Hydrol. Earth Syst. Sci., 29, 5535–5553, 2025
https://doi.org/10.5194/hess-29-5535-2025

Hydrology and Earth System Sciences
Open Access
EGU

Using century-long reanalysis and a rainfall-runoff model to explore multi-decadal variability in catchment hydrology at the European scale

Pierre Brigode[1] **and Ludovic Oudin**[2]

[1]Univ Rennes, CNRS, Géosciences Rennes – UMR 6118, Rennes 35000, France
[2]Sorbonne Université, Université PSL, EPHE, CNRS, Milieux Environnementaux, Transferts et Interactions dans les hydrosystèmes et les Sols, METIS, Case 105, 4 place Jussieu, 75005 Paris, France

Correspondence: Pierre Brigode (pierre.brigode@ens-rennes.fr)

Received: 24 October 2024 – Discussion started: 6 November 2024
Revised: 3 September 2025 – Accepted: 16 September 2025 – Published: 22 October 2025

Abstract. This study explores the ability of global reanalyses to simulate catchment hydrology at the European scale using a conceptual rainfall–runoff model. We used two reanalyses, NOAA 20CR and ERA-20C, to simulate daily streamflows for over 2000 catchments since the 1840s. Our findings show that both reanalyses perform well, particularly for mean flows, with simulation performance improving as catchment size increases, though challenges remain for Mediterranean and snow-dominated regions. Additionally, the study highlights significant multi-decadal variations in streamflow, revealing alternating wet and dry periods across Europe. These findings provide valuable insights into long-term hydrological trends and offer a useful framework for understanding future changes in both water resources and hydrological extremes, such as floods, under climate variability.

1 Introduction

Catchment hydrology varies across different time scales: Wetter- and drier-than-usual periods are observed on relatively "short" time scales such as days and seasons (e.g., flood and seasonal regimes) but also on "longer" time scales such as years and decades. For example, southeastern Australia faced a decade-long drought (named the "Millennium Drought") that started in the late 1990s, leading to significant changes in the rainfall–runoff relationship in certain catchments (Fowler et al., 2022). In Europe, different observations have documented flood-rich and flood-poor periods over the past decades and centuries (Blöschl et al., 2020; Wilhelm et al., 2022; Tarasova et al., 2023). Our understanding of such "long-term" variability is still limited compared with the understanding of the daily and seasonal variability (Montanari, 2012), mainly due to the relatively short period of continuous flow recordings. Yet, detecting and understanding the origin of these periods are essential in the context of climate change and for projections of changes in water resources and associated extreme events such as droughts and floods (Blöschl et al., 2019).

During the past few decades, changes in catchment time series were sought based on different assumptions and therefore different methods and tools. The first common assumption is that linear or monotonous trends may be present in hydrometeorological data. Stahl et al. (2010) performed one of the first pan-European analyses looking at streamflow trends in monthly streamflow for the period 1962–2004 across 441 catchments. They highlighted decreasing streamflow trends at the annual scale in the southern and eastern European regions and positive trends elsewhere. Masseroni et al. (2021) recently analyzed trends in the annual streamflow volume from 1950 using a larger set of catchments (more than 3400) and also showed significant negative trends for the Mediterranean catchments and positive trends in the northern regions. These positive trends were also detected by Teutschbein et al. (2022) in 50 catchments in Sweden over the past 60 years. Gudmundsson et al. (2019) used the GSIM dataset (Do et al., 2018; Gudmundsson et al., 2018) to discuss changes in low, mean, and high streamflow values at the

regional scale, and they identified negative trends in all flow indices for the southern regions of Europe and positive trends in the northern regions. Although Nasreen et al. (2022) reported negative trends when analyzing their 500-year annual flow reconstruction over 14 European catchments, the associated signal "is not linear," with wetter and drier periods identified for the catchments studied. These analyses reveal significant hydrological trends at the European scale since the 1950s, with wetter catchments in the north and drier catchments in the south; but these analyses lead to more nuanced conclusions when viewed from a deeper historical perspective.

Another common assumption is that there are potential periodicities in the variability in hydrological processes over several decades and that these periodicities can be identified using signal-processing techniques. Applying wavelet analysis to more than 1800 monthly streamflow series available since 1962 over western Europe, Lorenzo-Lacruz et al. (2022) identified a 7-year cycle in a large proportion of catchments since the mid-1980s. This cycle was not present in earlier periods, suggesting recent changes in the periodicities of streamflows over the study regions. A 7.5-year periodicity was also identified by Rust et al. (2022) when correlating monthly streamflow variations of 767 UK catchments and the North Atlantic Oscillation (NAO). Fossa et al. (2021) studied 152 French catchments since 1958 and identified three significant timescales of variability (1, 2–4, and 5–8 years). Thus, these studies reveal significant periodicities of streamflows over particular European regions, in relation to large-scale climate variability and periodicity. For example, Haslinger et al. (2021) highlighted a significant multi-decadal variability in summer precipitation, potentially due to changes in atmospheric circulation related to the Atlantic Multidecadal Variability (AMV), as was previously shown for the northern part of Europe (Ghosh et al., 2017). The multi-decadal variability in floods has been illustrated at the European scale by Blöschl et al. (2020) and Brönnimann et al. (2022), identifying "flood-rich" and "flood-poor" periods, linked to changes in air temperature, atmospheric circulation, and atmospheric moisture. Renard et al. (2023) also identified flood "hot moments" and "hot spots" at the global scale in their analysis of 180-year flood and heavy precipitation reconstructions. Giuntoli et al. (2013) found both a significant increasing trend in drought severity and a correlation with climate indices (e.g., the North Atlantic Oscillation (NAO) and the Atlantic Multidecadal Oscillation (AMO)) in southern France since 1948, and thus stated that "these observations highlight the difficulties in distinguishing between long-term trends and low-frequency variability based on relatively short series".

A major challenge in identifying trends, periodicity, or both in catchments at the European scale is how to cope with the variability in length and continuity of hydrometeorological observations. One way to extend the period of observation both in space and time is to use dedicated reanalyses as inputs of rainfall–runoff models. In this context, several global reanalyses such as the NOAA 20CR (20th Century Reanalysis, Slivinski et al., 2019) have been specifically produced for the assessment of the past century. In its third revision, 20CR is available over the period 1836–2015 and provides 3-hourly meteorological values across 75 km grids. This reanalysis has been used as boundary conditions for the simulation of an extreme event (Parodi et al., 2017), to discuss trends in weather patterns over specific regions (e.g., Blanc et al., 2022), or for the reconstitution of past precipitation using analog methods (Horton, 2022). Yet, such reanalyses are not widely used as inputs of rainfall–runoff models to study multi-decadal variations in catchment hydrology at the continental scale, being applied at the national scale instead, such as in France (e.g., Kuentz et al., 2015; Bonnet et al., 2020; Devers et al., 2021). Simulating catchment hydrology at the multi-decadal scale using such reanalyses faces two main limitations. Firstly, downscaling might be required to (i) bridge the scale gap between global reanalysis and catchment hydrology and also to (ii) correct long-term bias potentially present in the reanalysis, e.g., long-term biases in air temperature and precipitation over France were highlighted by Caillouet et al. (2016) and Bonnet et al. (2017). Secondly, the use of hydrological models over long and over hydrologically contrasting periods is limited and associated with uncertainty (e.g., Brigode et al., 2013; Trotter et al., 2023). Thus, the models need to be calibrated over the past decades (e.g., $\sim$1980–2020) using a reference hydrometeorological dataset before being used over multi-decadal periods (e.g., Brigode et al., 2016). Despite these two main limitations, such a modeling approach offers the opportunity to understand, at large spatial and temporal scales, the documented changes in the past in terms of rainfall–runoff relationships and processes. Moreover, hydrological models are useful to illustrate how catchments can play a role in amplifying or weakening air temperature and precipitation signals (e.g., Müller et al., 2021; Baulon et al., 2022; David and Frasson, 2023).

The general objective of this paper is to evaluate the ability of such a modeling methodology to identify trends and/or periodicities of catchment hydrology at the European scale despite the coarse spatial resolution of the global reanalyses and the rainfall–runoff modeling uncertainty. To this end, we used two global reanalyses (NOAA 20CR and ERA-20C; Poli et al., 2016) as inputs of a conceptual rainfall–runoff model (GR4J) over 2128 European catchments to simulate daily streamflows since the 1840s. More specifically, we aimed to answer these three questions:

- How well do these two global reanalyses perform in providing climate forcings that enable hydrological models to reproduce observed streamflow, both at daily and decadal timescales?
- Does the performance depend on the spatial scale (catchment size) and the hydrological processes studied (catchment regimes)?

– Is the low-frequency variability simulated using this methodology in agreement with observations and other simulation results?

2 Data

2.1 Climate forcings

Several meteorological databases were assembled for this study. The first objective was to have a reference meteorological forcing over the recent period (typically the past four decades) that would be homogeneous for all catchments. This meteorological forcing enables a "classic" calibration of rainfall–runoff models at a spatial resolution that fits the catchment area. A common forcing set for all catchments studied ruled out the approach of considering the forcings provided in certain CAMELS-type databases. Therefore, two meteorological forcings were extracted over each catchment:

– Catchment daily precipitations were estimated using the MSWEP (V2) dataset (Beck et al., 2018), providing daily precipitation over the period 1979–2019 and a 0.1° ($\sim 91\,km^2$) grid.

– Catchment mean daily air temperatures were estimated using the ERA5 reanalysis (Hersbach et al., 2020), providing hourly variables over the period 1980–2019 and a 0.25° ($\sim 580\,km^2$) grid.

The combination of MSWEP precipitation time series and ERA5 air temperature time series is denoted as "MSWEP and ERA5" hereafter.

It is worth noting that we also considered an alternative reference dataset, namely ERA5-Land precipitation and ERA5 air temperature. However, since the calibration performance obtained with this forcing was lower than that achieved with the MSWEP and ERA5 dataset (results not shown in this paper), we did not retain this forcing for the remainder of the study.

Then, daily precipitation and mean daily air temperature of two long-term historical forcings were extracted over each catchment studied:

– The ERA-20C reanalysis (Poli et al., 2016), available over the period 1900–2010 with a spatial resolution of 1.4° ($\sim 17\,000\,km^2$).

– The NOAA 20CR (v3) reanalysis (Slivinski et al., 2019), available over the period 1836–2015 with a spatial resolution of 1.0° ($\sim 10\,000\,km^2$).

While the two reanalyses differ in terms of their underlying models and data assimilation techniques, they also diverge in the types of assimilated observations and prescribed forcings. The NOAA 20CR reanalysis is based solely on surface pressure observations from the International Surface Pressure Databank (ISPD; Cram et al., 2015), assimilated into NOAA's Global Forecast System, with sea surface temperature (SST) and sea ice concentration (SIC) prescribed as boundary conditions. In contrast, the ERA-20C reanalysis assimilates surface pressure data from both ISPD and the International Comprehensive Ocean–Atmosphere Data Set (ICOADS; Woodruff et al., 2011), as well as marine wind observations from ICOADS. Additionally, ERA-20C prescribes not only SST and SIC, but also solar radiation, tropospheric and stratospheric aerosols, ozone, and greenhouse gas concentrations.

The common period of the four meteorological forcings is the period 1979–2010 (32 years).

2.2 Catchment set

2.2.1 Data source and catchment sample selection

An initial sample of 4396 European catchments was assembled from a collection of several "CAMELS-like" datasets (cf. Table 1). This sample groups catchments with daily streamflow series from national datasets (CAMELS-CH (Höge et al., 2023) in Switzerland, CAMELS-FR (Delaigue et al., 2025) in France, CAMELS-GB (Coxon et al., 2020) in the United Kingdom, the NVE (The Norwegian Water Resources and Energy Directorate) dataset (using the NVE Hydrological API (HydAPI), https://hydapi.nve.no/, last access: 25 June 2023) in Norway, the SMHI dataset (https://www.smhi.se/, last access: 12 May 2023) in Sweden, and SAIH-RODEA (Yeste et al., 2024) in Spain), transnational datasets (LamaH-CE (Klingler et al., 2021) for Austria, Germany, the Czech Republic, Switzerland, Slovakia, Italy, Liechtenstein, Slovenia and Hungary), or global datasets (GRDC, 2023). Note that 77 duplicate stations were removed from the CAMELS-FR, CAMELS-CH, GRDC, and LamaH-CE datasets.

Four criteria were applied to select a sub-sample of European catchments. We retained only catchments that met all of the following conditions:

– At least 10 years of daily streamflow data available during the calibration period (1996–2010);

– At least 10 years of daily streamflow data available during the evaluation period (1982–1995);

– A catchment area larger than $100\,km^2$, a subjective threshold applied to ensure compatibility with the daily time step and the spatial resolution of climate forcings;

– An equivalent water storage capacity from upstream dams below 10 mm, calculated using the GRanD dataset (Lehner et al., 2011) as the ratio between total reservoir storage and catchment area (see Delaigue et al., 2025), to limit the influence of human regulation.

After applying these criteria, 2128 catchments were selected.

Table 1. List of catchment datasets used.

Country or region	Initial number of catchments	Number of catchments	Reference	Period	Catchment boundaries?
Switzerland	211	108	CAMELS-CH (Höge et al., 2023)	1981–2020	Y
France	2201	783	CAMELS-FR (Delaigue et al.,2025)	1900–2021	Y
United Kingdom	661	394	CAMELS-GB (Coxon et al., 2020)	1970–2015	Y
Central Europe	799	466	LamaH-CE (Klingler et al., 2021)	1981–2017	Y
Norway	75	29	NVE, https://hydapi.nve.no/ (last access: 25 June 2023)	1912–2019	N
Sweden	256	192	SMHI, https://www.smhi.se/ (last access: 12 May 2023)	1900–2019	N
Europe	98	91	GRDC (GRDC, 2023)	1826–2021	N
Spain	95	65	SAIH-ROEA (Yeste et al., 2024)	1912–2020	N
Total	4396	2128			

2.2.2 Catchment delineations and hypsometric data

The catchment delineations were extracted from the "CAMELS-like" dataset when available (CAMELS-CH, CAMELS-FR, CAMELS-GB, and LamaH-CE) or were estimated using TauDEM routines (Tarboton, 2013) by positioning manually the hydrometric stations on the theoretical river network estimated using the EU-DEM (v1.1) dataset. This DEM is available at a spatial resolution of 25 m. A comparison between the reference catchment area (i.e., given by the data producer) and the DEM-derived catchment area was performed to ensure catchment area coherence (not shown here). Hypsometric data were also calculated for each catchment using the EU-DEM dataset.

2.2.3 Catchment characteristics

Catchments were grouped by their (i) hydrological regimes and (ii) regions. The catchment regimes were derived using the classification proposed by Hashemi et al. (2022) considering interannual monthly catchment air temperature (from ERA5 dataset; see Sect. 2.1), precipitation (from MSWEP dataset; see Sect. 2.1), and streamflow over the period 2001–2015. The catchment set is thus composed of (see Fig. 1b) 467 nival catchments, 269 nival-pluvial catchments, 902 pluvial catchments, 130 Mediterranean catchments and 108 uniform catchments.

Catchments are also assigned to one of the eight regions inspired by the eight European regions used by Christensen and Christensen (2007). The catchment set is thus composed of (see Fig. 1c): 442 catchments in the Alps, 395 catchments in the British Isles, 139 catchments in eastern Europe, 470 catchments in western France, 65 catchments in the Iberian Peninsula, 68 catchments in the Mediterranean region, 328 catchments in central Europe and 221 catchments in Scandinavia.

Depending on the data source, streamflow time series span over different periods. The streamflow time series starts in 1900 for 25 catchments (1 % of the total catchment set), in 1950 for around 200 catchments (10 % of the total catchment set), and in 1970 for half of the catchment set (Fig. 1d).

3 Methods

3.1 Rainfall–runoff model

The GR4J (Perrin et al., 2003) conceptual rainfall–runoff model and its snow accumulation and melting module CemaNeige (Valéry et al., 2014) were used for streamflow simulations. The inputs of this model are:

- Daily precipitation (in $\mathrm{mm\,j^{-1}}$)
- Daily potential evapotranspiration (in $\mathrm{mm\,j^{-1}}$), estimated using the Oudin et al. (2005) formulation, considering daily air temperature and mean catchment latitude.
- Daily air temperature (°C).

The CemaNeige module takes into account the hypsometric curve of each catchment to perform a downscaling of meteorological forcings by distributing daily precipitation and air temperatures across five (in our case) elevation bands. Thus, the forcings input into the model are considered representative of the median elevation of the catchment and are then distributed across the five zones according to the gradients described by Valéry et al. (2014).

3.2 Model parameter calibration

The GR4J and CemaNeige models have, respectively, four and two free parameters that need to be calibrated conjointly for each catchment, considering the observed daily streamflow available over a given time period. The parameter calibration was performed individually for each catchment over a common period comprising 10–15 years over the period 1996–2010, with a warm-up period of 3 years from 1993 to 1995. The objective function used for the model calibration

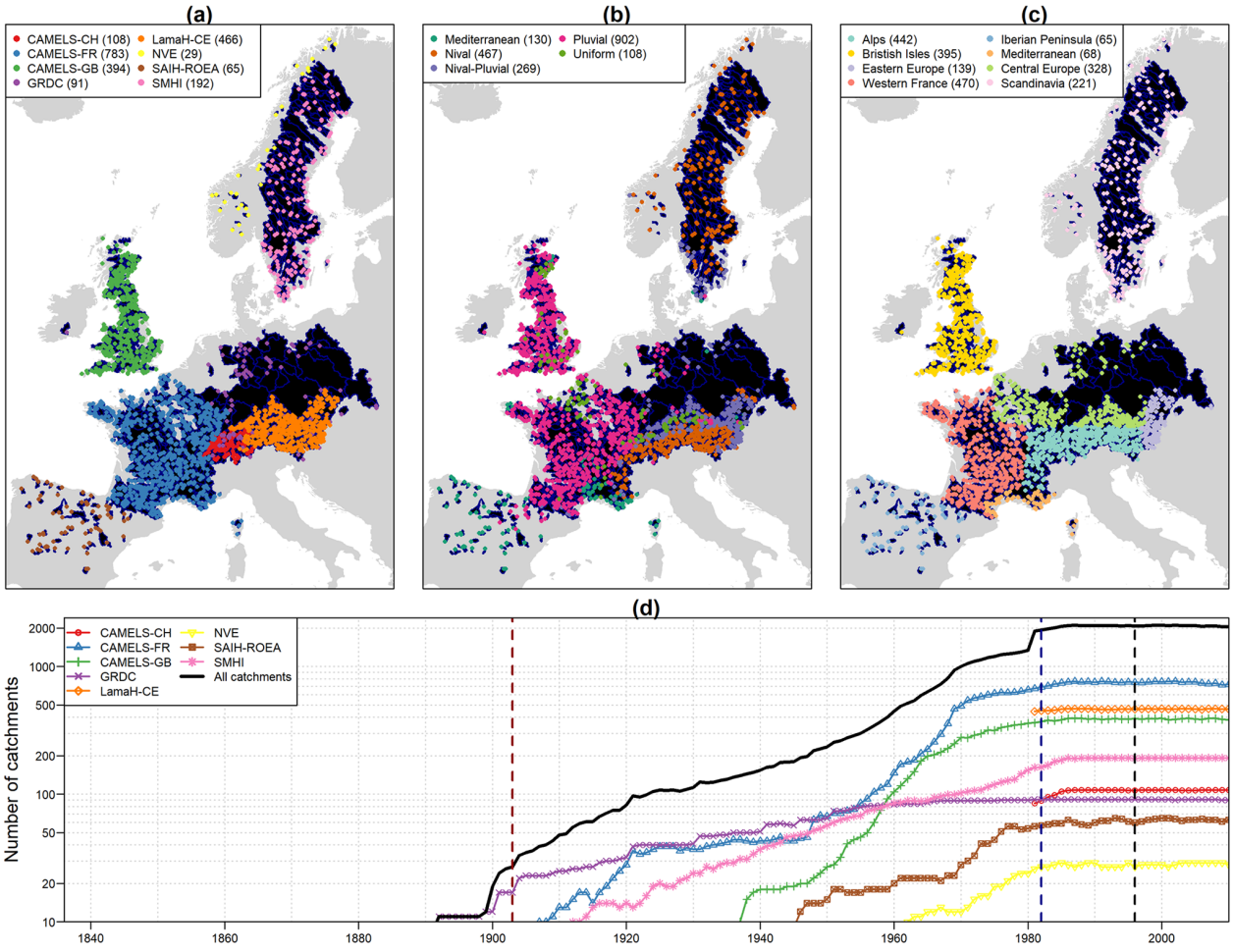

Figure 1. **(a)** Source, **(b)** hydrological regimes, **(c)** regions, and **(d)** temporal availability of daily streamflow data of the study catchments.

is the Kling and Gupta efficiency criterion (KGE; Gupta et al., 2009), ranging from $-\infty$ to 1 and estimated as follows:

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

where:

- β is the ratio between the means of the simulated and observed time series; this quantifies the simulation bias, and ranges between 0 and $+\infty$ (values > 1 indicate a model overestimation).
- α is the ratio between the standard deviations of the simulated and observed time series; this quantifies the ability of the simulation to reproduce the streamflow variability, and ranges between 0 and $+\infty$ (values > 1 indicate a model overdispersion).
- r is the coefficient of correlation between the simulated and the observed series; this quantifies the ability of the simulation to reproduce the observed temporal variations, and ranges between -1 and 1 (perfect correlation).

We used the 2009 version of the KGE in order to allow comparison of model performance and parameters with other similar studies conducted on sub-samples of European catchments.

Three different meteorological forcings were used for the model parameter calibration, thus resulting in three different parameter sets for each catchment:

- Parameter sets calibrated considering precipitation and air temperature from MSWEP and ERA5,
- Parameter sets calibrated considering precipitation and air temperature from ERA-20C,
- Parameter sets calibrated considering precipitation and air temperature from NOAA 20CR.

The model was implemented using the R (R Core Team, 2024) package airGR (Coron et al., 2017, 2023), using the default optimization algorithm included in the airGR package. This algorithm was specifically designed for GR models (Coron et al., 2017).

3.3 Model evaluation

3.3.1 Periods of daily streamflow evaluation

The daily streamflow simulations were evaluated over two different time periods (Fig. 2):

1. To compare the three meteorological forcings (MSWEP and ERA5, NOAA 20CR, and ERA-20C), the period 1982–1995 (14 years) was considered, with 3 years for model warm-up (1979–1981).

2. A longer period was considered to compare the two historical meteorological forcings (NOAA 20CR and ERA-20C) by using all available years from 1903 to 1995, with 3 years for model warm-up (1900–1902).

The evaluation metrics are the KGE values and its three components (β, α, and r; see Eq. 1).

To evaluate whether differences in model performance between the simulations were statistically significant, the Wilcoxon rank-sum test (Wilcoxon, 1945) was applied. This non-parametric test is used to compare two independent distributions and does not require the assumption of normality. It is therefore well suited for assessing differences in performance metrics, such as the KGE values, across a large set of catchments.

3.3.2 Annual time series at the catchment scale

For each catchment, the daily times series (observations, NOAA 20CR simulations, and ERA-20C simulations) are aggregated at the annual time step for the specific evaluation of:

- Mean flow: calculation for each year of the mean annual streamflow, named "Q_A" hereafter

- High flow: calculation for each year of the maximum streamflow value, named "Q_X" hereafter

- Low flow: calculation for each year of the minimum mean monthly streamflow value, named "Q_M" hereafter.

To assess the ability of the model to reproduce the long-term temporal variability of the streamflow time series, temporal correlations between simulations and observations were estimated for each variable (Q_A, Q_X, and Q_M) for each catchment individually (i) on the annual time series and (ii) on 10-year running mean time series. In each case, correlations are estimated only if the observed series length exceeds 30 years.

3.3.3 Annual time series at the regional scale

Finally, regional anomalies of mean, high, and low flows were calculated. For each region and each year, catchments with data were identified: If more than 10 catchments are available for a given year and region, an anomaly is calculated by dividing the annual flow values by the average of the annual flow values for the catchment subset studied. Thus, the catchment subset may change every year. Finally, the 10-year running mean is calculated for mean-, high-, and low-flow indices.

4 Results

4.1 Model calibration performance

Figure 3 presents the rainfall–runoff model performance over the calibration period for the three different climate forcings (MSWEP and ERA5, NOAA 20CR, and ERA-20C). The calibration performance obtained using the MSWEP and ERA5 meteorological forcings is relatively good, with a slightly worse performance for catchments in eastern Europe, the Iberian Peninsula, and Mediterranean regions. The performance obtained using NOAA 20CR and ERA-20C forcings is somewhat lower, being average to poor. The rainfall–runoff model achieves better performance with NOAA 20CR precipitation and temperature compared with ERA-20C. Nevertheless, the performance of the model in each region is similar depending on the meteorological dataset used, with the worst performance encountered for catchments in eastern Europe and Mediterranean regions.

4.2 Evaluation performance

4.2.1 Daily streamflow simulations

Figure 4 presents the rainfall–runoff model performance over the calibration period and first evaluation period, grouped according to the parameter sets used (calibrated using MSWEP and ERA5, NOAA 20CR, or ERA-20C forcings over the calibration period) and according to the climate forcing used (NOAA 20CR or ERA-20C). The performance obtained using parameters calibrated with MSWEP and ERA5 forcings is poor when using both NOAA 20CR and ERA-20C forcings. Logically, when using NOAA 20CR (ERA-20C) forcings, the performance obtained using parameters calibrated with NOAA 20CR (ERA-20C) forcings is better than using parameter sets obtained with ERA-20C (NOAA 20CR) forcings. Finally, the performance obtained over the two evaluation periods is higher when using NOAA 20CR forcings than ERA-20C forcings (p value of Wilcoxon rank test (1945) equal to 0 when comparing performance using NOAA 20CR parameter sets and climate forcings with performance using ERA-20C parameter sets and climate forcing).

Figure 5 presents the KGE and KGE components calculated over the two evaluation periods, considering each forcing for parameter calibration and as meteorological forcings. For the second evaluation period (1903–1995), only the 187 catchments with more than 50 years of observations are considered. This figure shows that NOAA 20CR simulations are

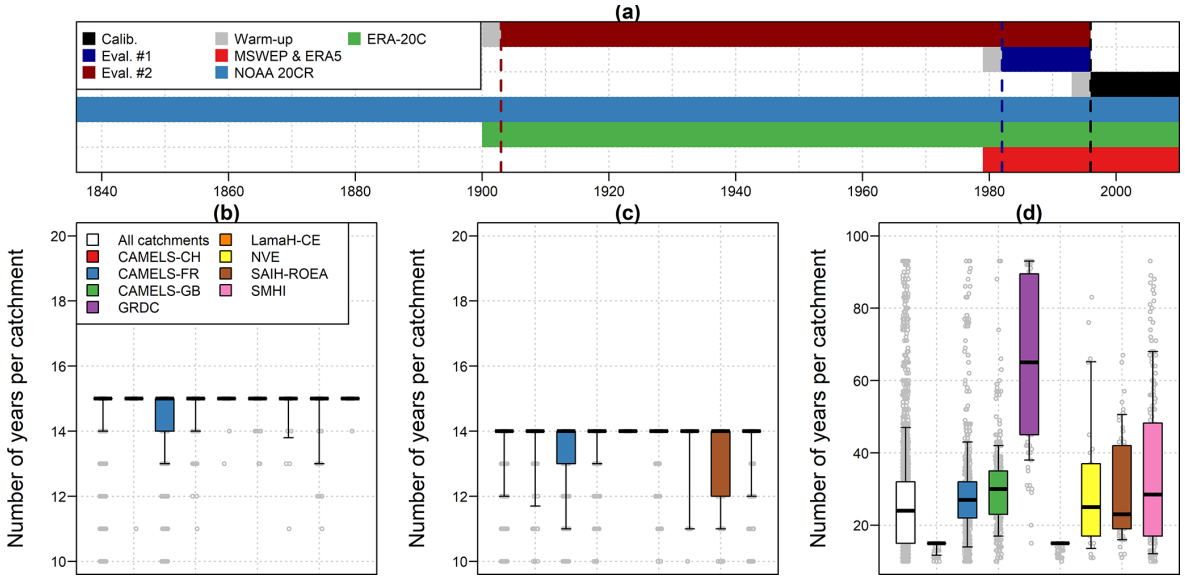

Figure 2. **(a)** Temporal availability of meteorological forcings and number of streamflow observations per catchment over **(b)** the calibration period (1996–2010), **(c)** evaluation period 1 (1982–1995), and **(d)** evaluation period 2 (1903–1995).

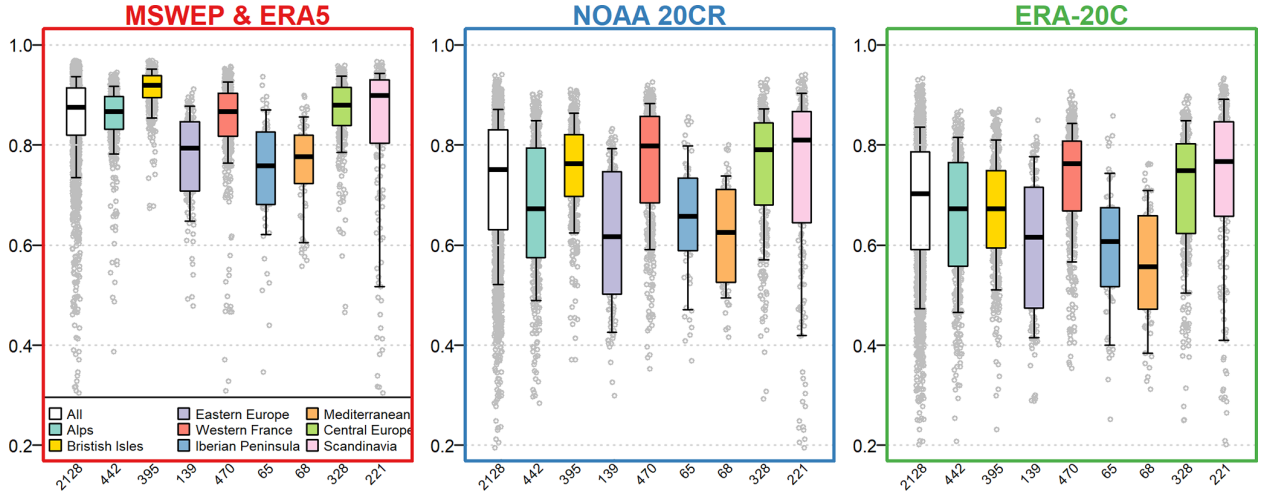

Figure 3. Calibration performance (KGE) for three different climate forcings (MSWEP and ERA5, NOAA 20CR, and ERA-20C, from left to right) according to region. Boxplots are constructed with the 0.10, 0.25, 0.50, 0.75, and 0.90 quantiles.

more closely correlated with observations (r) compared with ERA-20C simulations: (p value of Wilcoxon rank test (1945) equal to 0 when comparing r values obtained using NOAA 20CR parameter sets and climate forcings with r values obtained using ERA-20C parameter sets and climate forcings). Mean bias (beta) and deviation bias (alpha) reveal an overall slight underestimation of streamflow values and variance by ERA-20C simulations.

4.2.2 Simulations aggregated over time

Figure 6 shows the temporal correlations evaluated individually for each catchment over the 1903–1995 evaluation period for the daily simulations aggregated at the annual time step. The overall performance is better for mean flows than for low flows and high flows. The difference between the two simulation types (NOAA 20CR or ERA-20C) is not clear, with NOAA 20CR simulations being marginally better and in relative agreement when analyzed over the geographical regions. The performance is, for mean, low, and high flows, dependent on the region: For mean flows, the performance obtained over regions in western France is clearly the best, while the performance obtained over the Alps, eastern Europe, and the Iberian Peninsula is the worst. For low and high flows, the performance over the Alps, eastern Europe, and the Iberian Peninsula is also the worst for NOAA 20CR simula-

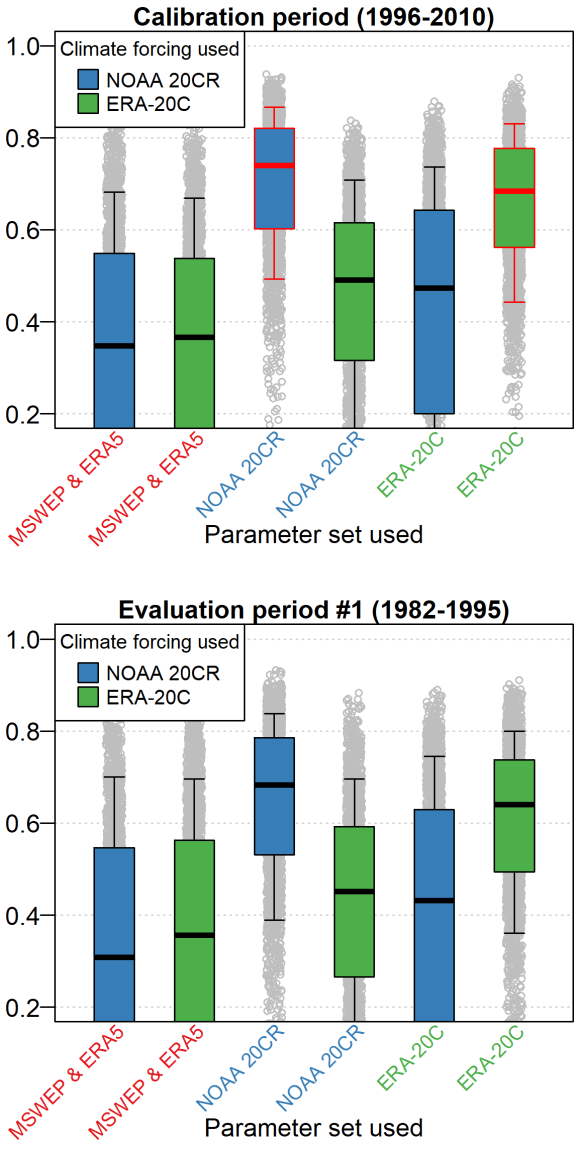

Figure 4. Evaluation performance (KGE) over the calibration period (top) and evaluation period 1 (bottom), grouped by the parameter sets considered and the climate forcing used. Boxplots are constructed with the 0.10, 0.25, 0.50, 0.75, and 0.90 quantiles. The boxplots framed in red summarize the performance obtained in calibration.

tions. For ERA-20C, the low-flow simulations are the worst for the Iberian Peninsula, while the high-flow simulations are the worst for the Alps, eastern Europe, Iberian Peninsula, and Mediterranean regions. For both NOAA 20CR and ERA-20C, the best performance for low and high flow is found in western France, central Europe, and Scandinavian regions. These general conclusions are also valid when looking at the flow annual time series smoothed with a 10-year time step (see Appendix A, Fig. A1).

4.2.3 Simulations aggregated over time and space

Figure 7 compares the annual anomalies of mean flow (Q_A) observations and simulations for each region studied. The two climate forcings reproduce relatively well the interannual variability in regional mean flows, with a good correlation between observed and simulated successions of wetter/drier sub-periods. No clear trend emerges in terms of dependence on the region or time period considered. Unfortunately, this analysis is limited for regions with few data available (e.g., eastern Europe and the Mediterranean region). Similar conclusions are reached when analyzing low and high flows (cf. Figs. 8 and 9), albeit with a higher amplitude of anomalies in observations and simulations.

5 Discussion

5.1 Method and data used for the multi-decadal hydrological simulation

The primary aim of this study was to evaluate the suitability of two global reanalyses as inputs for reconstructing catchment-scale hydrology through conceptual rainfall–runoff modeling. To this end, the methodological framework was deliberately kept simple and consistent across catchments, focusing on the effects of the input forcings rather than the modeling choices themselves. Consequently, aspects such as the choice of a single hydrological model, the use of a fixed objective function, or the daily temporal resolution were not explored in depth. Although a full quantification of the uncertainties associated with these methodological decisions lies beyond the scope of this study, it is nevertheless important to acknowledge and briefly discuss these limitations, as they can influence the interpretation of the results.

In comparing the performance obtained depending on the forcings and parameter sets used, we aimed to show that it was more appropriate to use forcings with the finest spatial resolution (in this case, MSWEP and ERA5 forcings) to calibrate the model parameters. However, this hypothesis proved to be false: When NOAA 20CR forcings are used as input to the model, the parameter sets obtained after calibration with MSWEP and ERA5 perform worse than those obtained after calibration with ERA-20C, which in turn are less effective than those obtained after calibration with NOAA 20CR. Thus, consistency between the forcings used in calibration and simulation appears to be more important than the spatial resolution of the forcings used during calibration. This result can be explained by the flexibility of the rainfall–runoff model during parameter calibration, allowing for an implicit adaptation of the model parameters to the spatial resolution of the forcings. Hydrological models have the ability to compensate for errors in forcing via parameter calibration, whether these errors are systematic (bias) or random (see e.g. Dawdy and Bergmann, 1969; Oudin et al., 2006). In

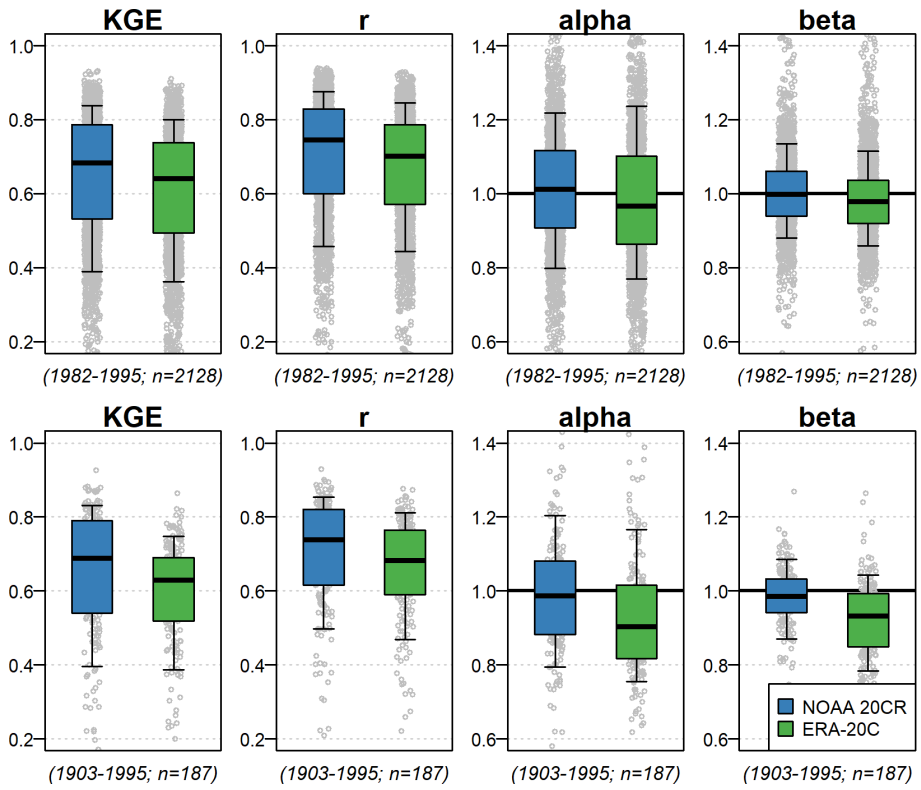

Figure 5. Evaluation performance (KGE and KGE components) over the two evaluation periods (1982–1995 and 1903–1995, from top to bottom), with catchments grouped by the climate forcing used for parameter calibration and simulation. For the second evaluation period (1903–1995), only the 187 catchments with more than 50 years of observations are considered. Boxplots are constructed with the 0.10, 0.25, 0.50, 0.75, and 0.90 quantiles.

our study, the calibration period was temporally restricted to be common across the different forcings. Nevertheless, this result advocates extending the calibration period to cover the entire period of available discharge data for each catchment.

The simulation methodology used in this study has several important limitations to be noted. First, it is based on a single conceptual rainfall–runoff model. Using a multi-model approach would make it possible to quantify the uncertainty associated with the model structure in the simulations performed (Wan et al., 2021; Martel et al., 2023; Thébault et al., 2024). The simulation method was applied at a daily time step, which may be too coarse for flood modeling in some small and/or Mediterranean catchments. The Oudin et al. (2005) formula used to estimate potential evapotranspiration series at the catchment scale is also a significant source of uncertainty in the context of streamflow simulation (Lemaitre-Basset et al., 2022a). The use of different formulations, for example, taking into account CO_2 (Lemaitre-Basset et al., 2022b) is an interesting perspective for this work.

The objective function used in this study is also an important methodological choice, and could have been adapted to the flows studied (e.g., by using an objective function better suited to reproducing low flows; Pushpalatha et al., 2012) and also adapted to the modeling exercise over a long period (e.g., Split KGE proposed by Fowler et al. (2018) for the simulation of drying climate in Australia).

Furthermore, the rainfall–runoff model parameters were estimated over a short and recent period relative to the entire period considered for streamflow reconstruction. Many studies have highlighted the significant difficulty hydrological models face in simulating periods with different climatic conditions than those considered for parameter calibration (e.g., by Fowler et al., 2020; Duethmann et al., 2020). This uncertainty could be quantified by considering, for each catchment, several parameter sets per hydrological model (e.g., after calibrations over different sub-periods through bootstrapping of the observations, Brigode et al., 2015; Arsenault et al., 2018).

Finally, the selection of catchments for this analysis raises some important points. Certain regions are over-represented while others are under-represented, and the uncertainty associated with the hydrometric data has not been addressed. Furthermore, the identification of catchments with minimal anthropogenic influence relied on a limited set of indicators – such as major water withdrawals, dams, or wastewater treatment plants – and did not account for the uncertainty associated with these data. Other long-term changes likely to affect

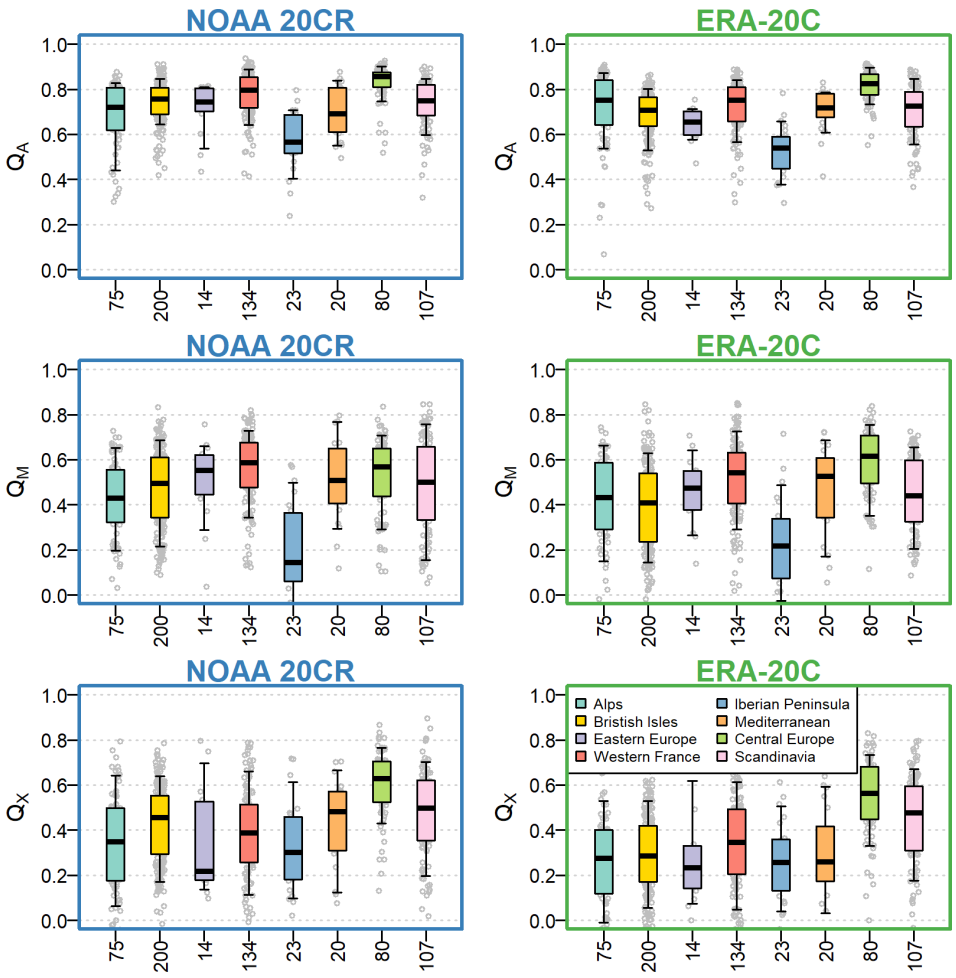

Figure 6. Temporal correlation (r) over the 1903–1995 evaluation period, with catchments grouped by the climate forcing used (NOAA 20CR or ERA-20C) and the catchment region. Boxplots are constructed with the 0.10, 0.25, 0.50, 0.75, and 0.90 quantiles. The x axis shows the number of catchments available for each region.

hydrological response, such as land use changes related to forestry, agriculture, or urbanization over the past century, were not considered in this context. Additionally, the availability of long streamflow time series (i.e., spanning several decades) is scarce, highlighting the need to explore the possibility of conducting this analysis with a dataset constructed at a monthly time scale.

5.2 Variability in performance within the catchment set

The analysis of the performance obtained during the calibration of the rainfall–runoff model showed poorer results for Mediterranean catchments. This poor performance can be explained by various factors, such as the spatial resolutions of the meteorological forcings used (greater or equal to 10 000 km^2), which may be too coarse relative to the size of the catchment to capture the relevant processes. Other factors include significant anthropogenic influences that may vary over time in the streamflow observations used or the model's difficulty in simulating the hydrological regime intermittency. The performance on eastern European basins is also weaker than for other regions; thus, further investigations are needed to explain this trend.

As expected, the simulation performance is better for annual mean flows than for minimum monthly flows and annual maximum daily flows, for both forcings tested. This result can probably be explained by (i) the method used for the hydrological simulation (discussed in Sect. 5.1) and more specifically the choice of the objective function used to calibrate the model (which can be adapted to a particular objective) and (ii) the quality of the climatic forcings used, characterized by a spatiotemporal resolution that is probably too coarse to represent the variability in the hydrological processes that generate floods.

Finally, no clear relationship is found between daily performance and the ability of the model to reproduce interannual variability in high, mean, and low-flows, suggesting

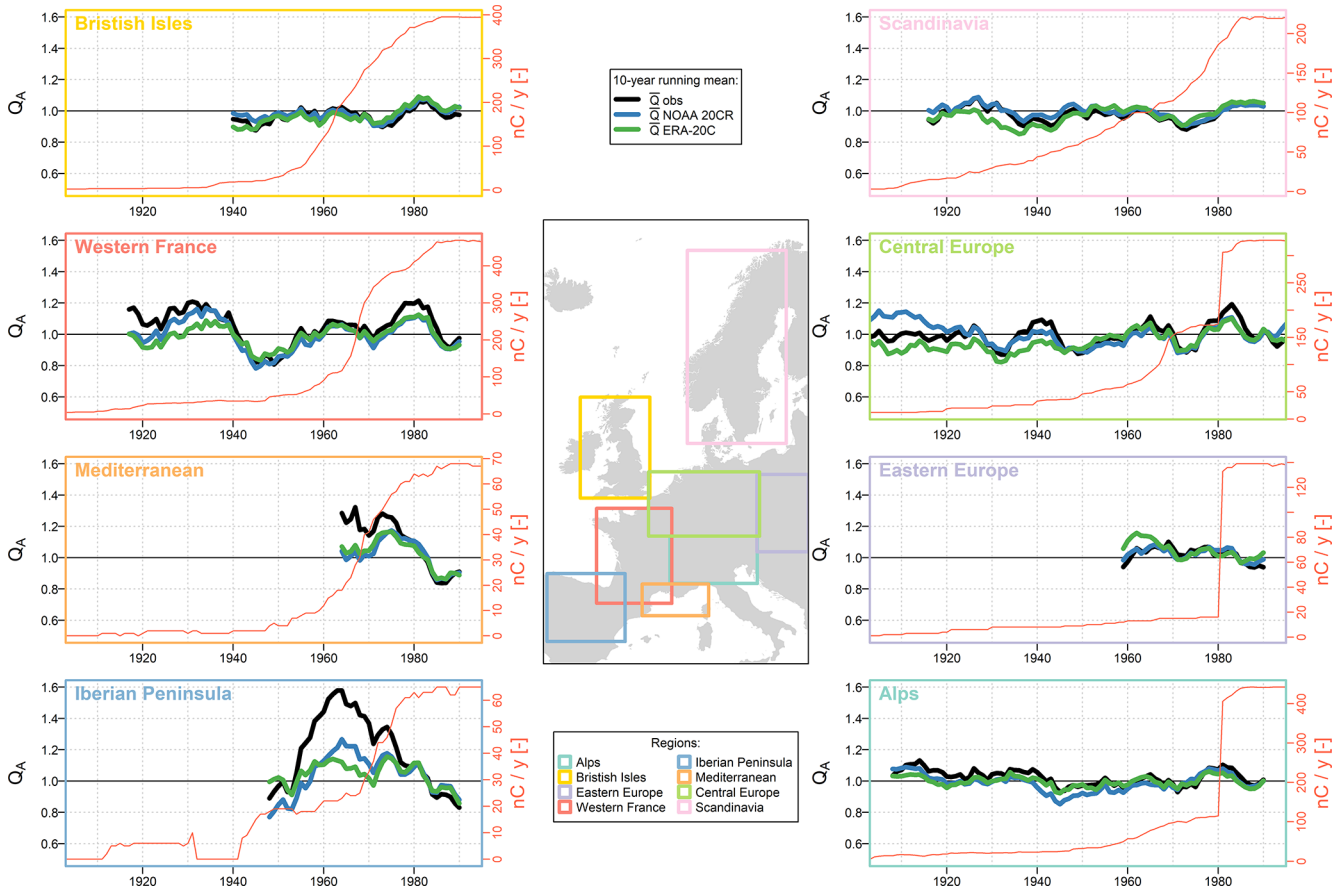

Figure 7. 10-year running means of the Q_A observations (black) and simulations (blue: NOAA 20CR simulations, green: ERA-20C simulations) for each region studied. Only years with at least 10 available catchments per region are considered. The red y axis shows the number of available catchments per year and region.

an independence between these aspects (see Appendix B, Fig. B1).

5.3 Consistency of multi-decadal hydrological variability

The interannual variability simulated using the two climate forcings (NOAA 20CR and ERA-20C) shows reasonable consistency with available observed data at the regional scale. When analyzing 10-year running means, the temporal correlations reach approximately 0.8 for mean flows and 0.6 for both low and high flows (Appendix C, Fig. C1). On average, the NOAA 20CR reanalysis performs slightly better than the ERA-20C reanalysis across the three flows studied. However, when zooming in at the regional scale, this ranking is reversed for the central European and Mediterranean regions for mean and low flows, as well as for the eastern European region in terms of high flows. The simulation performance for high flows is notably poor for both reanalyses in the Alps and eastern European regions. It is important to note that the analysis of consistency between simulations and observations was conducted over large, somewhat arbitrary regions, which may group together catchments with different hydrological regimes. A similar analysis could be conducted at a finer spatial scale using more precisely "hydrologically defined" regions. Moreover, explaining the differences in trends obtained with the two climate forcings requires further investigation, in order to attribute these discrepancies to differences in assimilated data or, for example, to differences in the underlying climatic models. In addition, a complementary analysis was performed by retaining only catchments with calibration and validation KGE values above 0.7. The results were very similar to those presented in Figs. 6 to 8, confirming that the conclusions are not significantly affected by poorly modelled catchments.

Reviews of hydrological trends in Europe since the 1950s suggest a general tendency for northern European catchments to become wetter, while southern European catchments tend to dry out (e.g., Masseroni et al., 2021). This drying trend is evident in both the observations and the simulations for the Mediterranean and Iberian Peninsula catchments, affecting mean, low, and high flows. Although the past few decades appear to have been relatively wetter for

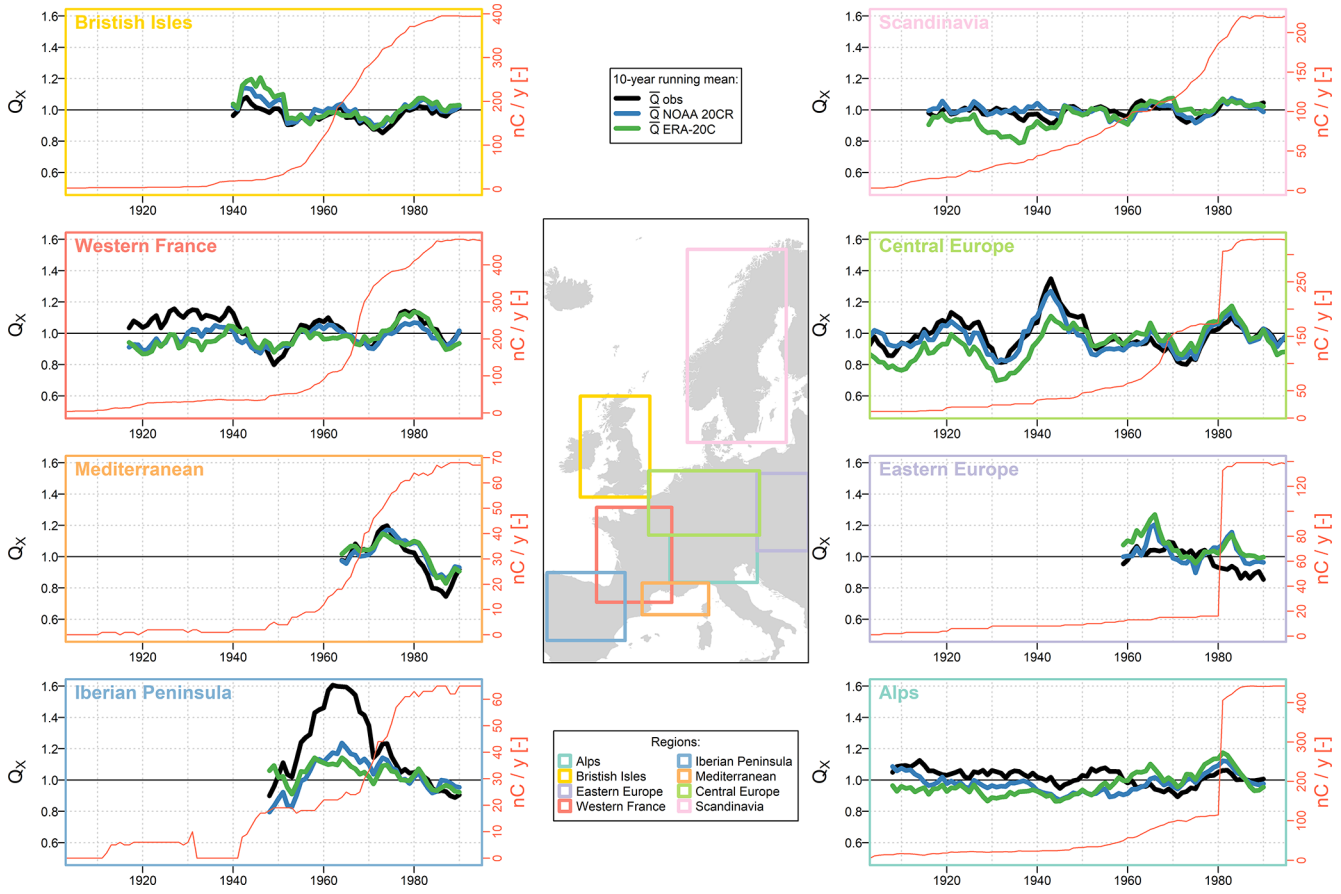

Figure 8. 10-year running means of the Q_M observations (black) and simulations (blue: NOAA 20CR simulations, green: ERA-20C simulations) for each region studied. Only years with at least 10 available catchments per region are considered. The red *y* axis shows the number of available catchments per year and region.

the Scandinavian catchment studied, no clear linear trends are observed for these basins.

In general, the analysis reveals alternating dry and wet periods across all regions and flow indicators studied, rather than consistent linear trends. It is noteworthy that within the same region, individual analyses of interannual variations for low, mean, and high flows can reveal different anomalies. For instance, in the British Isles, the years 1980 to 1985 appear relatively wet across low, mean, and high flows, while the period from 1940 to 1950 is only characterized as wet in terms of high flows. Similarly, for the central European catchments, there was a dry period for high flows between 1930 and 1935, followed by a wet period between 1940 and 1945. This anomaly is not reflected in the analysis of low flows, highlighting differing interannual variations between flow types. Most of these trends have already been identified in the literature. For instance, the wet periods (1920s, 1980–1990) and the dry decade (1970s) observed by Lindström and Bergström (2004) in Sweden, the multi-decadal variations simulated with a similar method by Devers et al. (2024) in France, or the "flood-poor" period identified after World War II by Brönnimann et al. (2022) at the European scale.

6 Conclusions

This study assessed the capacity of two global climate reanalyses – NOAA 20CR and ERA-20C – to drive a conceptual rainfall–runoff model (GR4J) for simulating catchment-scale streamflows across Europe since the 1840s. Despite the coarse spatial resolution of these datasets, the results show that both reanalyses can reproduce daily and multi-decadal streamflow variability reasonably well, particularly for mean flows.

An important result to be highlighted is the necessity of ensuring consistency between the climate forcings used for calibration and those used for simulation: the best performance with the ERA-20C forcing was obtained when using parameters calibrated with the same forcing, despite its relatively coarse spatial resolution. This suggests that consistency in meteorological inputs may play a more critical role than spatial resolution when working with centennial reanalyses. Al-

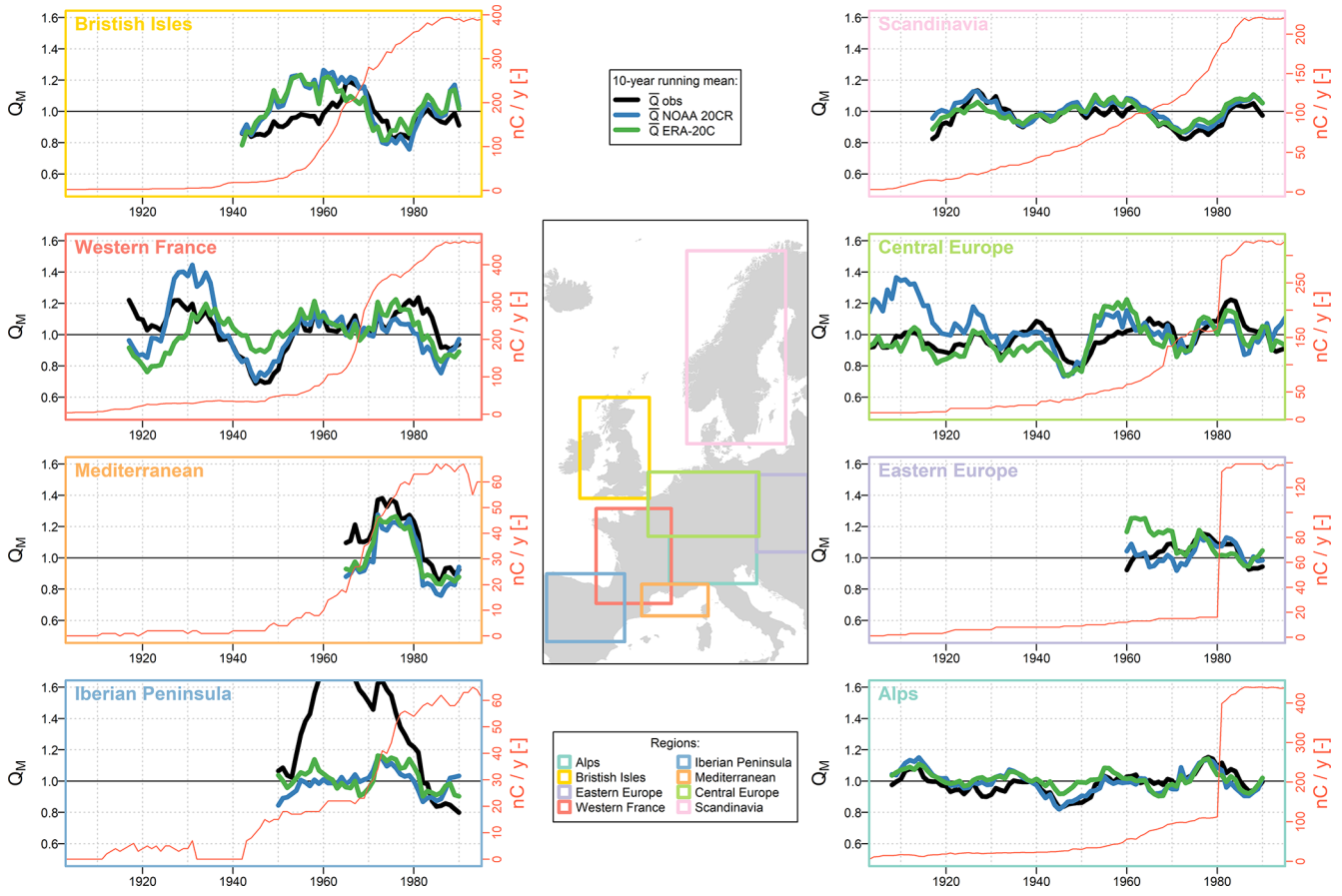

Figure 9. 10-year running means of the Q_X observations (black) and simulations (blue: NOAA 20CR simulations, green: ERA-20C simulations) for each region studied. Only years with at least 10 available catchments per region are considered. The red y axis shows the number of available catchments per year and region.

though higher-resolution datasets such as MSWEP would likely yield better performance if available over the full historical period, producing such datasets over the full 20th century requires downscaling methods that raise challenges in terms of temporal consistency and robustness. While downscaling remains a promising direction for improving local-scale performance, particular care must be taken to ensure the robustness of long-term hydrological trends.

In general, the performance increases with catchment size (except for Mediterranean and snow-dominated catchments), up to a threshold beyond which no further improvement is observed. Additionally, the performance varies across regions, reflecting the different hydrological processes in the catchments studied. The best performance is typically seen in catchments located in western France, Scandinavia, and the British Isles, while the lowest performance is observed in eastern Europe, the Mediterranean, and the Iberian Peninsula. Therefore, the ability of these two climate forcings and the modeling chain used to represent processes related to snow accumulation and melting, as well as the intermittency of the hydrological cycle, requires further investigation. Finally, the differences between the two reanalyses in certain regions are noteworthy. For example, the temporal correlation on 10-year running means for mean flows in eastern Europe is significantly higher for NOAA 20CR (~ 0.8) than for ERA-20C (~ 0.4), suggesting the potential for coupling the two reanalyses in specific regions.

The interannual variability simulated by NOAA 20CR and ERA-20C reanalyses shows good consistency with observed data. The analysis highlights alternating wet and dry periods across all regions, with different anomalies depending on flows, and suggests that the tested methodology holds promise for investigating mechanisms behind these variations to understand regional hydrological changes better. There are opportunities to refine the analysis of these consistencies by comparing it with databases focused on specific flow types, such as the flood dates compiled in the HANZE database (Paprotny et al., 2024).

Finally, this analysis highlights the significant multi-decadal variability in catchment streamflow, which may be underestimated when linear trends are sought in time series spanning only a few decades. In this context, the use of long-

term climatological and hydrological reanalyses is crucial, particularly for anticipating the effects of climate change on hydrosystems and for attributing changes at the catchment scale.

Appendix A

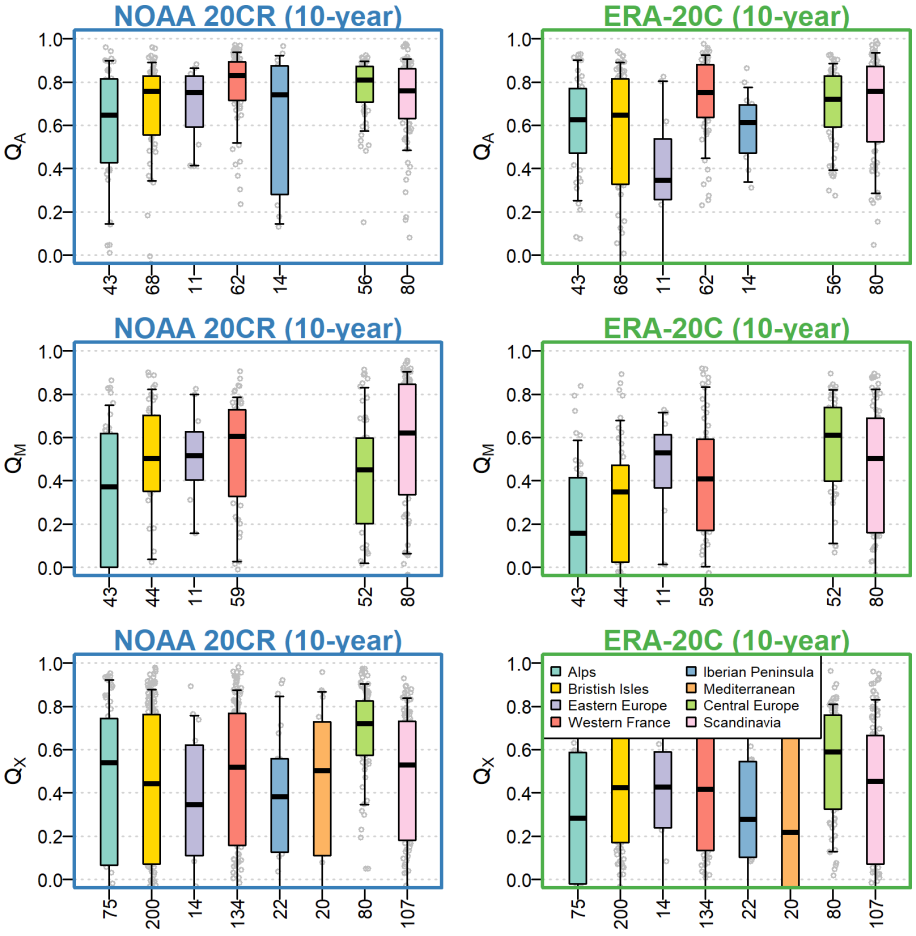

Figure A1. Temporal correlation (r) evaluated individually for each catchment over the 1903–1995 period for the daily simulations smoothed with a 10-year time step, with catchments grouped by the climate forcing used (NOAA 20CR or ERA-20C) and the catchment region. Boxplots are constructed with the 0.10, 0.25, 0.50, 0.75, and 0.90 quantiles. The x axis shows the number of catchments available for each region.

Appendix B

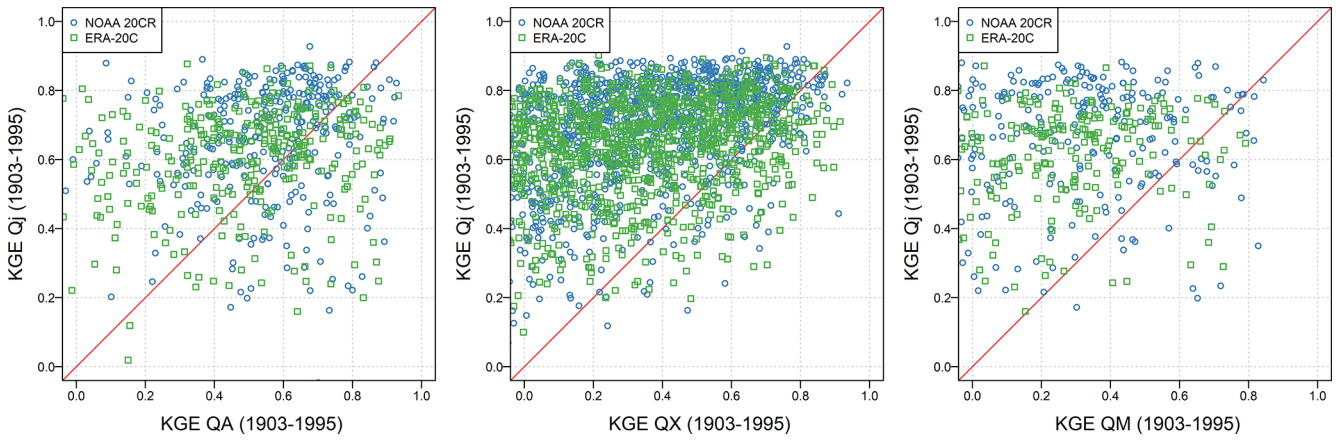

Figure B1. Comparison between simulation performance of mean- (left), high- (center), and low-flow (right) simulations (smoothed over 10 years) and the performance of simulations of daily streamflows.

Appendix C

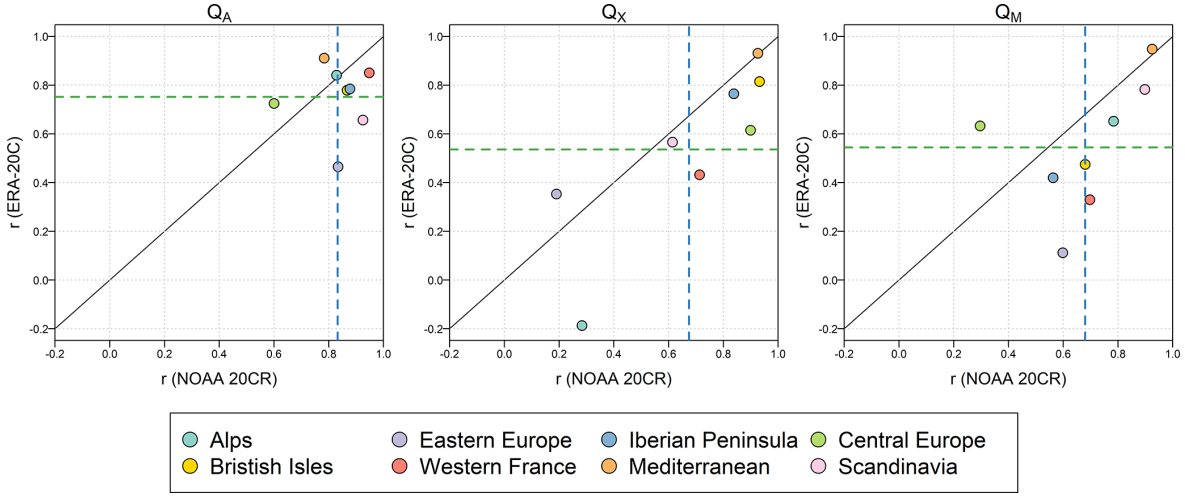

Figure C1. Temporal correlation (r) over the 1903–1995 evaluation period between 10-year running means of the Q_A, Q_X, and Q_M observations and simulations for each region studied. Only years with at least 10 available catchments per region are considered. Dashed blue (resp. green) line represents the average correlation coefficient over all regions obtained using NOAA 20CR (resp. ERA-20C).

Code availability. The R (R Core Team, 2024) package airGR (Coron et al., 2017, 2023) was used to perform all hydrological simulations. The R scripts required to use this package in this context can be provided upon request from the corresponding author.

Data availability. The climate forcings used in this article can be downloaded online (MSWEP: https://www.gloh2o.org/mswep/, last access: 1 February 2022; ERA-20C: https://reanalyses.org/, last access: 25 June 2023; NOAA-20CR: https://reanalyses.org/, last access: 25 June 2023). The streamflow data for the studied catchments were obtained from the open-access databases listed in Table 1 or downloaded online via the APIs described in Table 1.

Author contributions. PB conceptualized the work. PB wrote the computer codes to format the data, simulate streamflow from climate forcings and quantify model performance. PB and LO drafted the manuscript. All authors reviewed and edited the manuscript.

Competing interests. The contact author has declared that neither of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Also, please note that this paper has not received English language copy-editing. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. The authors thank the two reviewers and the editor, who provided constructive comments on an earlier version of the manuscript, which helped clarify the text.

Financial support. This work was supported by the French National program EC2CO (Ecosphère Continentale et Côtière).

Review statement. This paper was edited by Daniel Viviroli and reviewed by two anonymous referees.

References

Arsenault, R., Brissette, F., and Martel, J.-L.: The hazards of split-sample validation in hydrological model calibration, J. Hydrol., https://doi.org/10.1016/j.jhydrol.2018.09.027, 2018.

Baulon, L., Allier, D., Massei, N., Bessiere, H., Fournier, M., and Bault, V.: Influence of low-frequency variability on groundwater level trends, J. Hydrol., 606, 127436, https://doi.org/10.1016/j.jhydrol.2022.127436, 2022.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., McVicar, T. R., and Adler, R. F.: MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment, B. Am. Meteorol. Soc., https://doi.org/10.1175/BAMS-D-17-0138.1, 2018.

Blanc, A., Blanchet, J., and Creutin, J.-D.: Past evolution of western Europe large-scale circulation and link to precipitation trend in the northern French Alps, Weather Clim. Dynam., 3, 231–250, https://doi.org/10.5194/wcd-3-231-2022, 2022.

Blöschl, G., Bierkens, M. F. P., Chambel, A., et al.: Twenty-three Unsolved Problems in Hydrology (UPH) – a community perspective, Hydrolog. Sci. J., https://doi.org/10.1080/02626667.2019.1620507, 2019.

Blöschl, G., Kiss, A., Viglione, A., Barriendos, M., Böhm, O., Brázdil, R., Coeur, D., Demarée, G., Llasat, M. C., Macdonald, N., Retsö, D., Roald, L., Schmocker-Fackel, P., Amorim, I., Bělínová, M., Benito, G., Bertolin, C., Camuffo, D., Cornel, D., Doktor, R., Elleder, L., Enzi, S., Garcia, J. C., Glaser, R., Hall, J., Haslinger, K., Hofstätter, M., Komma, J., Limanówka, D., Lun, D., Panin, A., Parajka, J., Petrić, H., Rodrigo, F. S., Rohr, C., Schönbein, J., Schulte, L., Silva, L. P., Toonen, W. H. J., Valent, P., Waser, J., and Wetter, O.: Current European flood-rich period exceptional compared with past 500 years, Nature, 583, 560–566, https://doi.org/10.1038/s41586-020-2478-3, 2020.

Bonnet, R., Boé, J., Dayon, G., and Martin, E.: Twentieth-Century Hydrometeorological Reconstructions to Study the Multidecadal Variations of the Water Cycle Over France, Water Resour. Res., 53, 8366–8382, https://doi.org/10.1002/2017WR020596, 2017.

Bonnet, R., Boé, J., and Habets, F.: Influence of multidecadal variability on high and low flows: the case of the Seine basin, Hydrol. Earth Syst. Sci., 24, 1611–1631, https://doi.org/10.5194/hess-24-1611-2020, 2020.

Brigode, P., Oudin, L., and Perrin, C.: Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change?, J. Hydrol., 476, 410–425, https://doi.org/10.1016/j.jhydrol.2012.11.012, 2013.

Brigode, P., Paquet, E., Bernardara, P., Gailhard, J., Garavaglia, F., Ribstein, P., Bourgin, F., Perrin, C., and Andréassian, V.: Dependence of model-based extreme flood estimation on the calibration period: case study of the Kamp River (Austria), Hydrolog. Sci. J., 60, 1424–1437, https://doi.org/10.1080/02626667.2015.1006632, 2015.

Brigode, P., Brissette, F., Nicault, A., Perreault, L., Kuentz, A., Mathevet, T., and Gailhard, J.: Streamflow variability over the 1881–2011 period in northern Québec: comparison of hydrological reconstructions based on tree rings and geopotential height field reanalysis, Clim. Past, 12, 1785–1804, https://doi.org/10.5194/cp-12-1785-2016, 2016.

Brönnimann, S., Stucki, P., Franke, J., Valler, V., Brugnara, Y., Hand, R., Slivinski, L. C., Compo, G. P., Sardeshmukh, P. D., Lang, M., and Schaefli, B.: Influence of warming and atmospheric circulation changes on multidecadal European flood variability, Clim. Past, 18, 919–933, https://doi.org/10.5194/cp-18-919-2022, 2022.

Caillouet, L., Vidal, J.-P., Sauquet, E., and Graff, B.: Probabilistic precipitation and temperature downscaling of the Twentieth Century Reanalysis over France, Clim. Past, 12, 635–662, https://doi.org/10.5194/cp-12-635-2016, 2016.

Christensen, J. H. and Christensen, O. B.: A summary of the PRUDENCE model projections of changes in European climate by the end of this century, Climatic Change, 81, 7–30, https://doi.org/10.1007/s10584-006-9210-7, 2007.

Coron, L., Thirel, G., Delaigue, O., Perrin, C., and Andréassian, V.: The suite of lumped GR hydrological models in an R package, Environ. Modell. Softw., 94, 166–171, https://doi.org/10.1016/j.envsoft.2017.05.002, 2017.

Coron, L., Delaigue, O., Thirel, G., Dorchies, D., Perrin, C., and Michel, C.: airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling. R package version 1.7.6, Recherche Data Gouv, https://doi.org/10.15454/EX11NA, 2023.

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J. K., Lane, R., Lewis, M., Robinson, E. L., Wagener, T., and Woods, R.: CAMELS-GB: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, Earth Syst. Sci. Data, 12, 2459–2483, https://doi.org/10.5194/essd-12-2459-2020, 2020.

Cram, T. A., Compo, G. P., Yin, X., Allan, R. J., McColl, C., Vose, R. S., Whitaker, J. S., Matsui, N., Ashcroft, L., Auchmann, R., Bessemoulin, P., Brandsma, T., Brohan, P., Brunet, M., Comeaux, J., Crouthamel, R., Gleason, B. E., Groisman, P. Y., Hersbach, H., Jones, P. D., Jónsson, T., Jourdain, S., Kelly, G., Knapp, K. R., Kruger, A., Kubota, H., Lentini, G., Lorrey, A., Lott, N., Lubker, S. J., Luterbacher, J., Marshall, G. J., Maugeri, M., Mock, C. J., Mok, H. Y., Nordli, Ø., Rodwell, M. J., Ross, T. F., Schuster, D., Srnec, L., Valente, M. A., Vizi, Z., Wang, X. L., Westcott, N., Woollen, J. S., and Worley, S. J.: The International Surface Pressure Databank version 2, Geosci. Data J., 2, 31–46, https://doi.org/10.1002/gdj3.25, 2015.

David, C. H. and Frasson, R. P. D. M.: Blame the river not the rain, Nat. Geosci., 16, 282–283, https://doi.org/10.1038/s41561-023-01163-w, 2023.

Dawdy, D. R. and Bergmann, J. M.: Effect of rainfall variability on streamflow simulation, Water Resour. Res., 5, 958–966, https://doi.org/10.1029/WR005i005p00958, 1969.

Delaigue, O., Guimarães, G. M., Brigode, P., Génot, B., Perrin, C., Soubeyroux, J.-M., Janet, B., Addor, N., and Andréassian, V.: CAMELS-FR dataset: a large-sample hydroclimatic dataset for France to explore hydrological diversity and support model benchmarking, Earth Syst. Sci. Data, 17, 1461–1479, https://doi.org/10.5194/essd-17-1461-2025, 2025.

Devers, A., Vidal, J.-P., Lauvernet, C., and Vannier, O.: FYRE Climate: a high-resolution reanalysis of daily precipitation and temperature in France from 1871 to 2012, Clim. Past, 17, 1857–1879, https://doi.org/10.5194/cp-17-1857-2021, 2021.

Devers, A., Vidal, J.-P., Lauvernet, C., Vannier, O., and Caillouet, L.: 140-year daily ensemble streamflow reconstructions over 661 catchments in France, Hydrol. Earth Syst. Sci., 28, 3457–3474, https://doi.org/10.5194/hess-28-3457-2024, 2024.

Do, H. X., Gudmundsson, L., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata, Earth Syst. Sci. Data, 10, 765–785, https://doi.org/10.5194/essd-10-765-2018, 2018.

Duethmann, D., Blöschl, G., and Parajka, J.: Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change?, Hydrol. Earth Syst. Sci., 24, 3493–3511, https://doi.org/10.5194/hess-24-3493-2020, 2020.

Fossa, M., Dieppois, B., Massei, N., Fournier, M., Laignel, B., and Vidal, J.-P.: Spatiotemporal and cross-scale interactions in hydroclimate variability: a case-study in France, Hydrol. Earth Syst. Sci., 25, 5683–5702, https://doi.org/10.5194/hess-25-5683-2021, 2021.

Fowler, K., Peel, M., Western, A., and Zhang, L.: Improved Rainfall-Runoff Calibration for Drying Climate: Choice of Objective Function, Water Resour. Res., https://doi.org/10.1029/2017WR022466, 2018.

Fowler, K., Knoben, W., Peel, M., Peterson, T., Ryu, D., Saft, M., Seo, K.-W., and Western, A.: Many Commonly Used Rainfall-Runoff Models Lack Long, Slow Dynamics: Implications for Runoff Projections, Water Resour. Res., 56, e2019WR025286, https://doi.org/10.1029/2019WR025286, 2020.

Fowler, K., Peel, M., Saft, M., Peterson, T. J., Western, A., Band, L., Petheram, C., Dharmadi, S., Tan, K. S., Zhang, L., Lane, P., Kiem, A., Marshall, L., Griebel, A., Medlyn, B. E., Ryu, D., Bonotto, G., Wasko, C., Ukkola, A., Stephens, C., Frost, A., Gardiya Weligamage, H., Saco, P., Zheng, H., Chiew, F., Daly, E., Walker, G., Vervoort, R. W., Hughes, J., Trotter, L., Neal, B., Cartwright, I., and Nathan, R.: Explaining changes in rainfall–runoff relationships during and after Australia's Millennium Drought: a community perspective, Hydrol. Earth Syst. Sci., 26, 6073–6120, https://doi.org/10.5194/hess-26-6073-2022, 2022.

Ghosh, R., Müller, W. A., Baehr, J., and Bader, J.: Impact of observed North Atlantic multidecadal variations to European summer climate: a linear baroclinic response to surface heating, Clim. Dynam., 48, 3547–3563, https://doi.org/10.1007/s00382-016-3283-4, 2017.

Giuntoli, I., Renard, B., Vidal, J.-P., and Bard, A.: Low flows in France and their relationship to large-scale climate indices, J. Hydrol., 482, 105–118, https://doi.org/10.1016/j.jhydrol.2012.12.038, 2013.

GRDC: Data Portal, https://grdc.bafg.de/data/data_portal/, last access: 26 April 2023.

Gudmundsson, L., Do, H. X., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment, Earth Syst. Sci. Data, 10, 787–804, https://doi.org/10.5194/essd-10-787-2018, 2018.

Gudmundsson, L., Leonard, M., Do, H. X., Westra, S., and Seneviratne, S. I.: Observed Trends in Global Indicators of Mean and Extreme Streamflow, Geophys. Res. Lett., 46, 756–766, https://doi.org/10.1029/2018GL079725, 2019.

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

Hashemi, R., Brigode, P., Garambois, P.-A., and Javelle, P.: How can we benefit from regime information to make more effective use of long short-term memory (LSTM) runoff models?, Hydrol. Earth Syst. Sci., 26, 5793–5816, https://doi.org/10.5194/hess-26-5793-2022, 2022.

Haslinger, K., Hofstätter, M., Schöner, W., and Blöschl, G.: Changing summer precipitation variability in the Alpine region: on the role of scale dependent atmospheric drivers, Clim. Dynam., 57, 1009–1021, https://doi.org/10.1007/s00382-021-05753-5, 2021.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G. D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P. de, Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M. G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A. E., Addor, N., Brunner, M., Pool, S., Zappa, M., and Fenicia, F.: CAMELS-CH: hydro-meteorological time series and landscape attributes for 331 catchments in hydrologic Switzerland, Earth Syst. Sci. Data, 15, 5755–5784, https://doi.org/10.5194/essd-15-5755-2023, 2023.

Horton, P.: Analogue methods and ERA5: Benefits and pitfalls, Int. J. Climatol., 42, 4078–4096, https://doi.org/10.1002/joc.7484, 2022.

Klingler, C., Schulz, K., and Herrnegger, M.: LamaH-CE: LArge-SaMple DAta for Hydrology and Environmental Sciences for Central Europe, Earth Syst. Sci. Data, 13, 4529–4565, https://doi.org/10.5194/essd-13-4529-2021, 2021.

Kuentz, A., Mathevet, T., Gailhard, J., and Hingray, B.: Building long-term and high spatio-temporal resolution precipitation and air temperature reanalyses by mixing local observations and global atmospheric reanalyses: the ANATEM model, Hydrol. Earth Syst. Sci., 19, 2717–2736, https://doi.org/10.5194/hess-19-2717-2015, 2015.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management, Front. Ecol. Environ., 9, 494–502, https://doi.org/10.1890/100125, 2011.

Lemaitre-Basset, T., Oudin, L., and Thirel, G.: Evapotranspiration in hydrological models under rising CO_2: a jump into the unknown, Climatic Change, 172, 36, https://doi.org/10.1007/s10584-022-03384-1, 2022a.

Lemaitre-Basset, T., Oudin, L., Thirel, G., and Collet, L.: Unraveling the contribution of potential evaporation formulation to uncertainty under climate change, Hydrol. Earth Syst. Sci., 26, 2147–2159, https://doi.org/10.5194/hess-26-2147-2022, 2022b.

Lindström, G. and Bergström, S.: Runoff trends in Sweden 1807–2002, Hydrolog. Sci. J., 49, 69–83, https://doi.org/10.1623/hysj.49.1.69.54000, 2004.

Lorenzo-Lacruz, J., Morán-Tejeda, E., Vicente-Serrano, S. M., Hannaford, J., García, C., Peña-Angulo, D., and Murphy, C.: Streamflow frequency changes across western Europe and interactions with North Atlantic atmospheric circulation patterns, Global Planet. Change, 212, 103797, https://doi.org/10.1016/j.gloplacha.2022.103797, 2022.

Martel, J.-L., Arsenault, R., Lachance-Cloutier, S., Castaneda-Gonzalez, M., Turcotte, R., and Poulin, A.: Improved historical reconstruction of daily flows and annual maxima in gauged and ungauged basins, J. Hydrol., 129777, https://doi.org/10.1016/j.jhydrol.2023.129777, 2023.

Masseroni, D., Camici, S., Cislaghi, A., Vacchiano, G., Massari, C., and Brocca, L.: The 63-year changes in annual streamflow volumes across Europe with a focus on the Mediterranean basin, Hydrol. Earth Syst. Sci., 25, 5589–5601, https://doi.org/10.5194/hess-25-5589-2021, 2021.

Montanari, A.: Hydrology of the Po River: looking for changing patterns in river discharge, Hydrol. Earth Syst. Sci., 16, 3739–3747, https://doi.org/10.5194/hess-16-3739-2012, 2012.

Müller, M. F., Roche, K. R., and Dralle, D. N.: Catchment processes can amplify the effect of increasing rainfall variability, Environ. Res. Lett., 16, 084032, https://doi.org/10.1088/1748-9326/ac153e, 2021.

Nasreen, S., Součková, M., Vargas Godoy, M. R., Singh, U., Markonis, Y., Kumar, R., Rakovec, O., and Hanel, M.: A 500-year annual runoff reconstruction for 14 selected European catchments, Earth Syst. Sci. Data, 14, 4035–4056, https://doi.org/10.5194/essd-14-4035-2022, 2022.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2-Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, J. Hydrol., 303, 290–306, https://doi.org/10.1016/j.jhydrol.2004.08.026, 2005.

Oudin, L., Perrin, C., Mathevet, T., Andréassian, V., and Michel, C.: Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models, J. Hydrol., 320, 62–83, https://doi.org/10.1016/j.jhydrol.2005.07.016, 2006.

Paprotny, D., Terefenko, P., and Śledziowski, J.: HANZE v2.1: an improved database of flood impacts in Europe from 1870 to 2020, Earth Syst. Sci. Data, 16, 5145–5170, https://doi.org/10.5194/essd-16-5145-2024, 2024.

Parodi, A., Ferraris, L., Gallus, W., Maugeri, M., Molini, L., Siccardi, F., and Boni, G.: Ensemble cloud-resolving modelling of a historic back-building mesoscale convective system over Liguria: the San Fruttuoso case of 1915, Clim. Past, 13, 455–472, https://doi.org/10.5194/cp-13-455-2017, 2017.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, J. Hydrol., 279, 275–289, https://doi.org/10.1016/S0022-1694(03)00225-7, 2003.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G. H., Peubey, C., Thépaut, J.-N., Trémolet, Y., Hólm, E. V., Bonavita, M., Isaksen, L., and Fisher, M.: ERA-20C: An Atmospheric Reanalysis of the Twentieth Century, J. Climate, 29, 4083–4097, https://doi.org/10.1175/JCLI-D-15-0556.1, 2016.

Pushpalatha, R., Perrin, C., Le Moine, N., and Andréassian, V.: A review of efficiency criteria suitable for evaluating low-flow simulations, J. Hydrol., 420–421, 171–182, https://doi.org/10.1016/j.jhydrol.2011.11.055, 2012.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/ (last access: 17 October 2025), 2024.

Renard, B., McInerney, D., Westra, S., Leonard, M., Kavetski, D., Thyer, M., and Vidal, J.-P.: Floods and Heavy Precipitation at the Global Scale: 100-Year Analysis and 180-Year Reconstruction, J. Geophys. Res.-Atmos., 128, e2022JD037908, https://doi.org/10.1029/2022JD037908, 2023.

Rust, W., Bloomfield, J. P., Cuthbert, M., Corstanje, R., and Holman, I.: The importance of non-stationary multiannual periodic-

ities in the North Atlantic Oscillation index for forecasting water resource drought, Hydrol. Earth Syst. Sci., 26, 2449–2467, https://doi.org/10.5194/hess-26-2449-2022, 2022.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., Kennedy, J., Spencer, L. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Blancq, F. L., Lee, T.-C., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, G. W. K., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Smith, C. A., Tinz, B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., and Wyszyński, P.: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system, Quarterly Journal of the Royal Meteorological Society, 145, 2876–2908, https://doi.org/10.1002/qj.3598, 2019.

Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L. M., van Lanen, H. A. J., Sauquet, E., Demuth, S., Fendekova, M., and Jódar, J.: Streamflow trends in Europe: evidence from a dataset of near-natural catchments, Hydrol. Earth Syst. Sci., 14, 2367–2382, https://doi.org/10.5194/hess-14-2367-2010, 2010.

Tarasova, L., Lun, D., Merz, R., Blöschl, G., Basso, S., Bertola, M., Miniussi, A., Rakovec, O., Samaniego, L., Thober, S., and Kumar, R.: Shifts in flood generation processes exacerbate regional flood anomalies in Europe, Commun. Earth Environ., 4, 1–12, https://doi.org/10.1038/s43247-023-00714-8, 2023.

Tarboton, D. G.: TauDEM 5.1: Guide to using the TauDEM command line functions, https://hydrology.usu.edu/taudem/taudem5/documentation.html (last access: 17 October 2025), 2013.

Teutschbein, C., Quesada Montano, B., Todorović, A., and Grabs, T.: Streamflow droughts in Sweden: Spatiotemporal patterns emerging from six decades of observations, J. Hydrol. Reg. Stud., 42, 101171, https://doi.org/10.1016/j.ejrh.2022.101171, 2022.

Thébault, C., Perrin, C., Andréassian, V., Thirel, G., Legrand, S., and Delaigue, O.: Multi-model approach in a variable spatial framework for streamflow simulation, Hydrol. Earth Syst. Sci., 28, 1539–1566, https://doi.org/10.5194/hess-28-1539-2024, 2024.

Trotter, L., Saft, M., Peel, M. C., and Fowler, K. J. A.: Symptoms of Performance Degradation During Multi-Annual Drought: A Large-Sample, Multi-Model Study, Water Resour. Res., 59, e2021WR031845, https://doi.org/10.1029/2021WR031845, 2023.

Valéry, A., Andréassian, V., and Perrin, C.: "As simple as possible but not simpler": what is useful in a temperature-based snow-accounting routine? Part 2 – Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments, J. Hydrol., 517, 1176–1187, https://doi.org/10.1016/j.jhydrol.2014.04.058, 2014.

Wan, Y., Chen, J., Xu, C.-Y., Xie, P., Qi, W., Li, D., and Zhang, S.: Performance dependence of multi-model combination methods on hydrological model calibration strategy and ensemble size, J. Hydrol., 603, 127065, https://doi.org/10.1016/j.jhydrol.2021.127065, 2021.

Wilcoxon, F.: Individual Comparisons by Ranking Methods, Biometrics Bulletin, 1, 80–83, https://doi.org/10.2307/3001968, 1945.

Wilhelm, B., Rapuc, W., Amann, B., Anselmetti, F. S., Arnaud, F., Blanchet, J., Brauer, A., Czymzik, M., Giguet-Covex, C., Gilli, A., Glur, L., Grosjean, M., Irmler, R., Nicolle, M., Sabatier, P., Swierczynski, T., and Wirth, S. B.: Impact of warmer climate periods on flood hazard in the European Alps, Nat. Geosci., 1–6, https://doi.org/10.1038/s41561-021-00878-y, 2022.

Woodruff, S. D., Worley, S. J., Lubker, S. J., Ji, Z., Eric Freeman, J., Berry, D. I., Brohan, P., Kent, E. C., Reynolds, R. W., Smith, S. R., and Wilkinson, C.: ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive, Int. J. Climatol., 31, 951–967, https://doi.org/10.1002/joc.2103, 2011.

Yeste, P., García-Valdecasas Ojeda, M., Gámiz-Fortis, S. R., Castro-Díez, Y., Bronstert, A., and Esteban-Parra, M. J.: A large-sample modelling approach towards integrating streamflow and evaporation data for the Spanish catchments, Hydrol. Earth Syst. Sci., 28, 5331–5352, https://doi.org/10.5194/hess-28-5331-2024, 2024.