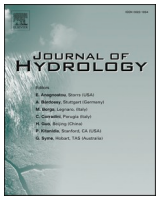
Journal of Hydrology 623 (2023) 129807

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol



Research papers

Uni- and multivariate bias adjustment of climate model simulations in Nordic catchments: Effects on hydrological signatures relevant for water resources management in a changing climate

Faranak Tootoonchi [a,*], Andrijana Todorović [b], Thomas Grabs [a], Claudia Teutschbein [a]

[a] Department of Earth Sciences, Uppsala University, Uppsala, Sweden
[b] University of Belgrade, Faculty of Civil Engineering, Institute for Hydraulic and Environmental Engineering, Belgrade, Serbia

ARTICLE INFO

Keywords:
Climate change
Bias adjustment
Bias correction
Climate model
Hydrological signatures
High latitudes

ABSTRACT

Hydrological climate-change-impact studies depend on climatic variables simulated by climate models. Due to parametrization and numerous simplifications, however, climate-model outputs come with systematic biases compared to the observations. In the past decade, several methods of different complexity and dimensionality for adjustment of such biases were introduced, but their benefits for impact studies and accurate streamflow projections are still debated. In this paper, we evaluated the ability of two state-of-the-art, advanced multivariate bias-adjustment methods to accurately reproduce 16 hydrological signatures, and compared their performance against two parsimonious univariate bias-adjustment methods based on a multi-criteria performance evaluation. The results indicated that all bias-adjustment methods considerably reduced biases and increased the consistency of simulated hydrological signatures. The added value of multivariate methods in maintaining dependence structures between precipitation and temperature was not systematically reflected in the resulting hydrological signatures, as they were generally outperformed by univariate methods. The benefits of multivariate methods only emerged for low-flow signatures in snowmelt-driven catchments. Based on these findings, we identified the most suitable bias-adjustment methods for water-resources management in Nordic regions under a changing climate, and provided practical guidelines for the selection of bias-adjustment methods given specific research targets and hydroclimatic regimes.

1. Introduction

Streamflow regimes are considerably impacted by climate change due to simultaneous effects of rising temperatures and changing precipitation patterns (Villarini and Wasko, 2021). Higher temperatures cause increasing atmospheric water demand, which, in turn, changes precipitation timing and amount (Cheng et al., 2019; Hanus et al., 2021; IPCC, 2014). At the same time, higher temperatures affect the timing and duration of snow season (Gergel et al., 2017), snowmelt (Hakala et al., 2018) and resulting streamflow (Clow, 2010). Concurrent shifts in precipitation and temperature also influence the availability and access to water (Clifton et al., 2018), as well as the timing and magnitude of extremes like floods or droughts (Dankers et al., 2014; Dankers and Feyen, 2009; Villarini and Wasko, 2021).

In Nordic countries, changes in hydroclimatic drivers of streamflow have already been observed through warming temperatures, resulting in shorter ice cover durations (Hallerbäck et al., 2021) and general wetting trends (Chen et al., 2021a). Current future projections indicate an intensification of climate change in the Northern Hemisphere by the end of the century (IPCC, 2021), which would inevitably influence flow regimes in Nordic regions (Teutschbein et al., 2015). Anticipated changes include e.g., a decrease in daily flows (Matti et al., 2017), shorter snow seasons (Hallerbäck et al., 2021), an earlier onset of spring floods with lower peaks, slightly lower summer flows, and higher winter base-flows in snowmelt-driven catchments (Teutschbein et al., 2015). Together, these changes could lead to a higher frequency of rainfall-driven floods in Nordic catchments (Bergström et al., 2001; Matti et al., 2016) and to a shift in the current flood regimes from spring/summer to autumn/winter (Arheimer and Lindström, 2015; Vormoor et al., 2015). In contrast, rainfall-driven catchments in southern regions of the Nordic countries will likely experience streamflow reductions during most of the year, except for winter, due to increasing evapotranspiration rates and longer

* Corresponding author.
E-mail address: faranak.tootoonchi@geo.uu.se (F. Tootoonchi).

https://doi.org/10.1016/j.jhydrol.2023.129807
Received 3 September 2022; Received in revised form 3 March 2023; Accepted 8 June 2023
Available online 14 June 2023

growing seasons (Teutschbein et al., 2017). Such shifts can severely limit the availability of water for ecosystems (Carpenter et al., 1992), and adversely affect human-livelihood by threatening water availability for agriculture and hydropower (Berghuijs et al., 2014). To develop efficient mitigation and adaptation strategies, water resource managers need reliable information generated by hydrological impact studies.

Hydrological impact studies rely on a modeling chain to simulate future streamflow projections (Hakala et al., 2019; Wagener et al., 2022). Climate models (CMs) provide global scale meteorological data representative for one or several greenhouse gas concentration pathways (Refsgaard et al., 2014). The data can be further utilized as initial and boundary conditions for regional CMs to obtain meteorological projects at finer spatial and temporal resolutions. Theoretically, the outputs of the regional CMs could be used as direct input to hydrological models to simulate climate impacts on the water cycle at the catchment scale (Hakala et al., 2019). Nevertheless, due to parametrization and numerical limitations, CMs generally provide biased simulations with systematic errors, compared to observations (Maraun, 2016; Teutschbein and Seibert, 2012). These biases can manifest in the amount, timing, and the dependence between the climate variables at hand (François et al., 2020; Hakala et al., 2018; Piani and Haerter, 2012; Tootoonchi et al., 2022a; Wilcke et al., 2013).

To reduce these biases, different bias-adjustment methods have been suggested. These range from simple univariate bias adjustment (BA) methods, such as delta-change, quantile mapping (QM) or the delta mapping method (DM), to more advanced multivariate BA methods such as copula and bias correction in multiple dimensions (MBCn).

Univariate BA methods adjust only one climate variable at the time (e.g. in Hakala et al., 2018; Teutschbein & Seibert, 2012). In particular, the QM methods, such as distribution scaling (DS) and quantile delta mapping (QDM), are often preferred due to their simple implementation procedure and their arguably robust capabilities to adjust various aspects of CM simulations (Berg et al., 2022; Gudmundsson et al., 2012; Teutschbein and Seibert, 2012; Tootoonchi et al., 2022a). Multivariate methods, however, account for the interlinkages between climate variables and adjust both univariate and multivariate characteristics of these series (e.g. in François et al., 2020; Meyer et al., 2019; Räty et al., 2018; Mehrotra et al., 2018; Mehrotra and Sharma, 2021), which can be relevant for extreme and compound events (Meyer et al., 2019; Tootoonchi et al., 2022a; Zscheischler et al., 2019). But these methods can be more challenging to implement compared to univariate ones, as they demand more computational resources, and advanced statistical and programming skills of the user. They may also adversely affect temporal characteristics of the climate variables: for example, they can distort autocorrelation in the series (François et al., 2020; Tootoonchi et al., 2022a; Van de Velde et al., 2022).

In a Nordic context, previous research showed that both multivariate methods (copula and MBCn) perform similarly to their univariate counterparts (DS and QDM, respectively) in adjustment of biases in univariate characteristics of precipitation and temperature series in numerous Swedish catchments (Tootoonchi et al., 2022a). However, the multivariate BA methods outperformed their univariate counterparts in reproducing the positive dependence between precipitation and temperature during colder winter months, the negative dependence during warm summer months, and the Clausius-Clapeyron relation. Correlations between precipitation and temperature varied across catchments (i.e., latitudes and climates) and seasons. Although both multivariate methods performed well in this regard, they were inferior to the univariate methods in adjusting temporal characteristics, such as cross-correlation.

Considering the advantages of multivariate methods in accounting for dependence between the adjusted variables, their impact on the accuracy of simulated streamflows is not clear. Specifically, the question if the use of advanced multivariate BA methods results in considerably more reliable streamflow projections compared to their univariate alternatives has remained opened. On one hand, it is well recognized that univariate behavior of climate variables, such as mean and variance of temperature and precipitation, should be well represented (Guo et al., 2020; Teutschbein and Seibert, 2012; Tootoonchi et al., 2022a) to properly simulate future streamflows. On the other hand, other characteristics, such as temporal behavior or the dependence between precipitation and temperature may also play a key role in the accuracy of streamflow simulations (Chen et al., 2021a). For instance, dependence between precipitation and temperature influences water holding capacity of the atmosphere (Panthou et al., 2014), which can affect both intensity and timing of extreme precipitation (Myhre et al., 2019), evapotranspiration (AghaKouchak et al., 2014), and snowmelt (Hakala et al., 2018). As a result, preserving the dependence between climatic variables could be critical to reliably project future water balance, as well as the timing and severity of future floods or droughts (Guo et al., 2020; Meyer et al., 2019; Zehe and Sivapalan, 2009). To reliably simulate streamflow from simulated climate data, hydrological models need to be calibrated (Pool et al., 2017). Model performance in the calibration (and evaluation) periods have traditionally been quantified in term of efficiency measures, such as Nash-Sutcliffe (Nash and Sutcliffe, 1970) or Kling-Gupta coefficients (Gupta et al., 2009; Yilmaz et al., 2010). While these measures are good indicators of the overall model performance, they do not indicate the ability of a hydrological model to reproduce specific features of streamflow regimes (i.e., hydrological signatures), such as the timing or magnitude of (seasonal) high or low flows (Addor and Fischer, 2015; Feng and Beighley, 2020; Mendoza et al., 2016). Such signatures are, however, essential for flood hazard assessment and water resource management (Brunner et al., 2021; Todorović et al., 2022). Thus, we argue that using a range of hydrological signatures (Kabuya et al., 2020; McMillan, 2021; Zhang et al., 2018) is an essential to complement traditional efficiency measures and enable a process–informed evaluation of BA methods.

While a number of studies evaluated the impact of univariate BA methods on simulated hydrological variables (e.g., Hakala et al., 2018; Muerth et al., 2013; Teutschbein & Seibert, 2012), relatively few studies evaluated multivariate BA methods in this context. For instance, Räty et al. (2018) compared streamflow simulations after applications of two univariate and two multivariate BA methods in four European catchments, but did not find any substantial improvement after application of multivariate BA methods in these catchments. Meyer et al. (2019), however, found that multivariate BA outperformed univariate BA when simulating water storage in snow and glaciers in two Swiss alpine catchments. A North American study found that multivariate BA methods improve accuracy of streamflow simulations more in catchments of arid and warm regions, than in catchments in colder regions with a prolonged snow season (Guo et al., 2020). It remains, however, unclear if these findings can be transferred to Nordic catchments. Additionally, research in this field has so far produced little guidance for impact modelers who are interested in one or several specific streamflow characteristics, and who need to select a BA method that is both easily applicable and fit for a specific task at hand.

To bridge these gaps, this study assessed the suitability of different BA methods for accurate reproduction of hydrological signatures relevant for water resources management in the Nordic regions under climate change. This assessment was based on the outputs of 10 CMs in 50 Swedish catchments, which covered a large variety of climatic and hydrologic conditions. Climate variables in this multi-catchment ensemble were adjusted with two parsimonious univariate, and two advanced multivariate bias adjustment methods. Bias-adjusted temperature and precipitation series were then used to force the hydrological HBV-light model (Bergström, 1976; Lindström et al., 1997; Seibert and Vis, 2012) and to simulate streamflows in each catchment.

Evaluation of the BA methods was based on a comprehensive set of hydrological signatures describing (a) the overall water balance and streamflow dynamics, (b) seasonal streamflows, (c) low- and (d) high-flow characteristics. The evaluation was tailored to enable addressing the following research questions:

1) *Performance and similarities of BA methods:* Can the application of BA methods improve the accuracy in reproducing hydrological signatures in Nordic catchments, and are there any considerable differences among the BA methods, or recurring similarities can be found?
2) *Robustness of BA methods:* Do more complex, multivariate BA methods outperform parsimonious univariate BA methods by frequent improvements and consistent performance across (a) climate models, (b) catchments and (c) hydroclimatic regimes?
3) *Comparison-based recommendations:* Can we identify one or several BA method(s) that consistently outperform(s) the other methods in simulating hydrological signatures relevant for water resources management under changing climate in the Nordic regions?

We synthetized all results to provide specific recommendations for the selection of BA methods depending on the hydrological signature of interest, and put these recommendations in the context of effective water resources management under changing climate conditions.

2. Material and methods

2.1. Study area

Sweden, a country with a land area of roughly 408,000 km^2, is heavily forested (69% of the land area) with a large number of lakes, wetlands and streams (9%). Roughly 8% of the land is covered by shrubs and grass land, 8% are agricultural land, 3% urban areas and 3% open land and glaciers. Sweden stretches over three Köppen-Geiger (Beck et al., 2018) climatic zones (Fig. 1a): (1) the polar tundra climate zone (ET) with monthly mean temperatures below 10 °C covers the Scandinavian Mountains in North-western Sweden, (2) the subarctic boreal climate (Dfc) with cool summers, and winters with persistent seasonal snow and soil frost covers Central and Northern Sweden, and (3) the warm summer hemiboreal climate zone (Dfb) covers Southern Sweden.

In this study, we used a set of 50 Swedish catchments (Fig. 1a) spread latitudinally from 55.95°N to 68.37°N and longitudinally between 11.54°E to 22.31°E. Catchment areas spanned from 2 to 8425 km^2 with average elevations between 12 and 942 m.a.s.l.. Selection of the catchments was primarily dictated by the availability of data, i.e., only those catchments for which continuous daily temperature, precipitation and streamflow observations were available for the period 1961–2004 were included. Additional selection criteria included a low percentage of lakes (<37%), glaciers and urbanized areas (<3.1%), and a low degree (<30.2%) of regulation, i.e., reservoir volume relative to the mean annual flow volume from its draining area (cf., Todorović et al., 2022). The resulting 50 catchments were spatially distributed across all three climatic zones of Sweden (Fig. 1a) and, thus, featured diverse hydroclimatic characteristics.

During the period 1961–2004, total annual precipitation over the chosen catchments ranged from 1500 mm·year^1 in north-western Sweden to 400 mm·year^{-1} in southern Sweden (Fig. 1b), while annual mean temperature varied from −3.0 °C in the North to + 7.7 °C in the South (Fig. 1c). Hydrological regimes also featured a north–south gradient (Fig. 1a, right panel), with snowmelt-driven regimes in northern regions and rainfall-driven regimes in southern regions (Arheimer and Lindström, 2015; Matti et al., 2017).

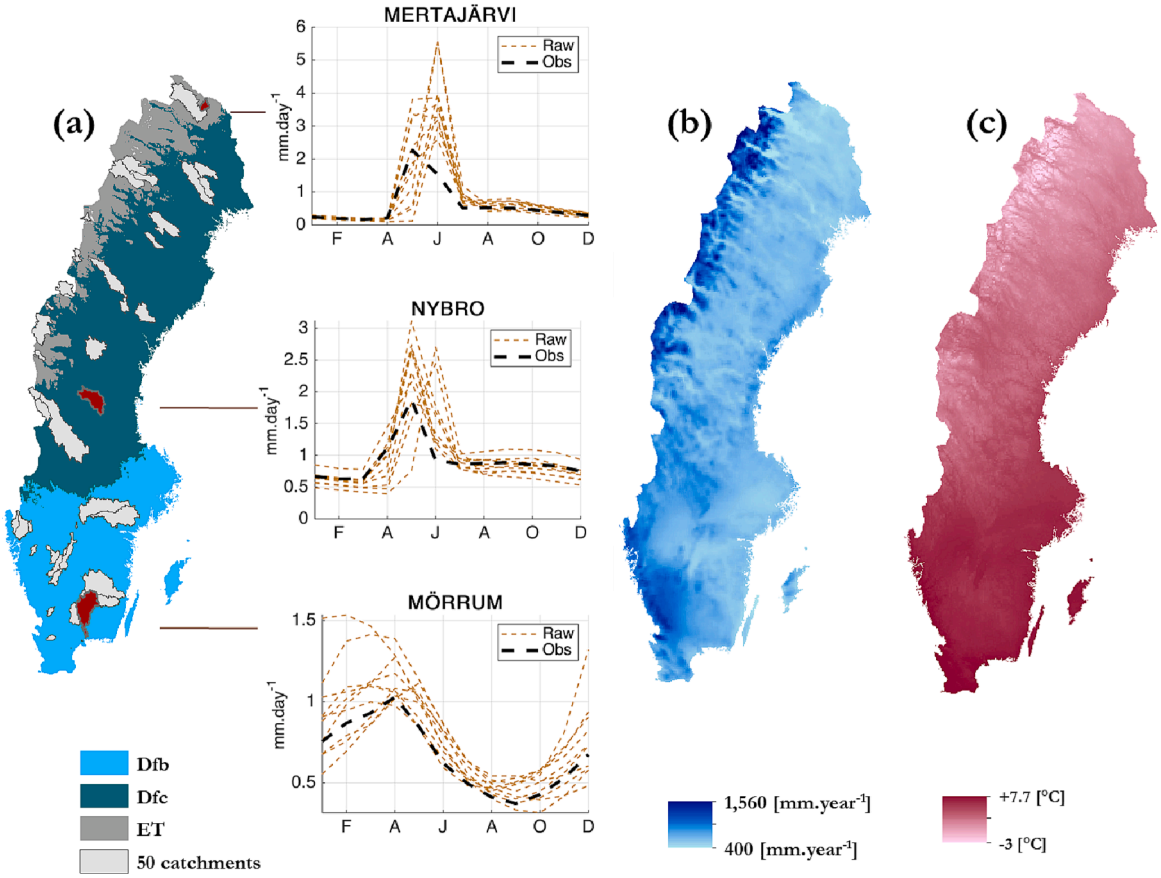


Fig. 1. Overview of the study area, including the (a) spatial span of the three Köppen-Geiger climatic zones covering Sweden (Beck et al., 2018) with the location of the 50 catchments and three selected catchments showcasing the different flow regimes in each climate zone, with typical runoff regimes in catchments of these three climate zones, (b) mean observed annual precipitation, and (c) mean annual temperature over the period 1961–2004.

2.2. Data

2.2.1. Observed variables: Precipitation, temperature, and streamflow

Gridded daily mean values of observed temperature and precipitation were obtained from the Swedish Meteorological and Hydrological Institute (SMHI) as spatially interpolated 4 km × 4 km national grids for the period 1961–2020 (Johansson, 2002). Daily streamflow records were downloaded from a publicly accessible streamflow database (https://vattenwebb.smhi.se/) hosted by SMHI, while geospatial data of the catchment boundaries for each of the 50 catchments were obtained from SMHI's SVAR database (Eklund, 2011).

2.2.2. Climate data

To cover a wide range of uncertainty, we used outputs from different combinations of global and regional climate models (CMs) for the historic period of 1961–2004 (Table 1), acquired from the Coupled Model Inter-comparison Project Phase 5 (CMIP5) together with Coordinated Regional Climate Downscaling experiment (CORDEX) initiative (Jacob et al., 2014). We selected only those CMs that had both historic and future precipitation and temperature simulations available until the end of the century (2100), for three greenhouse gas concentration trajectories (i.e., RCP 2.6, 4.5 and 8.5), and for the highest available resolution 0.11° (roughly 12.5 km). This selection protocol left us with 10 RCM-GCM combinations. If simulations were available in several versions (e.g., v1 and v2), we opted for the most recent version (i.e., v2). Gridded daily precipitation and temperature data were spatially averaged over each catchment from values in all grid cells (partly) covering the catchment of interest.

2.3. Bias adjustment of climate models

We selected four different BA methods: two commonly used univariate methods and two more recently developed multivariate methods. The selected methods either rely on a theoretical probably distribution (hereafter referred to as distribution-based methods) or follow a distribution-free approach. While other univariate methods, e.g., linear scaling (Lenderink et al., 2007) or variance scaling (Chen et al., 2011), as well as alternative multivariate methods, e.g., dynamical optimal transport correction (Robin et al., 2019), exist, the four selected methods represent the most frequently utilized to date. Note that multivariate methods can additionally be categorized according to their conceptual features (i.e., if they show stochastic or deterministic behavior), or the order in which the dependence is adjusted (François et al., 2020).

Each BA method was calibrated to adjust CM simulations to precipitation and temperature observed over 22 years (between 1961 and 1982). Precipitation and temperature data recorded in the subsequent 22 years (evaluation period 1983–2004) were adjusted directly based on the parameters estimated during the calibration period. This set-up was selected both due to data availability (i.e., historic simulations of CMIP5 CM outputs span until 2005) and due to the requirements of equally long time series for forming the empirical copula. For each period, BA was performed on a monthly basis to preserve seasonal variations. In accordance with the common practice, drizzle days with<1 mm·day^{-1} of precipitation were set to dry days (i.e., days with zero precipitation) before the evaluation of the BA methods (Teutschbein and Seibert, 2012; Tootoonchi et al., 2022a).

Table 1
Selected climate model combinations from CMIP5 - EURO CORDEX archive.

CM ID	Institute	Global Climate Models (GCMs)	Parameters	Regional Climate models (RCM)	Version
1	CLMcom	ICHEC-EC-EARTH	r12i1p1	CCLM4-8–17	v1
2	CNRM	CNRM-CERFACS-CNRM-CM5	r1i1p1	ALADIN53	v1
3	CNRM	CNRM-CERFACS-CNRM-CM5	r1i1p1	ALADIN63	v2
4	DMI	ICHEC-EC-EARTH	r3i1p1	HIRHAM5	v2
5	GERICS	NCC-NorESM1-M	r1i1p1	REMO2015	v1
6	KNMI	ICHEC-EC-EARTH	r12i1p1	RACMO22E	v1
7	KNMI	CNRM-CERFACS-CNRM-CM5	r1i1p1	RACMO22E	v2
8	MPI-CSC	MPI-M−MPI−ESM−LR	r2i1p1	REMO2009	v1
9	MPI-CSC	MPI-M−MPI−ESM−LR	r1i1p1	REMO2009	v1
10	RMIB-UGent	CNRM-CERFACS-CNRM-CM5	r1i1p1	ALARO-0	v1

2.3.1. Univariate methods

Quantile mapping (QM) is a category of univariate BA methods that adjusts the entire probability distributions of CM outputs to match the corresponding distribution of observations (Chen et al., 2013; Gudmundsson et al., 2012). QM methods have been shown to be superior to other univariate methods, such as linear scaling or variance scaling (Teutschbein and Seibert, 2012), and are, thus, considered the benchmark to compare multivariate methods with. We adopted two versions of QM: (1) the distribution-based method called *distribution scaling* (DS), and (2) the distribution-free method called *quantile delta mapping* (QDM).

Distribution scaling (DS) implies the selection of particular theoretical distribution families to form cumulative distribution functions (CDFs) of variables. Typically, a Gamma distribution with the shape parameter α and the scale parameter β is used to represent precipitation intensity on wet days (Thom, 1951), while a Gaussian distribution with mean μ and variance σ is often applied to temperature of the entire record. In line with the previous studies (e.g., Gudmundsson et al., 2012; Teutschbein & Seibert, 2012; Yang et al., 2010), our preliminary analysis confirmed the suitability of these two distribution families to represent CM outputs, and hence we used them in this paper.

In contrast to DS, the distribution-free **quantile delta mapping (QDM)** approach explicitly preserves relative changes in simulated CM outputs (Li et al., 2010), and, at the same time, adjusts biases in quantiles of the simulated variables by taking into account observations in the calibration period (Cannon et al., 2015; Olsson et al., 2009; Willems and Vrac, 2011). This method is, thus, beneficial if there is enough confidence in the plausibility of the change signal simulated by the CM (Cannon, 2018; Maraun, 2016). A visual representation of QDM approach in comparison to other quantile mapping methods can be found in the paper by Berg et al. (2022).

2.3.2. Multivariate methods

In contrast to univariate BA methods, multivariate methods additionally adjust the dependence between variables along with their univariate features (Piani and Haerter, 2012). Two frequently used multivariate methods in impact studies include distribution-based copula, and distribution-free multivariate bias correction in *n* dimensions methods (MBCn).

In **copula adjustment methods**, the joint rank dependence of two (or more) variables is represented by a mathematical function, which couples their marginal distributions. For two continuous random set of variables X and Y, the copula can be represented as (Sklar, 1959):

$$H(X,Y) = C[F(X), G(Y)] \quad (1)$$

where $H(X,Y)$ features the cumulative distribution function (CDF) of the bivariate distribution of X and Y, $F(X)$ and $G(Y)$ are their respective univariate (marginal) CDFs, and C the copula function that joins two marginal CDFs together. For a comprehensive introduction to the general concept and mathematical basis of copulas the reader is referred to seminal papers by Genest & Favre (2007) and Nelsen (2006).

For adjusting biases in CM simulations, copulas were either used by

fitting theoretical copula families to the series (Li et al., 2014; Räty et al., 2018) or in empirical form (Bárdossy and Pegram, 2012; Piani and Haerter, 2012; Vrac, 2018). Here, we tested four theoretical copula families, consisting of three Archimedean copulas (Frank, Clayton and Gumbel) and one elliptical copula (Gaussian).

To apply the copulas, the formal procedure by Tootoonchi et al. (2022b), was adopted in this study: first, the data was pre-processed and screened for non-stationarity, trends and ties. Ties are data with the same rank, such as days with zero precipitation (dry days). As daily data featured numerous dry days, a jittering algorithm similar to the one proposed by Pappadà et al. (2017) was applied to the precipitation series to break the ties. In this approach, adjustment of precipitation is performed by replacement of zero precipitation values (i.e., dry days) with extremely small random but nonzero values (Vrac et al., 2016).

Thereafter, theoretical copula families were fitted to the observed data. To ensure the suitability of the selected theoretical copula families, the Cramér-von-Mises goodness-of-fit test was applied (Genest et al., 2009). When the test failed to provide an admissible copula, i.e., the test yielded p-value < 0.05 (following e.g., Brunner et al., 2019b; Genest et al., 2009; Genest and Remillard, 2008), empirical copulas were used instead. Using the fitted copula parameters, a new copula was computed for the CM-simulated precipitation and temperature, creating adjusted dependence characteristics. Thereafter, the DS method was applied on the margins (i.e., precipitation and temperature) to adapt the univariate characteristics of these series. For more detailed explanation of the copula approach, the reader is referred to other papers (e.g., Genest and Favre, 2007; Tootoonchi et al., 2022b).

Recently, Cannon (2018) introduced **MBCn** to the bias adjustment community, which is available as a free R package (https://rdrr.io/cran/MBC/man/MBCn.html). Since its publication, MBCn has been adopted as the benchmark for testing multivariate methods in numerous studies (e.g., de Velde et al., 2022; François et al., 2020; Meyer et al., 2019; Singh & Reza Najafi, 2020). In this method, randomly generated orthogonal matrices are repeatedly applied on the variables, rotating them in space. At each rotation, a selected univariate method (e.g., QDM) is applied on each of variables separately to adjust biases in their univariate features. This iterative process is continued until the variables feature multivariate similarity according to the following energy-distance metric (Székely and Rizzo, 2013):

$$D^2(F,G) = 2E\|X-Y\| - E\|X-X'\| - E\|Y-Y'\| \quad (2)$$

where X=$\{X_1, X_2, \ldots, X_n\}$ and Y=$\{Y_1, Y_2, \ldots, Y_n\}$ are two independent random vectors (here precipitation and temperature series), and F and G are CDFs of X and Y, while E is the expected value of the Euclidean norm $\|.\|$, and X' and Y' are independent identically distributed copies of X and Y after rotation. For more information regarding this method, the reader is refer to Cannon (2018), and for specific details of energy distance method to Rizzo & Székely (2016).

2.4. Hydrological modeling with the HBV-light model

2.4.1. The HBV-light model

Daily streamflow was simulated in each of the 50 catchments with the standard version of the HBV-light model (Seibert and Vis, 2012), which has been widely adopted in many regions (Seibert and Bergström, 2022). The model consists of (1) snow, (2) soil moisture, (3) response, and (4) routing routines (Seibert and Vis, 2012). In this study, we used the spatially-lumped standard version of the HBV-light model. The model required daily precipitation, temperature and potential evapotranspiration as inputs. Series of daily potential evapotranspiration were computed from the temperature series (observed and simulated by the CMs, raw and bias-adjusted) by applying the Hamon method (Hamon, 1961).

2.4.2. Calibration and evaluation of HBV-light

Following a one-year warm-up period from 1961 to 1962, HBV-light was calibrated in each catchment against 21 years of daily streamflow observations (from 1963 to 1982). The calibration was performed in each catchment separately by using the built-in GAP optimization algorithm (Seibert, 2000).

The GAP-optimization was run 5000 times after randomly drawing 50 initial parameter sets. The model calibration was performed by maximizing a composite objective function (F_{obj}), which was selected to achieve a balanced model performance in high- and low flows, and in reproducing streamflow volume:

$$F_{obj} = 0.75(NSE) + 0.2(\log NSE) + 0.05(VE) \quad (3)$$

where *NSE* denotes the (high-flow sensitive) Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970) computed from daily flows, *logNSE* is the (low-flow sensitive) Nash-Sutcliffe efficiency computed from log-transformed flows to put more emphases on low flows (Oudin et al., 2006; Santos et al., 2018), and *VE* denotes volumetric efficiency (Criss and Winston, 2008).

Values of the three components of the objective function in the calibration and evaluation periods (Fig. 2) suggested a good model performance and, therefore, provide a solid ground for assessing the performance of the selected BA methods in reproducing hydrological signatures.

2.4.3. Streamflow simulations

The calibrated HBV-light model was then used to simulate daily streamflow in each of the 50 catchments for the period 1963–2004 with three different sets of input:

1. Observed precipitation and temperature (the reference simulations),
2. Raw precipitation and temperature simulated by the 10 CMs, and
3. Bias-adjusted precipitation and temperature simulated by the 10 CMs and the four BA methods.

To reduce the effect of model uncertainty on the evaluation of BA methods, a set of 16 hydrological signatures (Table 2, see section 2.4.3) obtained from streamflows simulated with the observed precipitation and temperature was used as reference. Hydrological signatures

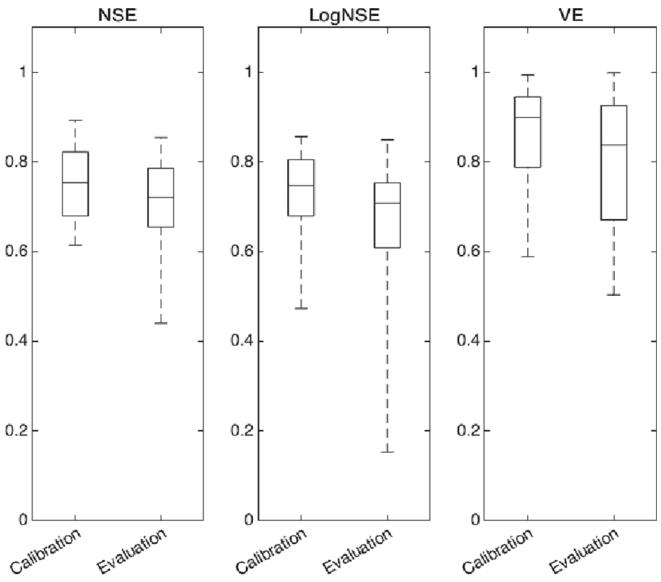


Fig. 2. Values of the three components of the objective function (NSE, logNSE and VE) in the calibration (1963–1982) and evaluation (1983–2004) periods. The ranges of the boxplots show the spread of the values across the 50 catchments within each period.

Table 2
List of 16 hydrological signatures used in this study. Signatures were categorized into 4 categories representing (a) water balance and flow dynamics (b) seasonal flows (c) low flows and (d) high flows. Within each category, the signature name is stated together with the abbreviation that is used throughout the paper, followed by the signature description (including equations) and the related references.

	Hydrological signatures (abbreviation)	Description	Reference	category
1	Mean annual flow, (*Q_mean*)	Average flow in a year (mm year^{-1}).		water balance and flow dynamics
2	Runoff coefficient, (Q_coeff)	Fraction of the total yearly precipitation that generates flow:$Q_{coeff} = \frac{\sum_i Q_i}{\sum_i P_i}$, Where Q_i represents the daily flow (in mm·day^{-1}) and P_i, daily precipitation, both of which were summed over a year.	(Merz and Blöschl, 2009)	
3	Timing of the center of mass of annual flow, (*COMt*)	Timing is computed from daily flows Q_i and for each year:$COMt = \frac{\sum_i Q_i t_i}{\sum_i Q_i}$where t_i represents ordinal day of a year.	(Kormos et al., 2016; Mendoza et al., 2016)	
4	Spring onset (spring "pulse day"), (*SPD*)	Spring onset is the ordinal number of the day in which the negative difference between the streamflow mass curve and the mean streamflow mass curve is the greatest. Spring onset series is obtained from values in each year	(Cunderlik and Ouarda, 2009)	
5	Mean spring flow, (*Q_spring*)	Flows in the spring (1st March through 31st May) in each year.	(Chen et al., 2017)	seasonal behavior of the flows
6	Mean summer flow, (*Q_summer*)	Flows in the summer (1st June through 31st August) in each year.		
7	Mean autumn flow, (*Q_autumn*)	Flows in the autumn (1st September through 30st November) in each year.		
8	Mean winter flow, (Q_winter)	Flows in the winter (1st December through 28th February) in each year.		
9	Low-flow frequency, *(LFfreq)*	Number of days in a year with flows smaller than 20% of the mean observed flow in the complete simulation period (1963–2004).	(Addor et al., 2018)	low flow characteristics
10	Timing of 30-day low flow, (T_minQ30)	Ordinal day in which 15th day of the 30-day minimum annual flow occurred, obtained in each year. If there were several consecutive days with the same minimum flows, the mean timing of these days in a year is adopted.	(Parajka et al., 2016; Vis et al., 2015)	
11	7-day low flow, (minQ_d7)	Minimum flows averaged over a given number of consecutive days (here, 7 and 30) obtained in each year.	(Olden and Poff, 2003; Richter et al., 1996)	
12	30-day low flow, (minQ_d30)			
13	High flow frequency, (*HFfreq*)	Number of days in a year with flows larger than 80% in the full simulation period (1963–2004) computed from the observed flows.	Following Krysanova et al. (2017) and Westerberg & McMillan (2015)	high flow characteristics
14	Timing of the 1-day high flow, (T_maxQ1)	Ordinal day in which the maximum annual flow occurred, obtained in each year of the simulation period (1963–2004).	(Richter et al., 1996)	
15	30-day high flow, (maxQ_d30)	Maximum flows averaged over a given number of days (here, 30 and 1) obtained in each year.	(Dankers et al., 2014; Vis et al., 2015)	
16	1-day high flow, (maxQ_d1)			

obtained from streamflows simulated with raw or bias-adjusted CM precipitation and temperature (i.e., inputs 2 and 3) were, thus, evaluated against this reference rather than against signatures of streamflow observations. This evaluation protocol eliminates the uncertainties caused by hydrological model structure and/or parameter estimates, and is in line with the recommendations by Hakala et al. (2019) for the assessment of uncertainty stemming from the BA methods.

2.5. Hydrological signatures

The 16 signatures were selected to capture complementary characteristics of streamflow regimes, which relate to (a) water balance and flow dynamics, (b) seasonal flows, (c) low flows, and (d) high flows (Table 2). Considering that the selected signatures generally reflect annual streamflow behavior (e.g., spring flows, mean annual flows or timing of the spring pulse day), they were firstly computed from daily streamflows separately for each year, and then averaged over the entire

period between 1963 and 2004. Such protocol, which is adopted in many hydroclimatic studies, suggests evaluate of BA methods' performance over long periods (François et al., 2020; Maraun and Widmann, 2018). Focusing on the entire 42-year period was recognized as the most robust split-sample strategy (Maraun and Widmann, 2018; Shen et al., 2022), and, as such, was also applied in numerous previous studies (e.g., Meyer et al., 2019; Zscheischler et al., 2019).

2.6. Evaluation framework for the comparison of bias correction methods

We compared the performance of four BA methods (section 2.3) in reproducing 16 hydrological signatures (section 2.5) obtained from simulated streamflows with the outputs of 10 different CMs (section 2.2.2) and observed climate series in the 50 Swedish catchments (section 2.1). The evaluation was based on absolute errors (MAE), computed as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|R_i - R_{ref}\right| \quad (4)$$

where R_i denotes a signature computed from the outputs of the HBV-light model forced with the raw and four bias-adjusted outputs of the climate modelling chains separately. Conversely, R_{ref} denotes a signature obtained with the reference model (i.e., model forced with observed precipitation and temperature).

We then addressed the research questions in a framework consisting of (1) performance, (2) robustness and (3) a comparison-based evaluation of BA methods, which guided the practical recommendations in the final section of this paper.

The ***performance of BA methods*** was visually analyzed with the help of separate heatmaps for raw and the four bias-adjusted CM outputs, which allowed an evaluation of emerging patterns across catchments and CMs. The Pearson correlation coefficient (Pearson, 1920) based on 500 pairs of values (50 catchments × 10 CMs) for each possible combination of the 5 CM simulations (1 raw and 4 BA methods) for each of the 16 hydrological signatures (Table 2) was then used to detect ***similarities in the performance*** of the applied BA methods across catchments and CMs. In this study, we considered BA-method pairs with correlation coefficients higher than 0.7 similar.

The overall ***robustness of BA methods*** was evaluated based on the concepts of accuracy and precision that have often been applied to measurements (McMillan et al., 2018) or model predictions (Wagener et al., 2022), and are here applied to the bias-adjusted model simulations. Accuracy here refers to the ability of a bias-adjusted model to yield simulations close to the observations (i.e., have a low bias), and it was quantified by averaging MAEs across the 50 catchments and 10 CMs (500 instances), and also by the frequency of improvements (i.e., how many times a BA method reduced the raw bias). Precision refers to the low variability of bias-adjusted simulations provided by multiple CMs (i.e., have a low variance). Thus, the variance across CMs and across catchments was computed under the assumption that robust BA methods perform consistently well (i.e., have a low variability) regardless of the climate models used, or catchment properties. To further check the consistency across hydroclimatic regimes, the robustness was analyzed separately for the snowmelt-driven regimes in northern regions (i.e., within ET and Dfb climate zones), and for rainfall-driven regimes in southern regions (i.e., Dfc climate zone). The differences in runoff regimes between these two groups of catchments are clearly shown in Fig. 1a: the snowmelt-driven catchments were characterized by a pronounced runoff seasonality, with the highest runoff in spring and early summer, as opposed to rainfall-dominated catchments that exhibited rather subtle runoff variation throughout the year.

The final ***comparison-based evaluation*** was based on the following eight criteria, which described each BA method's ability to:

- *strongly* reduce biases, i.e., to yield low MAE values **(criterion 1)**
- *frequently* reduce biases, i.e. to have a high frequency of improvement **(criterion 2)**
- *robustly* reduce biases *across various CMs* that had different boundary conditions and process representations, i.e., to yield a low variance in MAE values across CMs **(criterion 3)**
- *robustly* reduce biases *across* a multitude of *catchments* with various catchment characteristics and hydroclimatic properties, i.e., to yield a low variance in MAE values across catchments **(criterion 4)**
- *strongly* reduce biases *across hydroclimatic regimes dominated by snowfall and melt*, i.e., to have low MAE values in northern catchments **(criterion 5)**
- *frequently* reduce biases *across hydroclimatic regimes dominated by snowfall and melt*, i.e., to have a high frequency of improvements in the northern catchments **(criterion 6)**
- *strongly* reduce biases *across hydroclimatic regimes dominated by rainfall*, i.e., to have low MAE values in southern catchments **(criterion 7)**
- *frequently* reduce biases *across hydroclimatic regimes dominated by rainfall*, i.e., to have a high frequency of improvements in the southern catchments **(criterion 8)**

For each of these criteria, the performance of every BA method was categorized as high (≥75% of max–min range), medium (<75% and greater than 25%), or low (≤25%) to reveal if particular BA methods consistently outperformed the remaining ones in simulating signatures relevant for water resources management under changing climate.

3. Results

3.1. Performance of BA methods

3.1.1. Performance across the signature groups

3.1.1.1. Water balance and runoff dynamics signatures. Runoff signatures simulated with raw CMs typically featured considerable biases (expressed as MAE). For annual mean flow (Q_mean), raw biases were as high as 2 mm·day^{-1} (Fig. 3a, left panel), nearly three times the reference value, i.e., mean flow simulated with the observed meteorological input. The runoff coefficient (Q_coeff) demonstrated substantial biases of up to 0.3 (Fig. 3b). Signatures characterizing streamflow dynamics also showed considerable biases in raw simulations: the center of mass (COMt) was biased up to 35 days, or roughly one month (Fig. 3c), and biases in the timing of the spring flood (SPD) reached up to 200 days, i.e., more than 6 months (Fig. 3d). Considerably stronger biases in SPD were detected in southern catchments with latitudes below 60°N (Fig. 3d).

All BA methods were able to reduce the initial biases in these signatures obtained from raw CM outputs (Fig. 3a-d). Main differences were visible between the two groups of distribution-based (DS and copula) and distribution-free methods (QDM and MBCn), while the differences between uni- (DS and QDM) and multivariate methods (copula an MBCn) were less pronounced. Biases in raw Q_mean were reduced from 2 mm·day^{-1} down to a maximum of 0.2 mm·day^{-1} with distribution-based methods (DS and copula), and down to a maximum of 0.4 mm·day^{-1} with distribution-free approaches, i.e., QDM and MBCn (Fig. 3a). Biases in raw Q_coeff were considerably reduced after application of all the methods (Fig. 3b). The distribution-based methods reduced MAEs to maximum of 0.03, and distribution-free methods up to 0.06 (Fig. 3b). For COMt, all four BA methods performed similarly, resulting in biases between 13 days with DS, and 16 days with copula (Fig. 3c). Biases in SPD after applying the distribution-based methods were noticeably lower (approximately up to 70 days) than after applying the distribution-free methods that resulted in biases up to approximately 120 days (Fig. 3d).

Within both groups of distribution-based and distribution-free

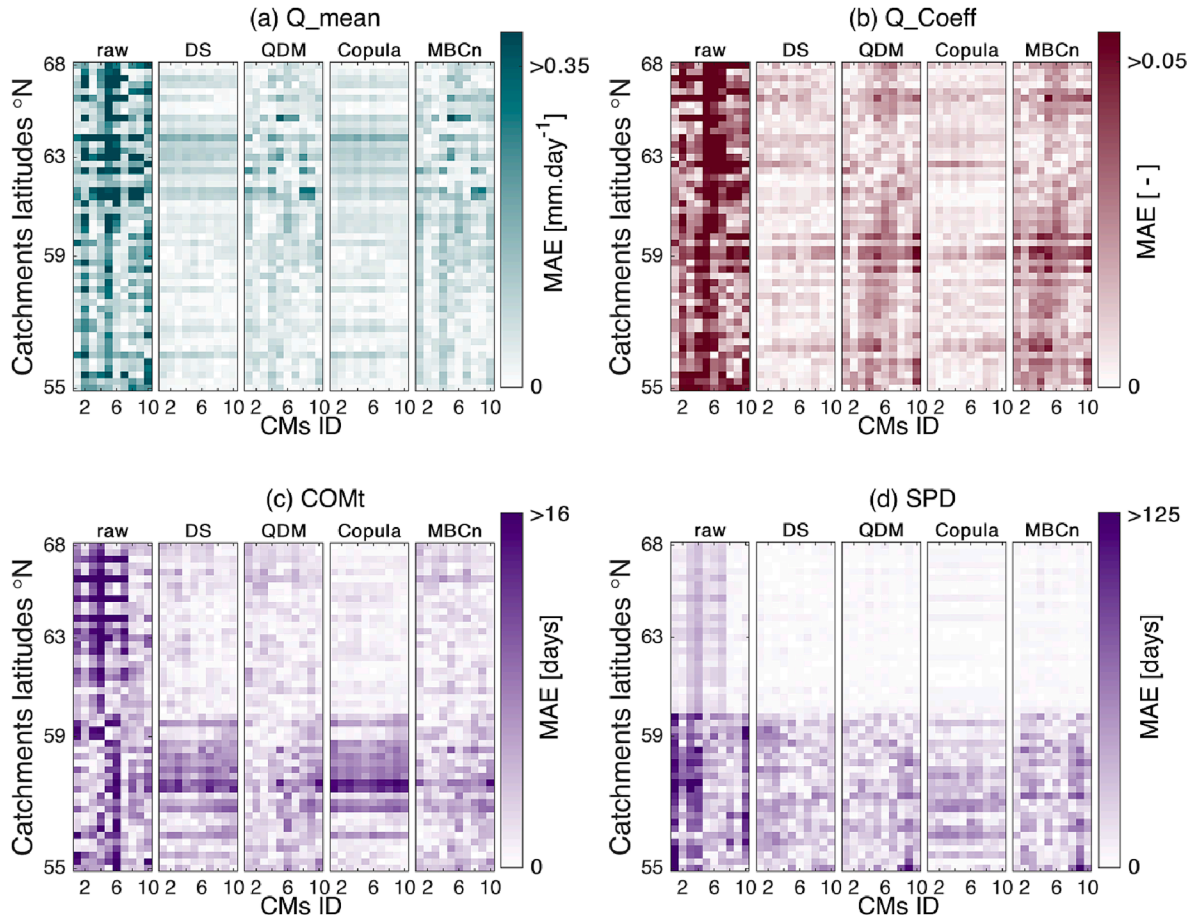


Fig. 3. MAEs for signatures representing water balance (a-b) and runoff dynamics (c-d) obtained from the flows simulated with the HBV-light model with uncorrected (i.e., raw) CM outputs (left panel) and with bias adjusted variables using four BA methods (the other four panels) over 1963–2004 period. Latitudes of the catchments are indicated on the panels' y-axis, while the x-axis represents CM-IDs (Table 1). Note that the colorbar range corresponds to the min–max ranges of MAEs in the signatures obtained from the simulations with the four BA methods (not including raw).

methods, similar patterns emerged in the remaining biases in all signatures of the group (Fig. 3). The exception in this regard was SPD (Fig. 3d), for which the patterns obtained with copula diverged from those of the remaining methods. Generally, patterns obtained with all BA methods differed from those obtained with the raw CM-outputs in all cases, with exception of SPD, which was characterized by pronounced change with the latitude. In contrast to the variation across latitudes, i.e., catchments, the variation in biases across CMs generally did not exhibit any apparent pattern, i.e., no CM consistently resulted in lower biases across all signatures of the group obtained and with all BA methods.

3.1.1.2. Seasonal flows signatures. Seasonal flows obtained with uncorrected CM outputs featured biases of up to 2.3 mm·day^{-1} in Q_spring (Fig. 4a) and up to 6 mm·day^{-1} in Q_summer (Fig. 4b). The MAEs were, however, lower for the other two seasons, only reaching as high as 1.6 mm·day^{-1} for Q_autumn (Fig. 4c) and 1.1 mm·day^{-1} for Q_winter (Fig. 4d). In all four cases, raw biases were substantial and at least twice the reference values (in case of Q_autumn even 7 times).

Biases in seasonal flows also featured a distinct spatial pattern, with much higher MAE values in the North in the spring and summer (Fig. 4a, b), whereas the opposite behavior was observed in the winter (Fig. 4d).

The distribution-free methods (QDM and MBCn) reduced biases in Q_spring (Fig. 4a) to a maximum of 0.7 mm·day^{-1}, while the distribution-based methods resulted in biases of up to 0.55 mm·day^{-1} and 0.35 mm·day^{-1} with DS and copula, respectively (Fig. 4a). For Q_summer (Fig. 4b), univariate BA methods resulted in MAEs of up to 1.8 mm·day^{-1}, while simulations based on the multivariate BA methods featured smaller errors (up to 1.3 mm·day^{-1}). Distribution-based methods performed similarly for Q_autumn (maximum MAE of 0.83 mm·day^{-1}) (Fig. 4c), and generally better than distribution-free approaches, which performed similarly taking values up to 1.3 mm·day^{-1}. Through simulations with all four methods, bias ranges in Q_winter (Fig. 4d) were somewhat similar, and took values of up to 0.45 mm·day^{-1} with distribution-based methods, and up to 0.55 mm·day^{-1} with the distribution-free ones.

In all signatures of this group (Fig. 4a-d), the two distribution-free methods exhibited similar MAE patterns. Furthermore, similarities in the MAE patterns could also be observed in the univariate methods (DS and QDM) for Q_summer (Fig. 4b) and Q_winter (Fig. 4d). The copula method resulted in a completely different MAE pattern in Q_winter (Fig. 4d), particularly in the southern catchments. Additionally, Q_summer and Q_winter exhibited a distinct spatial pattern, i.e., MAEs in Q_summer were notably higher in the northern catchments, while Q_winter showed the opposite pattern. The patterns obtained with the raw CM outputs were generally altered by applying the BA methods, although a slight resemblance in MAE patterns could be detected for the summer and winter flows. Visual inspection of the panels with the signatures obtained with bias-adjusted climate variables did not identify any CM that consistently yielded lower biases across the BA methods and signatures.

3.1.1.3. Low flows signatures. Biases in the low-flow frequency (LFfreq, Fig. 5a) and timing of low flows (T_minQ30, Fig. 5b) were generally high

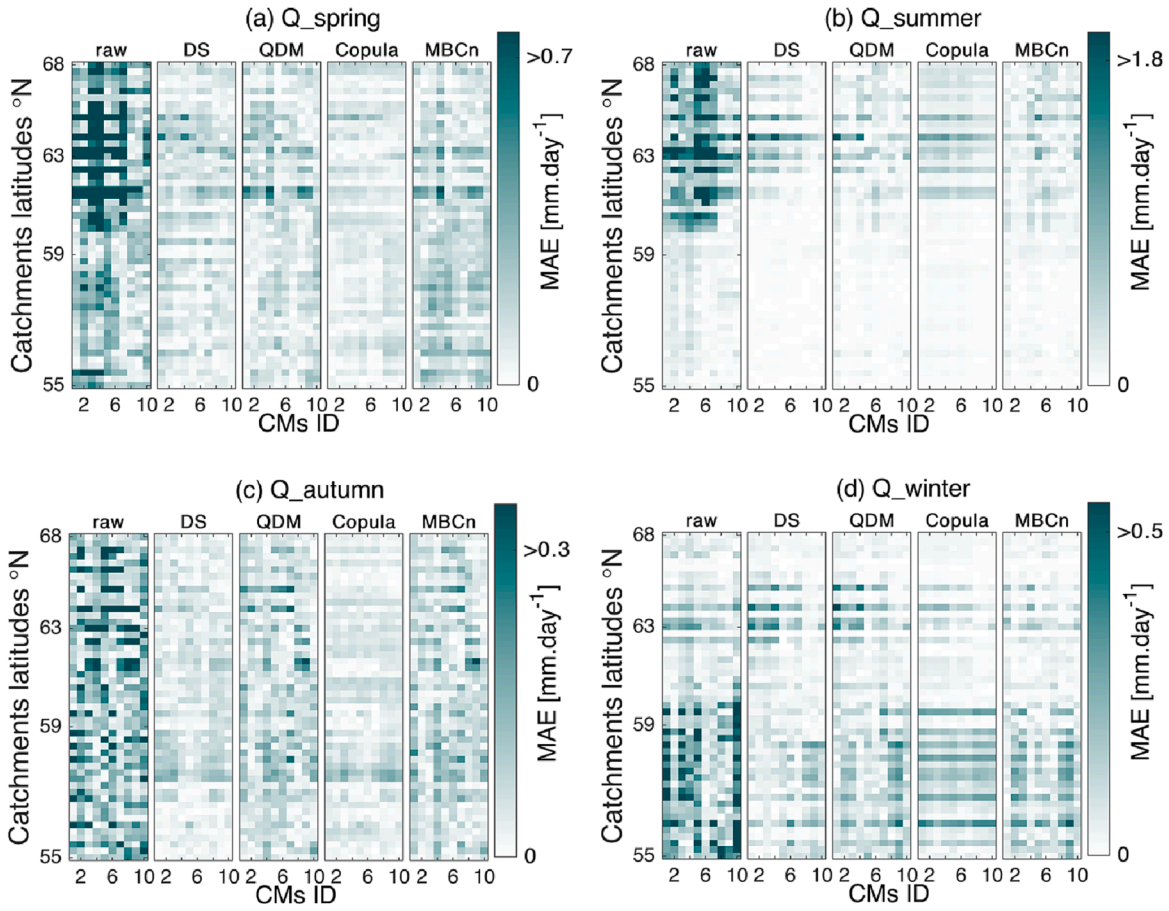


Fig. 4. MAEs for signatures representing seasonal flows (a-d) obtained from the flows simulated with the HBV-light model with uncorrected (i.e., raw) CM outputs (left panel) and simulations with bias adjusted variables using four BA methods (the other four panels) over 1963–2004 period. Latitudes of the catchments are indicated on the panels' y-axis, while the x-axis represents CM-IDs. Note that the colorbar range corresponds to the min–max ranges of MAEs for the simulations with the four BA methods (not including raw).

(up to 100 days) for simulations with uncorrected CM outputs. The other two signatures representing the magnitude of low flows featured biases as high as 0.22 mm·day^{-1} in minQ_d7 (Fig. 5c), and up to 0.25 mm·day^{-1} in minQ_d30 (Fig. 5d), both of which exceeded the reference values by more than three times.

Biases in all low-flow signatures with bias-adjusted CM simulations were considerably cut down. The MAE values in LFfreq (Fig. 5a) were similar across all four BA methods, ranging between 40 days after applying the distribution-based methods, and 45 days after the application of the distribution-free methods. The distribution-based methods resulted in lower ranges of MAEs in T_minQ30 (up to 40 days with DS and up to 55 days with copula, Fig. 5b) compared to distribution-free methods, which showed errors of up to 80 days after applying QDM, and up to 100 days after applying MBCn. For minQ_d7 (Fig. 5c) and minQ_d30 (Fig. 5d), DS, QDM and MBCn reduced the original raw biases to a similar extent, resulting in MAE values of up to 0.16 mm·day^{-1} and 0.18 mm·day^{-1}, respectively. However, the application of the copula method resulted in considerably lower biases of only up to 0.8 mm·day^{-1} in minQ_d7 (Fig. 5c), and of 0.1 mm·day^{-1} in minQ_d30 (Fig. 5d).

While similarities in the MAE patterns could be observed between distribution-free methods (Fig. 5a-d), the copula approach again stood out by featuring considerably different patterns in T_minQ30 (Fig. 5b). MAEs in none of the signatures obtained with bias-adjusted CM outputs exhibited any apparent pattern across CMs or latitudes, except for copula for T_minQ30 that yielded slightly higher MAEs in the South. Such behavior was consistent with the signatures obtained with raw outputs, which did not display any distinct north–south pattern. Nevertheless, close similarities between the patterns obtained from raw and adjusted CM outputs could not be detected in these signatures.

3.1.1.4. High flows signatures. High-flow signatures based on the uncorrected CM outputs featured considerable biases (Fig. 6a-d). Biases in the frequency-related signature HFfreq (Fig. 6a) and timing-related signature T_maxQ1 (Fig. 6b) were up to 100 days, particularly in southern catchments in the latter. MAEs in maxQ_d30 (Fig. 6c) and maxQ_d1 (Fig. 6d) featured higher MAEs in the northern catchments of up to 8 mm·day^{-1} respective 12 mm·day^{-1}, both of which amounted to over two times the references values.

After bias adjustment, biases in HFfreq (Fig. 6a) were considerably reduced with all four methods, taking values of up to 20 days with the distribution-based methods (DS and copula), and up to 35 days for the distribution-free methods (QDM and MBCn). A noteworthy pattern emerged: while all four BA methods were able to considerably reduce biases in T_maxQ1 in the northern catchments, the distribution-based methods struggled in the southern catchments and basically did not result in a distinct reduction of MAEs (Fig. 6b). On the contrary, after the application of DS, several catchments featured biases in T_maxQ1 as large as 100 days, which exceeded the original raw biases. This problem was even more pronounced for the copula method, which to a large extent inflated the raw biases in the southern catchments (Fig. 6b). For maxQ_d30 (Fig. 6c), all methods performed similarly, with copula featuring the lowest biases (up to 2 mm·day^{-1}), and QDM the highest biases (up to 3.5 mm·day^{-1}). Biases in maxQ_d1 (Fig. 6d) featured similar ranges for DS, copula and MBCn (up to 5.5 mm·day^{-1}), while

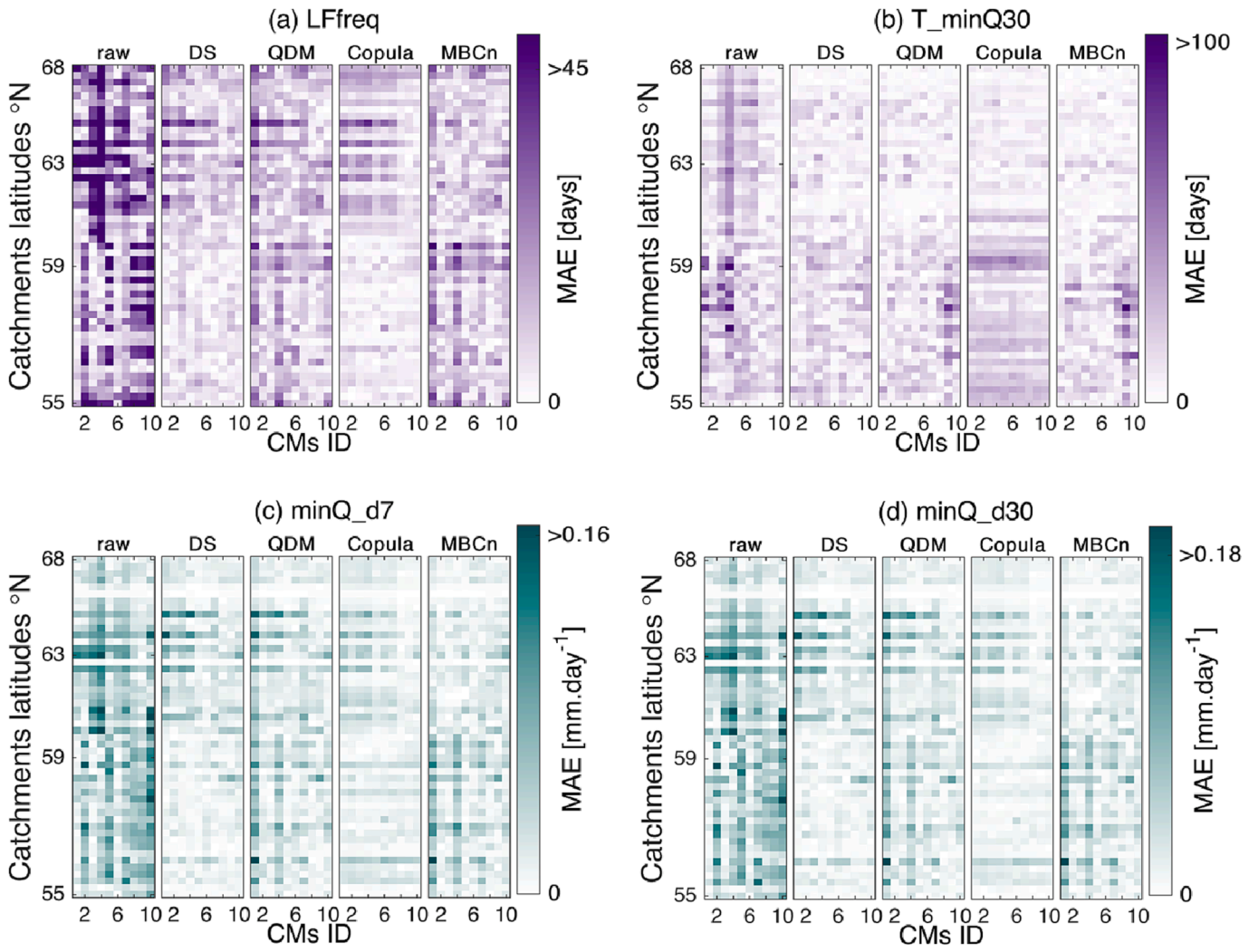


Fig. 5. MAEs for signatures representing low flows (a-d) obtained from the flows simulated with the HBV-light model with uncorrected (i.e., raw) CM outputs (left panel) and simulations with bias adjusted variables using four BA methods (the other four panels) over 1963–2004 period. Latitudes of the catchments are indicated on the panels' y-axis, while the x-axis represents CM-IDs. Note that the colorbar range corresponds to the min–max ranges of MAEs for the simulations with the four BA methods (not including raw).

MAEs with QDM were somewhat larger, exhibiting values of up to 8 mm·day^{-1} (Fig. 6d).

For all high-flow signatures, similarities could be observed in patterns obtained with the distribution-free methods (QDM and MBCn). Note that the patterns in MAEs after applying the copula approach were again somewhat different from the other methods (Fig. 6a-d). Most evidently, we could observe horizontal lines in Fig. 6a-d, which indicated rather low variation across the CMs. The MAE values in HFfreq (Fig. 6a) did not feature a distinct north–south gradient; however, spatial variations across latitudes were apparent in the other three high-flow signatures. For example, biases in T_maxQ1 (Fig. 6b) were mostly higher in the southern catchments, while the opposite behavior was exhibited by maxQ_d30 and maxQ_d1. As for the latter, larger biases in the northern catchments were also obtained with the raw CM outputs. No CMs could be identified as the most robust in the ensemble, since no CM consistently yielded the lowest residual biases.

4. Similarity of BA patterns

Similarities between the patterns in the signatures were detected in many instances, as outlined in section 3.1.1. First, we analyzed to what degree BA methods maintained the patterns obtained with raw CM data (i.e., correlations between raw and each of the BA methods). Thereafter, we examined the correlations between all pairwise combinations of the four BA methods. We considered the Pearson correlation coefficients greater than 0.70 to be an indication of a strong correlation.

Overall, correlations between the signatures obtained from adjusted CM outputs and those obtained with raw ones were weak, as indicated by rather low values of the Pearson correlation coefficients. Signatures that had the lowest correlations to MAE obtained from raw simulations included COMt (Fig. 7a, third panel), T_minQ30 (Fig. 7c, second panel) and HFfreq (Fig. 7d, first panel), with the copula method showing the lowest correlation with uncorrected CMs in these signatures. The highest correlation coefficients between MAEs in raw and the bias-adjusted values were detected in Q_mean, Q_summer, maxQ_d30 and maxQ_d1. In all cases, except for Q_mean, the multivariate methods (copula and MBCn) featured stronger correlations with raw CM than the univariate ones (DS and QDM).

As opposed to correlations to the signatures obtained from raw CM outputs, some rather strong correlations between signatures obtained with BA methods could be detected. More specifically, strong correlations were detected in three pairs of BA methods: (1) most frequently between distribution-free methods QDM and MBCn (for all 16 signatures), (2) between distribution-based methods (DS and copula) for seven signatures, and (3) between the univariate methods (DS and QDM) for four signatures. The correlation between the two multivariate

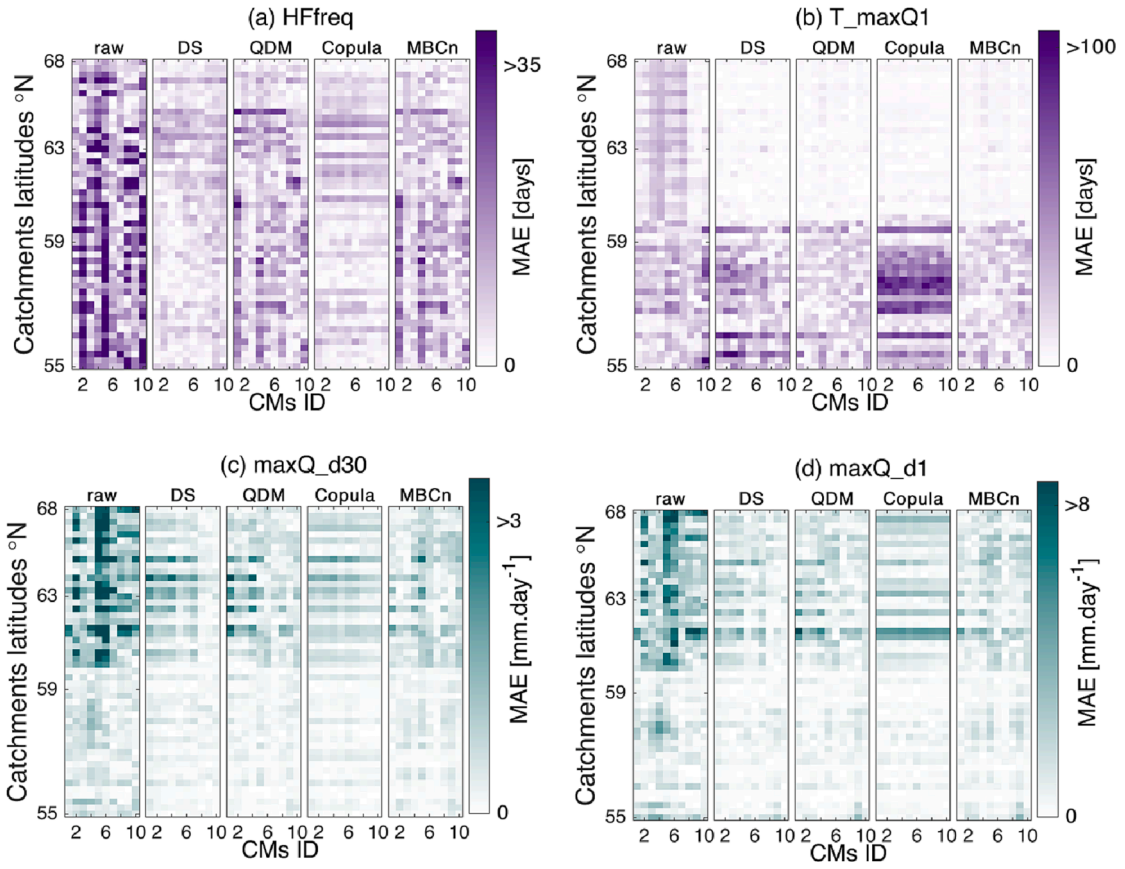


Fig. 6. MAEs for signatures representing high flows (a-d) obtained from the flows simulated with the HBV-light model with uncorrected (i.e., raw) CM outputs (left panel) and simulations with bias adjusted variables using four BA methods (the other four panels) over 1963–2004 period. Latitudes of the catchments are indicated on the panels' y-axis, while the x-axis represents CM-IDs. Note that the colorbar range corresponds to the min–max ranges of MAEs for the simulations with the four BA methods (not including raw).

methods was generally lower than for the other couples (i.e., in the range 0.03 – 0.63).

4.1. Robustness of BA methods

4.1.1. Overall robustness

All BA methods were able to consistently reduce the initial raw biases, which was demonstrated by considerably smaller MAE values (averaged over all 50 catchments and 10 CMs) compared to the simulations based on raw CM data (indicated by smaller circles in Fig. 8). Biases in the signatures related to water balance and flow dynamics (Fig. 8a), as well as in the seasonal flows (Fig. 8b), were generally reduced, i.e., the application of the BA methods resulted in average bias reductions of more than 50% of the initial bias. Biases in signatures related to low- (Fig. 8c) and high flows (Fig. 8d) were reduced less well, which was particularly pronounced in e.g., timings of 30-day low flows (T_minQ30, Fig. 8c) and of the annual maximum flows, T_maxQ1 (Fig. 8d).

To examine whether the BA methods reduced the biases not just in a few instances, but rather consistently across all 500 simulations (50 catchments × 10 CMs), we also analyzed the frequency of improvements. These results generally corroborated the MAE values. Water balance and flow dynamics signatures (Fig. 8a) as well as seasonal flows (Fig. 8b), had somewhat higher rates of bias reductions (on average, improvements in 81% and 78% of simulations, respectively), compared to the other signatures. Application of the BA methods led to improvements in 74% of the simulated low flow signatures (Fig. 8c), and 75% of high flow signatures (Fig. 8d). The lowest frequency of improvements was found in T_minQ30 and T_maxQ1 (both 57%), while the highest frequency was achieved in summer flows, Q_summer (93%, Fig. 8b), and in the runoff coefficients, Q_Coeff (92%, Fig. 8a).

In most cases, the best results were obtained with the distribution-based methods (DS and copula), which resulted in lower MAE values than their distribution-free counterparts, but also a higher frequency of improvements. The only two signatures for which the opposite pattern was obtained were T_minQ30 and T_maxQ1.

It is also noteworthy that uni- and multivariate methods within the same group of either distribution-based or distribution-free BA approaches generally resulted in rather similar values of averaged MAE and frequency of improvements, especially in the signatures related to water balance and flow dynamics (Fig. 8a) and the seasonal flows (Fig. 8b). The only exception in this regard was the copula method, which struggled to adjust biases in T_minQ30 (Fig. 8c), and even inflated existing biases in T_maxQ1 (Fig. 8d).

4.1.2. Consistency across climate models

The application of BA methods led to noticeably reduced MAE values (compared to raw) across all climate models for all 16 signatures, with exception of copula and T_maxQ1 (as elaborated in detail in the previous section). The reduction in biases is apparent not only in the median values, but also in the interquartile ranges of MAE values across the 10 CMs, which were consistently smaller than those of the raw simulations (Fig. 9).

Distribution-free methods generally showed higher variability of the errors (i.e., larger inconsistency) than distribution-based methods (Fig. 9, depicted by the circles on top of each boxplot representing the variance), with MBCn demonstrating the highest variance across CMs for half of the signatures and never featuring the lowest variance. This

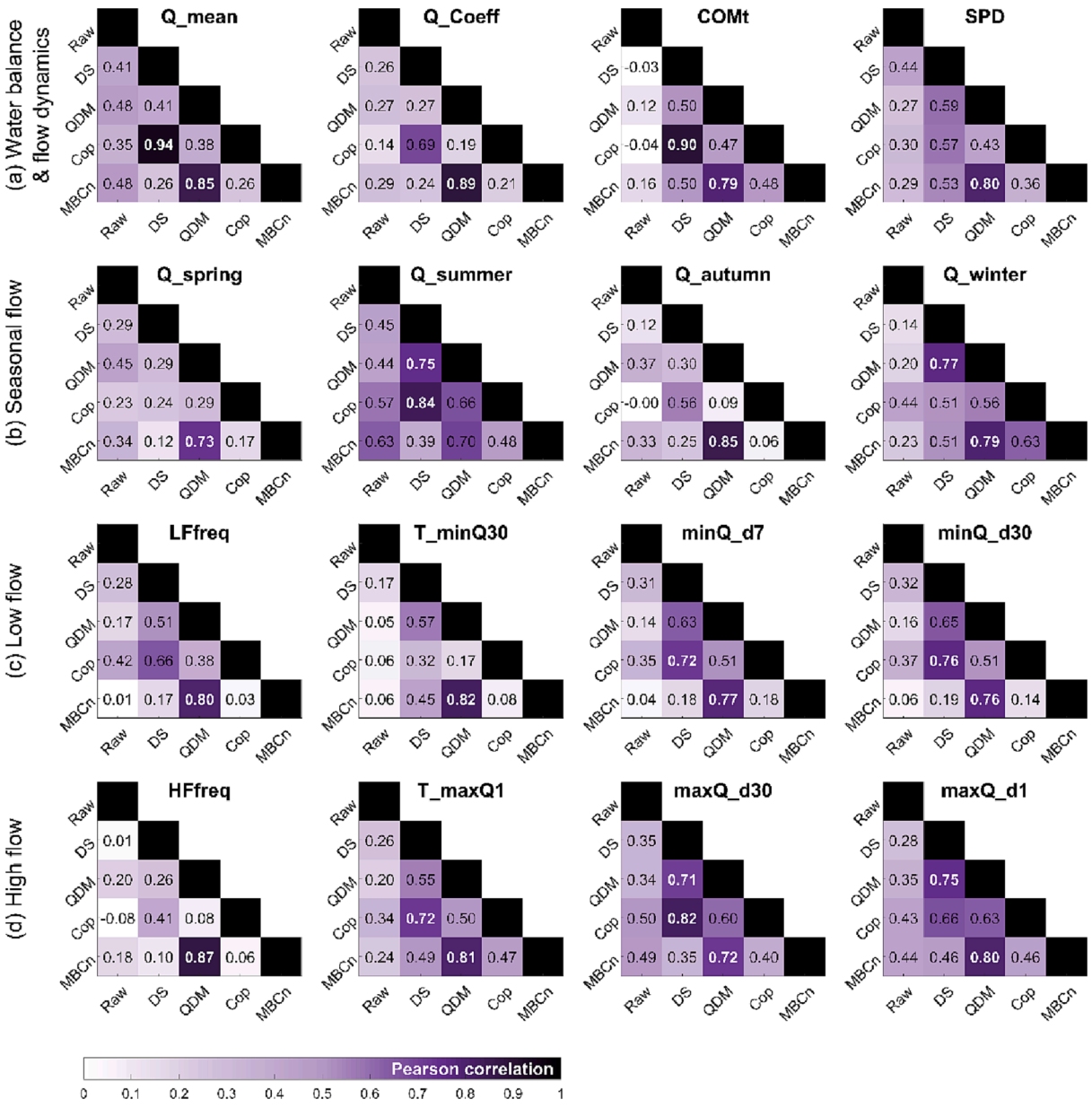


Fig. 7. Pearson correlation coefficients based on the 500 pairs of MAE values (50 catchments × 10 CMs) for each possible combination of the five simulations (one raw and four BA methods) for the 16 hydrological signatures. BA-method pairs with correlation coefficients higher than 0.7 are highlighted with darker shades of purple and white text color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

was closely followed by QDM, which featured the highest variance in five signatures and the lowest in one. In contrast, copula stood out by resulting in the lowest variance across CM (i.e., the highest consistency) for 14 of the 16 signatures, closely followed by DS (Fig. 9).

The largest variances (i.e., inconsistencies) were found in the low-flow signatures (Fig. 9c), for which the minimum–maximum ranges across the CMs were relatively large, and in particular for minQ_d7 and minQ_d30. Although these ranges were not excessive in the absolute terms (i.e., they took values up to 0.05 mm day^{-1}), they were nearly as large as for the raw simulations, especially with the QDM method.

4.1.3. *Consistency across catchments*

The performance of BA methods was generally more variable across catchments compared to their variance across CMs, which was reflected by a wider interquartile range and a larger variance (Fig. 10). However, the application of the BA methods still led to a visible reductions in the median MAE for all signatures, and in lower variability of MAE values across catchments, which was consistently smaller than raw values for the majority of signatures.

The distribution-based methods (DS and copula) featured generally a higher variability (i.e., larger inconsistencies), with copula having the largest variance in nine of the 16 signatures, and DS in four cases. Distribution-free methods were mostly associated with lower variance of MAEs, with QDM featuring only once and MBCn only twice the highest variance among the BA methods. However, differences across the BA methods in variances across catchments were less pronounced than in variances across CMs.

Variance across catchments was generally a little smaller for signatures representing the water balance and runoff dynamics (Fig. 10a) and seasonal streamflows (Fig. 10b) than for the signatures representing low- (Fig. 10c) and high flows (Fig. 10d). Inconsistencies were particularly large in T_minQ30, for which the raw variance was even inflated by the copula and MBCn methods (Fig. 10c, second panel), and in T_maxQ1, for which all methods except MBCn inflated the variance

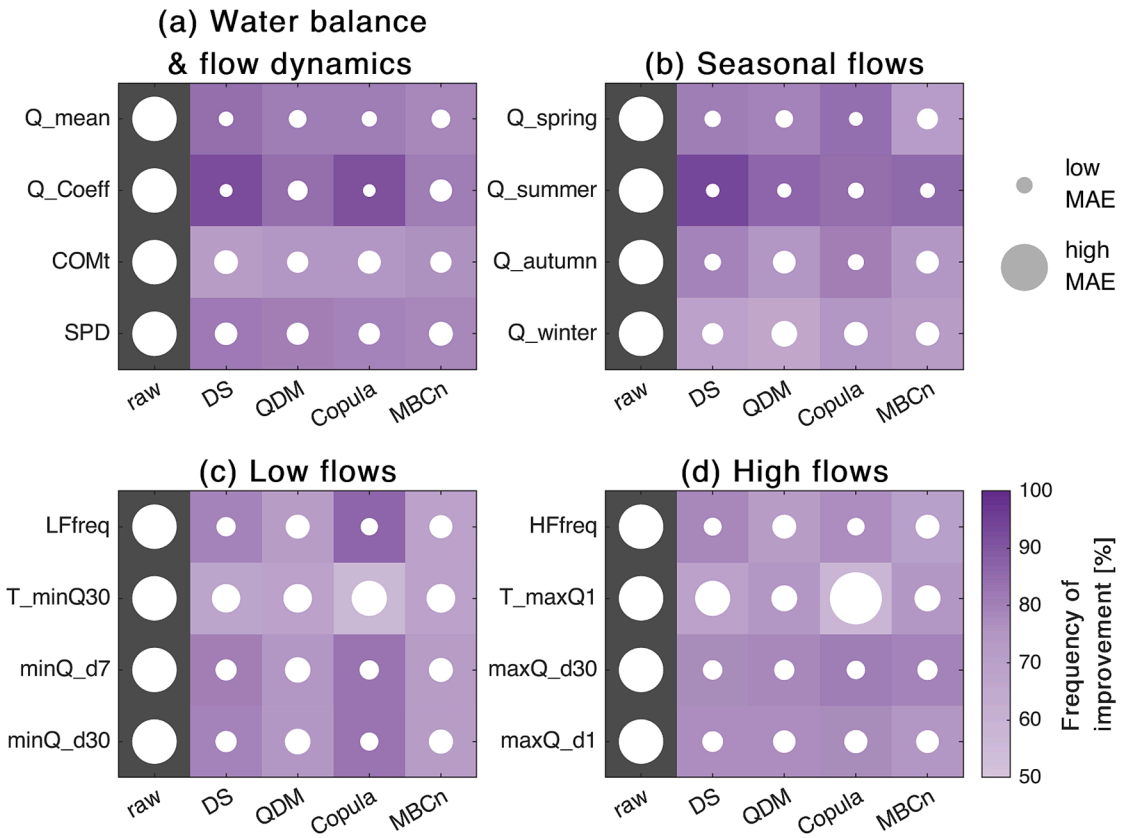


Fig. 8. Mean biases (MAE) and frequency of improvement in the four groups of hydrological signatures (a-d) obtained with the four BA methods relative to the raw CM outputs. Circles represent the MAE values (averaged over all CMs and catchments), where larger circles indicate higher MAE values, i.e., greater bias in the hydrological signature. Shadings of the cells indicate the frequency of improvement, i.e., the percentage of cases (with 10 CMs in 50 catchments) in which application of a BA method resulted in lower bias in a particular signature than the raw simulations. A dark background and small white circle indicate a robust BA method.

(Fig. 10d, second panel).

4.1.4. *Consistency across hydroclimatic regimes*

Given the variability in biases across catchments (Fig. 11), we further assessed whether systematic differences in performance of BA methods in the northern catchments with dominantly snowmelt-driven streamflow regimes (located above latitude 60°N), and southern, primarily rainfall-driven catchments (located below 60°N), could be detected.

In the northern catchments (Fig. 11a-d), initial raw biases were consistently reduced with all BA methods in all 16 signatures. Notable were the high remaining biases and relatively low frequency of improvements for Q_winter (Fig. 11b), particularly with the univariate methods. In contrast, Q_summer (Fig. 11b) and T_maxQ1 (Fig. 11d) only had small biases remaining, and featured a high rate of improvements. In signatures representing the water balance and runoff dynamics (Fig. 11a) in the northern catchments, distribution-based methods resulted in slightly lower MAEs than distribution-free methods. The same behavior was detected for Q_spring and Q_autumn (Fig. 11b). For Q_summer, however, univariate methods outperformed the multivariate ones, as opposed to Q_winter. Similarly, both multivariate BA methods reduced biases in low-flow signatures more than the univariate ones (Fig. 11c), except for T_minQ30, for which copula performed worse than the other three methods. Biases in high-flow signatures were reduced most by the distribution-based methods, with the exception of maxQ_d1, for which copula resulted in the largest remaining bias (Fig. 11d). Frequencies of improvement with the BA methods were largely consistent with the MAE values, i.e., lower remaining biases were accompanied with the highest rate of improvements.

The BA methods were generally able to reduce biases in the southern, mainly rainfall-driven catchments (Fig. 11e-h). In most cases, the distribution-based BA methods outperformed distribution-free ones. Both distribution-based methods performed similarly, i.e., DS and copula each performed best for half of the signatures. On the other hand, in the distribution-free methods, QDM outperformed MBCn in the majority of signatures (14 out of 16). For COMt (Fig. 11e), T_minQ30 (Fig. 11g) and T_maxQ1 (Fig. 11*h*), however, the opposite pattern emerged, i.e., the distribution-free methods resulted in lower MAEs and a higher frequency of improvements compared to distribution-based ones, especially copula. Especially the copula method showed a low performance with remaining biases nearly as high as raw biases (in COMt), or even higher (in T_minQ30 and T_maxQ1). Similarly to the northern catchments, frequencies of improvement with the BA methods were consistent with the reduction in biases. In other words, a higher frequency of improvement was typically connected to lower MAEs.

4.2. *A comparison-based evaluation*

Based on the eight robustness criteria (section 2.6) assessed in the previous sections, we here (Fig. 12) provide a comprehensive comparison of the BA methods aimed at identifying the most suitable method(s) for assessment of climate change impact on streamflow regime features represented by 16 signatures.

While application of each of the BA methods led to considerable improvements, their overall performance varied for each criterion across signatures. For example, distribution-based BA methods were generally more robust than the distribution-free methods for signatures related to water balance and flow dynamics (Fig. 12a), as well as for seasonal flows (Fig. 12b). In both groups of signatures, univariate DS frequently displayed a high performance across the criteria (in 20–21 out of 32 cases), and resulted in only 5 cases of low performance

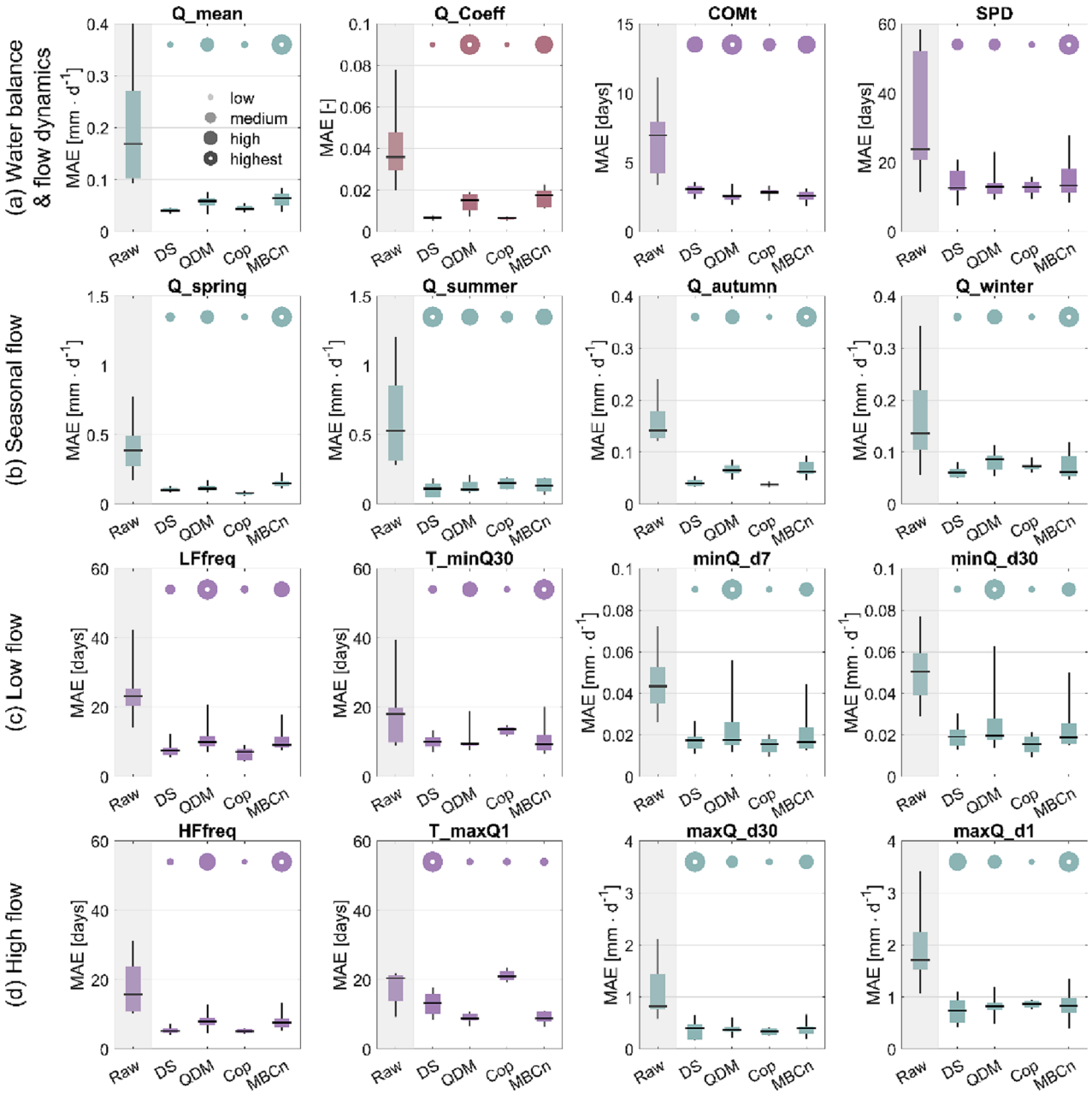


Fig. 9. Mean biases (MAE) and their variance across CMs in the four groups of hydrological signatures (a-d) obtained with the four BA methods relative to raw CM output (gray background). Boxplots represent the range of MAE values (averaged over all 50 catchments) across the 10 CMs, circles on top represent the variance, with smaller circles depicting lower variance. The BA method with the largest variance is highlighted as a reference with a hollow circle (normalized to have same size across signatures), the other BA methods are scaled accordingly.

(Fig. 12a,b, first panel), closely followed by the multivariate copula approach with 18–20 high-performance cases and 8–9 low-performance cases (Fig. 12a,b, third panel). The multivariate variant of distribution-free methods (MBCn), however, performed worse than its univariate counterpart (QDM), with MBCn showing a low performance in 23 (Fig. 12a, fourth panel) respective 22 (Fig. 12b, fourth panel) cases, whereas QDM ranked low only for 9 (Fig. 12a, third panel) and 14 (Fig. 12b, third panel) instances, respectively.

Remarkable was the robust performance of distribution-based methods in reproducing signatures related to low flows (Fig. 12c) and high flows (Fig. 12d). These methods resulted in many more instances of high performance (14 each for DS and 18–21 for copula) than the distribution-free methods (9–12 cases for MBCn and 7–8 for QDM). In the case of low flows (Fig. 12c), multivariate BA methods performed slightly better than their univariate counterparts (i.e., copula performed better than DS, and MBCn better than QDM). For the signatures representing high flows (Fig. 12d), the application of multivariate methods resulted in a larger number of high-performing instances compared to their univariate counterpart. However, the number of low-performing cases also increased, leaving fewer chances for medium performance across these methods. The two multivariate methods, namely, copula (Fig. 12d, third panel) and MBCn (Fig. 12d, fourth panel), had only 2 and 5 medium performing cases, while DS (Fig. 12d, first panel) and QDM (Fig. 12d, second panel) had as many as 11 and 15 of such cases, respectively.

Averaging the performance over the signatures within each signature group (Fig. 12e) revealed that the distribution-based methods yielded similar performances that were better than those of the distribution-free

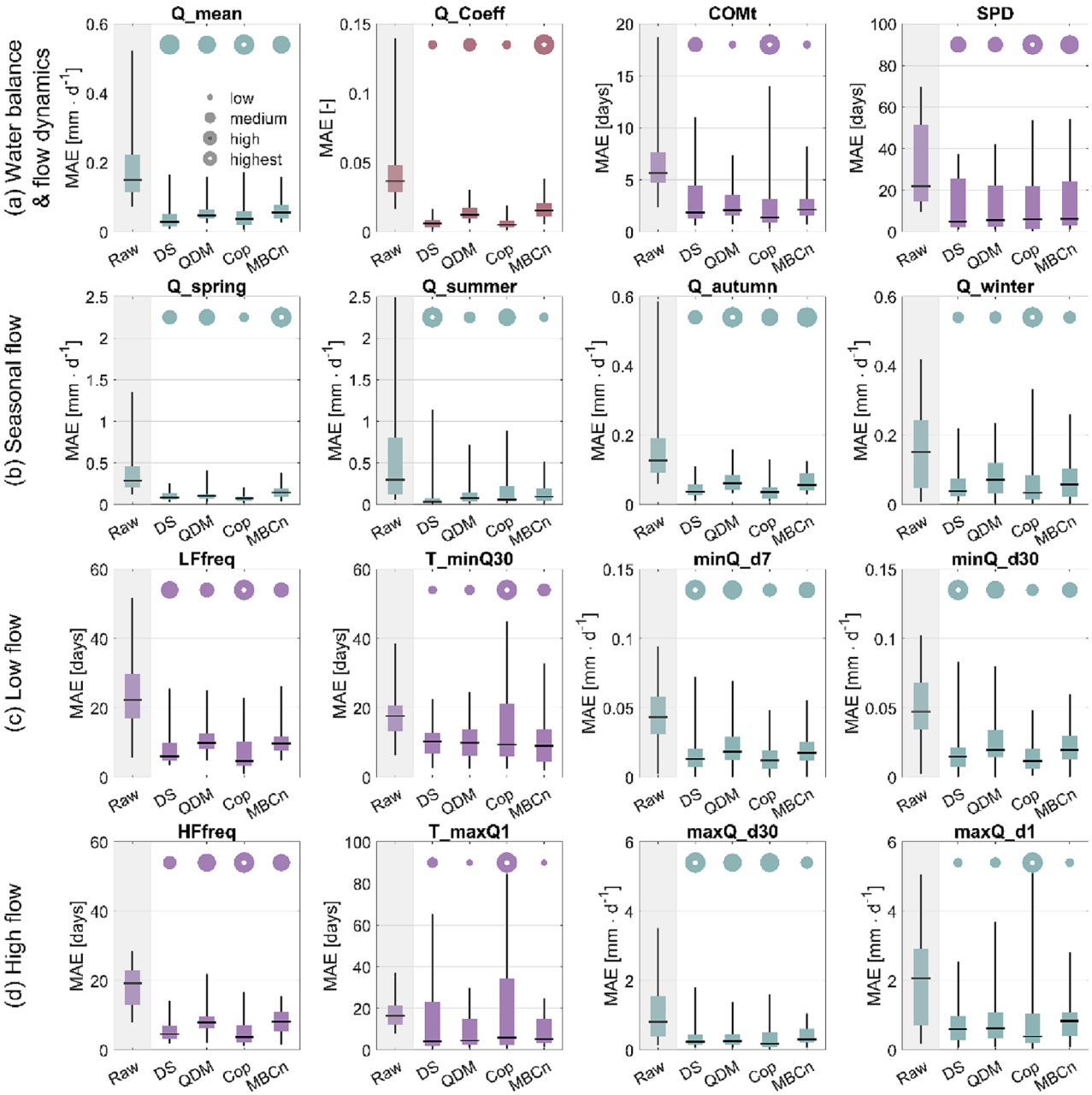


Fig. 10. Mean biases (MAE) and their variance across catchments in the four groups of hydrological signatures (a-d) obtained with the four BA methods relative to raw CM output (gray background). Boxplots represent the range of MAE values (averaged over all 10 CMs) across the 50 catchments, circles on top represent the variance, with smaller circles depicting lower variance. The BA method with the largest variance is highlighted as a reference with a hollow circle (normalized to have same size across signatures), the other BA methods are scaled accordingly.

methods, as indicated by their lower MAEs **(criterion 1)** and higher frequencies of improvements **(criterion 2).** While the distribution-based methods also showed higher consistency across CMs **(criterion 3),** they came with slightly lower consistency across catchments **(criterion 4)**. Generally, there were slight differences between univariate and multivariate methods, which were often overshadowed by the strong differences between distribution-free and distribution-based methods. Noticeable differences between uni- and multivariate methods only emerged for the snowmelt-driven catchments (located above 60°N), where advanced multivariate methods resulted in frequently better performance compared to their univariate counterparts **(criteria 5–6)**. Nonetheless, distribution-based methods resulted in better performance compared to distribution-free methods. In the rainfall-driven catchments (located below 60°N), the differences were again rather found between distribution-based and distribution-free methods **(criteria 7–8)**, where the former performed similar to each other, and better than the latter, with low performance of the advanced MBCn method across all four categories of signatures.

5. Discussion

5.1. Performance and similarity of BA methods

Our results indicated that in the studied Swedish catchments, the HBV-light model driven by uncorrected CM outputs (i.e., precipitation and temperature) resulted in substantial biases in all hydrological signatures representing (1) water balance and flow dynamics, (2) seasonal flows, (3) low flows and (4) high flows. The magnitude of these biases

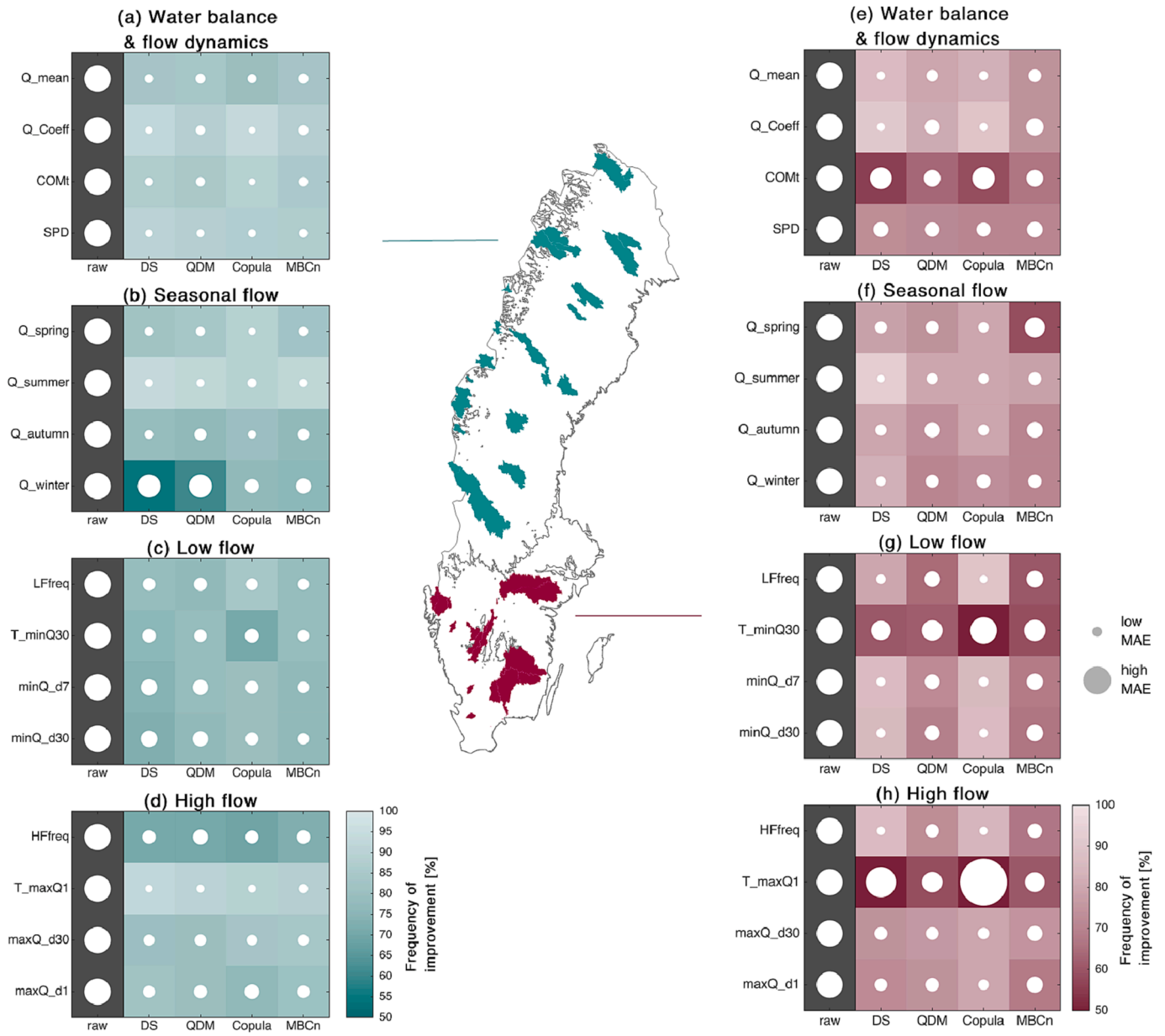


Fig. 11. Mean biases (MAE) and frequency of improvement in the four groups of hydrological signatures in snowmelt-driven catchments (a-d) and rainfall-driven catchments (e-h) obtained with the four BA methods relative to raw CM output. Circles represent the MAE values (averaged over all CMs and catchments), where larger circles indicate higher MAE values, i.e., greater bias in the hydrological signature. Shadings of the cells indicate the frequency of improvement, i.e., the percentage of cases (with 10 climate models in 50 catchments) in which application of a BA method resulted in lower bias in a particular signature than the raw simulations. A light-colored background and small white circle indicate a robust BA method.

varied across signatures, CMs, catchments and their hydroclimatic regimes. It is difficult to isolate signatures with the largest biases, since they all have different units. They take values that differ in orders of magnitude (e.g., signatures related to timings can take value up to 366, as opposed to runoff coefficient, which is <1), which also hinders comparisons across the signatures. Nevertheless, all signatures obtained with raw CM outputs have in common great raw biases that can exceed the reference values by several times, which brings informativeness and potential applicability of such signatures into question.

Variations in biases across the CMs did not exhibit any distinct behavior (i.e., no CM consistently resulted in the highest/lowest biases across signatures), but certain patterns across catchments, i.e., their hydroclimatic regions, could be detected. For example, seven signatures related to seasonal and extreme flows showed clear spatial patterns. Among them were the 1-day and 30-day high flows, as well as spring and summer flows, which demonstrated higher systematic errors in northern snowfall-dominated catchments located above 60°N in the ET or Dfb climate zones. Winter flows, timing of 1-day high flow and spring pulse day featured the opposite pattern by having larger biases in the southern catchments (below latitude 60°N in the Dfc climate zone). This behavior can be a consequence of different dominating hydrological regimes: specifically, regimes in northern Sweden are generally snowmelt-driven, resulting in a pronounced flow seasonality with low flows during winters due to snow accumulation, and a flood peak that typically occurs over springs and summers (Arheimer and Lindström, 2015; Matti et al., 2017; Teutschbein et al., 2022, Teutschbein et al., 2015). There are, thus, higher absolute flows in the north over spring and summer, which consequently result in larger biases (expressed in absolute terms in this study). In contrast, flows in southern Sweden are generally rainfall-driven (Matti et al., 2017) and are affected by higher evaporative

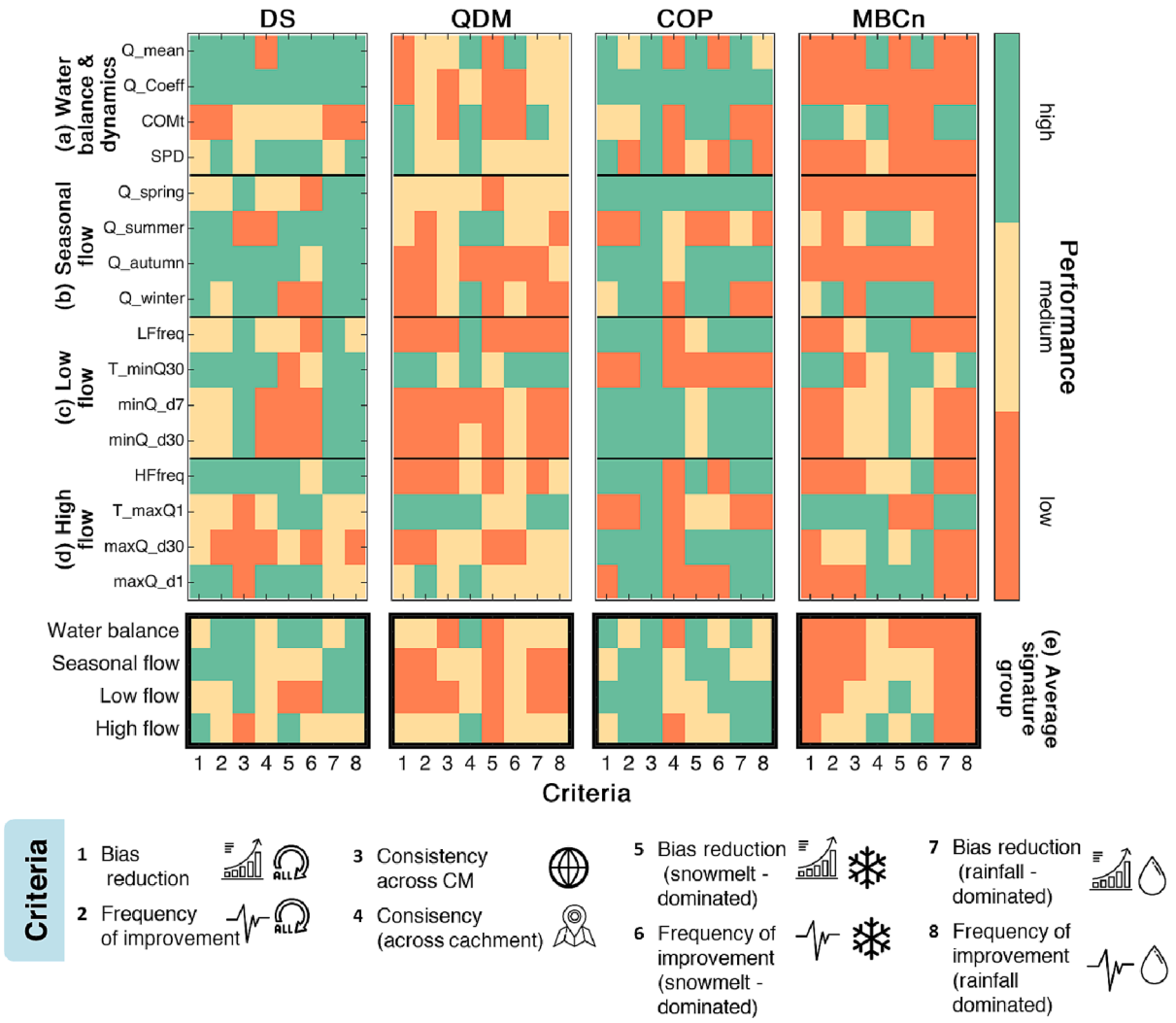


Fig. 12. Visual guidelines for selecting the most suitable BA methods for assessment of climate change impact on 16 hydrological signatures within four categories (a-d), derived from 8 evaluation criteria. The final row represents a holistic evaluation within each of these categories. Note that all BA methods generally resulted in improvement and the ranking here is only relates to the direct comparison of BA methods with each other. For each criterion, high and low performance is considered as the 75th and 25th percentile of values obtained from the four BA methods.

demands in the summer. As a consequence, there are generally lower flows in the summer and higher flows in winter in this region, implying also higher biases (in absolute terms) in winter flows, as demonstrated by our results (section 3.1.1.2). Larger biases in timings of annual maxima in these catchments can be attributed to spatiotemporal variability of extreme precipitation (Barlow et al., 2019) that affects the timing of rainfall-driven flow peaks, which also results in less pronounced flow seasonality (i.e., "flatter" intra-annual flow distribution) than in the northern catchments with snowmelt-driven regimes. Biases in the other nine signatures did not show distinct spatial patterns.

Generally, the detected biases in hydrological signatures obtained with the raw CM outputs can be attributed to simplified representations of, e.g., atmospheric convection (Lucas-Picher et al., 2021), aerosol behavior (Huang and Ding, 2021) or humidity (Bastin et al., 2019) in CMs, which are then translated into simulated streamflows through the impact modelling chain (Brunner and Slater, 2022; Chen et al., 2021b; Dobler et al., 2012; Teutschbein and Seibert, 2012). Biases in simulated streamflows or hydrolgic signatures stemming from these simplifications cannot be expected to get alleviated with the application of the latest generation of CMs (i.e., CMIP6), since there is no evidence for their superiority in representing different characteristics of climatic variables over the previous CM versions (Bourdeau-Goulet and Hassanzadeh, 2021; Seneviratne and Hauser, 2020).

Application of bias adjustment methods has been debated over the past decade (e.g., Chen et al., 2021b; Ehret et al., 2012). Some studies indicated that BA methods can potentially impair spatiotemporal characteristics of the CM outputs (Ehret et al., 2012), whereas some studies clearly demonstrated their merits in minimizing overall uncertainty associated with hydrological simulations (Muerth et al., 2013; Teutschbein and Seibert, 2013; Velázquez et al., 2013), including even field-scale agricultural impact studies (Glotter et al., 2014; Laux et al., 2021). This paper corroborates conclusions presented in both groups of studies and specifically shows that application of BA methods can considerably modify simulated signals by raw CMs. Our results also demonstrated that the application of BA methods in the vast majority of cases substantially decreases biases in the simulated hydrological signatures in the Scandinavian catchments, regardless of their hydroclimatic or geomorphological characteristics. Considering that CM-induced biases can adversely affect results of the hydrological impact studies and, hence, hinder identification of adequate climate change adaptation strategies, the results of our study support common practice, which implies bias adjustment of simulated climate variables to be an integral part of impact studies (Dobler et al., 2012; Hakala et al., 2019; Teutschbein and Seibert, 2012).

The four BA methods used in this study can be either categorized by the number of variables simultaneously being adjusted (i.e., uni- versus multivariate), or by their underlying approach to map the distributions/quantiles (i.e., distribution-free versus distribution-based). DS and copula both rely on the distribution-based DS method for adjustment of univariate (marginal) biases in precipitation and temperature, while QDM and MBCn share the distribution-free QDM method for adjustment of univariate characteristics. Therefore, by construction, the performance of each pair of methods was similar in reducing biases in univariate characteristics of precipitation and temperature, which are drivers of hydrological modeling (Tootoonchi et al., 2022a).

Our study revealed strong correlations in the remaining biases in simulated hydrological signatures obtained with the BA methods. The strongest and most frequent correlations in this study were obtained between QDM and MBCn. This demonstrates that the additional adjustment of the (quite weak) precipitation-temperature dependence, which is conducted with MBCn, does not affect the similarities between these two distribution-free methods, since MBCn does not rely on heavy stationarity assumptions for dependence or temporal sequencing of climate models (Cannon, 2016). Such strong correlations between QDM and MBCn could be explained by the fact that the distribution-free methods do not require identification of appropriate distribution functions at any step, as opposed to the DS and copula methods that exhibited strong correlations less frequently. This can be seen as a great flexibility in selecting a distribution-free BA method, which is relevant for future impact studies (Berg et al., 2022; Gudmundsson et al., 2012).

Strong correlations were also found between the residual biases of DS and copula in many signatures. The largest discrepancies between these two BA methods emerged in seasonal streamflows. These results can indicate that seasonality of simulated streamflows is affected by accuracy in bias adjustment of precipitation-temperature dependence, which is a multivariate feature that is better adjusted with copula than with univariate DS (Tootoonchi et al., 2022a).

Correlations of residual biases obtained with distribution-based and distribution-free BA methods to raw biases were largely similar but low. However, the weakest correlations in this regard were obtained with the copula method (e.g., winter flows or timing of annual minima, section 3.1.1). The copula method includes a shuffling algorithm (Vrac, 2018) that is typically stationary and based on the reference precipitation signal (François et al., 2020). Thus, the original signal of the raw CM simulations is completely lost for precipitation and temperature, which further reflects on simulated streamflow and hydrological signatures. This might also explain some of the differences between DS and the copula approach. To overcome this issue, Vrac & Thao (2020) recently introduced a flexible shuffling algorithm that takes different shuffling options into consideration. They, however, also highlighted that in the context of impact studies further evaluations are needed to find optimal temporal reference characteristics at different locations.

Rather strong correlations between QDM and MBCn, together with the fact that QDM slightly outperformed its multivariate counterpart in most instances in this study, point to a conclusion that the added complexity and computation time needed to implement the multivariate MBCn method might not necessarily provide added value over QDM in simulating hydrological signatures. Interestingly, previous research on the robustness of the BA methods in Nordic catchments pinpointed MBCn as a rather robust method when it comes to bias adjustment of various (including multivariate) features of precipitation and temperature series (Tootoonchi et al., 2022a). However, results presented in this study indicated that adding a hydrological model to the modeling chain can suppress benefits gained with a particular BA method (in this case, MBCn). In other words, nonlinearity of the hydrological models (Beven, 2012), which is inevitable to describe the highly nonlinear nature of catchment responses (Todini, 1996), can completely overturn conclusions on the robustness of BA methods based on their performance in reproducing features of precipitation and temperature series alone, regardless of how thorough and exhaustive those evaluations are. We, thus, argue that the adjustment of the dependence between precipitation and temperature might come with the cost of hampering the accurate representation of the temporal features of these series, for which the resulting impacted variable (here, streamflows and hydrological signatures) might pay the price. More specifically, the copula method significantly distorts temporal characteristics of the climatic variables (Tootoonchi et al., 2022a), which affected the resulting streamflow simulations. In other words, multivariate BA methods do correct the dependencies between climatic variables, but this does not necessarily translate into the streamflow simulations. Our results clearly suggest that evaluations of the ability of BA methods in reproducing climatic variables are simply not sufficient to justify their selection for a hydrological impact study at hand in Nordic catchments. Thus, examination of BA methods based on impact variables that integrate the climatic variables are crucial, and we recommend further assessments of hydrological signatures in other catchments with various physiographic characteristics and hydroclimatic regimes.

5.2. Robustness of BA methods

Previous studies (e.g., François et al., 2020; Tootoonchi et al., 2022a) showed merits of multivariate BA methods for adjusting multiple features of CM outputs. In term of hydrological simulations, however, the consequences of applying intricate BA methods varied across different studies and in different regions. Some studies demonstrated superior performance of the multivariate BA methods in specific regions (Singh and Najafi, 2020) or in reproducing specific components of runoff (Meyer et al., 2019). Other studies detected modest performance of the multivariate BA methods that led to deteriorations in the hydrological simulations (Van de Velde et al., 2022), or found no substantial difference between the performance of uni- and multivariate BA methods (Guo et al., 2020; Räty et al., 2018). However, many of these studies analyzed only performance measures like Nash-Sutcliffe (Nash and Sutcliffe, 1970), which could provide only limited information on the complex role of BA methods in streamflow simulations. Thus, to provide deeper and more detailed insights into the effects of BA methods on hydrological modeling results, we here used a suite of 16 hydrological signatures that described various streamflow characteristics, which in turn represent an integration of the temporal and bivariate aspects of precipitation and temperature at the catchment scale (Hakala et al., 2018). Our evaluation in this study was tailored to accommodate concepts of accuracy and precision, which were here applied to bias-adjusted model simulations. As elaborated in section 2.6, accuracy was represented by low average biases and high frequency of improvement (in comparison to raw CM outputs), while precision was represented by low values of variance across CMs and catchments.

Based on our holistic evaluation of performance of the BA methods in all 50 Swedish catchments, we found no considerable differences between the accuracy of uni- and multivariate BA methods. The multivariate methods performed similarly to their univariate counterparts both in reducing biases and in the frequency of improvement. The only noticeable difference was the higher frequency of improvement by multivariate methods in winter flows, compared to the univariate ones. Larger dissimilarities were, however, detected between distribution-based and distribution-free methods. Distribution-based methods consistently resulted in lower MAEs and higher frequency of improvement, except for the signatures representing temporal aspects (i.e., COMt or timing of minimum and maximum flows). This might be attributed to their poorer performance in reproducing temporal features of climate variables, i.e., cross- (copula) and autocorrelation (DS) in the precipitation and temperature series (Tootoonchi et al., 2022a). These features might result in an improperly reproduced “chronology” in these series, which reflects in timings of simulated streamflow series and, consequently, in these signatures.

Robust BA methods are expected to consistently reduce biases in simulated streamflows and hydrological signatures, i.e., to demonstrate

high precision by yielding low variability in biases across signatures, CMs or catchments, including different hydroclimatic regimes (Teutschbein and Seibert, 2012). Application of all four BA methods generally reduced the variability in biases across both CMs and catchments in comparison to raw biases. The distribution-based methods showed lower variability across CMs, meaning that distribution-based methods (especially copula) are less "sensitive" to the selection of CMs. The lowest variability was obtained with copula, which almost cancelled out any variability across CMs in many signatures. Although this can be considered to be an appealing property of the copula method, it also gives rise to concerns that in this way signals in raw CM outputs are completely concealed, which might result in loss of information offered by CMs (Cannon, 2018; Maraun, 2016).

On the contrary, copula had the highest variability across catchments, while the distribution-free BA methods generally showed lower variability. This in turn suggests that the performance of distribution-free approaches in reproducing hydrological signatures is less sensitive to hydroclimatic or other properties of catchments. This outcome was further confirmed by dividing the catchments into two distinct groups that represent snowmelt-driven (northern) and rainfall-driven (southern) streamflow regimes. Indeed, distribution-based methods in the southern catchments performed poorly in reducing biases in timing-related signatures and even inflated the biases in timing of 1-day annual maxima. In general, the exact timing of extreme precipitation is typically not accurately captured in CM simulations (Lind et al., 2020). This can impair the predictably of extreme streamflows in southern catchments, where floods are mainly generated by heavy and/or prolonged rainfall events. In contrast, high flows in the northern catchments are better simulated by the modelling chain, as they mostly depend on the snow accumulation over a period of several months, during which the exact timing of precipitation events is of less importance. We here found that catchments with typically thin snowpacks (located in the southwest of Sweden) are at particular risk of large biases in the timing of flood peaks, mostly because slight melting might entirely diminish the snowpack in these catchments, which then causes the flood peaks to become entirely rainfall-dependent, and to potentially shift to a different season.

The separate evaluation of the BA methods in the northern and southern catchments also shed light on the added value of application of the multivariate methods. For example, the multivariate methods were superior in the cold winter season in representation of all signatures depicting low flows in the northern snowmelt-driven catchments. Apparently, in these snowmelt-driven environments, the dependence between precipitation and temperature distinguishes between liquid precipitation and snow (Meyer et al., 2019), and snowmelt timing defines runoff generation, which consequently influences the seasonal cycle of streamflow. Therefore, multivariate bias-adjustment methods, such as copula and MBCn, both of which perform quite well in reproducing the dependence between precipitation and temperature (Tootoonchi et al., 2022a), also shine in the reproduction of low-flow signatures in these regions.

With global warming causing shorter snow duration, less snow accumulation on the ground and earlier snowmelt onset (Addor et al., 2014; Coppola et al., 2018; Klein et al., 2016; Ye and Cohen, 2013), the magnitude of high flows are expected to drastically decrease (Hakala et al., 2020), resulting in lower streamflows after a normal snowmelt period, and causing so-called warm snow season droughts (Van Loon and Van Lanen, 2012). Such changes may have considerable negative impacts on the environment, e.g., by jeopardizing aquatic life and impeding sediment transport. Northern Sweden (Perers et al., 2007) also hosts many hydropower plants, which require reliable operational low flow forecasts, at both short and long time scales for their optimal functionality (Abgottspon and Andersson, 2016; Boucher and Ramos, 2018), or assessment of required environmental flows (Gain et al., 2013). Shifting hydrological regimes require adaptation of farming practices to mitigate risks of loss in crops (Cervantes-Godoy et al., 2014). Thus, even slight improvements in accuracy of low-flow forecasts and/or projections by applying the multivariate BA methods may prevent considerable ecological and financial losses in these regions. For riparian studies that analyze 30-day annual maxima, which are essential for assessment of riparian forest patch persistence, distribution-based BA methods were proven superior in this study.

In rainfall-driven southern catchments, multivariate methods did not demonstrate any distinct superiority in performance. The univariate DS method even outperformed other three methods, closely followed by copula. Therefore, greater reduction in biases in the south was obtained by BA methods that adjust well univariate and temporal features of precipitation (Tootoonchi et al., 2022a). Good performance was obtained in reproducing annual maxima and minima of different durations in these catchments. These results suggest that application of distribution-based methods is preferred for studies aimed at e.g., flood hazard or flood risk assessment, or design of water intakes and estimation of environmental flows. Although not directly tested in their study, this outcome was already suspected by Zscheischler et al. (2019), who argued that in cases of (1) two climate variables not being strongly correlated, or (2) the process of interest (e.g., rainfall-driven floods) being predominantly dependent on one climate variable, the impact of adjusting dependence might be of less importance, compared to the proper representation of the most influential climate variable. Therefore, analyzing the dependence between the contributing hydrometeorological variables at the relevant spatiotemporal scale might be crucial and time-saving (Tootoonchi et al., 2022b). If there is little to no evidence of dependent behaviour, application of rather intricate multivariate methods might not add any significant value.

5.3. A comparison-based evaluation

While there was not a single method that exceled according to all eight evaluation criteria, the application of any BA method generally improved streamflow simulations by reducing systematic errors and increasing consistency in performance across CMs and catchments. These results emphasize the need for bias adjustment as an essential step that should be taken prior to any hydrological modeling exercise to ensure long-term sustainability of water resources management. The choice of an appropriate BA method is, however, not straightforward, and it depends on the study location, data, computation resources, as well as water management issue and related research question at hand. Concerning the two univariate methods (DS and QDM), distribution-based DS generally outperformed distribution-free QDM, except for the timings of high flows and low flows. Between the two multivariate methods, copula resulted in considerably better performance than MBCn. More specifically, in rainfall-driven catchments, MBCn performed poorly across all four categories of signatures. Nonetheless, both multivariate methods resulted in better performance in low-flows in snowmelt-driven catchments. MBCn outperformed copula in all four signatures within this category, and generally showed a higher skill in adjusting signatures depicting timing of flows.

Overall, the two distribution-based methods performed rather similarly, and outperformed distribution-free methods in rainfall-driven catchments. Simple univariate DS even resulted in slightly better performance in the signatures that represent water balance and flow dynamics, as well as seasonal flow, whereas copula demonstrated its merits in reproducing signatures related to low- and high flows. This behavior was also consistent between the two distribution-free BA methods, where simple applications (such as those for reproducing annual or seasonal flows) were adjusted adequately with univariate QDM, while extreme flows benefited from the application of more complex MBCn method. As discussed is section 4.1, although evaluation of the BA methods clearly revealed robustness of MBCn in reproducing various features of precipitation and temperature series (Tootoonchi et al., 2022a), when applied within the entire modeling chain that includes also hydrological models, this robustness fades out. Nevertheless, it

should be noted that multivariate methods that account for dependence between precipitation and temperature might be crucial for assessing compound extreme events (AghaKouchak et al., 2014; Brunner et al., 2019a; Teutschbein et al., 2023).

6. Concluding remarks

This study evaluated the ability of four bias adjustment methods to reproduce various features of hydrologic regimes in a 42-year long record period (1962–2004) in 50 Swedish catchments. To this end, the bias adjustment methods were applied within the ensemble of modelling chains comprising ten climate models and the HBV-light hydrological model. Special emphases in the evaluation were placed on the differences and similarities among these methods (research question 1). The results elucidate that the application of BA methods is crucial and still an essential step for hydroclimatic impact studies, and that simulation outputs (specifically, hydrological signatures) for water resources management planning strongly depend on the selection of a specific method. However, no method was proven superior for all signatures and according to all evaluation criteria considered in this study.

Additionally, the robustness of the BA methods across climate models, catchments and hydroclimatic regimes was assessed to provide deep insights into different aspect of BA performance (research question 2). We here demonstrated robust performance of distribution-based methods across CMs, indicating that they are less sensitive to the selection of climate models with different boundary conditions and process representations than distribution-base approaches. All BA methods were highly sensitive to the selection of catchments, which emphasizes the need to incorporate process understanding to the bias-adjustment step to provide bias-adjusted and simultaneously physically meaningful data for hydrological impact studies in different parts of the world. However, these results remain to be confirmed for other study areas with different hydroclimatic conditions and streamflow regimes.

Finally, in this study we sought to identify the most suitable BA method(s) for the simulation of hydrological signatures relevant for water resources management under changing climate in the Nordic regions (research question 3). To conclude, in most cases the simple and straightforward univariate distribution scaling method is the best choice as it has the highest potential to robustly reduce biases across a variety of signatures. Application of the multivariate methods (copula or MBCn) might be worth the additional efforts only in snowmelt-driven catchments or for particular purposes that require correct timing of high- or low flows. In these cases, however, one should be aware that these multivariate methods might add further uncertainties due to various assumptions for adjustment of dependence between precipitation and temperature, and consequent modifications of temporal characteristics.

CRediT authorship contribution statement

Faranak Tootoonchi: Conceptualization, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Andrijana Todorović:** Investigation, Validation, Writing – review & editing. **Thomas Grabs:** Investigation, Validation, Writing – review & editing. **Claudia Teutschbein:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The CORDEX data used is available at http://www.cordex.org.

Acknowledgments

This research was supported by the Swedish Research Council (VR) with a starting grant in the domain of Natural and Engineering Sciences (registration number 2017-04970). The Swedish Meteorological and Hydrological Institute (SMHI) is acknowledged for maintenance of the PTHBV database with meteorological data and for making the geospatial data available on their web page, which has been funded by the Swedish water authorities. We also acknowledge the free availability of the EURO-CORDEX data through the ESGF datanode at the National Supercomputer Centre, Linköping. The authors would like to thank two anonymous reviewers whose feedback on earlier drafts of this work helped to substantially improve this article.

References

Abgottspon, H., Andersson, G., 2016. Multi-horizon modeling in hydro power planning. Energy Procedia 87, 2–10. https://doi.org/10.1016/j.egypro.2015.12.351.

Addor, N., Rössler, O., Köplin, N., Huss, M., Weingartner, R., Seibert, J., 2014. Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments. Water Resour. Res. 50, 7541–7562. https://doi.org/doi:10.1002/2014WR015549.

Addor, N., Fischer, E.M., 2015. The influence of natural variability and interpolation errors on bias characterization in RCM simulations. J. Geophys. Res. 120, 10180–10195. https://doi.org/10.1002/2014JD022824.

Addor, N., Nearing, G., Prieto, C., Newman, A.J., Le Vine, N., Clark, M.P., 2018. A ranking of hydrological signatures based on their predictability in space. Water Resour. Res. 54, 8792–8812. https://doi.org/10.1029/2018WR022606.

AghaKouchak, A., Cheng, L., Mazdiyasni, O., Farahmand, A., 2014. Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 California drought. Geophys. Res. Lett. 41, 8847–8852. https://doi.org/10.1002/2014GL062308.

Arheimer, B., Lindström, G., 2015. Climate impact on floods: changes in high flows in Sweden in the past and the future (1911–2100). Hydrol. Earth Syst. Sci. 19, 771–784. https://doi.org/10.5194/hess-19-771-2015.

Bárdossy, A., Pegram, G., 2012. Multiscale spatial recorrelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small. Water Resour. Res. 48 https://doi.org/10.1029/2011WR011524.

Barlow, M., Gutowski, W.J., Gyakum, J.R., Katz, R.W., Lim, Y.-K., Schumacher, R.S., Wehner, M.F., Agel, L., Bosilovich, M., Collow, A., Gershunov, A., Grotjahn, R., Leung, R., Milrad, S., Min, S.-K., 2019. North American extreme precipitation events and related large-scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends, Climate Dynamics. Clim Dyn 53 (11), 6835–6875.

Bastin, S., Drobinski, P., Chiriaco, M., Bock, O., Roehrig, R., Gallardo, C., Conte, D., Domínguez Alonso, M., Li, L., Lionello, P., Parracho, A.C., 2019. Impact of humidity biases on light precipitation occurrence: Observations versus simulations. Atmos. Chem. Phys. 19, 1471–1490. https://doi.org/10.5194/acp-19-1471-2019.

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future köppen-geiger climate classification maps at 1-km resolution. Sci. Data 5, 1–12. https://doi.org/10.1038/sdata.2018.214.

Berg, P., Bosshard, T., Yang, W., Zimmermann, K., 2022. MIdAS — MultI-scale bias AdjuStment 1–25.

Berghuijs, W.R., Woods, R.A., Hrachowitz, M., 2014. A precipitation shift from snow towards rain leads to a decrease in streamflow. Nat. Clim. Chang. 4, 583–586. https://doi.org/10.1038/nclimate2246.

Bergström, S., 1976. Development and application of a conceptual runoff model for Scandinavian catchments. Norrköping, Sweden.

Bergström, S., Carlsson, B., Gardelin, M., Lindström, G., Petterson, A., Rummukainen, M., 2001. Climate change impacts on runoff in Sweden - assessments by global climate models, dynamical downscalling and hydrological modelling. Clim. Res. 16, 101–112.

Beven, K., 2012. Rainfall-Runoff Modelling: The Primer: Second Edition, Rainfall-Runoff Modelling: The Primer: Second Edition. https://doi.org/10.1002/9781119951001.

Boucher, M.-A., Ramos, M.-H., 2019. In: Handbook of Hydrometeorological Ensemble Forecasting. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1289–1306.

Bourdeau-Goulet, S.C., Hassanzadeh, E., 2021. Comparisons between CMIP5 and CMIP6 models: simulations of climate indices influencing food security, infrastructure resilience, and human health in Canada. Earth's Futur. 9, 1–17. https://doi.org/10.1029/2021EF001995.

Brunner, M.I., Slater, L.J., 2022. Extreme floods in Europe: Going beyond observations using reforecast ensemble pooling. Hydrol. Earth Syst. Sci. 26, 469–482. https://doi.org/10.5194/hess-26-469-2022.

Brunner, M.I., Furrer, R., Favre, A.-C., 2019a. Modeling the spatial dependence of floods using the Fisher copula. Hydrol. Earth Syst. Sci 23, 107–124. https://doi.org/10.5194/hess-23-107-2019.

Brunner, M.I., Liechti, K., Zappa, M., 2019b. Extremeness of recent drought events in Switzerland: dependence on variable and return period choice. Hazards Earth Syst. Sci 19, 2311–2323. https://doi.org/10.5194/nhess-19-2311-2019.

Brunner, M.I., Melsen, L.A., Wood, A.W., Rakovec, O., Mizukami, N., Knoben, W.J.M., Clark, M.P., 2021. Flood spatial coherence, triggers, and performance in hydrological simulations: large-sample evaluation of four streamflow-calibrated

models. Hydrol. Earth Syst. Sci. 25, 105–119. https://doi.org/10.5194/hess-25-105-2021.
Cannon, A.J., 2016. Multivariate bias correction of climate model output: matching marginal distributions and intervariable dependence structure. J. Clim. 29, 7045–7064. https://doi.org/10.1175/JCLI-D-15-0679.1.
Cannon, A.J., 2018. Multivariate quantile mapping bias correction : an N - dimensional probability density function transform for climate model simulations of multiple variables. Clim. Dyn. 50, 31–49. https://doi.org/10.1007/s00382-017-3580-6.
Cannon, A.J., Sobie, S.R., Murdock, T.Q., 2015. Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes? J. Clim. 28, 6938–6959. https://doi.org/10.1175/JCLI-D-14-00754.1.
Carpenter, S.R., Fisher, S.G., Grimm, N.B., Kitchell, J.F., 1992. Global change and freshwater ecosystems. Annu. Rev. Ecol. Syst. 23, 119–139. https://doi.org/10.1146/annurev.es.23.110192.001003.
Cervantes-Godoy, D., Dewbre, J., Amegnaglo, C.J., Soglo, Y.Y., Akpa, A.F., Bickel, M., Sanyang, S., Ly, S., Kuiseu, J., Ama, S., Gautier, B.P., 2014. The future of food and agriculture: trends and challenges, The future of food and agriculture: trends and challenges. Food Agricult. Org, United Nations, Rome, Italy.
Chen, J., Brissette, F.P., Leconte, R., 2011. Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. J. Hydrol. 401, 190–202. https://doi.org/10.1016/j.jhydrol.2011.02.020.
Chen, J., Brissette, F.P., Chaumont, D., Braun, M., 2013. Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. Water Resour. Res. 49, 4187–4205. https://doi.org/10.1002/wrcr.20331.
Chen, J., Brissette, F.P., Lucas-Picher, P., Caya, D., 2017. Impacts of weighting climate models for hydro-meteorological climate change studies. J. Hydrol. 549, 534–546. https://doi.org/10.1016/j.jhydrol.2017.04.025.
Chen, J., Arsenault, R., Brissette, F.P., Zhang, S., 2021b. Climate change impact studies: should we bias correct climate model outputs or post-process impact model outputs? Water Resour. Res. 57, 1–22. https://doi.org/10.1029/2020WR028638.
Chen, D., Zhang, P., Seftigen, K., Ou, T., Giese, M., Barthel, R., 2021a. Hydroclimate changes over Sweden in the twentieth and twenty-first centuries: a millennium perspective. Geogr. Ann. Ser. A Phys. Geogr. 103, 103–131. https://doi.org/10.1080/04353676.2020.1841410.
Cheng, L., Ma, L., Yang, M., Wan, G., Wang, X., 2019. Changes of temperature and precipitation and their impacts on runoff in the upper Taohe River in northwest China from 1956 to 2014. Environ. Earth Sci. 78, 1–14. https://doi.org/10.1007/s12665-019-8399-5.
Clifton, C.F., Day, K.T., Luce, C.H., Grant, G.E., Safeeq, M., Halofsky, J.E., Staab, B.P., 2018. Effects of climate change on hydrology and water resources in the Blue Mountains, Oregon. USA. Clim. Serv. 10, 9–19. https://doi.org/10.1016/j.cliser.2018.03.001.
Clow, D.W., 2010. Changes in the timing of snowmelt and streamflow in Colorado: A response to recent warming. J. Clim. 23, 2293–2306. https://doi.org/10.1175/2009JCLI2951.1.
Coppola, E., Raffaele, F., Giorgi, F., 2018. Impact of climate change on snow melt driven runoff timing over the Alpine region. Clim. Dyn. 51, 1259–1273. https://doi.org/10.1007/s00382-016-3331-0.
Criss, R.E., Winston, W.E., 2008. Do Nash values have value ? Discussion and alternate proposals. Hydrol. Process. 22, 2723–2725. https://doi.org/10.1002/hyp.
Cunderlik, J.M., Ouarda, T.B.M.J., 2009. Trends in the timing and magnitude of floods in Canada. J. Hydrol. 375, 471–480. https://doi.org/10.1016/j.jhydrol.2009.06.050.
Dankers, R., Arnell, N.W., Clark, D.B., Falloon, P.D., Fekete, B.M., Gosling, S.N., Heinke, J., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., Wisser, D., 2014. First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble. Proc. Natl. Acad. Sci. U. S. A. 111, 3257–3261. https://doi.org/10.1073/pnas.1302078110.
Dankers, R., Feyen, L., 2009. Flood hazard in Europe in an ensemble of regional climate scenarios. J. Geophys. Res. Atmos. 114, 1–16. https://doi.org/10.1029/2008JD011523.
Dobler, C., Hagemann, S., Wilby, R.L., StÃtter, J., 2012. Quantifying different sources of uncertainty in hydrological projections in an Alpine watershed. Hydrol. Earth Syst. Sci. 16, 4343–4360. https://doi.org/10.5194/hess-16-4343-2012.
Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., Liebert, J., 2012. HESS Opinions "should we apply bias correction to global and regional climate model data?" Hydrol. Earth Syst. Sci. 16, 3391–3404. https://doi.org/10.5194/hess-16-3391-2012.
Eklund, A., 2011. SVAR, Svenskt vattenarkiv.
Feng, D., Beighley, E., 2020. Identifying uncertainties in hydrologic fluxes and seasonality from hydrologic model components for climate change impact assessments. Hydrol. Earth Syst. Sci. 24, 2253–2267. https://doi.org/10.5194/hess-24-2253-2020.
François, B., Vrac, M., Cannon, A., Robin, Y., Allard, D., 2020. Multivariate bias corrections of climate simulations: Which benefits for which losses? Earth Syst. Dyn. Discuss. 1–41 https://doi.org/10.5194/esd-2020-10.
Gain, A.K., Apel, H., Renaud, F.G., Giupponi, C., 2013. Thresholds of hydrologic flow regime of a river and investigation of climate change impact-the case of the Lower Brahmaputra river Basin. Clim. Change 120, 463–475. https://doi.org/10.1007/s10584-013-0800-x.
Genest, C., Favre, A.-C., 2007. Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. J. Hydrol. Eng. 12, 347–368. https://doi.org/10.1061/(asce)1084-0699(2007)12:4(347).
Genest, C., Remillard, B., 2008. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. Ann. l'institut Henri Poincare Probab. Stat. 44, 1096–1127. https://doi.org/10.1214/07-AIHP148.
Genest, C., Rémillard, B., Beaudoin, D., 2009. Goodness-of-fit tests for copulas: A review and a power study. Insur. Math. Econ. 44, 199–213. https://doi.org/10.1016/j.insmatheco.2007.10.005.
Gergel, D.R., Nijssen, B., Abatzoglou, J.T., Lettenmaier, D.P., Stumbaugh, M.R., 2017. Effects of climate change on snowpack and fire potential in the western USA. Clim. Change 141, 287–299. https://doi.org/10.1007/s10584-017-1899-y.
Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., Moyer, E.J., 2014. Evaluating the utility of dynamical downscaling in agricultural impacts projections. Proc. Natl. Acad. Sci. U. S. A. 111, 8776–8781. https://doi.org/10.1073/pnas.1314787111.
Gudmundsson, L., Bremnes, J.B., Haugen, J.E., Engen-Skaugen, T., 2012. Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – A comparison of methods. Hydrol. Earth Syst. Sci. 16, 3383–3390. https://doi.org/10.5194/hess-16-3383-2012.
Guo, Q., Chen, J., Zhang, X.J., Xu, C., Chen, H., 2020. Impacts of Using State-of-the-art multivariate bias correction methods on hydrological modeling over North America. Water Resour. Res. 56 https://doi.org/10.1029/2019WR026659.
Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.
Hakala, K., Addor, N., Seibert, J., 2018. Hydrological modeling to evaluate climate model simulations and their bias correction. J. Hydrometeorol. 19, 1321–1337. https://doi.org/10.1175/JHM-D-17-0189.1.
Hakala, K., Addor, N., Teutschbein, C., Vis, M., Dakhlaoui, H., Seibert, J., 2019. Hydrological Modeling of Climate Change Impacts. Encycl. Water 1–20. https://doi.org/10.1002/9781119300762.wsts0062.
Hakala, K., Addor, N., Gobbe, T., Ruffieux, J., Seibert, J., 2020. Risks and opportunities for a Swiss hydroelectricity company in a changing climate. Hydrol. Earth Syst. Sci. 24, 3815–3833. https://doi.org/10.5194/hess-24-3815-2020.
Hallerbäck, S., Huning, L., Love, C., Persson, M., Stensen, K., Gustafsson, D., AghaKouchak, A., 2021. Warming climate shortens ice durations and alters freeze and breakup patterns in Swedish water bodies. Cryosph. Discuss. 5, 1–20.
Hamon, W.R., 1961. Estimating potential evaporation, in: Proceedings of the American Society of Civil Engineers, Division, J.o.H. pp. 107–120.
Hanus, S., Hrachowitz, M., Zekollari, H., Schoups, G., Vizcaino, M., Kaitna, R., 2021. Timing and magnitude of future annual runoff extremes in contrasting Alpine catchments in Austria. Hydrol. Earth Syst. Sci. Discuss. 1–35.
Huang, X., Ding, A., 2021. Aerosol as a critical factor causing forecast biases of air temperature in global numerical weather prediction models. Sci. Bull. 66, 1917–1924. https://doi.org/10.1016/j.scib.2021.05.009.
IPCC, 2014. IPCC, 2014: Climate change 2014: synthesis report.
IPCC, 2021. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen,. Cambridge Univ. Press 3949.
Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O.B., Bouwer, L.M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., Yiou, P., 2014. EURO-CORDEX: New high-resolution climate change projections for European impact research. Reg. Environ. Chang. 14, 563–578. https://doi.org/10.1007/s10113-013-0499-2.
Johansson, B., 2002. Estimation of areal precipitation for hydrological modelling in Sweden. Göteborg University, Göteborg, Sweden.
Kabuya, P.M., Hughes, D.A., Tshimanga, R.M., Trigg, M.A., Bates, P., 2020. Establishing uncertainty ranges of hydrologic indices across climate and physiographic regions of the Congo River Basin. J. Hydrol. Reg. Stud. 30, 100710 https://doi.org/10.1016/j.ejrh.2020.100710.
Klein, G., Vitasse, Y., Rixen, C., Marty, C., Rebetez, M., 2016. Shorter snow cover duration since 1970 in the swiss alps due to earlier snowmelt more than to later snow onset. Clim. Change 139, 637–649. https://doi.org/10.1007/s10584-016-1806-y.
Kormos, P.R., Luce, C.H., Wenger, S.J., Berghuijs, W.R., 2016. Trends and sensitivities of low streamflow extremes to discharge timing and magnitude in Pacific Northwest mountain streams. Water Resour. Res. 52, 4990–5007. https://doi.org/10.1002/2015WR018125.
Krysanova, V., Vetter, T., Eisner, S., Huang, S., Pechlivanidis, I., Strauch, M., Gelfan, A., Kumar, R., Aich, V., Arheimer, B., Chamorro, A., van Griensven, A., Kundu, D., Lobanova, A., Mishra, V., Plötner, S., Reinhardt, J., Seidou, O., Wang, X., Wortmann, M., Zeng, X., Hattermann, F.F., 2017. Intercomparison of regional-scale hydrological models and climate change impacts projected for 12 large river basins worldwide - A synthesis. Environ. Res. Lett. 12 (10), 105002.
Laux, P., Rötter, R.P., Webber, H., Dieng, D., Rahimi, J., Wei, J., Faye, B., Srivastava, A. K., Bliefernicht, J., Adeyeri, O., Arnault, J., Kunstmann, H., 2021. To bias correct or not to bias correct? An agricultural impact modelers' perspective on regional climate model data. Agric. For. Meteorol. 304-305, 108406.
Lenderink, G., Buishand, A., Van Deursen, W., 2007. Estimates of future discharges of the river Rhine using two scenario methodologies: Direct versus delta approach. Hydrol. Earth Syst. Sci. 11, 1145–1159. https://doi.org/10.5194/hess-11-1145-2007.
Li, H., Sheffield, J., Wood, E.F., 2010. Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. J. Geophys. Res. Atmos. 115 https://doi.org/10.1029/2009JD012882.
Li, C., Sinha, E., Horton, D.E., Diffenbaugh, N.S., Michalak, A.M., 2014. Joint bias correction of temperature and precipitation in climate model simulations.

J. Geophys. Res. Atmos. 119, 13153–13162. https://doi.org/10.1002/2014JD022514.
Lind, P., Belušić, D., Christensen, O.B., Dobler, A., Kjellström, E., Landgren, O., Lindstedt, D., Matte, D., Pedersen, R.A., Toivonen, E., Wang, F., 2020. Benefits and added value of convection - permitting climate modeling over Fenno - Scandinavia. Clim. Dyn. 55, 1893–1912. https://doi.org/10.1007/s00382-020-05359-3.
Lindström, G., Johansson, B., Persson, M., Gardelin, M., Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. J. Hydrol. 201, 272–288. https://doi.org/10.2166/wst.2009.488.
Lucas-Picher, P., Argüeso, D., Brisson, E., Tramblay, Y., Berg, P., Lemonsu, A., Kotlarski, S., Caillaud, C., 2021. Convection-permitting modeling with regional climate models: Latest developments and next steps. Wiley Interdiscip. Rev. Clim. Chang. 12, 1–59. https://doi.org/10.1002/wcc.731.
Maraun, D., 2016. Bias Correcting Climate Change Simulations - a Critical Review. Curr. Clim. Chang. Reports 2, 211–220. https://doi.org/10.1007/s40641-016-0050-x.
Maraun, D., Widmann, M., 2018. Cross-validation of bias-corrected climate simulations is misleading. Hydrol. Earth Syst. Sci. 22, 4867–4873. https://doi.org/10.5194/hess-22-4867-2018.
Matti, B., Dahlke, H.E., Lyon, S.W., 2016. On the variability of cold region flooding. J. Hydrol. 534, 669–679. https://doi.org/10.1016/j.jhydrol.2016.01.055.
Matti, B., Dahlke, H.E., Dieppois, B., Lawler, D.M., Lyon, S.W., 2017. Flood seasonality across Scandinavia—Evidence of a shifting hydrograph? Hydrol. Process. 31, 4354–4370. https://doi.org/10.1002/hyp.11365.
McMillan, H.K., 2021. A review of hydrologic signatures and their applications. Wiley Interdiscip. Rev. Water 8, 1–23. https://doi.org/10.1002/wat2.1499.
McMillan, H.K., Westerberg, I.K., Krueger, T., 2018. Hydrological data uncertainty and its implications. WIREs Water 5, 1–14. https://doi.org/10.1002/wat2.1319.
Mehrotra, R., Johnson, F., Sharma, A., 2018. A software toolkit for correcting systematic biases in climate model simulations. Environ. Model. Softw. 104, 130–152. https://doi.org/10.1016/j.envsoft.2018.02.010.
Mehrotra, R., Sharma, A., 2021. A robust alternative for correcting systematic biases in multi-variable climate model simulations. Environ. Model. Softw. 139, 105019 https://doi.org/10.1016/j.envsoft.2021.105019.
Mendoza, P.A., Clark, M.P., Mizukami, N., Gutmann, E.D., Arnold, J.R., Brekke, L.D., Rajagopalan, B., 2016. How do hydrologic modeling decisions affect the portrayal of climate change impacts? Hydrol. Process. 30, 1071–1095. https://doi.org/10.1002/hyp.10684.
Merz, R., Blöschl, G., 2009. A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria. Water Resour. Res. 45, 1–19. https://doi.org/10.1029/2008WR007163.
Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J., Cannon, A.J., 2019. Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments. Hydrol. Earth Syst. Sci. 23, 1339–1354. https://doi.org/10.5194/hess-23-1339-2019.
Muerth, M.J., Gauvin St-Denis, B., Ricard, S., Velázquez, J.A., Schmid, J., Minville, M., Caya, D., Chaumont, D., Ludwig, R., Turcotte, R., 2013. On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff. Hydrol. Earth Syst. Sci. 17, 1189–1204. https://doi.org/10.5194/hess-17-1189-2013.
Myhre, G., Alterskjær, K., Stjern, C.W., Hodnebrog, Ø., Marelle, L., Samset, B.H., Sillmann, J., Schaller, N., Fischer, E., Schulz, M., Stohl, A., 2019. Frequency of extreme precipitation increases extensively with event rareness under global warming. Sci. Rep. 9 (1) https://doi.org/10.1038/s41598-019-52277-4.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I - A discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.
Nelsen, R.B., 2006. An introduction to Copulas. Springer.
Olden, J.D., Poff, N.L., 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Res. Appl. 19, 101–121. https://doi.org/10.1002/rra.700.
Olsson, J., Berggren, K., Olofsson, M., Viklander, M., 2009. Applying climate model precipitation scenarios for urban hydrological assessment: A case study in Kalmar City. Sweden. Atmos. Res. 92, 364–375. https://doi.org/10.1016/j.atmosres.2009.01.015.
Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., Michel, C., 2006. Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. Water Resour. Res. 42 (W07410), 1–10. https://doi.org/10.1029/2005WR004636.
Panthou, G., Mailhot, A., Laurence, E., Talbot, G., 2014. Relationship between surface temperature and extreme rainfalls: A multi-time-scale and event-based analysis. J. Hydrometeorol. 15, 1999–2011. https://doi.org/10.1175/JHM-D-14-0020.1.
Pappadà, R., Durante, F., Salvadori, G., 2017. Quantification of the environmental structural risk with spoiling ties: is randomization worthwhile? Stoch. Environ. Res. Risk Assess. 31, 2483–2497. https://doi.org/10.1007/s00477-016-1357-9.
Parajka, J., Paul Blaschke, A., Blöschl, G., Haslinger, K., Hepp, G., Laaha, G., Schoner, W., Trautvetter, H., Viglione, A., Zessner, M., 2016. Uncertainty contributions to low-flow projections in Austria. Hydrol. Earth Syst. Sci. 20, 2085–2101. https://doi.org/10.5194/hess-20-2085-2016.
Pearson, K., 1920. Notes on the History of {Correlation}. Biometrika 13, 25. https://doi.org/10.2307/2331722.
Perers, R., Lundin, U., Leijon, M., 2007. Development of synchronous generators for Swedish hydropower: A review. Renew. Sustain. Energy Rev. 11, 1008–1017. https://doi.org/10.1016/j.rser.2005.07.007.
Piani, C., Haerter, J.O., 2012. Two dimensional bias correction of temperature and precipitation copulas in climate models. Geophys. Res. Lett. 39, 1–6. https://doi.org/10.1029/2012GL053839.
Pool, S., Vis, M.J.P., Knight, R.R., Seibert, J., 2017. Streamflow characteristics from modeled runoff time series - Importance of calibration criteria selection. Hydrol. Earth Syst. Sci. 21, 5443–5457. https://doi.org/10.5194/hess-21-5443-2017.
Räty, O., Räisänen, J., Bosshard, T., Donnelly, C., 2018. Intercomparison of Univariate and Joint Bias Correction Methods in Changing Climate From a Hydrological Perspective. Climate 6, 33. https://doi.org/10.3390/cli6020033.
Refsgaard, J.C., Madsen, H., Andréassian, V., Arnbjerg-Nielsen, K., Davidson, T.A., Drews, M., Hamilton, D.P., Jeppesen, E., Kjellström, E., Olesen, J.E., Sonnenborg, T. O., Trolle, D., Willems, P., Christensen, J.H., 2014. A framework for testing the ability of models to project climate change and its impacts. Clim. Change 122, 271–282. https://doi.org/10.1007/s10584-013-0990-2.
Richter, B.D., Baumgartner, J.V., Powell, J., Braun, D.P., 1996. A Method for Assessing Hydrologic Alteration within Ecosystems. Conserv. Biol. 10, 1163–1174. https://doi.org/10.1046/j.1523-1739.1996.10041163.x.
Rizzo, M.L., Székely, G.J., 2016. Energy distance. Wiley Interdiscip. Rev. Comput. Stat. 8, 27–38. https://doi.org/10.1002/wics.1375.
Robin, Y., Vrac, M., Naveau, P., Yiou, P., 2019. Multivariate stochastic bias corrections with optimal transport. Hydrol. Earth Syst. Sci. 23, 773–786. https://doi.org/10.5194/hess-23-773-2019.
Santos, L., Thirel, G., Perrin, C., 2018. Technical note: Pitfalls in using log-transformed flows within the KGE criterion. Hydrol. Earth Syst. Sci. 22, 1–14. https://doi.org/10.5194/hess-2018-298.
Seibert, J., 2000. Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. Hydrol. Earth Syst. Sci. Discuss. 4, 215–224. https://doi.org/10.5194/hess-4-215-2000.
Seibert, J., Bergström, S., 2022. A retrospective on hydrological catchment modelling based on half a century with the HBV model. Hydrol. Earth Syst. Sci. 26, 1371–1388. https://doi.org/10.5194/hess-26-1371-2022.
Seibert, J., Vis, M.J.P., 2012. Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. Hydrol. Earth Syst. Sci. 16, 3315–3325. https://doi.org/10.5194/hess-16-3315-2012.
Seneviratne, S.I., Hauser, M., 2020. Regional Climate Sensitivity of Climate Extremes in CMIP6 Versus CMIP5 Multimodel Ensembles. Earth's Futur. 8, 1–12. https://doi.org/10.1029/2019EF001474.
Shen, H., Tolson, B.A., Mai, J., 2022. Time to Update the Split-Sample Approach in Hydrological Model Calibration. Water Resour. Res. 58, 1–26. https://doi.org/10.1029/2021WR031523.
Singh, H., Najafi, M.R., 2020. Evaluation of gridded climate datasets over Canada using univariate and bivariate approaches: Implications for hydrological modelling. J. Hydrol. 584, 124673 https://doi.org/10.1016/j.jhydrol.2020.124673.
Sklar, A., 1959. Fonctions de Répartition à n Dimensions et Leurs Marges. Paris Publ. l'Institut Stat. L'Université Paris. 8, 229–231.
Székely, G.J., Rizzo, M.L., 2013. Energy statistics: A class of statistics based on distances. J. Stat. Plan. Inference. 143 (8), 1249–1272.
Teutschbein, C., Jonsson, E., Todorović, A., Tootoonchi, F., Stenfors, E., Grabs, T., 2023. Future drought propagation through the water-energy-food-ecosystem nexus – A Nordic perspective. J. Hydrol. 617, 128963.
Teutschbein, C., Seibert, J., 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. J. Hydrol. 456–457, 12–29. https://doi.org/10.1016/j.jhydrol.2012.05.052.
Teutschbein, C., Quesada Montano, B., Todorović, A., Grabs, T., 2022. Streamflow droughts in Sweden: Spatiotemporal patterns emerging from six decades of observations. J. Hydrol. Reg. Stud. 42, 101171.
Teutschbein, C., Seibert, J., 2013. Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions? Hydrol. Earth Syst. Sci. 17, 5061–5077. https://doi.org/10.5194/hess-17-5061-2013.
Teutschbein, C., Grabs, T., Karlsen, R.H., Laudon, H., Bishop, K., 2015. Hydrological response to changing climate conditions: Spatial streamflow variability in the boreal region. Water Resour. Res. 51, 9425–9446. https://doi.org/10.1002/2015WR017337.
Teutschbein, C., Sponseller, R.A., Grabs, T., Blackburn, M., Boyer, E.W., Hytteborn, J.K., Bishop, K., 2017. Future Riverine Inorganic Nitrogen Load to the Baltic Sea From Sweden: An Ensemble Approach to Assessing Climate Change Effects. Global Biogeochem. Cycles 31, 1674–1701. https://doi.org/10.1002/2016GB005598.
Thom, H.C., 1951. A frequency distribution for precipitation. Bull. Am. Meteorol. Soc. 32, 397.
Todini, E., 1996. The ARNO rainfall-runoff model. J. Hydrol. 175, 339–382. https://doi.org/10.1016/S0022-1694(96)80016-3.
Todorović, A., Grabs, T., Teutschbein, C., 2022. Advancing traditional strategies for testing hydrological model fitness in a changing climate. Hydrol. Sci. J. 67 (12), 1790–1811.
Tootoonchi, F., Haerter, J.O., Todorović, A., Räty, O., Grabs, T., Teutschbein, C., 2022a. Uni- and multivariate bias adjustment methods in Nordic catchments: Complexity and performance in a changing climate. Sci. Total Environ. 853, 158615.
Tootoonchi, F., Sadegh, M., Haerter, J.O., Räty, O., Grabs, T., Teutschbein, C., 2022b. Copulas for hydroclimatic analysis: A practice-oriented overview. Wiley Interdiscip. Rev. Water 9. https://doi.org/10.1002/wat2.1579.
Van de Velde, J., Demuzere, M., De Baets, B., Verhoest, N.E.C., 2022. Impact of bias nonstationarity on the performance of uni- and multivariate bias-adjusting methods 26, 2319–2344. https://doi.org/10.5194/hess-2020-639.
Van Loon, A.F., Van Lanen, H.A.J., 2012. A process-based typology of hydrological drought. Hydrol. Earth Syst. Sci. 16, 1915–1946. https://doi.org/10.5194/hess-16-1915-2012.
Velázquez, J.A., Schmid, J., Ricard, S., Muerth, M.J., Gauvin St-Denis, B., Minville, M., Chaumont, D., Caya, D., Ludwig, R., Turcotte, R., 2013. An ensemble approach to

assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources. Hydrol. Earth Syst. Sci. 17, 565–578. https://doi.org/10.5194/hess-17-565-2013.

Villarini, G., Wasko, C., 2021. Humans, climate and streamflow. Nat. Clim. Chang. 11, 725–726. https://doi.org/10.1038/s41558-021-01137-z.

Vis, M., Knight, R., Pool, S., Wolfe, W., Seibert, J., 2015. Model calibration criteria for estimating ecological flow characteristics. Water (Switzerland) 7, 2358–2381. https://doi.org/10.3390/w7052358.

Vormoor, K., Lawrence, D., Heistermann, M., Bronstert, A., 2015. Climate change impacts on the seasonality and generation processes of floods – Projections and uncertainties for catchments with mixed snowmelt/rainfall regimes. Hydrol. Earth Syst. Sci. 19, 913–931. https://doi.org/10.5194/hess-19-913-2015.

Vrac, M., 2018. Multivariate bias adjustment of high-dimensional climate simulations: the Rank Resampling for Distributions and Dependences R2D2 bias correction. Hydrol. Earth Syst. Sci. 22, 3175–3196. https://doi.org/10.5194/hess-22-3175-2018.

Vrac, M., Thao, S., 2020. R2D2 v2.0: Accounting for temporal dependences in multivariate bias correction via analogue rank resampling. Geosci. Model Dev. 13, 5367–5387. https://doi.org/10.5194/gmd-13-5367-2020.

Vrac, M., Noël, T., Vautard, R., 2016. Bias correction of precipitation through singularity stochastic removal: Because occurrences matter. J. Geophys. Res. 121, 5237–5258. https://doi.org/10.1002/2015JD024511.

Wagener, T., Reinecke, R., Pianosi, F., 2022. On the evaluation of climate change impact models. Wiley Interdiscip. Rev. Clim. Chang. 1–13 https://doi.org/10.1002/wcc.772.

Westerberg, I.K., McMillan, H.K., 2015. Uncertainty in hydrological signatures. Hydrol. Earth Syst. Sci. 19, 3951–3968. https://doi.org/10.5194/hess-19-3951-2015.

Wilcke, R.A.I., Mendlik, T., Gobiet, A., 2013. Multi-variable error correction of regional climate models. Clim. Change 120, 871–887. https://doi.org/10.1007/s10584-013-0845-x.

Willems, P., Vrac, M., 2011. Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change. J. Hydrol. 402, 193–205. https://doi.org/10.1016/j.jhydrol.2011.02.030.

Yang, W., Andréasson, J., Graham, L.P., Olsson, J., Rosberg, J., Wetterhall, F., 2010. Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies. Hydrol. Res. 41, 211–229. https://doi.org/10.2166/nh.2010.004.

Ye, H., Cohen, J., 2013. A shorter snowfall season associated with higher air temperatures over northern Eurasia. Environ. Res. Lett. 8 (1), 014052.

Yilmaz, K.K., Vrugt, J.A., Gupta, H.V., Sorooshian, S., 2010. Model calibration in watershed hydrology. In: Sivakumar, B., Berndtsson, R. (Eds.), Advances in Data-Based Approaches for Hydrologic Modeling and Forecasting. WORLD SCIENTIFIC, pp. 53–105.

Zehe, E., Sivapalan, M., 2009. Threshold behaviour in hydrological systems as (human) geo-ecosystems: Manifestations, controls, implications. Hydrol. Earth Syst. Sci. 13, 1273–1297. https://doi.org/10.5194/hess-13-1273-2009.

Zhang, Y., Chiew, F.H.S., Li, M., Post, D., 2018. Predicting Runoff Signatures Using Regression and Hydrological Modeling Approaches. Water Resour. Res. 54, 7859–7878. https://doi.org/10.1029/2018WR023325.

Zscheischler, J., Fischer, E.M., Lange, S., 2019. The effect of univariate bias adjustment on multivariate hazard estimates. Earth Syst. Dyn. 10, 31–43. https://doi.org/10.5194/esd-10-31-2019.