



Uni- and multivariate bias adjustment methods in Nordic catchments: Complexity and performance in a changing climate

Faranak Tootoonchi^{a,*}, Jan O. Haerter^{b,c,d}, Andrijana Todorović^e, Olle Rätty^f, Thomas Grabs^a, Claudia Teutschbein^a

^a Department of Earth Sciences, Uppsala University, Uppsala, Sweden

^b Niels Bohr Institute, Copenhagen University, Blegdamsvej 17, 2100 Copenhagen, Denmark

^c Complexity and Climate, Leibniz Centre for Tropical Marine Research, Fahrenheitstrasse 6, 28359 Bremen, Germany

^d Physics and Earth Sciences, Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany

^e University of Belgrade, Faculty of Civil Engineering, Institute of Hydraulic and Environmental Engineering, Belgrade, Republic of Serbia

^f Finnish Meteorological Institute, P.O. BOX 503, FI-00101 Helsinki, Finland

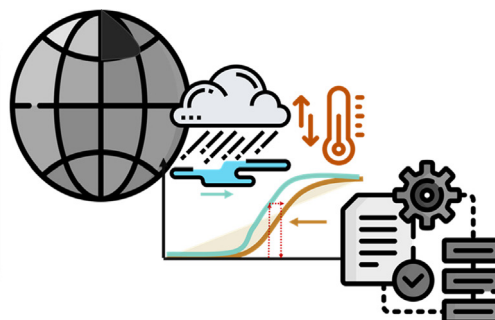


HIGHLIGHTS

- Biases in climate model outputs of precipitation and temperature are substantial.
- Bias adjustment generally improves various aspects of raw climate model outputs.
- Not a single bias adjustment method excels in all of the investigated statistical features.
- Simple univariate distribution scaling (DS) performs well for simple hydroclimatic applications.
- Multivariate bias correction (MBCn) performs better for complex applications, but is computationally more demanding.

GRAPHICAL ABSTRACT

Uni- and multivariate bias adjustment methods in Nordic catchments: Complexity and performance in a changing climate



ARTICLE INFO

Editor: Martin Drews

Keywords:

Bias adjustment
Bias correction
Univariate and multivariate methods
Precipitation and temperature
Climate change
Sweden

ABSTRACT

For climate-change impact studies at the catchment scale, meteorological variables are typically extracted from ensemble simulations provided by global and regional climate models, which are then downscaled and bias-adjusted for each study site. For bias adjustment, different statistical methods that re-scale climate model outputs have been suggested in the scientific literature. They range from simple univariate methods that adjust each meteorological variable individually, to more complex and more demanding multivariate methods that take existing relationships between meteorological variables into consideration. Over the past decade, several attempts have been made to evaluate such methods in various regions. There is, however, still no guidance for choosing appropriate bias adjustment methods for a study at hand. In particular, the question whether the benefits of potentially improved adjustments outweigh the cost of increased complexity, remains unanswered.

This paper presents a comprehensive evaluation of the performance of two commonly used univariate and two multivariate bias adjustment methods in reproducing numerous univariate, multivariate and temporal features of precipitation and temperature series in different catchments in Sweden. The paper culminates in a discussion on trade-offs between the potential benefits (i.e., skills and added value) and disadvantages (complexity and computational demand) of each method to offer plausible, defensible and actionable insights from the standpoint of climate-change impact studies in high latitudes.

We concluded that all selected bias adjustment methods generally improved the raw climate model simulations, but that not a single method consistently outperformed the other methods. There were, however, differences in the

* Corresponding author.

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

methods' performance for particular statistical features, indicating that other practical aspects such as computational time and heavy theoretical requirements should also be taken into consideration when choosing an appropriate bias adjustment method.

1. Introduction

Climate models (CMs) are primary tools for reconstructing past and predicting future climate. It is common procedure to use global climate models (GCMs) for large-scale studies, and a combination of GCMs and regional climate models (RCMs) for climate-change impact studies at a finer spatial resolution. Such studies encompass a large variety of research topics, ranging from hydroclimatic impact studies that involve, e.g., estimation of future droughts (Masud et al., 2017), floods (Brunner et al., 2019) or wildfires (Yang et al., 2015), to fundamental research projects that advance the understanding of the climate system beyond what could be achieved from observations or theories alone (Edwards, 2011). Therefore, a precise representation of the climate system is important not only for climate modeling, but also for many other research disciplines, including for example impact studies on Earth's natural systems and the interconnected socio-economic consequences, as well as research on mitigation and adaptation strategies (Bonan and Doney, 2018).

However, CMs typically provide a somewhat biased representation of reality, i.e., the statistical features of their outputs (such as mean, variance, quantiles, dependence between variables or temporal properties) do not necessarily comply with the features of the observed variables such as precipitation or temperature (Hakala et al., 2019; Maraun, 2016). Formally, such biases can be defined as “systematic differences between model simulations and observations” (Jung, 2005; Teutschbein and Seibert, 2012), and they may occur because CMs rely on the parametrization of processes (e.g., atmospheric convection) that exist at a smaller scale than what the models are typically able to resolve (Gentine et al., 2018; Maher et al., 2018). During parametrization, average or expected values of the variables are specified or determined through empirical relations, which causes uncertainty in CM outputs and is considered one of the main reasons for biases (McFarlane, 2011). Thus, CM outputs cannot readily be used for impact studies (Teutschbein and Seibert, 2010), and some adjustments to reduce the existing biases are required prior to any impact study (Hakala et al., 2019; Maraun et al., 2017; Teutschbein et al., 2011). Formally, this process is called ‘bias correction’ or ‘bias adjustment’ (BA). The latter term prevails in recent scientific publications (François et al., 2020; Schmith et al., 2020; Zscheischler et al., 2017) and is, thus, adopted in this paper.

In the past two decades, various BA methods have been developed. All methods rely on some kind of algorithm that is established (i.e., calibrated) to improve agreement between the CM outputs and the corresponding observations over a given climatic period (frequently named as control or calibration period) in which observations are available. Then the established algorithm serves as basis to adjust the entire record of the simulated climate variable, including both calibration and future periods (also referred to as projection period).

Traditionally, BA methods were only applied to one simulated variable at a time to adjust one or several univariate statistical attributes, without considering the dependence structure among multiple climate variables. Such simple univariate methods were already introduced around the turn of the millennium and include for instance linear scaling, the delta approach or quantile mapping. For detailed reviews on their underlying mathematical assumptions and their performance in different regions, we refer the reader to other references (e.g., Fang et al., 2015; Gudmundsson et al., 2012; Gutjahr and Heinemann, 2013; Maraun and Widmann, 2018a; Rätty et al., 2014; Teutschbein and Seibert, 2012; Watanabe et al., 2012). One of the most commonly used univariate BA methods is quantile mapping (QM), which adjusts the full probability distribution of a climate variable. Originally introduced by Panofsky et al. (1958), QM shows better

performance compared to other univariate methods that only adjust biases in the mean or variance (Teutschbein and Seibert, 2012).

A changing climate will influence not solely univariate characteristics of variables, but also their dependences (i.e., multivariate features), and temporal aspects, which are of critical importance for properly modeling and projecting climate change impacts on human society and ecosystems (Chen et al., 2020). In fact, shifts in temporal and multivariate aspects are already being observed (Ettinger et al., 2021; Singh et al., 2020). But univariate BA methods do not consider biases in cross-dependence attributes among multiple CM outputs and may, thus, generate unrealistic multivariate situations, causing misinterpretations and distortions of the result of impact studies (Meyer et al., 2019; Zscheischler et al., 2019). Multivariate characteristics are especially relevant for extreme hydroclimatic compound events that result from the interplay between different variables (Zscheischler et al., 2019). For example, droughts may occur because of the compound effect of a precipitation deficit coupled with high temperatures that lead to increased evapotranspiration (AghaKouchak et al., 2014). In high latitudes or at high elevations, the precipitation-temperature dependence also influences the transition between rainfall and snowfall as well as snowmelt processes, and therefore plays a key role in controlling spring flood peak, volume and timing (Meyer et al., 2019). Thus, adjusting simulated variables separately might lead to inaccurate estimates of their combined influence and the risk of compound events.

Over the past few years, new BA methods, which specifically consider the multivariate properties of CM outputs, have been proposed. These methods differ from the traditional univariate methods in their level of complexity, the number of observed/adjusted climate variables, the statistical attributes to be adjusted, and the underlying assumptions. Among the proposed methods, two approaches are frequently adopted: (1) different variations of the copula approach in its empirical form (Piani and Haerter, 2012; Vrac, 2018) or theoretical form (Laux et al., 2011; C. Li et al., 2014), and (2) multivariate bias correction in n dimensions (MBCn) (Cannon, 2018), which is a multivariate extension of the widely used univariate QM approach.

Given the relatively recent development of multivariate BA methods, most of them are not yet fully understood and their ability to adjust the CM outputs, added value and suitability for impacts studies have not been thoroughly examined, mainly because of their complexity and sometimes challenging underlying statistical assumptions. In fact, the scientific literature only provides few studies that systematically compare their performance in adjusting multivariate and temporal statistics of climate variables. For example, Vrac and Friederichs (2015) analyzed the performance of two multivariate BA methods to adjust coarse-scale gridded re-analysis data (temperature and precipitation) over southern France. François et al. (2020) compared the ability of several multivariate BA methods to adjust coarse-scale gridded temperature and precipitation simulations of a single CM over France. Both studies concluded that application of multivariate BA methods is beneficial for intervariable dependences.

Proper representation of the complex relationships among climate variables are particularly important for hydrological impact studies such as climate change or land use/land cover impact assessments, or for water management in general. Therefore, good performance of BA methods in reproducing these relationships, especially at spatial scales that are relevant for the hydrological studies, is essential. These scales are determined by the spatial resolutions at which runoff generation processes are being simulated.

For this reason, the present study aims at adding another piece to the yet unsolved puzzle of suitability of multivariate BA methods for practical

tasks, and at providing more plausible, defensible and actionable information primarily for future climate-change impact studies at the catchment scale (Hewitson et al., 2014; Maraun et al., 2015). Based on a multi-catchment and multi-climate-model approach, we here present a novel systematic evaluation of the ability of the four commonly used uni- and multivariate BA methods to adjust a wide variety of statistical features of simulated precipitation and temperature series aggregated over 55 catchments of different sizes and climatic regimes across Sweden. We conclude this paper by discussing the trade-offs between potential benefits and disadvantages of each method from the standpoint of climate-change impact studies in high latitudes.

2. Methods

2.1. Study area

Sweden is a country in northern Europe with a land area of roughly 408,000 km², and an elevation range of -2 to 2100 m.a.s.l.. It covers three of the Köppen-Geiger climate zones (Fig. 1a) (Beck et al., 2018): (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, very cold winter, persistent seasonal snow cover and soil frost during winters, stretches over Central and Northern Sweden, and (3) the warm summer hemiboreal climate zone (Dfb) can be found in Southern Sweden.

Sweden is a heavily forested country (69 % of the land area) with a large number (9 %) of lakes, wetlands and streams. Roughly 8 % of the area is covered by shrubs and grass land, 8 % is agricultural land, 3 % urban areas and the remaining 3 % open land and glaciers.

The present study used a well-established data set consisting of 55 Swedish catchments with different sizes (ranging from 2 km² to 22,600 km²), which has been the basis for various hydroclimatic studies (e.g., Teutschbein et al., 2022; Todorović et al., 2022). The selected sites

cover an extensive range of latitudes from 55.9°N to 68.4°N and, thus, spread over all three Swedish climate regions (Fig. 1a). Daily mean temperature in the catchments in period 1961–2004 varied from -3.0 °C in the North to +7.7 °C in the South (Fig. 1b), while average daily precipitation ranged from 3.2 mm·day⁻¹ in north-western Sweden to 1.4 mm·day⁻¹ in southern Sweden (Fig. 1c). Temperature exhibited a pronounced seasonality, with the highest temperatures in July (Fig. 1b). Precipitation also displayed a distinct pattern with the highest precipitation rates in summers and autumns (Fig. 1c). Variability in precipitation across the catchments was considerable, especially during the colder autumn and winter months.

2.2. Study design

To evaluate the ability of the selected bias adjustment methods to reduce biases in various features of the simulated series of precipitation and temperature in Sweden, the following procedure consisting of four steps (each of which is explained in more detail in the sections below) was implemented:

1. A set of 10 different CM outputs of precipitation and temperature in 55 Swedish catchments of varying sizes and climatic features was selected (see Section 2.3 Data Selection).
2. Biases in univariate, multivariate and temporal features were estimated (see Section 2.4 Statistical Features)
3. Two widely used traditional *univariate* BA methods (variations of quantile mapping: distribution scaling and quantile delta mapping) and two prevailing *multivariate* BA methods (copula-based method and n-dimensional multivariate bias correction method, MBCn) were calibrated over 22 years of observations (1961–1982), thus representing the calibration period. The calibrated BA methods were then applied to adjust the biases over the full record period of observations (1961–2004, see Section 2.5 Bias Adjustment (BA) methods).

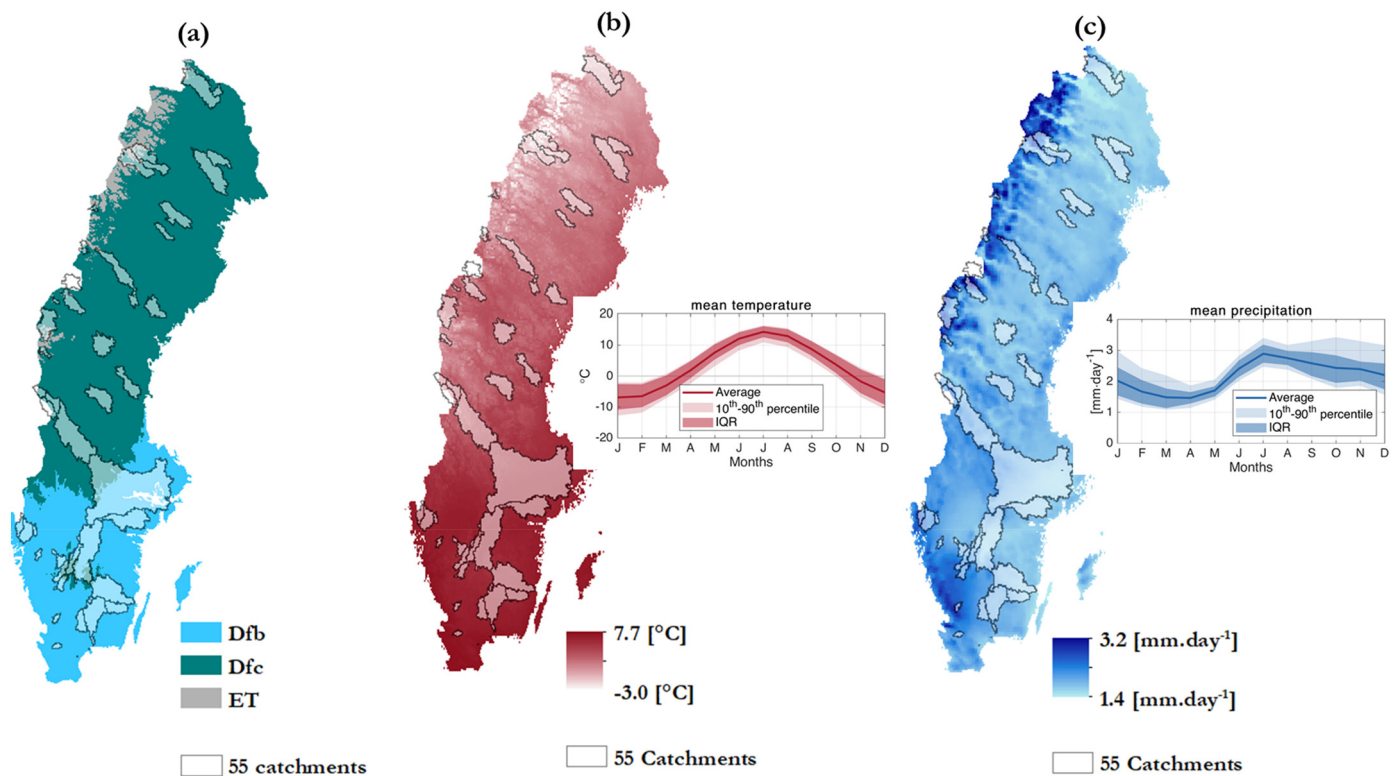


Fig. 1. Location of the 55 Swedish study sites and (a) the spatial extent of the 3 different climate zones covering the country, as well as spatial and seasonal variability of (b) temperature and (c) precipitation over the period 1961–2004. The shaded ranges in the seasonal plots in (b) and (c) show the interquartile range (IQR) and the 10th–90th percentile ranges of all catchments.

4. Performance of the selected BA methods in reducing biases in the considered features (i.e., univariate, multivariate and temporal statistical properties) of the series was assessed during a 22-yr verification period (1983–2004) that was different from the data that BA methods were calibrated to (see Section 2.6 Performance Assessment and Ranking).

2.3. Data selection

2.3.1. Observations (reference data set)

Gridded daily mean values of observed temperature and precipitation were obtained from SMHI's PTHBV database (SMHI, 2005), which provides a spatially interpolated 4 km × 4 km national grid for the period 1961–2020 (Johansson, 2002). Catchment-specific temperature and precipitation values were calculated from area-weighted averages of all grid cells partly or fully lying within the catchment boundaries. Geospatial data of the catchment boundaries for each of the 55 catchments was obtained from SMHI's SVAR database (Eklund, 2011).

2.3.2. Climate model simulations

Daily precipitation and temperature series simulated by an ensemble of 10 combinations of global and regional climate models (detailed in the supplementary material, Table S1) for the historic period 1961–2004 were obtained from the EURO CORDEX initiative (Jacob et al., 2014). We selected only those CMs that had available both the historical and future simulations for all three-greenhouse gas concentration trajectories (i.e., RCP 2.6, 4.5 and 8.5), and in the highest available horizontal resolution of 0.11 degree (roughly 12.5 km). To cover a wide range of uncertainties, we selected simulations from different families of global climate models. If simulations were available in more than one version (e.g., v1 and v2), then the most up-to-date and corrected version was selected. The gridded daily precipitation and temperature data was averaged for each catchment following the same approach as for the averaging of the gridded observations.

2.4. Statistical features of precipitation and temperature series

Following the formal definition of bias by Jung (2005), we measured systematic differences between statistical attributes of climate simulations and observations (Teutschbein and Seibert, 2012). Biases were estimated for (1) univariate, (2) multivariate and (3) temporal statistical features as listed below.

Univariate aspects of climate variables included common summary statistics, such as mean and variance of precipitation and temperature, the 10th and 90th percentiles of temperature, the 90th percentile of precipitation and the number of dry days in a considered period.

Additionally, we analyzed several *multivariate* and *temporal* aspects outlined below that were not directly calibrated when establishing the statistical relationships for each BA method (Maraun and Widmann, 2018a). *Multivariate* features covered the Pearson correlation coefficient ρ (Pearson, 1920) for the entire range of dependence between precipitation and temperature, and the Clausius-Clapeyron (C-C) relation (a measure for the relationship between heavy rainfall and temperature). Pearson correlation was here chosen instead of other rank correlation measures (e.g., Spearman correlation coefficient (Spearman, 1904)), as it can be more robust for data with same ranks (Vrac and Thao, 2020), which was the case for our precipitation data that contained many drizzle (i.e., zero precipitation) days. Detailed information on the concept and implementation of the C-C relation can be found in the supplementary material S2.1. *Temporal* characteristics encompassed the cross-correlation between precipitation and temperature for lags between ± 5 days and the autocorrelation of precipitation time series until a lag of +5 days. Cross-correlation is a common measure to detect the underlying interaction between precipitation and temperature processes, and possible time delays in those. The correct representation of cross-correlation has been shown to be crucial for simulations of streamflow (in particular low flows) and other land-surface interactions (Seo et al., 2019). In contrast, the autocorrelation of precipitation quantifies the persistence of rainfall/snow events and has critical

implications for assessments of drought and flood risks. Furthermore, due to the importance of wet or dry spells for many types of hydroclimatic impact assessments, we followed the suggestion of Maraun et al. (2019) to analyze precipitation transition probabilities. These transition probabilities are conditional probabilities for a state at time t , given the state at time $t-1$ (Wilks, 2006). More specifically, we investigated the two transition probabilities most relevant for hydrological extreme events: (1) P_{11} , a wet day following a wet day, which is relevant for flood events and (2) P_{00} , a dry day following a dry day, relevant for drought events. For a detailed description of the computation of transition probabilities, we refer the reader to the supplementary material S2.2.

2.5. Bias adjustment (BA) methods

We applied a set of four BA methods varying in their complexity and the number of climate variables they consider. Methods used in this study ranged from simple traditional univariate BA methods (Section 2.5.1) to rather advanced, modern multivariate BA methods (Section 2.5.2). These methods were further categorized into (1) so-called distribution-based methods if they were based on known probability distributions (Ivanov and Kotlarski, 2017; Luo et al., 2018), and (2) into distribution-free methods. All four methods are shortly described below. In-depth explanations and equations can be found in the supplementary material, S3.

2.5.1. Traditional univariate BA methods

In the bias adjustment literature, different versions of quantile mapping (QM) are commonly adopted. QM methods are specifically advantageous because of their ability to adjust the overall distribution of the data rather than adjusting just mean or variance. The idea behind these methods is to correct the theoretical probability distribution of CM outputs to agree with probability distributions of the corresponding observations. Here we adopted two versions of QM as benchmark univariate methods to which the multivariate methods were compared to. The benchmark univariate methods were the distribution scaling and quantile delta mapping.

Distribution scaling (DS) is a distribution-based quantile mapping approach introduced by Yang et al. (2010), which is widely adopted in the scientific literature (Gudmundsson et al., 2012; Teutschbein and Seibert, 2012). In this method, particular distribution families are chosen to form a cumulative distribution function (CDF) of variables. The Gamma distribution with shape parameter α and scale parameter β is commonly used to represent the precipitation intensity (i.e., the precipitation amount on wet days) (Teutschbein and Seibert, 2012; Thom, 1951). For temperature, the Gaussian distribution with mean μ and variance σ is assumed to provide a plausible fit (Teutschbein and Seibert, 2012). Therefore, these two distributions were adopted in this paper. For the detailed formulation of the DS approach refer to Teutschbein and Seibert (2012), Eqs. (23)–(26).

Quantile delta mapping (QDM) was first introduced by H. Li et al. (2010). It is a distribution-free combination with the simple delta approach, that explicitly preserves relative changes in simulated CM quantiles (Cannon et al., 2015). QDM adjusts CM output variables in the projection (verification) period first towards the CM output variables in the calibration period, and then towards the observed variables in the calibration period. By comparing QDM to other QM methods, Cannon et al. (2015) showed the merits of QDM to preserve the trends of CM outputs.

2.5.2. Modern multivariate BA methods

In addition to adjustments of univariate characteristics of the CM outputs, multivariate methods also adjust the dependence structure between variables. Different multivariate methods have recently been introduced, among which are copula approaches (e.g., C. Li et al., 2014; Rätty et al., 2018) and MBCn (e.g., Cannon, 2018; Singh and Reza Najafi, 2020), both widely used in impact studies.

Copula adjustment methods rely – as the name indicates – on a copula, which is a mathematical function that links marginal distributions of two or more sets of random variables to represent their joint probability distribution (Sklar, 1959). In the bias-adjustment literature, *empirical* copulas

(e.g., Bárdossy and Pegram, 2012; Vrac and Friederichs, 2015) and distribution-based *Gaussian* copulas (e.g., C. Li et al., 2014; Rätty et al., 2018) are typically used. However, several studies have shown that Gaussian copulas do not always guarantee the best fit, and that other copula families should also be tested (Brunner et al., 2019; Singh et al., 2020; Tootoonchi et al., 2022). Thus, we here tested one elliptical copula (i.e., Gaussian), and three widely used Archimedean copulas (i.e., Clayton, Frank and Gumbel copula), all of which were applied separately on daily precipitation and temperature of each month. Only the best-fitting theoretical copula according to the results of the Cramer-von Mises test (Genest et al., 2009) was used for the bias adjustment of the variables. Whenever the test failed to provide an admissible theoretical copula, the empirical copula was used for modeling the dependence. In this study, empirical copula was used most often (38 %), followed by Clayton (24 %), Gaussian (16 %), Frank (11 %) and Gumbel (10 %) copulas. After adjusting the dependence, the DS method was applied on the margins to adjust univariate characteristics.

It should also be noted that the copula approach cannot directly be applied if the data contains values of the same rank (known as 'ties' in the copula literature), which is typically the case in daily precipitation data that contains occasional dry days (Tootoonchi et al., 2022). Therefore, a method to handle ties should first be applied on the data to produce a dataset free from ties. In this study, a jittering algorithm (following Pappadà et al., 2017) was applied on the data with the same rank to break these ties.

The **n-dimensional multivariate bias correction (MBCn)** method, inspired by image processing techniques, was introduced by Cannon (2018) and has by now been adopted as a benchmark for multivariate bias adjustment performance assessments (Van de Velde et al., 2022; François et al., 2020; Rätty et al., 2018; Singh and Reza Najafi, 2020; Zscheischler et al., 2018). This method is distribution-free and randomly generates orthogonal rotation matrices, and repeatedly applies them to multivariate data. At each rotation, a selected univariate BA method (we used the QDM method) is applied on each of the variables separately to adjust for biases. This iterative procedure is applied until the distribution of the target variables converges to the corresponding distribution for observations according to the energy-distance metric, which is a measure for multivariate similarity. Originally, Cannon coded this approach in R and the algorithm is freely available in the 'MBCn' package (<https://cran.r-project.org/web/packages/MBC/index.html>).

2.5.3. Calibration and verification of BA methods

Both observations and climate simulations were divided into two intervals of 22 years each to allow for a split-sample evaluation (Klemeš, 1983). The split-sample test was preferred over k-fold cross validation (Maraun, 2016) to resemble evaluation protocols typically used by impact modelers (e.g., for the evaluation for hydrological models). This setup was further dictated by the data availability, and by the fact that the empirical copula can only be applied to series of equal length. The first period (1961–1982) was considered as the calibration period to set (calibrate) parameters of the bias adjustment methods. The simulated precipitation and temperature series in the second (verification) period (1983–2004) were adjusted directly based on the parameters estimated during the calibration period. For each period, BA was performed separately for each of the 12 months to preserve seasonal variations. In accordance with common practice, drizzle days with less than 1 mm·day⁻¹ of precipitation were set to dry days (zero precipitation) before evaluation.

2.6. Performance assessment and ranking

Evaluation of the selected BA methods in this paper was based on the similarity between the statistical features of the simulated and observed precipitation and temperature (Gudmundsson et al., 2012). The performance of BA methods was judged based on the results in the verification period, which was generally warmer and wetter than the calibration period. We applied the *t*-test under the null hypothesis of equal means of the data over the two periods at 5 % significance level, which was rejected for

most catchments for both annual mean precipitation and temperature. Thus, the vast majority of the selected data indicates differences in the climatic properties for calibration and verification periods. We, therefore, considered that such a setup enables a comprehensive evaluation of biases in the simulated precipitation and temperature, and fair comparison of the BA methods.

The statistical features considered in this study are elaborated in Section 2.4. In this paper, we utilized the mean absolute error (MAE) as a performance measure of bias in univariate, multivariate and temporal features of the simulated temperature and precipitation series.

The MAE value represents the mean absolute difference between the simulated and corresponding observed statistical features (Eq. (1)). As such, MAE is also suitable for features that take rather small values (i.e., close to zero), which was the case for several statistical features we analyzed. The MAE values were computed as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |R_{obs,i} - R_{sim,i}|, \quad (1)$$

where, $R_{obs,i}$ represents a statistical feature calculated from observations, while $R_{sim,i}$ stands for the statistical feature of the corresponding model simulation, and N is the total number of observations. Lower values of MAE suggest lower biases.

It should be noted that MAE cannot show the sign of bias (e.g., if the simulated mean temperature shows a cold or warm bias). However, we here focused solely on the magnitude of bias (rather than on the sign) and, therefore, considered MAE an appropriate indicator to quantify errors in the statistical features. The MAE values were computed for each statistical feature (Section 2.4) for the raw and for bias-adjusted series simulated by the 10 CMs (Table 1) in each of the 55 catchments, separately for the calibration and verification periods. For univariate metrics and Pearson correlation, biases (MAEs) were computed separately for each month, whereas MAEs in the C-C relation and temporal features were calculated over the entire period under consideration.

While we carefully analyzed the performance of BA methods for both calibration and verification period, we found the results to be mostly consistent across periods and chose to only present results for the independent verification period to ensure consistency, brevity and transferability of the results to future climate conditions (i.e., conditions that the BA methods have not been calibrated to).

To provide a holistic evaluation and fair comparison of the applied BA methods, a final scoring matrix was developed based on the remaining biases using the MAE values that were averaged over all catchments and CM outputs. For the final scoring, MAE of raw CM outputs was taken as benchmark and all other MAE values (obtained after applying the various BA methods) were scaled accordingly by dividing them with the benchmark MAE value. Therefore, the raw CM outputs always received a value of 1, whereas all other methods either received a better score (i.e., below 1) when they reduced the bias or a worse score (i.e., above 1) when they increased the bias. For each category of univariate, multivariate and temporal features, the BA methods were thus ranked based on their remaining biases (i.e., based on the magnitude of MAE), and a final average rank was computed for each method. Additionally, a ranking of the methods' complexity based on their computational times was also included in the evaluation. It should be noted, however, that such a ranking is highly sensitive to coding style and might be speeded up with further coding enhancements.

3. Results

3.1. Statistical features of the observed precipitation and temperature

Observed hydroclimatic features of the selected catchments were calculated over both the calibration and the verification periods, and minimum, mean and maximum values over the catchments for the statistics introduced in Section 2.4 are presented in Table 1. In accordance with their

Table 1

Statistical features of observed precipitation and temperature time series in the selected catchments calculated over the entire calibration (1961–1982) and verification (1983–2004) periods. Note that these values are based on daily values.

Spatial characteristics	Calibration period (1961–1982)			Verification period (1983–2004)		
	Min	Mean	Max	Min	Mean	Max
Univariate statistics						
<i>Temperature</i>						
Mean [°C]	−3.5	2.6	7.2	−2.4	3.5	8.1
Variance [°C ²]	50.3	83.1	133.1	47.34	72.6	120.4
10 th percentile [°C]	−18.8	−9.6	−1.9	−16.8	−7.5	−0.6
90 th percentile [°C]	9.2	13.5	16.2	10.1	14.2	16.9
<i>Precipitation</i>						
Mean [mm·day ^{−1}]	1.4	2.0	3.0	1.5	2.2	3.4
Variance [mm ² ·day ^{−2}]	8.5	13.9	24.3	9.2	16.9	30.9
The 90 th percentile [mm ² ·day ^{−1}]	4.2	6.1	9.	4.6	6.8	9.8
Number of dry days per year	33.6	96.0	169.8	37.9	94.0	171.6
Multivariate statistics						
Pearson correlation between daily variables [−]	0.03	0.09	0.17	0.03	0.10	0.16
C-C relation [%/°C ^{−1}]	0.3 %	2.5 %	6.3 %	0.07 %	3.2 %	4.7 %
Temporal statistics						
Cross-correlation lag +1 day [−]	−0.04	0.20	0.37	−0.02	0.24	0.38
Autocorrelation of precipitation lag +1 day [−]	0.35	0.45	0.57	0.35	0.43	0.56
Dry-day to dry-day transition probability (P ₀₀) [−]	0.42	0.57	0.68	0.42	0.55	0.69
Wet-day to wet-day transition probability (P ₁₁) [−]	0.72	0.84	0.93	0.72	0.83	0.93

wide geographical spread across the country, the study catchments showed a wide range of different hydroclimatic characteristics.

3.1.1. Univariate statistical features of precipitation and temperature

During the calibration period (1961–1982), mean temperature ranged from −3.5 °C in the North to +7.2 °C in the South. Temperature featured a clear north-south pattern in its univariate characteristics (not shown here). Generally, all temperature-related features were relatively higher in the southern catchments, and lower in the northern ones, except for the variance, which exhibited the opposite behavior. The mean precipitation varied from 3.0 mm·day^{−1} in the Scandinavian mountains (North-Western Sweden) to 1.4 mm·day^{−1} in the drier southern regions (not shown here). During the later verification period (1983–2004), all temperature-related statistics increased, except for the variance. The average annual temperature was considerably warmer as it increased by roughly 1 °C in the selected catchments, while precipitation rose slightly by +10%. Although the underlying causes of these shifts from calibration to verification period cannot clearly be attributed to man-made climate change (given the relatively short length of these periods) and could potentially be caused by internal climate variability, the verification period can be considered a reasonable proxy for future climate conditions, as the observed changes in mean temperature and precipitation are within the range of likely climate-change scenarios in Sweden (IPCC, 2014).

3.1.2. Multivariate statistical features of precipitation and temperature

Dependence between daily temperature and precipitation (here expressed in terms of the Pearson correlation between daily variables, Table 1) was weakly positive in all catchments during both simulation periods with values between 0.03 and 0.17. However, dependence also showed considerable monthly variations (Fig. 2a), with generally negative values during the months of May to August and positive ones from October to March. The spatiotemporal pattern detected during the calibration period (Fig. 2a, left) generally persisted over the verification period (Fig. 2a, right), but slight shifts occurred in some months. For example, the dependence in July became weaker in the southern catchments, while positive dependences in February and October became stronger.

The Clausius-Clapeyron (C-C) relation between the precipitation greater than 95th percentile and the temperature at which it occurred had an average of 2.5 %/°C in the calibration period, and increased to 3.2 %/°C in the verification period (Table 1). This increase, however, was not detected in all catchments (denoted by the decrease in the minimum and maximum

values in Table 1). Furthermore, The C-C relation did not feature a pronounced north-south gradient.

3.1.3. Temporal statistical features of precipitation and temperature

Cross-correlation of precipitation and temperature was assessed for lags between −5 days to +5 days (Fig. 2b). The overall pattern was consistent across all lags from −5 to +5: positive cross-correlation was higher in the southern catchments and slightly increased in the verification period. In the northern catchments, cross-correlation was weaker or even negative in some catchments across all lags. In a few northern catchments, a shift towards positive values in the verification period was observed.

Precipitation autocorrelation (Fig. 2c) took values between 0 and 0.57 with decreasing values with increasing lag length. Generally, higher values were found in the northern catchments in all lags (Fig. 2c). No clear alteration in the magnitude was observed in the verification period (Table 1).

The dry-to-dry-day transition probability (P₀₀) took values between 0.4 and 0.7, with slightly lower values in the northern catchments in both periods. These probabilities decreased over the verification period (Fig. 2d). The wet-to-wet-day transition probability (P₁₁) was higher than P₀₀ (Fig. 2e), indicating a higher probability of two consecutive rainy days than two consecutive dry days. The values of P₁₁ were between 0.72 and 0.93 in both periods (Table 1), without any distinct spatial pattern, including the north-south gradient. There was also no clear difference in this probability between the calibration and verification periods (Fig. 2e).

3.2. Bias estimation (raw CM outputs)

To quantify biases in the simulated precipitation and temperature series and to enable comparison of the four BA methods, we calculated MAE from the univariate, multivariate and temporal statistics of these series, by comparing them to the corresponding features of the observations (as explained in Section 2.4). We evaluated the BA methods over the verification period (1983–2004) and, for the sake of brevity, presented biases in the raw CM outputs only in that period.

3.2.1. Univariate statistical features of precipitation and temperature

Biases in univariate characteristics of raw simulated temperature (Fig. 3a–d) and precipitation (Fig. 3e–h) showed considerable seasonal variations, and, in many instances, large spread across the CMs.

Simulated monthly mean temperatures (Fig. 3a) featured the largest (absolute) biases (i.e., MAE) in spring with values up to 7.27 °C in April.

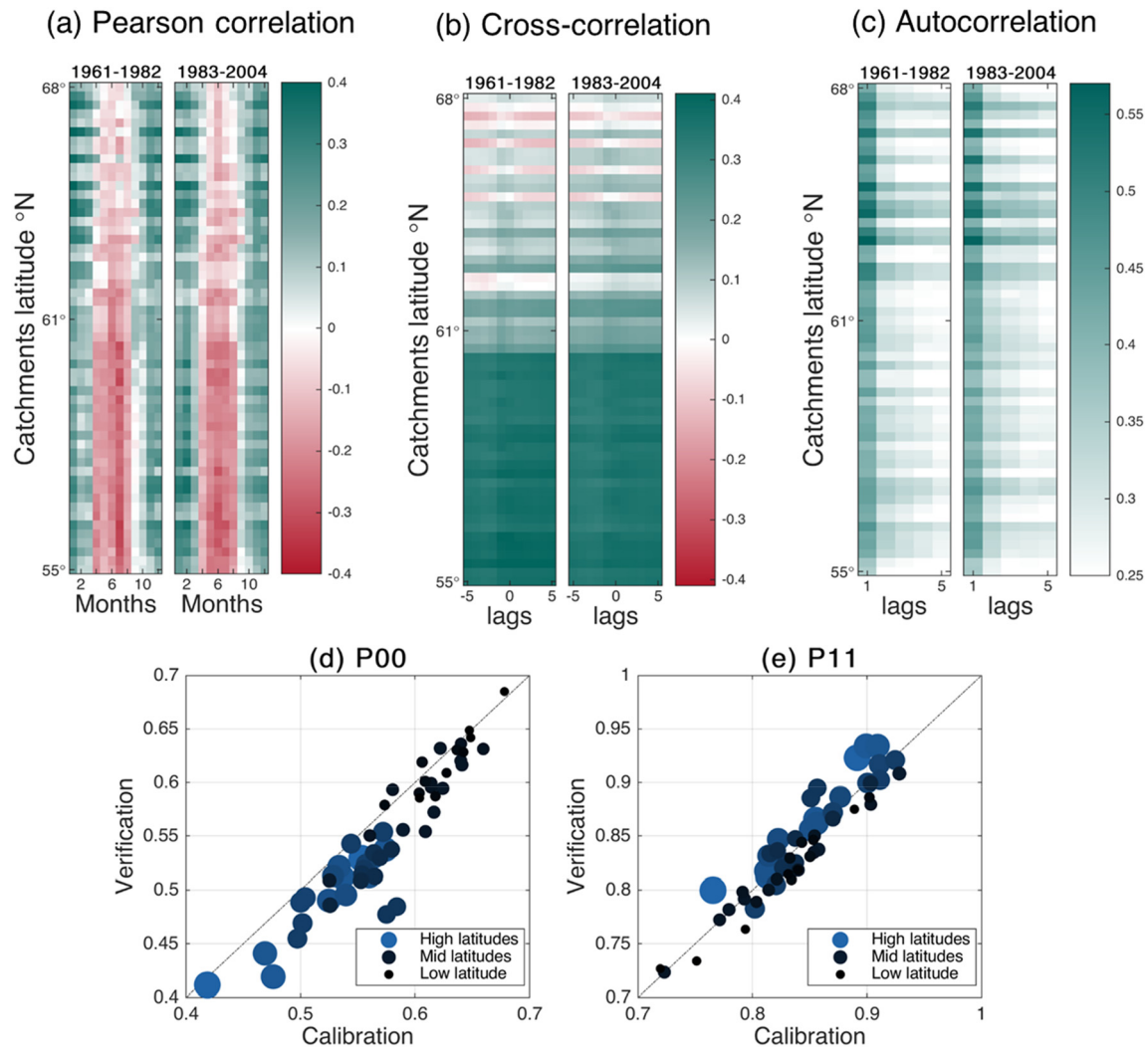


Fig. 2. Multivariate and temporal statistical features in the 55 catchments over the calibration and verification periods: (a) Pearson correlation coefficients obtained from daily precipitation and temperature and is shown for different months, (b) cross-correlation between daily precipitation and temperature from lag -5 to lag $+5$ days, (c) autocorrelation of precipitation for lags 1 to 5 days, (d) transition probabilities for dry-to-dry days (P_{00}), (e) transition probabilities for wet-to-wet days (P_{11}). Latitude of the catchments is indicated on the y-axis in panels (a)–(c).

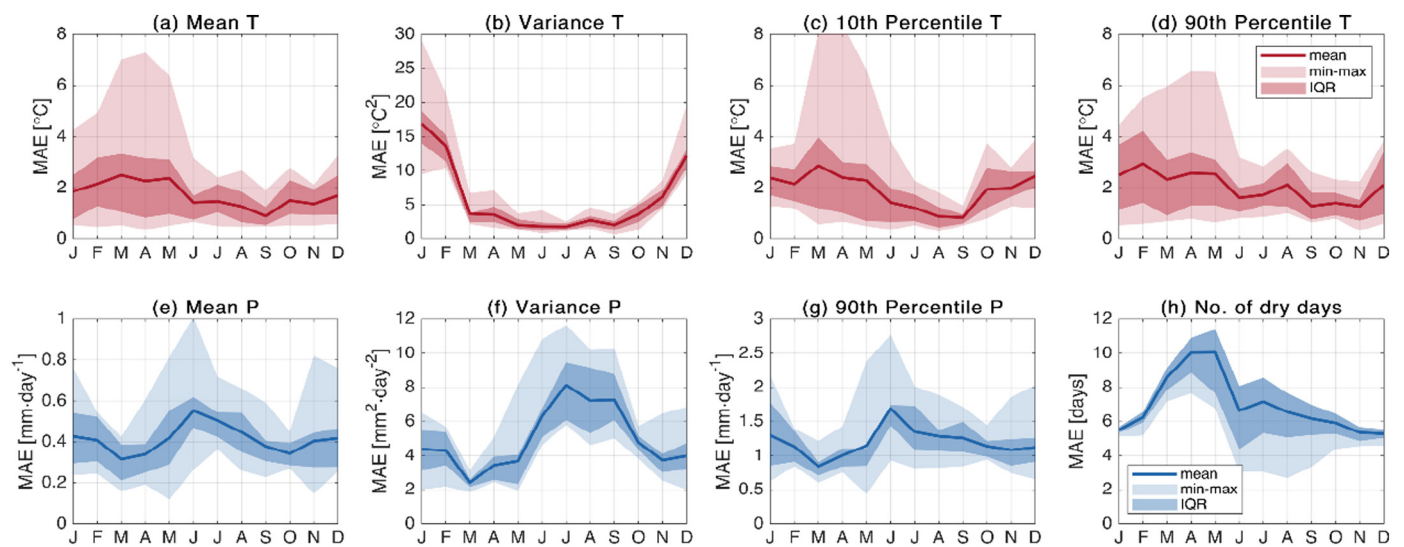


Fig. 3. The MAE values in the univariate statistics obtained for each month from the raw simulated temperature (upper row) and precipitation (lower row). The ranges show the spread (minimum and maximum values, as well as the interquartile range, IQR) in MAE across the 10 CMs after averaging MAE values over all 55 catchments in the verification period.

During the first half of the year (January–May), simulated temperatures exhibited both higher uncertainty and larger average bias than in the remaining months (Fig. 3a). Bias in the variance of the simulated temperature (Fig. 3b) exhibited a clear seasonal pattern as well, featuring both larger average biases and a larger spread during winter (December–February) than in the period from March to November. The largest biases were detected in January with values up to 29°C^2 . The ensemble mean of MAE in the 10th and 90th percentiles of temperature (Fig. 3c, d) generally had similar values and followed a similar seasonal pattern as the mean temperature (Fig. 3a). Both percentiles featured a larger spread of biases in CM outputs in the first half of the year, especially during spring (the highest values were detected in April). However, biases in the 10th percentile of temperature (Fig. 3c) featured a larger spread (with maximum values up to 8.36°C in April) than biases in the 90th percentile (Fig. 3d, maximum values were up to 6.3°C in April).

For mean precipitation, the largest biases (up to $1\text{ mm}\cdot\text{day}^{-1}$) as well as pronounced uncertainty of CM outputs were detected during warmer and wetter summer months (in June in particular) (Fig. 3e). On average,

however, the lowest biases and lowest variability were observed in 2 months of different seasons: March and October. Similar to the mean precipitation, variance had larger biases (as large as $11.63\text{ mm}^2\cdot\text{day}^{-2}$) and higher variability during summer (Fig. 3f). Extremely high precipitation (i.e., the 90th precipitation percentile) yielded considerably higher biases than the mean precipitation (nearly $3\text{ mm}\cdot\text{day}^{-1}$). The largest MAE values were found June–September, while the highest spread of biases in CM outputs was observed in the period May–July (Fig. 3g). The bias in CM-simulated number of dry days (i.e., days with precipitation less than 1 mm) was up to 12 days in the spring, while the variability was generally highest during the summer (Fig. 3h).

3.2.2. Multivariate statistical features of precipitation and temperature

To describe biases in multivariate characteristics in detail, the Pearson correlation coefficient between precipitation and temperature, and the C-C relation for the uncorrected CM output were compared against the corresponding features obtained from the observations (Fig. 4a, c) and quantified in terms of MAE (Fig. 4b, d).

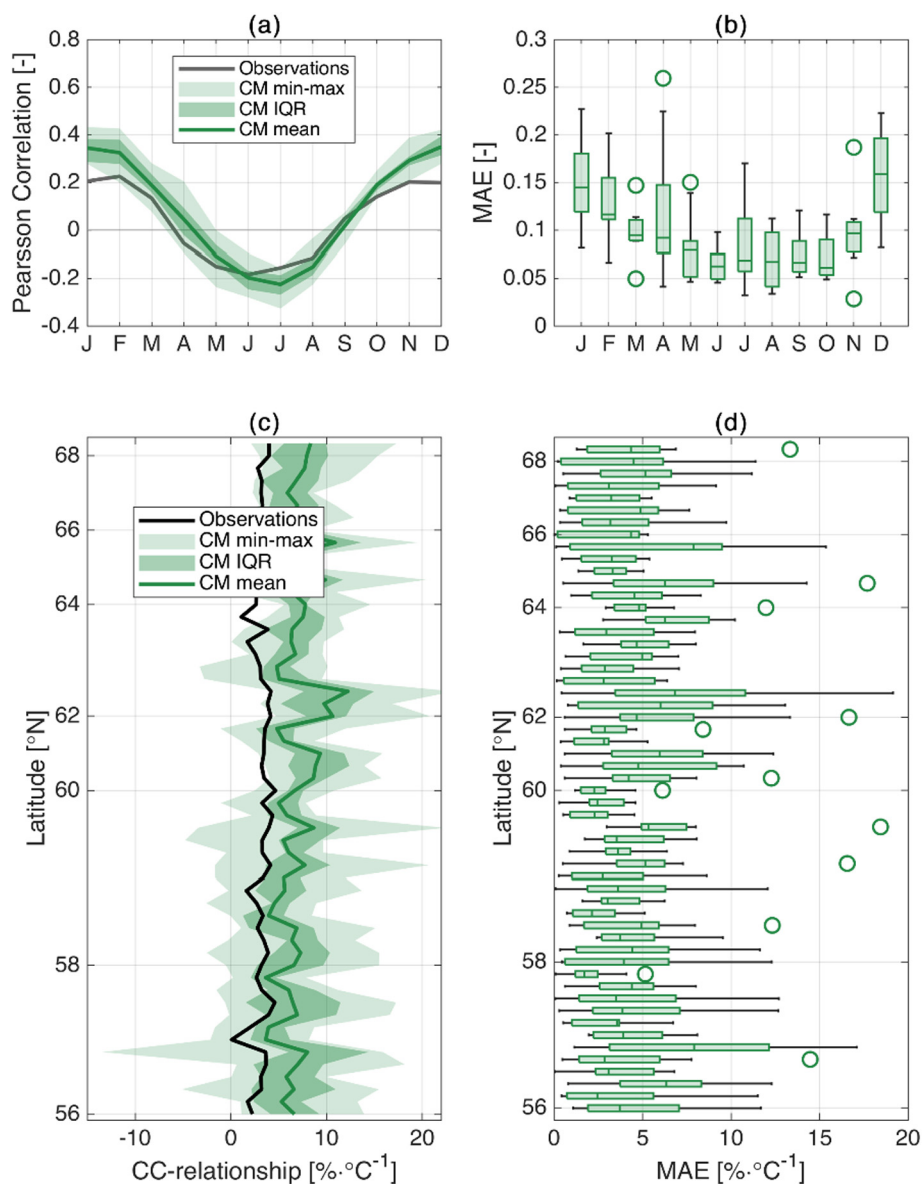


Fig. 4. Overview of the multivariate statistics obtained from the observed and raw simulated temperature and precipitation series. The upper row shows the seasonal patterns in (a) Pearson correlation coefficient obtained from raw CM outputs and observations, and (b) the corresponding MAE values. The ranges show the spread across the 10 CMs after averaging over 55 catchments. The lower row shows the spatial pattern (c) of the C-C relation obtained from raw CM outputs and observations, and (d) corresponding MAE values. Ranges show the spread across the 10 CMs in the verification period.

Observed correlation coefficients across the catchments exhibited a pronounced seasonal pattern, with negative correlation in the warm season (May–August) and positive correlation in the cold season (October–March). This pattern was generally reproduced by the CMs (Fig. 4a), with exception of a slight shift in negative correlation towards the end of the warm season. The correlation of CM outputs was somewhat stronger (i.e., more extreme) than the correlation featured by the observations (Fig. 4a). Stronger correlations in CM outputs could specifically be observed for the period November–February (positive correlation) and in July (negative correlation). This pattern of over-/underestimation was also reflected in the MAE (Fig. 4b), which demonstrated the largest median biases in the winter months (December through February) and the lowest biases in the summer (June through August). The greatest uncertainty (i.e., spread across the CMs) was observed in April.

The relationship between extreme precipitation and the temperature at which it occurred was mostly overestimated by CMs (Fig. 4c). Biases were overall high, with median MAE values across the catchments ranging from 1.7 %/°C to 7.9 %/°C (Fig. 4d).

3.2.3. Temporal statistical features of precipitation and temperature

Considerable disagreement between the temporal characteristics of raw simulated and observed series was detected (Fig. 5). The observations displayed the highest cross-correlation between precipitation and temperature at lag -1, but also continued to show relatively high values for lags 0 to 5 (Fig. 5a). In contrast, raw CM outputs consistently demonstrated highest cross-correlation at lag 0 (Fig. 5a), and consistently underestimated the observed cross-correlation at all lags. However, the lowest biases were found at lag 0 (Fig. 5b). The spread of MAE at different lags did not exhibit strong variations, spanning from 0.02 to 0.12 over all lags (Fig. 5b).

Concerning autocorrelation of precipitation over different lags, CM outputs mostly overestimated the observations, particularly at higher lags

(here of maximum 5 days, Fig. 5c). While no specific pattern could be detected in MAE over different lags (Fig. 5d), autocorrelation featured lower average biases and variability compared to cross-correlation.

For the simulated transition probabilities, MAE took relatively high values. Dry-to-dry-day transition probability (P_{00}) (Fig. 5e) featured considerably higher biases (almost twice as high) compared to wet-to-wet-day transition probability (P_{11}) (Fig. 5f). No clear north-south gradient could be detected in the MAE value (Fig. 5e, f).

3.3. Performance assessment of the BA methods

3.3.1. Univariate statistical features of precipitation and temperature

All BA methods considerably reduced MAE in the univariate attributes of temperature (Fig. 6a–d). Across all univariate features (i.e., mean, variance, the 10th and 90th percentiles), adjustments done through distribution-based BA methods (i.e., DS and copula) featured a lower median and spread of bias. For the mean temperature, the spread in MAE and especially the most extreme biases are reduced considerably, from a maximum bias of 7.8 °C to a maximum of approximately 4 °C with the distribution-free BA methods (i.e., QDM and MBCn), and to a maximum bias of less than 2 °C with the distribution-based DS and copula method (Fig. 6a). The median of MAE in the variance of temperature (Fig. 6b) was somewhat reduced, with better performance of the distribution-based BA methods (i.e., DS and copula). The spread of MAE, however, showed a considerable decrease from 15 °C in raw CM outputs to roughly 6.2 °C after application of DS and copula, and approximately 8.1 °C via the QDM and MBCn BA methods. While MAEs in both raw 10th and 90th percentiles of temperature (Fig. 6c, d) were roughly as large as MAE in the mean temperature (Fig. 6a), the remaining biases after bias adjustment were slightly larger compared to corresponding MAEs in mean temperature, i.e., application of the BA methods did not reduce biases in extreme temperatures to the same degree as

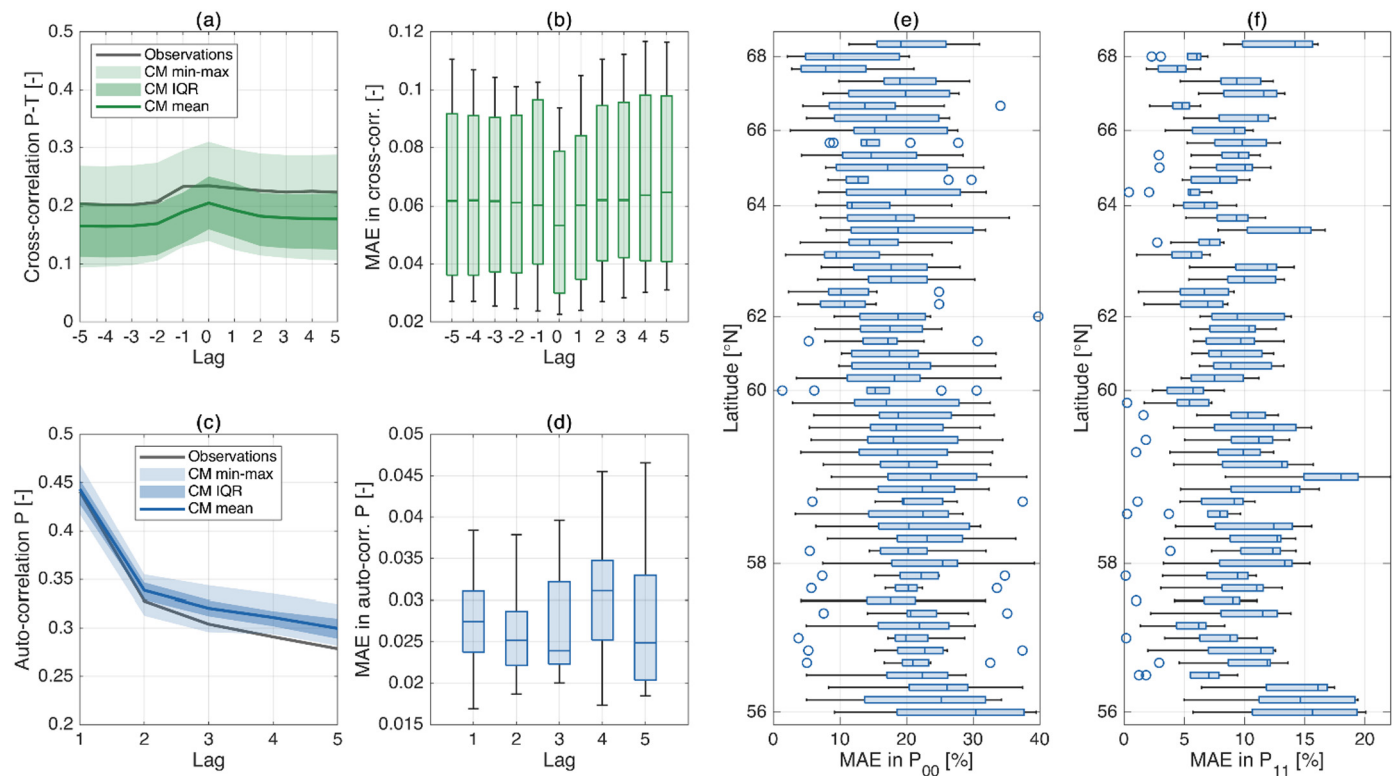


Fig. 5. Temporal statistics of the observed and raw simulated precipitation and temperature series: (a) raw CM outputs versus observed cross-correlation between precipitation and temperature, and (b) the corresponding MAE values, (c) raw CM outputs versus observed autocorrelation in precipitation, and (d) the corresponding MAE estimates. The two figures on the right-hand side show spatial variability of MAE in (e) dry-to-dry-day transition probability P_{00} , (f) wet-to-wet-day transition probability P_{11} . The ranges show the spread in the CM outputs (a and c), MAE (b and d) across 10 CMs after averaging MAE over all catchments and temporal variability of MAE across the catchments (e and f), in the verification period.

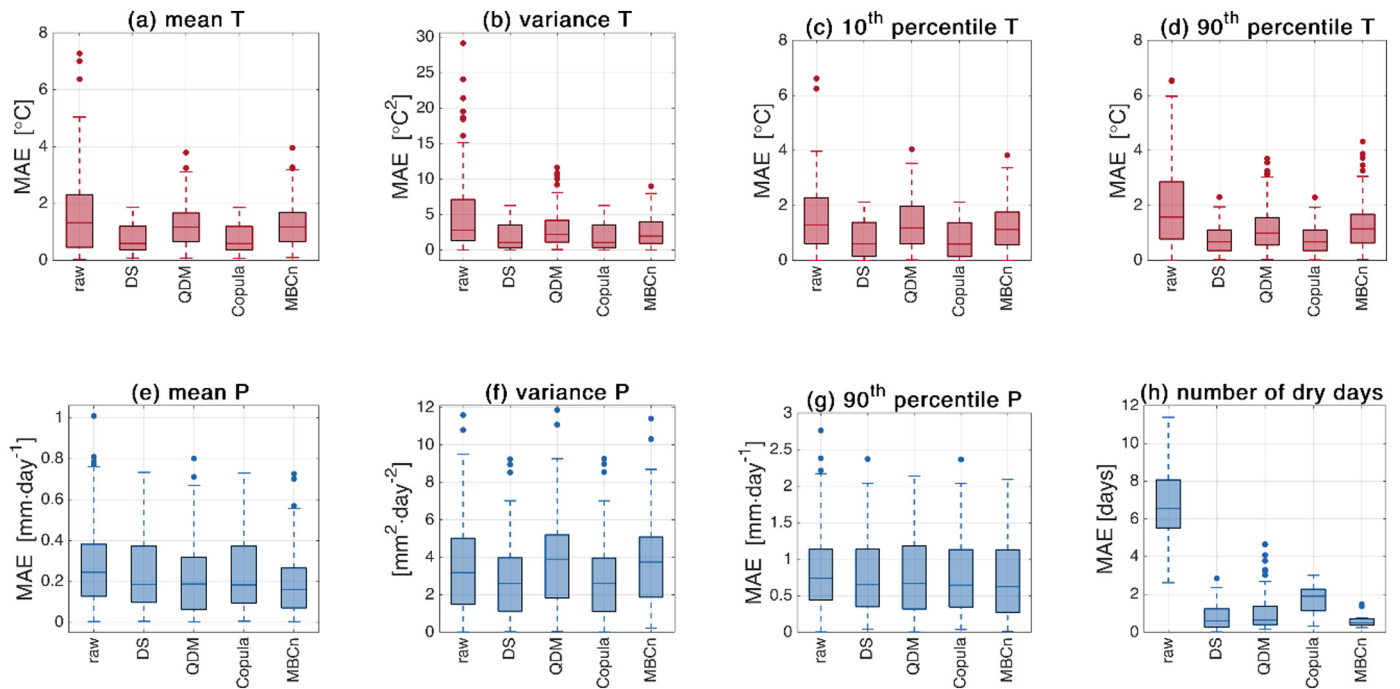


Fig. 6. MAE in the univariate statistics of raw CM outputs and bias adjusted temperature (upper row) and precipitation (lower row). The ranges show the spread in MAE across the 10 CMs after averaging MAE over all catchments and 12 months in the verification period.

in mean temperatures (Fig. 6c–d). When comparing MAE of the 10th percentile to the 90th percentile, the former exhibited a slightly higher spread of bias with all BA methods (Fig. 6c–d).

The performance of the BA methods for precipitation (Fig. 6e–h) varied across the statistical features and the methods. In the case of mean precipitation, all methods reduced the median MAE to some extent, but the remaining biases featured a considerable spread across the CMs (Fig. 6e). Outperforming the other three BA methods, MBCn yielded the lowest median bias, which decreased from 0.24 mm·day⁻¹ in raw CM outputs to roughly 0.16 mm·day⁻¹ after the adjustment. Variance of precipitation (Fig. 6f), on the other hand, did not exhibit any improvements after the application of QDM or MBCn, with the overall bias even increasing. However, both the spread and median of bias was reduced after adjusting the results with the distribution-based methods DS and copula, reducing the median of bias from 3.18 mm²·day⁻² in raw CM outputs to roughly 2.6 mm²·day⁻². Similar to mean precipitation, biases in extreme precipitation were generally reduced from 0.73 mm·day⁻¹ in raw CM outputs to roughly 0.65 mm·day⁻¹ after application of the BA methods (Fig. 6h). All BA methods considerably reduced biases in the number of dry days per month (Fig. 6h). Raw CM outputs ranged from approximately 2 to 12 days, while biases after the adjustment ranged from 0 to 4 days. The best result was obtained with the MBCn method that reduced the biases to almost zero days. Unlike previous statistics, however, there was no distinct difference between the distribution-based and distribution-free methods.

3.3.2. Multivariate statistical features of precipitation and temperature

As highlighted earlier (Section 3.2.2), the raw CM outputs (Fig. 7a) overestimated the magnitude of correlation in summer and winter, which led to large biases across the year, especially in the colder seasons (Fig. 7b). The two univariate BA methods DS (Fig. 7c, d) and QDM (Fig. 7e, f) were not able to noticeably reduce the existing biases throughout the year. Only the multivariate BA methods based on the copula approach (Fig. 7g, h) and MBCn (Fig. 7i, j) considerably reduced biases in the correlation over different months, but also reduced the spread across the CM outputs.

The copula method resulted in the lowest MAE values and the lowest variability both across the CM outputs (Fig. 7g) and across the seasons (Fig. 7h), with biases in some months even been completely removed after the

adjustment. The copula-corrected CMs generally followed observational variations in the calibration period (Fig. 7g) and the original signal from the raw CM outputs seemed to be almost completely lost, which was apparent in the weaker correlation in February and October. MBCn brought the raw CM outputs correlation closer to those of the observations (Fig. 7i), which resulted in overall lower MAE values (Fig. 7j). In contrast to the copula method, MBCn still preserved the original variation in MAE produced by the CMs (Fig. 7j).

When averaged over all months of the year (Fig. 7k), the differences among BA methods became more apparent: the univariate methods (DS and QDM) only slightly reduced existing biases, whereas multivariate methods (copula and MBCn) reduced the MAE in correlation from a median of 0.075 in raw outputs, down to 0.026 (copula) or 0.037 (MBCn).

The estimated C-C relation after bias adjustment (Fig. 8a) was generally improved after the application of all four BA methods. Corrected CM outputs through both univariate methods (DS and QDM) and one multivariate method (MBCn) generally resulted in a slight overestimation of the C-C relation (Fig. 8a), whereas copula somewhat underestimated the C-C relation, specifically in the northern catchments. While all four methods were able to reduce biases by lowering the median of MAE (Fig. 8b), multivariate methods succeeded to reduce MAE the most, with copula outperforming MBCn by both lowering the spread of error and the median of MAE from 4 %/°C in raw simulations to 1 %/°C in the corrected ones (Fig. 8b).

3.3.3. Temporal statistical features of precipitation and temperature

BA methods showed substantial differences in changing the biases of temporal features (Fig. 9). After applying the univariate BA methods, cross-correlation between precipitation and temperature (Fig. 9a) generally showed lower MAE values. In contrast, MAE increased from a median of 0.06 in raw CM outputs to approximately 0.09 in those corrected with the multivariate methods (Fig. 9a).

Autocorrelation of precipitation (Fig. 9b) exhibited less bias in comparison to the cross-correlation (Fig. 9a). Application of all BA methods except for DS resulted in some decrease of biases (Fig. 9b). Only the application of DS caused an increase of MAE from a median value of 0.012 in raw CM simulations to 0.019 in the adjusted ones.

All four BA methods substantially reduced biases in both transition probabilities P_{00} and P_{11} (Fig. 9c–d). Corrected CM outputs with univariate

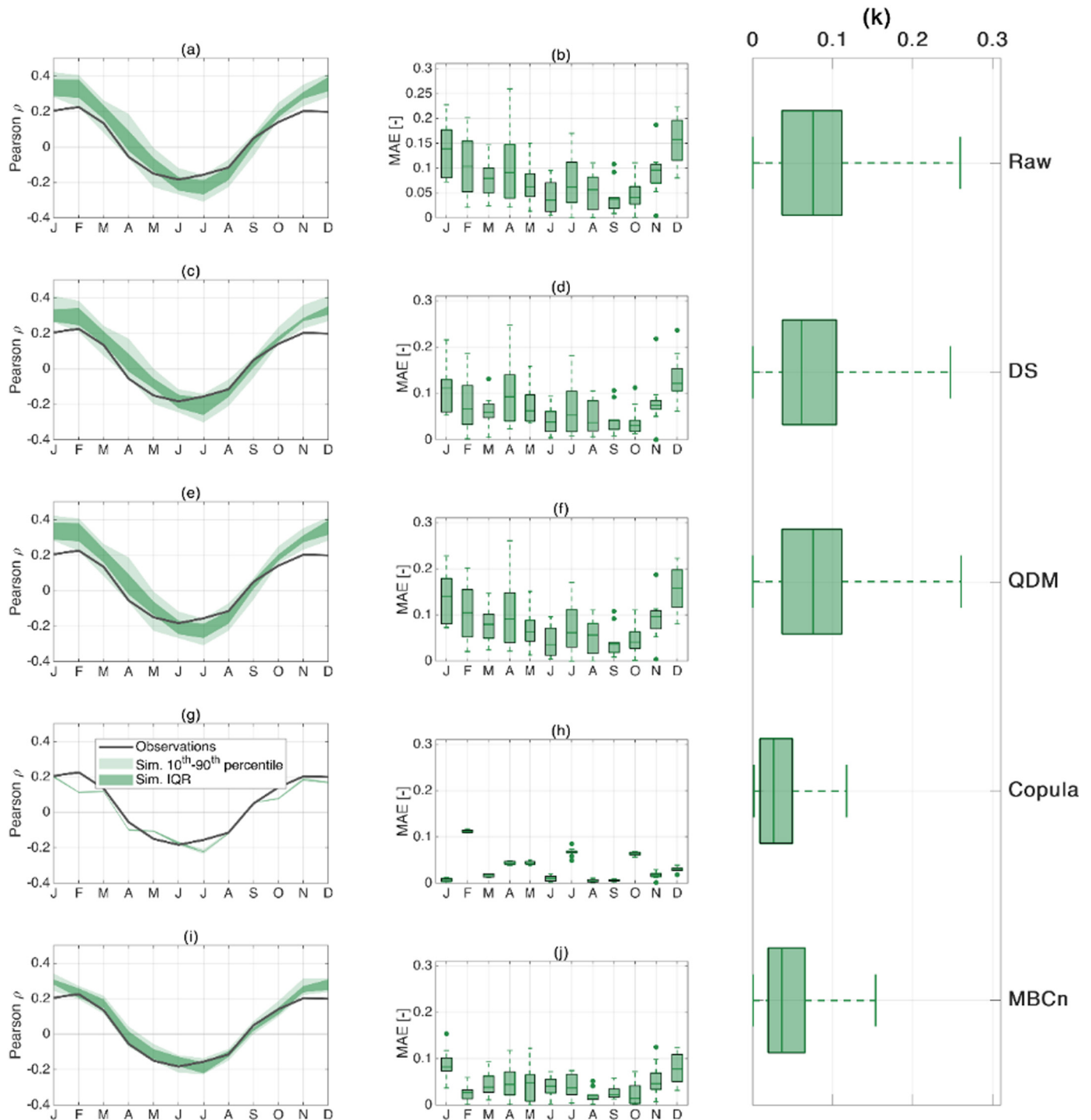


Fig. 7. Ability of BA methods to adjust the bias in dependence between temperature and precipitation. Left panels (a, c, e, g, i) show the seasonal dependence patterns in raw and bias-adjusted CM outputs versus observations, central panels (b, d, f, h, j) show the resulting MAE values. Right panel (k) indicates overall performance after averaging over all catchments for all months and CMs, in the verification period.

methods (DS and QDM) were associated with a higher spread of MAE in comparison to corrections done through multivariate methods (copula and MBCn). In both P_{00} and P_{11} as well as across all BA methods, the copula-based BA method featured the highest median, but the lowest spread of MAE (Fig. 9c, d).

3.3.4. Overall performance and ranking

All BA methods considerably reduced biases in univariate temperature metrics (Fig. 10a). It should be highlighted, though, that distribution-

based methods (univariate DS and multivariate copula) consistently outperformed QDM and MBCn (Fig. 10a).

For precipitation, the improvements were less pronounced. The distribution-based BA methods (DS and copula) were again able to reduce biases in all univariate metrics more than their distribution-free counterparts (QDM and MBCn). In case of variation, QDM and MBCn even slightly increased biases (Fig. 10a).

Concerning the multivariate features, univariate BA methods (DS and QDM) only slightly modified biases in both the dependence

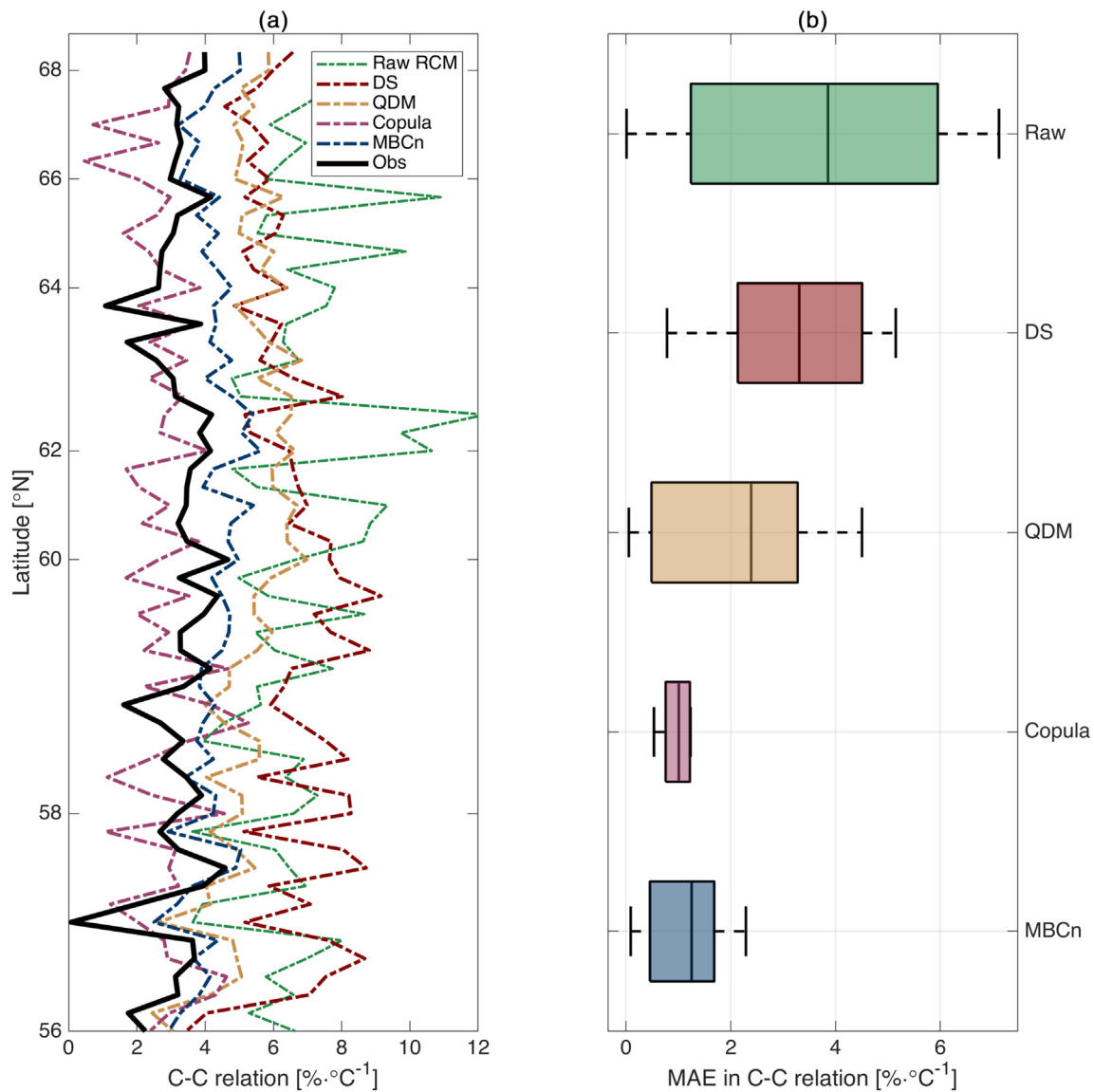


Fig. 8. Ability of BA methods to reduce the bias in the C-C relation: (a) spatial variability of the C-C relation obtained from raw and bias-adjusted CM outputs (averaged over the 10 CMs in each catchment) in direct comparison to those by the observations, and (b) the MAE values. The ranges of boxplots show the spread across the 10 CMs after averaging over all catchments.

structure and C-C relation (Fig. 10b), while both multivariate BA methods considerably reduced MAEs, with copula outperforming MBCn.

BA methods exhibited noticeable differences in their ability to modify temporal features (Fig. 10c). For cross-correlation, both multivariate methods resulted in larger biases than raw simulations, whereas for

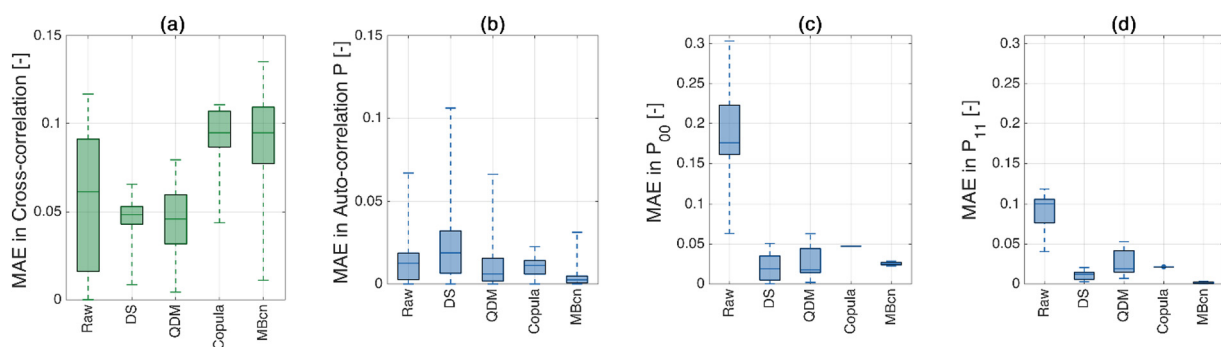


Fig. 9. MAE in temporal features of the raw and bias adjusted CM outputs variables in the verification period. The ranges show the spread in MAE across the 10 CMs, after averaging the MAE values over all catchments in the verification period. In (a) and (b) values were additionally averaged over all lags.

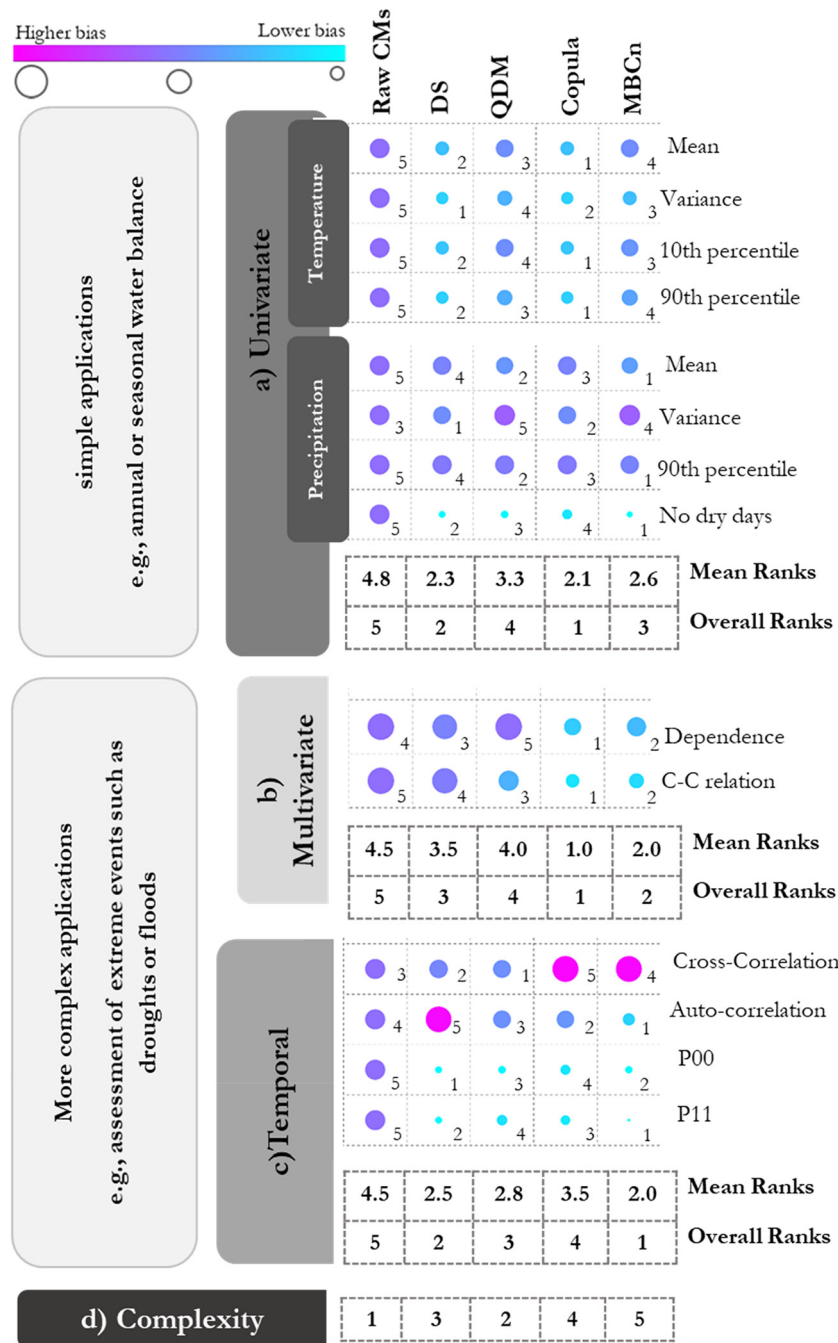


Fig. 10. Holistic performance of all BA methods in comparison to raw CM outputs across numerous statistical features of simulated precipitation and temperature series categorized into (a) univariate, (b) multivariate and (c) temporal features. Color and size of the circles showing raw CMs (left column) are fixed to the same value of one, while circle size and color of the bias adjusted CM represent the remaining biases relative to that. Bias reductions (good performances) are shown by smaller blue circles, bias amplification (poor performance) by larger pink circles. Performance ranks are included in the lower right corner of each BA method. Overall efficiency of each BA method for reducing biases for each category of (a) univariate, (b) multivariate and (c) temporal features was quantified by averaging their performance ranks within each category and shown together with the ranks for (d) complexity.

autocorrelation biases only increased with DS (Fig. 10c). Biases in both transition probabilities (P_{00} and P_{11}) consistently decreased after application of all four BA methods, with the copula approach performing slightly worse than the other methods (Fig. 10c).

Differences in the complexity of the underlying assumptions and statistical concepts of each BA method were also reflected in the time to conceptually understand and implement each BA method (not directly measured here), and in the computational costs to run them (Fig. 10d). Computations of the two simpler univariate methods were considerably faster, with QDM taking only 6 seconds for all catchments and DS twice as much. While the

run-time of the multivariate copula approach (30 seconds) had a similar order of magnitude, MBCn stood out by requiring approximately 120 times more time to run.

4. Discussion

4.1. Characteristics of observed variables

Our analysis of observed univariate characteristics of temperature and precipitation, their multivariate features, and their temporal aspects

highlighted a large variability across the study sites. Mean temperature followed the prevailing north-south climate gradient with colder temperatures in the North and warmer in the South. Precipitation exhibited a more heterogeneous behavior in different sites across the latitudes, but also in monthly or daily temporal resolutions. Heterogeneous variability of precipitation can be explained by natural variability of precipitation and its response to local forcings (A. Berg et al., 2015), intrinsic complexity of precipitation process (Franzke et al., 2020), and significant heterogeneity of humidity over entire Sweden (Lind and Erik, 2008) that contributes to precipitation events.

Both the magnitude and the north-south gradient of correlation between precipitation and temperature observed in our study were in accordance with other studies, e.g., C. Li et al. (2014), while the C-C relation for extreme precipitations deviated from the theoretical relation of 6–7 %/°C (Lenderink and Van Meijgaard, 2008; Myhre et al., 2019). Such deviations from the theoretical C-C value are a matter of discussion in numerous studies (e.g., Aleshina et al., 2021; P. Berg et al., 2009; Hardwick Jones et al., 2010; Panthou et al., 2014; Singh et al., 2020). For example, Hardwick Jones et al. (2010) argue that in high temperatures, the amount of extreme precipitation decreases because it depends not only on the water holding capacity of the atmosphere, but also on the available moisture, which naturally decreases at higher temperatures.

Temporal features of precipitation and temperature typically followed a clear north-south gradient. Higher values of autocorrelation were found in northern catchments, which form a pattern that was also observed over the UK (Wilby et al., 2003). This might be a consequence of different dominating types of precipitation, e.g., cyclonic atmospheric conditions at the large-scale versus local convective systems (Stehlik and Bárdossy, 2002). Cross-correlation and dry-to-dry-day transition probabilities showed the opposite pattern by being stronger in southern catchments. Temporal variability of hydroclimatic variables is highly affected by the spatiotemporal scale at hand (Franzke et al., 2020), but has essential implications for long-term dry or wet spells, which play a key role in vegetation growing season or the hydrological cycle.

We detected an overall warming and wetting trend during the study period 1961–2004, which is in line with other studies in this region (e.g., Nygren et al., 2021; Teutschbein et al., 2022). In agreement with previous studies, we found a declining variance in mean temperature (Tamarin-Brodsky et al., 2019) and an increasing variance in precipitation (Dore, 2005). However, this aspect needs further evaluation, because the behavior of variance is rather heterogeneous and strongly depends on the chosen temporal resolution and the length of study periods (Lewis and King, 2017). Our results also indicate an increasing value of the C-C relation, which might, according to Posch and Ludwig (2021), be a direct result of a shift from less intense large-scale events towards more intense convective events (Lenderink and Van Meijgaard, 2009). But this change in the ratio between the two types of precipitation did not alter the temporal characteristics (i.e., cross-correlation, autocorrelation and the transition probabilities) for all of our catchments.

The observed changes in temperature and precipitation during our study period were of comparable magnitude to the projected future trends in Sweden (IPCC, 2014). Thus, regardless of the underlying causes for these changes, our chosen verification period can be considered as a reasonable proxy for future climate conditions in our evaluation, serving as a basis for testing the ability of BA methods to correct several statistical features in climate conditions different from those that they were calibrated to.

Evaluating BA methods only with respect to their ability to reproduce marginal aspects (i.e., univariate statistics) of the simulated series increases the risk of not being able to identify unskillful BA methods (Maraun, 2016). We, therefore, thoroughly analyzed performances of the BA methods with respect to numerous aspects, including also those that the BA methods were not calibrated to reproduce (e.g., multivariate for DS and QDM or temporal for all four methods). Thus, we provided a rather robust and reliable performance assessment of the BA methods. Robustness of the evaluation in our study was further increased by including catchments with a wide

variability in all features of the observed precipitation and temperature, as well as in the changes in them between the two periods.

4.2. Bias estimation (raw CM)

The raw simulated temperature and precipitation with the 10 CMs used in this research showed substantial biases in their univariate, multivariate and temporal characteristics. These biases are mainly caused by imperfect representations of processes within CMs (Jaeger et al., 2008; Kotlarski et al., 2005; Teutschbein and Seibert, 2010, 2012), for instance atmospheric convection, cloud-microphysical, and aerosol behavior (IPCC, 2014).

Similar to previous studies in Sweden (Lind and Erik, 2008; Olsson et al., 2016), temperature showed considerable bias, more specifically over winter.

For precipitation, CM outputs featured considerable biases in all univariate features, including their magnitudes and, occasionally, seasonal variations. In fact, CMs face major challenges in capturing the spatiotemporal pattern and extreme precipitation at regional or smaller (e.g., catchment) scales (Lind and Erik, 2008), because of significant heterogeneity of humidity and dependence of precipitation processes on local characteristics. In our study, the number of dry days was consistently underestimated by all CMs. This is a known issue with climate simulations called *drizzle effect*, which is a consequence of CMs producing several days with close-to-zero precipitation values (Olsson et al., 2016; Teutschbein and Seibert, 2012).

Raw CM outputs resulted in a somewhat stronger correlation in comparison to the observations in several months, but seasonal variations were mostly well-reproduced. Stronger dependence may potentially have negative consequences on the simulations of water-cycle components and might result in overestimation of extreme events such as droughts or floods (Singh et al., 2020). Then, adjusting multivariate characteristics is relevant for impact studies.

Biases in the temporal structure of CM outputs series were substantial, too. This was also confirmed by previous studies, e.g., Rajczak et al. (2016), which also observed substantial overestimation of the frequency of wet days, as well as other systematic biases in precipitation persistence over time. This can be explained by the lower level of confidence in dynamic aspects of climate change, such as location and timing of the events (Shepherd, 2019), due to fundamental modeling errors in synoptic-scale atmospheric circulation in CMs (Addor et al., 2016; Maraun et al., 2017) and also under-representation of persistence of atmospheric circulation patterns (Maraun et al., 2021).

4.3. Performance assessment

The ability of each BA method to correct CM outputs strongly depended on the climate variable at hand (i.e., temperature or precipitation), the statistical feature being adjusted, and the capability of raw CMs to present that particular feature.

4.3.1. Univariate statistical features of precipitation and temperature

For univariate features of both temperature and precipitation, no clear pattern could be detected as to whether uni- or multivariate BA methods performed better. But we detected similarities in the ability of the distribution-based methods (DS and copula) to adjust biases, which on average performed better than the distribution-free BA methods (QDM and MBCn).

It should be noted that the performance of QDM and MBCn strongly depended both on the raw CM outputs and on the estimated transfer function. Thus, uncertainties associated with the robustness of the transfer function must be considered. Piani and Haerter (2012) argued: ‘*Transfer functions with a high number of parameters perform well, by construction, when applied to model output in the calibration period. However, they are likely to do poorly in the cross-validation period because of the variation of the climate bias over long time scales.*’

Changes in higher statistical moments like variance and skewness after application of quantile mapping have been the focus of previous studies (e.g., Hempel et al., 2013; Maraun et al., 2019; Switanek et al., 2017). Remaining biases in such higher-order moments are problematic as they might cause over- or underestimation of the exceedance probabilities for a given warming or drying/wetting, and, thus, any subsequent estimates of hydroclimatic compound events, such as floods or droughts (Fischer and Knutti, 2015). To overcome this issue, Hempel et al. (2013) suggested using QDM. However, this method was not able to preserve variance in both precipitation and temperature in our study for 55 Swedish catchments.

4.3.2. Multivariate statistical features of precipitation and temperature

For the 55 study sites, adjustments made by the univariate BA methods (DS and QDM) typically reproduced the dependence characteristics of uncorrected CMs and, thus, were not able to adjust the existing biases considerably. This has been also observed by other studies (Cannon, 2018; François et al., 2020; Wilcke et al., 2013), and can be explained by the fact that each variable is adjusted separately. Slight differences after the application of the univariate methods mainly occurred due to the adjustment of drizzle days.

In contrast, multivariate methods have been shown to be able to adjust dependences in previous studies (Guo et al., 2020; C. Li et al., 2014; Piani and Haerter, 2012; Singh and Reza Najafi, 2020). However, for the studied Swedish catchments, there were fundamental differences between the two multivariate methods (i.e., copula and MBCn) and their performance in adjusting correlation or the C-C relation.

The copula method considerably reduced correlation biases, but at the same time the original signal of the raw CMs seemed to vanish completely. Simulations yielded almost the exact same values for all CMs (denoted by exceptionally narrow ranges), but not always accurate, which is shown by large MAE values in February, July and October. The reason behind this lies in the statistical framework for the application of copulas: in the classic (stationary) copula framework, the copula parameter, which indicates the strength of dependence (Tootoonchi et al., 2022), is calculated based on the calibration period and is considered to be stationary over time (Bárdossy and Pegram, 2012; François et al., 2020; Tootoonchi et al., 2022; Vrac and Friederichs, 2015). By design, the dependence is therefore time-invariant. Vrac (2018) argued that capturing the baseline dependence structure of temperature and precipitation during the calibration period might be sufficient even for future projections as it is mainly bounded by regional constraints. As we can see from our results, considering the magnitude of climate change in our study sites, this assumption might not hold for all regions and for all dependence features. In fact, Mahony and Cannon (2018) found that the dependence between temperature and precipitation may globally change more significantly in the future. This was also the case for the Swedish study sites: We found that dependence was specifically modified in southern catchments in warmer months over the verification period. Naturally, this alteration of dependence was not reflected by the copula method and, consequently, considerable biases appeared in some months.

To overcome the issue of stationary copula parameters, nonstationary algorithms such as dynamic copulas (Rakonczi et al., 2012; van den Goobergh et al., 2005) can be adopted, but the added uncertainty of these methods and the challenging mathematical requisites to employ such methods considerably limits their applicability, specifically for practical climate change studies.

Another important source of uncertainty in the copula approach is the added jittering algorithm that is used to perturb ties, e.g., dry days (Y. Li et al., 2020; Pappadà et al., 2017; Salvadori et al., 2014). This algorithm may compromise the dependence structure, typically in regions where there is a large portion of data with ties. This was also the case in southern Sweden, where catchments experienced more dry days. Of course, copula can also be adopted only on a portion of data, where variables do not have ties (most typically wet days).

Simulations after application of MBCn showed better agreement with observed correlation and C-C relation than after application of the univariate methods. The ability to perform well even in the verification period in combination with the remaining spread across bias-adjusted CMs indicates that MBCn could maintain the climate-change signal projected by the CMs, which the copula method was not able to uphold (spread across CMs vanished). This is a great advantage of MBCn, which does not make strong stationarity assumptions about climate models (Cannon, 2018), specifically if the raw simulations are preserving dependence plausibly and are in line with changes over the two periods. But MBCn relies on a stochastic rotation matrix (Cannon, 2018), which can impact the adjustments. The algorithm incorporated in the MBCn method can potentially lead to modifications of the climate signal (François et al., 2020) and the statistics of adjusted variables can slightly differ from one adjustment to another. However, the effect of this stochastic procedure is not yet entirely understood and needs further investigation.

Whether or not climatic shifts in multivariate features simulated by raw CMs should be preserved needs to be carefully considered based on the study at hand, but also depends largely on the performance of the CMs. Wilcke et al. (2013) argued that the ability to maintain the raw CM-simulated dependence structure can be advantageous, because the underlying assumption for bias-adjusting CMs is that these models are able to simulate the spatiotemporal features of regional climate in a physically correct way. However, the same authors also highlight that the problem of CM-caused deficiencies in those features are potentially being retained as well. In line with the latter, several other authors argue that BA methods can compensate for such deficiencies and improve the simulated climate-change signal if the model error characteristics can be considered stationary (Bellprat et al., 2013; Boberg and Christensen, 2012; Gobiet et al., 2015). For instance, Gobiet et al. (2015) argued that methods based on quantile mapping have the potential to serve as an empirical constraint on model uncertainty in climate projections under the assumption of time-invariant model bias. Maraun (2016), on the other hand, stated that one should adopt a trend-preserving BA method (such as QDM and MBCn) if the simulated change signal can be trusted, while the way forward is rather difficult in case of implausible trend simulations.

4.3.3. Temporal statistical features of precipitation and temperature

The BA methods in this study are not specifically aimed at adjusting temporal behavior of CM outputs and, thus, this aspect of their performance was not included in the calibration. Therefore, further investigation in these aspects can serve as basis for understanding the performance of BA methods (Maraun, 2016). If the observed temporal behavior of climate variables is fundamentally misrepresented by CMs, the BA methods cannot mitigate such errors (Haerter et al., 2011; Maraun et al., 2017, 2021). However, our study clearly demonstrated that all BA methods (even unintentionally) modified the temporal characteristics of CM outputs, temperature and precipitation.

The ability of BA methods to modify temporal behavior has only been discussed in few studies. For example, Wilcke et al. (2013) found some improvement or no clear effect on autocorrelation in the simulated series of precipitation and temperature using univariate QM for a few stations over Austria and Switzerland, while François et al. (2020) found deterioration in autocorrelation of precipitation for France and stated that 'univariate correction could have a non-negligible effect on Pearson autocorrelation'. In our case, it seemed like some methods were able to reduce biases in autocorrelation of precipitation, while application of DS clearly increased such biases. Furthermore, modification of cross-correlation after application of any BA method has also been observed in Van de Velde et al. (2022), which was also the case in our study where univariate methods resulted in some reduction of errors, while both multivariate methods even inflated the biases in cross-correlation.

Owing to the fact that autocorrelation and cross-correlation are highly affected by the chosen spatiotemporal scales, correct representation of such features is not straightforward, and slight modification of time series can have huge impacts on both features (François et al., 2020). Generally

false improvement (i.e., apparent ‘improvement’ that is a mere artefact of the bias adjustment) may give the wrong impression of the method’s robustness (Maraun and Widmann, 2018b). For instance, differences between autocorrelation of corrected and uncorrected simulations may potentially be a combined result of incorrect rank structure of the model (François et al., 2020) and the drizzle effect. This is also the case for cross-correlation, which is affected both by its univariate features as well as chronology of the variables. The adjustment of drizzle days is, thus, crucial to improve temporal precipitation features. In particular, we demonstrated that the adjustment of drizzle days as part of all four BA methods applied in our study strongly improved the ability to simulate transition probabilities in all study sites.

4.3.4. Summary of findings

There was not a single BA method that excelled in all of the investigated statistical features, and variations across individual statistics could be observed. Nonetheless, the average performance of each BA method was relatively consistent: Raw simulations performed consistently worse across the analyzed univariate, multivariate and temporal features, while application of any of the four BA methods generally reduced existing biases.

Considering the two univariate methods, DS outperformed QDM in all temperature-related univariate statistics, half of the precipitation-related univariate statistics as well as half of the multivariate and temporal statistics, and consistently ranked higher than QDM within each of the statistical categories (univariate, multivariate and temporal).

When comparing the two multivariate methods with each other, copula outperformed MBCn for the temperature-related univariate statistics as well as for precipitation variance. While both methods were able to considerably reduce biases in multivariate features, their performance varied for temporal features: Among all four BA methods, MBCn showed the best performance (rank 1), whereas copula was least suitable (rank 4) for reducing biases in temporal features.

Among the two distribution-based BA methods (DS and copula), the simpler univariate DS method ranked only slightly lower than copula (average rank 2.3 versus rank 2.1) in adjusting univariate features. While copula showed better performance (rank 1) for adjusting multivariate statistics, it ranked the lowest (rank 4) for reducing biases in temporal characteristics and was outperformed by DS (rank 2). Among the distribution-free approaches, MBCn exceeded QDM in the vast majority of statistics, though they performed similarly in adjusting univariate features.

It should, however, be noted that the practical applicability of a bias adjustment method should not only be judged on its performance, but also on its complexity – or in other words on the simplicity to be implemented. While MBCn was able to reduce biases in many intricate aspects (i.e., multivariate and temporal), compared to simpler univariate methods and even the advanced copula method, it was computationally rather expensive. This can be drawback when bias adjustment has to be applied to multiple catchments, CMs, RCPs and variables at the same time. François et al. (2020) even stated that a higher number of dimensions to be corrected can result in a deterioration of rank variability. Additionally, the iterative rotation that is applied on variables may result in overfitting in the calibration period or cause higher uncertainties in comparison to easier univariate methods.

5. Concluding remarks

The evaluation presented in this paper provides an overview of the performance of two prevailing multivariate bias adjustment methods (i.e., copula and MBCn) and compares them against two popular univariate methods (i.e., DS and QDM).

By using a large hydroclimatic dataset across 55 Swedish study sites in combination with an ensemble of 10 climate models, we evaluated the ability of the four bias adjustment (BA) methods to adjust multiple univariate, multivariate and temporal aspects of daily simulated precipitation and temperature series. Our large multi-climate-model and multi-site setup of the study enabled a thorough evaluation of practical benefits and drawbacks of adopting various BA methods with different levels of complexity, under climate change, and under conditions representative for high latitudes.

In our study, biases in raw CM outputs of precipitation and temperature were substantial. Thus, raw series of CM simulations should not be used directly in (hydrologic) impact studies without prior adjustment. While all selected BA methods generally improved various aspects of the raw climate signal, there was not a single method that was consistently able to adjust all characteristics and outperform all other methods. In some instances, BA methods even inflated existing biases.

For adjusting univariate characteristics of both temperature and precipitation, DS outperformed the other univariate method (QDM). As it is computationally cheaper than multivariate methods, we recommend the adoption of DS for simple hydroclimatic applications that rely for instance on annual or seasonal water balance computations. When accurate simulations of more complex properties of hydroclimatic variables (e.g., the correct timing or magnitude of extreme events) are needed, we instead recommend to apply multivariate BA methods and in particular MBCn, as it performs similar to copula in reducing biases in multivariate features, but is rather superior in modifying temporal aspects. The multivariate methods tested in this study were, however, rather complex and may result in further unwanted modifications of temporal features. Therefore, prior to adopting these methods, other practical aspects, such as computational time and heavy theoretical requirements should be considered.

Additionally, we would like to highlight that improvement through these methods might not always be distinct in a specific region or at a given spatiotemporal scale due to multiple sources of added uncertainties. Thus, careful assessment of these bias adjustment methods for a specific purpose, region or on other hydroclimatic variables is crucial to ensure robust future projections and adjustments of systematic errors in them.

CRedit authorship contribution statement

Faranak Tootoonchi: Conceptualization, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Jan Olaf Haerter:** Investigation, Validation, Writing – review & editing. **Andrijana Todorović:** Investigation, Validation, Writing – review & editing. **Olle Rätty:** 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.

Data availability

The CORDEX data used in this work is freely available (<http://www.cordex.org>)

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.

Acknowledgements

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. The CORDEX data used in this work (<http://www.cordex.org>) were funded by the EU FP6 Integrated Project ENSEMBLES (contract 505539) whose support is gratefully acknowledged. JOH gratefully acknowledges funding from the Villum Foundation (grant no. 13168) and funding from the European Research Council under the European Union's Horizon 2020 Research and Innovation programme (grant no. 771859). FT would also like to thank Alex Cannon, whose valuable suggestions on how to apply MBCn method enabled the lead author to code this method. The authors would like to thank three

anonymous reviewers whose feedback on earlier drafts of this work helped to substantially improve this article.

Appendix A. Supplementary data

Supplementary material to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2022.158615>.

References

- Addor, N., Rohrer, M., Furrer, R., Seibert, J., 2016. Propagation of biases in climate models from the synoptic to the regional scale: implications for bias adjustment. *J. Geophys. Res. Atmos. Res.* 121 (5), 2075–2089. <https://doi.org/10.1002/2015JD024040>.
- AghaKouchak, A., Cheng, L., Mazdiyasi, O., Farahmand, A., 2014. Global warming and changes in risk of concurrent climate extremes: insights from the 2014 California drought. *Geophys. Res. Lett.* 41 (24), 8847–8852. <https://doi.org/10.1002/2014GL023088>.
- Aleshina, M.A., Semenov, V.A., Chernokulsky, A.V., 2021. A link between surface air temperature and extreme precipitation over Russia from station and reanalysis data. *Environ. Res. Lett.* 16 (10). <https://doi.org/10.1088/1748-9326/ac1c3a>.
- Bárdossy, A., Pegram, G., 2012. Multiscale spatial recorelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small. *Water Resour. Res.* 48 (9). <https://doi.org/10.1029/2011WR011524>.
- 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>.
- Bellprat, O., Kotlarski, S., Lüthi, D., Schär, C., 2013. Physical constraints for temperature biases in climate models. *Geophys. Res. Lett.* 40 (15), 4042–4047. <https://doi.org/10.1002/grl.50737>.
- Berg, P., Haerter, J.O., Thejll, P., Piani, C., Hagemann, S., Christensen, J.H., 2009. Seasonal characteristics of the relationship between daily precipitation intensity and surface temperature. *J. Geophys. Res.* 114 (D18), D18102. <https://doi.org/10.1029/2009JD012008>.
- Berg, A., Lintner, B.R., Findell, K., Seneviratne, S.I., Denhurk, B.Van, Ducharme, A., Chéruy, F., Hagemann, S., Lawrence, D.M., Malyshev, S., Meier, A., Gentine, P., 2015. Interannual coupling between summertime surface temperature and precipitation over land: processes and implications for climate change. *J. Clim.* 28 (3), 1308–1328. <https://doi.org/10.1175/JCLI-D-14-00324.1>.
- Boberg, F., Christensen, J.H., 2012. Overestimation of Mediterranean summer temperature projections due to model deficiencies. *Nat. Clim. Chang.* 2 (6), 433–436. <https://doi.org/10.1038/nclimate1454>.
- Bonan, G.B., Doney, S.C., 2018. Climate, ecosystems, and planetary futures: the challenge to predict life in Earth system models. *Science* 359 (6375). <https://doi.org/10.1126/science.aam8328>.
- Brunner, M.I., Hingray, B., Zappa, M., Favre, A., 2019. Future trends in the interdependence between flood peaks and volumes: hydro-climatological drivers and uncertainty. *Water Resour. Res.* 55 (6), 2019WR024701. <https://doi.org/10.1029/2019WR024701>.
- 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 (1), 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 (17), 6938–6959. <https://doi.org/10.1175/JCLI-D-14-00754.1>.
- Chen, D., Rodhe, H., Emanuel, K., Seneviratne, S.I., Zhai, P., Allard, B., Berg, P., Björck, S., Brown, I.A., Bärring, L., Chafik, I., Deng, K., Gaillard-Lemdhah, M.J., Hieronymus, M., Kjellström, E., Linderholm, H.W., May, W., Näslund, J.O., Ou, T., Zhang, Q., 2020. Summary of a workshop on extreme weather events in a warming world organized by the Royal Swedish Academy of Sciences. *Tellus Ser. B Chem. Phys. Meteorol.* 72 (1), 1–13. <https://doi.org/10.1080/16000889.2020.1794236>.
- Dore, M.H.L., 2005. Climate change and changes in global precipitation patterns: what do we know? *Environ. Int.* 31 (8), 1167–1181. <https://doi.org/10.1016/j.envint.2005.03.004>.
- Edwards, P.N., 2011. History of climate modeling. *Wiley Interdiscip. Rev. Clim. Chang.* 2 (1), 128–139. <https://doi.org/10.1002/wcc.95>.
- Eklund, A., 2011. SVAR, Svenskt vattenarkiv. 53. <https://www.smhi.se/publikationer/svar-svenskt-vattenarkiv-1.17833>.
- Ettinger, A.K., Buonaiuto, D.M., Chamberlain, C.J., Morales-Castilla, I., Wolkovich, E.M., 2021. Spatial and temporal shifts in photoperiod with climate change. *New Phytol.* 230 (2), 462–474. <https://doi.org/10.1111/nph.17172>.
- Fang, G.H., Yang, J., Chen, Y.N., Zammit, C., 2015. Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in China. *Hydrol. Earth Syst. Sci.* 19 (6), 2547–2559. <https://doi.org/10.5194/hess-19-2547-2015>.
- Fischer, E.M., Knutti, R., 2015. Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nat. Clim. Chang.* 5 (6), 560–564. <https://doi.org/10.1038/nclimate2617>.
- 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>.
- Franzke, C.L.E., Barbosa, S., Blender, R., Fredriksen, H.B., Laepple, T., Lambert, F., Nilsen, T., Rypdal, K., Rypdal, M., Scotto, M.G., Vannitsem, S., Watkins, N.W., Yang, L., Yuan, N., 2020. The structure of climate variability across scales. *Rev. Geophys.* 58 (2). <https://doi.org/10.1029/2019RG000657>.
- Genest, C., Rémillard, B., Beaudoin, D., 2009. Goodness-of-fit tests for copulas: a review and a power study. *Insurance Math. Econom.* 44 (2), 199–213. <https://doi.org/10.1016/j.insmatheco.2007.10.005>.
- Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., Yacalis, G., 2018. Could machine learning break the convection parameterization deadlock? *Geophys. Res. Lett.* 45 (11), 5742–5751. <https://doi.org/10.1029/2018GL078202>.
- Gobiet, A., Suklitsch, M., Heinrich, G., 2015. The effect of empirical-statistical correction of intensity-dependent model errors on the temperature climate change signal. *Hydrol. Earth Syst. Sci.* 19 (10), 4055–4066. <https://doi.org/10.5194/hess-19-4055-2015>.
- 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 (9), 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 (5). <https://doi.org/10.1029/2019WR026659>.
- Gutjahr, O., Heinemann, G., 2013. Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM: effects on extreme values and climate change signal. *Theor. Appl. Climatol.* 114 (3–4), 511–529. <https://doi.org/10.1007/s00704-013-0834-z>.
- Haerter, J.O., Hagemann, S., Moseley, C., Piani, C., 2011. Climate model bias correction and the role of timescales. *Hydrol. Earth Syst. Sci.* 15 (3), 1065–1079. <https://doi.org/10.5194/hess-15-1065-2011>.
- Hakala, K., Addor, N., Teutschbein, C., Vis, M., Dakhlaoui, H., Seibert, J., 2019. Hydrological Modeling of Climate Change Impacts. *Encyclopedia of Water*, pp. 1–20. <https://doi.org/10.1002/9781119300762.wsts0062> February.
- Hardwick Jones, R., Westra, S., Sharma, A., 2010. Observed Relationships Between Extreme Sub-daily Precipitation, Surface Temperature, and Relative Humidity. <https://doi.org/10.1029/2010GL045081>.
- Hempel, S., Frieler, K., Warszawski, L., Schewe, J., Piontek, F., 2013. A trend-preserving bias correction – the ISI-MIP approach. *Earth Syst. Dyn.* 4 (2), 219–236. <https://doi.org/10.5194/esd-4-219-2013>.
- Hewitson, B.C., Daron, J., Crane, R.G., Zermoglio, M.F., Jack, C., 2014. Interrogating empirical-statistical downscaling. *Clim. Chang.* 122 (4), 539–554. <https://doi.org/10.1007/s10584-013-1021-z>.
- IPCC, 2014. IPCC, 2014: climate change 2014: synthesis report. https://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_FINAL_full_wcover.pdf.
- Ivanov, M.A., Kotlarski, S., 2017. Assessing distribution-based climate model bias correction methods over an alpine domain: added value and limitations. *Int. J. Climatol.* 37 (5), 2633–2653. <https://doi.org/10.1002/joc.4870>.
- 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., Yiou, P., 2014. EURO-CORDEX: new high-resolution climate change projections for European impact research. *Reg. Environ. Chang.* 14 (2), 563–578. <https://doi.org/10.1007/s10113-013-0499-2>.
- Jaeger, E.B., Anders, I., Lüthi, D., Rockel, B., Schär, C., Seneviratne, S.I., 2008. Analysis of ERA 40-driven CLM simulations for Europe. *Meteorol. Z.* 17 (4), 349–367. <https://doi.org/10.1127/0941-2948/2008/0301>.
- Johansson, B., 2002. Estimation of Areal Precipitation for Hydrological Modelling in Sweden. Göteborg University, Göteborg, Sweden.
- Jung, T., 2005. Systematic errors of the atmospheric circulation in the ECMWF forecasting system. *Q. J. R. Meteorol. Soc.* 131 (607), 1045–1073. <https://doi.org/10.1256/qj.04.93>.
- Klemeš, V., 1983. Conceptualization and scale in hydrology. *J. Hydrol.* 65 (1–3), 1–23. [https://doi.org/10.1016/0022-1694\(83\)90208-1](https://doi.org/10.1016/0022-1694(83)90208-1).
- Kotlarski, S., Block, A., Böhm, U., Jacob, D., Keuler, K., Knoche, R., Rechid, D., Walter, A., 2005. Regional climate model simulations as input for hydrological applications; evaluation of uncertainties. *Adv. Geosci.* 5, 119–125. <https://doi.org/10.5194/adgeo-5-119-2005>.
- Laux, P., Vogl, S., Qiu, W., Knoche, H.R., Kunstmann, H., 2011. Copula-based statistical refinement of precipitation in RCM simulations over complex terrain. *Hydrol. Earth Syst. Sci.* 15 (7), 2401–2419. <https://doi.org/10.5194/hess-15-2401-2011>.
- Lenderink, G., Van Meijgaard, E., 2008. Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nat. Geosci.* 1 (8), 511–514. <https://doi.org/10.1038/ngeo262>.
- Lenderink, G., Van Meijgaard, E., 2009. Unexpected rise in extreme precipitation caused by a shift in rain type? *Nat. Geosci.* 2 (6), 373. <https://doi.org/10.1038/ngeo524>.
- Lewis, S.C., King, A.D., 2017. Evolution of mean, variance and extremes in 21st century temperatures. *Weather Clim. Extremes* 15 (July 2016), 1–10. <https://doi.org/10.1016/j.wace.2016.11.002> February.
- 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 (10). <https://doi.org/10.1029/2009JD012882>.
- Li, C., Sinha, E., Horton, D.E., Duffenbaugh, N.S., Michalak, A.M., 2014. Joint bias correction of temperature and precipitation in climate model simulations. *J. Geophys. Res. Atmos.* 119 (23), 13153–13162. <https://doi.org/10.1002/2014JD022514>.
- Li, Y., Li, Y., Qin, Y., Yan, J., 2020. Copula modeling for data with ties. *Stat. Interface* 13 (1), 103–117. <https://doi.org/10.4310/SII.2020.v13.n1.a9>.
- Lind, P., Erik, K., 2008. Reports Meteorology and Climatology, Temperature and Precipitation Changes in Sweden, A Wide Range of Model-based Projections for the 21st Century.
- Luo, M., Liu, T., Meng, F., Duan, Y., Frankl, A., Bao, A., De Maeyer, P., 2018. Comparing bias correction methods used in downscaling precipitation and temperature from regional climate models: a case study from the Kaidu River Basin in Western China. *Water (Switzerland)* 10 (8). <https://doi.org/10.3390/w10081046>.
- Maher, P., Vallis, G.K., Sherwood, S.C., Webb, M.J., Sansom, P.G., 2018. The impact of parameterized convection on climatological precipitation in atmospheric global climate models. *Geophys. Res. Lett.* 45 (8), 3728–3736. <https://doi.org/10.1002/2017GL076826>.
- Mahony, C.R., Cannon, A.J., 2018. Wetter summers can intensify departures from natural variability in a warming climate. *Nat. Commun.* 9 (1). <https://doi.org/10.1038/s41467-018-03132-z>.

- Maraun, D., 2016. Bias correcting climate change simulations - a critical review. *Curr. Clim. Chang. Rep.* 2 (4), 211–220. <https://doi.org/10.1007/s40641-016-0050-x>.
- Maraun, D., Widmann, M., 2018a. *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press 2018. Cambridge University Press.
- Maraun, D., Widmann, M., 2018b. Cross-validation of bias-corrected climate simulations is misleading. *Hydrol. Earth Syst. Sci.* 22 (9), 4867–4873. <https://doi.org/10.5194/hess-22-4867-2018>.
- Maraun, D., Widmann, M., Gutiérrez, J.M., Kotlarski, S., Chandler, R.E., Hertig, E., Wibig, J., Huth, R., Wilcke, R.A.I., 2015. VALUE: a framework to validate downscaling approaches for climate change studies. *Earth's Future* 3 (1), 1–14. <https://doi.org/10.1002/2014EF000259>.
- Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J.M., Hagemann, S., Richter, L., Soares, P.M.M., Hall, A., Mearns, L.O., 2017. Towards process-informed bias correction of climate change simulations. *Nat. Clim. Chang.* 7 (11), 764–773. <https://doi.org/10.1038/nclimate3418>.
- Maraun, D., Huth, R., Gutiérrez, J.M., Martín, D.S., Dubrovsky, M., Fischer, A., Hertig, E., Soares, P.M.M., Bartholy, J., Pongrácz, R., Widmann, M., Casado, M.J., Ramos, P., Bedia, J., 2019. The VALUE perfect predictor experiment: evaluation of temporal variability. *Int. J. Climatol.* 39 (9), 3786–3818. <https://doi.org/10.1002/joc.5222>.
- Maraun, D., Truhetz, H., Schaffer, A., 2021. Regional climate model biases, their dependence on synoptic circulation biases and the potential for bias adjustment: a process-oriented evaluation of the Austrian regional climate projections. *J. Geophys. Res. Atmos.* 126 (6). <https://doi.org/10.1029/2020JD032824>.
- Masud, M.B., Khaliq, M.N., Wheatler, H.S., 2017. Future changes to drought characteristics over the Canadian Prairie Provinces based on NARCCAP multi-RCM ensemble. *Clim. Dyn.* 48 (7–8), 2685–2705. <https://doi.org/10.1007/s00382-016-3232-2>.
- Mcfarlane, N., 2011. Parameterizations: representing key processes in climate models without resolving them. *Wiley Interdiscip. Rev. Clim. Chang.* 2 (4), 482–497. <https://doi.org/10.1002/wcc.122>.
- 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 (3), 1339–1354. <https://doi.org/10.5194/hess-23-1339-2019>.
- Myrhe, G., Alterskjær, K., Stjern, C.W., Hodnebrog, M., 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), 1–10. <https://doi.org/10.1038/s41598-019-52277-4>.
- Nygren, M., Giese, M., Barthel, R., 2021. Recent trends in hydroclimate and groundwater levels in a region with seasonal frost cover. *J. Hydrol.* 602 (July), 126732. <https://doi.org/10.1016/j.jhydrol.2021.126732>.
- Olsson, J., Arheimer, B., Borris, M., Donnelly, C., Foster, K., Nikulin, G., Persson, M., Perttu, A.M., Uvo, C.B., Viklander, M., Yang, W., 2016. Hydrological climate change impact assessment at small and large scales: key messages from recent progress in Sweden. *Climate* 4 (3). <https://doi.org/10.3390/cli4030039>.
- Panofsky, H.A., Brier, G.W., Best, W.H., 1958. *Some Application of Statistics to Meteorology*. Earth and Mineral Sciences Continuing Education, College of Earth and Mineral Sciences, Pennsylvania State University.
- 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 (5), 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. Env. Res. Risk A.* 31 (10), 2483–2497. <https://doi.org/10.1007/s00477-016-1357-9>.
- Pearson, K., 1920. Notes on the history of {correlation}. *Biometrika* 13 (1), 25. <https://doi.org/10.2307/2331722>.
- Piani, C., Haerter, J.O., 2012. Two dimensional bias correction of temperature and precipitation copulas in climate models. *Geophys. Res. Lett.* 39 (20), 1–6. <https://doi.org/10.1029/2012GL053839>.
- Poschold, B., Ludwig, R., 2021. Internal variability and temperature scaling of future sub-daily rainfall return levels over Europe. *Environ. Res. Lett.* 16 (6). <https://doi.org/10.1088/1748-9326/ac0849>.
- Rajczak, J., Kotlarski, S., Schär, C., 2016. Does quantile mapping of simulated precipitation correct for biases in transition probabilities and spell lengths? *J. Clim.* 29 (5), 1605–1615. <https://doi.org/10.1175/JCLI-D-15-0162.1>.
- Rakonczai, P., Márkus, L., Zempléni, A., 2012. Autocopulas: investigating the interdependence structure of stationary time series. *Methodol. Comput. Appl. Probab.* 14 (1), 149–167. <https://doi.org/10.1007/s11009-011-9230-2>.
- Räty, O., Räisänen, J., Ylhäisi, J.S., 2014. Evaluation of delta change and bias correction methods for future daily precipitation: intermodel cross-validation using ENSEMBLES simulations. *Clim. Dyn.* 42 (9–10), 2287–2303. <https://doi.org/10.1007/s00382-014-2130-8>.
- 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 (2), 33. <https://doi.org/10.3390/cli6020033>.
- Salvadori, G., Tomasicchio, G.R., D'Alessandro, F., 2014. Practical guidelines for multivariate analysis and design in coastal and off-shore engineering. *Coast. Eng.* 88, 1–14. <https://doi.org/10.1016/j.coastaleng.2014.01.011>.
- Schmith, T., Thejll, P., Berg, P., Boberg, F., Christensen, O.B., Christiansen, B., Christensen, J.H., Steger, C., Madsen, M.S., 2020. Identifying robust bias adjustment methods for extreme precipitation in a pseudo-reality setting. *Hydrol. Earth Syst. Sci. Discuss.* (July), 1–22. <https://doi.org/10.5194/hess-2020-318>.
- Seo, S.B., Das Bhowmik, R., Sankarasubramanian, A., Mahinthakumar, G., Kumar, M., 2019. The role of cross-correlation between precipitation and temperature in basin-scale simulations of hydrologic variables. *J. Hydrol.* 570 (October 2018), 304–314. <https://doi.org/10.1016/j.jhydrol.2018.12.076>.
- Shepherd, T.G., 2019. Storyline approach to the construction of regional climate change information. *Proc. R. Soc. A: Math. Phys. Eng. Sci.* 475 (2225). <https://doi.org/10.1098/rspa.2019.0013>.
- 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 (January), 124673. <https://doi.org/10.1016/j.jhydrol.2020.124673>.
- Singh, H., Pirani, F.J., Najafi, M.R., 2020. Characterizing the temperature and precipitation covariability over Canada. *Theor. Appl. Climatol.* 139 (3–4), 1543–1558. <https://doi.org/10.1007/s00704-019-03062-w>.
- Sklar, A., 1959. *Fonctions de Répartition à n Dimensions et Leurs Marges*. 8. Publications de l'Institut de Statistique de L'Université de Paris, Paris, pp. 229–231.
- SMHI, 2005. PTHBV Klimatdatabas för hydrologiska beräkningar PTHBV - en areellt högupplöst klimatdatabas för hydrologiska modellberäkningar.
- Spearman, C., 1904. The proof and measurement of association between two {things}. *Am. J. Psychol.* 15 (1), 72–101. <https://doi.org/10.1093/ije/dyq191>.
- Stehlik, J., Bárdossy, A., 2002. Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation. *J. Hydrol.* 256 (1–2), 120–141. [https://doi.org/10.1016/S0022-1694\(01\)00529-7](https://doi.org/10.1016/S0022-1694(01)00529-7).
- Switaneck, M.B., Troch, P.A., Castro, C.L., Leuprecht, A., Chang, H.-I., Mukherjee, R., Demaria, E.M.C., 2017. Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes. *Hydrol. Earth Syst. Sci.* 21 (6), 2649–2666. <https://doi.org/10.5194/hess-21-2649-2017>.
- Tamarin-Brodsky, T., Hodges, K., Hoskins, B.J., Shepherd, T.G., 2019. A dynamical perspective on atmospheric temperature variability and its response to climate change. *J. Clim.* 32 (6), 1707–1724. <https://doi.org/10.1175/JCLI-D-18-0462.1>.
- Teutschbein, C., Seibert, J., 2010. University of Zurich Zurich Open Repository and Archive Regional climate models for hydrological impact studies at the catchment scale: a review of recent modeling strategies Regional Climate Models for Hydrological Impact Studies at the Catchment-Scale. *Geogr. Compass* 4 (7), 834–860.
- 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., Wetterhall, F., Seibert, J., 2011. Evaluation of different downscaling techniques for hydrological climate-change impact studies at the catchment scale. *Clim. Dyn.* 37 (9–10), 2087–2105. <https://doi.org/10.1007/s00382-010-0979-8>.
- 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 (June). <https://doi.org/10.1016/j.ejrh.2022.101171>.
- Thom, H.C., 1951. A frequency distribution for precipitation. *Bull. Am. Meteorol. Soc.* 32 (10), 397.
- Todorović, A., Grabs, T., Teutschbein, C., 2022. Advancing traditional strategies for testing hydrological model fitness in a changing climate. *Hydrol. Sci. J.* <https://doi.org/10.1080/02626667.2022.2104646>.
- Tootoonchi, F., Sadegh, M., Haerter, J.O., Rätty, O., Grabs, T., Teutschbein, C., 2022. Copulas for hydroclimatic analysis: a practice-oriented overview. *Wiley Interdiscip. Rev. Water* 9 (2). <https://doi.org/10.1002/wat2.1579>.
- van den Goorbergh, R.W.J., Genest, C., Werker, B.J.M., 2005. Bivariate option pricing using dynamic copula models. *Insurance Math. Econom.* 37 (1 SPEC. ISS.), 101–114. <https://doi.org/10.1016/j.insmatheco.2005.01.008>.
- 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(9), pp. 2319–2344. <https://doi.org/10.5194/hess-2020-639>.
- 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 (6), 3175–3196. <https://doi.org/10.5194/hess-22-3175-2018>.
- Vrac, M., Friederichs, P., 2015. Multivariate—intervariable, spatial, and temporal—bias correction*. *J. Clim.* 28 (1), 218–237. <https://doi.org/10.1175/JCLI-D-14-00059.1>.
- Vrac, M., Thao, S., 2020. R2D2 v2.0: accounting for temporal dependences in multivariate bias correction via analogue rank resampling. *Geosci. Model Dev.* 13 (11), 5367–5387. <https://doi.org/10.5194/gmd-13-5367-2020>.
- Watanabe, S., Kanae, S., Seto, S., Yeh, P.J.F., Hirabayashi, Y., Oki, T., 2012. Intercomparison of bias-correction methods for monthly temperature and precipitation simulated by multiple climate models. *J. Geophys. Res. Atmos.* 117 (23), 1–13. <https://doi.org/10.1029/2012JD018192>.
- Wilby, R.L., Tomlinson, O.J., Dawson, C.W., 2003. Multi-site simulation of precipitation by conditional resampling. *Clim. Res.* 23 (3), 183–194. <https://doi.org/10.3354/cr023183>.
- Wilcke, R.A.I., Mendlik, T., Gobiet, A., 2013. Multi-variable error correction of regional climate models. *Clim. Chang.* 120 (4), 871–887. <https://doi.org/10.1007/s10584-013-0845-x>.
- Wilks, D.S., 2006. *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series Vol. 91.
- 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 (3–4), 211–229. <https://doi.org/10.2166/nh.2010.004>.
- Yang, W., Gardelin, M., Olsson, J., Bosshard, T., 2015. Multi-variable bias correction: application of forest fire risk in present and future climate in Sweden. *Nat. Hazards Earth Syst. Sci.* 15 (9), 2037–2057. <https://doi.org/10.5194/nhess-15-2037-2015>.
- Zscheischler, J., Orth, R., Seneviratne, S.I., 2017. Bivariate Return Periods of Temperature and Precipitation Explain a Large Fraction of European Crop Yields. <https://doi.org/10.3929/ethz-b-000190400>.
- Zscheischler, J., Westra, S., Van Den Hurk, B.J.J.M., Seneviratne, S.I., Ward, P.J., Pitman, A., Aghakouchak, A., Bresch, D.N., Leonard, M., Wahl, T., Zhang, X., 2018. Future climate risk from compound events. *Nature Climate Change*. Vol. 8. Nature Publishing Group, pp. 469–477. <https://doi.org/10.1038/s41558-018-0156-3> February.
- Zscheischler, J., Fischer, E.M., Lange, S., 2019. The effect of univariate bias adjustment on multivariate hazard estimates. *Earth Syst. Dyn.* 10 (1), 31–43. <https://doi.org/10.5194/esd-10-31-2019>.