulated under two distinct scenarios: a static baseline case ($Q_{\text{Karakum}} = R_{\text{annual}}/365$) and a seasonally fluctuating scenario governed by a cosine-based wave-form function:

$$Q_{\text{Karakum}}(\text{doy}) = \left(1 - \cos\left(\frac{2\pi(\text{doy} - 1)}{365}\right)\right) R_{\text{daily}} \tag{4.10}$$

Where $Q_{\text{Karakum}}(\text{doy})$ is the seasonally adjusted volumetric extraction rate (m^3/day), doy is the day of the year (1 to 365), and R_{daily} is the reference baseline daily removal target (m^3/day). Given that the available data on the canal abstraction are given in approximate annual volumes, the $(1 - \cos(\theta))$ formulation ensures that the targeted total annual water extraction is the same between both modeling scenarios, while allowing for water demand to better follow agricultural growing season.

Limitations

The implementation involves several simplifications. The abstraction scheme assumes instantaneous removal of water from a single river reach. Spatial redistribution of diverted water outside the model domain and irrigation return flows are not represented. Operational constraints, including water allocation agreements and transboundary water management rules, are also not included, and abstractions follow prescribed demand scenarios only.

In addition, regulating reservoirs along the Karakum Canal network are not explicitly represented. These structures may influence the timing of water deliveries by buffering or redistributing flows along the canal system, which is not captured in the present implementation. The 80% channel storage constraint is a numerical safeguard rather than a physically based threshold and may influence abstraction behaviour under low-flow conditions. Finally, the prescribed seasonal distribution of demand is an idealised forcing and does not represent operational decision-making in the Karakum Canal-Amu Darya system.

4.3.4. PCR-GLOBWB2 Experiments

The PCR-GLOBWB2 model was executed across a suite of historical and future scenarios spanning 1950–2100. This experimental framework serves a dual purpose: first, it functions as a diagnostic tool to evaluate climate forcing pre-processing techniques and overall model behavior; second, it provides the essential downstream inflow data required for the broader Aral Sea modeling chain. A comprehensive overview of these simulation configurations is structured by objective in Table 4.3.

Baseline Historical Simulations

Initial historical baseline runs (1950–2014) were driven by raw ERA5 meteorological forcing. An uncalibrated simulation was first established to quantify baseline model performance and evaluate uncalibrated model behavior. Following the primary calibration phase, the "Overall Calibrated" parameter set was implemented as the core basin-wide benchmark. Finally, to secure optimal downstream inflow data for subsequent Aral Sea modeling, a post-calibration analysis was performed to isolate the specific parameter sets that yielded the highest predictive performance for the downstream gauges of the Amu Darya and Syr Darya sub-basins respectively.

Climate Model Bias correction evaluation

To isolate the impacts of climate forcing preprocessing and assess model sensitivity to forcing uncertainty, a series of historical GCM-driven simulations (1950–2014) were performed using four models from the CMIP6 ensemble: MIROC6, CanESM5, MPI-ESM1.2-HR, and MRI-ESM2.0. These simulations compared uncorrected (regridded raw) configurations directly against bias-corrected forcing data, allowing for a systematic evaluation of how bias-correction preprocessing alters simulated hydrological behavior.

Climate Model and Future Projections

For future projections (2015–2100), all four GCMs were implemented using bias-corrected forcing under the Shared Socioeconomic Pathway SSP2-4.5 to establish a multi-model medium emission ensemble. To isolate system sensitivity across a broader spectrum of future climate trajectories, the MPI-ESM1.2-HR model was further simulated under all remaining available scenarios, namely SSP1-2.6, SSP3-7.0, and SSP5-8.5.

Anthropogenic Sensitivity Experiments

Beyond climatic variations, a final set of experiments was conducted to evaluate the hydrological impact of the Karakum Canal. Because this massive 1,400 km water diversion is not natively captured by the default PCR-GLOBWB2 routing architecture, these runs serve to evaluate alternative methods for representing unmodeled human infrastructure and its downstream propagation. Abstraction schemes were evaluated by manipulating either a single grid cell or multiple grid cells, testing both constant withdrawal and dynamic seasonal wave abstraction regimes. The specific mathematical formulations, numerical constraints, and implementation logic for these abstraction schemes are detailed in Section 4.3.3.

Computational Workflow and Simulation Design

To optimize computational efficiency and streamline the development cycle, the continuous long-term simulations (1950-2015 and 2015-2100) were divided into three roughly equal temporal blocks. This strategy provided quicker feedback loops during the initial testing phases, allowing for more rapid verification of model setups without awaiting the completion of a full run. Furthermore, this method reduced individual segment runtime to approximately 24 hours, minimizing time and data loss during a simulation failure. Each independent simulation block included 5-year initialization period to establish stable hydrological storage states prior to the data collection period. No discontinuities in streamflow were observed.

4.3.5. Evaluation metrics

Model performance is evaluated using a set of complementary metrics designed to capture different aspects of hydrological behaviour, depending on the objective of each experimental setup. Given the focus on long-term water balance and cumulative system behaviour, particular emphasis is placed on metrics describing long-term discharge evolution, while short-term discharge metrics are used to assess hydrological dynamics where a reference simulation is available.

For bias-correction experiments, model performance is evaluated against an ERA5-driven reference simulation using the Kling–Gupta Efficiency (KGE), computed on daily discharge. The KGE captures agreement in timing, variability, and bias, and is used here as a measure of hydrograph similarity. As the forcing data originate from fundamentally different sources (reanalysis versus climate model output), perfect agreement is not expected; instead, KGE is interpreted as an indicator of how well the simulations reproduce the seasonal dynamics and flow amplitude of discharge.

Long-term water balance consistency is assessed using cumulative discharge diagnostics. Daily discharge is converted to volumetric units and accumulated over time, and deviations from the reference trajectory are quantified using the Root Mean Square Error (RMSE) and relative differences in cumulative discharge at selected checkpoint years.

For future climate projections, no observational reference is available; therefore, evaluation focuses on relative changes in discharge between early and late century periods, cumulative discharge totals, and inter-model spread across the CMIP6 ensemble.

In addition, flow duration curves (FDCs) are used as a diagnostic tool to assess the similarity of flow regimes, particularly the representation of high- and low-flow conditions independent of timing.

Table 4.3: Summary matrix of historical, future, and anthropogenic PCR-GLOBWB2 experiments for use in the Aral Sea modeling chain (1950–2100).

Forcing Data	Experimental Variable (Setup)	Research Objective
Baseline Historical Runs (1950–2014 / ERA5)		
ERA5 (Raw)	Uncalibrated	Establish baseline uncalibrated model performance.
	Overall Calibrated	Primary benchmark representing optimized basin-wide performance.
	Sub-basin Calibrated (Amu/Syr)	Post-calibration selection optimized for downstream flow to support Aral Sea modeling.
GCM Historical Ensemble (1950–2014: MIROC6, MPI-ESM1.2-HR, CanESM5, MRI-ESM2.0)		
CMIP6 Historical	Regridded (Raw)	Evaluate baseline climate model structural uncertainty.
	Bias Corrected	Quantify the impact and performance of bias-correction methods.
GCM Future Projections (2015–2100: Bias Corrected)		
GCM Ensemble (above)	SSP2-4.5 Scenario	Evaluate multi-model future sensitivity under a medium emission pathway.
MPI-ESM1.2-HR	SSP1-2.6 Scenario SSP2-4.5 Scenario SSP3-7.0 Scenario SSP5-8.5 Scenario	Isolate system sensitivity to a wide range of future climate trajectories (all available SSP scenarios).
Anthropogenic Sensitivity Experiments (Default Parameters)		
ERA5	Single Grid Cell: Constant Multiple Grid Cells: Constant Multiple Grid Cells: Seasonal Wave	Evaluate various methods for representing unmodeled anthropogenic water abstraction infrastructure and its downstream propagation.

4.4. Aral Sea model

The primary objective of this study is to assess the response of the Aral Sea to different climate forcing scenarios. As the terminal lake of an endorheic basin, the Aral Sea integrates all upstream hydrological processes, including river discharge generation, local atmospheric fluxes, and anthropogenic water use. Its temporal evolution therefore provides a basin-scale measure of hydrological change under varying climate conditions.

To translate climate-driven hydrological changes into lake dynamics, a conceptual volume balance model of the Aral Sea was developed. In this framework, river discharges simulated by PCR-GLOBWB under historical and future climate scenarios provide the primary inflow boundary conditions. The modelling chain thus links climate forcing to runoff generation and ultimately to lake storage dynamics.

4.4.1. Conceptual model structure

The Aral Sea is represented as a lumped system in which water storage evolves according to inflows from major rivers, direct precipitation, evaporation losses, and groundwater exchange. The model explicitly accounts for the dominant hydrological fluxes controlling lake volume, while neglecting minor or poorly constrained processes.

The model components are summarised in Table 4.4.

Governing equations

The Aral Sea is a closed basin system in which river discharge does not reach the ocean but terminates in an inland lake. The dominant inflows are provided by the Amu Darya and Syr Darya rivers, supplemented by direct precipitation and to a lesser extent groundwater exchange. The primary loss term is evaporation from the lake surface.

The dynamic water volume balance of the Aral Sea is mathematically expressed as:

$$V(t) = V(t-1) + \left[Q_{\text{Amu}}(t) + Q_{\text{Syr}}(t) + Q_P(t) - Q_E(t) + Q_{\text{gw}}\right] \Delta t \tag{4.11}$$

where $V(t)$ is the total lake volume (km^3) at time step t, $Q_{\text{Amu}}(t)$ and $Q_{\text{Syr}}(t)$ are the respective river discharge fluxes, $Q_P(t)$ is the volumetric direct precipitation rate, $Q_E(t)$ is the volumetric open-water evaporation rate,

Table 4.4: Components of the Aral Sea mass and volume balance model.

Type	Component	Description / Source
Forcing	Amu Darya discharge	PCR-GLOBWB / literature-based estimates
Forcing	Syr Darya discharge	PCR-GLOBWB / literature-based estimates
Forcing	Precipitation over lake	ERA5 / CMIP forcing data
Forcing	Potential evaporation	ERA5 / CMIP forcing data
Forcing	Groundwater exchange	Literature-based estimate (Micklin, 2007)
State	Lake volume	Modelled state variable
Output	Lake area	Derived from hypsometry
Output	Water level	Derived from volume–area relation
Output	Evaporation flux	Makkink-based formulation
Parameter	Bathymetry	(Khazaei et al., 2022; Nachtnebel, 2006)

and Q_{gw} represents a constant net groundwater exchange flux. All flux terms are evaluated in units of km^3/day over the discrete temporal step Δt. Positive fluxes denote net gains to the lake system.

The initial state boundary condition at the simulation baseline $(t = 0)$ is defined as:

$$V(0) = V_0 \tag{4.12}$$

where V_0 represents the initial baseline storage volume configuration. Depending on the temporal window and experimental design of the simulation run, V_0 is parameterized using two distinct configurations:

- Historical Baseline (1950-2015): $t = 0$ corresponds to 1 January 1950, where $V_0 = 1060$ km^3
- Future Projections (2015-2100): $t = 0$ corresponds to 1 January 2015, where $V_0 = 75$ km^3

River inflow

River inflow to the Aral Sea is dominated by discharge from the Amu Darya and Syr Darya rivers. Daily discharge time series derived from PCR-GLOBWB simulations (section 4.3.4) are used as input forcing for the lake model.

Precipitation

Direct precipitation contributes to lake inflow and is computed as:

$$Q_P(t) = 10^{-6}\, P(t)\, A(t) \tag{4.13}$$

where $P(t)$ is precipitation over the lake surface (mm) and $A(t)$ is lake surface area. The factor 10^{-6} converts units to km^3. Precipitation over the upstream catchment is not included explicitly, as it is already represented in river discharge simulations.

Evaporation

Evaporation is the dominant loss term in the Aral Sea water balance. Total lake evaporation is computed as:

$$Q_E(t) = E_p(t)\, A(t)\, C_{\text{Makkink}} \tag{4.14}$$

where $E_p(t)$ is potential evaporation derived using the Makkink formulation, $A(t)$ is lake surface area, and C_{Makkink} is a conversion factor representing open-water evaporation correction. The Makkink approach was selected due to consistency with available climate forcing data.

Groundwater exchange

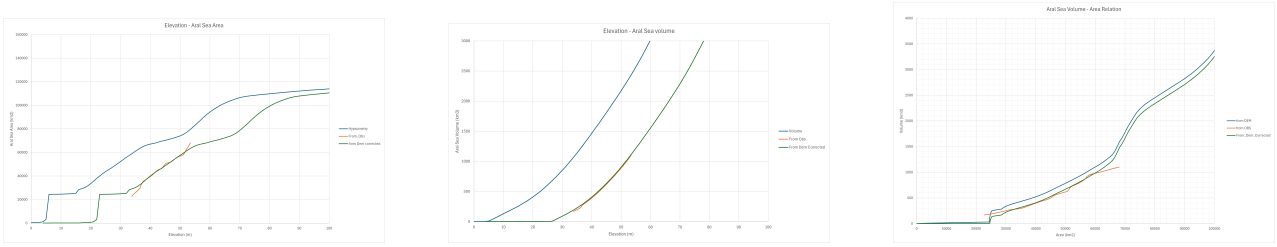
Groundwater exchange is uncertain and relatively small compared to river inflow and evaporation. Literature estimates range between 0.5–1 km^3 yr^{-1} (Cretaux et al., 2013; Micklin, 2007). Given its limited magnitude, groundwater exchange is represented as a constant term in the reference simulations.

Lake hypsometry

Because evaporation and precipitation depend on lake surface area, a dynamic relationship between lake volume, surface area, and water level is required. This relationship is derived from high-resolution bathymetry data and defines the hypsometric characteristics of the Aral Sea.

A digital elevation model is used to construct the relationships between elevation, surface area, and lake volume. From these relations, both lake area and water level are derived diagnostically from simulated volume (Figure 4.4).

This formulation allows consistent coupling between hydrological fluxes and lake geometry without prescribing lake shape evolution.



(**a**) Elevation–area relationship (**b**) Elevation–volume relationship (**c**) Volume–area relationship

Figure 4.4: Hypsometric relationships of the Aral Sea derived from bathymetry data, linking elevation, surface area, and volume.

Workflow integration

The Aral Sea model serves as the final integration step in the modelling workflow, linking climate forcing scenarios to basin-scale hydrological response to dynamics of water level in Aral Sea. While PCR-GLOBWB simulates runoff generation and river discharge, the lake model translates these inflows into changes in lake storage, surface area, and water level. This enables direct assessment of how climate variability and anthropogenic water use propagate through the hydrological system to affect the terminal basin.

4.4.2. Model validation and inflow parameters

The long-term average surface water inflow into the Aral Sea from the combined discharges of the Amu Darya and Syr Darya is established at approximately 56 km^3/year (Duan et al., 2024; Micklin et al., 2014; Nachtnebel, 2006).

To account for total water inputs, groundwater flux must also be included. The majority of existing hydrological research estimates baseline groundwater inflow to be relatively minor, ranging between 1 and 3 km^3/year (Crétaux et al., 2005; Micklin et al., 2014). However, a small number of outlier studies propose significantly higher sub-surface contributions, with upper limits reaching up to 30 km^3/year as noted by Cretaux et al. (2013). Aligning the model configuration primarily with the mainstream literature consensus yields a combined long-term average inflow target range of 55 to 65 km^3/year.

The model was validated against the pre-diversion Aral Sea lake levels using ERA5 meteorological forcing and synthetic constant inflows within the 55 to 65 km^3/year range, starting from an initial volume of 1,060 km^3 on 1 January 1950. Under these conditions, the simulated Aral Sea should stabilize near its pre-diversion water levels, even if the model does not fully replicate specific historical internal dynamics.

5

Results

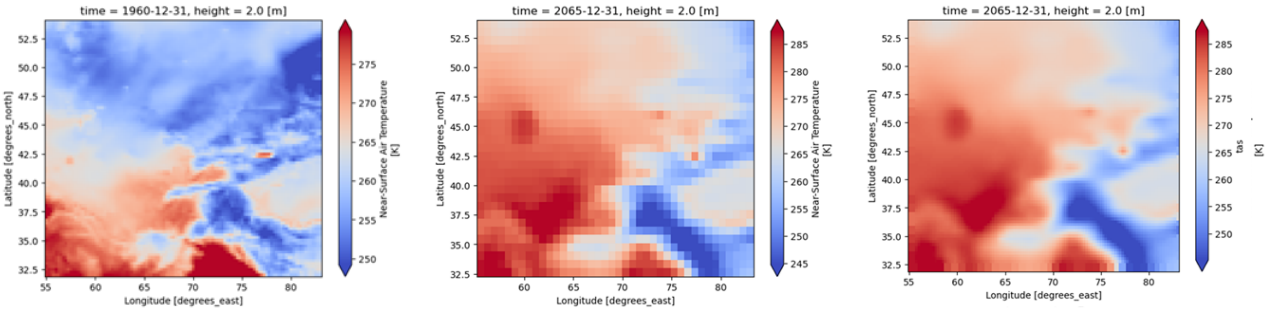
This chapter presents the results of the modelling framework, including forcing generation, calibration performance, and scenario simulations. The findings are organised following the structure of the modelling workflow introduced in the previous chapter.

5.1. Forcing Generation

5.1.1. Forcing Harmonization

The CMIP6 forcing datasets were transformed following the workflow described in section 4.2.1. Initial spatial resolution and grid structure depended on the climate model output generated using the ESMValTool (Eyring, Righi, et al., 2016) within eWaterCycle. The final CMIP6 forcing datasets were produced at the same spatial resolution, grid definition, and variable structure as the ERA5 forcing.

Figure 5.1 illustrates the forcing harmonization workflow with CMIP6 model MPI-ESM1-2-HR. Figure 5.1a shows ERA5 forcing data already compatible with PCR-GLOBWB2, Figure 5.1b presents MPI-ESM1-2-HR forcing prior to preprocessing, and Figure 5.1c shows the MPI-ESM1-2-HR forcing after transformation into a PCR-GLOBWB2 ready format.



(**a**) ERA5 meteorological forcing (PCR-GLOBWB2 compatible)

(**b**) MPI-ESM1-2-HR meteorological forcing before pre-processing

(**c**) MPI-ESM1-2-HR meteorological forcing after pre-processing (PCR-GLOBWB2 compatible)

Figure 5.1: Workflow and results of meteorological forcing generation.

Validation

Several checks were performed to validate consistency between the transformed CMIP forcing and the ERA5 reference dataset:

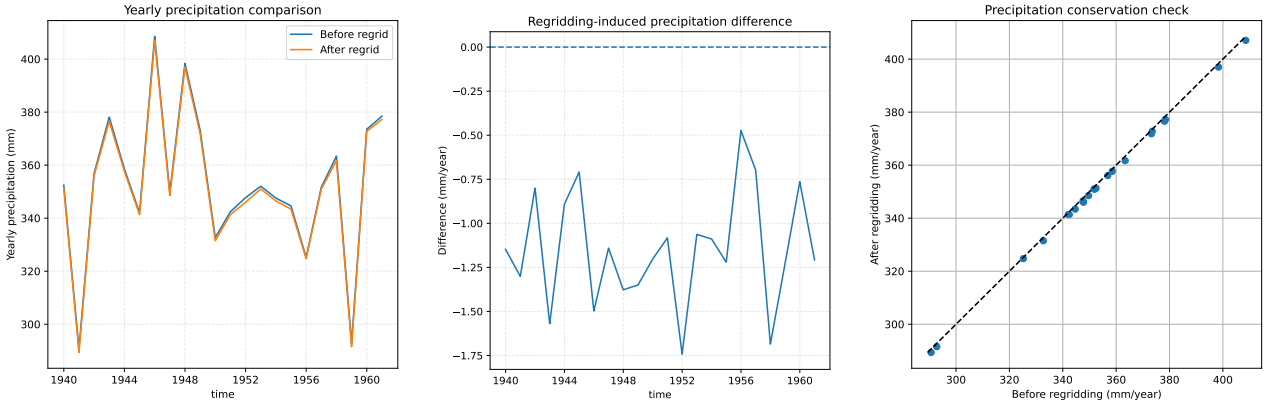
- Spatial consistency was verified by comparing grid cell counts and basin coverage between datasets (see Table 5.1).
- Basin-averaged precipitation totals were evaluated to assess preservation of large-scale water input. Figures 5.2 shows that the of yearly area-weighted precipitation before and after regridding is consistent.

The scatter plot further confirms this consistency, with values closely aligned along the 1:1 line. The mean difference is -0.33% or ≈ 1mm of yearly precipitation.

Table 5.1: Comparison of spatial resolution of ERA5 and CMIP forcing data before and after regridding and resampling.

	ERA5	MPI-ESM1-2-HR	
	(Reference)	(Before regridding)	(After regridding)
Latitude range (°N)	54.0 - 32.0	53.77 - 32.26	54.0 - 32.0
Longitude range (°E)	55.0 - 83.0	55.31 - 82.50	55.0 - 83.0
Latitude resolution (°)	0.25	0.94	0.25
Longitude resolution (°)	0.25	0.94	0.25
Latitude points	89	24	89
Longitude points	113	30	113
Total grid cells	10 057	720	10 057

These validation steps indicate that the processed CMIP forcing is consistent with ERA5 forcing and suitable for use in subsequent PCR-GLOBWB2 simulations.



(**a**) Yearly precipitation (**b**) Difference (**c**) Conservation check

Figure 5.2: Assessment of precipitation conservation during regridding. (a) Yearly basin-averaged precipitation, (b) differences after regridding, and (c) scatter-based conservation check against the 1:1 line.

5.1.2. Bias-Corrected Climate Forcing

Prior to their use in the hydrological simulations, the historical and future CMIP6 climate data from the selected GCMs were bias-corrected using ERA5 reanalysis as the reference baseline, following the methodology outlined in Section 4.2.2. The impact of this adjustment on the historical period and the future projections for MPI-ESM1-2-HR is illustrated in figure 5.3 with both the uncorrected and corrected precipitation and temperature data shown.

Historical Evaluation: Uncorrected vs. Bias-Corrected Forcing

For the historical baseline period (1940–2015), the correction methodology substantially reduces systematic errors across all evaluated models. In the case of temperature, the baseline warm bias characteristic of the raw CMIP6 GCMs is fully eliminated, aligning the mean temperature perfectly with the ERA5 reference data. For instance, the correction removes a substantial bias of 4.4°C for MIROC6, bringing the model average down to the baseline reanalysis level.

For precipitation the systemic bias is minimized, but it is not entirely removed. For instance the wet bias for MRI-ESM2-0 is reduced from +35.6% compared to ERA5 down to just +3.9% mean annual precipitation. Similarly, for MPI-ESM1-2-HR this reduced the dry bias from -8.1% to -4.1%. This indicates that even if the bias correction significantly aligns the datasets, minor structural differences in mean climatology between the climate models and ERA5 persist, as quantified in table 5.2.

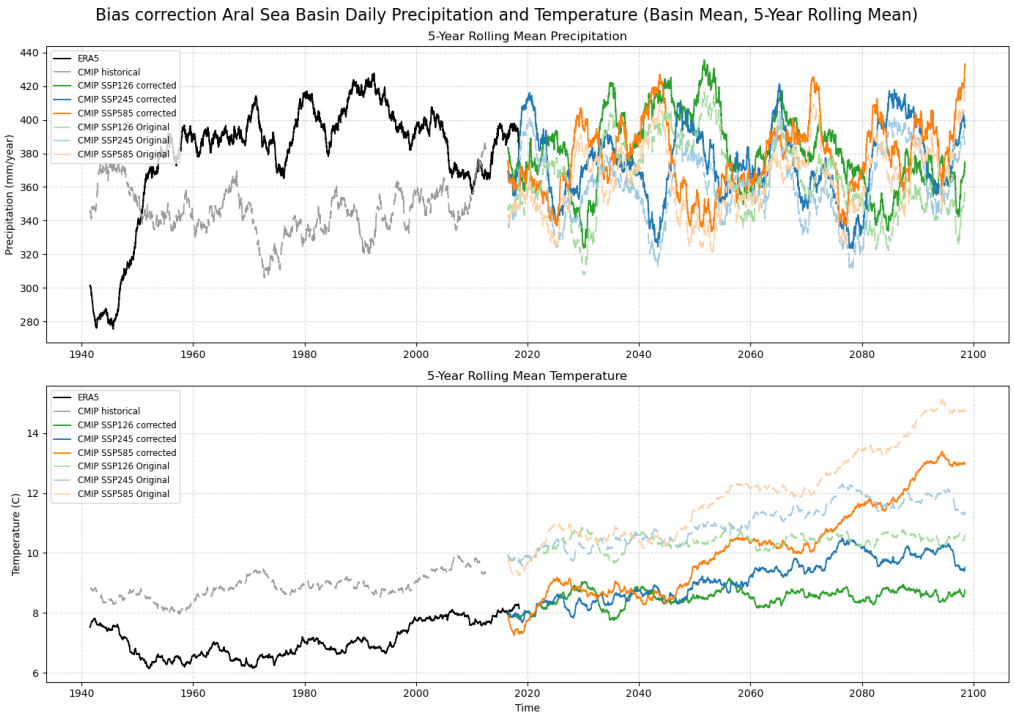


Figure 5.3: Example of raw and bias-corrected precipitation and temperature forcing derived from the MPI-ESM1-2-HR model. Future projections span the SSP1-2.6, SSP2-4.5, and SSP5-8.5 scenarios to capture a representative range of low- to high-emission pathways.

It should be noted that these calculated values serve primarily as a methodological validation of the bias correction workflow. They represent grid-wide averages over the entire region shown in figure 5.1, rather than a catchment-specific or topography-specific evaluation. However, these remaining discrepancies must be factored into the interpretation of the subsequent hydrological discharge simulations made on the basis of these forcing datasets.

Because this workflow successfully forces the historical simulations to align closely with the reference ERA5 baseline, it validates the use of their corresponding future projections, which are adjusted using the same correction procedure for integration into the hydrological model chain.

Table 5.2: Basin-averaged annual precipitation and temperature statistics for ERA5, CMIP6 uncorrected historical forcing, and bias-corrected CMIP6 data over 1940–2014. Bias is expressed relative to ERA5, which are used as pseudo-observations in this study. Of note is the large difference between precipitation for MRI-ESM2-0 and ERA5

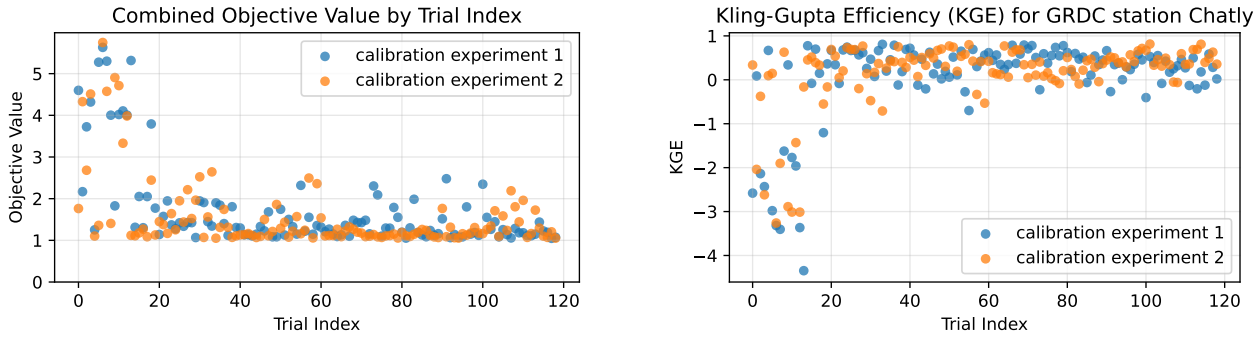
		Precipitation			Temperature		
Dataset	Type	Mean (mm/yr)	Std	Bias (%)	Mean (K)	Std	Bias (%)
ERA5	Reference	377.066	45.623	-	280.174	0.743	-
CanESM5	Raw	306.700	43.595	-18.662	278.780	0.963	-0.497
	Corrected	371.032	59.546	-1.600	280.177	0.944	0.001
MIROC6	Raw	382.185	44.786	1.357	284.606	0.672	1.582
	Corrected	386.181	48.864	2.417	280.184	0.614	0.004
MPI-ESM1-2-HR	Raw	346.309	36.632	-8.157	282.042	0.638	0.667
	Corrected	362.408	38.136	-3.887	280.184	0.667	0.003
MRI-ESM2-0	Raw	510.835	49.941	35.476	281.454	0.648	0.457
	Corrected	392.522	41.395	4.099	280.184	0.682	0.004

5.2. PCR-GLOBWB2 - Calibration

This section presents the results of the PCR-GLOBWB2 calibration procedure, including the resulting parameter sets and the associated performance across individual gauging stations.

5.2.1. Calibration Performance

PCR-GLOBWB2 was calibrated using two independent runs, each consisting of 119 simulations or trials, using the calibration method described in Section 4.3.2. Figure 5.4 shows that there are similar convergence patterns across the calibration runs. The majority of performance gains of the calibration objective (equation (4.7)) are achieved during the early phase, indicating efficient exploration of parameter space, while later generations primarily perform local refinement.



(**a**) Calibration objective function, 0 is best

(**b**) KGE performance at Chatly station, 1 is best

Figure 5.4: Calibration performance results showing (a) the overall objective function and (b) KGE performance at the Chatly station during calibration.

Figure 5.4a shows the performance of the combined objective function during calibration as defined in equation (4.7). The lowest achieved objective value across all calibration runs is $J = 1.048$, where $J = 0$ corresponds to a perfect fit. The objective function combines KGE and NSE, such that larger values reflect poorer overall agreement with observed discharge across both metrics.

As an example of station-level performance, figure 5.4b shows the Kling–Gupta Efficiency (KGE) score for the Chatly station, calculated using equation (4.3). Chatly is the most downstream station on the Amu Darya, located near the river's mouth at the Aral Sea, and as such the most important for monitoring inflow to the Aral Sea. It has as a KGE of 0.284 for the calibration trial with the best calibration score. However among all calibration trials performend the Chatly KGE has a value of 0.814, much closer to the optimum.

In the best-objective trial, of the ten highest weighted stations only three stations (Tyumen-Aryk, Bekabad, and Khorog) achieve KGE values above 0.5, indicating reasonable model performance at only a limited number of locations. However, when considering the best-performing simulations per station among all calibration runs this threshold is met by all stations except two. One of these station scores only slightly below 0.5 and another that shows such low scores that it likely indicates a station location misalignment or data quality issue (Table 5.3).

These results show a fundamental trade-off for multi-site calibration. Individual station performance can reach high KGE values when evaluating the full pool of calibration runs after the fact. However, the overall system performance, as well as the individual station performance for the trial with the best overall objective score (equation (E.1)), remain low. By trying to simultaneously balance multiple highly weighted stations across two separate river systems, the Amu Darya and the Syr Darya, the single objective function is forced to make trade-offs. This leads to a single compromise value that yields moderate performance for both river systems instead of high performance for either system individually.

To evaluate the impact of these trade-offs on the subsequent modeling chain, three parameter sets were selected for further simulation. The first is the overall best parameter set based on the global objective function. For comparison, two basin-specific parameter sets were also used: the trial yielding the best individual score (equation (4.4)) at the Chatly station, representing the Amu Darya inflow into the Aral Sea, and the trial yielding the highest performance at the Kazalinsk station, representing the Syr Darya inflow. These two stations

Table 5.3: Top 10 stations by weight with KGE and NSE at the best-objective trial (run 3, objective value 1.048) combined with each station's overall best scores across all trials from runs 3 and 4. All values calculated on monthly timeseries for 1950-1954.

Station	Weight	KGE (best trial)	NSE* (best trial)	NSE (best trial)	KGE (highest)	NSE* (highest)	NSE (highest)
Chatly	670.820	0.284	-0.072	-0.156	0.814	0.631	0.774
Tyumen-aryk	467.974	0.566	-0.005	-0.010	0.597	0.097	0.176
Bekabad	376.829	0.684	0.475	0.644	0.796	0.604	0.753
Kal	300.000	0.251	-0.154	-0.364	0.695	0.260	0.413
Uch-kurgan	241.661	0.208	-0.031	-0.064	0.791	0.552	0.712
Ust. Kekirim	186.011	0.106	-0.110	-0.247	0.790	0.508	0.674
Tutkaul	176.635	-0.157	-0.304	-0.873	0.585	0.444	0.615
Garm	141.421	-0.193	-0.286	-0.803	0.438	0.306	0.468
Barchadiv	129.228	-2.943	-0.918	-22.438	-0.052	-0.885	-15.436
Khorog	117.047	0.818	0.709	0.829	0.858	0.709	0.829

were selected because they are the most downstream gauging points for the two major tributary systems and therefore directly control the inflow boundary conditions of the Aral Sea water balance. This is one of the experiments described in section 4.3.4.

5.2.2. Impact of calibration parameters on model performance

Parameter importance was evaluated using mutual information (MI) (Kraskov et al., 2004), which quantifies the dependency between individual parameters and the calibration objective and is capable of capturing non-linear, non-monotonic relationships (figure 5.5). A value of 0 indicates no statistical dependence on the objective, while higher values indicate stronger dependence.

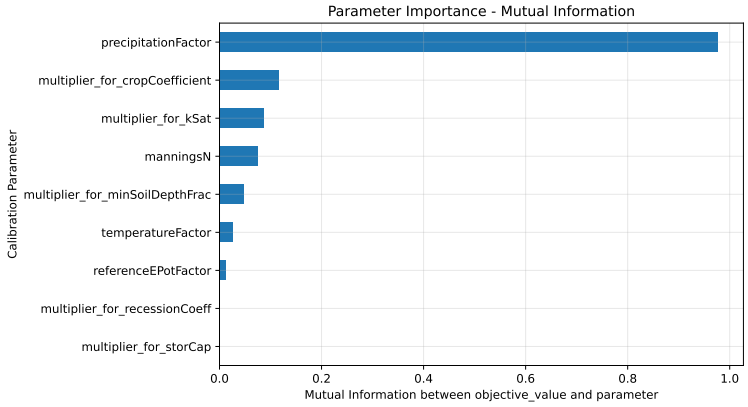


Figure 5.5: parameter importance: Mutual information based on (Kraskov et al., 2004) using ("Mutual_info_regression", n.d.)

The results show that `precipitationFactor` is by far the most influential parameter, with a mutual information value of 0.98, vastly exceeding all other parameters. This strong dependency between `precipitationFactor` and model performance can also be seen in the parameter distributions shown in the violin plots of the best-performing simulations (Figure 5.6). In these plots `precipitationFactor` shows a narrow concentration of high-performing parameter values between 0.4-0.5.

All other parameters have substantially lower mutual information values. The second most important parameter, `cropCoefficient`, has a much smaller MI of only 0.12, followed by `kSat` and `manningsN`, which contribute only weakly to model performance. Several parameters, including `stor Cap` and `recessionCoeff`, have (near-)zero mutual information, suggesting negligible influence on the calibration objective within the explored parameter space.

Overall, both the mutual information analysis and the violin plots consistently indicate that model performance is strongly dominated by `precipitationFactor`, while most other parameters have only a minor or negligible influence. The implications of which are discussed in sections sections 6.1 and 6.3

The violin plots also show that different combinations of exist that lead to similar outcomes, see for example `cropCoefficient`, or `minSoilDepth` where top performing calibration trials have very different values for

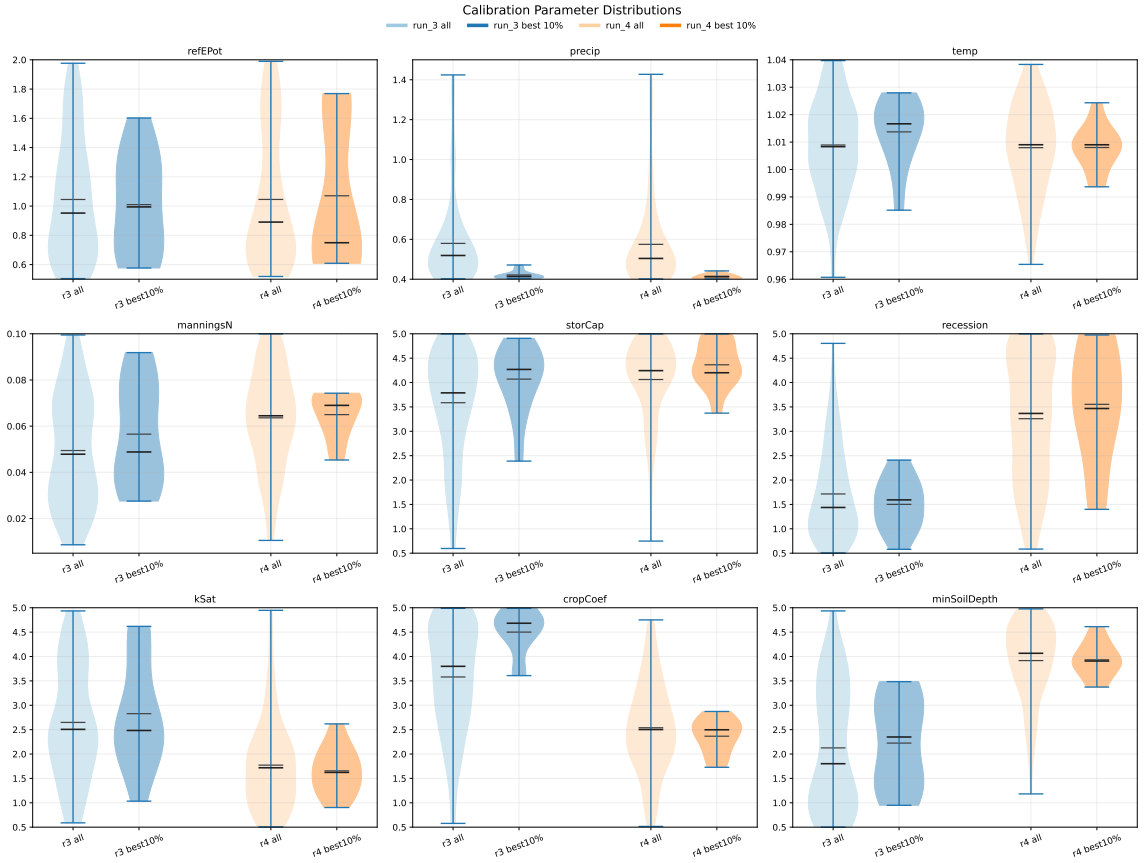


Figure 5.6: Distribution of PCR-GLOBWB2 calibration parameters across the two parallel calibration experiments optimizing for the overall objective (equation (4.7)). For each parameter, the four violins illustrate the complete parameter space sampled alongside the top 10% highest-performing subset for both parallel runs, demonstrating the variability within the optimization process.

these parameters.

Parameter	Uncalibrated	Overall Calibrated	Amu Calibrated	Syr Calibrated
potential evapotranspiration	1.0000	1.3657	1.5881	0.8277
precipitation	1.0000	0.4037	0.4844	0.4543
temperature	1.0000	1.0167	1.0067	1.0354
Manning's N	0.0400	0.0380	0.0507	0.0402
Storage Capacity	1.0000	4.3460	3.3180	4.4588
Recession Coefficient	1.0000	1.1505	2.6188	0.5080
Saturated Hydraulic Conductivity	1.0000	4.1716	2.8140	1.9315
Crop coeffecient	1.0000	4.7073	2.0625	4.9587
minimum soil depth fraction	1.0000	1.5050	3.8834	2.2830

Table 5.4: Comparison of calibrated parameters against the uncalibrated baseline configuration for the overall best scoring calibration trail and the parameters for the trails that had best score for Amy Darya Aral Sea inflow and the Syr Darya Aral Sea inflow.

5.3. PCR-GLOBWB2 experiments results

The experiments described in section 4.3.4 and summarized in table 4.3 form the basis of the results presented in this section. First, the historical baseline and calibration results are evaluated. Next, the impacts of the Karakum Canal experiments are examined, followed by an assessment of the GCM bias-correction. Finally, future hydrological projections under various climate models and scenarios are analyzed.

5.3.1. Baseline Performance and Historical Validation

Mean Annual Discharge

Mean annual discharge for the 1950-1960 period is summarised in table 5.6 for the two most downstream GRDC stations located near the Aral Sea inflow zone. The uncalibrated model shows a substantial positive bias at both locations. At Chatly on the Amu Darya, simulated discharge of 169.0 km km^3/year exceeds the historical reference of 39.0 km^3/year (Micklin et al., 2014) by 333%. At Kazalinsk on the Syr Darya, the overestimation is even larger, with simulated discharge of 133.1 km^3/year against a reference of 16.0 km^3/year, corresponding to an excess of 732%.

All three calibrated parameter sets substantially reduce this error, though performance varies by sub-basin. The Amu Darya calibrated set achieves the closest match at Chatly, with a simulated volume of 39.9 km^3/year and a error of only 2.4%. However, it overshoots the Kazalinsk reference by 74.4%, indicating poor transferability to the Syr Darya basin. Conversely, the Syr Darya calibrated set performs reasonably at Kazalinsk (31.5% underestimation) but strongly underestimates discharge at Chatly (66.1%), reducing simulated flow to just 13.2 km^3/year. The Overall calibrated set also has large error of 49.1% at Chatly, but is the closest at Kazalinsck with an error of 9.1%.

Calibration and Validation Performance

Monthly KGE scores for the calibration (1950–1954) and validation (1955-1959) periods are presented in table 5.5. The scores are consistent across both periods for all parameter sets, indicating that none of the configurations overfit to the calibration window.

The Amu Darya calibrated parameter set achieves the highest KGE at the Amu Darya river stations, with scores of 0.82 at Chatly and 0.34-0.36 at Kerki. This aligns with the near-zero volume bias reported in table 5.6. However, performance degrades significantly at the Syr Darya stations, with KGE values dropping to between -1.80 and -1.92 at Kazalinsk. This drop highlights that parameters optimized for one sub-basin do not transfer well to the other.

The Syr Darya calibrated set is the only configuration yielding a positive KGE at Kazalinsk (0.48-0.50), despite showing a larger volume bias (-31.5%) relative to the overall calibrated set (-9.1%). This suggests that while the Overall calibrated set accurately reproduces the long-term mean discharge magnitude at Kazalinsk, it misrepresents the temporal variability of the hydrograph.

The uncalibrated configuration performs poorly across all stations, yielding KGE values as low as -24 at Karaozek and massive volume overestimations exceeding 300% (table 5.6). This confirms that calibration

is an essential prerequisite for scenario analysis.

Table 5.5: Monthly KGE scores for selected stations for calibration (1950–1954) and validation (1955–1959) periods.

Station	Period	Uncalibrated	Overall cal.	Amu Darya cal.	Syr Darya cal.
Chatly	Calibration	−2.94	0.28	0.82	−0.15
Amu Darya	Validation	−2.66	0.25	0.82	−0.20
Kerki	Calibration	−0.02	0.13	0.34	−0.12
Amu Darya	Validation	−0.09	0.15	0.36	−0.05
Kazalinsk	Calibration	−15.41	0.06	−1.92	0.50
Syr Darya	Validation	−15.13	−0.08	−1.80	0.48
Karaozek	Calibration	−24.05	−1.09	−4.09	0.00
Syr Darya	Validation	−23.30	−1.34	−3.88	−0.24
Tyumen-aryk	Calibration	−7.70	0.57	−0.19	0.42
Syr Darya	Validation	−8.95	0.39	−0.52	0.45

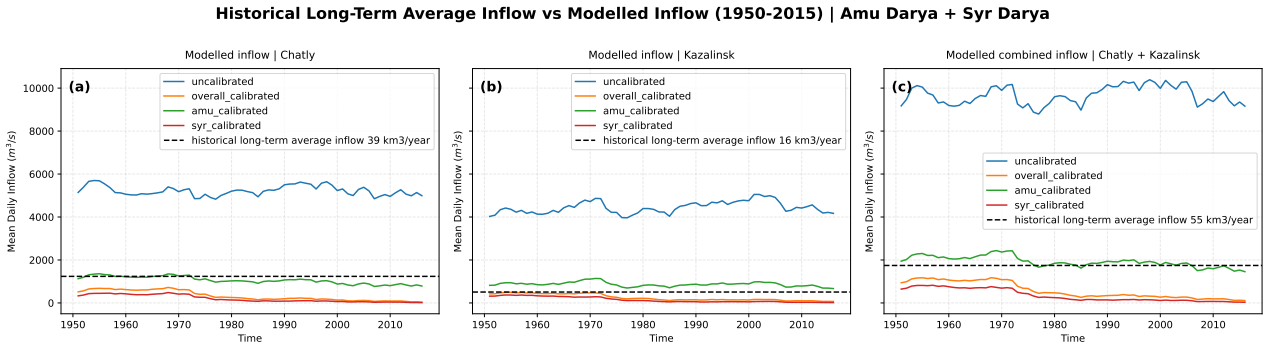


Figure 5.7: discharge for different parameter sets

5.3.2. Karakum Canal Experiments

The Karakum Canal experiments modelling the extra water taken from the Amu Darya are described in sections 4.3.3 and 4.3.4. The simulation logs indicate successful completion of the corresponding simulation for 1990-2020, however, three output files were expected, but only two are present, and both are incomplete. This suggests instability in the simulation method or output generation process. The analysis is therefore limited to the 1962-1990 period. Over this interval, the single-cell configuration achieves an abstraction efficiency of approximately 80% under low target volumes, decreasing to roughly 40% as the target approaches the peak threshold of 12 km^3/year reported by Aus Der Beek et al. (2011) and Touge et al. (2024). Distributing extraction across multiple grid cells raises efficiency to approximately 85%, while the seasonal wave configuration achieves approximately 95% of the target volume (figure 5.8).

The cumulative impact of each configuration on Aral Sea inflow up to 1980 is summarized in table 5.7. All three configurations reduce total inflow relative to the reference run, with deficits ranging from 131.70 km^3/year (11.3%) for the single-cell configuration to 192.47 km^3/year (16.5%) for the seasonal wave configuration.

5.3.3. Effect of Climate Bias Correction

The impact of meteorological bias correction on simulated discharge was evaluated against the ERA5-driven reference run for four CMIP6 GCMs (CanESM5, MIROC6, MPI-ESM1-2-HR, and MRI-ESM2-0). Model performance is summarised using KGE on daily discharge (KGE_Q), RMSE on cumulative discharge ($RMSE_\Sigma$), and cumulative discharge volumes at 1980 and 2010 (table 5.8). The results are also visualized in figures 5.9 and 5.10.

Overall, bias correction improved model performance across all evaluation metrics and stations. Improvements were most consistent in the long-term water balance representation, as indicated by substantial reduc-

Table 5.6: Comparison of historical mean discharge (1950–1960) against reference baseline values (Micklin et al., 2014) for the Chatly (Amu Darya) and Kazalinsk (Syr Darya) stations across different calibration parameter sets.

Station	Parameter Set	Simulated Discharge		Reference	Difference	
		($m^3\,s^{-1}$)	($km^3\,year^{-1}$)	($km^3\,year^{-1}$)	($km^3\,year^{-1}$)	(%)
Chatly	Uncalibrated	5358.8	169.0	39.0	130.0	333.3
	Overall Calibrated	630.1	19.9	39.0	−19.1	−49.1
	Amu Calibrated	1266.1	39.9	39.0	0.9	2.4
	Syr Calibrated	419.5	13.2	39.0	−25.8	−66.1
Kazalinsk	Uncalibrated	4220.7	133.1	16.0	117.1	731.9
	Overall Calibrated	461.2	14.5	16.0	−1.5	−9.1
	Amu Calibrated	884.9	27.9	16.0	11.9	74.4
	Syr Calibrated	347.4	11.0	16.0	−5.0	−31.5

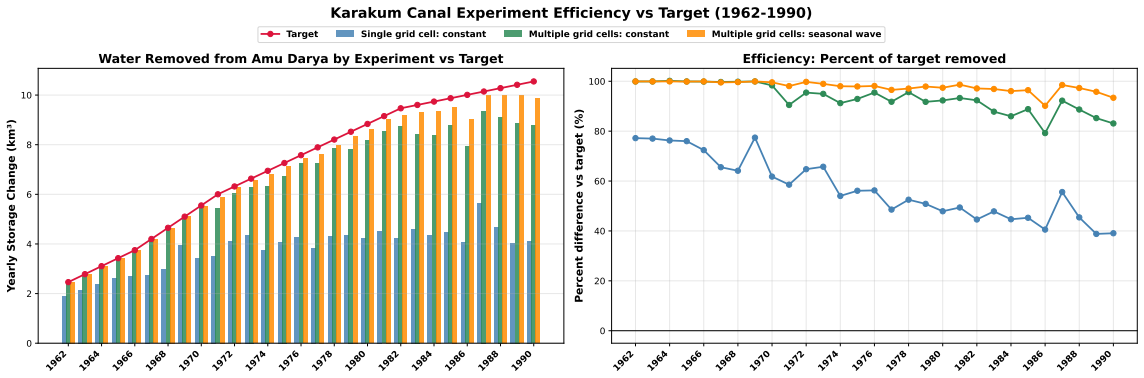


Figure 5.8: Karakum Canal abstraction results (1962–1990): (a) simulated yearly storage change for the single-cell, multi-cell, and seasonal-wave configurations against the target demand; (b) corresponding abstraction efficiency as percentage of target.

Table 5.7: Cumulative Amu Darya storage and volumetric deficit relative to the calibrated reference run (1166.49 km^3/year) across Karakum Canal experiment configurations.

Experiment Configuration	Total Storage (km^3/year)	Difference (km^3/year)	Difference (%)
Reference (Amu Darya Calibration)	1166.49	—	—
Single grid cell: constant	1034.79	−131.70	−11.29
Multiple grid cells: constant	981.01	−185.48	−15.90
Multiple grid cells: seasonal wave	974.03	−192.47	−16.50

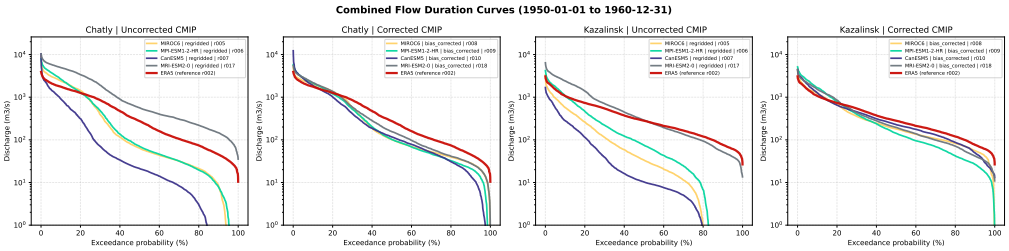


Figure 5.9: Flow Duration Curves for uncorrected and corrected data. The FDCs show that not only the long term statistics improve but also that the distribution of high and low flows aligns much more to the refence ERA5 simulation

tions in RMSE_{Σ} and marked convergence of cumulative discharge trajectories towards the ERA5 reference.

At the daily scale, KGE_Q increased for most model-station combinations, with the largest gains occurring for models that initially exhibited poor discharge dynamics (notably MRI-ESM2-0 at Chatly and MIROC6 at Kazalinsk). One exception was MPI-ESM1-2-HR at Kazalinsk, where a slight decrease in KGE_Q was observed, although this did not translate into degraded long-term performance.

In terms of cumulative discharge, bias correction substantially reduces volumetric errors in nearly all cases, with percentage deviations from the ERA5 reference decreasing the error by orders of magnitude down to single digits for most model-station combinations. A notable exception is MIROC6 at Chatly, where ΣQ_{1980} moves further from the ERA5 reference after bias correction ($+3.4\%$ uncorrected versus -10.3% bias-corrected), suggesting that bias correction does not guarantee improvement in all cases.

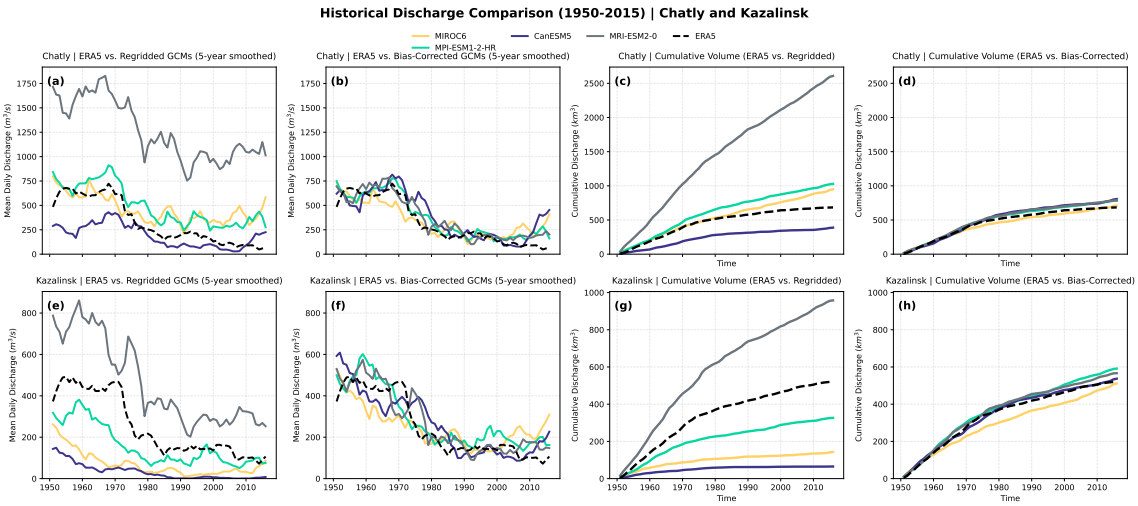


Figure 5.10: Figure showing the results of modelling with uncorrected and corrected data for the two downstream stations. Both smoothed mean discharge and cumulative discharge is shown.

Table 5.8: Comparison of uncorrected and bias-corrected meteorological forcing against the ERA5-driven reference run. Performance is evaluated using KGE on daily discharge (KGE_Q), RMSE on cumulative discharge ($RMSE_\Sigma$), and cumulative discharge volumes at 1980 and 2010, and their percentage difference relative to the ERA5-driven reference run. All runs use the overall-calibrated parameter set. Bias correction yields substantial improvements across nearly all models and metrics at both stations: KGE_Q increases markedly (e.g. MRI-ESM2-0 at Chatly from -2.49 to $+0.50$), cumulative volume errors are reduced by up to an order of magnitude, and percentage deviations from the ERA5 reference decrease by orders of magnitude down to single digits in most cases.

	KGE_Q		$RMSE_\Sigma$ (km^3)		ΣQ_{1980} (km^3)				ΣQ_{2010} (km^3)			
	Uncorr.	BC	Uncorr.	BC	Uncorr.	(Δ%)	BC	(Δ%)	Uncorr.	(Δ%)	BC	(Δ%)
Chatly (*Amu Darya*)												
ERA5	—	—	—	—	527.8	—	527.8	—	677.5	—	677.5	—
CanESM5	0.311	0.310	239.991	60.021	287.2	−45.6	593.1	12.4	362.9	−46.4	756.1	11.6
MIROC6	0.110	0.531	105.677	36.349	545.7	3.4	473.3	−10.3	895.1	32.1	674.9	−0.4
MPI-ESM1-2-HR	−0.086	0.424	183.867	47.570	678.8	28.6	574.0	8.7	971.7	43.4	747.5	10.3
MRI-ESM2-0	−2.485	0.495	1148.634	55.411	1490.2	182.3	565.0	7.0	2477.7	265.7	762.3	12.5
Kazalinsk (*Syr Darya*)												
ERA5	—	—	—	—	379.6	—	379.6	—	512.4	—	512.4	—
CanESM5	−0.209	0.530	320.620	12.152	59.6	−84.3	386.6	1.8	64.5	−87.4	513.4	0.2
MIROC6	0.003	0.569	269.972	48.658	107.4	−71.7	309.5	−18.5	135.2	−73.6	479.3	−6.5
MPI-ESM1-2-HR	0.453	0.333	142.909	32.618	229.1	−39.7	392.5	3.4	312.7	−39.0	564.4	10.2
MRI-ESM2-0	−0.378	0.499	279.029	25.689	629.8	65.9	397.8	4.8	924.7	80.5	548.7	7.1

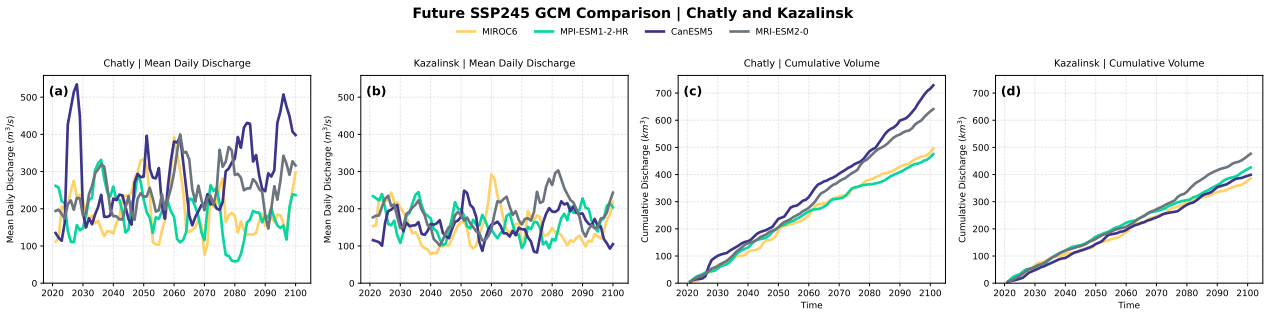


Figure 5.11: Result of climate model ensemble experiment where for different CMIP6 models the discharge for 2015-2099 is modelled. For all models SSP2-4.5 was used. The results show that For the Amu Darya the spread is much larger compared to Syr Darya.

5.3.4. Future Projections: Climate Model Uncertainty

Simulated discharge over the 2020-2099 period was compared across four GCMs (CanESM5, MIROC6, MPI-ESM1-2-HR, and MRI-ESM2-0) forced under SSP2-4.5 with bias-corrected meteorological input (Figure 5.11, Table 5.9). At Chatly, early-period mean discharge (2020-2040) ranged from 6.39 km^3/year (MIROC6) to 7.90 km^3/year (CanESM5), with total cumulative discharge spanning 148.5 km^3/year (MPI-ESM1-2-HR) to 226.6 km^3/year (CanESM5). A clear directional split was observed between models: CanESM5 and MRI-ESM2-0 projected increasing discharge towards the end of the century, with late-period means of 10.96 and 8.56 km^3/year respectively, representing increases of 38.4% and 17.5% relative to the early period. In contrast, MIROC6 and MPI-ESM1-2-HR projected declining discharge, with late-period means of 5.29 and 4.40 km^3/year (-7.4% and -36.8%, respectively). At Kazalinsk, the same directional split was present, though absolute differences were smaller, with cumulative discharge ranging from 120.0 km^3/year (MIROC6) to 149.0 km^3/year (MRI-ESM2-0).

Table 5.9: Model uncertainty in future discharge projections for the Chatly (Amu Darya) and Kazalinsk (Syr Darya) stations under SSP2-4.5 with bias-corrected forcing and the overall-calibrated parameter set. $\bar{Q}_{2020–2040}$ and $\bar{Q}_{2075–2095}$ are mean discharges for the early and late projection periods respectively; ΔQ (%) is the relative change between the two periods; ΣQ is the total cumulative discharge over 2020–2099.

Model	*Chatly (Amu Darya)*				*Kazalinsk (Syr Darya)*			
	$\bar{Q}_{2020–2040}$ ($m^3\,s^{-1}$)	$\bar{Q}_{2075–2095}$ ($m^3\,s^{-1}$)	ΔQ (%)	ΣQ (km^3)	$\bar{Q}_{2020–2040}$ ($m^3\,s^{-1}$)	$\bar{Q}_{2075–2095}$ ($m^3\,s^{-1}$)	ΔQ (%)	ΣQ (km^3)
CanESM5	250.1	346.1	38.4	719.9	153.8	182.4	18.6	396.6
MIROC6	180.2	166.9	−7.4	487.8	165.9	125.1	−24.6	380.0
MPI-ESM1-2-HR	219.7	138.8	−36.8	471.1	192.1	162.0	−15.7	422.7
MRI-ESM2-0	229.8	270.0	17.5	636.7	188.0	212.7	13.1	471.9

5.3.5. Future Projections: Emission Scenarios

The sensitivity of simulated discharge to emission scenario was assessed using MPI-ESM1-2-HR across four SSPs (Figure 5.12, Table 5.10). At Chatly, early-period mean discharge ranged from 5.62 km^3/year (SSP5-8.5) to 8.48 km^3/year (SSP3-7.0), with all scenarios projecting a decline towards the end of the century. The largest relative declines were simulated under SSP2-4.5 and SSP3-7.0 (-36.8% and -36.3%, respectively), while SSP1-2.6 produced the most moderate reduction (-11.6%). Total cumulative discharge at Chatly ranged from 133.0 km^3/year (SSP5-8.5) to 178.3 km^3/year (SSP3-7.0). At Kazalinsk, scenario sensitivity was considerably weaker, with cumulative discharge ranging from 129.5 km^3/year to 136.4 km^3/year across all four scenarios and late-period means remaining broadly stable relative to the early period.

5.3.6. Model Spread versus Scenario Spread

Across both experiments, structural uncertainty arising from GCM choice exceeded scenario uncertainty in terms of projected cumulative discharge. For the Amu Darya, at station Chatly , the spread for all four CMIP6 models using SSP2-4.5 was 248.8 km^3, compared to a spread of 143.2 km^3 for all four SSP scenarios for MPI-ESM1-2-HR. At Kazalinsk, the corresponding ranges were 91.9 and 22.2 km^3, respectively. Notably, the addition of MRI-ESM2-0 to the ensemble more than doubled the inter-model spread at Kazalinsk relative to the

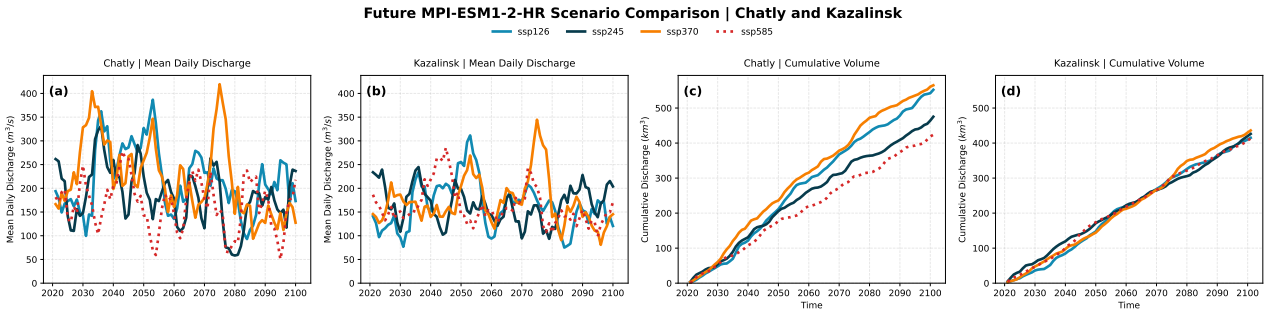


Figure 5.12: Result of climate model ensemble experiment where for one CMIP6 model the discharge for 2015-2099 is modelled, using 4 different SSP scenarios. SSP5-8.5 is dashed this scenario is nog longer considered feasible in CMIP7. The results show that For the Amu Darya the spread is much larger compared to Syr Darya.

Table 5.10: Scenario uncertainty in future discharge projections for the Chatly (Amu Darya) and Kazalinsk (Syr Darya) stations using MPI-ESM1-2-HR with bias-corrected forcing and the overall-calibrated parameter set. $\bar{Q}_{2020–2040}$ and $\bar{Q}_{2075–2095}$ are mean discharges for the early and late projection periods respectively; ΔQ (%) is the relative change between the two periods; ΣQ is the total cumulative discharge over 2020–2099.

	Chatly (Amu Darya)				*Kazalinsk (Syr Darya)*			
Scenario	$\bar{Q}_{2020–2040}$ ($m^3\,s^{-1}$)	$\bar{Q}_{2075–2095}$ ($m^3\,s^{-1}$)	ΔQ (%)	ΣQ (km^3)	$\bar{Q}_{2020–2040}$ ($m^3\,s^{-1}$)	$\bar{Q}_{2075–2095}$ ($m^3\,s^{-1}$)	ΔQ (%)	ΣQ (km^3)
SSP1-2.6	206.2	182.3	−11.6	543.5	141.4	140.1	−0.9	412.2
SSP2-4.5	219.7	138.8	−36.8	471.1	192.1	162.0	−15.7	422.7
SSP3-7.0	267.8	170.7	−36.3	564.8	157.0	161.5	2.9	432.3
SSP5-8.5	177.6	140.4	−20.9	421.6	165.7	131.7	−20.5	410.1

three-model ensemble, highlighting the sensitivity of uncertainty estimates to ensemble composition. In both experiments, projections for Kazalinsk remained more constrained than those for Chatly, suggesting that cumulative inflow from the Syr Darya is less sensitive to both scenario and model uncertainty than that from the Amu Darya.

5.4. Aral Sea Dynamics

5.4.1. Aral Sea model validation

Table 5.11 and figure 5.13 show the results of the validation method of the Aral sea model described in section 4.4.2. The results demonstrate that for the 59 and 61 km^3/year scenarios, simulated water levels remain within 0.3–0.5 m of the pre-diversion baseline of ∼53 m, and volumes stay within 5% of ∼1,060 km^3, confirming that the model reproduces pre-diversion equilibrium conditions under historically plausible inflows. The more conservative figure of km^3/year also leads to plausible results with a equilibrium at approximately 52 m elevation and a volume still within the 5% of the estimated equilibrum volume for the period up to 1990. The subsequent decrease will be discussed in the discussion section.

Table 5.11: Simulated Aral Sea water level (m) and volume (km^3) under selected validation cobmined river-groundwater inflow scenarios. The 1800-1960 observed pre-expansion baseline is approximately 53 m and 1,060 km^3.

	57		**59**		**61**		**63**	
Year	**m**	**km^3**	**m**	**km^3**	**m**	**km^3**	**m**	**km^3**
1950	52.87	1063.9	52.87	1064.0	52.87	1064.2	52.87	1064.3
1960	52.34	1029.9	52.60	1045.4	52.83	1060.8	53.03	1076.2
1970	52.24	1022.8	52.63	1046.9	52.96	1071.2	53.31	1095.0
1980	52.15	1016.6	52.61	1045.7	53.02	1074.9	53.42	1102.6
1990	52.13	1015.3	52.63	1047.1	53.08	1079.2	53.51	1108.6
2000	52.03	1008.8	52.54	1042.0	53.03	1075.5	53.46	1105.5
2014	51.60	982.9	52.12	1015.1	52.65	1048.9	53.09	1080.1

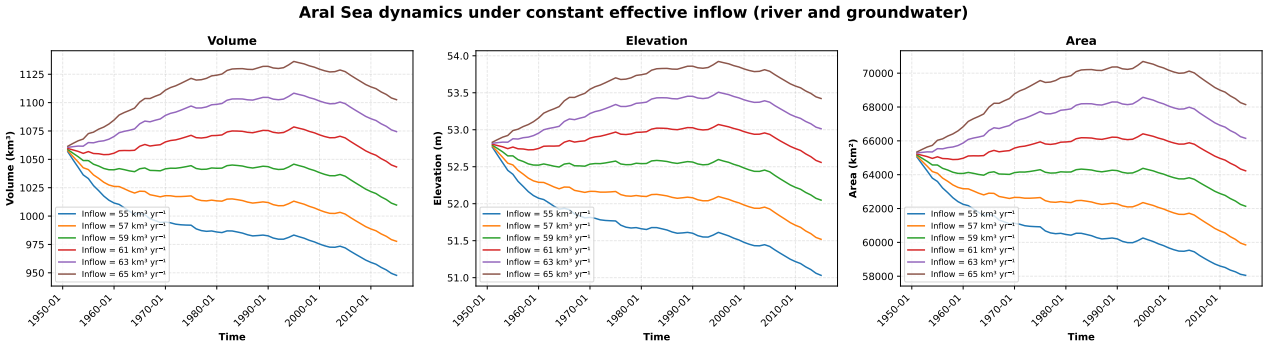
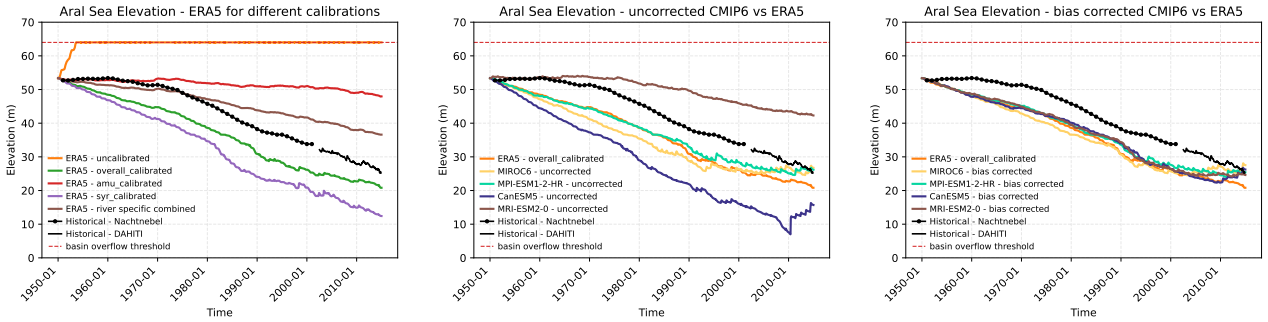


Figure 5.13: Simulated Aral Sea water level (m.a.s.l.) area (km^2) and volume (km^3) under selected validation cobmined river-groundwater inflow scenarios. The 1800-1960 observed pre-expansion baseline is approximately 53 m and 1,060 km^3.

5.4.2. Aral Sea Response to PCR-GLOBWB2 Simulated Inflows

By integrating the simulated river discharge from the PCR-GLOBWB2 experiments described in section 4.3.4 and presented in section 5.3, this section evaluates the downstream impacts on the Aral Sea.



(a) (b) (c)

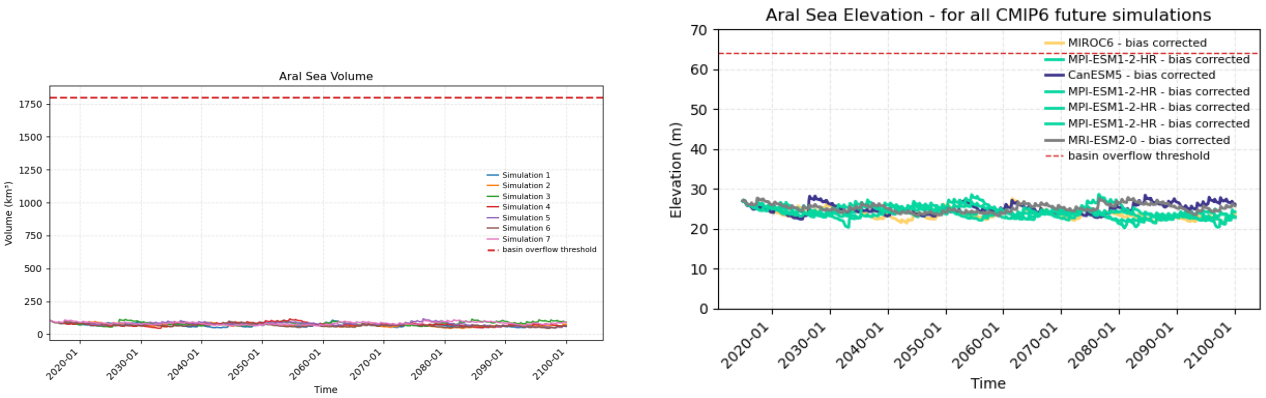
Figure 5.14: Panels a, b and c show the results of using the discharge from the PCR-GLOBWB2 experiments in modelling the Aral Sea lake levels.

When using PCR-GLOBWB2 modelled outputs as input to the Aral Sea model, most experiments show a much more severe drying trend than observed in reality (figure 5.14). First, the uncalibrated setup rapidly fills the Aral Sea within approximately five years and even leads to an unrealistic overflow towards the Caspian Sea. The Amu Darya–calibrated configuration follows the observed trajectory reasonably well up to around 1965, but fails to reproduce the subsequent strong drying trend. In contrast, both the fully calibrated and Syr Darya–calibrated configurations already show substantial drying during the so-called "stable period," which is not consistent with observations.

When using uncorrected CMIP6 scenarios with the same calibration configuration, the spread between models is large. CanESM5 in particular shows an extreme drying signal, eventually reaching near-complete desiccation of the lake and exhibiting numerical artefacts due to interpolation over an almost dry lake bed. Other models also show immediate drying during the so-called stable period, with the exception of MRI-ESM2-0, which, due to a strong precipitation bias, is the only model that approximately reproduces the observed stable-period behaviour. It is important to note that all simulations in this comparison use the same overall best-calibrated hydrological setup, which itself already shows structural deficiencies, as discussed earlier.

When applying bias-corrected inputs, the model outputs converge again towards the behaviour of the overall best-calibrated configuration, with all scenarios exhibiting drying during the stable period. Interestingly, despite differences in early-time inflow, all simulations converge to nearly the same lake level by 2015. This suggests that although initial modelled river discharges are lower than observed, the long-term integrated water balance is actually much more equal than expected.

When running all CMIP6 future scenarios (four climate models and, for one model, four SSPs), all simulations



(**a**) modelled volume

(**b**) modelled elevation

Figure 5.15: Aral Sea model results for CMIP6 future climate scenarios. Four climate models, each with SSP2-4.5 and four SSP scenarios for MPI-ESM1-2-HR. Across all experiments, lake levels remain approximately stable, and no significant refilling of the Aral Sea is simulated.

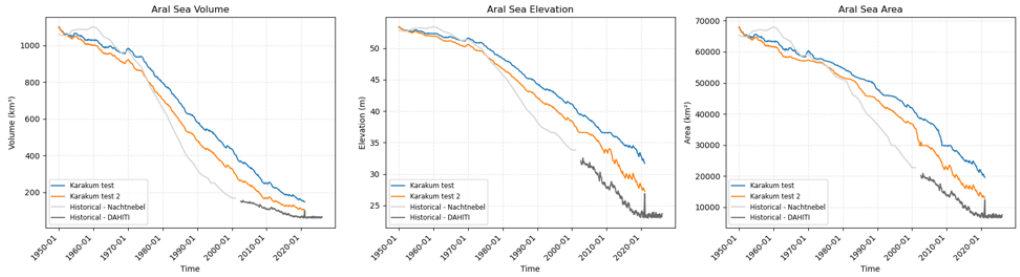


Figure 5.16: Artificial Karakum Experiment

show more or less constant lake levels and no refilling and stay close to the level of the current day. (figure 5.15)

Unfortunately, because the Karakum Canal experiments failed, a solid analysis could not be carried out. However, during testing of the Aral Sea model, an alternative experiment was conducted. By combining the best-performing ERA5 inflows for the Amu Darya and Syr Darya with an artificial river using negative discharges, to represent different estimated diversions for the Karakum Canal, the results closest to real-world values were obtained, as shown in figure 5.16.

5.5. Workflow as a Result

An important outcome of this research is the development of a reproducible and modular modeling workflow implemented using *Snakemake*. This workflow enables transparent experimentation and can be seamlessly integrated and reused within the broader *eWaterCycle* ecosystem. Several components were developed as reusable modules, including:

- **Shapefile-based forcing generation:** A tool compatible with PCR-GLOBWB2 that automatically determines the required spatial latitude and longitude dimensions based on a provided shapefile (`.shp`).
- **Forcing regridding module:** A module designed for the spatial harmonization of diverse datasets into grids compatible with ERA5 and PCR-GLOBWB2.
- **Forcing bias-correction module:** A component developed to improve consistency with historical observations and ERA5 pseudo-observations.
- **PCR-GLOBWB calibration workflow:** A workflow enabling model parameter calibration on the basis of GRDC observations.
- **Scenario-based automation:** A robust framework tailored for running large-scale ensembles of concurrent PCR-GLOBWB instances across different simulation scenarios.

- **Visualization and analysis helpers:** Built-in utilities to facilitate the visualization and evaluation of driving environmental forcing fields.

5.5.1. Supporting Tools

Outside of the primary *Snakemake* workflow, additional independent supporting tools were developed to assist with data pre-processing and exploratory analysis. These include:

- **Köppen-Geiger visualizer and analyzer:** A tool dedicated to classifying and visualizing regional Köppen - Geiger climate zones based on an input shapefile. This tool is already being used in ongoing research projects within the eWaterCycle project (Rolf Hut & Mark Melotto, 2026)
- **NASA visualization tool:** A utility designed to fetch, process, and display NASA earth observation satellite data for specific geographic regions defined by a shapefile.
- all kinds of exploratory notebooks that can potentially be (partially) reused or repurposed by others or turned into more self contained scripts.

5.5.2. Functional but unused components

The following components were fully developed and verified to be functional, but were ultimately not used in the final workflow:

- **Automated map generation script:** A utility that automatically constructs a PCR-GLOBWB-compatible `clonemap` and `landmask`, determining the valid computational model domain directly from an arbitrary shapefile.
- **HBV optimization workflows:** Various calibration and optimization setups specifically tailored for the HBV hydrological model.

6

Discussion

This study developed and applied a reproducible modelling framework to reconstruct the historical evolution of the Aral Sea, disentangle climatic and anthropogenic drivers of change, and explore potential future trajectories under climate scenarios. The results demonstrate that the coupled modelling framework is capable of reproducing the overall desiccation trend of the Aral Sea, while revealing important limitations related to hydrological forcing, model structure, and representation of human water use.

This chapter interprets the obtained results in relation to the research objectives. First, the simulated system behaviour is evaluated along the modelling chain from forcing generation to hydrological simulation and Aral Sea dynamics. Subsequently, methodological aspects of the modelling framework are assessed, followed by a discussion of uncertainties, limitations, and directions for future research.

6.1. Interpretation of Results

The modelling workflow consists of a sequential chain of forcing generation, hydrological simulation using PCR-GLOBWB, and Aral Sea water balance modelling (Section 1.4). Evaluating model performance therefore requires interpreting results at each stage of this chain, as uncertainties propagate downstream and influence the simulated lake response.

6.1.1. Forcing generation

The forcing generation workflow performs robustly and provides consistent climate forcing for subsequent modelling stages. CMIP historical and future climate projections were statistically downscaled to the ERA5 spatial resolution, resulting in well-aligned grids that enabled direct integration with the hydrological model. The downscaling procedure introduced only a small area-averaged error (<1%), indicating that interpolation artefacts remain limited at the large-basin scale.

However, this downscaling error estimate was evaluated over the entire modelling domain (i.e. the full square grid) rather than specifically over the Aral Sea catchments, where local deviations may be larger and be masked by domain wide averages. Downscaling may introduce artificial gradients due to interpolation across grid cells that do not fully represent regional climate variability, for example for mountain ranges adjacent to lowlands.

Subsequently, bias correction was applied using ERA5 as a pseudo-observational reference dataset. This approach successfully reduced systematic temperature differences between CMIP climate forcing data and the ERA5 reference forcing conditions. However, precipitation biases remained partly unresolved after correction.

The four CMIP6 models exhibit a striking range of disagreement over the study region, with long-term average temperature differences of up to 4.4 °C and precipitation spreads exceeding 50% between models. This wide inter-model spread is itself an informative result, reflecting genuine structural uncertainty in how CMIP6 models represent Central Asian climate. Without pre-defined exclusion criteria this risks cherry-picking, selectively removing models that produce inconvenient results.

Precipitation and temperature were bias-corrected separately, which preserves their individual distributions but does not take into account potential dependencies between variables. While this approach appears to be sufficient for long-term mean hydrological simulations, it may not fully preserve joint variability relevant for compound or extreme events, where interactions between precipitation and temperature are important. As this study primarily focuses on mean and seasonal discharge and water balance closure, this limitation is not expected to strongly affect the main conclusions, but it may be relevant in work targeting hydrological extremes or drought characteristics.

The models included in this study were selected based on practical considerations rather than performance metrics. Specifically, models were required to provide a sufficient number of SSP scenarios, represent a geographically diverse set of modelling centres (including centres from three different continents), and be compatible with the PCR-GLOBWB workflow. Following the resolution of a leap-year-related issue in the processing chain, an additional model was included to evaluate the robustness of the results.

An additional practical consideration in model selection was data availability. Although many CMIP6 models are nominally accessible through the ESGF infrastructure, in practice some datasets were unavailable, incomplete, or difficult to retrieve reliably using the available eWaterCycle tooling. For example, I was unable to obtain a complete set of required simulations for EC-Earth3 despite repeated attempts. Consequently, model selection was constrained not only by scientific considerations, but also by the availability of datasets that could be successfully downloaded, processed, and integrated into the PCR-GLOBWB workflow.

Given that this study serves both as a proof of concept and as a case study, this pragmatic model selection approach was considered appropriate. The objective was not to identify a subset of "best-performing" climate models, but rather to explore the range of plausible outcomes and demonstrate the applicability of the modelling framework.

Because ERA5 serves as the statistical reference, any deficiencies in the ERA5 records are implicitly transferred to the corrected CMIP forcing. In particular, ERA5 precipitation values prior to approximately 1950 appear on visual inspection unrealistically low compared to 1950-2020, suggesting reduced reliability of early reanalysis forcing. As precipitation is the dominant control on runoff generation in PCR-GLOBWB, uncertainties originating from the reference climatology likely propagate through the hydrological simulations and influence modeled discharge and Aral Sea dynamics. Restricting the bias-correction period to a more reliable ERA5 interval (e.g., post-1950) could improve the bias correction results for CMIP.

6.1.2. Hydrological model performance

Calibration

Calibration significantly improved model performance compared to the uncalibrated baseline, primarily by correcting a large precipitation overestimation. This improvement is also visible in the difference between Kling–Gupta Efficiency (KGE) the uncalibrated configuration to the three different calibrated configurations. The values go from strongly negative to near-zero and positive for most stations for at least one calibrated configuration, tables 5.5 and G.1, however these improvements are uneven across basins and configurations, and indicates a clear trade off between performance for the Amu Darya and Syr Darya systems. The overall best performing calibration for the score defined in equation (4.7) also optimises for the Syr Darya while the Amu Darya is more important for the lake levels of the Aral Sea. Several factors may explain this behavior.

The calibration objective combined KGE and NSE scores across both basins into a single metric, without explicitly penalizing errors in cumulative volumes. Given the substantially higher discharges in the Amu Darya, this may have caused the optimization to implicitly favor the Syr Darya, where dimensionless metrics normalize discharge magnitudes, reducing the weight of absolute volumetric errors and thereby favoring basins with lower flows in the aggregated objective function. The spatial distribution and number of GRDC gauging stations used may have compounded this effect, further tilting the algorithm toward the Syr Darya.

A multi-objective optimization approach, where basin or specific individual station scores are optimized simultaneously rather than expressed as a single value, would make these trade-offs explicit, exposing the Pareto front of solutions and allowing an informed choice between basin-specific and balanced parameter sets.

While all station-level performance metrics were retained for all calibration runs, they were not incorporated into the optimization algorithm itself beyond their combined objective score. The subsequent identification of

basin-specific best-performing parameter sets therefore is a post-calibration approximation of multi-objective behavior and not a formal Pareto optimization.

The parameter space explored was also limited to PCR-GLOBWB2 calibration parameters reported in the literature. Exposing additional model parameters, such as scaling factors for agricultural land fractions, could expand the solution space. The mutual information analysis further revealed that some parameters had near-zero influence on model output. Fixing these and redirecting calibration effort toward more sensitive or unexplored parameters could improve results.

The validation results confirm that the calibration outcomes are robust over time. KGE values remain broadly the same between calibration and validation periods, indicating limited overfitting and suggesting temporal stability of the calibrated parameter sets. However, basin-specific calibrations consistently lead to worse outcomes when transferred between basins, reinforcing the conclusion that the Amu Darya and Syr Darya exhibit distinct hydrological regimes that cannot be simultaneously captured well with a single shared parameterisation. Of note is that the overall-calibrated parameter set outperforms the Syr Darya-specific calibration at the downstream Syr Darya station in terms of cumulative discharges. This may indicate that the station-specific calibration is more sensitive to the limited calibration period and potentially overfits to local conditions, whereas the overall calibration provides a more robust representation of basin-scale behaviour. However, the underlying cause of this difference remains unclear and warrants further investigation.

bias corrected CMIP forcing in PCR-GLOBWB2

Bias correction of the CMIP6 forcing had a moderate effect on daily discharge performance but a more pronounced effect on cumulative volumes. This is consistent with the inherent nature of the two forcing types. ERA5 is a reanalysis product constrained by historical observations, meaning it approximates the actual sequence of meteorological events. CMIP6 models, by contrast, simulate the climate system freely and are not constrained to reproduce observed weather, they may capture the statistics and long-term trends of the climate, but whether a given model simulates precipitation on a specific date bears no necessary relation to what was observed.

As such, high KGE values for daily discharge should not be expected when using CMIP6 forcing, even after bias correction: if the model places a wet spell in November rather than October, the day-to-day phasing will be poor regardless of how well the precipitation totals are adjusted. Bias correction can address systematic biases in mean volumes or seasonal distributions, and this is reflected in the improved cumulative discharge agreement, but it cannot reproduce the same temporal response because none exists due to the underlying forcing. The appropriate benchmark for CMIP6-forced runs is therefore long-term statistics and volumetric trends rather than event-scale reproduction, consistent with the results in section 5.3.3

These effects are reflected in the results (table 5.8) where overall KGE improvement is shown, but does not approach the optimal value of 1. The Root Mean Square Error of cumulative discharge ($RMSE_{\Sigma}$)is improved significantly, showing a much larger alignment of cumulative discharge. This indicates that bias correcting is particularly good at eliminating systemic errors.

The flow duration curves (figure 5.9) suggest that the overall structure of the discharge distributions is reasonably well preserved after bias correction. However, this similarity has not been quantitatively evaluated and is only supported by visual inspection. To confirm that the underlying flow regimes are accurately modeled a more extensive assessment would be required.

A known limitation of bias correction is the implicit assumption that the bias estimated over the control period remains stationary into the future. If a model is anomalously dry during the control period but behaves more normally under future forcing, the correction will over-correct, artificially inflating future precipitation and discharge. This breakdown of the stationary assumption may explain the behaviour of CanESM5, which exhibited the largest dry bias during the control period yet produced by far the highest discharges across future scenarios. In this case, bias correction may have compounded rather than reduced errors in the future period, and results from strongly bias-corrected models should therefore be interpreted with caution.

Climate model versus scenario uncertainty

The future discharge projections reveal a clear hierarchy of uncertainty sources, with structural differences between climate models exerting a stronger influence on simulated river discharge than differences between emission scenarios. Across the SSP2-4.5 ensemble, inter-model spread in cumulative discharge substantially exceeds the range induced by alternative emission pathways, particularly at the Amu Darya (Chatly) station.

In contrast, scenario uncertainty derived from a single GCM (MPI-ESM1-2-HR) is comparatively limited. While emission pathways introduce systematic differences in long-term trends, the overall magnitude of variation remains smaller than the spread induced by switching between climate models. This suggests that, for this area and study, hydrological projections are more sensitive to structural climate model uncertainty than to emissions uncertainty.

The Amu Darya system appears more responsive to differences in climate forcing, resulting in larger inter-model spread, whereas the Syr Darya shows comparatively constrained variability. One potential contributing factor is the strong topographic contrast within the basins, where mountainous headwaters may enhance sensitivity to precipitation variability and model-dependent differences in orographic precipitation. In addition, differences between models in the representation of regional circulation patterns, such as large-scale moisture transport into Central Asia, may alter the spatial distribution of precipitation across the two basins. However, these mechanisms are not explicitly evaluated in this study and therefore remain highly speculative.

Overall, these results highlight that uncertainty in future discharge projections is primarily driven by GCM structural differences rather than emission scenarios, and that this effect is strongly basin-dependent. For impact assessments in endorheic basins such as the Aral Sea system, this implies that ensemble selection and climate model diversity are more critical than scenario selection when characterising projection uncertainty.

6.1.3. Aral Sea Dynamics

The model demonstrates a strong capability to replicate real-world behavior under plausible discharge scenarios as shown during model validation. The primary challenge stemmed from PCR-GLOBWB yielding significantly underestimated inflows, which limited the depth of a definitive analysis. Nevertheless, general trends can still be identified. For relatively stable flow regimes of the Amu Darya, the artificial Karakum Canal experiment yielded simulated flows that most closely matched observations. The fact that that was needed indicates the importance of the human factor.

The future CMIP6 projection scenarios (figure 5.15) provide an insight into the future trajectory of the basin. Across four distinct climate models and multiple Shared Socioeconomic Pathways (SSPs), the simulated lake levels flatline, showing more or less constant modern-day volumes with zero significant refilling. The fact that the Aral Sea remains locked in its current state across a wide range of future temperature and precipitation projections demonstrates that natural climate variations alone are insufficient to trigger recovery. This lack of sensitivity to climate scenarios strongly indicates that anthropogenic water management, rather than shifting climate regimes, dictates the success or failure of regional refilling strategies.

6.2. Methodological Evaluation

This section shifts the focus from the interpretation of results to a methodological evaluation of the modelling framework itself. While the previous sections discussed the behaviour of the model outputs and their implications for the Aral Sea system, the following discussion reflects on how these results are produced. In particular, it examines the calibration strategy, workflow design, data dependencies, and sources of structural uncertainty that influence the robustness and reproducibility of the findings. The aim is not to reinterpret the results, but to critically assess the methods that generated them and to identify limitations and potential improvements for future work.

6.2.1. Workflow Orchestration and FAIR Principles

The modelling framework developed in this study was designed to be FAIR: findable, accessible, interoperable, and reusable. The integration of Snakemake as a workflow management tool is part of this. By managing the full modelling pipeline, from forcing preprocessing and bias correction through to PCR-GLOBWB2 simulation and Aral Sea volume modelling, the framework makes clear exactly which steps were run, in which order, and on which data while automatically handling parallel execution across available CPU cores.

Snakemake was integrated into the workflow at a relatively late stage of the project. Although the forcing preprocessing, spatial downscaling, bias correction, and PCR-GLOBWB2 simulations are fully orchestrated within the rule system at the time of writing, earlier adoption would have ensured complete coverage of all pipeline steps from the outset and reduced the need to retroactively integrate or refactor existing scripts. Future workflows of this type would benefit from establishing structure before any substantial scripting begins.

Because Snakemake monitors file integrity using input/output timestamps, modifications to any upstream

dataset or configuration parameters can automatically trigger downstream reruns. This introduces a practical risk for computationally expensive steps such as multi-decadal PCR-GLOBWB2 simulations. To mitigate unintended reruns, operational 'firewalls' were implemented within the rule dependencies to decouple preprocessing from core simulation blocks. Although this barrier structure remains somewhat fragile under default parsing conditions, it can be reliably managed through disciplined use of Snakemake's dry-run functionality.

"The major practical advantage of defining all rules, dependencies, and data flows upfront is that adding a new climate model or scenario requires only a single change in the settings file. Snakemake then automatically generates the forcing, preprocesses it, and runs it through PCR-GLOBWB2 without further intervention.

Splitting the workflow into Snakemake rules, with standardised `.nc` files saved between each stage, means that intermediate results are always available for inspection and individual components can be rerun or reused independently. Each rule is designed to do one thing only, which prevents the accumulation of long, hard-to-maintain scripts. This does require some familiarity with how Snakemake structures its rules and dependencies, which may present an initial learning curve even for those with some coding experience.

It should also be noted that only a subset of Snakemake's functionality was used in this study. Features such as automated reporting, integration of per-rule `conda` environments via a `requirements.yaml`, and the use of container images per rule exist and could further improve the portability and reproducibility of the workflow.

The forcing generation, regridding, downscaling, and bias correction steps are not specific to the Aral Sea and can be applied to other catchments within the eWaterCycle framework with minimal modification. The same applies to the calibration workflow, which is built around a general optimization procedure rather than any case-specific logic

6.2.2. Reproducibility and Software Dependencies

To make the project as reproducible as possible, model parameters and simulation specifics are defined in `.yaml` configuration files, and random seeds are fixed for the calibration model components to limit run-to-run variability. The use of Docker containers means that the PCR-GLOBWB2 model version and its immediate dependencies are preserved exactly as used.

Nevertheless, there are factors that may still introduce differences when reproducing the workflow. First, the calibration procedure relies on parallel computation to manage the demands of PCR-GLOBWB2. Model evaluations complete asynchronously, meaning that the optimisation algorithm updates its internal state based on the order in which results become available. Small variations in worker runtime can therefore lead to differences in the sequence of learning steps and potentially to slightly different final parameter sets. More deterministic algorithms such as CMA-ES (Hansen, 2006) reduce this effect. However, reproducibility of the main results is preserved through the availability of calibrated configuration files, which allow simulations to be rerun without repeating the calibration step.

Second, although the Docker containers preserves the core model, PRG-GLOBWB, external Python dependencies and components of the eWaterCycle framework itself continue to evolve. Reproducing the workflow with newer versions of these tools may therefore require minor adjustments to wrapper scripts, even when the underlying model configuration is unchanged.

A further practical limitation is computational cost: ensemble simulations require approximately seven to ten days of wall-clock time, while calibration adds an additional week on the infrastructure described in appendix D. This limits full end-to-end reproducibility due to the substantial computational resources required. However, reproducibility of the main results is preserved through the availability of calibrated configuration files, which allow simulations to be rerun without repeating the calibration step.

In addition to software and methodological considerations, data availability introduces an external constraint on reproducibility that lies outside the control of the modelling framework itself. Access to CMIP6 datasets via the Earth System Grid Federation (ESGF) infrastructure can be inconsistent, as experienced with EC-EARTH. Similar limitations may apply to reanalysis products such as ERA5. These practical limitations can therefore influence model selection and data completeness, even when the methodological framework itself is as reproducable as possible.

The modelling workflow evolved iteratively over the course of the study. Early versions relied on Python-

based configuration files (e.g. constants.py and paths.py) to define parameters and file locations, while later stages transitioned towards a more structured YAML-based configuration system aligned with the Snakemake workflow. This shift improved transparency and reproducibility by centralising key experimental settings within the workflow definition.

However, due to the incremental nature of this transition, some parameters and settings may still reside outside the central configuration files. While all core components of the modelling chain are controlled through the Snakemake YAML workflow specification, isolated "stray" configuration values may persist in auxiliary scripts from earlier development stages.

6.2.3. Calibration Framework

The calibration method used in the study optimizes a single combined objective, which, as discussed in section 6.1.2, introduces implicit trade-offs between basin performance that were partially resolved post-calibration rather than during optimization. A true multi-objective approach would make these trade-offs explicit by simultaneously optimizing basin-specific objectives and exposing the Pareto front.

The choice of optimization algorithm also warrants reflection. The method used here adopts a Bayesian approach through the Tree-structured Parzen Estimator (TPE), which builds a probabilistic model of the objective function and focuses sampling on promising regions. This is well suited to the non-linear parameter space of PCR-GLOBWB2 and avoids some of the convergence risks of evolution strategies such as Hansen (2006). However, asynchronous parallelisation of Optuna trials means that the sampler state at the time of each trial submission depends on which previous evaluations have already returned, introducing a degree of non-determinism as discussed in section 6.2.

The current calibration used 238 model evaluations, whereas comparable studies have used an order of magnitude more (Guan et al., 2026). Optuna supports chained sampling strategies: for example using an initial random or quasi-random phase to broadly explore the parameter space, followed by a TPE phase that concentrates effort in high-performing regions, which could improve convergence within a comparable budget. Inspecting the hyperparameter importance distributions between phases would additionally enable more informed decisions about where to focus subsequent calibration effort. Combined with altering the parameter space to include additional sensitive parameters identified by the mutual information analysis or from literature, while fixing near-zero sensitivity parameters to reduce the effective dimensionality, this could substantially improve calibration outcomes.

6.3. Uncertainties & limitations

7

conclusion

This study set out to achieve two objectives: to model the effects of climate change and human activities on the Aral Sea, and to do so within a reproducible computational framework. Against both, the answer is yes; with an asterisk

On the hydrology: the model successfully reproduced the long-term desiccation trend of the Aral Sea, confirming that the dominant underlying processes are captured in the framework. The inclusion of the Karakum Canal and the relatively high importance of the crop coefficient parameter strongly suggest that anthropogenic agricultural expansion, rather than climate change, has been the primary driver of the Aral Sea's historical decline. Future projections reveal that structural uncertainty from GCM choice exceeds scenario uncertainty, suggesting that expanding the model ensemble is a higher priority than adding emission scenarios in future work. However, due to the calibration mismatch between the Syr Darya and Amu Darya, where parameter sets optimised for one basin systematically degrade performance in the other, quantifying the relative contributions precisely remains challenging. The remaining shortcomings point to incomplete representation of agricultural water use and uncertainties in the meteorological forcing for this terrain, rather than fundamental limitations of the modelling approach itself. The framework is therefore well-positioned for more definitive attribution analysis once these components are improved.

On the methodology: the study designed and executed an automated workflow combining the eWaterCycle platform with Snakemake orchestration. The workflow enforces strict data lineage and FAIR data principles, and demonstrates that multi-stage hydrological simulations can be bundled into a transferable open-science framework. In theory, all results can be deleted and reproduced from scratch with a single command. In practice, full reproduction requires access to the eWaterCycle infrastructre or a comparable HPC environment, as PCR-GLOBWB2 demands significant computational resources. A single historical run takes approximately 16 hours using parallelised segments; reproducing the full experiment set across all climate models and scenarios requires approximately 7–10 days of compute time, excluding calibration. The PCR-GLOBWB2 parameter set used is publicly available but amounts to 300 GB of data. Reproducibility is therefore guaranteed in principle and documented in full, but is computationally non-trivial. Model parameters are specified in version-controlled configuration files, random seeds are fixed and documented, and intermediate outputs are stored as standardised NetCDF files at each stage. Forcing generation, downscaling, and bias correction are location-independent, making the workflow reusable beyond this study. The same holds for running PCR-GLOBWB2 model for different climate scenarios. Model state manipulation is now possible for other to use even if the exact anthropogenic water experiments are not directly transferable. Where full FAIR compliance is not possible due to licensing constraints, the documentation explains how to obtain the relevant data. The workflow is fully transparent and shared on the eWaterCycle GitHub.

Beyond the geographic context of the Aral sea and that of Central Asia, the primary value of this work lies in its methodological blueprint. By successfully bundling a complex, multi-stage simulation into a transparent, FAIR-compliant pipeline, this study tries to provide a practical template for those using eWaterCycle and the for the hydrological community as a whole.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019, July 25). *Optuna: A Next-generation Hyperparameter Optimization Framework*. arXiv: 1907.10902 [cs]. https://doi.org/10.48550/arXiv.1907.10902

Allen, R., Pereira, L., Raes, D., & Smith, M. (1998). Crop evapotranspiration guidelines for computing crop requirements. FAO Irrig. Drain. Report modeling and application. *J. Hydrol.*, *285*, 19–40.

Aus Der Beek, T., Voß, F., & Flörke, M. (2011). Modelling the impact of Global Change on the hydrological system of the Aral Sea basin. *Physics and Chemistry of the Earth, Parts A/B/C*, *36*(13), 684–695. https://doi.org/10.1016/j.pce.2011.03.004

Baker, A., C. Hellstrom, J., Kelly, B. F. J., Mariethoz, G., & Trouet, V. (2015). A composite annual-resolution stalagmite record of North Atlantic climate over the last three millennia. *Scientific Reports*, *5*(1), 10307. https://doi.org/10.1038/srep10307

Beck, H. E., McVicar, T. R., Vergopolan, N., Berg, A., Lutsko, N. J., Dufour, A., Zeng, Z., Jiang, X., Van Dijk, A. I. J. M., & Miralles, D. G. (2023). High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. *Scientific Data*, *10*(1), 724. https://doi.org/10.1038/s41597-023-02549-6

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 180214. https://doi.org/10.1038/sdata.2018.214

Berdimbetov, T., Pushpawela, B., Murzintcev, N., Nietullaeva, S., Gafforov, K., Tureniyazova, A., & Madetov, D. (2024). Unraveling the Intricate Links between the Dwindling Aral Sea and Climate Variability during 2002–2017. *Climate*, *12*(7), 105. https://doi.org/10.3390/cli12070105

Berdimbetov, T. T., Ma, Z.-G., Liang, C., & Ilyas, S. (2020). Impact of Climate Factors and Human Activities on Water Resources in the Aral Sea Basin. *Hydrology*, *7*(2), 30. https://doi.org/10.3390/hydrology7020030

Bergstra, J., Bardenet, R., Bengio, Y., & Kegl, B. (2011). Algorithms for Hyper-Parameter Optimization.

Bergström, S. (1976, January). *Development and application of a conceptual runoff model for scandinavian catchments* (Vol. 134 pp.).

Boomer, I., Wünnemann, B., Mackay, A., Austin, P., Sorrel, P., Reinhardt, C., Keyser, D., Guichard, F., & Fontugne, M. (2009). Advances in understanding the late Holocene history of the Aral Sea region. *Quaternary International*, *194*(1–2), 79–90. https://doi.org/10.1016/j.quaint.2008.03.007

Boomer, I., Aladin, N., Plotnikov, I., & Whatley, R. (2000). The palaeolimnology of the Aral Sea: A review. *Quaternary Science Reviews*, *19*(13), 1259–1278. https://doi.org/10.1016/S0277-3791(00)00002-0

Brite, E. B. (2016). Irrigation in the Khorezm oasis, past and present: A political ecology perspective. *Journal of Political Ecology*, *23*(1). https://doi.org/10.2458/v23i1.20177

Byrne, A., Norris, K., Chadwick, M. A., Avery, S., Olaka, L., & Tebbs, E. J. (2024). Rising lake levels in central East Africa are driven by increasing rainfall and land-use intensification. *Journal of Hydrology: Regional Studies*, *56*, 101999. https://doi.org/10.1016/j.ejrh.2024.101999

C3S. (2018). *ERA5 hourly data on single levels from 1940 to present*. https://doi.org/10.24381/CDS.ADBB2D47

Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *Journal of Climate*, *28*(17), 6938–6959. https://doi.org/10.1175/JCLI-D-14-00754.1

Cretaux, J.-F., Letolle, R., & Bergé-Nguyen, M. (2013). History of Aral Sea level variability and current scientific debates. *Global and Planetary Change*, *110*, 99–113. https://doi.org/10.1016/j.gloplacha.2013.05.006

Crétaux, J.-F., Kouraev, A. V., Papa, F., Bergé-Nguyen, M., Cazenave, A., Aladin, N., & Plotnikov, I. S. (2005). Evolution of Sea Level of the Big Aral Sea from Satellite Altimetry and Its Implications for Water Balance. *Journal of Great Lakes Research*, *31*(4), 520–534. https://doi.org/10.1016/S0380-1330(05)70281-1

De Visser, C., Johansson, L. F., Kulkarni, P., Mei, H., Neerincx, P., Joeri Van Der Velde, K., Horvatovich, P., Van Gool, A. J., Swertz, M. A., Hoen, P. A. C. T., & Niehues, A. (2023). Ten quick tips for building FAIR workflows (P. M. Palagi, Ed.). *PLOS Computational Biology*, *19*(9), e1011369. https://doi.org/10.1371/journal.pcbi.1011369

Deng, H., & Chen, Y. (2017). Influences of recent climate change and human activities on water storage variations in Central Asia. *Journal of Hydrology*, *544*, 46–57. https://doi.org/10.1016/j.jhydrol.2016.11.006

Duan, Z., Afzal, M. M., Liu, X., Chen, S., Du, R., Zhao, B., Yuan, W., & Awais, M. (2024). Effects of climate change and human activities on environment and area variations of the Aral Sea in Central Asia. *International Journal of Environmental Science and Technology*, *21*(2), 1715–1728. https://doi.org/10.1007/s13762-023-05072-8

Encyclopædia Britannica. (2026, March 28). *Shrinkage of the Aral Sea, 1960–2014*. Retrieved April 23, 2026, from %7Bhttps://www.britannica.com/place/Aral-Sea%7D

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Eyring, V., Righi, M., Lauer, A., Evaldsson, M., Wenzel, S., Jones, C., Anav, A., Andrews, O., Cionni, I., Davin, E. L., Deser, C., Ehbrecht, C., Friedlingstein, P., Gleckler, P., Gottschaldt, K.-D., Hagemann, S., Juckes, M., Kindermann, S., Krasting, J., ... Williams, K. D. (2016). ESMValTool (v1.0) – a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geoscientific Model Development*, *9*(5), 1747–1802. https://doi.org/10.5194/gmd-9-1747-2016

FAO. (2012). *AQUASTAT Transboundary River Basin Overview – Aral Sea*. Food and Agriculture Organization of the United Nations. Rome, Italy. https://www.fao.org/aquastat/en/countries-and-basins/transboundary-river-basins/aral-sea

Frenken, K. (2013). *Irrigation in Central Asia in figures - AQUASTAT Survey 2012*. Food and Agriculture Organization of the United Nations (FAO). https://doi.org/10.13140/RG.2.1.2770.8247

GRDC: The global runoff data centre. (n.d.). *Global runoff data centre*. 56068 Koblenz, Germany, Federal Institute of Hydrology (BfG). Retrieved September 25, 2025, from https://www.bafg.de/GRDC/

Guan, Y., Lian, Y., Ling, M., Gao, Y., Ren, Y., & Sun, X. (2026). Improving PCR-GLOBWB 2 to model regional water stress dynamics under climate and socioeconomic change scenarios. *Journal of Hydrology*, *664*, 134472. https://doi.org/10.1016/j.jhydrol.2025.134472

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Haasnoot, D. (2024, July 11). *Analysing deficiencies in hydrological models using data assimilation* [Master's thesis, Delft University of Technology]. https://resolver.tudelft.nl/uuid:a405fd22-72a8-49d0-a90b-9a020f00579b

Hall, C. A., Saia, S. M., Popp, A. L., Dogulu, N., Schymanski, S. J., Drost, N., Van Emmerik, T., & Hut, R. (2022). A hydrologist's guide to open science. *Hydrology and Earth System Sciences*, *26*(3), 647–664. https://doi.org/10.5194/hess-26-647-2022

Hansen, N. (2006). The CMA Evolution Strategy: A Comparing Review. In J. A. Lozano, P. Larrañaga, I. Inza, & E. Bengoetxea (Eds.), *Towards a New Evolutionary Computation* (pp. 75–102, Vol. 192). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-32494-1_4

Hansen, N., Akimoto, Y., & Baudis, P. (2019, February). *CMA-ES/pycma on Github*. https://doi.org/10.5281/zenodo.2559634

He, H., Hamdi, R., Luo, G., Cai, P., Zhang, M., Shi, H., Li, C., Termonia, P., De Maeyer, P., & Kurban, A. (2022). Numerical study on the climatic effect of the Aral Sea. *Atmospheric Research*, *268*, 105977. https://doi.org/10.1016/j.atmosres.2021.105977

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., ... Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hoch, J. M., Sutanudjaja, E. H., Wanders, N., Van Beek, R. L. P. H., & Bierkens, M. F. P. (2023). Hyper-resolution PCR-GLOBWB: Opportunities and challenges from refining model spatial resolution to 1 km over the European continent. *Hydrology and Earth System Sciences*, *27*(6), 1383–1401. https://doi.org/10.5194/hess-27-1383-2023

Hoogelander, V. (2022, December 7). *Assessing the impact of climate change on streamflow in catchments across the United States, using an easily reproducible modeling approach* [Master's thesis, Delft University of Technology].

Hut, R. (2022). FAIR Models. *Groundwater*, *60*(3), 309–310. https://doi.org/10.1111/gwat.13180

Hut, R., Drost, N., Van De Giesen, N., Van Werkhoven, B., Abdollahi, B., Aerts, J., Albers, T., Alidoost, F., Andela, B., Camphuijsen, J., Dzigan, Y., Van Haren, R., Hutton, E., Kalverla, P., Van Meersbergen, M., Van Den Oord, G., Pelupessy, I., Smeets, S., Verhoeven, S., ... Weel, B. (2022). The eWaterCycle platform for open and FAIR hydrological collaboration. *Geoscientific Model Development*, *15*(13), 5371–5390. https://doi.org/10.5194/gmd-15-5371-2022

Hut, R., & Hall, C. (2025). More efficient reproducible research in hydrology: Moving research down the academic career scale (MRDTACS). *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *383*(2302), 20240296. https://doi.org/10.1098/rsta.2024.0296

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., & Arheimer, B. (2016). Most computational hydrology is not reproducible, so is it really science?: REPRODUCIBLE COMPUTATIONAL HYDROLOGY. *Water Resources Research*, *52*(10), 7548–7555. https://doi.org/10.1002/2016WR019285

Hutton, E., Piper, M., & Tucker, G. (2020). The Basic Model Interface 2.0: A standard interface for coupling numerical models in the geosciences. *Journal of Open Source Software*, *5*(51), 2317. https://doi.org/10.21105/joss.02317

Khazaei, B., Read, L. K., Casali, M., Sampson, K. M., & Yates, D. N. (2022). GLOBathy, the global lakes bathymetry dataset. *Scientific Data*, *9*(1), 36. https://doi.org/10.1038/s41597-022-01132-9

Krapivin, V., Mkrtchyan, F., & Rochon, G. (2019). Hydrological Model for Sustainable Development in the Aral Sea Region. *Hydrology*, *6*(4), 91. https://doi.org/10.3390/hydrology6040091

Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. *Physical Review E*, *69*(6), 066138. https://doi.org/10.1103/PhysRevE.69.066138

Krivonogov, S. K. (2009). Extent of the Aral Sea drop in the Middle Age. *Doklady Earth Sciences*, *428*(1), 1146–1150. https://doi.org/10.1134/S1028334X09070241

Krivonogov, S., Burr, G., Kuzmin, Y., Gusskov, S., Kurmanbaev, R., Kenshinbay, T., & Voyakin, D. (2014). The fluctuating Aral Sea: A multidisciplinary-based history of the last two thousand years. *Gondwana Research*, *26*(1), 284–300. https://doi.org/10.1016/j.gr.2014.02.004

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*(15), 2171–2186. https://doi.org/10.1002/hyp.9740

Lehner, B., Verdin, K., & Jarvis, A. (2008). New Global Hydrography Derived From Spaceborne Elevation Data. *Eos, Transactions American Geophysical Union*, *89*(10), 93–94. https://doi.org/10.1029/2008EO100001

Létolle, R., Micklin, P., Aladin, N., & Plotnikov, I. (2007). Uzboy and the Aral regressions: A hydrological approach. *Quaternary International*, *173–174*, 125–136. https://doi.org/10.1016/j.quaint.2007.03.003

Lindström, G., Johansson, B., Persson, M., Gardelin, M., & Bergström, S. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, *201*(1–4), 272–288. https://doi.org/10.1016/S0022-1694(97)00041-3

López López, P., Sutanudjaja, E. H., Schellekens, J., Sterk, G., & Bierkens, M. F. P. (2017). Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products. *Hydrology and Earth System Sciences*, *21*(6), 3125–3144. https://doi.org/10.5194/hess-21-3125-2017

Maraun, D., & Widmann, M. (2018, January 18). Model Output Statistics. In *Statistical Downscaling and Bias Correction for Climate Research* (1st ed., pp. 170–200). Cambridge University Press. https://doi.org/10.1017/9781107588783

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, *7*(1), 13603. https://doi.org/10.1038/ncomms13603

Micklin, P. (2007). The Aral Sea Disaster. *Annual Review of Earth and Planetary Sciences*, *35*(1), 47–72. https://doi.org/10.1146/annurev.earth.35.031306.140120

Micklin, P. (2016). The future Aral Sea: Hope and despair. *Environmental Earth Sciences*, *75*(9), 844. https://doi.org/10.1007/s12665-016-5614-5

Micklin, P., Aladin, N., & Plotnikov, I. (Eds.). (2014). *The Aral Sea: The Devastation and Partial Rehabilitation of a Great Lake*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-02356-9

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Van Dyken, P. C., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Vieira, F. G., Meesters, C., Lee, S., Twardziok, S. O., Kanitz, A., VanCampen, J., Malladi, V., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2025). Sustainable data analysis with Snakemake. *F1000Research*, *10*, 33. https://doi.org/10.12688/f1000research.29032.3

Murodov, A., Cuo, L., Li, N., Murodov, D., Hou, M., & Hussain, G. (2023). Extreme Hydrometeorological Conditions and Changes in the Amu Darya River Basin in Central Asia. *Journal of Hydrometeorology*, *24*(2), 315–334. https://doi.org/10.1175/JHM-D-22-0025.1

Mutual_info_regression. (n.d.). scikit-learn. Retrieved April 30, 2026, from https://scikit-learn/stable/modules/generated/sklearn.feature_selection.mutual_info_regression.html

Nachtnebel, H. (2006, December). *The rehabilitation of the ecosystem and bioproductivity of the Aral Sea under conditions of water scarcity* (Final Report). International Association for the promotion of co-operation with scientists from the New Independent States of the former Soviet Union, INTAS Project – 0511 REBASOWS. Vienna - Tashkent. https://www.cawater-info.net/library/eng/intas_2006_en.pdf

NASA Earth Science Data and Information System (ESDIS). (2026, May 23). *Global Imagery Browse Services (GIBS) API*. NASA Earthdata. Retrieved May 23, 2026, from https://www.earthdata.nasa.gov/engage/open-data-services-software/earthdata-developer-portal/gibs-api

Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Oberhänsli, H., Boroffka, N., Sorrel, P., & Krivonogov, S. (2007). Climate variability during the past 2,000 years and past economic and irrigation activities in the Aral Sea basin. *Irrigation and Drainage Systems*, *21*(3–4), 167–183. https://doi.org/10.1007/s10795-007-9031-5

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., & Sanderson, B. M. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

OpenStreetMap contributors. (2026). Karakum canal GeoJSON dataset extracted via overpass turbo. https://openstreetmap.org

Pan, X., Wang, W., Liu, T., Akmalov, S., De Maeyer, P., & Van De Voorde, T. (2022). Integrated modeling to assess the impact of climate change on the groundwater and surface water in the South Aral Sea area. *Journal of Hydrology*, *614*, 128641. https://doi.org/10.1016/j.jhydrol.2022.128641

Portmann, F. T., Siebert, S., & Döll, P. (2010). MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochemical Cycles*, *24*(1), 2008GB003435. https://doi.org/10.1029/2008GB003435

Riahi, K., Van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., Kc, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., ... Tavoni, M. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., Vegas-Regidor, J., Bock, L., Brötz, B., De Mora, L., Diblen, F., Dreyer, L., Drost, N., Earnshaw, P., Hassler, B., Koldunov, N., Little, B., Loosveldt Tomas, S., & Zimmermann, K. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0 – technical overview. *Geoscientific Model Development*, *13*(3), 1179–1199. https://doi.org/10.5194/gmd-13-1179-2020

Rohwer, J., Gerten, D., & Lucht, W. (2007). *Development of functional types of irrigation for improved global crop modelling* (No. 104). PIK.

Rolf Hut & Mark Melotto. (2026, June 3). *Climate Change Impact Analysis using eWaterCycle*. Retrieved June 3, 2026, from https://github.com/eWaterCycle/CCI-analysis-seamless

Ruijsch, J., Verstegen, J. A., Sutanudjaja, E. H., & Karssenberg, D. (2021). Systemic change in the Rhine-Meuse basin: Quantifying and explaining parameters trends in the PCR-GLOBWB global hydrological model. *Advances in Water Resources*, *155*, 104013. https://doi.org/10.1016/j.advwatres.2021.104013

Saltelli, A. (Ed.). (2008). *Global sensitivity analysis: The primer*. John Wiley.

Schwatke, C., Dettmering, D., Bosch, W., & Seitz, F. (2015). DAHITI – an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. *Hydrology and Earth System Sciences*, *19*(10), 4345–4364. https://doi.org/10.5194/hess-19-4345-2015

Schwertfeger, B. (2025). *Btsschwertfeger/python-cmethods: V2.3.2*. https://doi.org/10.5281/zenodo.17606049

Seibert, J., & Bergström, S. (2022). A retrospective on hydrological catchment modelling based on half a century with the HBV model. *Hydrology and Earth System Sciences*, *26*(5), 1371–1388. https://doi.org/10.5194/hess-26-1371-2022

Siebert, S., & Döll, P. (2010). Quantifying blue and green virtual water contents in global crop production as well as potential production losses without irrigation. *Journal of Hydrology*, *384*(3–4), 198–217. https://doi.org/10.1016/j.jhydrol.2009.07.031

Sorrel, P., Popescu, S.-M., Klotz, S., Suc, J.-P., & Oberhänsli, H. (2007). Climate variability in the Aral Sea basin (Central Asia) during the late Holocene based on vegetation changes. *Quaternary Research*, *67*(3), 357–370. https://doi.org/10.1016/j.yqres.2006.11.006

Stagge, J. H., Rosenberg, D. E., Abdallah, A. M., Akbar, H., Attallah, N. A., & James, R. (2019). Assessing data availability and research reproducibility in hydrology and water resources. *Scientific Data*, *6*(1), 190030. https://doi.org/10.1038/sdata.2019.30

Stulina, G., & Eshchanov, O. (2013). Climate change impacts on hydrology and environment in the Pre-Aral region. *Quaternary International*, *311*, 87–96. https://doi.org/10.1016/j.quaint.2013.07.015

Sutanudjaja, E., van Beek, R., Wanders, N., Wada, Y., Bosmans, J., Drost, N., van der Ent, R., de Graaf, I., Hoch, J., de Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M., Vannametee, E., Wisser, D., & Bierkens, M. (2017, November 11). *Pcr-Globwb 2 Input Files Version 2017_11_Beta_1*. Zenodo. https://doi.org/10.5281/ZENODO.1045338

Sutanudjaja, E. H., Van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H. C., Drost, N., Van Der Ent, R. J., De Graaf, I. E. M., Hoch, J. M., De Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M. W., Vannametee, E., Wisser, D., & Bierkens, M. F. P. (2018). PCR-GLOBWB 2: A 5 arcmin global hydrological and water resources model. *Geoscientific Model Development*, *11*(6), 2429–2453. https://doi.org/10.5194/gmd-11-2429-2018

Tao, D., Shi, H., Gao, C., Zhan, J., & Ke, X. (2020). Water Storage Monitoring in the Aral Sea and its Endorheic Basin from Multisatellite Data and a Hydrological Model. *Remote Sensing*, *12*(15), 2408. https://doi.org/10.3390/rs12152408

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., Knutti, R., Lowe, J., O'Neill, B., Sanderson, B., Van Vuuren, D., Riahi, K., Meinshausen, M., Nicholls, Z., Tokarska, K. B., Hurtt, G., Kriegler, E., Lamarque, J.-F., Meehl, G., … Ziehn, T. (2021). Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6. *Earth System Dynamics*, *12*(1), 253–293. https://doi.org/10.5194/esd-12-253-2021

Tong, Y., Gao, X., Han, Z., Xu, Y., Xu, Y., & Giorgi, F. (2021). Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods. *Climate Dynamics*, *57*(5–6), 1425–1443. https://doi.org/10.1007/s00382-020-05447-4

Toonen, W. H. J., Macklin, M. G., Dawkes, G., Durcan, J. A., Leman, M., Nikolayev, Y., & Yegorov, A. (2020). A hydromorphic reevaluation of the forgotten river civilizations of Central Asia. *Proceedings of the National Academy of Sciences*, *117*(52), 32982–32988. https://doi.org/10.1073/pnas.2009553117

Touge, Y., Kobayashi, G., Khujanazarov, T., & Tanaka, K. (2024). Reproduction of historical water balance in the Aral Sea Basin: The physically-based framework to quantify water consumption components in endorheic lake. *Journal of Hydrology*, *640*, 131711. https://doi.org/10.1016/j.jhydrol.2024.131711

TU Delft OPEN Publishing policies. (n.d.). TU Delft. Retrieved May 4, 2026, from https://www.tudelft.nl/open-publishing/publish-with-us/resources-support/tu-delft-open-publishing-policies

Van Beek, L. P. H., Wada, Y., & Bierkens, M. F. P. (2011). Global monthly water stress: 1. Water balance and water availability. *Water Resources Research*, *47*(7), 2010WR009791. https://doi.org/10.1029/2010WR009791

van den Oord, G., Verhoeven, S., & Pelupessy, I. (2023, October 30). *gRPC wrapper for model with a Basic modeling interface* (Version v0.5.0). https://doi.org/10.5281/ZENODO.1462641

van der Veen, A. B. (2026). *Reproducible eWaterCycle-based modeling framework for assessing climate change and human activity impacts on the Aral Sea*.
Includes all scripts, notebooks, and workflows used in the MSc thesis Modeling the effects of climate change and human activities on the Aral Sea in a reproducible framework.

van der Ven, M. (2021, July 30). *Matching streamflow river gauges with hydrologic models* [Master's thesis, Delft University of Technology]. https://resolver.tudelft.nl/uuid:6f9643e8-474f-42f8-a895-b9074e9869a6

Van Vuuren, D. P., O'Neill, B. C., Tebaldi, C., Sanderson, B. M., Chini, L. P., Friedlingstein, P., Hasegawa, T., Riahi, K., Govindasamy, B., Bauer, N., Eyring, V., Fall, C. M. N., Frieler, K., Gidden, M. J., Gohar, L. K., Högner, A., Jones, A. D., Kikstra, J., King, A., … Ziehn, T. (2026). The Scenario Model Intercomparison Project for CMIP7 (ScenarioMIP-CMIP7). *Geoscientific Model Development*, *19*(7), 2627–2656. https://doi.org/10.5194/gmd-19-2627-2026

Varis, O., & Kummu, M. (2012). The Major Central Asian River Basins: An Assessment of Vulnerability. *International Journal of Water Resources Development*, *28*(3), 433–452. https://doi.org/10.1080/07900627.2012.684309

Wada, Y., De Graaf, I. E. M., & Van Beek, L. P. H. (2016). High-resolution modeling of human and climate impacts on global water resources. *Journal of Advances in Modeling Earth Systems*, *8*(2), 735–763. https://doi.org/10.1002/2015MS000618

Wang, J., Song, C., Reager, J. T., Yao, F., Famiglietti, J. S., Sheng, Y., MacDonald, G. M., Brun, F., Schmied, H. M., Marston, R. A., & Wada, Y. (2018). Recent global decline in endorheic basin water storages. *Nature Geoscience*, *11*(12), 926–932. https://doi.org/10.1038/s41561-018-0265-7

Watanabe, S. (2025, September 30). *Tree-Structured Parzen Estimator: Understanding Its Algorithm Components and Their Roles for Better Empirical Performance*. arXiv: 2304.11127 [cs]. https://doi.org/10.48550/arXiv.2304.11127

Wiersma, P., Aerts, J., Zekollari, H., Hrachowitz, M., Drost, N., Huss, M., Sutanudjaja, E. H., & Hut, R. (2022). Coupling a global glacier model to a global hydrological model prevents underestimation of glacier runoff. *Hydrology and Earth System Sciences*, *26*(23), 5971–5986. https://doi.org/10.5194/hess-26-5971-2022

Wikimedia Commons contributors. (2015, October 6). *Ocean drainage basins of the world*. Retrieved January 16, 2026, from https://upload.wikimedia.org/wikipedia/commons/b/b1/Ocean_drainage.png

Wu, F., Yang, X., Cui, Z., Ren, L., Jiang, S., Liu, Y., & Yuan, S. (2024). The impact of human activities on blue-green water resources and quantification of water resource scarcity in the Yangtze River Basin. *Science of The Total Environment*, *909*, 168550. https://doi.org/10.1016/j.scitotenv.2023.168550

Yang, X., Wu, F., Yuan, S., Ren, L., Sheffield, J., Fang, X., Jiang, S., & Liu, Y. (2024). Quantifying the Impact of Human Activities on Hydrological Drought and Drought Propagation in China Using the PCR-GLOBWB v2.0 Model. *Water Resources Research*, *60*(1), e2023WR035443. https://doi.org/10.1029/2023WR035443

Yapiyev, V., Koriche, S. A., Snow, D. D., Samarkhanov, K., Mao, X., & Li, J. (2024). Editorial: Climate change, land surface, and critical zone processes in endorheic basins. *Frontiers in Environmental Science*, *12*, 1479504. https://doi.org/10.3389/fenvs.2024.1479504

Yapiyev, V., Sagintayev, Z., Inglezakis, V., Samarkhanov, K., & Verhoef, A. (2017). Essentials of Endorheic Basins and Lakes: A Review in the Context of Current and Future Water Resource Management and Mitigation Activities in Central Asia. *Water*, *9*(10), 798. https://doi.org/10.3390/w9100798

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F., & Arino, O. (2022, October 28). *ESA WorldCover 10 m 2021 v200*. Zenodo. https://doi.org/10.5281/zenodo.7254221

Zhang, X., Kurbaniyazov, A., & Kirillin, G. (2021). Changing Pattern of Water Level Trends in Eurasian Endorheic Lakes as a Response to the Recent Climate Variability. *Remote Sensing*, *13*(18), 3705. https://doi.org/10.3390/rs13183705

Zhuang, J., raphael dussin, Bourgault, P., Huard, D., Banihirwe, A., Raynaud, S., Malevich, B., Schupfner, M., Gauthier, C., Filipe, Levang, S., Almansi, M., Jüling, A., RichardScottOZ, RondeauG, Rasp, S., Smith, T. J., Meyer, A. G., Mares, B., … Li, X. (2025, November 27). *Pangeo-data/xESMF: V0.9.2 - Third time's a charm* (Version v0.9.2). https://doi.org/10.5281/ZENODO.4294774

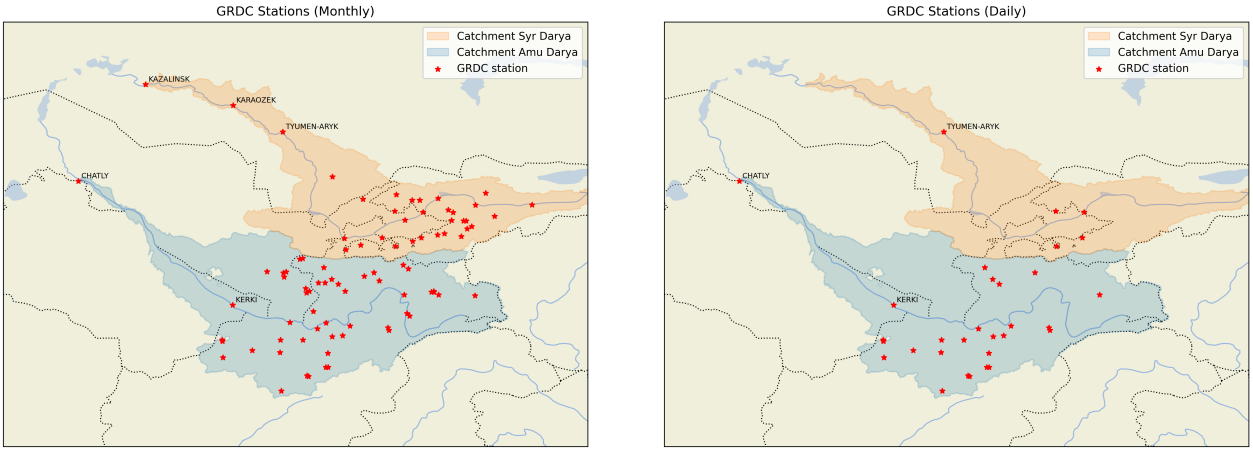
Zon, I. S., Glantz, M. H., Kostianoy, A. G., & Kosarev, A. N. (2009). *The Aral Sea Encyclopedia*. Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-540-85088-5

A

GRDC data overview

A.1. description

there are 85 stations, all of them have monthly values, some only daily. See figures A.1a and A.1b.



(**a**) Location of all GRDC stations with [MEAN MONTHLY DISCHARGE] in the Aral Sea Basin. Amount of available data per stations differs.

(**b**) Location of all GRDC stations with [MEAN DAILY DISCHARGE] in the Aral Sea Basin. Amount of available data per stations differs.

Figure A.1: Overview of GRDC stations in the Aral Sea Basin. [REF]

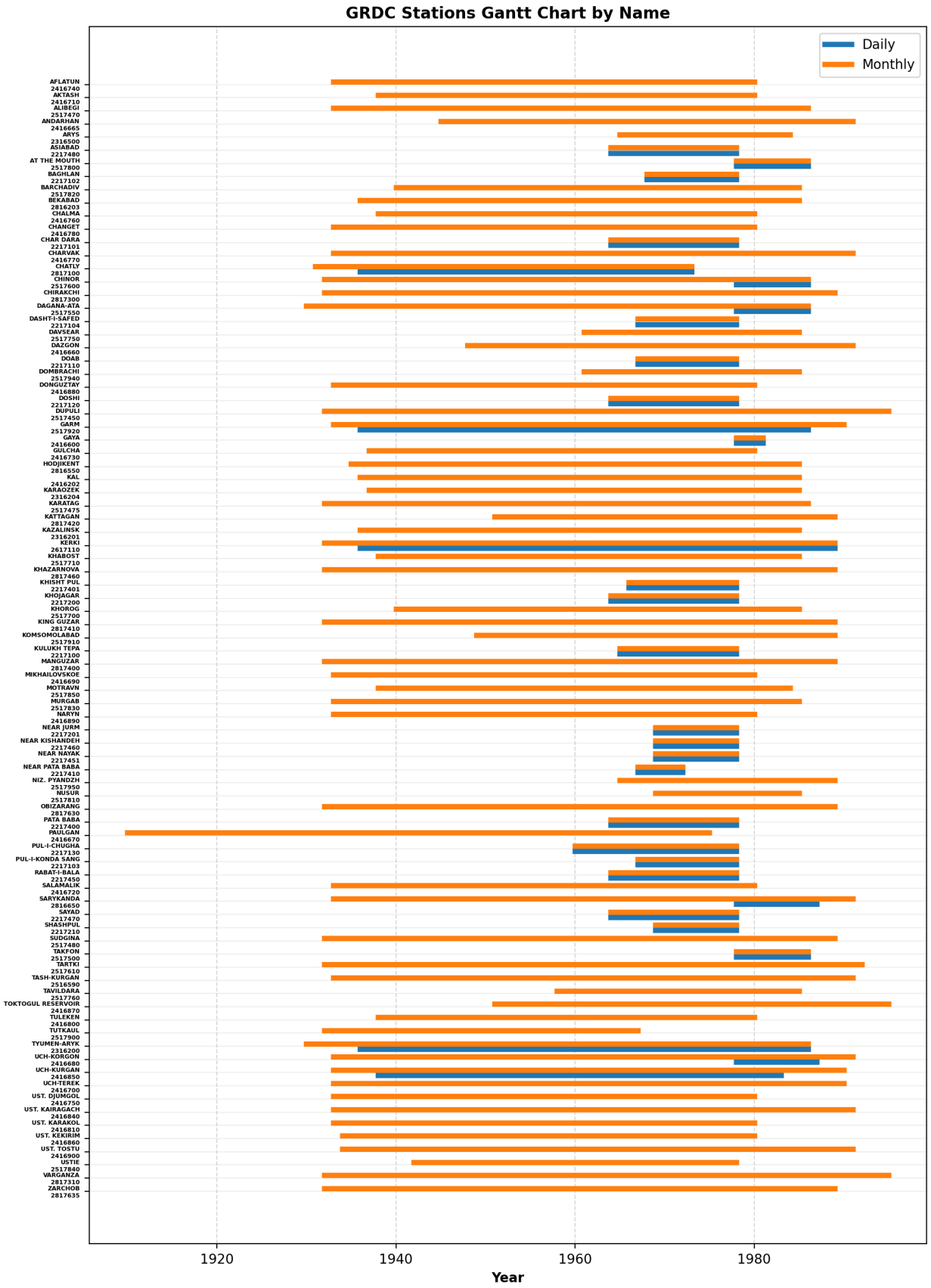


Figure A.2: placeholder

A.2. Monthly Data

Table A.1: Selected GRDC Station Metadata (Monthly)

GRDC No.	Station	Lat (°N)	Lon (°E)	Area (km²)	Time Series	Data Lines
2217100	KULUKH TEPA	36.98	68.30	37100	1965 - 1978	156
2217101	CHAR DARA	36.70	68.83	24820	1964 - 1978	170
2217102	BAGHLAN	36.10	68.67	19740	1968 - 1978	126
2217103	PUL-I-KONDA SANG	35.60	68.60	12610	1967 - 1978	132
2217104	DASHT-I-SAFED	35.30	67.92	3795	1967 - 1978	132
2217110	DOAB	35.27	67.98	5005	1967 - 1978	132
2217120	DOSHI	35.60	68.68	3705	1964 - 1978	168
2217130	PUL-I-CHUGHA	36.73	69.20	9760	1960 - 1978	222
2217200	KHOJAGAR	37.08	69.47	20645	1964 - 1978	174
2217201	NEAR JURM	36.93	70.87	7670	1969 - 1978	108
2217210	SHASHPUL	37.02	70.83	4485	1969 - 1978	108
2217400	PATA BABA	36.58	64.87	11775	1964 - 1978	173
2217401	KHISHT PUL	35.95	64.90	3280	1966 - 1978	153
2217410	NEAR PATA BABA	36.53	64.88	6685	1967 - 1972	59
2217450	RABAT-I-BALA	36.58	66.97	18035	1964 - 1978	174
2217451	NEAR NAYAK	34.75	67.00	1460	1969 - 1978	117
2217460	NEAR KISHANDEH	36.13	66.95	3070	1969 - 1978	117
2217470	SAYAD	36.58	67.78	8170	1964 - 1978	173
2217480	ASIABAD	36.20	65.95	6950	1964 - 1978	168
2316200	TYUMEN-ARYK	44.05	67.05	219000	1930 - 1986	684
2316201	KAZALINSK	45.75	62.12	-999	1936 - 1985	600
2316204	KARAOZEK	45.00	65.27	-999	1937 - 1985	585
2316500	ARYS	42.44	68.84	13100	1965 - 1984	240
2416202	KAL	40.88	71.45	90000	1936 - 1985	600
2416600	GAYA	41.20	71.08	657	1978 - 1981	48
2416660	DAZGON	39.82	69.32	712	1948 - 1991	528
2416665	ANDARHAN	39.98	69.85	1740	1945 - 1991	564
2416670	PAULGAN	40.12	71.72	1300	1910 - 1975	786
2416680	UCH-KORGON	40.25	72.03	2220	1933 - 1991	708
2416690	MIKHAILOVSKOE	41.16	73.18	1010	1933 - 1980	576
2416700	UCH-TEREK	41.80	71.14	-999	1933 - 1990	696
2416710	AKTASH	40.65	73.85	907	1938 - 1980	504
2416720	SALAMALIK	40.85	73.65	1180	1933 - 1980	576
2416730	GULCHA	40.30	73.47	2740	1937 - 1980	519
2416740	AFLATUN	41.60	71.98	863	1933 - 1980	576
2416750	UST. DJUMGOL	41.85	74.35	5290	1933 - 1980	570
2416760	CHALMA	40.57	73.68	3840	1938 - 1980	512
2416770	CHARVAK	41.25	73.00	1300	1933 - 1991	708
2416780	CHANGET	40.87	73.12	381	1933 - 1980	576
2416800	TULEKEN	40.40	72.87	2430	1938 - 1980	512
2416810	UST. KARAKOL	40.35	72.62	1680	1933 - 1980	576
2416840	UST. KAIRAGACH	41.02	74.67	530	1933 - 1991	708
2416850	UCH-KURGAN	41.17	72.10	58400	1933 - 1990	696
2416860	UST. KEKIRIM	41.42	73.98	34600	1934 - 1980	561
2416870	TOKTOGUL RESERVOIR	41.66	72.64	-999	1951 - 1995	531
2416880	DONGUZTAY	40.85	73.55	166	1933 - 1980	576
2416890	NARYN	41.43	76.02	10500	1933 - 1980	576
2416900	UST. TOSTU	41.60	71.70	366	1934 - 1991	696
2516590	TASH-KURGAN	40.25	70.62	1560	1933 - 1991	708

Continued on next page

Table A.1: Selected GRDC Station Metadata (Monthly)

GRDC No.	Station	Lat (°N)	Lon (°E)	Area (km²)	Time Series	Data Lines
2517450	DUPULI	39.51	67.77	10200	1932 - 1995	759
2517470	ALIBEGI	38.63	68.57	362	1933 - 1986	644
2517475	KARATAG	38.63	68.33	684	1932 - 1986	660
2517480	SUDGINA	39.48	67.67	1100	1932 - 1989	696
2517500	TAKFON	39.18	68.53	1490	1978 - 1986	108
2517550	DAGANA-ATA	38.76	68.81	1270	1930 - 1986	678
2517600	CHINOR	38.58	69.05	3040	1932 - 1986	660
2517610	TARTKI	37.60	68.15	9780	1932 - 1992	727
2517700	KHOROG	37.53	71.52	13700	1940 - 1985	552
2517710	KHABOST	37.44	71.61	4180	1938 - 1985	574
2517750	DAVSEAR	39.13	71.57	6550	1961 - 1985	300
2517760	TAVILDARA	38.70	70.52	5390	1958 - 1985	326
2517800	AT THE MOUTH	38.20	72.66	335	1978 - 1986	108
2517810	NUSUR	38.30	72.40	21400	1969 - 1985	204
2517820	BARCHADIV	38.30	72.48	16700	1940 - 1985	552
2517830	MURGAB	38.17	73.97	10500	1933 - 1985	636
2517840	USTIE	38.33	72.48	4500	1942 - 1978	437
2517850	MOTRAVN	38.20	71.42	1940	1938 - 1984	561
2517900	TUTKAUL	38.33	69.30	31200	1932 - 1967	425
2517910	KOMSOMOLABAD	38.87	69.98	29500	1949 - 1989	492
2517920	GARM	39.00	70.33	20000	1933 - 1990	696
2517940	DOMBRACHI	39.27	71.38	8370	1961 - 1985	300
2517950	NIZ. PYANDZH	37.19	68.61	113000	1965 - 1989	300
2617110	KERKI	37.83	65.25	309000	1932 - 1989	696
2816203	BEKABAD	40.22	69.27	142000	1936 - 1985	600
2816550	HODJIKENT	41.63	69.94	10900	1935 - 1985	612
2816650	SARYKANDA	39.94	71.11	2480	1933 - 1991	708
2817100	CHATLY	42.28	59.70	450000	1931 - 1973	516
2817300	CHIRAKCHI	39.03	66.48	4550	1932 - 1989	696
2817310	VARGANZA	38.84	67.10	511	1932 - 1995	759
2817400	MANGUZAR	37.21	67.31	13500	1932 - 1989	696
2817410	KING GUZAR	38.43	67.88	901	1932 - 1989	696
2817420	KATTAGAN	38.99	67.07	435	1951 - 1989	464
2817460	KHAZARNOVA	39.02	67.17	845	1932 - 1989	696
2817630	OBIZARANG	38.33	68.00	2630	1932 - 1989	696
2817635	ZARCHOB	38.27	67.90	2200	1932 - 1989	696

A.3. Daily Data

Table A.2: Selected GRDC Station Metadata (Daily)

GRDC No.	Station	Lat (°N)	Lon (°E)	Area (km²)	Time Series	Data Lines
2217100	KULUKH TEPA	36.98	68.30	37100	1965 - 1978	4,748
2217101	CHAR DARA	36.70	68.83	24820	1964 - 1978	5,170
2217102	BAGHLAN	36.10	68.67	19740	1968 - 1978	3,829
2217103	PUL-I-KONDA SANG	35.60	68.60	12610	1967 - 1978	4,018
2217104	DASHT-I-SAFED	35.30	67.92	3795	1967 - 1978	4,007
2217110	DOAB	35.27	67.98	5005	1967 - 1978	4,018
2217120	DOSHI	35.60	68.68	3705	1964 - 1978	5,113
2217130	PUL-I-CHUGHA	36.73	69.20	9760	1960 - 1978	6,739
2217200	KHOJAGAR	37.08	69.47	20645	1964 - 1978	5,271
2217201	NEAR JURM	36.93	70.87	7670	1969 - 1978	3,287
2217210	SHASHPUL	37.02	70.83	4485	1969 - 1978	3,287
2217400	PATA BABA	36.58	64.87	11775	1964 - 1978	5,244
2217401	KHISHT PUL	35.95	64.90	3280	1966 - 1978	4,656
2217410	NEAR PATA BABA	36.53	64.88	6685	1967 - 1972	1,786
2217450	RABAT-I-BALA	36.58	66.97	18035	1964 - 1978	5,296
2217451	NEAR NAYAK	34.75	67.00	1460	1969 - 1978	3,546
2217460	NEAR KISHANDEH	36.13	66.95	3070	1969 - 1978	3,552
2217470	SAYAD	36.58	67.78	8170	1964 - 1978	5,254
2217480	ASIABAD	36.20	65.95	6950	1964 - 1978	5,113
2316200	TYUMEN-ARYK	44.05	67.05	219000	1936 - 1986	18,613
2416600	GAYA	41.20	71.08	657	1978 - 1981	1,461
2416680	UCH-KORGON	40.25	72.03	2220	1978 - 1987	3,652
2416850	UCH-KURGAN	41.17	72.10	58400	1938 - 1983	16,801
2517500	TAKFON	39.18	68.53	1490	1978 - 1986	3,287
2517550	DAGANA-ATA	38.76	68.81	1270	1978 - 1986	3,287
2517600	CHINOR	38.58	69.05	3040	1978 - 1986	3,287
2517800	AT THE MOUTH	38.20	72.66	335	1978 - 1986	3,287
2517920	GARM	39.00	70.33	20000	1936 - 1986	18,323
2617110	KERKI	37.83	65.25	309000	1936 - 1989	19,724
2816650	SARYKANDA	39.94	71.11	2480	1978 - 1987	3,652
2817100	CHATLY	42.28	59.70	450000	1936 - 1973	13,880

B

Bathymetry

B.1. Generic Bathymetry

(Khazaei et al., 2022) provides estimated bathymetric maps and head–area–volume (H–A–V) relationships for all 1.4 million water bodies in the HydroLAKES database (Messager et al., 2016). While these estimates are inherently uncertain and, as will be shown for the Aral Sea in Section B.2, may be inaccurate for specific lakes—they provide a useful first-order approximation for modeling purposes.

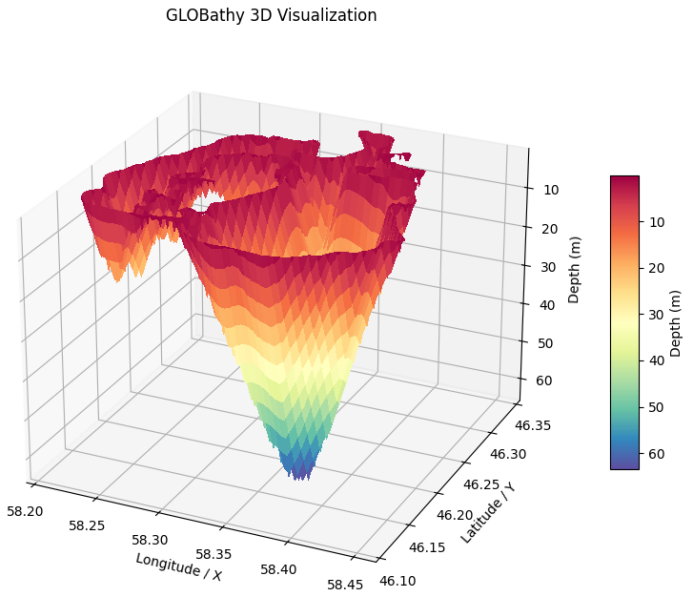


Figure B.1: Example bathymetry derived from the GLOBathy dataset (Khazaei et al., 2022).

B.2. Aral Sea-Specific Bathymetry

For the Aral Sea, higher-accuracy bathymetric information can be derived from multiple sources, including (Crétaux et al., 2005), high-resolution DEMs, and observational data from (Nachtnebel, 2006). Using these data, head–area–volume relationships can be constructed separately for the North (smaller) and South (larger) Aral Sea.

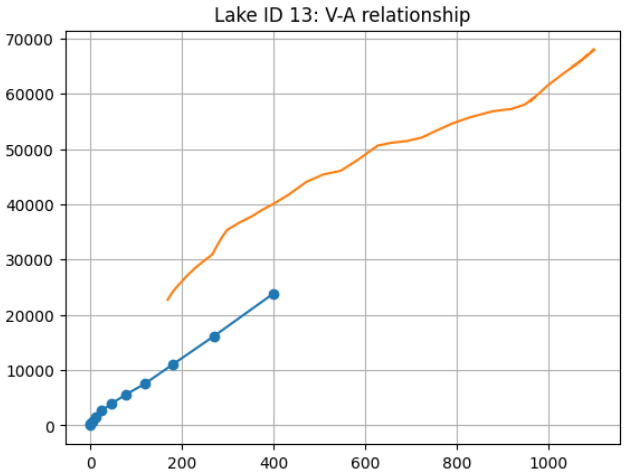
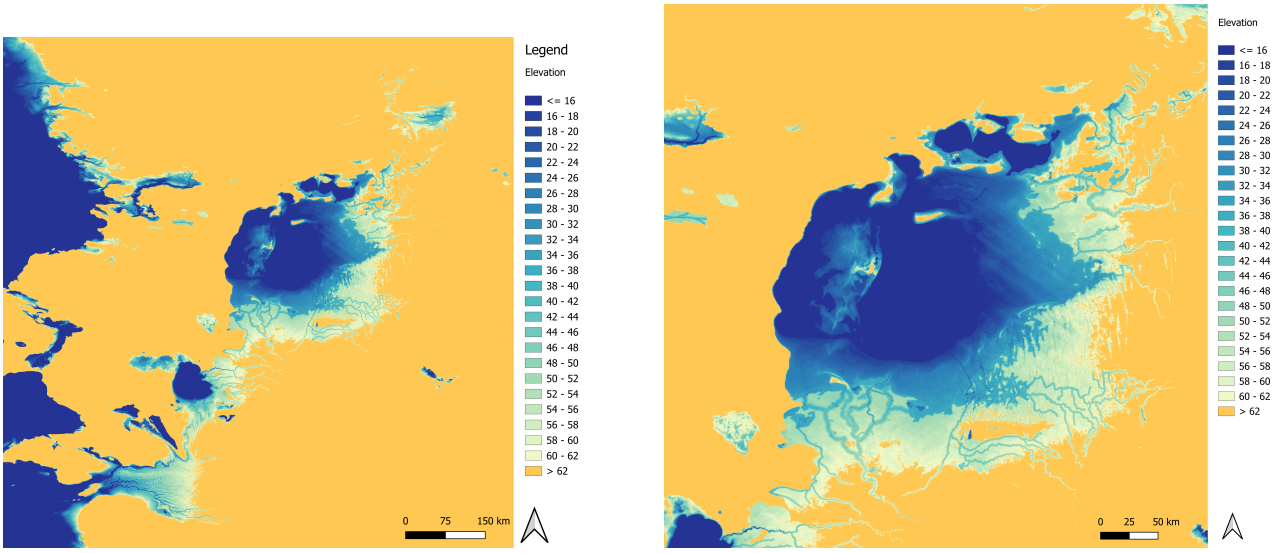


Figure B.2: placeholder ahv relation

Figure B.2 shows a comparison between the generic H–A–V relationship from Section B.1 and the Aral-specific relationships, highlighting discrepancies and potential limitations of the global dataset.

B.3. max extent



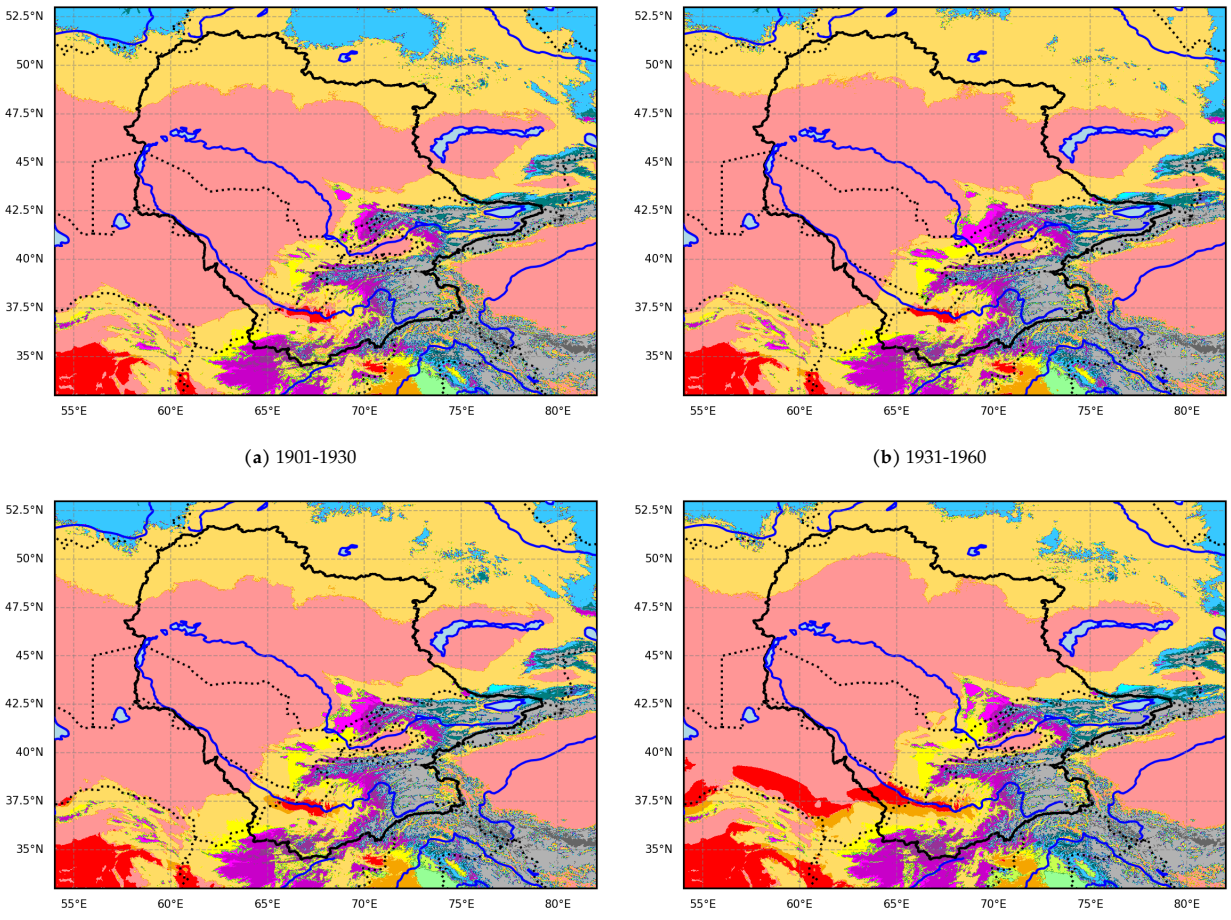
(**a**) Max extent Aral Sea

(**b**) Max extent Aral Sea zoomed in

Figure B.3: Max extent Aral Sea combined

C

Köppen-Geiger



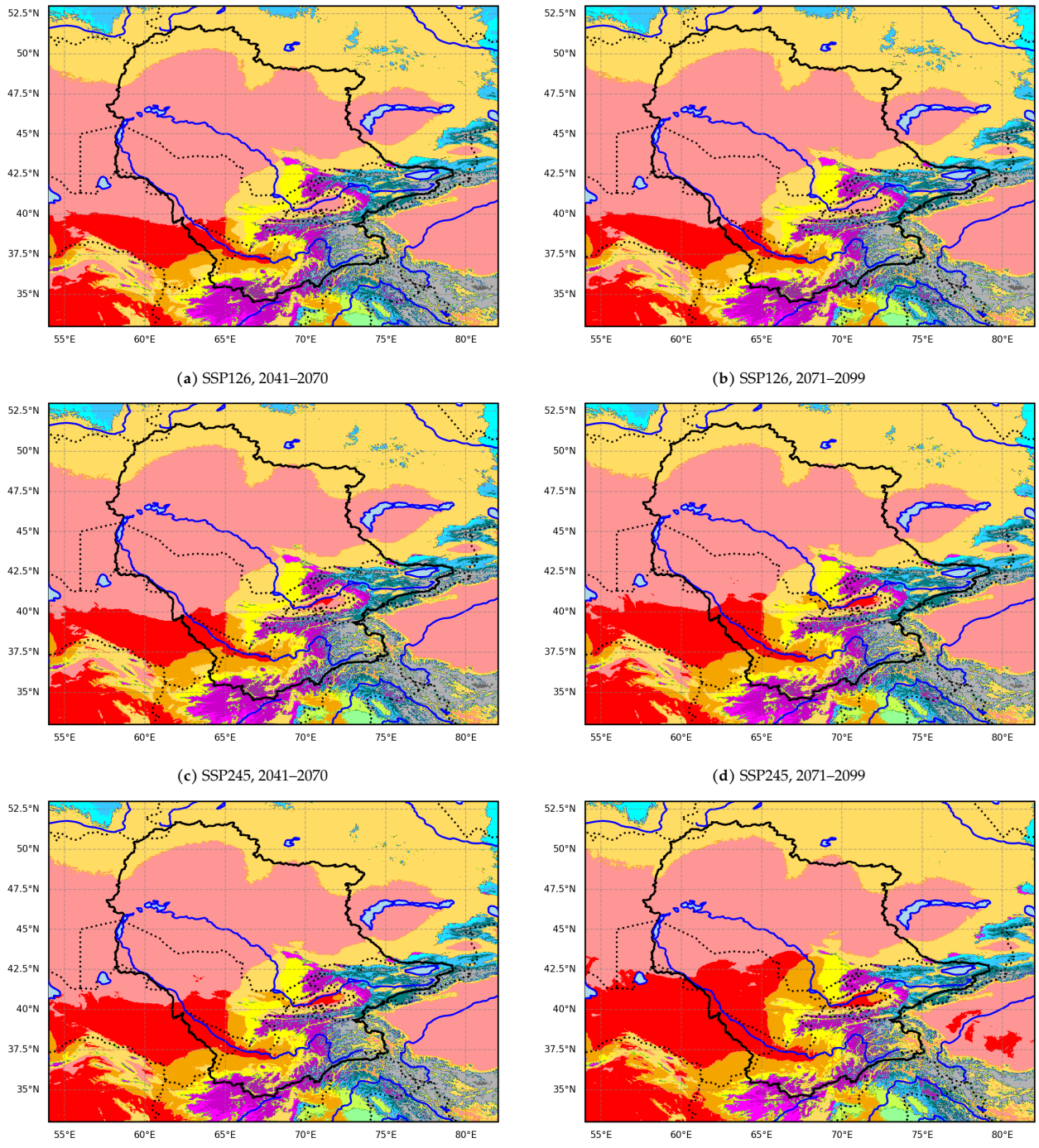
(**a**) 1901-1930

(**b**) 1931-1960

(**c**) 1961-1990

(**d**) 1991-2020)

Figure C.1: Köppen–Geiger climate classification of the Aral Sea Basin (black outline) for four historical time periods (1901–1930, 1931–1960, 1961–1990, and 1991–2020) at 1 km spatial resolution. The maps show the spatial distribution of Köppen–Geiger climate classes across the basin, with arid and semi-arid types prevailing over most areas and cold mountain climates occurring primarily at higher elevations in the southeastern part of the basin. Based on the high-resolution dataset by Beck et al. (Beck et al., 2023).



(**a**) SSP126, 2041–2070

(**b**) SSP126, 2071–2099

(**c**) SSP245, 2041–2070

(**d**) SSP245, 2071–2099

(**e**) SSP585, 2041–2070

(**f**) SSP585, 2071–2099

Figure C.2: Projected Köppen–Geiger climate classification of the Aral Sea Basin (black outline) for three SSP scenarios representing low, medium, and high greenhouse gas emissions, and for two future time periods, at 1 km spatial resolution. The scenarios shown are SSP1-2.6 (low emissions), SSP2-4.5 (medium emissions), and SSP5-8.5 (high emissions). Relative to the historical reference period (1961–1990), the projections indicate an increase in the spatial extent of hot desert (BWh) and hot arid (BWk) climate classes (red and dark orange), with larger extents under higher emission scenarios. Based on the high-resolution dataset by Beck et al. (Beck et al., 2023).

D

Computing Infrastructure

Virtual machines were made available through the eWaterCycle project on the SURF Research Cloud. Two environments were used in this study: one for general tasks such as development, testing, data processing, and visualisation, and a second environment dedicated to computationally intensive simulations.

Table D.1: Computational environment comparison of the two virtual machines used in this study.

Component	**Machine 1**	**Machine 2**
Virtualization platform	OpenStack (KVM)	OpenStack (KVM)
Operating system	Ubuntu 22.04.5 LTS	Ubuntu 22.04.5 LTS
Kernel	5.15.0-177-generic	5.15.0-176-generic
Architecture	x86_64	x86_64
CPU model	AMD EPYC-Rome	AMD EPYC-Milan
vCPUs allocated	4	15
Threads per core	1	1
NUMA nodes	1	1
Memory (RAM)	32 GiB	180 GiB
Storage	250 GB, XFS	1 TB, XFS
Host environment	RDO OpenStack Compute	RDO OpenStack Compute

E

Usage of the HBV Model

E.1. Model Calibration Workflow

The Hydrologiska Byråns Vattenbalansavdelning (HBV) model was utilized primarily to develop and validate the initial calibration workflow for this study. While the primary analysis shifted to PCR-GLOBWB and HBV was no longer used for the analysis, this implementation details the steps used during the testing phase. This material is preserved in this appendix to serve as a practical reference and methodological framework for future calibration tasks.

E.1.1. Lumped Model calibration

Based on earlier work within the eWaterCycle framework (Hoogelander, 2022), the HBV model was calibrated using Covariance Matrix Adaptation Evolution Strategy (CMA-ES) (Hansen, 2006) with the Python implementation provided by (Hansen et al., 2019). CMA-ES is a numerical, population-based evolutionary optimization algorithm used to estimate the parameters of the HBV model, as described in Section 3.2.2.

The algorithm starts from an initial guess and iteratively generates a population of candidate parameter sets. Each candidate set is used to run the HBV model, and the simulated output is compared to observations using the objective function J (Eq. E.1). Based on their performance, the best parameter sets are used to generate a new combination of parameter sets in the next iteration. This process continues until convergence criteria are met or a set of maximum function evalution is reached.

The objective function J was defined as a weighted combination of three performance metrics:

$$J = w_{\text{NSE}} \left(1 - \text{NSE}\right) + w_{\text{KGE}} \left(1 - \text{KGE}_{2009}\right) + w_V \, E_V \tag{E.1}$$

with weights $w_{\text{NSE}} = 0.3$, $w_{\text{KGE}} = 0.3$, and $w_V = 0.4$.

The NSE (Nash and Sutcliffe, 1970) was computed as:

$$\text{NSE} = 1 - \frac{\sum_{t=1}^{T} \left(Q_{\text{sim},t} - Q_{\text{obs},t}\right)^2}{\sum_{t=1}^{T} \left(Q_{\text{obs},t} - \overline{Q}_{\text{obs}}\right)^2} \tag{E.2}$$

The Kling–Gupta Efficiency (KGE; Gupta et al., 2009) defined as:

$$\text{KGE}_{2009} = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \tag{E.3}$$

with

$$r = \text{corr}(Q_{\text{sim}}, Q_{\text{obs}}), \quad \alpha = \frac{\sigma_{\text{sim}}}{\sigma_{\text{obs}}}, \quad \beta = \frac{\mu_{\text{sim}}}{\mu_{\text{obs}}}. \tag{E.4}$$

A volume error term E_V was included to guide the model towards reproducing the long term water balance:

$$E_V = \left| 1 - \frac{\sum_{t=1}^{T} Q_{\text{sim},t}}{\sum_{t=1}^{T} Q_{\text{obs},t}} \right|. \tag{E.5}$$

While NSE and KGE_{2009} assess the hydrograph shape, the volume error E_V specifically helps to guide the total cumulative simulated discharge to match observations. This is particularly important in endorheic basins, where water does not exit the system via rivers. Even small differences in yearly cumulative discharge accumulate over time, leading to large errors in for example lake levels.

Combining these three metrics in the objective function helps reduce equifinality, as parameter sets must perform well across multiple aspects of hydrological behaviour. The best performing parameter sets from CMA-ES runs are then used to create an ensemble of simulations for further analysis.

GRDC station	River Name	Calibration Period	Validation Period
Chatly	Lower Amu Darya	1950–1954	1955–1959
Kerki	Upper Amu Darya	1950–1954	1955–1959
Tyumen Aryk	Syr Darya	1950–1954	1955–1959

Table E.1: Overview of calibration and validation periods for the main rivers and catchments in the Aral Sea basin.

Meteorological forcing for the calibration was derived from the ERA5 reanalysis dataset (see Section 3.4.1).

The calibration, and validation focused on a period with sufficient streamflow observations (GRDC: The global runoff data centre, n.d.) and prior to large-scale agricultural interventions, which was used to define a baseline representing the catchment's natural hydrological behaviour (Table E.1).

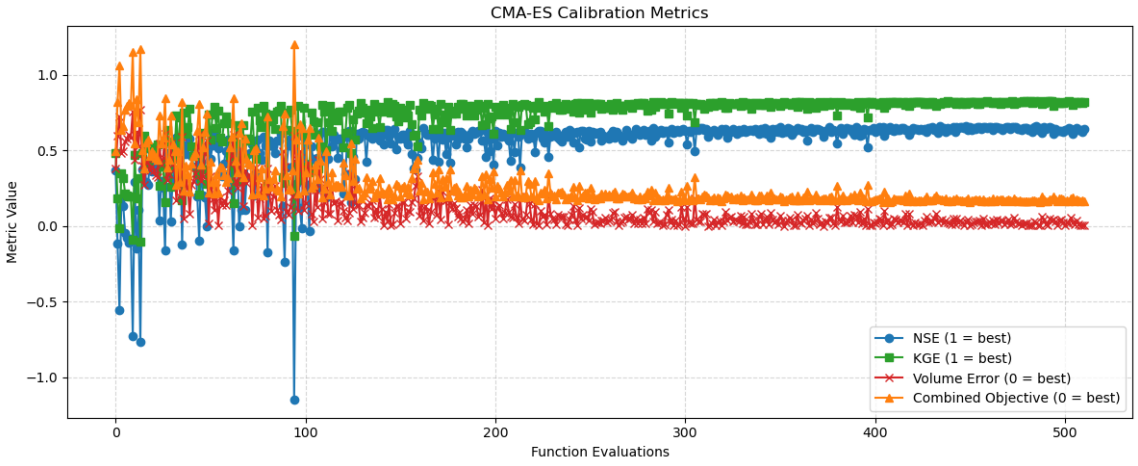


Figure E.1: Image of CMA-ES HBV parameter calibraiton

This ensemble-based CMA-ES approach enables robust parameter estimation while explicitly acknowledging uncertainties arising from model structure, forcing data, and the equifinality of hydrological model solutions.

E.1.2. HBV calibration

The HBV model was calibrated using the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) from (Hansen, 2006) following the procedure described in section E.1.1 for the basins selected in section. The optimisation procedure typically showed rapid improvement of the objective function during the initial iterations, followed by progressively slower convergence as the algorithm approached an optimal parameter region.

Figure E.2 illustrates the typical convergence behaviour. Most performance gains are achieved within the first three iterations, indicating efficient exploration of the parameter space, while later iterations primarily refine parameter estimates.

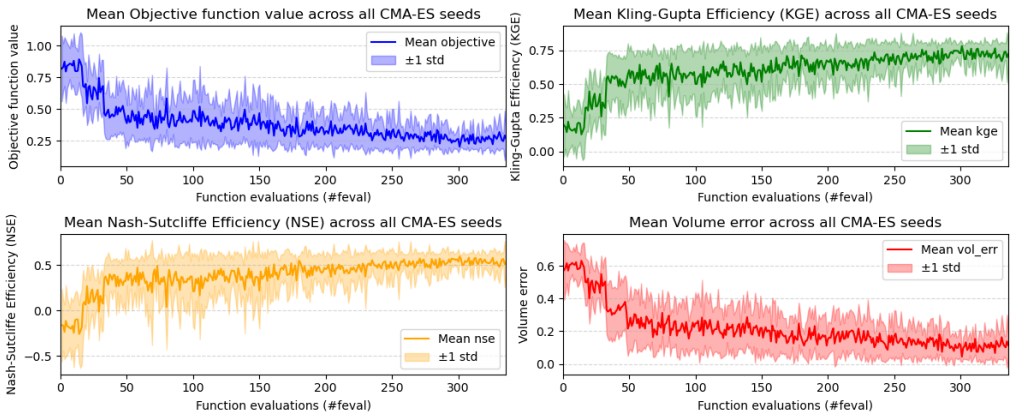


Figure E.2: Evolution of the objective function during multi particle CMA-ES optimisation for a single location. Rapid convergence occurs during the first iterations, after which improvements become gradual. Top left the objective functions and the other three panels show the components of the objective function described section E.1.1

The calibration performance was evaluated using the multi-objective function described in Section E.1.1, combining NSE, KGE, and a volume error term.

Table E.2 shows the calibration performance obtained for the different river basins.

Basin	NSE	KGE	Volume error	Combined score
Basin A	0.82	0.79	0.12	0.80
Basin B	0.76	0.74	0.11	0.75
Basin C	0.88	0.85	0.22	0.86
Basin D	0.71	0.69	0.33	0.70

Table E.2: Placeholder overview of HBV calibration performance using NSE, KGE, and volume error metrics. The combined score represents the objective function used during CMA-ES optimisation.

The distribution of calibrated parameter values across multiple optimisation seeds is shown in Figure E.3, while summary statistics are provided in Table E.3. The results indicate varying levels of parameter identifiability.

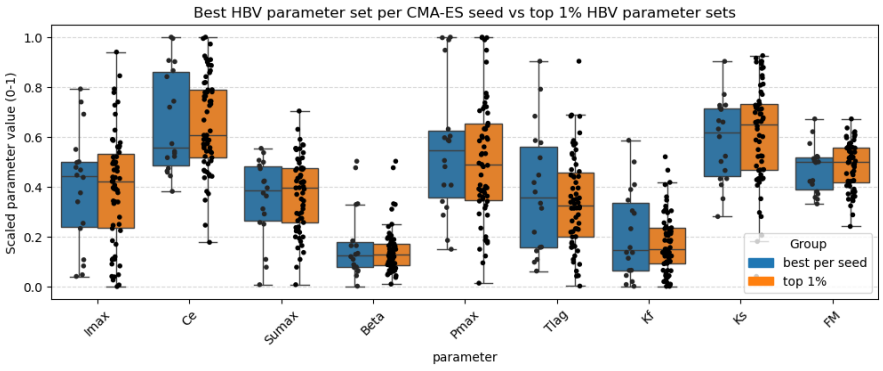


Figure E.3: Distribution of HBV parameter sets obtained from multiple independent CMA-ES calibration runs (shown in normalised parameter space). Parameters such as β and FM show strong convergence, whereas a wide range of acceptable parameter values is observed for P_{max} and T_{lag}.

Table E.3: Summary statistics of calibrated parameters

Statistic	Imax	Ce	Sumax	Beta	Pmax	Tlag	Kf	Ks	FM
mean	9.845	0.728	307.824	1.097	0.167	6.367	0.012	0.006	0.382
std	5.757	0.168	122.513	0.506	0.082	3.529	0.002	0.002	0.076
min	1.035	0.506	46.269	0.508	0.046	1.881	0.010	0.003	0.272
25%	5.964	0.588	242.176	0.771	0.108	3.212	0.011	0.004	0.317
50%	11.064	0.646	333.193	0.939	0.164	6.001	0.011	0.006	0.404
75%	12.503	0.888	406.333	1.127	0.188	8.867	0.013	0.007	0.420
max	19.808	1.000	461.243	2.259	0.300	13.647	0.016	0.009	0.541

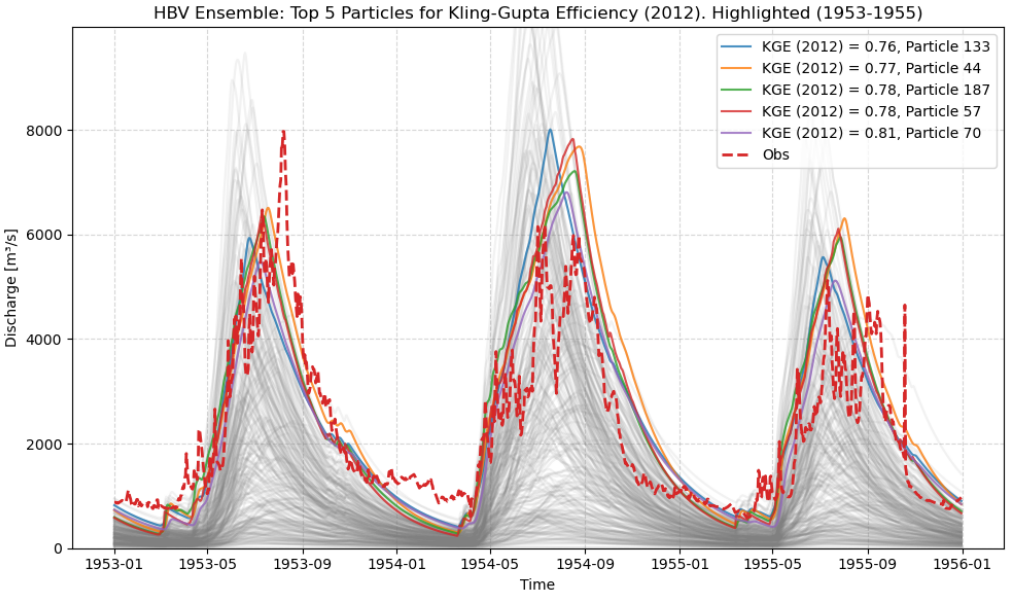


Figure E.4: Results of best HBV calibration

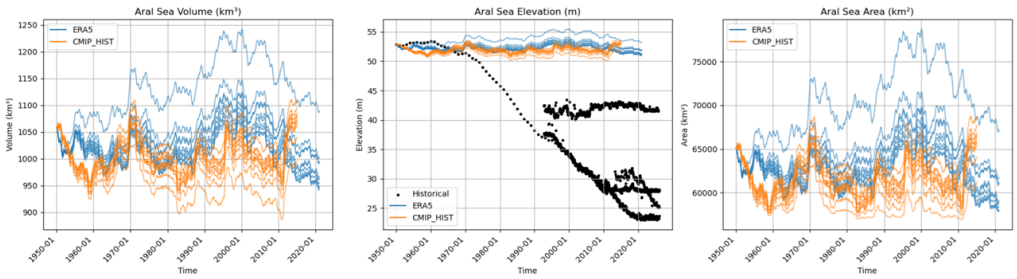


Figure E.5: Results of best HBV calibration on Aral sea water levels

F

FAIR

Table F.1: compliance to FAIR guidelines from De Visser et al. (2023)

#	Theme	FAIR Guideline	Implementation	Status
1	Findability	Register the workflow	WorkflowHub	Not yet
2	Findability	Describe the workflow with rich metadata	`CITATION.cff` and `README`	Somewhat
3	Accessibility	Make source code available in a public code repository	Whole project published on GitHub with a Zenodo DOI	In progress
4	Accessibility	Provide example input data and results along with the workflow	Raw input data not included due to redistribution restrictions; detailed sources in `README` and tutorial notebooks	Mostly there
5	Interoperability	The tools integrated in a workflow should adhere to file format standards	Core data files are `.nc` and `.csv`; some auxiliary files may use other formats (e.g. `.pickle` used during testing)	Yes
6	Interoperability	Make the workflow portable	Via Snakemake + `env.yaml`	In progress
7	Reusability	Provide a reproducible computational environment to run the workflow	`env.yaml` provided	In progress
8	Reusability	Add a configuration file with defaults	`config.yaml`	Done (needs check)
9	Reusability	Modularize the workflow	Automatically from Snakemake; calibration is the only current exception	Mostly there
10	Reusability	Provide clear and concise workflow documentation	Jupyter notebooks and READMEs	Needs expansion

G

Calibration Validation

Table G.1: Monthly KGE scores for calibration and validation periods across parameter sets.

Station	Period	Uncalibrated	Overall cal.	Amu Darya cal.	Syr Darya cal.
Chatly	Calibration	−2.94	0.28	0.82	−0.15
Chatly	Validation	−2.66	0.25	0.82	−0.20
Kerki	Calibration	−0.02	0.13	0.34	−0.12
Kerki	Validation	−0.09	0.15	0.36	−0.05
Tyumen-aryk	Calibration	−7.70	0.57	−0.19	0.42
Tyumen-aryk	Validation	−8.95	0.39	−0.52	0.45
Bekabad	Calibration	−6.43	0.68	0.13	0.54
Bekabad	Validation	−11.29	0.09	−0.99	0.18
Kal	Calibration	−2.73	0.27	0.58	0.11
Kal	Validation	−3.43	0.36	0.60	0.16
Uch-kurgan	Calibration	−1.16	0.25	0.41	0.12
Uch-kurgan	Validation	−1.36	0.27	0.46	0.09
Kazalinsk	Calibration	−15.41	0.06	−1.92	0.50
Kazalinsk	Validation	−15.13	−0.08	−1.80	0.48
Karaozek	Calibration	−24.05	−1.09	−4.09	0.00
Karaozek	Validation	−23.30	−1.34	−3.88	−0.24
Ust. Kekirim	Calibration	−0.95	0.20	0.29	0.11
Ust. Kekirim	Validation	−1.02	0.12	0.24	0.00
Tutkaul	Calibration	0.21	−0.17	−0.20	−0.24
Tutkaul	Validation	0.09	−0.16	−0.21	−0.26
Garm	Calibration	0.06	−0.23	−0.26	−0.23
Garm	Validation	−0.03	−0.24	−0.29	−0.26
Barchadiv	Calibration	−28.16	−2.88	−5.21	−2.41
Barchadiv	Validation	−30.94	−3.53	−6.30	−2.57
Khorog	Calibration	−1.67	0.86	0.49	0.67
Khorog	Validation	−0.96	0.88	0.68	0.75
Manguzar	Calibration	−4.15	0.29	−0.22	0.20
Manguzar	Validation	−5.11	−0.01	−0.58	−0.16

Continued on next page...

...continued from previous page

Station	Period	Uncalibrated	Overall cal.	Amu Darya cal.	Syr Darya cal.
Hodjikent	Calibration	−0.71	0.39	0.60	0.31
Hodjikent	Validation	−0.31	0.34	0.55	0.24
Naryn	Calibration	0.45	−0.03	−0.07	−0.05
Naryn	Validation	0.63	−0.11	−0.15	−0.14
Dupuli	Calibration	−0.88	0.43	0.69	−0.05
Dupuli	Validation	−0.72	0.43	0.69	0.01
Tartki	Calibration	0.32	0.35	0.39	0.34
Tartki	Validation	0.25	0.42	0.43	0.46
Ust. Djumgol	Calibration	−0.99	0.33	0.42	0.23
Ust. Djumgol	Validation	−0.66	0.23	0.34	0.14
Chirakchi	Calibration	−8.20	−1.41	−1.81	−1.14
Chirakchi	Validation	−10.41	−2.35	−2.67	−2.19
Ustie	Calibration	−7.74	0.38	−0.22	0.65
Ustie	Validation	−6.34	0.50	0.05	0.48
Khabost	Calibration	−1.21	0.56	0.86	0.52
Khabost	Validation	−0.75	0.52	0.79	0.53
Chalma	Calibration	0.68	−0.31	−0.23	−0.36
Chalma	Validation	0.61	−0.35	−0.24	−0.40
Chinor	Calibration	0.33	0.11	−0.09	0.07
Chinor	Validation	0.25	0.13	−0.12	0.14
Gulcha	Calibration	0.43	−0.39	−0.26	−0.42
Gulcha	Validation	0.24	−0.50	−0.37	−0.51
Obizarang	Calibration	−3.45	−0.15	−0.26	−0.38
Obizarang	Validation	−4.20	−0.52	−0.58	−0.85
Sarykanda	Calibration	0.62	0.59	0.68	0.41
Sarykanda	Validation	0.62	0.72	0.62	0.52
Tuleken	Calibration	0.29	−0.28	−0.14	−0.40
Tuleken	Validation	0.17	−0.29	−0.13	−0.52
Uch-korgon	Calibration	−1.03	0.53	0.54	0.42
Uch-korgon	Validation	−0.46	0.56	0.60	0.25
Motravn	Calibration	−0.58	−0.71	−0.63	−0.78
Motravn	Validation	−0.55	−0.66	−0.57	−0.81
Andarhan	Calibration	−1.81	0.25	0.56	−0.16
Andarhan	Validation	−2.30	0.45	0.61	−0.07
Ust. Karakol	Calibration	0.11	0.15	0.36	0.01
Ust. Karakol	Validation	−0.19	0.39	0.52	0.08
Tash-kurgan	Calibration	0.28	0.02	0.22	−0.03
Tash-kurgan	Validation	0.33	−0.03	0.26	−0.10
Charvak	Calibration	0.07	−0.38	−0.34	−0.40
Charvak	Validation	0.11	−0.38	−0.33	−0.41
Paulgan	Calibration	−5.35	−1.93	−2.40	−1.66
Paulgan	Validation	−6.96	−2.97	−3.37	−2.78
Dagana-ata	Calibration	0.47	0.32	0.21	0.42

Continued on next page...

...continued from previous page

Station	Period	Uncalibrated	Overall cal.	Amu Darya cal.	Syr Darya cal.
Dagana-ata	Validation	0.46	0.35	0.16	0.50
Salamalik	Calibration	0.66	−0.27	−0.22	−0.34
Salamalik	Validation	0.57	−0.28	−0.25	−0.36
Sudgina	Calibration	−4.00	−1.34	−1.92	−1.80
Sudgina	Validation	−4.13	−1.71	−1.66	−2.10
Mikhailovskoe	Calibration	0.55	−0.15	−0.04	−0.21
Mikhailovskoe	Validation	0.59	−0.17	−0.07	−0.28
Aktash	Calibration	−0.46	−0.11	−0.01	−0.17
Aktash	Validation	−0.37	−0.21	−0.09	−0.28
King Guzar	Calibration	−14.35	−4.79	−5.25	−5.20
King Guzar	Validation	−15.17	−5.52	−5.73	−6.25
Khazarnova	Calibration	−5.59	−1.57	−2.25	−1.84
Khazarnova	Validation	−5.18	−1.50	−1.95	−1.83
Dazgon	Calibration	−6.99	−3.96	−4.18	−6.59
Dazgon	Validation	−5.47	−4.39	−4.08	−7.54
Karatag	Calibration	0.36	−0.10	−0.04	−0.22
Karatag	Validation	0.34	−0.02	0.05	−0.20
Ust. Kairagach	Calibration	−0.01	−0.35	−0.29	−0.35
Ust. Kairagach	Validation	−0.20	−0.42	−0.38	−0.37
Varganza	Calibration	−3.81	0.17	−0.05	0.43
Varganza	Validation	−4.76	−0.14	−0.24	0.09
Changet	Calibration	−0.37	−0.46	−0.50	−0.62
Changet	Validation	−0.38	−0.44	−0.60	−0.51
Ust. Tostu	Calibration	0.68	−0.44	−0.21	−0.40
Ust. Tostu	Validation	0.65	−0.43	−0.26	−0.40
Alibegi	Calibration	0.31	0.58	0.38	0.67
Alibegi	Validation	0.26	0.59	0.28	0.58
Donguztay	Calibration	−0.06	−0.50	−0.46	−0.50
Donguztay	Validation	−0.08	−0.41	−0.41	−0.45