Exploring the impact of the recent global warming on extreme weather events in Central Asia using the counterfactual climate data ATTRICI v1.1

Bijan Fallah[1] · Masoud Rostami[1,2]

Abstract

We study the impact of recent global warming on extreme climatic events in Central Asia (CA) for 1901-2019 by comparing the composite representation of the observational climate with a hypothetical counterfactual one that does not include the long-term global warming trend. The counterfactual climate data are produced based on a simple detrending approach, using the global mean temperature (GMT) as the independent variable and removing the long-term trends from the climate variables of the observational data. This trend elimination is independent of causality, and the day-to-day variability in the counterfactual climate remains preserved. The analysis done in the paper shows that the increase in frequency and magnitude of extreme temperature and precipitation events can be attributed to global warming. Specifically, the probability of experiencing a +7 K temperature anomaly event in CA increases by up to a factor of seven in some areas due to global warming. The analysis reveals a significant increase in heatwave occurrences in Central Asia, with the observational climate dataset GSWP3-W5E5 (later called also factual) showing more frequent and prolonged extreme heat events than hypothetical scenarios without global warming. This trend, evident in the disparity between factual and counterfactual data, underscores the critical impact of recent climatic changes on weather patterns, highlighting the urgent need for robust adaptation and mitigation strategies. Additionally, using the self-calibrated Palmer drought severity index (scPDSI), the sensitivity of dry and wet events to the coupled precipitation and temperature changes is analyzed. The areas under dry and wet conditions are enhanced under the observational climate compared to a counterfactual scenario, especially over the largest deserts in CA. The expansion of the dry regions aligns well with the pattern of desert development observed in CA in recent decades.

✉ Bijan Fallah
fallah@pik-potsdam.de

1 Potsdam Institute for Climate Impact Research (PIK), Telegrafenberg A62, D-14473 Potsdam, Brandenburg, Germany

2 Laboratoire de Météorologie Dynamique (LMD), Sorbonne University (SU), Ecole Normale Supérieure (ENS), CNRS, Paris, France

Springer

1 Introduction

The recent climate warming is contributing to extreme climate events in Central Asia (CA), a region which is facing concerns about the water availability, hydropower, and food security (Fallah et al. 2023a). The selected domain in this study covers five former Soviet Union countries in CA: Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan. This is the definition of the CA in the Green Central Asia project funded by the German Federal Foreign Office (https://greencentralasia.org/en, last visited 28.Aug.2023). Those five countries consume 85 to 97% of the withdrawn water for agriculture (Unger-Shayesteh et al. 2013). Therefore, water scarcity has been one of the significant challenges in CA in recent years (Siegfried et al. 2010). Additionally, the arid desert climate in CA has expanded up to 100 kilometers to the North since 1980 (Guglielmi 2022; Hu and Han 2022). Unfortunately, the most crucial cause of the recent water shortage in CA is the water management issues like the diversion of the rivers, ineffective irrigation leading to water wasting, and dams' reconstruction (Micklin 2007; Pala 2005). Additional to the management impact, there are shreds of evidence of climate impact in the water resource shortage, like less snow during the cold seasons (Jiang and Zhou 2021; Hu et al. 2014).

Temperature reconstruction of the last Millennium in CA shows that the recent warming is the highest on record (Davi et al. 2015). It has been shown that under a climate with increased greenhouse gases, the subtropical westerly jet weakens and moves southward, causing a water deficit in the hydrological cycle (Jiang and Zhou 2021). Recent studies have shown a significant rise in the frequency and magnitude of extreme events (floods, droughts, and heatwaves) in CA in recent decades, primarily based on precipitation and temperature changes (Zhou and Huang 2010; Zhang et al. 2017; Hu et al. 2016; Dilinuer et al. 2021; Peng et al. 2020). However, the attribution research on the magnitude of extreme events in CA to the global warming is still in its infancy (Zou et al. 2021).

With the emergence of more frequent extreme weather events in recent decades, event attribution studies gained lots of attention (van Oldenborgh et al. 2021). One of the main challenges of attribution studies is identifying the impact of climate change on humans and the environment (Allan et al. 2021). Therefore, one needs a climate without human-induced changes as the baseline state for identifying and quantifying the associated climate impacts. Two methodologies exist for producing a so-called counterfactual climate: 1- using a large ensemble of general circulation models (GCMs) driven by the natural drivers only (i.e., volcanic and solar forcings), usually labeled as the *histNat* family in the Coupled Model Intercomparison Project (Gillett et al. 2016) and 2- applying empirical models to eliminate the climate change trend from the observational climate dataset. The former focuses on attributing the observed changes to anthropogenic emissions (Zou et al. 2021) and the latter on eliminating the factual climate trends independently of climate change drivers. Method 1, for example, allows us to identify the impact of anthropogenic climate change on any extreme individual event like floods, landslides, and heatwaves (Otto 2017; Lewis and Karoly 2015).

In this study we use a counterfactual data generated by method 2. Given that the detrending of the variables is done independently in method 2, it can not guarantee the preservation of the physical consistency between the atmospheric variables (Mengel et al. 2021). The variables are detrended independently based on their statistical relationships to the global mean temperature (GMT). However, the ranks are preserved in the variables, i.e., relatively high values in the factual climate stay highly ranked after the detrending. Therefore, the risk of physical inconsistency is reduced, which is crucial for extreme event attribution studies.

Although the impacts of future climate change might be very vital in CA, there exist very few studies focusing on this topic (Fallah et al. 2023a). By employing bias-adjusted Coupled Model Intercomparison Project phase 5 (CMIP5) simulations, it has been shown that the extreme climate events in CA will increase robustly in frequency and magnitude under future warming scenarios (Peng et al. 2020). The main motivation of this paper is to quantify the impact of climate change on extreme weather events in CA using a counterfactual data generated by method 2.

By definition, the impact of climate change on a single event is "detected" when that event only exists in the factual climate and not in a counterfactual baseline state (Liverman 2008). Following this definition (method 2, non-probabilistic approach), the frequency and magnitude of observed events which could be attributed to global warming are detected. By comparing the factual climate state (obsclim) and the counterfactual one (counterclim), the local dependencies of the observed local extreme events on the GMT are studied.

Here, the observational climate dataset GSWP3-W5E5 (later called also factual) and a detrended climate dataset produced by ATTRICI v1.1 method as the counterfactual climate (Mengel et al. 2021) are used and the following scientific questions are answered:

I) How does the probability of extreme temperature and precipitation events change under the factual (or observational) climate in CA?
II) What are the local impacts of the recent climate change in CA?
III) To what extent are CA drying/wetting trends linked to global warming?

We hypothesize that the recent global warming has increased the frequency and magnitude of extreme temperature and precipitation events in Central Asia. This hypothesis is based on the observed trends in global mean temperature increases and their documented impacts on regional climate patterns. By comparing factual climate data with counterfactual scenarios that exclude anthropogenic warming, we aim to quantify the specific contribution of global warming to the changing dynamics of extreme weather events in the region.

To our knowledge, most studies investigating the past climate change impacts in CA are based on comparing the trends between a reference climate (far past) and the recent climate (Unger-Shayesteh et al. 2013; Fallah et al. 2023a). Here, the factual climate state is compared against an alternative without long-term global warming in both single and multivariate frameworks. Furthermore, analyzing the quality of the detrended counterfactual data for any limited area has been suggested before applying it in any climate impact study (Mengel et al. 2021). Therefore the analysis presented here will pave the path for further impact modeling studies using more complicated process-based impact models in CA.

2 Data and methods

2.1 Study domain

Central Asia stretches from the Caspian Sea in the west to China in the east, and from Russia in the north to Afghanistan and Iran in the south, and includes five countries: Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan and Uzbekistan. Figure 1 shows the rectangular domain selected for this study. As can be seen, a large portion of the CA has an elevation of less than 500 m, and it also includes high passes and mountains (Tian Shan, Pamir). Tajikistan and Kyrgyzstan are the countries with higher elevations, and they provide water resources for the lowland areas in other countries. Some of the world's driest deserts are in CA, like Karakum desert in Turkmenistan, Kyzylkum desert in Uzbekistan and Kazakhstan, and Moiynkum and
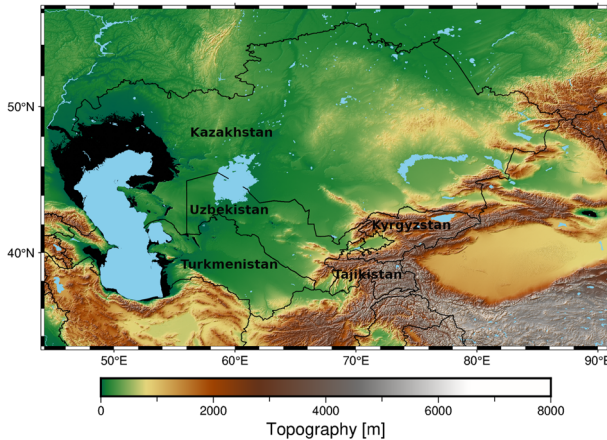
Fig. 1 Study region over Central Asia and the topography (m). Note that values under zero are shown in black around the Caspian Sea

Betpak-Dala deserts in eastern Kazakhstan. Around 60% of CA is covered by an arid climate zone and, therefore, is vulnerable even to tiny deviations of yearly precipitation.

2.2 Climate and other data and software

In this study, the observational climate dataset GSWP3-W5E5 was used as factual climate data, and a detrended climate dataset obtained from it by applying the ATTRICI v1.1 method was used as a counterfactual climate for CA.

It is crucial to elucidate that the GSWP3-W5E5 dataset, referenced as *obsclim* or *factual* in this study, is a synthesis that extends beyond direct observations from meteorological stations. Specifically, it incorporates a blend of observational data, reanalysis outputs, and modelling efforts to generate a global coverage of climate variables. The GSWP3-W5E5 dataset is segmented into two periods: for 1979-2019, W5E5 v2.0 (Lange 2021) is utilized, whereas for 1901-1978, GSWP3 v1.09 is harmonized with W5E5 (see details at https://doi.org/10.48364/ISIMIP.342217). For this amalgamation, the ISIMIP3BASD v2.5.0 bias adjustment method was applied for the earlier period (Lange 2019), aiming in refining the dataset for enhanced reliability and consistency.

By employing the GSWP3-W5E5 dataset as a baseline (obsclim), this study acknowledges its role as a composite representation of the climate, synthesizing observed data, reanalyses, and model simulations to provide a global perspective on climate variables. However, as suggested by the authors, for regional usage, the control plots have to be checked, adjusted, and analyzed regionally before any climate impact investigation.

For developing counterfactual climate data, the long-term climate trend was removed from the climate dataset GSWP3-W5E5, while preserving the intrinsic day-to-day variability. Creating the Phase 3 Inter-Sectoral Impact Model Inter-comparison Project (ISIMIP3) counterfactual data for attributing climate impacts to global warming leverages the ATTRICI v1.1 methodology. This detrending method transcends simple linear regression by allowing for the randomness of day-to-day variability and considering the Global Mean Temperature (GMT) as the independent variable instead of chronological time. This approach permits the accommodation of unexplained random variabilities annually to follow a non-normal dis-

tribution. Consequently, the day-to-day fluctuations are maintained smoothly over time and space, making this dataset suitable forcing data for impact models within ISIMIP3a (Mengel et al. 2021).

The ATTRICI v1.1 counterfactual climate data (counterclim) is accessible online at https://doi.org/10.5281/zenodo.5036364. A comprehensive description of the data creation methodologies and their applications is provided in the companion paper by Mengel et al. (2021).

Other data used in this study are as follows. The global annual mean CO2 data is downloaded via https://www.gml.noaa.gov. The available water holding capacity of the soil (AWC) dataset is downloaded from https://daac.ornl.gov (last visited 29.08.2022). The Global Landslide Catalog is available at https://data.nasa.gov/Earth-Science/Global-Landslide-Catalog-Not-updated-/h9d8-neg4. The Global Fatal Landslide Database is freely available at https://svs.gsfc.nasa.gov/4710. The list of datasets used in this study is shown in Table 1.

For manipulation of the NetCDF files, we use CDO (version 2.0.3; https://mpimet.mpg.de/cdo) and NCO (version 5.0.6). For data analysis and plotting, we use the jupyterhub at German Climate Computing Center (DKRZ) and the following python3.9.9 packages: basemap, dask, matplotlib, numpy, rasterio, earthpy, IPython, and joblib. For finding the best fit for the histograms, we use the Gaussian Kernel-Density Estimate (KDE) in a non-parametric way via the scipy.stats.gaussian_kde python library. The Python code for comparing the factual and station observations is available at https://github.com/bijanf/get_station_data/tree/main.

2.3 Method: Heat Wave Duration Index (HWDI)

The Heat Wave Duration Index (HWDI) is a critical measure in understanding and quantifying the frequency and intensity of heat waves. The methodology for calculating HWDI involves several key steps, each designed to ensure that the index accurately reflects the severity of heat waves over a given period and location. Here is a breakdown of the process:

- Calculation of Climatological Daily Maximum Value: The climatological daily maximum value is calculated as the average maximum temperatures for each calendar day over a 30-year reference period (1901-1930) and each grid cell. For example for any given day,

Table 1 Datasets used in this study

Data	Time range	Horizontal resolution	Variables
Carbon Dioxide Data	1979-2021	global time-series	CO_2 (ppm)
Available Water holding Capacity of the soil	1950-1995	0.5°	AWC (mm water per 1 m soil depth)
Global Landslide Catalog	2007-2018	–	hazard events
ATTRICI V1.1	1901-2019	0.5°	daily total precipitation / daily near surface air temperature
GSWP3-W5E5	1901-2019	0.5°	daily total precipitation / daily near surface air temperature
Global Historical Climatology Network daily (GHCNd)	1901-2019	station data	daily near surface air temperature

like July 4th, we look at the highest temperatures on all the July $4^{th}s$ from 1901 to 1930, add them up, and then divide by the number of years, which is 30, to get the average high temperature for July 4th. We establish a robust baseline for day-to-day variability within and across seasons. For each dataset, factual and counterfactual, we consider the first 30 years (1901-1930) as a reference, during which anthropogenic warming was not that high.

- Selection of Baseline Temperature: The reference temperature acts as a climatological threshold. This threshold is often determined based on historical temperature data, allowing for identifying significant deviations indicative of heat waves. Our study defines this threshold as 7K above the 30-year average (1901-1930) of daily maximum temperatures, offering a standardized benchmark across different geographical regions. This is done for every grid cell of the dataset.
- Definition of Heat Wave: A heat wave is characterized by the number of consecutive days the temperature exceeds the reference temperature. Our methodology distinguishes heat waves based on duration, setting thresholds at 5, 10, 15, and 20 consecutive days. This differentiation allows for a nuanced analysis of heat wave patterns, from shorter events to prolonged extremes.
- HWDI Calculation: For each location and each duration threshold (n days), the HWDI is calculated for the whole duration (1901-2019) of the two datasets (factual and counterfactual) by counting the occurrences when the temperature exceeds the reference temperature for at least n consecutive days. This calculation yields a quantitative measure of heat wave frequency and duration.
- Creation of Difference Maps: The final step involves generating difference maps by subtracting the counterfactual HWDI values (representing a hypothetical scenario without specific climatic influences) from the factual HWDI values (based on actual climate data). Positive values on these maps indicate locations where the factual climate data exhibited more frequent heat waves compared to the counterfactual scenario. These maps are invaluable for visualizing how heat wave patterns vary across locations and under varying climatic conditions.

By comparing the HWDI difference maps across different duration thresholds, researchers can gain insights into the frequency and intensity of heat waves that may change. This methodology provides a detailed account of heat wave patterns and supports broader climate change research by highlighting areas of increased vulnerability to extreme heat events. The Python code for calculating the HWDI and detailed documentation are available for further reference and application in related research projects (Fallah 2021).

2.4 Method: drought index

The Palmer Drought Severity Index (scPDSI) is calculated using the climate indices in python libraries (Adams 2017). The original PDSI is a suitable index for quantifying droughts' severity in a region (Palmer 1965). It is based on a simple two-layer bucket water balance model that estimates the soil moisture anomaly from the precipitation and potential evapotranspiration data. The PDSI uses empirical constants to calculate the climatic coefficient and the duration factors, which are derived from the historical records of Kansas and Iowa in the US. However, these constants may not be representative of other regions with different climate characteristics, and thus limit the spatial comparability of the PDSI values. The scPDSI uses an automatic calibration methodology that adjusts the empirical constants with dynamically calculated values based on the local climate conditions (Wells et al. 2004). This makes the

scPDSI more suitable for comparing droughts between different climates. The formula for calculating the monthly scPDSI is as follows:

$$\text{scPDSI}_i = p * \text{scPDSI}_{i-1} + q * Z_i \tag{1}$$

where scPDSI_i is the scPDSI value for month i, p and q are weighting factors, and Z_i is the moisture anomaly index for month i, which is calculated as:

$$Z_i = K_i * log((PE_i + R_i + RO_i)/(P_i + L_i)) \tag{2}$$

where K_i is the climatic coefficient for month i, PE_i is the potential evapotranspiration for month i, R_i is the recharge to the upper soil layer for month i, RO_i is the runoff from the lower soil layer for month i, P_i is the precipitation for month i, and L_i is the loss from both soil layers for month i.

The climatic coefficient K_i is calculated as:

$$K_i = K'_i / \sum (D_j * K_j) \tag{3}$$

where K′ is a calibration factor that depends on the local climate conditions, D_j is the duration factor for month j, and K_j is the original climatic coefficient for month j.

The calibration factor K′ and the duration factors D_j are determined by using a self-calibrating procedure that ensures that the mean scPDSI value for a given calibration period is zero, and that the frequency distribution of scPDSI values matches a predefined pattern (Wells et al. 2004).

For the calculation of the self-calibrated Palmer Drought Severity Index (scPDSI), we utilized daily near-surface air temperature and total precipitation data from the ATTRICI V1.1 and GSWP3-W5E5 datasets, along with the Available Water Holding Capacity (AWC) dataset for the soil, ensuring a comprehensive representation of climatic variables essential for accurate drought assessment in the study region.

3 Results

3.1 Evaluation of factual dataset

The motivation for comparing factual climate datasets with observational data arises from the need to validate climate models and assess their accuracy in reproducing observed climate conditions. Observational data collected from weather stations across CA provide a ground truth that can be used to evaluate the performance of climate datasets derived from satellite observations, reanalysis products, or climate model outputs. Such comparisons are essential for identifying potential discrepancies and biases in the datasets, which, in turn, can inform improvements to climate modelling and forecasting efforts.

Before analysing the temperature extremes in CA, we compared the factual dataset with the observational station dataset within the domain for the period when more data was available (1979-2019). We use the Global Historical Climatology Network daily (https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily) and plot the probability density function (PDF) of temperature and precipitation for all available days for predefined elevation ranges of 0-1000, 1000-2000 and greater than 2000 meters (Fig. 2). Unfortunately, in CA many stations do not have continuous record of daily temperature and precipitation. Especially for precipitation, the number of stations having more than 20 years of data in CA region is very limited. Therefore, we have selected different thresholds for data
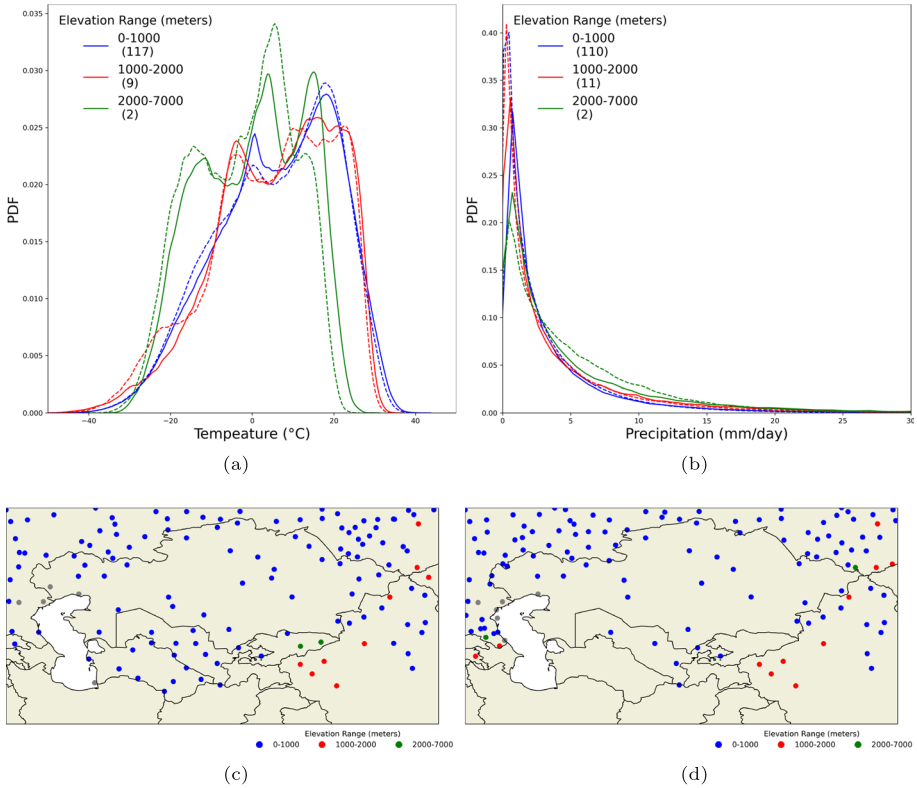
Fig. 2 Comparison of probability density functions (PDFs) of daily temperature (°C) and precipitation by elevation range (blue:0-1000m, red:1000-2000m and green 2000-7000m) for the whole study domain shown in Fig. 1 from the factual data (dashed lines) and observational stations (solid lines) shown in (a) and (b). Number of stations by elevation range for temperature and precipitation for the period of 1979-2019 are shown in the paranthesis. The geographical distribution of the stations for temperature and precipitation are shown in (c) and (d), respectively. The grey dots show the stations with unknown elevation

availability of precipitation and temperature station data. If more than 30 days in a year have missing values, we assign that year as missing. We ignore the stations with more than five (twenty) years of missing data in the 1979-2019 period for temperature (precipitation). We do not apply height correction to compare the gridded and station datasets. Accordingly, we find the nearest neighbour grid cell from the factual dataset to each chosen observation and plot the corresponding PDF for the elevation ranges.

Figure 2a and b show the related PDFs for temperature and precipitation from observations in solid lines and from factual data in dashed lines. We only selected the rainy days for precipitation, i.e., precipitation more than 0 mm/day. Our analysis shows that except for the stations located higher than 2000 m altitude, the PDFs of factual temperature follow well the ones estimated from the observed temperature at the stations. The estimated PDF of the factual climate dataset well represents the seasonality of the yearly temperature values.

For precipitation, the visible differences happen for values around 2 mm/day. At higher altitudes, more minor disagreements happen at precipitation ranges from ca. 8 to 20 mm/day. We must mention that the number of observations in CA reduces drastically with height, and most observations exist at lower altitudes (Fig. 2c and d). For precipitation, there exists

almost no proper station within the Central Asian countries if we limited the missing year's number to 5; therefore, we selected the stations having at least 20 years of data within the 1979-2019 period (see Fig. 2d).

For temperature, there are 117 stations with at least 35 years of data for the 1979-2019 period within the Central Asian countries in lowland (<1000 m), 9 stations for 1000-2000 m and only 2 for the 2000-7000 m elevation range. For precipitation, there are 110 stations with at least 20 years of data for the 1979-2019 period within the Central Asian countries in lowland (<1000 m), 11 stations for 1000-2000 m and only 2 for the 2000-7000 m elevation range.

The factual dataset can reproduce the distribution of the observed climate for the frequency of different temperature and precipitation values for the regions with higher station coverage. For every local analysis over higher elevations covering a small portion of the CA domain selected in this study, there might be significant biases in the factual dataset that must be corrected. However, our primary goal here is to compare the factual and counterfactual datasets and find differences among them on a large scale.

Generally, the factual dataset represents quite well the tails of the PDF of the temperatures of the observational dataset, where the extreme values might happen. We compare the factual and counterfactual climates for extreme temperature and precipitation events in the following.

3.2 Extreme near surface temperature events

Comparing the time-series of near surface air temperatures in CA from the counterfactual and factual climates, might provide insights on the effect of global warming. Figure 3a shows the evolution of the yearly average of daily near-surface air temperature (tas) over the CA domain calculated from obsclim and counterclim data sets. The positive trend within the factual climate follows well the rising global mean CO_2 concentration. The key observation is that the temperature in the factual (observed) climate scenario shows a positive trend, aligning with the increase in global CO2 levels, while the counterfactual scenario shows no such trend. Figure 3b examines temperature anomalies (differences from a historical baseline period of 1901-1930) for the period 1990-2019 from counterclim and obsclim datasets. The right tail of the histogram of the obsclim is shifted, indicating an increased risk of extreme warm spells for the factual climate. It presents a smoothed, normalized histogram for the factual climate, showing a significant shift towards higher temperature anomalies. This indicates not just a general warming but also an increased likelihood of extremely warm periods. The histogram line presents the KDE, following a non-parametric way to estimate a random variable's PDF (see Section 2.2). This method helps visualize the underlying distribution of data points, especially when the nature of the distribution is unknown or when we prefer not to assume a specific distributional form (like normal, exponential, etc.).

By comparing the likelihoods (the probability of each level of temperature change, i.e. PDF values) for temperatures > 1 in Fig. 3b, we can see that higher PDF values are attributed to obsclim, i.e. to the factual climate. The attribution ratio (AR), which compares the probabilities of warming in factual vs counterfactual climates, is calculated for temperature anomalies > 1 as follows:

$$AR = \frac{PDF(\text{obsclim})}{PDF(\text{counterclim})} \tag{4}$$

and presented in Fig. 3c. For example the AR of temperature anomalies of 3, 5 and 7 K are 2.1, 3 and 7. This figure allows to further explore the probability of various levels of temperature change, showing that the risk of higher temperature anomalies (extreme warmth) increases
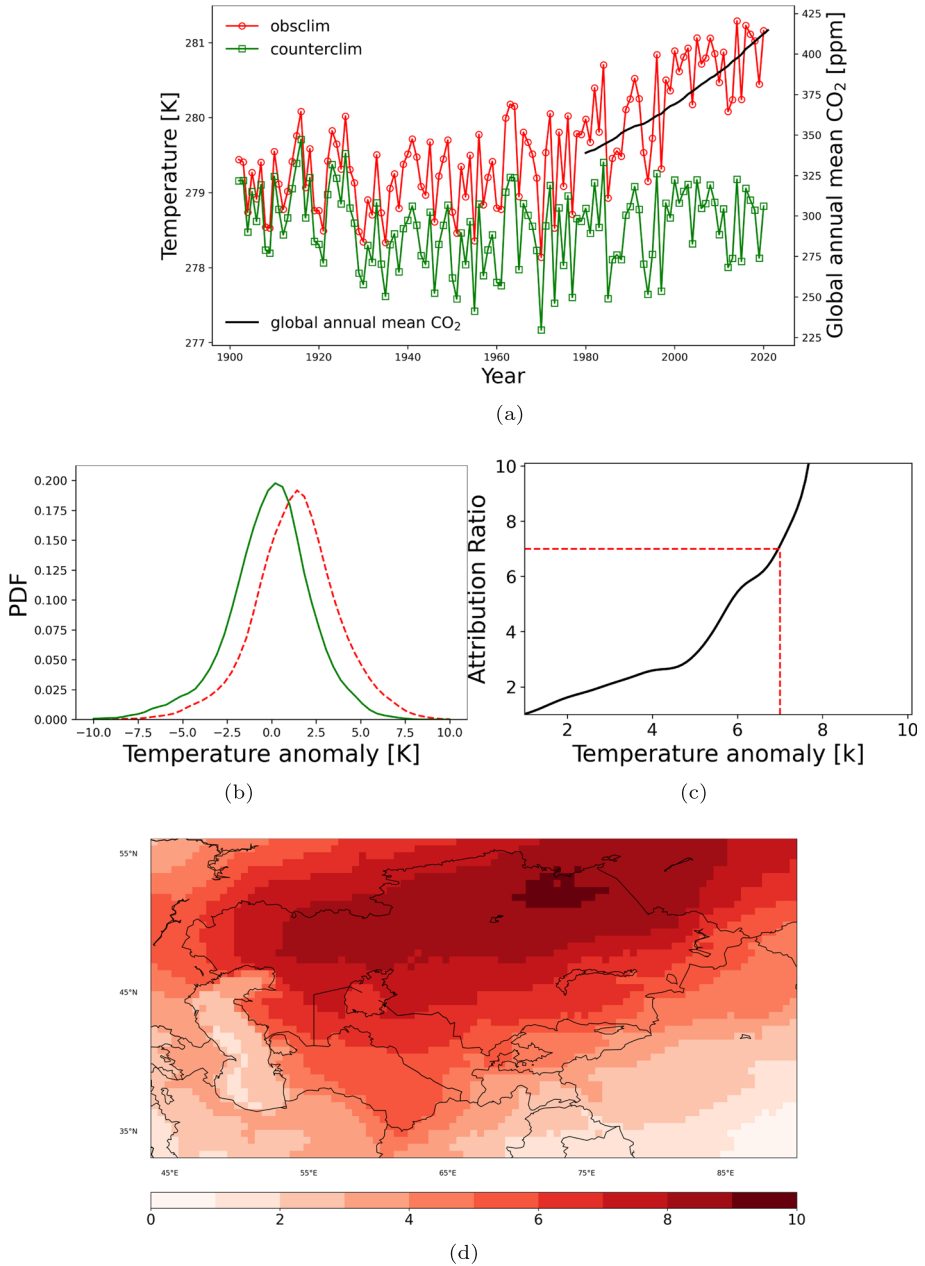
(a)

(b)

(c)

(d)

Fig. 3 (a) Annual average time series of daily near-surface air temperature (tas) over the CA domain calculated from obsclim (red line with circles) and counterclim (green line with squares) data sets. The global CO_2 concentrations are shown in the black line and the right y-axis. (b) The smoothed best fit of the histogram of daily temperature changes (1990-2019 w.r.t. 1901-1930) for obsclim (red-dashed line) and counterclim (green solid line). The black line in (c) shows the attribution ratio calculated from densities shown in (b). d) presents the composite pattern of tas anomalies when Central Asian temperature trend (values averaged over the entire CA) is higher than the +7K threshold from the factual climate (obsclim) (temperature anomalies to the right of vertical dashed line in c)

almost exponentially in the factual climate. For example, the PDF of the 5K anomaly is less than 0.025 in the counterfactual and around 0.05 in the factual dataset (almost triple the value). However, it should be noted that extreme temperature anomalies are rare, which affects the certainty of these risk calculations for higher anomaly values. For example, the risk of tas anomalies more than 7 K increases from 30 days in the counterfactual to 168 days in the factual climate.

The associated pattern of tas anomaly larger than +7 K from the factual climate is presented in Fig. 3d. This shows the average of temperature anomaly during the days in 1990-2019, which are +7 K warmer than the 1901-1930 average. A large warming pattern is associated with the +7K anomalies. Entire Kazakhstan shows warming of more than 5K w.r.t. 1901-1930 baseline period in the obsclim dataset. Not only the lowlands of CA but also the elevated regions of Kazakhstan and Kyrgyzstan show warming patterns of around 4 K. Given the stationary relationship between the anomalous warming patterns and the GMT this widespread warming shall be of concern if the GMT continues to increase. The overall message of Fig. 3 is that global warming influences have led to a noticeable and significant warming trend in CA, including an increased risk of extreme warm spells. This warming is not uniform but is particularly pronounced in certain regions, indicating a need for concern if global mean temperatures continue to rise.

Now, we move from analyzing extreme temperature anomalies to examining heatwave patterns. It is imperative to consider how these elevated temperatures manifest over sustained periods. The transition from examining isolated instances of extreme warmth to scrutinizing prolonged periods of intense heat underscores the nuanced impact of climate change on weather patterns. For a heatwave defined as a period of at least five consecutive days with temperatures exceeding the reference value by 7K (Fig. 4a), there is a noticeable distribution of positive values when comparing factual climate data to counterfactual scenario during 1901-2019. This suggests that short-duration heatwaves are more common under factual climate than in the counterfactual climate. As the required duration for a heatwave increases to 10 days (Fig. 4b), 15 days (Fig. 4c), and 20 days (Fig. 4d), the positive values across the maps indicate that longer heatwaves also occur more frequently than the counterfactual data would suggest. However, these longer heatwaves are understandably less frequent than the shorter-duration events, as indicated by the relative intensity of the positive values on the maps. The consistency of positive values across all maps underscores the reality that heatwaves, irrespective of duration, are more prevalent in the factual climate data, pointing to a tangible shift in climate patterns. The spatial distribution of these positive values highlights regions particularly susceptible to heatwaves, which is crucial information for developing targeted strategies for climate adaptation and mitigation.

3.3 Heavy precipitation events

According to the intergovernmental panel on climate change (IPCC) sixth assessment report (AR6), attributing regional extreme precipitation events at regional scales to human-induced drivers is still limited and the methods are under investigation (Wehner et al. 2021). This motivated us to contribute to the few studies in CA (Fallah et al. 2023a; Didovets et al. 2024; Fallah et al. 2023b), which deal with the possibility of attributing the heavy precipitation events to climate change using the counterfactual climate. Here, the ninety-eighth percentile of daily precipitation over the whole period (1901-2019) and for each grid point (PR98, hereafter) is used as the threshold for finding the heavy precipitation events (Jacobeit et al. 2009). Figure 5a shows the PR98 pattern of obsclim data set with maximum values over the
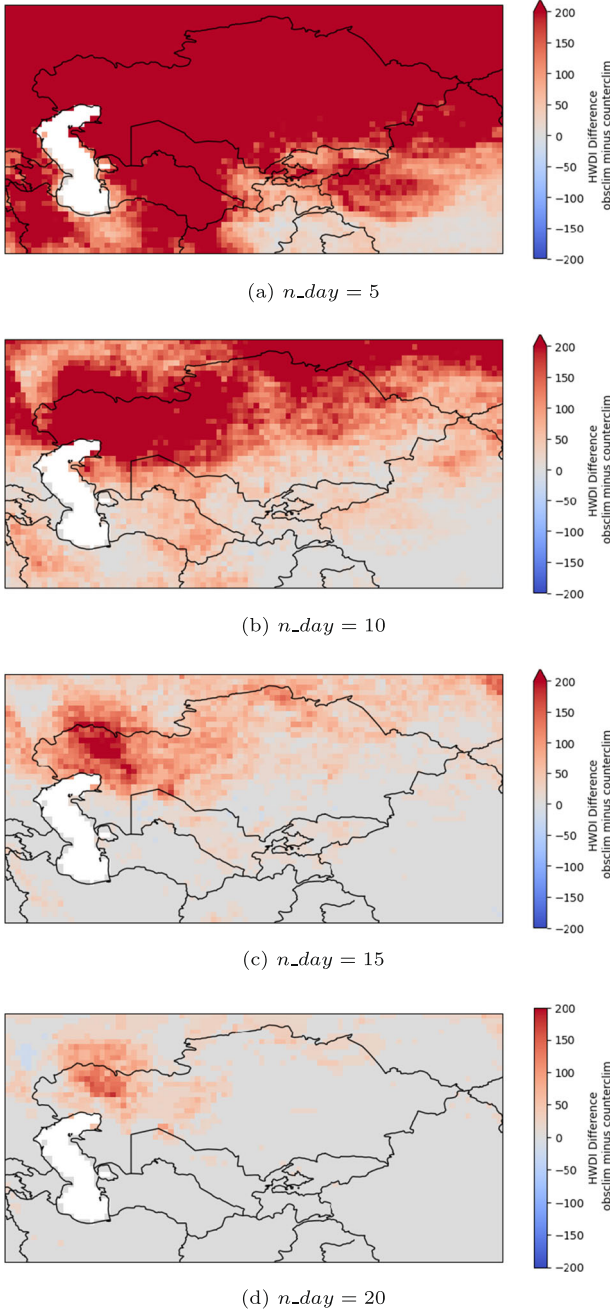
(a) $n\_day = 5$

(b) $n\_day = 10$

(c) $n\_day = 15$

(d) $n\_day = 20$

Fig. 4 Heat Wave Duration Index (HWDI) difference maps for different number of consecutive days (1901-2019) used to define a heatwave $n\_day = 5$, 10, 15, 20 and 7 K above the reference value ($T = 7K$) in a,b,c and d, respectively
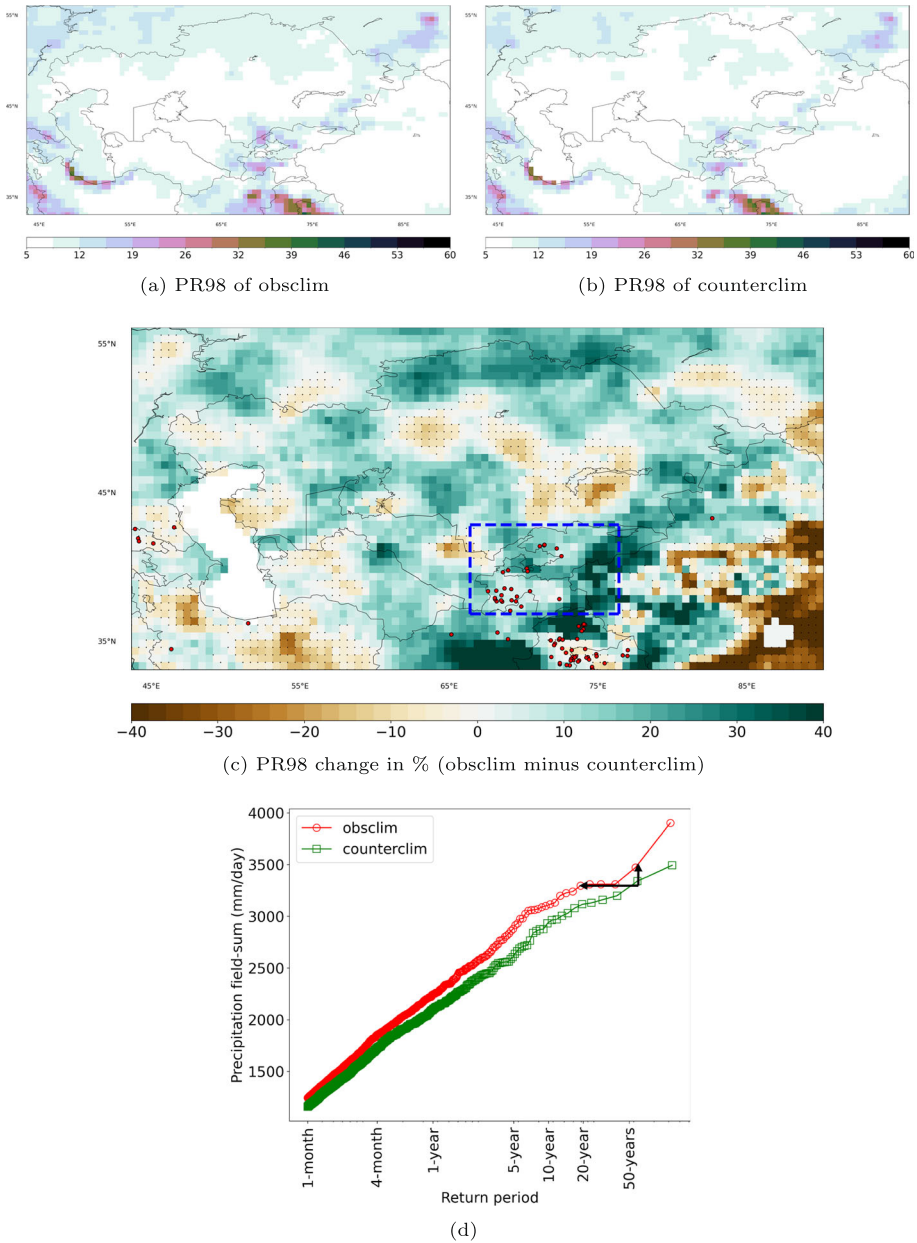
(a) PR98 of obsclim

(b) PR98 of counterclim

(c) PR98 change in % (obsclim minus counterclim)

(d)

Fig. 5 PR98 of the obsclim and counterclim (a and b) and PR98 changes [%] for 1901-2019 (c). The blue dashed box in (c) shows the selected region for calculating return periods. (d) Return periods of precipitation accumulated over the region around landslide events for obsclim (red circles) and counterclim (green squares). The vertical and horizontal arrows in (d) show the enhancing magnitude and frequency of 50-year events, respectively

South Caspian Sea, West Iran, Tajikistan, Kyrgyzstan, East Afghanistan, North India and East Uzbekistan and Kazakhstan. Figure 5b shows the PR98 pattern of counterclim data set. The patterns are very similar, which point to the consistent patterns of extreme precipitation. The percentages of differences in PR98 between the obsclim and counterclim caluclated according to Eq. 5 are shown in Fig. 5c.

$$\triangle PR98 = \frac{(PR98_{obsclim} - PR98_{counterclim}) \times 100}{PR98_{counterclim}} \tag{5}$$

An increase of PR98 over Tajikistan, Kyrgyzstan, West China and East Afghanistan, areas where the PR98 shows high values, might contribute to extreme high-impact events such as landslides and floods. The enhanced PR98 pattern linked to GMT lies over the location of rainfall-triggered landslide events in CA (red dots in Fig. 5b). On the other hand, the increase in extreme precipitation followed by large-scale warming over CA leads to melting ice in high elevations. It contributes to the shrinkage of glaciers in Tian Shan and Pamir (as reported by Hu and Han 2022).

To discover a potential shift in the frequency and magnitude of such events, a region covering the well-documented landslide events over Tajikistan and Kyrgyzstan (blue-dashed box in Fig. 5c) is chosen. The return periods of the monthly precipitation sum over this region are calculated for the factual and counterfactual climate states from 1901 to 2019. The separation between the statistics of the two climates for rare events is clear. A 50-year event in a counterfactual climate is a less-than-20-year event in a factual climate. Additionally, the magnitude of the rare events increases (vertical arrow in Fig. 5d). Under the linearity assumption of the trends, sporadic events (with a 100-year return period) might happen more than once in the lifetime of a person living in the 21st century in CA.
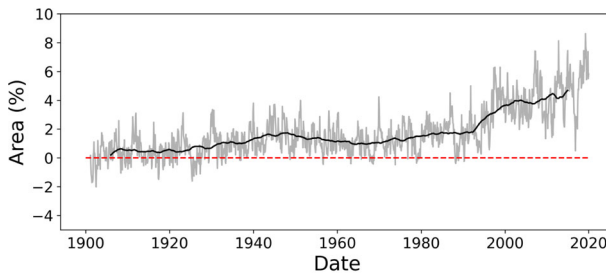
(a) Area under dry conditions
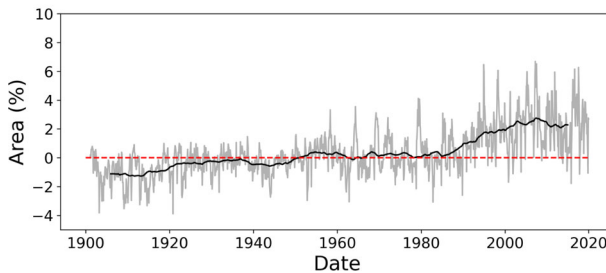
(b) Area under wet conditions

Fig. 6 Percentage of changes (obsclim minus counterclim) in the area under dry (a) and wet (b) conditions measured by scPDSI < -2 and > 2, respectively. Solid lines show the 10-year running mean

3.4 The impact of temperature and precipitation change on the dry and wet conditions

It was shown that the fingerprint of global warming was detectable in an extreme event with concurrent precipitation and temperature changes. As a simple climate impact study, those variables are used and a sensitivity test is conducted using the scPDSI (Wells et al. 2004). Daily near-surface temperature, precipitation, and AWC are the primary inputs for calculating the scPDSI. In our test, we isolate the impact of AWC (we use the same AWC for both datasets) and use the tas and pr variables from obsclim and counterclim to calculate the monthly scPDSI values. In the next step, the counterclim scPDSI is subtracted from obsclim scPDSI and the contribution of GMT to their changes is calculated. Figure 6 shows the percentages of changes in the area under drought/wet conditions, defined by the number of grid points with scPDSI values lower/higher than -2/+2. The areas under dry and wet conditions have been expanding in recent decades in CA, especially after the 1990s. Figure 7 shows the spatial pattern of the time-averaged scPDSI for 2010-2019 over CA. An expansion area of the dry condition is observed around the Aral Sea, i.e., over Betpaqdala desert in Kazakhstan, East of the Caspian Sea, Karakum desert, Ustyurt desert and Western China over the Xinjiang Taklamakan desert. This agrees with the previously observed desert expansion in these areas
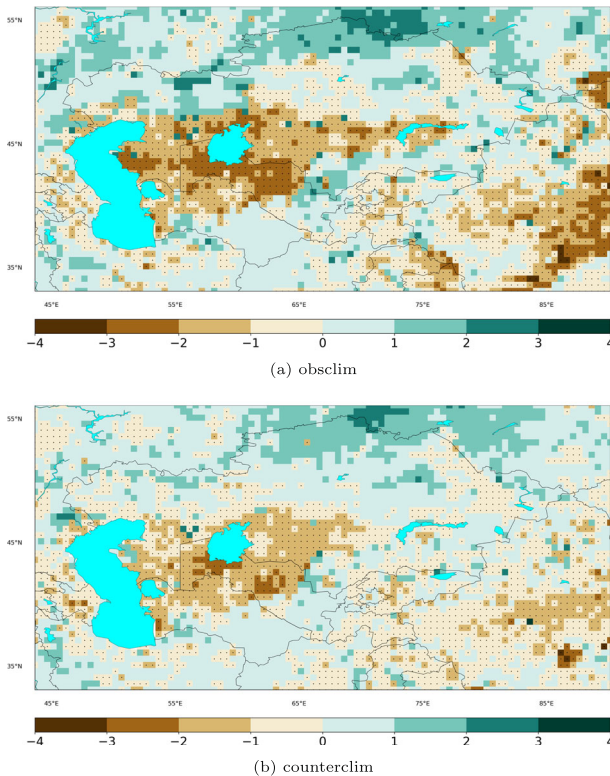

Fig. 7 Average of scPDSI for 2010-2019 from obsclim (a) and counterclim (b). Black dots indicate the negative values

(Hu and Han 2022). This tendency to drier conditions in downstream countries of CA will increase the water demand and put the agriculture sector under severe risks.

4 Discussion and conclusion

Detrending methodologies play a crucial role in climate change attribution studies by isolating the influence of long-term trends, such as those associated with global warming, from natural variability in climate data. By removing these trends from the dataset, researchers can compare current climate conditions with a hypothetical "counterfactual" scenario that might have existed if human-induced global warming did not occur. This comparison helps to assess the impact of global warming on the frequency, intensity, and magnitude of extreme weather events, including heavy precipitation.

For example, in the context of heavy precipitation events, detrending allows us to assess how much more likely these events have become in the current climate compared to a counterfactual scenario. Precisely, by comparing the statistical properties of precipitation events in the factual climate data with those in the detrended (counterfactual) climate data, we can quantify the contribution of global warming to changes in the magnitude and frequency of these events. This involves analyzing the distribution of precipitation events above certain thresholds (e.g., the 98th percentile) in both datasets and identifying shifts in these distributions that can be attributed to the warming trend.

By comparing the factual climate data with the detrended counterfactual data, we identified an increase in the frequency and magnitude of rare precipitation events that can be attributed to global warming. Specifically, we observed that heavy precipitation events, which are rare in the counterfactual scenario, have become more frequent in the observed climate, indicating that global warming is already modifying the hydrological cycle in Central Asia.

The analysis showed that comparing factual and counterfactual climate states could detect anomalous patterns in the recent climate. One can contribute a fraction of those changes (frequencies and magnitudes) to the global warming trend. The observed increase in heatwave occurrences in the factual climate data relative to the counterfactual scenarios indicates a significant shift in climatic conditions. The positive values across all HWDI difference maps for various heatwave durations underscore a clear trend towards more frequent and prolonged periods of extreme heat. This trend has profound implications for ecological systems, human health, and socioeconomic structures.

The factual and counterfactual data disparity suggests that recent climatic changes, potentially driven by anthropogenic factors, already impact weather patterns that could exacerbate vulnerability across different regions. The analysis of HWDI difference maps presents compelling evidence of the increasing prevalence of heatwaves in the current climate as opposed to what might have been expected under counterfactual conditions. This finding is critical for informing adaptation and mitigation strategies. It emphasizes the urgency of addressing climate change and highlights the need for robust, data-driven policies to manage the risks associated with rising temperatures. As the frequency and intensity of heatwaves continue to surpass historical norms, our societal resilience will increasingly depend on our ability to anticipate, prepare for, and respond to these extreme weather events. Therefore, the increased heatwave activity captured in the factual data versus the counterfactual scenarios represents a vital alarm signal that warrants immediate attention from policymakers, researchers, and the global community.

The analysis also suggests that data sets using detrending methodologies can be used for finding the contribution of global warming to the event magnitude (for example, heavy precipitation) and their changing trends (i.e., scPDSI changes). The impact modellers could benefit from detrending methodologies if a computationally affordable statistical surrogate of such models, like the ATTRICI counterfactual data, can capture the large-scale climate change impacts. However, such methodologies suffer from the problem of stationarity process. They can find relationships only observed within their period and can never be the perfect replacement for the histNat CMIP-like simulations. Another main caveat of detrending methodologies is their inability to find the contribution of the diverse drivers of climate change and their specific impacts.

Here, several further shortcomings of the presented study are discussed. For the calculation of the scPDSI, fixed AWC information is used. However, the human-induced driver might partially change the land-use, land-cover or soil characteristics. Further research could explore incorporating dynamic AWC information that reflects changes in land use, land cover, and soil characteristics over time. This approach would allow for a more accurate representation of the soil's ability to retain moisture, which is crucial for assessing drought severity in the context of changing environmental conditions. Additionally, integrating land-use change scenarios into the scPDSI calculation could provide insights into how human-induced landscape alterations, such as deforestation, urbanization, and agricultural expansion, affect local and regional water balance and drought risk.

Station observations and gridded datasets for temperature and precipitation can exhibit differences due to several factors, including spatial resolution, data processing, and the physical location of measurements. Gridded datasets, such as the bias-adjusted reanalysis with 0.5-degree horizontal resolution, provide a smoothed, interpolated view of climate variables over a regular grid, which can average local variations. In contrast, as collected by the Global Historical Climatology Network daily, station data represent point measurements that can be highly influenced by local geography, such as valleys, elevation, and urban heat islands.

Discrimination can arise in temperature because gridded datasets might not capture microclimates or extreme local weather events as accurately as station observations, particularly in complex terrains. Elevation plays a significant role; for stations above 2000 meters, the lack of height correction can lead to noticeable differences in temperature profiles between the datasets.

Precipitation differences are often more pronounced, especially for stations at lower altitudes (<1000m), where local topography, such as valleys, may lead to significant precipitation events that are not captured by the coarser resolution of gridded datasets. This is due to the gridded data's inability to resolve small-scale features that can significantly impact precipitation patterns.

Additionally, the station dataset's higher rate of missing values and the concentration of observations at lower altitudes complicate direct comparisons with gridded data. The underestimation of extreme temperatures and lower precipitation amounts in the factual dataset, particularly at lower altitudes, underscores the limitations of using gridded data to represent local climate variability without additional corrections. This contrast highlights the importance of considering dataset characteristics and potential biases when analyzing climate variables across different elevations and terrains.

By addressing these challenges, future studies can enhance our understanding of drought dynamics and improve our projection and mitigation of future climate change impacts on water scarcity. For calculating the return periods or similar statistical values, 120 years of daily information might be very short to sample enough rare events. Studies using large ensembles of climate simulations might give more realistic return period values for rare

events. It should be noted that the analysis presented here is not an alternative to a classical attribution methodology where a causal relationship is being investigated, and the impacts are connected to specific drivers. In contrast, here, the specific drivers of the rare events were not important, and the focus was only on climate change and not its drivers. There is no homogenized open access observation network of daily weather stations in CA which covers at least 30 years of the recent climate with adequate spatial coverage. Therefore, the gridded observational data sets and reanalysis might lack some important inter-seasonal variabilities.

An alternative to the observation data is using high-resolution regional climate models. It has been shown that such models could bypass the accuracy of the observation network, especially for regions with complex topography (Lundquist et al. 2019). However, the computational costs of those models are very high. The sparse coverage and uneven distribution of meteorological stations, coupled with the complex topography of CA, pose significant challenges to accurately capturing climate variability and trends through gridded datasets. These limitations underscore the need for a homogenized open-access observation network to provide comprehensive and reliable climate data across CA. Such a network would facilitate improved climate modelling and impact assessments by ensuring data consistency and covering gaps in the existing observation infrastructure.

Integrating data from high-resolution regional climate models (RCMs) could overcome the limitations of gridded observation datasets. RCMs can provide detailed climate projections at a scale more relevant to regional and local climate impacts, capturing the influences of topography, land use, and other regional factors that coarse-resolution global models might miss. While RCMs represent a significant advancement in climate modelling capabilities, they also present computational challenges. High-resolution simulations require substantial computational resources and expertise, making them less accessible for some research institutions.

Despite these challenges, the advantages of incorporating RCM data into climate studies are manifold. RCMs can enhance the understanding of regional climate processes, improve the accuracy of climate impact assessments, and inform more targeted adaptation and mitigation strategies. To leverage the benefits of RCMs while managing computational demands, researchers can adopt collaborative approaches, sharing computational resources and expertise across institutions. Furthermore, advances in computational technology and developing more efficient modelling techniques are gradually lowering the barriers to high-resolution climate modelling. By integrating RCM data with homogenized observation networks, the climate research community can significantly improve the quality and applicability of climate impact assessments in CA and beyond.

Author Contributions BF was responsible for data preparation, analysis, writing and original Draft preparation. MR has reviewed and discussed the reviewers' comments, scientific content, and supervised the production of the final version.

Availability of data and materials Please refer to Data and Methods (Sec.2).

Declarations

Competing interests There are no competing interests.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

References

Adams J (2017) climate_indices, an open source Python library providing reference implementations of commonly used climate indices. https://github.com/monocongo/climate_indices

Allan RP, Hawkins E, Bellouin N, et al. (2021) Ipcc, 2021: summary for policymakers. IPCC

Davi NK, D'Arrigo R, Jacoby G et al (2015) A long-term context (931–2005 ce) for rapid warming over central asia. Quatern Sci Rev 121:89–97

Didovets I, Krysanova V, Nurbatsina A et al (2024) Attribution of current trends in streamflow to climate change for 12 central asian catchments. Clim Change 177(1):16

Dilinuer T, Yao JQ, Chen J et al (2021) Regional drying and wetting trends over central asia based on köppen climate classification in 1961–2015. Adv Clim Chang Res 12(3):363–372

Fallah B (2024) Heatwave analysis. https://doi.org/10.5281/zenodo.10723413

Fallah B, Russo E, Menz C et al (2023a) Anthropogenic influence on extreme temperature and precipitation in central asia. Sci Rep 13(1):6854

Fallah B, Menz C, Russo E et al (2023b) Climate model downscaling in central asia: a dynamical and a neural network approach. Geosci Model Dev Discuss 2023:1–27

Gillett NP, Shiogama H, Funke B et al (2016) The detection and attribution model intercomparison project (damip v1.0) contribution to cmip6. Geosci Model Dev 9(10):3685–3697

Guglielmi G (2022) Climate change is turning more of central asia into desert. Nature

Hu Q, Han Z (2022) Northward expansion of desert climate in central asia in recent decades. Geophys Res Lett 49(11):e2022GL098,895

Hu Z, Zhang C, Hu Q et al (2014) Temperature changes in central asia from 1979 to 2011 based on multiple datasets. J Clim 27(3):1143–1167

Hu Z, Hu Q, Zhang C et al (2016) Evaluation of reanalysis, spatially interpolated and satellite remotely sensed precipitation data sets in central asia. J Geophys Res Atmos 121(10):5648–5663

Jacobeit J, Rathmann J, Philipp A, et al. (2009) Central european precipitation and temperature extremes in relation to large-scale atmospheric circulation types. Meteorologische Zeitschrift, pp 397–410

Jiang J, Zhou T (2021) Human-induced rainfall reduction in drought-prone northern central asia. Geophys Res Lett 48(7):e2020GL092,156

Lange S (2021) Isimip3 bias adjustment fact sheet

Lange S (2019) Trend-preserving bias adjustment and statistical downscaling with isimip3basd (v1.0). Geosci Model Dev 12(7):3055–3070

Lewis SC, Karoly DJ (2015) Are estimates of anthropogenic and natural influences on australia's extreme 2010–2012 rainfall model-dependent? Clim Dyn 45(3):679–695

Liverman D (2008) Assessing impacts, adaptation and vulnerability: Reflections on the working group ii report of the intergovernmental panel on climate change

Lundquist J, Hughes M, Gutmann E et al (2019) Our skill in modeling mountain rain and snow is bypassing the skill of our observational networks. Bull Am Meteor Soc 100(12):2473–2490

Mengel M, Treu S, Lange S et al (2021) Attrici v1.1-counterfactual climate for impact attribution. Geosci Model Dev 14(8):5269–5284

Micklin P (2007) The aral sea disaster. Annu Rev Earth Planet Sci 35(1):47–72

Otto FE (2017) Attribution of weather and climate events. Annu Rev Environ Resour 42:627–646

Pala C (2005) To save a vanishing sea

Palmer WC (1965) Meteorological drought, vol 30. US Department of Commerce, Weather Bureau

Peng D, Zhou T, Zhang L, et al (2020) Observationally constrained projection of the reduced intensification of extreme climate events in central asia from 0.5 ° less global warming. Climate Dynamics 54(1):543–560

Siegfried T, Bernauer T, Guiennet R, et al (2010) Will climate change exacerbate or mitigate water stress in Central Asia?, vol 1

Unger-Shayesteh K, Vorogushyn S, Farinotti D et al (2013) What do we know about past changes in the water cycle of central asian headwaters? a review. Global Planet Change 110:4–25

van Oldenborgh GJ, van der Wiel K, Kew S et al (2021) Pathways and pitfalls in extreme event attribution. Clim Change 166(1):1–27

Wehner M, Seneviratne S, Zhang X, et al (2021) Weather and climate extreme events in a changing climate, vol 2021

Wells N, Goddard S, Hayes MJ (2004) A self-calibrating palmer drought severity index. J Clim 17(12):2335–2351

Zhang M, Chen Y, Shen Y et al (2017) Changes of precipitation extremes in arid central asia. Quatern Int 436:16–27

Zhou LT, Huang RH (2010) Interdecadal variability of summer rainfall in northwest china and its possible causes. Int J Climatol 30(4):549–557

Zou S, Abuduwaili J, Duan W et al (2021) Attribution of changes in the trend and temporal non-uniformity of extreme precipitation events in central asia. Sci Rep 11(1):1–11