Environmental Research Letters

ACCEPTED MANUSCRIPT • OPEN ACCESS

One simulation, different conclusions - the baseline period makes the difference!

To cite this article before publication: Stefan Liersch *et al* 2020 *Environ. Res. Lett.* in press https://doi.org/10.1088/1748-9326/aba3d7

One simulation, different conclusions – the baseline period makes the difference!

S Liersch[1], M Drews[2], T Pilz[1], S Salack[3], D Sietz[1], V Aich[1,4], M A D Larsen[2], A Gädeke[1], K Halsnæs[2], W Thiery[5], S Huang[6], A Lobanova[1], H Koch[1], F F Hattermann[1]

[1] Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, P.O. Box 60 12 03, D-14412 Potsdam, Germany

[2] Technical University of Denmark (DTU), Produktionstorvet, 2800 Kongens Lyngby, Denmark

[3] WASCAL Competence Center, Blvd Moammar El-Khadafi, 06BP 9507, Ouagadougou 06, Ouagadougou, Burkina Faso

[4] Global Climate Observing System (GCOS) Secretariat, c/o World Meteorological Organization (WMO)

[5] Vrije Universiteit Brussel (VUB), Department of Hydrology and Hydraulic Engineering, Brussels, Belgium

[6] The Norwegian Water Resources and Energy Directorate (NVE), PO Box 5091, Majorstua, 0301 Oslo, Norway

E-mail: `liersch@pik-potsdam.de`

Abstract.

The choice of the base period, intentionally chosen or not, as a reference for assessing future changes of any projected variable can play an important role for the resulting statement. In regional climate impact studies, well-established or arbitrarily chosen baselines are often used without being questioned. Here we investigated the effects of different baseline periods on the interpretation of discharge simulations from eight river basins in the period 1960–2099. The simulations were forced by four bias-adjusted and downscaled Global Climate Models under two radiative forcing scenarios (RCP 2.6 and RCP 8.5). To systematically evaluate how far the choice of different baselines impacts the simulation results, we developed a similarity index that compares two time series of projected changes. The results show that 25% of the analysed simulations are sensitive to the choice of the baseline period under RCP 2.6 and 32% under RCP 8.5. In extreme cases, change signals of two time series show opposite trends. This has serious consequences for key messages drawn from a basin-scale climate impact study. To address this problem, an algorithm was developed to identify flexible baseline periods for each simulation individually, which better represent the statistical properties of a given historical period.

Keywords: Baseline period, Climate impacts, Climate projections, Flexible baseline

1. Introduction

In the context of climate change mitigation and adaptation, decision-makers generally call for information about impacts of projected changes in a specific region at different global warming levels or in certain future periods. They need answers to questions like: "Can we expect an increase or decrease in water availability, extreme events, such as floods, droughts, storm surges or heatwaves, around the year 2030, 2050 or by the end of the 21st century? And what will be the consequences for, e.g. crop production, renewable electricity generation?" To answer such questions, regional climate impact modellers face a variety of challenges, which relate to technical, methodological, and communication issues of simulation results [1] and corresponding recommendations under uncertainties in a comprehensible way.

Adding to technical and methodological challenges includes, e.g. the choice of climate scenarios, climate and impact models, the use of bias-adjustment methods, and model calibration and validation periods. The performance of a climate model is usually measured against it's ability to represent spatial patterns and trends in the historical climate. Sometimes the performance is used to assign weights to individual models within a model ensemble [2, 3, 4, 5, 6, 7]. The uncertainty cascade in the impact modelling is basically associated with model structure, model parameterisation, and input data quality [8, 9, 10, 11, 12, 13, 14, 15].

After the simulations have been carried out, the question about the baseline period used to compare future simulation results to, will arise. Where future scenario periods are usually defined to reflect the decision maker's planning horizon, baseline periods are often chosen arbitrarily or are based on existing standards. However, choosing a baseline period is a sensitive issue and can be easily instrumentalized to support specific conclusions, whether intentionally or not.

The World Meteorological Organization (WMO) recommends to use the 30-years period of 1961-1990 as the climate normal when comparing with future periods and that this should be maintained as a reference for monitoring long-term climate variability and change [16, 17]. Beyond that, a regularly updated 30-years baseline period, currently 1981–2010, should be employed to give people a more recent context for understanding weather and climate extremes and forecasts [17, 18, 19]. The Intergovernmental Panel on Climate Change (IPCC) used the 20-years period 1986–2005 as the baseline in many graphs in the Fifth Assessment Report [20] and will use the years 1995–2014 in its Sixth Assessment Report. So, what are climate impact modellers supposed to do? Which baseline should they select and does it actually matter?

At the global or continental scale, it is virtually impossible to choose a baseline period whose climate is represented realistically by all climate models. An arbitrary determination of global baselines is therefore justifiable. However, global and regional climate simulations are often not designed to synchronize with real year to year patterns and events [21], which creates a communication challenge, particularly in regional impact studies. For example, some climate models depict the mid-1980s as a period with above-normal rainfall, when in reality a drought hit West Africa. Others simulate the extraordinary wet 1950s and 1960s as very dry. Nevertheless, well-established global baseline periods are often used unquestioningly in regional impact studies, although the real-life statistical properties of the specific historical period may not be adequately represented by climate model simulations, therefore, also not in subsequent applications.

Even though the implications of the choice of baseline periods for the interpretation of simulation results are well known, little attention has been paid to them in the climate impact community. Ruokolainen and Räisänen (2007) [22] analyse the sensitivity of forecasts to the choice of different baselines in Southern Finland. Razavi et al. (2015) [23] emphasize that different length of baseline periods may lead to different conclusions about stationarity/non-stationarity. Hawkins and Sutton (2016) [18] discuss the choice of climate reference periods when comparing global air temperature projected by climate models with observations. Huang et al. (2018) [24] depict future flood characteristics in future periods in four river basins based on different 30-years baseline periods. Snell et al. (2018) [25] highlight the sensitivity to the choice of baseline climate in dynamic forest modeling in the Alps. Baker et al. (2016) [26] assessed the impact of six different climate baselines on projections of African bird species' responses to future climate change. Although this issue has been addressed as a side effect in several other studies, it has generally not been considered important to form the focal point for systematic research.

The present study systematically investigates the effect of the choice of the baseline period on the interpretation of simulation results. It provides examples from eight river basins located in various climate zones, where changes in projected future discharge are estimated based on WMO and IPCC baselines using four bias-adjusted and downscaled Global Climate Models (GCMs) from the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2) [27, 28, 29]. An index measuring the similarity between two time series is introduced and was used to assess the sensitivity of choosing different baseline periods. We developed an algorithm to overcome the problem in cases of substantial deviations. It identifies a baseline period, which consists of similar basic statistical properties as the historical period and is flexible in terms of length and timing. Although the main focus of this study is the analysis of river discharge, the method is in principle applicable to any time series variable, such as meteorological data, crop yields, emissions of greenhouse gases, hydropower potentials and so on.

2. Materials and methods

2.1. Study sites

The impact of different baseline periods was investigated by using simulated river discharge from eight exemplary river basins located in various climate zones from equatorial to polar (Figure 1 and Table 1). The simulations were carried out within the framework of various research projects (see references in Table 1). What they have in common is the hydrological model, the same four forcing GCMs, and the simulation period they cover (1960–2099), which guarantees consistency across the study basins.

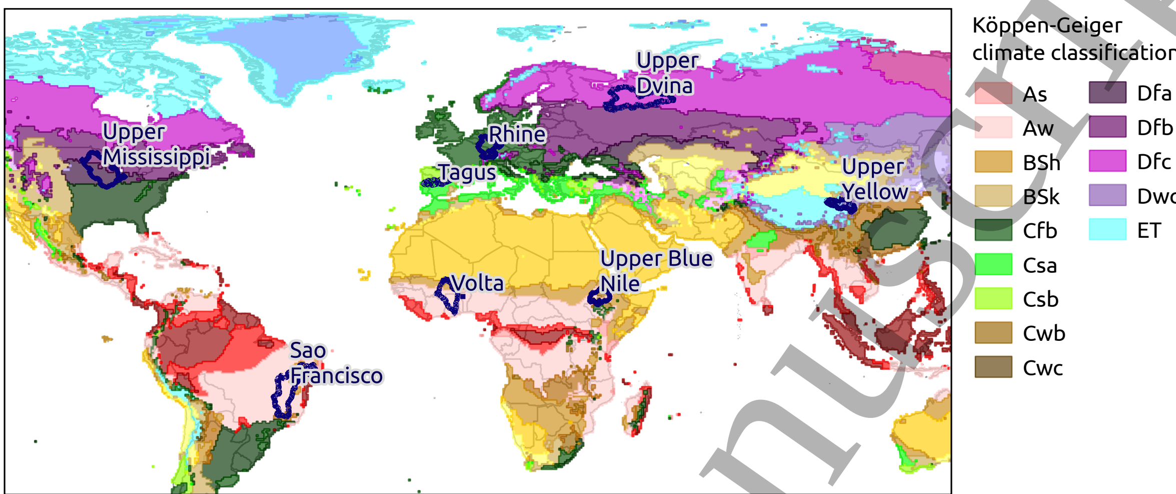


Figure 1. Map of case study river basins and climate zones. Source climate zones: Kottek et al. (2016) [30]

Table 1. River basins

River basin	Gauge	Area $[km^2]$	Region.	Climate Zone
Northern Dvina (DVI) [31]	Ust-Pinega	348.000	Europe, Russia	Snow (Dfc)
Rhine (RHI) [24]	Lobith	160.000	Europe	Warm temperate (Cfb)
São Francisco (SFC) [32]	Outlet	640.000	S. America	Equatorial (Aw, As), arid (BSh)
Tagus (TAG) [31]	Almourol	70.000	S. Europe	Warm temperate (Csa, Csb)
Upper Blue Nile (UBN) [4, 33, 34]	El Diem	175.000	E. Africa	Arid (Aw), warm temperate (Cwb)
Upper Mississippi (UMI) [24]	Alton	440.000	N. America	Snow (Dfa, Dfb)
Volta (VOL)	Outlet	403.000	W. Africa	Equatorial (Aw), arid (BSh)
Upper Yellow (YEL) [24]	Tangnaihai	121.000	China	Polar (ET), snow (Dwc)

2.2. Data

The investigation was conducted using annual mean discharge MQ, derived from simulated daily discharge from eight river basins, based on climate model input from four GCMs in the period 1960–2099. The discharge was simulated with the semi-distributed, eco-hydrological Soil and Water Integrated Model (SWIM) [35, 36]. The downscaled and bias-adjusted GCM climate simulation data were provided by ISIMIP2 [27, 28, 29, 37] for the GFDL-ESM2M, HadGEM2-ES, IPSL-CMA5-LR, and MIROC5 models. The aim is to provide harmonized climate simulation input to impact modelers and thereby to support the intercomparison of global and regional impact studies.

2.3. Baseline periods

The impact of the baseline period on the interpretation of changes in simulated future river discharge was investigated by using two baselines established by the WMO [17] and the IPCC [20]. The *WMO baseline* covers the 30 years 1961–1990 and the *IPCC baseline* the 20 years 1986–2005. Other IPCC reports use also different and longer baseline periods. However, we chose the above-mentioned baseline here, as it is used in many graphs in the IPCC AR5 report [20] and is therefore likely to tempt impact modellers to use it as a standard in their studies. Interestingly, the central limit theorem dictates that at least 30 samples are needed if we assume a normal distribution and to ample natural variability [38] as in the case of the WMO [19, 16]. From this perspective, the *IPCC baseline* is thereby too short, especially if variables with a high degree of natural variability are considered, e.g. river discharge. However, one could argue that the sample size is sufficiently large, if the combination of years in the baseline period times the number of models exceeds a critical threshold, which is given in the case of the IPCC (20 years times 40^+ GCMs). In addition, the selection of the baseline period should strike the balance between being statistically robust and representative of the target conditions (e.g. "present-day climate"). For rapidly changing variables, such as for instance extreme temperatures, reference periods of 30 years or longer might be considered insufficiently representative of the target conditions.

In this study, we hypothesize that the baseline period is a subset that accurately represents some basic statistical properties of a historical period, here defined as 1960–2005. An algorithm was developed to identify for each simulation a baseline period of variable length within a given historical period. The algorithm searches for a baseline period whose mean, minimum, and maximum values correspond to those of the historical period. In line with common practice of hydro-climatic impact studies, the baseline period should cover at least 30 years. The statistical properties of the baseline period are allowed to deviate from those of the historical period by not more than a user-defined threshold, e.g. 5%. If the algorithm is not able to find an appropriate baseline with $n = 30$ years, n is incremented by 1. The resulting baseline period is therefore flexible in terms of its length and starting year and is called hereafter "flexible baseline". The corresponding function, implemented in R, is provided in Appendix A. It works only for annual series but can be easily adapted for monthly or daily series.

To account for the possibility of a linear climate change trend in the historical discharge, the algorithm was tested using a time series detrended using the first (linear) differencing method (Appendix B). In general, the differences in the results were found to be minor and the identified baseline periods to be longer. To avoid accidentally removing or suppressing some of the extreme years by applying a linear operation, results shown below are all based on the original data.

2.4. Similarity index

The MQ time series was used to compute the relative change between a specific baseline and a corresponding future period as follows:

$$\Delta MQ_{i,j} = \frac{\overline{MQ}_{future,i} - \overline{MQ}_{base,j}}{\overline{MQ}_{base,j}}, \tag{1}$$

where $\overline{MQ}_{base}$ is the average of the annual values of a specific baseline period and $\overline{MQ}_{future}$ the average of a future period. The index i refers to different future periods with central years between 2020 and 2080, i.e. 61 time steps. The index j represents the different baselines (WMO, IPCC, and flexible baseline), where the length of the baseline determines the number of years around the central years in corresponding $\overline{MQ}_{future}$ periods.

The mean absolute deviation between two ΔMQ time series, e.g. $\Delta MQ_{WMO,i}$ for the WMO and $\Delta MQ_{IPCC,i}$ for the IPCC baseline, over all $k = 61$ time steps was then quantified as:

$$D = \frac{1}{k}\sum_{i=1}^{k} \mid \Delta MQ_{WMO,i} - \Delta MQ_{IPCC,i} \mid . \tag{2}$$

The deviation was then re-scaled by a user-defined deviation threshold D_{max} to an agreement score value

$$AS = \begin{cases} 1 - \dfrac{D}{D_{max}} & \text{if D} \le D_{max} \\ 0 & \text{if } D > D_{max} \end{cases} \tag{3}$$

This *Agreement Score* ranges between one (no deviation, perfect agreement) and zero (deviation larger than the threshold). In this study a threshold value of $D_{max} = 25\%$ was defined, because deviations in discharge projections $\geq 25\%$ that are solely based on different baselines, were considered to be very large and indicative for a substantial difference. For other applications (e.g. greenhouse gas emissions, temperature, precipitation, wind speed) or by using not relative but absolute changes for $\Delta MQ_{i,j}$, other threshold values might be more appropriate.

Apart from the deviation based on the choice of different baselines, we quantified the direction of change signals CS as

$$CS_{i,j} = \begin{cases} 1 & \text{if } \Delta MQ_{i,j} > 0.01 \\ 0 & \text{if } -0.01 \le \Delta MQ_{i,j} \le 0.01, \\ -1 & \text{if } \Delta MQ_{i,j} < -0.01 \end{cases} \tag{4}$$

compared the agreement between two baselines for the future periods by setting

$$AC_i = \begin{cases} 1 & \text{if } CS_{WMO,i} = CS_{IPCC,i} \\ 0 & \text{if } CS_{WMO,i} \neq CS_{IPCC,i} \end{cases} \tag{5}$$

to eventually derive the average agreement in the direction of the change signal by

$$AC = \frac{1}{k}\sum_{i=1}^{k} AC_i. \tag{6}$$

Finally, a *Similarity Index* was defined as

$$SI = \frac{AS + AC}{2}. \tag{7}$$

A value of zero is derived if the selection of a baseline has a large impact on the interpretation of results of an impact study while the optimum value of $SI = 1$ is achieved if the choice of the baseline would have no influence at all. An intermediate value of $SI = 0.5$ can be derived if the absolute deviations are large with respect to the user-defined D_{max} value, but both baselines show the same directions of change over all possible future periods. Likewise, a value of approximately 0.5 is obtained when the choice of different baselines results in a bias with small absolute deviation with respect to D_{max}, although the directions of change deviate for all possible future periods.

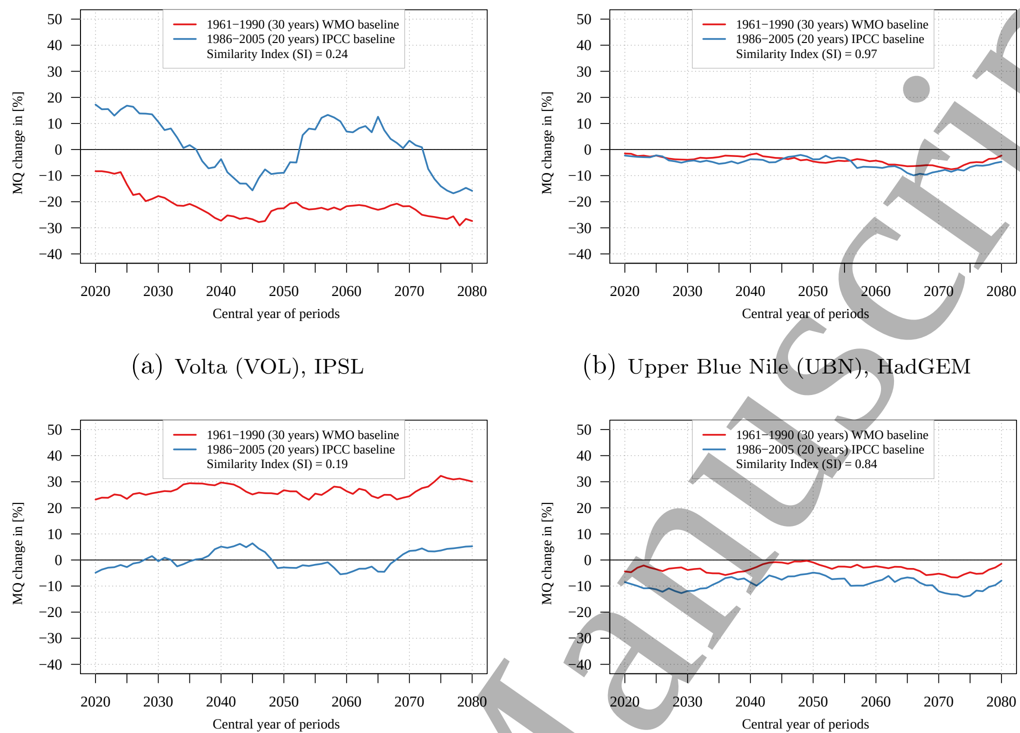
The computation of SI was also tested by integrating other factors, such as agreement in standard deviation or R^2, but the results achieved with a more complex indicator were not considered to be more meaningful than those achieved with the simplistic approach. The SI was also used to assess the sensitivity of the choice of the baseline depending on the GCM and the climate zone.

3. Results and discussion

This section shows to what extent the choice of the baseline alone can influence the interpretation of simulation results.

Figure 2 shows future ΔMQ series for selected river basins relative to MQs in the WMO and IPCC baselines. Future change signals and magnitudes of change can be extremely different between the two ΔMQ series (Figures 2 a and c). Both examples are therefore characterized by low SI values of 0.24 and 0.19, respectively, which indicate large differences of MQ values in the respective baselines. They also demonstrate that neither the results based on the one nor the other baseline generally tends to suggest higher or lower future ΔMQ, a phenomenon also found in river basins shown in Appendix C. Considering the example of the Northern Dvina River basin (Figure 2 c), one would conclude that future ΔMQ does not change substantially but rather fluctuates around the historical mean if the IPCC baseline is used. A completely different conclusion would be drawn with the WMO baseline, where ΔMQ is projected to increase between 22-32%. This illustrates how the choice of the baseline period, based on the same model simulation, would lead to conflicting recommendations for adaptation strategies.

Figures 2 b and d show examples of future ΔMQ where it apparently does not matter which baseline is used as reference. The corresponding SI values of 0.97 and 0.84 are therefore much higher than in the other two examples. Recommendations for



(a) Volta (VOL), IPSL (b) Upper Blue Nile (UBN), HadGEM

(c) Northern Dvina (DVI), GFDL (d) Rhine (RHI), MIROC5

Figure 2. Relative change in MQ of central years in four river basins using four different GCMs under RCP 2.6. Changes are relative to MQ in WMO and IPCC baselines

adaptation strategies would consequently be much less dependent on the choice of the baseline period in these cases.

A visual assessment of the ΔMQ series with the corresponding SI values in Figure 2 is conclusive, where low SI values indicate a high sensitivity to the choice of the baseline period and high SI values a low sensitivity. As with model performance indicators (e.g. R^2, PBIAS), an evaluation of which value ranges indicate actually a good or poor fit, or in the case of the SI, which values represent high or low sensitivity, remains somehow subjective. In the context of simulated river discharge, we propose SI values below 0.5 to indicate high sensitivity.

The choice of the baseline period has the highest impact on the interpretation of simulation results performed with the IPSL model and the lowest impact with the MIROC5 model. However, the average GCM SI value (Table 2) does not imply that this assumption is true for all basins and all RCPs. The results for RCP 8.5 are slightly different, where the highest SI value is also achieved with the MIROC5 model, but the lowest values with GFDL (Table D1).

Assuming an SI threshold of 0.5, it mattered in 25% of the simulations under RCP 2.6 (Table 2) and in 32% under RCP 8.5 (Table D1), whether the one or the other baseline was used to assess future changes. There are basically two options to deal with simulations resulting in $SI \leq 0.5$: (i) discuss the uncertainties and/or (ii) choose a different baseline that represents the basic statistical properties of the historical period more consistently, e.g. by using the proposed algorithm in Appendix A.

To exemplify, the projected discharge changes with an additional flexible baseline for the Volta River basin is shown in Figure 3. It explains why the results derived with both baselines are so different. The MQ values for the WMO and IPCC baselines are $1454\,m^3s^{-1}$ and $1115\,m^3s^{-1}$, respectively. The algorithm identifies the 30-years flexible baseline 1972–2001 with an MQ value of $1361\,m^3s^{-1}$, which is much closer to the MQ of the WMO than to the MQ of the IPCC baseline. Furthermore, the range of values of the IPCC baseline is much smaller. Where the years with the lowest discharges are identical, the highest MQ is only $1800\,m^3s^{-1}$ but $3150\,m^3s^{-1}$ in the WMO and flexible baselines. This would make a significant difference in an analysis of the distribution of wet, dry, and extreme years.

Results from all river basins under both RCPs show that the projected ΔMQ series using flexible baselines lie either in between or outside WMO and IPCC ΔMQ. But, in all cases, they resemble the WMO more than the IPCC ΔMQ series (Figures in Appendix C), which is an indication that also the length of the baseline period matters.

Table 3 shows relative differences of ensemble ΔMQ between WMO, IPCC, and flexible baselines for all river basins around the central years 2040, 2060, and 2080 for RCP 2.6 and Table E1 for RCP 8.5. In the Northern Dvina River basin (DVI) in 2040 and 2060 and in the São Francisco River basin (SFC) in 2080, the ensemble mean projects opposing change signals between WMO and IPCC baselines, with absolute differences up to 13.9% under RCP 2.6 and almost 20% under RCP 8.5. Relative differences between WMO and IPCC baselines are lower if the ensemble mean is considered (Appendix E), but can be very high for individual models, as was shown in Figure 2 a and c. As with individual models, the ensemble mean ΔMQ series of the flexible baseline are always more similar to the WMO than to the IPCC ΔMQ series.

The sensitivity ($SI \leq 0.5$) of the choice of the baseline period for different climate zones is inconclusive (Table 2 and Table E1). A larger sample size of catchments from various climate zones is required to make more robust statements. However, the lowest sensitivity was achieved in warm temperate climates (C) represented by the Rhine, Tagus, and Upper Blue Nile River basins.

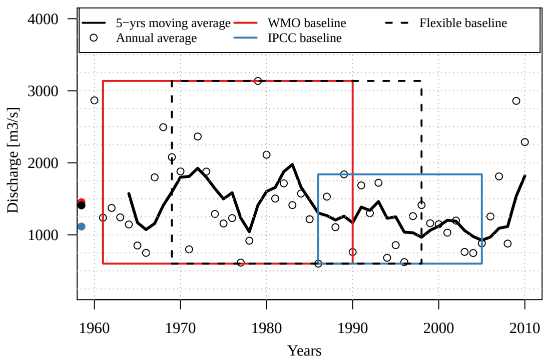
Table 2. Similarity index SI between WMO and IPCC ΔMQ series (mean of period 2020–2080), RCP 2.6

	DVI	RHI	SFC	TAG	UBN	UMI	VOL	YEL	Average
GFDL	0.19	0.82	0.82	0.68	0.93	0.49	0.54	0.44	0.61
HadGEM	0.93	0.53	0.62	0.65	0.97	0.35	0.75	0.62	0.68
IPSL	0.43	0.65	0.15	0.55	0.88	0.73	0.24	0.95	0.57
MIROC5	0.86	0.84	0.81	–	0.19	0.92	0.8	0.62	0.72
Average	0.60	0.71	0.60	0.62	0.74	0.62	0.58	0.66	

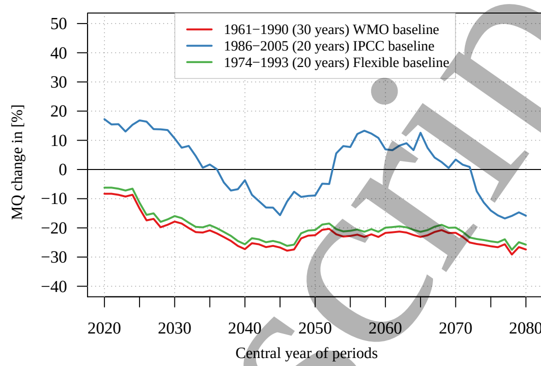
Table 3. Ensemble mean ΔMQ in selected future periods in [%] relative to MQ of WMO, IPCC, and flexible baseline, RCP 2.6

	2040			2060			2080		
Basins	WMO	IPCC	Flex.	WMO	IPCC	Flex.	WMO	IPCC	Flex.
DVI	9.2	-4.7	7.9	7.2	-4.7	6.0	8.0	0.7	6.6
RHI	-4.2	-5.4	-4.8	-0.8	-3.1	-1.4	3.0	2.1	2.3
SFC	-12.8	-6.3	-12.1	-14.0	-8.0	-10.4	-5.0	8.8	-4.5
TAG	-5.6	-3.7	-5.6	0.3	0.0	-0.3	7.6	11.5	7.0
UBN	18.5	14.7	16.0	22.4	15.7	19.8	17.9	13.3	15.2
UMI	-8.5	-7.9	-8.4	-4.1	-0.7	-3.5	1.1	3.9	2.2
VOL	10.6	8.8	11.2	9.3	16.6	10.1	2.4	7.3	3.3
YEL	1.8	5.5	4.0	7.7	13.3	9.7	5.6	11.8	7.7
Mean diff.	4.2			5.5			5.5		
Min. diff.	0.6			0.3			0.8		
Max. diff.	13.9			11.9			13.9		
Median diff.	2.8			5.8			4.7		

Last four rows indicate differences between WMO and IPCC ΔMQ series
Flex. = flexible baseline



(a) Baseline periods



(b) Discharge change

Figure 3. a) MQ in the Volta River basin simulated with IPSL and the range and timing of baselines, b) Relative change in MQ of central years in the Volta basin with IPSL under RCP 2.6 (WMO, IPCC, and flexible baselines)

4. Conclusions

This study demonstrates how solely the choice of a baseline period can influence the interpretation of discharge projections in eight river basins using climate input from four bias-adjusted GCMs. To evaluate whether the choice of either the well established 30-years (1961–1990) WMO [17] or the more recent 20-years (1986–2005) IPCC [20] baseline matters, a similarity index SI was introduced as a measure to compare the two resulting time series of future change. In about 25% of the simulations under RCP 2.6 and in 32% under RCP 8.5, large quantitative differences and/or opposite signals of change were found, with at least one case of major discrepancies in each river basin. The deviations for selected future periods can be so large that they range from -5% to +45% for a given central year. These figures indicate that different recommendations for action could possibly be derived in at least every fourth case.

No systematic differences in the direction of change using either baseline period could be identified. Neither the results based on the WMO nor those based on the IPCC baseline tend to generally project higher or lower future river discharge.

4.1. Choosing baseline periods

Given that a baseline period is normally a subset of the historical period, it should represent its basic statistical properties. From a formal statistical perspective, a minimum length of 30 years is highly recommended for regional impact studies, particularly when using integrated variables, such as river discharge. We developed an algorithm, which identifies for each simulation a flexible baseline of variable length and variable start year representing the basic statistical properties of a given historical period. In about 20% of the 32 simulations, the flexible baselines were longer than 30 years, highlighting the importance of longer-term perspectives to more confidently quantify historic reference variability when developing adaptation strategies. The use of flexible baselines helps to reduce uncertainty in the interpretation of model simulations in cases where standard baseline periods do not capture the variability of the historical period. If multiple ranges of uncertainty, such as those implied by the impact modelling cascade and multi-criteria baseline selection, are combined, the central limit theorem implies that central tendencies are favoured at the expense of extremes [39].

4.2. Regional context

At the local and regional scales, it is important to take region-specific characteristics into account, where other factors that are largely independent of past climate variability may also influence the choice of a representative baseline period, e.g., degree of human impact (land use / cover change, reservoirs, irrigation). In this context, it is reasonable to question whether the baseline period should represent rather natural conditions (far back in time with low human impact) or more recent conditions (with strong human impact). Another reason why the application of standard baseline periods is questionable is that they are often detached from reality. If a baseline period is chosen that, for example, was characterized by severe droughts in reality and future simulations project relatively drier conditions (even though the simulated baseline was above normal), stakeholders may interpret that the future will be drier than the driest period they have experienced in their lives. Using flexible baselines is a solution to better tailor information to the needs of decision makers while addressing the challenge of uncertainty transparently and efficiently.

4.3. Ensemble mean versus single model simulations

Generally, the interpretation of results based on model ensembles is less sensitive to the choice of baseline periods than for single model simulations. Nevertheless, in three cases, even the ensemble mean using the WMO and IPCC baselines projected opposite change signals in selected future periods.

4.4. Outlook

An analysis of results based on monthly or daily time series or a focus on extremes rather than the average might reveal an even higher sensitivity to chosen baseline periods than the annual time series used in this study. An improvement of the algorithm to identify flexible baseline periods, in terms of incorporating more sophisticated statistical parameters and tests, might be necessary if applied to monthly or daily time series.

Acknowledgments

This research was funded in the frame of the CIREG project (https://cireg.pik-potsdam.de/en/) by ERA-NET Co-fund action initiated by JPI Climate, funded by BMBF (DE), FORMAS (SE), BELSPO (BE), and IFD (DK) with co-funding by the European Union‘s Horizon 2020 Framework Program (Grant 690462). We thank our colleagues Dr. Valentina Krysanova, Iulii Didovets, Samuel Fournet, and the two anonymous reviewers for their inspiring ideas.

Appendix A. Baseline algorithm (R)

```
#---------------------------------------------------------------------------
assess_deviations <- function(ts, ncycle, base_length, ts.stats) {

   # numeric vectors storing deviations between
   # mean maximumm, and minimum values
   # between flexible baseline and entire historical period
   dev.mean <- numeric(ncycle)
   dev.max  <- numeric(ncycle)
   dev.min  <- numeric(ncycle)

   for ( y in 1:ncycle ) {
      sel         <- ts[y:(y + base_length - 1)]
      dev.mean[y] <- (mean(sel, na.rm = T) -
                     as.numeric(ts.stats[1])) / as.numeric(ts.stats[1]) * 100
      dev.max[y]  <- (max(sel, na.rm = T)  -
                     as.numeric(ts.stats[2])) / as.numeric(ts.stats[2]) * 100
      dev.min[y]  <- (min(sel, na.rm = T)  -
                     as.numeric(ts.stats[3])) / as.numeric(ts.stats[3]) * 100
   }

   assess_deviations <- list(dev.mean, dev.max, dev.min)
}
#---------------------------------------------------------------------------

#---------------------------------------------------------------------------
select_baseline <- function(ts, first_year, last_year, base_length, thresh.dev) {
   # This function returns a logical vector,
   # where TRUE-values indicate the years of the "best" baseline  period.
   # The "best" baseline period is defined as
   # the period of a given length (base_length)
   # with the lowest deviations between mean, max, and min values
   # between the flexible baseline period and the entire historical period.

   # INCOMING VARIABLES
   # ts          = time series of (annual) values
   # first_year  = first year of time series
   # last_year   = last year of time series
   # base_length = minimum number of years in baseline periods

   # number of cycles from first to last year, depending on base_length
   ncycle   <- last_year - first_year + 2 - base_length
   nyears   <- last_year - first_year + 1

   # statistics of entire time series
   ts.mean  <- mean(ts, na.rm = T)
   ts.max   <- max(ts, na.rm = T)
   ts.min   <- min(ts, na.rm = T)
   ts.stats <- list(ts.mean, ts.max, ts.min)
```

```
#--------------------------------------------------------------------------------
# iterations
#--------------------------------------------------------------------------------
bs_length         <- base_length - 1
nc                <- ncycle
index             <- NULL
ids.valid         <- NULL
select_baseline   <- vector(mode = "logical", length = nyears)

while ( length(ids.valid) == 0 ) {
   # add 1 year to baseline period if necessary during iterations
   bs_length   <- bs_length + 1

   # if length of baseline period equals entire period:
   #   - return entire period and exit function
   if ( bs_length == nyears ) {
      print("length of baseline period equals entire period")
      select_baseline[1:length(select_baseline)] <- T
      return(select_baseline)
   }

   # compute the number of cycles possible to iterate the baseline
   # period through the entire period
   nc          <- last_year - first_year + 2 - bs_length

   # compute mean, max., and min. deviations between
   # baseline and entire period
   dev.stats   <- assess_deviations(ts, nc, bs_length, ts.stats)

   # evaluate results against given threshold
   ids.valid   <- which(abs(unlist(dev.stats[1])) < thresh.dev &
                        abs(unlist(dev.stats[2])) < thresh.dev &
                        abs(unlist(dev.stats[3])) < thresh.dev)

   if ( length(ids.valid) >= 1 ) {
      if ( length(ids.valid) == 1 ) { index <- ids.valid }
      if ( length(ids.valid) > 1  ) {

         # find tuple with lowest sum
         sum.tuple<- abs(unlist(dev.stats[1]))[ids.valid] +
                     abs(unlist(dev.stats[2]))[ids.valid] +
                     abs(unlist(dev.stats[3]))[ids.valid]

         index    <- ids.valid[which(sum.tuple == min(sum.tuple))]
      }
   }
}

select_baseline[index:(index + bs_length - 1)] <- T
return(select_baseline)
}
```

#

Appendix B. Using detrended historical data

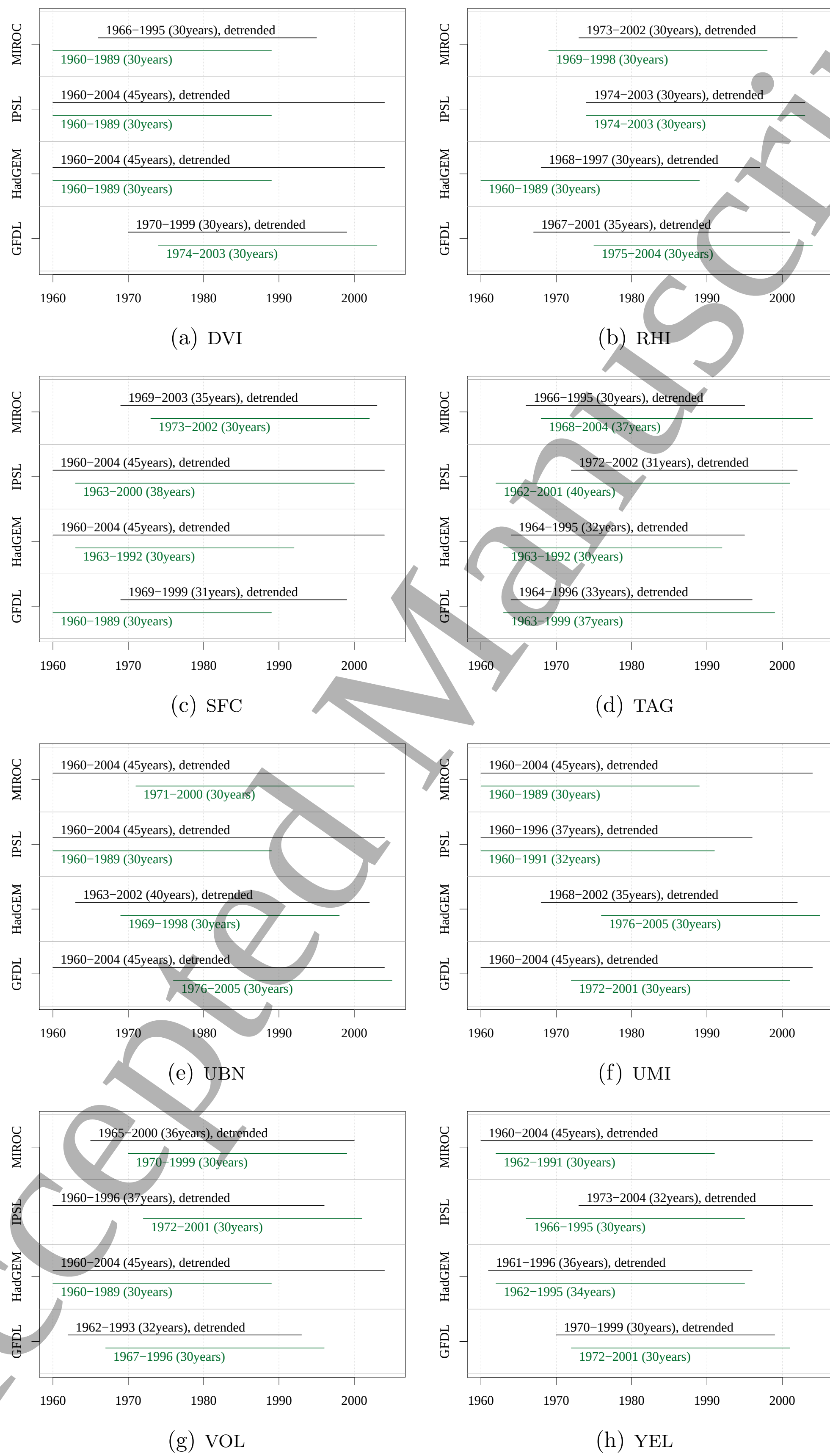
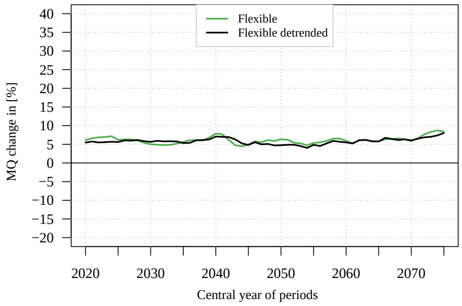
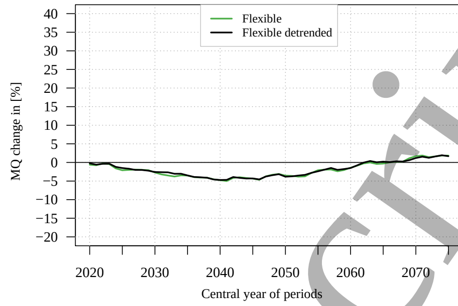


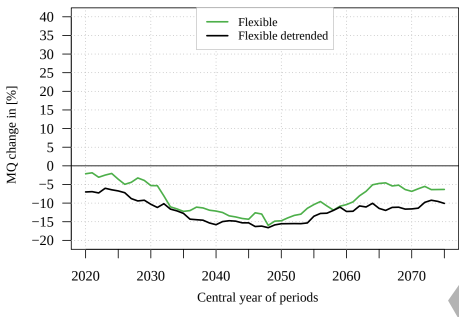
Figure B1. Length and timing of baseline periods "original" and detrended.



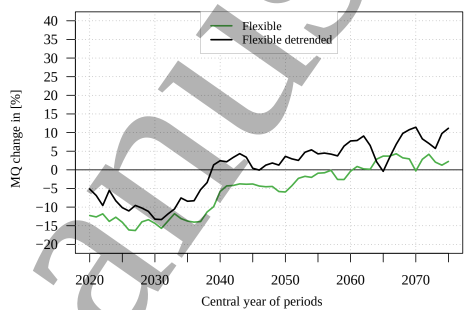
(a) DVI



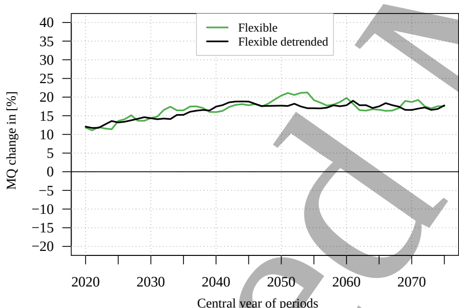
(b) RHI



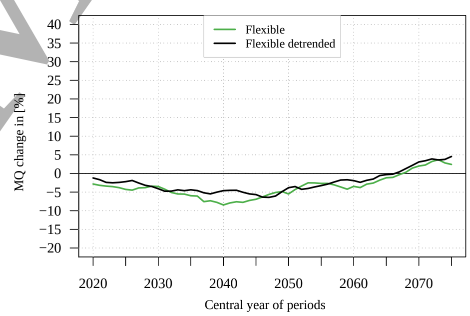
(c) SFC



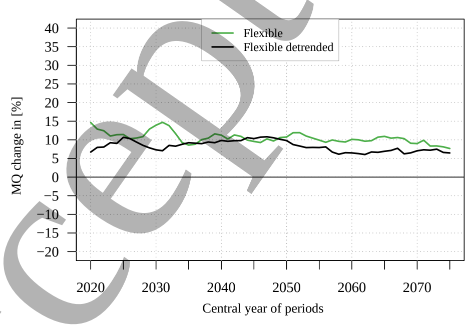
(d) TAG



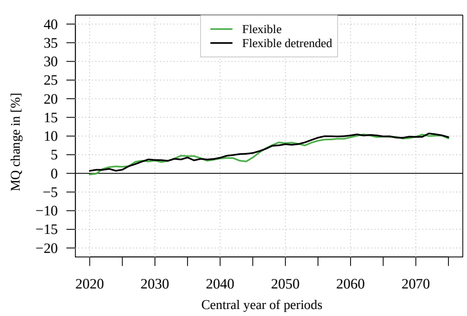
(e) UBN



(f) UMI



(g) VOL



(h) YEL

Figure B2. Ensemble mean MQ changes using the "original" flexible baseline and the detrended baseline.

Appendix C. Relative discharge changes

Appendix C.1. RCP 2.6

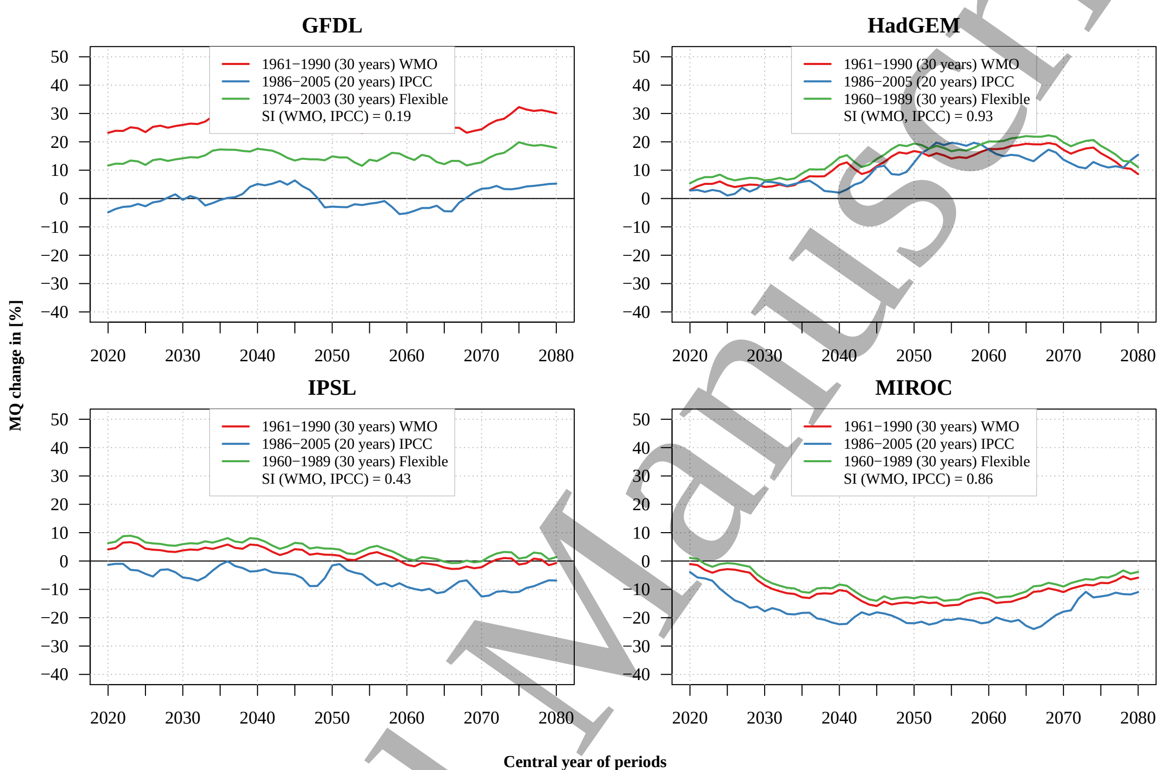


Figure C1. Relative change in mean discharge of central years. Northern Dvina River basin (DVI) under RCP 2.6. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

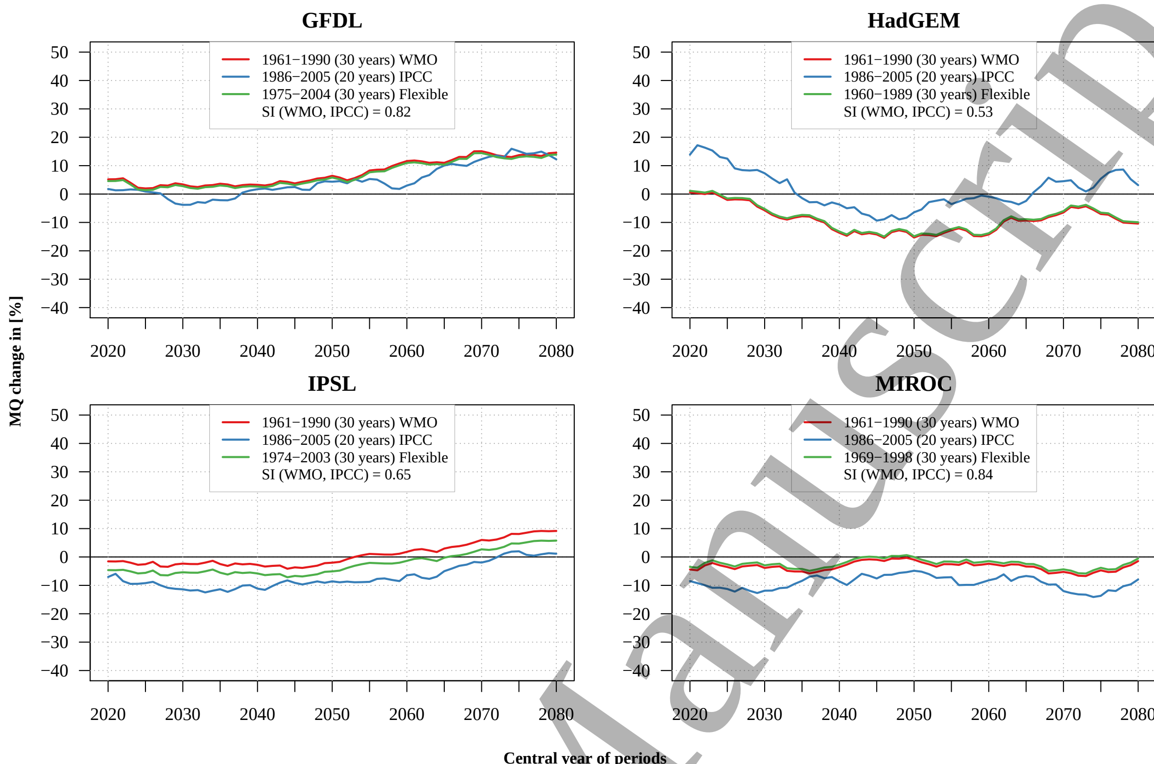


Figure C2. Relative change in mean discharge of central years. Rhine River basin (RHI) under RCP 2.6. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

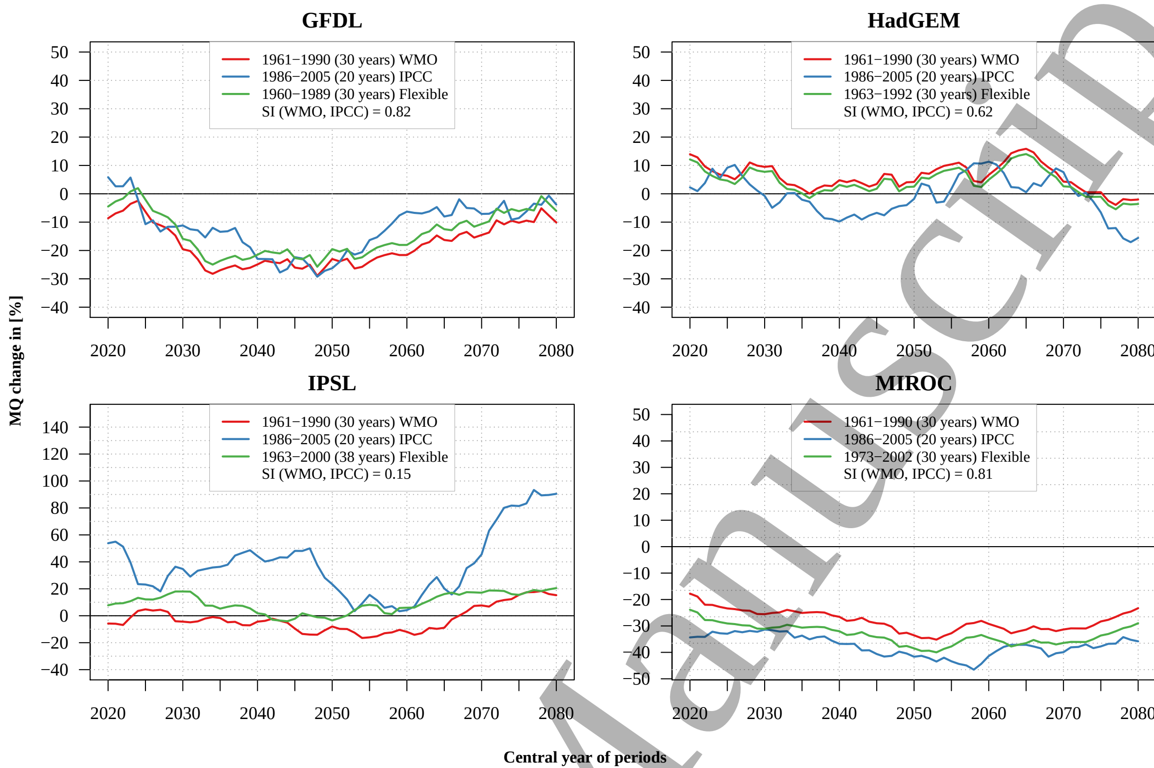


Figure C3. Relative change in mean discharge of central years. São Francisco River basin (SFC) under RCP 2.6. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

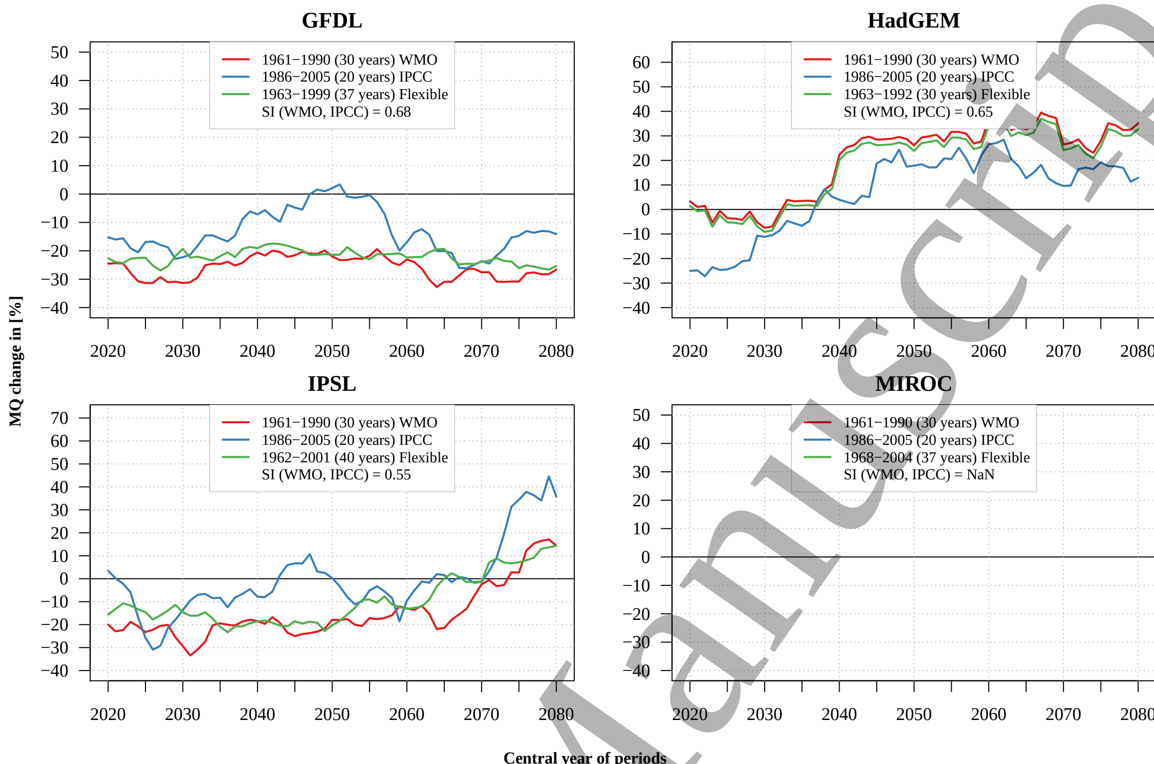


Figure C4. Relative change in mean discharge of central years. Tagus River basin (TAG) under RCP 2.6. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

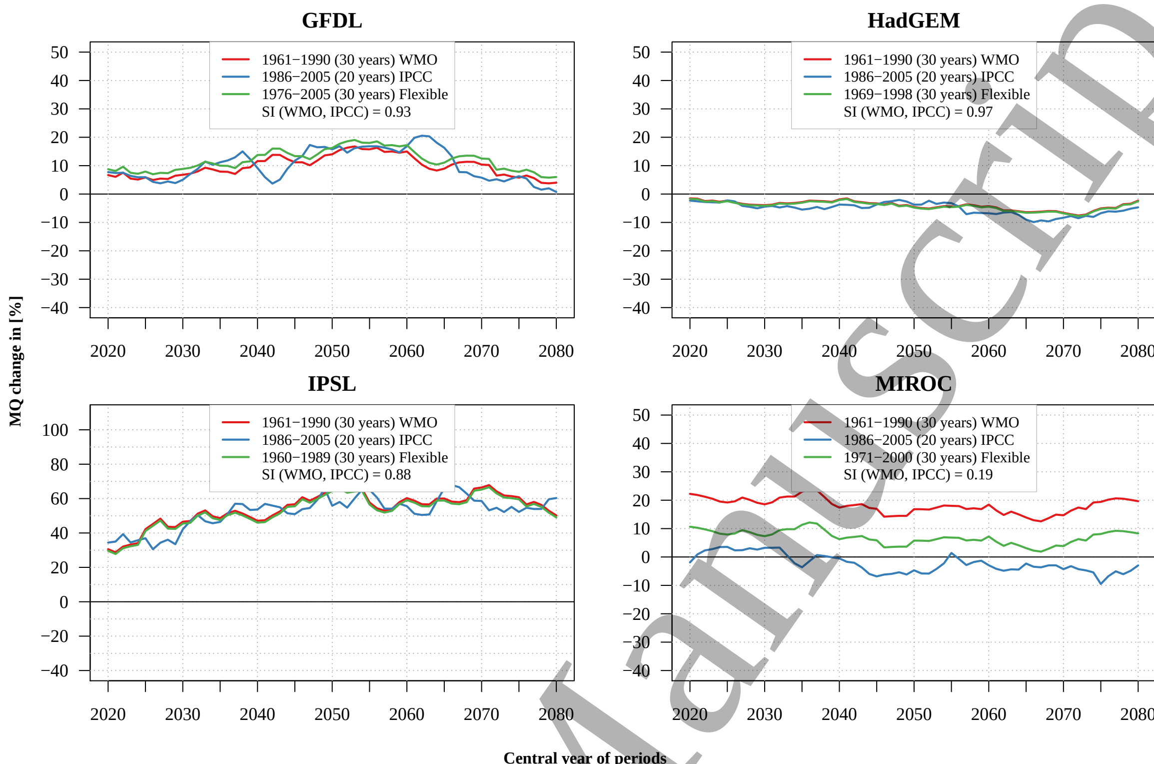


Figure C5. Relative change in mean discharge of central years. Upper Blue Nile River basin (UBN) under RCP 2.6. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

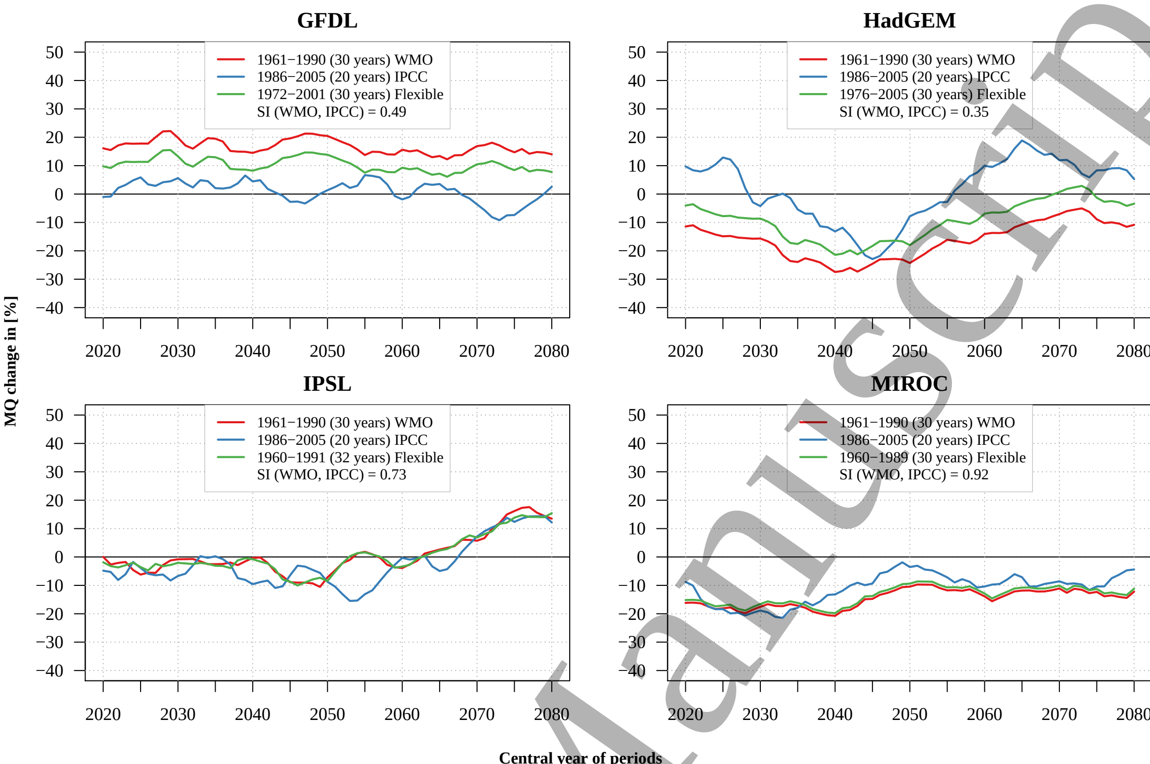


Figure C6. Relative change in mean discharge of central years. Upper Mississippi River basin (UMI) under RCP 2.6. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

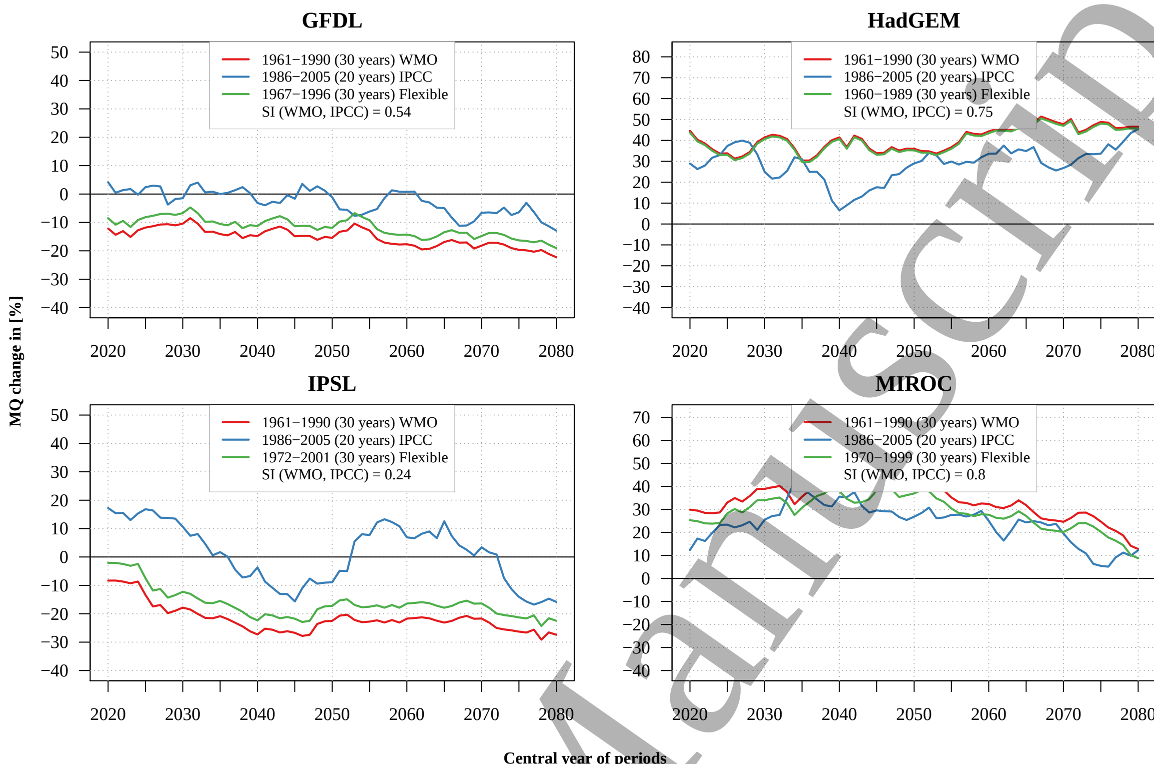


Figure C7. Relative change in mean discharge of central years. Volta River basin (VOL) under RCP 2.6. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

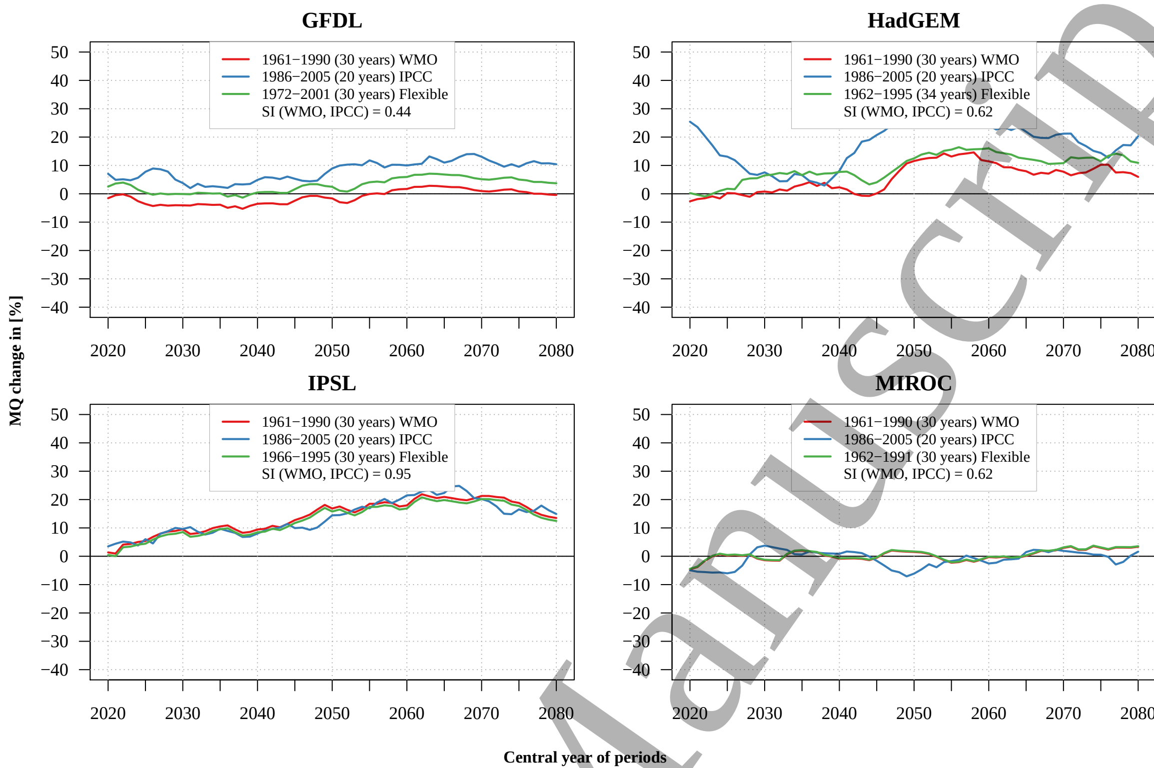


Figure C8. Relative change in mean discharge of central years. Upper Yellow River basin (YEL) under RCP 2.6. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

Appendix C.2. RCP 8.5

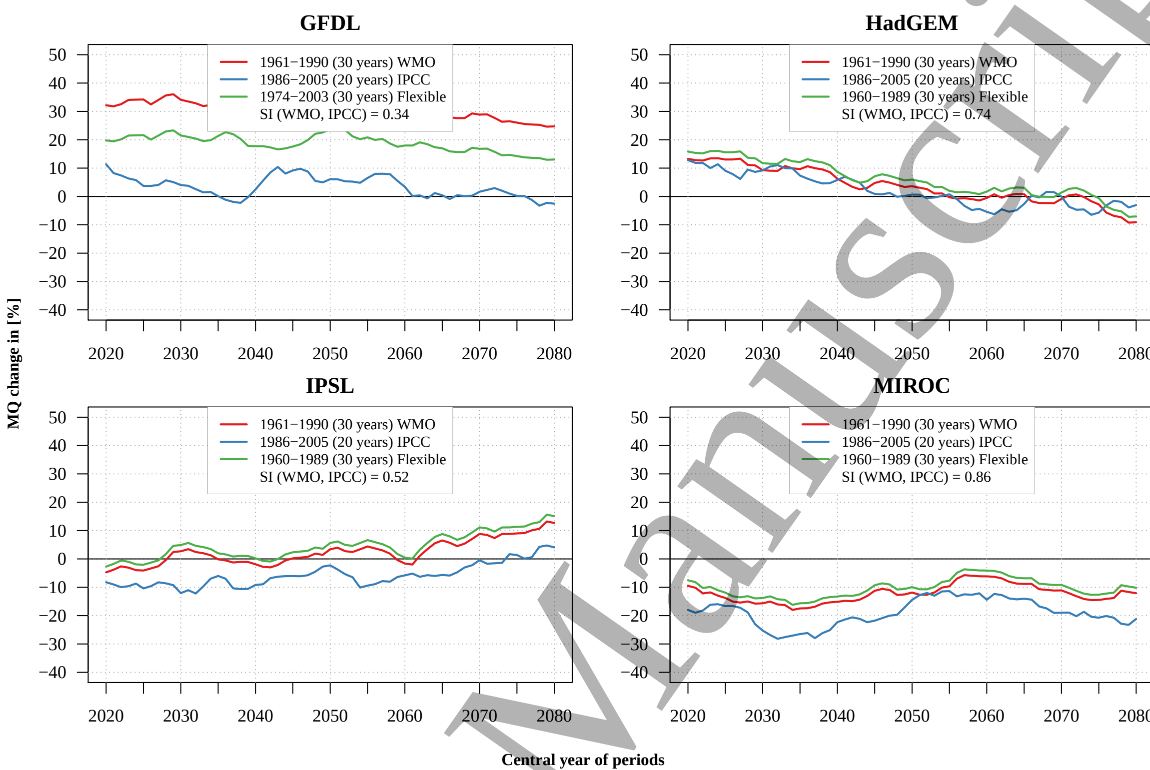


Figure C9. Relative change in mean discharge of central years. Northern Dvina River basin (DVI) under RCP 8.5. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

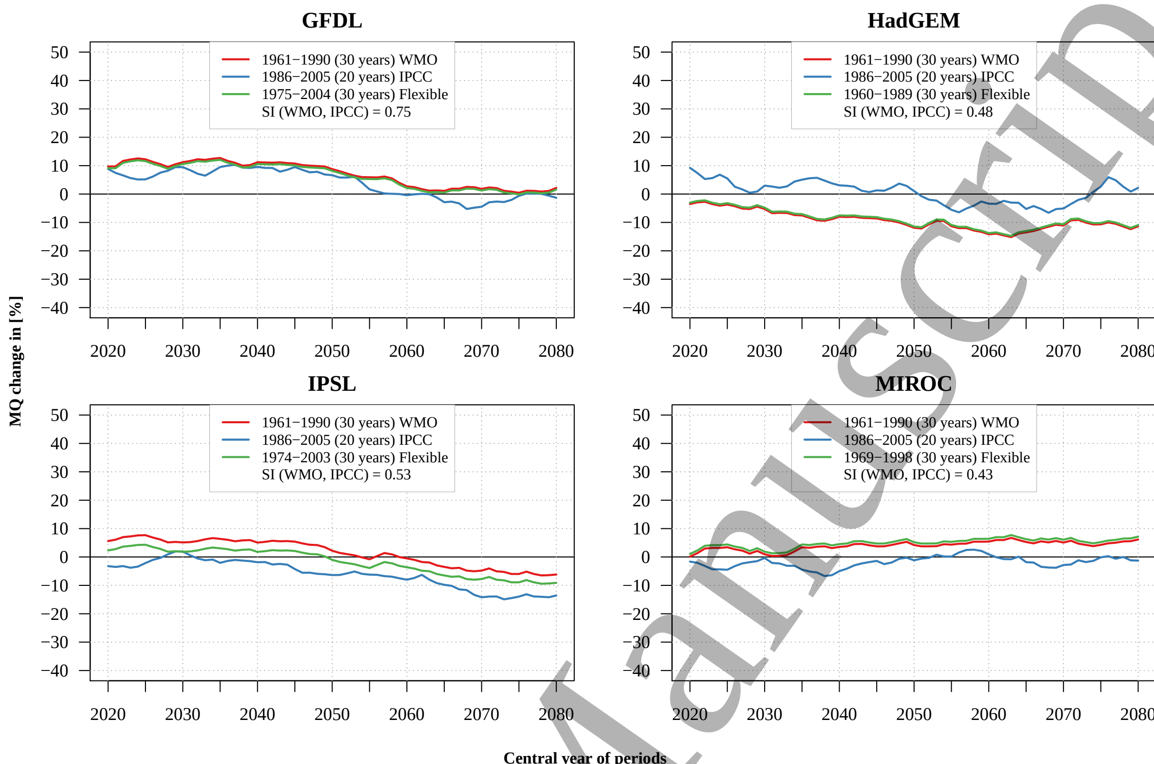


Figure C10. Relative change in mean discharge of central years. Rhine River basin (RHI) under RCP 8.5. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

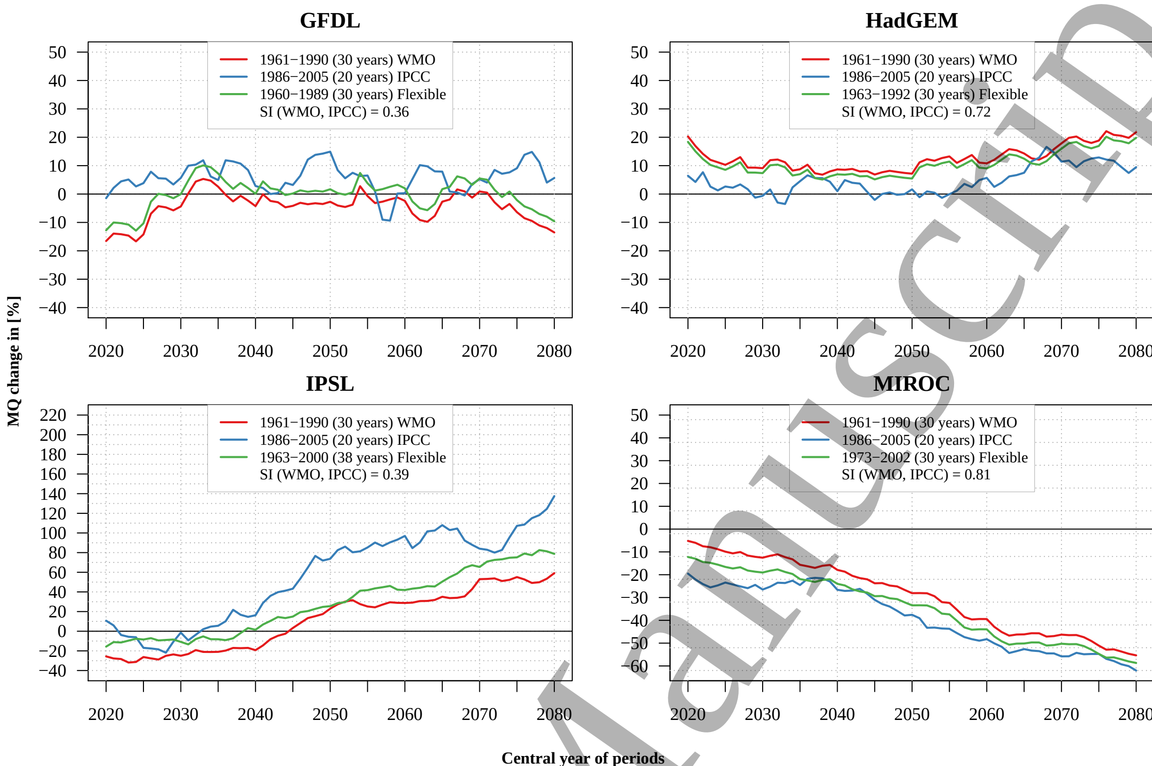


Figure C11. Relative change in mean discharge of central years. São Francisco River basin (SFC) under RCP 8.5. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

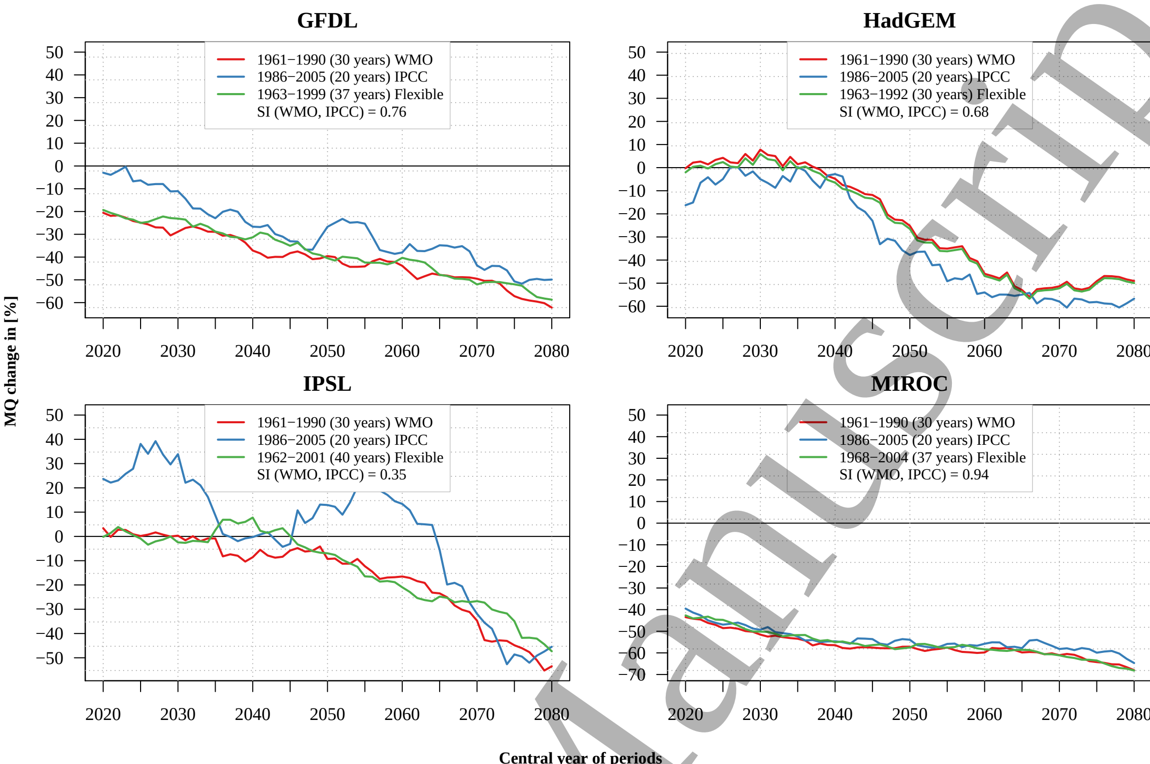


Figure C12. Relative change in mean discharge of central years. Tagus River basin (TAG) under RCP 8.5. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

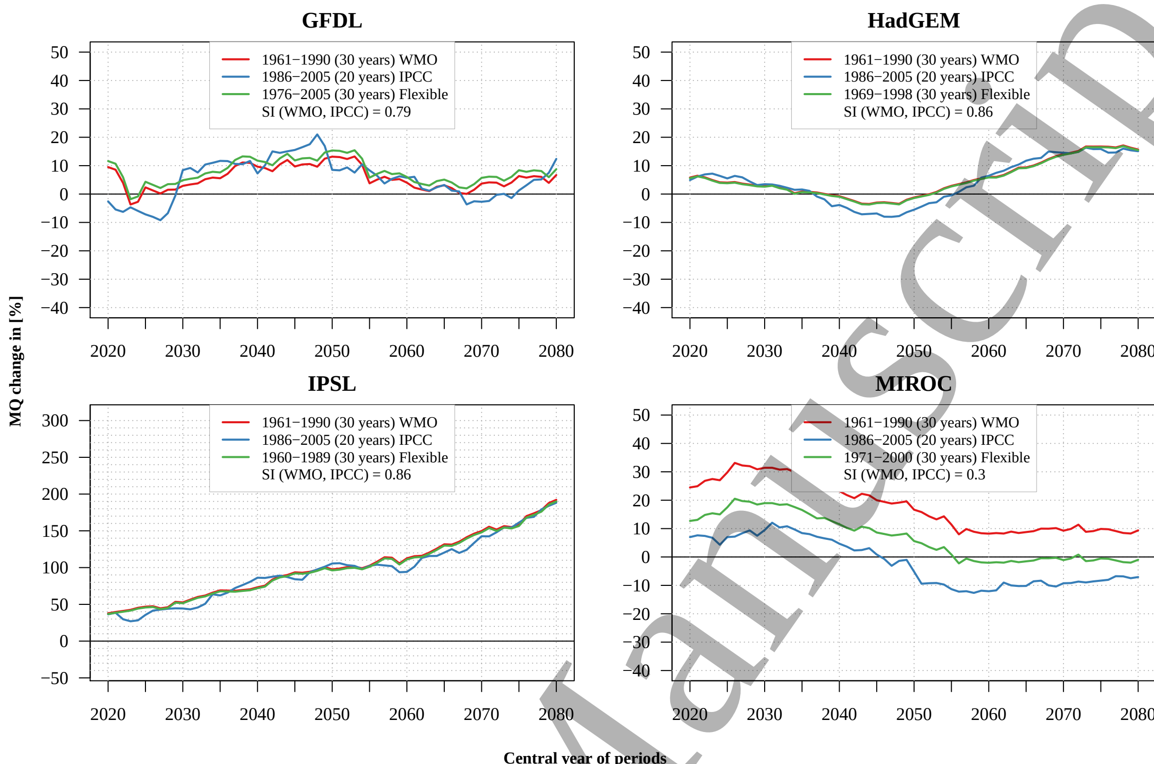


Figure C13. Relative change in mean discharge of central years. Upper Blue Nile River basin (UBN) under RCP 8.5. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

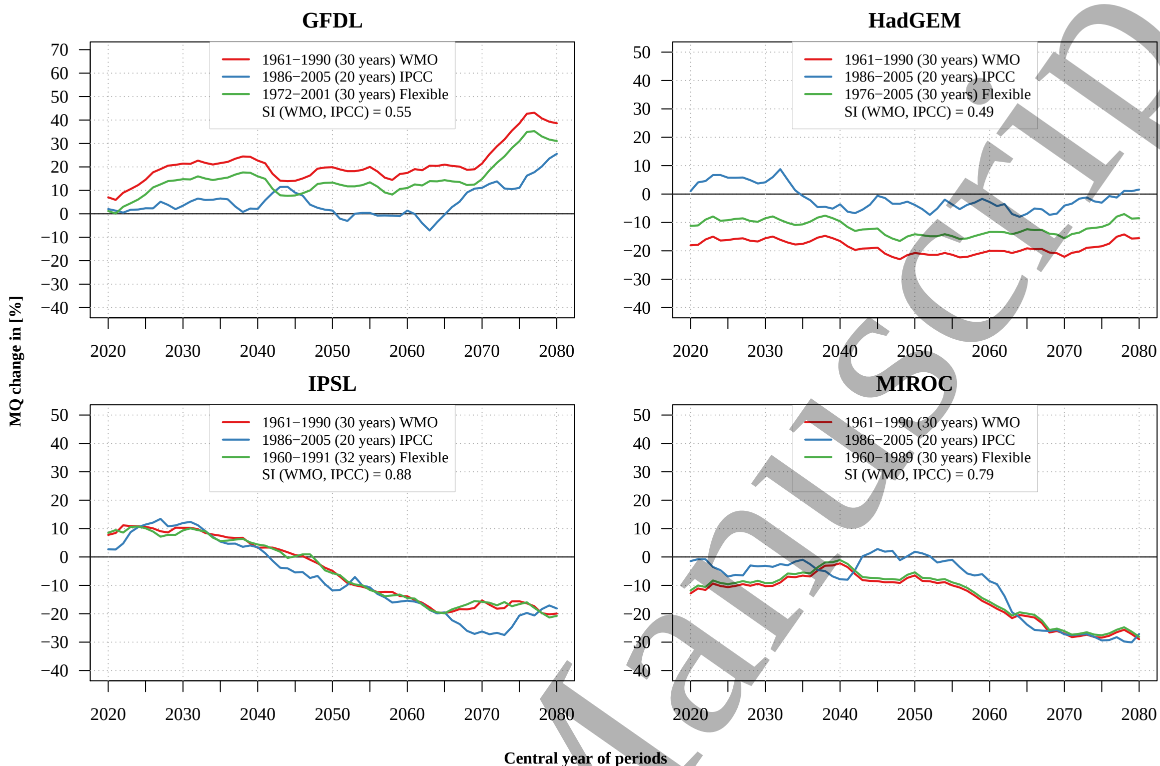


Figure C14. Relative change in mean discharge of central years. Upper Mississippi River basin (UMI) under RCP 8.5. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

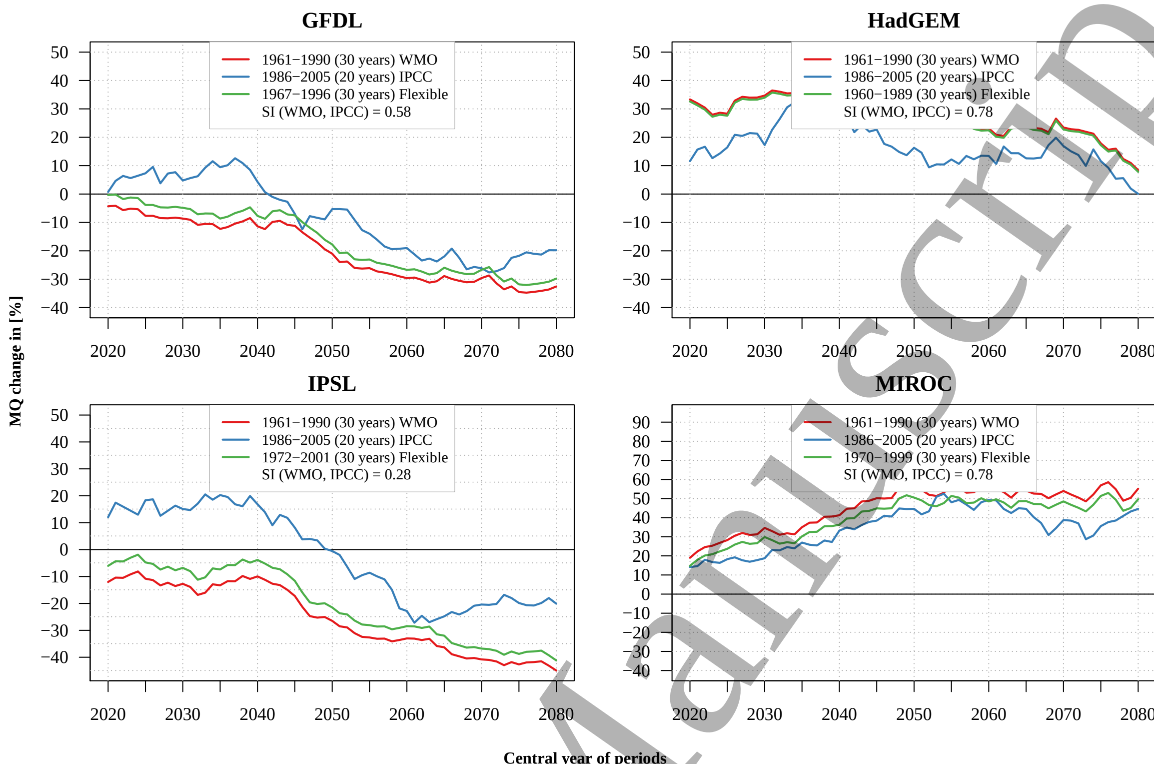


Figure C15. Relative change in mean discharge of central years. Volta River basin (VOL) under RCP 8.5. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

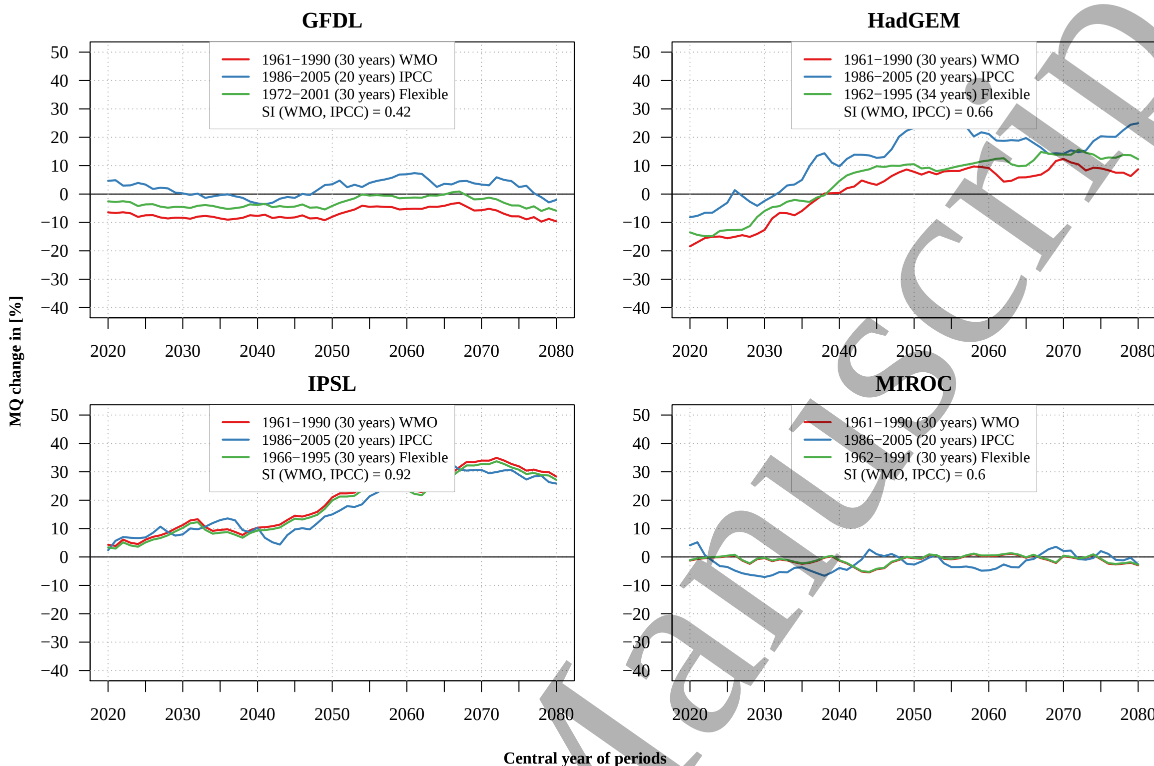


Figure C16. Relative change in mean discharge of central years. Upper Yellow River basin (YEL) under RCP 8.5. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

Appendix D. Similarity Index *SI*

Table D1. Similarity index SI between WMO and IPCC ΔMQ series, RCP 8.5

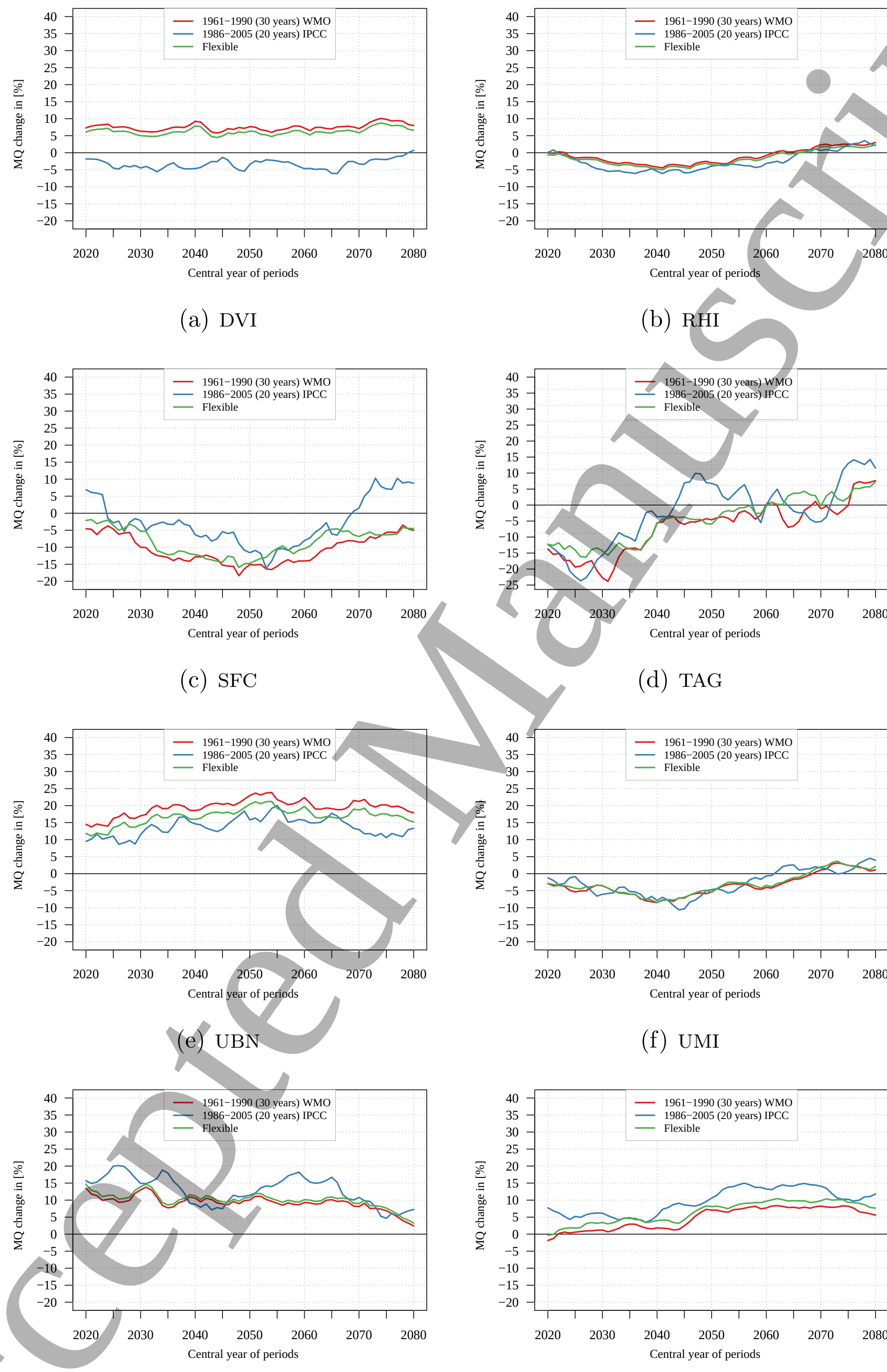
	DVI	RHI	SFC	TAG	UBN	UMI	VOL	YEL	Average
GFDL	0.34	0.75	0.36	0.76	0.79	0.55	0.58	0.42	0.57
HadGEM	0.74	0.48	0.72	0.68	0.86	0.49	0.78	0.66	0.68
IPSL	0.52	0.53	0.39	0.35	0.86	0.88	0.28	0.92	0.59
MIROC5	0.86	0.43	0.81	0.94	0.30	0.79	0.78	0.60	0.69
Average	0.62	0.55	0.57	0.68	0.70	0.68	0.61	0.65	

Appendix E. Ensemble mean

Table E1. Ensemble mean ΔMQ in selected future periods in [%] relative to MQ of WMO, IPCC, and flexible baseline, RCP 8.5

	2040			2060			2080		
Basins	WMO	IPCC	Flex.	WMO	IPCC	Flex.	WMO	IPCC	Flex.
DVI	4.8	-5.8	3.3	5.5	-5.6	4.0	4.1	-5.7	2.7
RHI	3.0	1.5	2.3	-1.6	-2.7	-2.2	-2.3	-3.5	-2.8
SFC	-8.2	-1.6	-3.9	-0.6	13.7	2.3	3.0	22.6	7.6
TAG	-26.7	-21.1	-21.8	-41.5	-33.6	-41.8	-58.2	-54.2	-55.8
UBN	26.3	23.6	23.6	32.7	23.6	30.2	56.0	52.2	53.3
UMI	1.8	-1.5	2.4	-8.3	-6.3	-8.2	-6.5	-4.5	-6.6
VOL	14.6	20.4	15.6	3.6	5.2	3.9	-3.5	1.2	-3.4
YEL	0.4	3.2	2.2	7.3	13.4	8.7	6.1	11.5	7.7
Mean diff.	4.9			6.7			6.3		
Min. diff.	1.5			1.1			1.2		
Max. diff.	10.6			14.3			19.6		
Median diff.	4.5			7.0			4.4		

Last four rows indicate differences between WMO and IPCC ΔMQ series
Flex. = flexible baseline periods



(a) DVI (b) RHI

(c) SFC (d) TAG

(e) UBN (f) UMI

(g) VOL (h) YEL

Figure E1. Relative change in mean discharge of central years using the ensemble mean under RCP 2.6. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

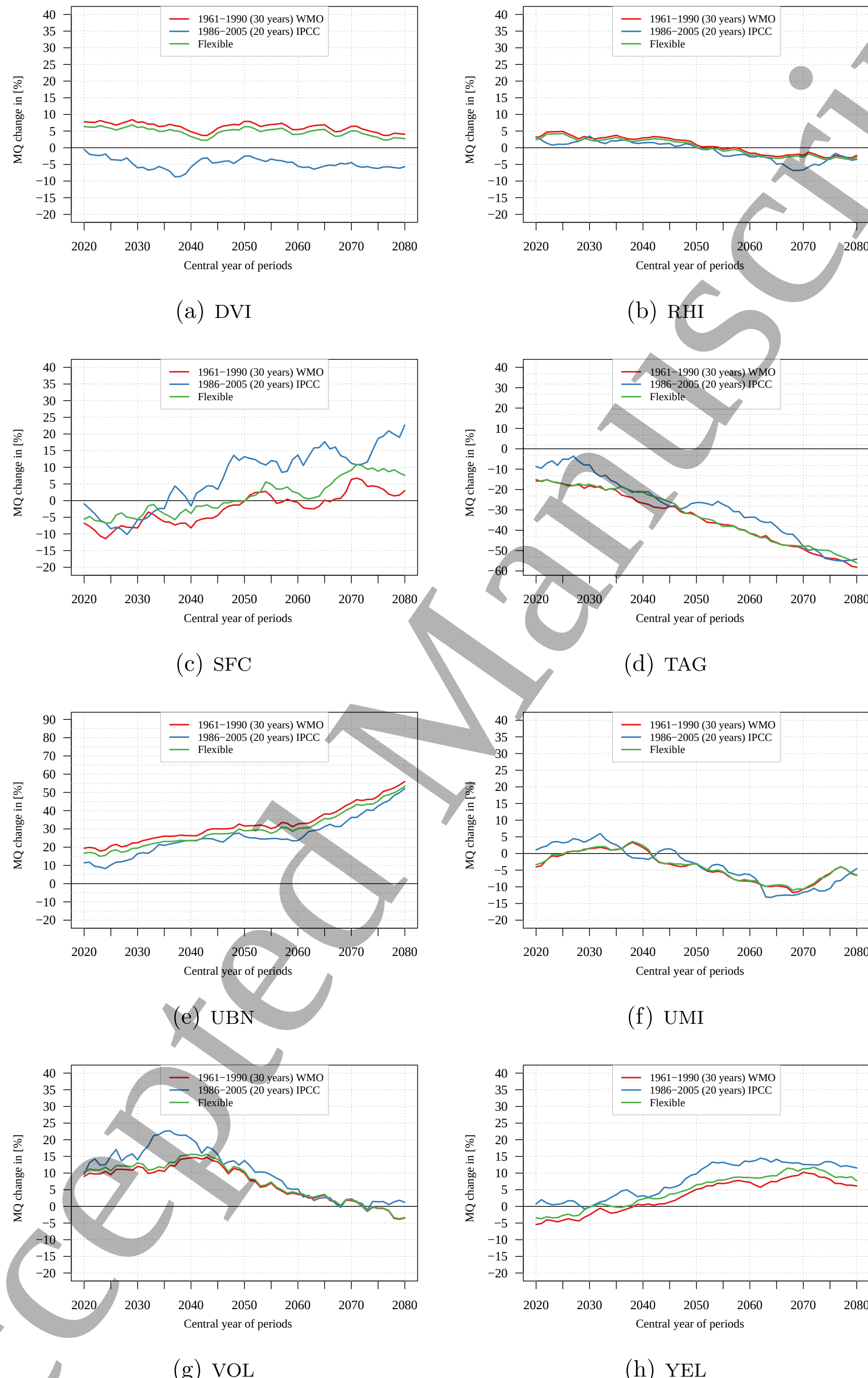


Figure E2. Relative change in mean discharge of central years using the ensemble mean under RCP 8.5. Changes are relative to mean annual discharge in WMO and IPCC baseline periods

References

[1] Porter J J and Dessai S 2017 *Environmental Science & Policy* **77** 9–14 ISSN 1462-9011 URL http://www.sciencedirect.com/science/article/pii/S1462901116308875

[2] Eyring V, Cox P M, Flato G M, Gleckler P J, Abramowitz G, Caldwell P, Collins W D, Gier B K, Hall A D, Hoffman F M, Hurtt G C, Jahn A, Jones C D, Klein S A, Krasting J P, Kwiatkowski L, Lorenz R, Maloney E, Meehl G A, Pendergrass A G, Pincus R, Ruane A C, Russell J L, Sanderson B M, Santer B D, Sherwood S C, Simpson I R, Stouffer R J and Williamson M S 2019 *Nature Climate Change* **9** 102–110 ISSN 1758-6798 URL https://doi.org/10.1038/s41558-018-0355-y

[3] Lorenz R, Herger N, Sedláček J, Eyring V, Fischer E M and Knutti R 2018 *Journal of Geophysical Research: Atmospheres* **123** 4509–4526 (*Preprint* https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2017JD027992) URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017JD027992

[4] Liersch, S, Tecklenburg, J, Rust, H, Dobler, A, Fischer, M, Kruschke, T, Koch, H, Hattermann and F F 2018 *Hydrology and Earth System Sciences* **22** 2163–2185 URL https://www.hydrol-earth-syst-sci.net/22/2163/2018/

[5] Rowell, P D, Senior, A C, Vellinga, Michael, Graham and J R 2016 *Climatic Change* **134** 621–633 ISSN 1573-1480

[6] Latif M 2011 *Journal of Geochemical Exploration* **110** 1 – 7 ISSN 0375-6742 sustainability of Geochemical Cycling URL http://www.sciencedirect.com/science/article/pii/S0375674210001433

[7] Bronstert, Axel, Kolokotronis, Vassilios, Schwandt, Daniel, Straub and Helmut 2007 *International Journal of Climatology* **27** 1579–1594 ISSN 1097-0088 URL http://dx.doi.org/10.1002/joc.1621

[8] Krysanova V, Donnelly C, Gelfan A, Gerten D, Arheimer B, Hattermann F and Kundzewicz Z W 2018 *Hydrological Sciences Journal* **63** 696–720 URL https://doi.org/10.1080/02626667.2018.1446214

[9] Hattermann F F, Vetter T, Breuer L, Su B, Daggupati P, Donnelly C, Fekete B, Flörke F, Gosling S N, Hoffmann P, Liersch S, Masaki Y, Motovilov Y, Müller C, Samaniego L, Stacke T, Wada Y, Yang T and Krysnaova V 2018 *Environmental Research Letters* **13** 015006 URL http://stacks.iop.org/1748-9326/13/i=1/a=015006

[10] Vetter, Tobias, Reinhardt, Julia, Flörke, Martina, van Griensven, Ann, Hattermann, Fred, Huang, Shaochun, Koch, Hagen, Pechlivanidis, G I, Plötner, Stefan, Seidou, Ousmane, Su, Buda, Vervoort, Willem R, Krysanova and Valentina 2017 *Climatic Change* **141** 419–433 ISSN 1573-1480

[11] Kent C, Chadwick R and Rowell D P 2015 *Journal of Climate* **28** 4390–4413 (*Preprint* https://doi.org/10.1175/JCLI-D-14-00613.1) URL https://doi.org/10.1175/JCLI-D-14-00613.1

[12] Remesan R and Holman I P 2015 *Journal of Hydrology* **528** 631 – 642 ISSN 0022-1694 URL http://www.sciencedirect.com/science/article/pii/S0022169415004394

[13] Elshamy, M, di Baldassarre, G and van Griensven A 2013 *Journal of hydrologic engineering ASCE* **18** 582–589

[14] Knutti R and Sedlacek J 2012 *Nature Clim. Change* **advance online publication** – ISSN 1758-6798 URL http://dx.doi.org/10.1038/nclimate1716

[15] Teutschbein C and Seibert J 2012 *Journal of Hydrology* **456-457** 12 – 29 ISSN 0022-1694 URL http://www.sciencedirect.com/science/article/pii/S0022169412004556

[16] WMO 2007 The role of climatological normals in a changing climate Tech. Rep. WMO/TD-No. 1377; WCDMP-No. 61 World Meteorological Organization (WMO) URL https://library.wmo.int/doc_num.php?explnum_id=4546

[17] WMO 2014 URL https://bit.ly/3aN7Zdr

[18] Hawkins E and Sutton R 2016 *Bulletin of the American Meteorological Society* **97** 963–980 (*Preprint* `https://doi.org/10.1175/BAMS-D-14-00154.1`)

[19] WMO 2017 WMO Guidelines on the Calculation of Climate Normals Tech. Rep. WMO- No. 1203 World Meteorological Organization (WMO) URL `https://library.wmo.int/doc_num.php?explnum_id=4166`

[20] IPCC 2014 Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change Tech. rep. Intergovernmental Panel on Climate Change URL `https://www.ipcc.ch/report/ar5/syr/`

[21] Kravtsov S, Grimm C and Gu S 2018 *npj Climate and Atmospheric Science* **1** 34 ISSN 2397-3722 URL `https://doi.org/10.1038/s41612-018-0044-6`

[22] Ruokolainen L and Räisänen J 2007 *Tellus A: Dynamic Meteorology and Oceanography* **59** 309–320

[23] Razavi S, Elshorbagy A, Wheater H and Sauchyn D 2015 *Water Resources Research* **51** 1813–1830 (*Preprint* `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014WR015696`) URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014WR015696`

[24] Huang S, Kumar R, Rakovec O, Aich V, Wang X, Samaniego L, Liersch S and Krysanova V 2018 *Environmental Research Letters* **13** 124005 URL `http://stacks.iop.org/1748-9326/13/i=12/a=124005`

[25] Snell R S, Elkin C, Kotlarski S and Bugmann H 2018 *Regional Environmental Change* **18** 2145–2159 ISSN 1436-378X URL `https://doi.org/10.1007/s10113-018-1337-3`

[26] Baker D J, Hartley A J, Butchart S H M and Willis S G 2016 *Global Change Biology* **22** 2392–2404 (*Preprint* `https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.13273`) URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.13273`

[27] Warszawski L, Frieler K, Huber V, Piontek F, Serdeczny O and Schewe J 2014 *Proceedings of the National Academy of Sciences* **111** 3228–3232

[28] Huber V, Schellnhuber H J, Arnell N W, Frieler K, Friend A D, Gerten D, Haddeland I, Kabat P, Lotze-Campen H, Lucht W, Parry M, Piontek F, Rosenzweig C, Schewe J and Warszawski L 2014 *Earth System Dynamics* **5** 399–408 URL `https://www.earth-syst-dynam.net/5/399/2014/`

[29] Hempel S, Frieler K, Warszawski L, Schewe J and Piontek F 2013 *Earth System Dynamics Discussions* **4** 49–92 URL `http://www.earth-syst-dynam-discuss.net/4/49/2013/`

[30] Kottek M, Grieser J, Beck C, Rudolf B and Rubel F 2006 *Meteorologische Zeitschrift* **15** 259–263 URL `http://dx.doi.org/10.1127/0941-2948/2006/0130`

[31] Lobanova A, Liersch S, Nunes J P, Didovets I, Stagl J, Huang S, Koch H, López M R, Maule C F, Hattermann F and Krysanova V 2018 *Journal of Hydrology: Regional Studies* **18** 15–30 ISSN 2214-5818 URL `https://www.sciencedirect.com/science/article/pii/S2214581817303877`

[32] Koch H, Liersch S, Azevedo J R G D, Silva A L C and Hattermann F F 2018 *Hydrology Research* **49** 1831–1846 URL `https://iwaponline.com/hr/article-abstract/49/6/1831/41218/Assessment-of-observed-and-simulated-low-flow?redirectedFrom=fulltext`

[33] Liersch S, Koch H and Hattermann F F 2017 *Water* **9** ISSN 2073-4441 URL `http://www.mdpi.com/2073-4441/9/10/728`

[34] Aich V, Liersch S, Vetter T, Huang S, Tecklenburg J, Hoffmann P, Koch H, Fournet S, Krysanova V, Müller E N and Hattermann F F 2014 *Hydrology and Earth System Sciences* **18** 1305–1321 URL `http://www.hydrol-earth-syst-sci.net/18/1305/2014/`

[35] Krysanova, V, Müller-Wohlfeil, DI, Becker and A 1998 *Ecological Modelling* **106** 261–289

[36] Krysanova V, Hattermann F and Wechsung F 2005 *Hydrological Processes* **19** 763–783

[37] Frieler K, Lange S, Piontek F, Reyer C P O, Schewe J, Warszawski L, Zhao F, Chini L, Denvil S, Emanuel K, Geiger T, Halladay K, Hurtt G, Mengel M, Murakami D, Ostberg S, Popp A, Riva R, Stevanovic M, Suzuki T, Volkholz J, Burke E, Ciais P, Ebi K, Eddy T D, Elliott J, Galbraith E, Gosling S N, Hattermann F, Hickler T, Hinkel J, Hof C, Huber V, Jägermeyr J, Krysanova V, Marcé R, Müller Schmied H, Mouratiadou I, Pierson D, Tittensor D P, Vautard R,

van Vliet M, Biber M F, Betts R A, Bodirsky B L, Deryng D, Frolking S, Jones C D, Lotze H K, Lotze-Campen H, Sahajpal R, Thonicke K, Tian H and Yamagata Y 2017 *Geoscientific Model Development* **10** 4321–4345 URL https://gmd.copernicus.org/articles/10/4321/2017/

[38] Trenberth K E, Dai A, van der Schrier G, Jones P D, Barichivich J, Briffa K R and Sheffield J 2014 *Nature Climate Change* **4** 17–22 ISSN 1758-6798 URL https://doi.org/10.1038/nclimate2067

[39] Wigley T M L and Raper S C B 2001 *Science* **293** 451–454 ISSN 0036-8075 (*Preprint* https://science.sciencemag.org/content/293/5529/451.full.pdf) URL https://science.sciencemag.org/content/293/5529/451