



Revisiting the bias correction of climate models for impact studies

Thi Lan Anh Dinh, Filipe Aires

► To cite this version:

Thi Lan Anh Dinh, Filipe Aires. Revisiting the bias correction of climate models for impact studies. Climatic Change, 2023, 176 (10), pp.140. 10.1007/s10584-023-03597-y . hal-04284419

HAL Id: hal-04284419

<https://hal.science/hal-04284419v1>

Submitted on 14 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Revisiting the bias correction of climate models for impact studies

Thi Lan Anh Dinh¹ · Filipe Aires²

Received: 21 April 2022 / Accepted: 8 August 2023
© Springer Nature B.V. 2023

Abstract

Climate models are widely used in climate change impact studies. However, these simulations often cannot be used directly due to inherent limitations, such as structural biases or parametric uncertainties. Nevertheless, several so-called “bias correction” (B-C) or “bias adjustment” methods have been proposed to get these simulations closer to real observations. Various studies have reviewed available methods; however, numerous innovative methods have been developed in recent years. An up-to-date review of the B-C methods is presented here. To compare these complex methods, a focus is placed on the pedagogy of the presentation. The main lines of thought are presented based on the method assumptions, mathematical form, properties, and applicative purposes. Six representative quantile-based methods are compared for temperature and precipitation monthly time series over the European area, for a climate change scenario with a strong CO₂ forcing which is chosen here to facilitate the analysis of the differences among the methods. New, simple, and easy-to-understand diagnostic tools are recommended to measure the impact of the adjustment on the ability of B-C methods to: (1) bring the model outputs closer to observations over the historical record, (2) exploit as much as possible the climate change signal provided by the model. Each B-C method is intended to find the best compromise between these two objectives. A discussion on potential pathways for future developments is finally proposed.

Keywords Climate model · Calibration · Bias correction · Quantile mapping

1 Introduction

Global and regional circulation models are our major source of information for climate change impact studies, e.g. in hydrology or agriculture. However, they often suffer from substantial biases and errors compared to the real-world climate as described by observations (Flato et al. 2013; Kotlarski et al. 2014; Doblas-Reyes et al. 2021). These biases originate from several sources, for instance, due to structural and parametric uncertainties. Consequently,

✉ Thi Lan Anh Dinh
lananh.dinh@hotmail.com

Filipe Aires
filipe.aires@obspm.fr

¹ LERMA, Observatoire de Paris, Sorbonne Université, Paris, France

² LERMA, Observatoire de Paris, Université PSL, CNRS, Paris, France

post-processing techniques are needed to correct climate model simulations so that they can appropriately be used in climate impact studies (Doblas-Reyes et al. 2021). Many previous studies have applied several bias correction/adjustment strategies to remove systematic model errors and calibrate model outputs against observations (Wood et al. 2004; Ines and Hansen 2006; Lenderink et al. 2007; Li et al. 2010; Hawkins et al. 2013; Cannon et al. 2015; Kim et al. 2019; Maraun and Widmann 2018).

In the past decades, the number of B-C techniques has not ceased to increase. Several studies have attempted to review or inter-compare them. For example, Teutschbein and Seibert (2012) reviewed and evaluated six different methods for temperature and precipitation over five catchments in Sweden. Watanabe et al. (2012) inter-compared six different methods (five existing methods and one new method) for monthly temperature and precipitation. Maraun (2016) gave a critical review on common approaches used for bias correcting and downscaling climate change simulations. Gutjahr and Heinemann (2013) compared different precipitation B-C methods for high-resolution regional climate model simulations. Since then, numerous novel methods (e.g. quantile- or distribution-based methods) have been proposed to correct and minimise climate model errors. For instance, some studies may aim at adjusting the wet-day frequencies, while others may prefer to take into account the recurrence interval value (i.e. return period value). This variety is indeed an advantage as more choices allow the climate modellers to choose the best one for their own particular needs. However, the variety of solutions can also become a difficulty when the impact modellers need to choose an adequate calibration method. Consequently, an up-to-date and synthetic review of existing methods should be welcomed at this time. Although a recent study of Casanueva et al. (2020) inter-compared some standard and novel (i.e. trend-preserving) methods, the study aims to assess the influence of observational uncertainty and resolution mismatch rather than provide a synthetic review. In another example, Maraun and Widmann (2018) provided a general discussion of B-C approaches; however, a detailed presentation (e.g. with mathematical form) was not included yet.

Our study aims to provide a comprehensive reference of available B-C methods for the readers, e.g. the beginners in the field or the experienced users. Methods will be presented pedagogically—based on understanding the principles and requirements that lead to mathematical properties—to convey better their ideas, hypotheses, advantages, and inconveniences. Practical diagnostic tools that measure inherent statistical properties are necessary to assess various available methods. These tools will be presented in this study.

In the following, Section 2 introduces the study area, the datasets, and notations used in the study. Section 3 proposes a synthetic review of the three main B-C “families”: direct, delta, and combined. The general principle of each method is described, together with the mathematical formulas, resulting in properties and literature references. In addition, a general table synthesising all this information is provided. Section 4 analyses, compares and evaluates six commonly used quantile-based methods. Finally, Section 6 concludes the study and Sect. 7 discusses further prospects.

2 Materials

Study area This study will focus on the [8°W–31.5°E; 33°N–70°N] area that covers Europe and part of northern Africa. It is a very contrasted domain with cold and very hot regions and elevations.

Observations The re-analysis data will be considered in the following as the observations. We chose here the ERA5-Land, i.e. a replay of the land component of ERA5 re-analysis of the

European Center for Medium-Range Weather Forecasts (ECMWF) (Hersbach et al. 2018). The data are available at a resolution of $0.1^\circ \times 0.1^\circ$ (about $10\text{ km} \times 10\text{ km}$ at the Equator) from 1981 to 2018 (Muñoz Sabater 2019) and can be freely downloaded from <https://cds.climate.copernicus.eu/#!/home>. These data are upscaled to $1^\circ \times 1^\circ$ resolution to be consistent with the climate model data. We studied here the gridded monthly-mean 2 m temperature (T) and total precipitation (P) during the 1981–2014 period, which will be referred to in the following as the historical (i.e. present-day) period. Typically, reanalysis precipitation can be strongly biased as rainfall observations are not assimilated. However, in the context of this study, the potential bias in the data used as observations is not a major concern since the B-C methods can be tested using any reference dataset with reasonably realistic statistical properties.

Climate model data We considered the monthly mean of temperature and precipitation data from CNRM-CM6-1-HR (Voldoire 2019; Voldoire et al. 2019) as the climate model data. CNRM-CM6-1-HR is the high resolution of the fully coupled atmosphere-ocean general circulation model of sixth-generation jointly developed by CNRM (Centre National de Recherches Météorologiques) and Cerfacs (Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique). These data belong to the High-Resolution Model Intercomparison Project (Haarsma et al. 2016) of the sixth phase of the Coupled Model Intercomparison Project 6 (Eyring et al. 2016) and can be downloaded from <https://esgf-node.ipsl.upmc.fr/projects/cmip6-ipsl/>. They are available from 1850 to 2100 at a resolution of $1^\circ \times 1^\circ$.

As mentioned above, the data from 1981 to 2014 are studied for the historical period. The model provides the data from 2015 to 2100 as the future projections. These future data are available for several Shared Socioeconomic Pathways (SSPs) (O'Neill et al. 2016; Riahi et al. 2017). In the following, we will perform the tests on the SSP585 scenario. This scenario indicates a future in a high fossil-fuel development world throughout the 21st century, i.e. a target radiative forcing of $8.5\text{ W}\cdot\text{m}^{-2}$ in 2100. The CNRM model is considered here, but other models could easily be used. This will be done with an ensemble of models in the future.

Notations In the following, the subscript O stands for Observation, M for Model, C for Bias Correction; h for historical, f for future; and X here refers to T or P variables.

To perform a B-C, three datasets are considered, including

- A dataset of historical observations $X_{O,h}$ (i.e. the 1981–2014 period);
- A dataset of historical simulations from the considered climate model $X_{M,h}$ (i.e. the same period as the observations);
- A dataset of model simulations of the future climate (i.e. climate model projections) $X_{M,f}$ (e.g. the 2066–2099 period).

The calibrated values will be denoted as $X_{C,h}$ for values in the historical period and $X_{C,f}$ in the future period.

3 Bias correction methods

3.1 Previous classifications of the B-C methods

In the literature, B-C methods have been classified into distinct categories. Gudmundsson et al. (2012) defined parametric and non-parametric transformations as the two common B-C methods. Ho et al. (2012) determined two adjustment pathways: bias correction (i.e. mapping of future climate model simulations to find corresponding calibrated data) and change factor (i.e. mapping of the observations to find future calibrated data). The same classification was identified in the latter studies, but with different names, e.g. bias correction and delta change in Räisänen and Rätty (2013); Doblas-Reyes et al. (2021), or direct and delta change in Maraun (2016). Another study of Watanabe et al. (2012) used two distinct categories to classify available B-C methods: by whether future statistics are included in the equation for B-C (constant and variable types), and by assumptions on the statistical distributions (parametric and non-parametric types).

Here, we divided available methods into three main approaches: direct,¹ delta,² and combined (as shown in Fig. 1). While the direct approach (Section 3.2.1) applies the B-C directly to the climate model data, the delta approach (Section 3.2.2) considers only the model change signals (e.g. the change or difference between model simulations and projections). In addition, the combined approach (Section 3.2.3) associates the two former ones: it calibrates the climate model data but also accounts for the model change signals (e.g. changes in the mean, variance, long-term trend, or changes in quantiles). For example, Fig. 1c presents an illustration of a particular combined method (named as Equidistant Quantile Mapping (EqQM), which will be detailed in Section 3.2.3): this method calibrates the raw future model simulations by using the projected model changes in quantiles.

3.2 Review of three B-C approaches

3.2.1 Direct approach

The direct approach first estimates the statistical/distributional properties of the climate data (both observations and climate model data) during the historical record (step (1) in Fig. 1a). Then it tries to correct the model data to get closer to the observations (step (2) in Fig. 1a) to obtain the same statistical properties. The choice of statistical properties depends on the requirements of the considered applications; it can be the mean, variance, frequency, quantiles, or other complex properties. Moreover, several models can be used to calibrate these properties, from linear scaling to non-linear transformations. This chosen model will be referred to as the transfer model or “transfer function” in the following.

(a) Linear Scaling (LS) The LS method (Lenderink et al. 2007; Graham et al. 2007a; Berg et al. 2012) is one of the simplest B-C methods which aims to preserve the mean value of the considered variable: the mean of calibrated model data will be equal to that of the observations. A linear transformation model is used. An additive (or multiplicative) term corresponding to the difference of historical mean observations and historical model simulations is applied to the model temperatures (or precipitations).

¹ Also known as the bias correction (Ho et al. 2012; Hawkins et al. 2013) or the nudging approach (Hawkins et al. 2013).

² Also known as the change factor (Diaz-Nieto and Wilby 2005; Chen et al. 2011a; Ho et al. 2012; Hawkins et al. 2013) or the delta change (Hay et al. 2000; Maraun 2016).

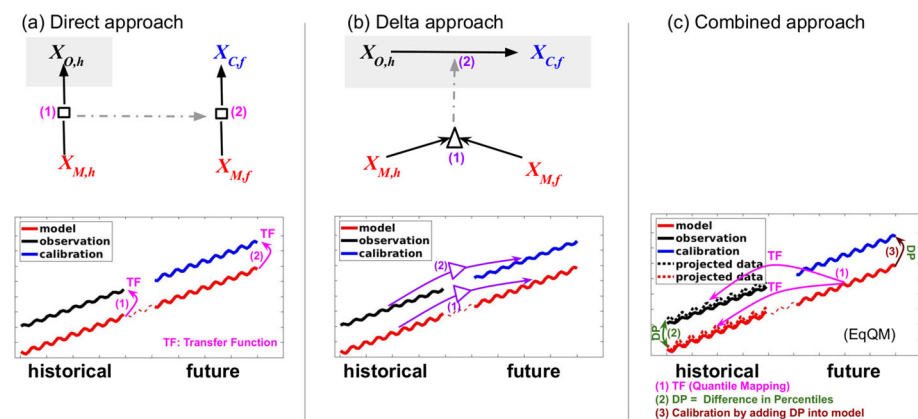


Fig. 1 Schematic (top) and time series illustration (bottom) of the three B-C approaches: **a** Direct approach applies the adjustment directly to the climate model data. **b** Delta approach considers only the model change signals and then applies this change to the observed data. **c** An illustration of a particular combined method (Equidistant Quantile Mapping (EqQM) in Section 3.2.3) includes three steps: project the future model values to historical time, compute the projected model changes (in quantiles), superimpose these changes to the raw future model data. The notations include: $X_{O,h}$ - Observations in historical period; $X_{M,h}$ - Model data in historical period; $X_{M,f}$ - Model data in future period; and $X_{C,f}$ - Bias Correction in future period

The calibrated temperatures can be described as follows:

$$\begin{aligned} T_{C,h}(i) &= T_{M,h}(i) + [\mu_{(T_{O,h})} - \mu_{(T_{M,h})}] \\ T_{C,f}(i) &= T_{M,f}(i) + [\mu_{(T_{O,h})} - \mu_{(T_{M,h})}] \end{aligned} \quad (1)$$

where μ represents for the historical monthly mean, and i is the time step. The calibrated precipitations are as follows:

$$\begin{aligned} P_{C,h}(i) &= P_{M,h}(i) \times \left[\frac{\mu_{(P_{O,h})}}{\mu_{(P_{M,h})}} \right] \\ P_{C,f}(i) &= P_{M,f}(i) \times \left[\frac{\mu_{(P_{O,h})}}{\mu_{(P_{M,h})}} \right] \end{aligned} \quad (2)$$

Since these transformations are additive and/or multiplicative, the term “linear scaling” (LS) is used here. LS corrects the temperature mean, and both precipitation mean and variance (but keeps their ratio constant). This method uses the model data as the baseline; thus, climate variability is more consistent with the model simulations. However, for that reason, the calibrated data is sensitive to the quality of climate model data used as inputs.

Figure 2a1 presents the scatter plot between the observed and simulated historical temperatures. Then the Probability Density Functions (PDFs) of the observed, modelled, and calibrated temperatures are plotted in Figs. 2a3 and a4 for the historical and future periods.

In addition to the LS method, the Local intensity scaling method was introduced by Schmidli et al. (2006) and successfully applied by Moron et al. (2008) or Themeßl et al. (2011) to correct the wet-day frequencies and wet-day intensities.

(b) Non-linear scaling While the LS method focuses on the bias in the mean, other attempts to correct both the mean and variance were considered in the past, for example, using a non-linear transformation model. The non-linearity is not related to the fact that we look

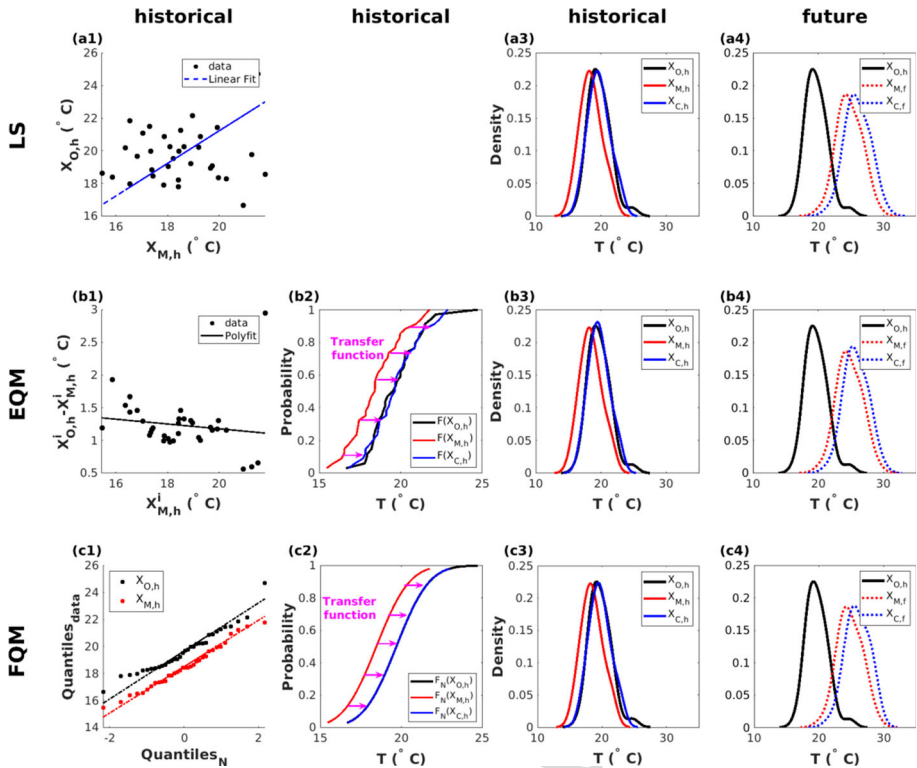


Fig. 2 Direct B-C methods applied for August temperature, at the pixel $[48^{\circ}N - 7^{\circ}E]$, for the historical (1981–2014) and future (2066–2099) periods. LS method: **a1** plots the scatter between the observed ($X_{O,h}$) and historical model data ($X_{M,h}$), and the linear fit line as for $X_{C,h}$ showed in Eq. 1; **a3** shows the PDFs of the observed, modelled and calibrated data ($X_{C,h}$); **a4** shows the observed data (for comparison purpose), the future model and corresponding calibrated data ($X_{C,f}$). EQM method: **b1** presents the polynomial fit between the historical model quantiles and the difference between the observed and historical model quantiles; **b2** plots the CDFs of the observed, historical model, and calibrated data. From the polynomial fit in **b1**, we obtain a transfer function to correct the CDF of model data to match the CDF of the observations; **b3** and **b4** are the same as **a3** and **a4**. FQM method: similar to the EQM, but **c1** presents a quantile-quantile (Q-Q) plot between the distribution of the observed and historical model data against the expected normal distribution, **c2** shows the fitted CDFs instead of the empirical CDFs as in **b2**

at something else than the bias. The Power Transformation (PT) and modified PT methods (Leander and Buishand 2007; Leander et al. 2008; van Pelt et al. 2009; Terink et al. 2010; Teutschbein and Seibert 2012; Smitha et al. 2018) consider an exponential form of precipitation by transforming precipitation value (P) into a corrected amount $P^c = a \times P^b$. The two parameters a and b in the transformation model make it possible to fit both the mean and the variance.

With a similar idea, Chen et al. (2011a, b) proposed a method aiming at conserving the mean and standard deviation of temperature variable, named as Variance Scaling. A detailed description of these methods can be found in Teutschbein and Seibert (2012).

(c) Quantile Mapping (QM) The QM³ is the most widely used technique (Panofsky and Brier 1968; Wood et al. 2004; Ines and Hansen 2006; Déqué 2007; Boé et al. 2007; Piani et al. 2009; Watanabe et al. 2012; Cannon et al. 2015). There are numerous QM-based or quantile-based adaptations (as presented in the following sections). The main idea of QM is to correct the distribution function of climate model simulations to match the distribution function of the observations. This is more ambitious than trying to adjust only the mean and variance; we expect here to adjust the whole distribution of the considered variable. A more complex transfer function needs to be used to calibrate data for that purpose. The transfer function will non-linearly adjust all events and thus, be able to correct the wet/dry day frequencies and intensities.

The general QM can be expressed as follows:

$$\begin{aligned} X_{C,h}(i) &= F_{O,h}^{-1} [F_{M,h}(X_{M,h}(i))] \\ X_{C,f}(i) &= F_{O,h}^{-1} [F_{M,h}(X_{M,f}(i))] \end{aligned} \quad (3)$$

where F is the Cumulative Distribution Function (CDF) and F^{-1} is the inverse of F .

In Eq. 3, if F is an empirical transformation, without any assumption on the variable distribution (e.g. temperature or precipitation), the method is known as the **Empirical** (or non-parametric) **Quantile Mapping (EQM)** method (Thiemeßl et al. 2011). In this case, F and F^{-1} are non-parametric and can be quite complex. Figure 2 illustrates the EQM method in the second row. Here, the transfer function is a polynomial regression that fits the historical model quantiles and the difference between the observed and historical model quantiles (Fig. 2b1). This transfer function corrects the model CDF to match the observed CDF during the historical period (Fig. 2b2). Then, Fig. 2 plots the PDFs of the observed, modelled, and calibrated data in panels b3 and b4. By definition, this method is set up during the historical period, and this will condition the range of values that can be considered by the B-C model. Thus, the problem arises when a simulated value is out-of-range of the historical values. For instance, the fit on this over-range data point can cause “new extremes”. To deal with these extremes, Wood et al. (2004) and Themeßl et al. (2012) proposed some form of extrapolation using a parametric distribution based on a hypothesis for the variable distribution. Boé et al. (2007) employed the constant correction approach to handle this new extreme issue. In another study, Michelangeli et al. (2009) suggested a specific extreme value method to model extreme wind intensities.

Instead of using the non-parametric CDF (i.e. EQM), another suggestion is to fit the CDF to the common distribution laws, which we call here a **Fitted** (or parametric) **Quantile Mapping (FQM)** method. This method performs best when the distribution of the data fits well the distribution type. For example, the Normal (or Gaussian) distribution $N(\mu, \sigma)$ is often used to fit the temperature distribution (Thom 1954; Hay and Clark 2023; Haerter et al. 2011; Ho et al. 2012; Teutschbein and Seibert 2012). On the other hand, the two-parameter gamma distribution (Thom 1958) $\Gamma(k, \theta)$ has been shown to be appropriate for precipitation distribution (Watterson and Dix 2003; Hay and Clark 2023; Ines and Hansen 2006; Block et al. 2009; Piani et al. 2009).

³ Also known as “probability mapping” (Ines and Hansen 2006; Block et al. 2009), “quantile-quantile mapping/quantile-based mapping” (Wood et al. 2004; Déqué 2007; Boé et al. 2007), “statistical bias correction” (Piani et al. 2009; Haerter et al. 2011), or “histogram equalisation” (Rojas et al. 2011).

For temperature, the calibrated model data are given by

$$\begin{aligned} T_{C,h}(i) &= F_N^{-1} [F_N(T_{M,h}(i); \mu_{(T_{M,h})}, \sigma_{(T_{M,h})}); \mu_{(T_{O,h})}, \sigma_{(T_{O,h})}] \\ T_{C,f}(i) &= F_N^{-1} [F_N(T_{M,f}(i); \mu_{(T_{M,h})}, \sigma_{(T_{M,h})}); \mu_{(T_{O,h})}, \sigma_{(T_{O,h})}] \end{aligned} \quad (4)$$

where F_N is the Normal CDF and F_N^{-1} its inverse; μ is the mean or expectation of the distribution and σ is its standard deviation.

For precipitation, we use

$$\begin{aligned} P_{C,h}(i) &= F_\Gamma^{-1} [F_\Gamma(P_{M,h}(i); k_{(P_{M,h})}, \theta_{(P_{M,h})}); k_{(P_{O,h})}, \theta_{(P_{O,h})}] \\ P_{C,f}(i) &= F_\Gamma^{-1} [F_\Gamma(P_{M,f}(i); k_{(P_{M,h})}, \theta_{(P_{M,h})}); k_{(P_{O,h})}, \theta_{(P_{O,h})}] \end{aligned} \quad (5)$$

where F_Γ is the gamma CDF and its inverse is denoted as F_Γ^{-1} ; k is a shape parameter and θ is a scale parameter.

The methodology of the FQM method is also presented in the third row of Fig. 2: panel c1 presents a quantile-quantile (Q-Q) plot between the distribution of the observed and model data during the historical record against the expected normal distribution. After that, the fitted model CDF is transferred to the observed CDF in panel c2. Finally, the PDFs of all data are plotted for the historical and future periods in panels c3 and c4.

Other distribution laws (e.g. Gumbel, Weibull, Lognormal, etc.) can also be considered to better fit the variables of interest. For example, the Generalised Extreme Value (GEV) distribution can be used to fit the temperature and precipitation extremes (Coles 2001; Kharin et al. 2007). Watterson (2008) showed that the four-parameter beta distribution provides a smooth PDF matching the mean and range of the simulated data.

The Gamma-Pareto Quantile Mapping—based on the combination of a gamma distribution and a Generalised Pareto Distribution $GPD(k, \theta, \zeta)$ (Dargahi-Noubary 1989; Coles 2001)—is widely used in the literature (Gutjahr and Heinemann 2013; Volosciuk et al. 2017; Kim et al. 2019). The main idea of this method is to better correct the higher percentiles. For precipitation, F_Γ is fitted for values between the 5th and 95th percentiles, while the GPD is used to fit the upper and lower 5%:

$$\begin{aligned} P_{C,f}(i) &= F_\Gamma^{-1} [F_\Gamma(P_{M,f}(i); k_{(P_{M,h})}, \theta_{(P_{M,h})}); k_{(P_{O,h})}, \theta_{(P_{O,h})}], \\ &\text{if } 5^{th} \text{ percentile} < P_{M,f}(i) < 95^{th} \text{ percentile;} \\ P_{C,f}(i) &= F_{GPD}^{-1} [F_{GPD}(P_{M,f}(i); \mu_{(P_{M,h})}, \sigma_{(P_{M,h})}, \zeta_{(P_{M,h})}); \mu_{(P_{O,h})}, \sigma_{(P_{O,h})}, \zeta_{(P_{O,h})}], \\ &\text{if } P_{M,f}(i) \geq 95^{th} \text{ percentile or } P_{M,f}(i) \leq 5^{th} \text{ percentile.} \end{aligned} \quad (6)$$

Here, the GPD is specified by three parameters: location μ , scale σ , and shape ζ .

The formula of Eq. 6 can also be used to calibrate the temperature variable, but a Normal distribution F_N will be preferred over F_Γ for the 5th and 95th percentiles.

We have shown that a different transfer function needs to be used to in the direct approach, depending on the statistical properties that we aim to preserve in the calibrated data (e.g. mean, variance, quantiles, extremes).

3.2.2 Delta approach

The delta approach extracts the model change signals, i.e. the differences between historical and future climate simulations. These signals are then added onto the observations to obtain the adjusted climate projections. Thus, this delta approach is not a bias correction of climate models (Teutschbein and Seibert 2012; Maraun 2016). However, as it has a long history and

has been widely used in climate studies (Hay et al. 2000; Graham et al. 2007a, b; Olsson et al. 2009; Bosshard et al. 2011; Hawkins et al. 2013; Navarro-Racines et al. 2020), we will therefore include it in this review study as one of the available B-C approaches.

Like the direct approach (Section 3.2.1), the delta approach can be used to adjust different statistical properties (e.g. mean, variance, or quantiles). All methods presented in Section 3.2.1 can be adapted to the delta approach in practise. Here, we present the two main delta methods: LS_{Δ} and QM_{Δ} (the subindex Δ will be added to avoid confusion with the direct approach). In addition, an adjustment of QM—named the Adjusted Quantile Mapping (AQM)—will be introduced.

Since, by definition, this approach does not aim to correct the historical model data, only the adjustments for the future period $X_{C,f}$ will be presented in the following of this section.

(a) Linear Scaling (LS_{Δ}) In the LS_{Δ} method (Graham et al. 2007a; Chen et al. 2011a; Hawkins et al. 2013), the absolute (or relative) changes in temperature (or precipitation) characteristics, estimated from climate model data, are transferred to the observed temperature (or precipitation) time series by additive (or multiplicative) factors:

$$T_{C,f}(i) = T_{O,h}(i) + [\mu_{(T_{M,f})} - \mu_{(T_{M,h})}] \quad (7)$$

for temperature, and:

$$P_{C,f}(i) = P_{O,h}(i) \times \left[\frac{\mu_{(P_{M,f})}}{\mu_{(P_{M,h})}} \right] \quad (8)$$

for precipitation.

By utilising the observed climate as a baseline, this method ensures that the adjusted data aligns with the statistical properties of the observed climate in the historical record. For instance, the number of rainy days will not change for future climate compared to the historical climate (i.e. the observations). In addition, extreme events are modified by the same factors as all other precipitation events.

(b) Quantile Mapping (QM_{Δ}) The QM_{Δ} method assumes that the changes from historical to future in the observation distribution are the same as the changes in the model distribution (Olsson et al. 2009; Ho et al. 2012; Hawkins et al. 2013). This is achieved by using QM (Section 3.2.1(c)): the transfer function is first estimated from the changes in the model distribution (between future and historical data); then, this transfer function is applied to the observed time series. While the LS_{Δ} method corrects only temperature mean and applies the same factors for precipitation variable, the QM_{Δ} method is able to correct both frequencies and intensities.

In general, the calibrated data (both precipitation and temperature) can be estimated as follows:

$$X_{C,f}(i) = F_{M,f}^{-1} [F_{M,h}(X_{O,h}(i))] \quad (9)$$

As for the direct approach, the CDF function F can be empirical (EQM_{Δ}) or follow a known distribution law (FQM_{Δ}) (see Section 3.2.1).

(c) Adjusted Quantile Mapping (AQM) As an adaptation of the QM_{Δ} method, Amengual et al. (2012) proposed the AQM method (or quantile-quantile adjustment) to obtain the future model simulations. The method first detects the changes of each quantile in the CDFs of the climate model outputs and then applies these changes, after calibration based on the historical period, to the observed time series.

The mathematical formula of the calibrated future model data is as follows:

$$X_{C,f}^j = X_{O,h}^j + g\bar{\delta} + s(\delta^j - \bar{\delta}) \quad (10)$$

where:

- the index j here refers to the j^{th} ranked value of the corresponding CDFs (e.g. for observations or historical and future model data),
- $\delta^j = X_{M,f}^j - X_{M,h}^j$,
- $\bar{\delta} = \frac{\sum_{j=1}^N \delta^j}{N}$,
- $g = 1$ for temperature and $g = \frac{\mu(X_{O,h})}{\mu(X_{M,h})}$ for precipitation,
- $s = \frac{IQR_{O,h}}{IQR_{M,h}}$. Here, $IQR_{O,h}$ and $IQR_{M,h}$ are the interquartile ranges of the observed and modelled data during the historical record, respectively.

3.2.3 Combined approach

Various quantile-based methods are based on the QM technique (direct approach in Sect. 3.2.1) but also account for the model change signals (e.g. delta approach). This is intended to benefit from the advantages of the two previously described approaches (Sects. 3.2.1 and 3.2.2). These methods are called here “combined approach”.

In this approach, the calibrated model simulations over the historical record are identical to the results obtained by the direct QM method (i.e. $X_{C,h}$ in Eq. 3), except for the Scaled Distribution Mapping (SDM) method. Thus, only calibrated model simulations over the future record (i.e. calibrated model projections) will be considered in the following.

(a) New Quantile Mapping (NewQM) Suggested by Watanabe et al. (2012), a NewQM method—which is based on FQM—takes into account the changes in mean and standard deviation (or coefficient of variation) for temperature (or precipitation).

Mathematically, the model temperatures over the future period are calibrated following:

$$T_{C,f}(i) = F_N^{-1} [F_N(T_{M,f}(i); \mu_{(T_{M,h})}, \sigma_{(T_{M,h})}); \mu_{cor}, \sigma_{cor}] \quad (11)$$

in which $\mu_{cor} = \mu_{(T_{O,h})} + \mu_{(T_{M,f})} - \mu_{(T_{M,h})}$ and $\sigma_{cor} = \frac{\sigma_{(T_{O,h})} \times \sigma_{(T_{M,f})}}{\sigma_{(T_{M,h})}}$.

The non-precipitation months ($P = 0$) are first removed, and the model precipitations over the future period are then corrected following:

$$P_{C,f}(i) = F_\Gamma^{-1} [F_\Gamma(P_{M,f}(i); k_{(P_{M,h})}, \theta_{(P_{M,h})}); k_{cor}, \theta_{cor}] \quad (12)$$

where k_{cor} and θ_{cor} are induced from the corrected mean and coefficient of variation (CV),

in which: $\mu_{cor} = \frac{\mu_{(P_{O,h})} \times \mu_{(P_{M,f})}}{\mu_{(P_{M,h})}}$ and $CV_{cor} = \frac{CV_{(P_{O,h})} \times CV_{(P_{M,f})}}{CV_{(P_{M,h})}}$.

(b) Detrended Quantile Mapping (DetQM) Based on the idea that trends should be preserved during the B-C process (Hempel et al. 2013), the DetQM was suggested by Cannon et al. (2015). This method is developed from the EQM method (Sect. 3.2.1), but it explicitly accounts for the modelled changes in trend. In detail, the model data are used to obtain a climate trend. Then, QM is used to calibrate model projections that have their long-term trend removed. After that, the removed climate trend is then reintroduced. This DetQM method is designed to preserve the long-term absolute (or relative) trend of the model temperature (or precipitation) data.

This method follows three steps:

- Step (1): Trend removal: $T_{C,f}^{Det1}(i) = T_{M,f}(i) - [\mu_{(T_{M,f})} - \mu_{(T_{M,h})}]$; and $P_{C,f}^{Det1}(i) = P_{M,f}(i) \times \left[\frac{\mu_{(P_{M,h})}}{\mu_{(P_{M,f})}} \right]$.
- Step (2): Quantile mapping: $T_{C,f}^{Det2}(i) = F_{O,h}^{-1} [F_{M,h}(T_{C,f}^{Det1}(i))]$; and $P_{C,f}^{Det2}(i) = F_{O,h}^{-1} [F_{M,h}(P_{C,f}^{Det1}(i))]$.
- Step (3): Trend reimposition: $T_{C,f}(i) = T_{C,f}^{Det2}(i) + [\mu_{(T_{M,f})} - \mu_{(T_{M,h})}]$; and $P_{C,f}(i) = P_{C,f}^{Det2}(i) \times \left[\frac{\mu_{(P_{M,f})}}{\mu_{(P_{M,h})}} \right]$.

Here, zero values in the observed and modelled data are first replaced by nonzero uniform random values below a tracing threshold prior to calibration. Then, values—after the adjustment—that are lower than the predefined threshold are set back to zero. This measure, therefore, can adjust the wet-day frequencies.

Initially, EQM is considered in Step (2), but we can also apply FQM instead, depending on the considered application.

(c) Equidistant Quantile Mapping (EqQM) The EqQM method (Li et al. 2010; Pierce et al. 2015) focuses on the historical bias, i.e. the difference between observed and modelled data, at a given percentile during the historical period will apply to the future period:

$$X_{C,f}(i) = X_{M,f}(i) + F_{O,h}^{-1} [F_{M,f}(X_{M,f}(i))] - F_{M,h}^{-1} [F_{M,f}(X_{M,f}(i))]. \quad (13)$$

Initially, Li et al. (2010) used EqQM to correct both temperature and precipitation data. However, the method was shown to be inefficient for precipitation; for example, it was problematic over some dry regions (Li et al. 2010) or resulted in negative precipitation values (Wang and Chen 2014). Consequently, a multiplicative factor (i.e. ratio) should be considered instead of an additive form, as often for precipitation. This adaptation was first proposed by Li et al. (2010) and then introduced as the equiratio CDF matching method by Wang and Chen (2014):

$$P_{C,f}(i) = P_{M,f}(i) \times \frac{F_{O,h}^{-1} [F_{M,f}(P_{M,f}(i))]}{F_{M,h}^{-1} [F_{M,f}(P_{M,f}(i))]} \quad (14)$$

For simplicity reasons, we will consider Eqs. 13 and 14 as the additive and multiplicative forms of the EqQM method.

(d) Quantile Delta Mapping (QDM) The QDM method (Cannon et al. 2015) aims to preserve the change signals of the model in terms of quantiles. The method considers the relative changes of a ratio variable. QDM follows three main steps:

- Step (1): The future model simulations are detrended by quantile and adjusted to observations by QM:

$$P_{C,f}^{QDM1}(i) = F_{O,h}^{-1} [F_{M,f}(P_{M,f}(i))] \quad (15)$$

- Step (2): The model changes (relative for precipitation) in terms of quantiles are computed following:

$$\Delta_{m,p}(i) = \frac{F_{M,f}^{-1} [F_{M,f}(P_{M,f}(i))]}{F_{M,h}^{-1} [F_{M,f}(P_{M,f}(i))]} = \frac{P_{M,f}(i)}{F_{M,h}^{-1} [F_{M,f}(P_{M,f}(i))]} \quad (16)$$

- Step (3): The relative changes $\Delta_{m,p}(i)$ are then superimposed on the corrected data $P_{C,f}^{QDM1}(i)$:

$$P_{C,f}(i) = P_{C,f}^{QDM1}(i) \times \Delta_{m,p}(i) \quad (17)$$

In short, Eq. 17 can be rewritten as $P_{C,f}(i) = P_{M,f}(i) \times \frac{F_{O,h}^{-1}[F_{M,f}(P_{M,f}(i))]}{F_{M,h}^{-1}[F_{M,f}(P_{M,f}(i))]}$, which is similar to the equiratio CDF matching method (Wang and Chen 2014) showed in Eq. 14. Also, the additive form of the QDM method had been showed to be equivalent to the EqQM as in Eq. 13 (Li et al. 2010; Cannon et al. 2015). Although, EqQM and QDM start with different motivations (i.e. the historical biases and the future change signals respectively), they are found to be similar, as concluded in Cannon et al. (2015). In the following sections, we will consider these two methods (EqQM and QDM) as EqQM.

In addition, QDM uses the same correction for the wet-day frequencies as the DetQM method (shown previously in Sect. 3.2.3(b)).

(e) Scaled Distribution Mapping (SDM) The SDM method was first proposed by Switanek et al. (2017) and recently used to correct a regional climate model by Maraun et al. (2021). This method is similar to QDM but also considers the frequency of rain days (which is essential when dealing with the daily precipitation data) and the likelihood of individual events.

SDM can be implemented by following seven steps for both temperature and precipitation:

- Step (1): preprocess data: detrend the raw modelled and observed temperature time series, and set a precipitation threshold to separate days with and without rain;
- Step (2): fit a PDF on the preprocessed data;
- Step (3): calculate the scaling between the fitted raw future model distribution and the fitted raw historical distribution at each probability corresponding to the events of the raw future model time series;
- Step (4): calculate the recurrence intervals (RI);
- Step (5): find the scaled or adjusted RI (RI_{scaled}) and then the corresponding scaled CDF values (F_{scaled}) for the future simulation;
- Step (6): calculate the initial array of corrected values ($BC_{initial}$);
- Step (7): reinsert $BC_{initial}$ back into the right time series, and then correct the precipitation and trend of temperature.

A detailed algorithm of SDM can be found in Switanek et al. (2017).

The $BC_{initial}$ can be mathematically written as:

$$BC_{initial}(i) = F_{O,h}^{-1}[F_{scaled,T}(T_{M,f}(i))] + \Delta_{m,T}(i) \text{ for temperature,} \quad (18)$$

$$\text{in which } \Delta_{m,T}(i) = \left\{ F_{M,f}^{-1}[F_{M,f}(T_{M,f}(i))] - F_{M,h}^{-1}[F_{M,f}(T_{M,f}(i))] \right\} \times \left[\frac{\sigma(T_{O,h})}{\sigma(T_{M,h})} \right].$$

The multiplicative form of SDM for precipitation is given by:

$$BC_{initial}(i) = F_{O,h}^{-1}[F_{scaled,P}(P_{M,f}(i))] \times \Delta_{m,P}(i), \quad (19)$$

in which $\Delta_{m,p}(i)$ is given in Eq. 16. As we can see by comparing Eqs. 17 and 19, the two methods (QDM and SDM) are very similar, except that SDM considers F_{scaled} instead of the CDF of the model projections $F_{M,f}$. The F_{scaled} reflects the scaling of the model change in likelihood event corresponding to the observed likelihoods. In addition, SDM includes the preprocessing step that also helps dealing with the trend in the temperature data or the non-precipitation day. As a result, SDM may work better when working with daily precipitation data.

3.3 General comments

Table 1 synthesises the methods discussed in Sect. 3.2, providing the main ideas, their advantages and disadvantages, and references. The presented B-C methods differ fundamentally in how they apply corrections for a specific quantile to future climate data. Some methods, notably direct approaches, determine the percentile of a simulated value in the future climate based on the CDF of simulated values in the present climate and then map onto the observed values for the same percentile (e.g. in Eq. 3). This means a simulated value is always mapped onto the same corrected value regardless of whether it occurs in the present or future climate. While this stationary assumption may not be the optimal option, it can be considered reasonable due to the lack of knowledge on how biases may change from present to future climate. In contrast, alternative methods, such as combined approaches, calculate the percentile of a simulated future value using the future simulated CDF and apply the correction derived from the present climate specifically for this percentile (see, for instance, in Eqs. 13 and 14). In addition, each method can also be used to correct different statistical properties of climate model data (except the delta approach), e.g. simply correcting the mean (LS), the mean and variance (Non-Linear Scaling), or adjusting all moments of the climate model distribution (QM and its variants). Therefore, the choice of using a particular B-C method depends on the applications (Ho et al. 2012; Teutschbein and Seibert 2012; Switanek et al. 2017) and on what the impact modellers aim to emphasise on the observations (e.g. average mean, high-resolution pattern) and on the model data (e.g. trend, distributions of changes).

4 Comparison of quantile-based B-C methods

The delta methods-especially LS_{Δ} and QM_{Δ} -do not adjust the climate model outputs. However, among three delta methods listed in Sect. 3.2.2, it can be interesting to analyse the AQM method, which is based on QM_{Δ} but also accounts for the relationship between observations and model simulations in the historical period (i.e. the factors g and s in Eq. 10). In the following sections, our analyses will focus on six quantile-based methods (i.e. QM and its variants): one direct method (FQM), one delta method (AQM), and four combined methods (NewQM, DetQM, EqQM, SDM). Here, the QM refers to the FQM, where a normal distribution is used to fit temperature values, and a gamma distribution is for precipitation values. Other distribution functions or EQM could have been considered. For the detailed analysis among direct methods (and LS_{Δ}), we refer the readers to the previous review study by Teutschbein and Seibert (2012).

By definition, the historical calibrated data of the combined methods (except for SDM) will be similar to the one from FQM, i.e. $X_{C,h}(FQM)$. Also, to be consistent, $X_{C,h}(AQM)$ will be considered equivalent to the observations $X_{O,h}$. In brief, the results include $X_{C,h}(FQM)$, $X_{C,h}(AQM)$, and $X_{C,h}(SDM)$ for historical period and six B-C methods for the future period.

4.1 Metrics of comparison

The B-C for the future period cannot be validated. Previous studies often tried to evaluate the B-C performance on the historical periods, but this evaluation might not be valid for the future (Teutschbein and Seibert 2012). Switanek et al. (2017) tested the impact of different historical periods (i.e. control periods) and showed that the control period strongly influences the B-C. This is equivalent in machine learning; they try to learn too much on the training data

Table 1 Summary of the B-C methods

	METHODS	MAIN IDEAS & COMMENTS	REFERENCES
DIRECT APPROACH	Applies the calibration to the model simulations; follows the stationary assumption (i.e. the model discrepancies are time-invariant).		
Linear	Linear	LS adjusts the mean values by applying to the raw model data an additive (or multiplicative) term, corresponding to the difference between mean observed and historical modelled data.	Lenderink et al. (2007); Graham et al. (2007a); Berg et al. (2012)
Scaling (LS)	Scaling	<i>LS corrects the temperature mean and both precipitation mean and variance (but keeps their ratio constant). This method offers calibrated data with a variability more consistent with the raw model data. But it adjusts all events by the same correction factor and thus cannot correct the frequencies.</i>	
	Local intensity scaling	This method has the same idea as LS, but also considers the wet-day frequencies and intensities	Schmidli et al. (2006); Moron et al. (2008); Themeßl et al. (2011)
Non-Linear	(Modified) Power	PT and modified PT use an exponential form to adjust the variance statistics of precipitation time series	Leander and Buishand (2007); Leander et al. (2008); van Pelt et al. (2009); Terink et al. (2010); Smitha et al. (2018)
Scaling	Variance Scaling	This method first corrects the mean, as for LS, then accounts for the variance correction of temperature time series	Chen et al. (2011a, b); Teutschbein and Seibert (2012)
	<i>Non-Linear Scaling method can correct both the mean and variance of the raw model data but not the wet-day frequencies and intensities.</i>		
Quantile	The main idea of QM is to correct the climate modelled distribution function to match the observed distribution function by using a transfer function.		
Mapping	Empirical Quantile Mapping (EQM)	The transfer function is done without any assumption of the variable distribution (i.e. non-parametric method). To improve the performance of EQM on its ability to generate new extremes, some studies suggested some form of extrapolations or a constant correction approach	Panofsky and Brier (1968); Déqué (2007); Boé et al. (2007); Michelangeli et al. (2009); Wood et al. (2004); Themeßl et al. (2012)

Table 1 continued

METHODS	MAIN IDEAS & COMMENTS	REFERENCES
Fitted Quantile	FQM try to fit the CDFs to the common known distribution laws	Watterson and Dix (2003); Ines and Hansen (2006);
Mapping (FQM)	(e.g. normal, gamma, GEV, gamma-pareto, or beta)	Block et al. (2009); Ho et al. (2012); Piani et al. (2009)
<i>QM non-linearly adjusts all events and thus, be able to correct the frequencies and intensities. EQM may raise the problem of functional robustness, e.g. causing the new extremes, since its overlarge degree of freedom is equal to the number of parameters. FQM performs best when the data distribution fit well the chosen distribution law. Otherwise, unrealistic values can be produced.</i>		
DELTA APPROACH	Assumes that the observed change from historical/present-day to future (e.g. change in mean or distribution) will be similar to the model change. In detail, this approach employs the model response to climate change to modify observations; therefore, it is not a B-C of a climate model.	
Linear Scaling	LS _Δ calibrates the future model data by adding (or multiplying) the model change	Graham et al. (2007a);
(LS _Δ)	signals (mean bias or ratio) to the observed time series	Chen et al. (2011a); Hawkins et al. (2013)
Quantile Mapping	QM _Δ assumes that the change from historical to future in the observation	Olsson et al. (2009); Ho et al. (2012);
(QM _Δ)	distribution is the same as the change in the model distribution	Hawkins et al. (2013)
Adjusted Quantile	AQM first detects the changes of each quantile in the CDFs of model data and	Amengual et al. (2012)
Mapping (AQM)	then applies these changes, after reschedules based on the historical period to the observed time series.	
<i>By definition, delta methods have the advantages of conserving all small details of the observations and the change in variabilities of the climate model. However, the method artificially accounts for all observed variabilities that, in fact, should not be resolved in the climate model data.</i>		
COMBINED APPROACH	Aims to calibrate directly the climate model data but also accounts for the model change signals; is not constrained by the stationarity assumption (except DetQM).	
New Quantile	NewQM corrects the mean and std (or mean and coefficient of variation),	Watanabe et al. (2012)
Mapping (NewQM)	instead of using directly the ones from observations, prior to QM.	

Table 1 continued

METHODS	MAIN IDEAS & COMMENTS	REFERENCES
Detrended Quantile Mapping (DetQM)	<p><i>This method conserves the changes in several statistical parameters from the historical to the future period between uncalibrated and calibrated data.</i></p> <p>DetQM first removes (or rescales) the modelled trend in the long-term mean before QM and then reimposes it afterwards.</p>	Cannon et al. (2015)
Equidistant/Equiratio Quantile Mapping (EqQM)	<p><i>DetQM better preserves the model's long-term trend in the mean.</i></p> <p>EqQM tries to estimate the correction on the historical period but only on the subspace of the variability that is predictable by a climate model.</p> <p><i>Using the QM that links the future climate to the historical, EqQM can isolate only the long-term tendency from the climate model, but not the other variabilities. In contrast, e.g. for the direct QM, the adjustment may take into account many variabilities from observations (both temporal and spatial) that we do not expect to have in the model.</i></p> <p>QDM shares the same ideas/results as EqQM</p>	Li et al. (2010); Pierce et al. (2015); Wang and Chen (2014)
Quantile Delta Mapping (QDM) Scaled Distribution	SDM is conceptually similar to QDM (or EqQM) but more explicitly considers the likelihood of events (i.e. recurrent interval). Also, SDM includes the preprocessing step that deals with the rain-day frequencies and trends in temperature values.	Switanek et al. (2017); Maraun et al. (2021)
Mapping (SDM)	<i>SDM focuses on preserving the raw climate model change in terms of magnitude, rain-day frequencies, and the likelihood of events.</i>	

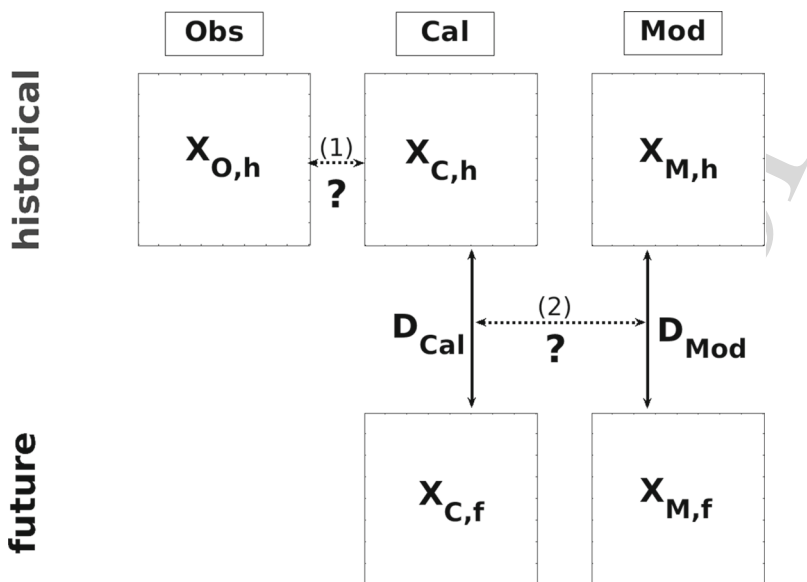


Fig. 3 Schematic presenting the two main objectives of a B-C method: (1) Do the bias corrected data ($X_{C,h}$) get closer to the observed data ($X_{O,h}$)? (2) How the bias corrected change (D_{Cal}) is compared to the raw model change (D_{Mod}). D represents the climate change signal (between historical and future climate), these changes can be measured using several statistical characteristics (e.g. mean, standard deviation, skewness)

at the detriment to the generalisation ability of the model (i.e. overtraining or overfitting). As a result, this study will not only evaluate B-C methods based on the historical period. Instead, we aim to balance the two main objectives (shown in Fig. 3) that the B-C is designed for: (1) Do the bias corrected data ($X_{C,h}$) get closer to the observed data ($X_{O,h}$)? (2) How the bias corrected change (D_{Cal}) is compared to the raw model change (D_{Mod}). D represents the climate change signal or the change between historical and future climate. The question of preserving the change signal should be addressed case by case. Ideally, B-C should not alter a credibly simulated trend. However, if the simulated climate change is implausible, it may be necessary to consider a method that modifies the trend. Further discussions on this climate change signal can be found in Maraun et al. (2017) and Maraun and Widmann (2018). Our analysis here assumes that the model provides a realistic climate change signal. We then evaluate these changes (D_{Cal} and D_{Mod}) using several characteristics of the distribution, for instance:

- mean (μ)—an indication of the general size of the data,
- standard deviation or std (σ)—an indication of variability,
- 10th and 90th percentiles (X_{10} , X_{90})—an indication of extreme values,
- skewness (γ)—a measurement of the asymmetry of the distribution of values.

We also included the trend for temperature analysis and coefficient of variation (CV) for precipitation analysis. Furthermore, two diagnostics are considered: the spatial correlation (COR^4) and Root Mean Square Error (RMSE) between D_{Mod} and D_{Cal} for all grid cells.

⁴ This refers to Pearson's correlation between two variables within a given spatial domain. The correlation is closer to 1, indicating a strong relationship between the two variables; however, the two variables are not necessarily identical if $COR = 1$.

4.2 Performance

Figure 4 presents an example of July statistics for standard deviation (a) and skewness (b) over Europe and part of northern Africa. Here, we can see that the raw climate model data ($X_{M,h}$ and $X_{M,f}$) change significantly over time. For example, the July temperature variability increases about 2°C over the study area (shown in Fig. 4a). In addition, the skewness varies greatly between the two considered periods. In Fig. 4b, the skewness of the July temperature

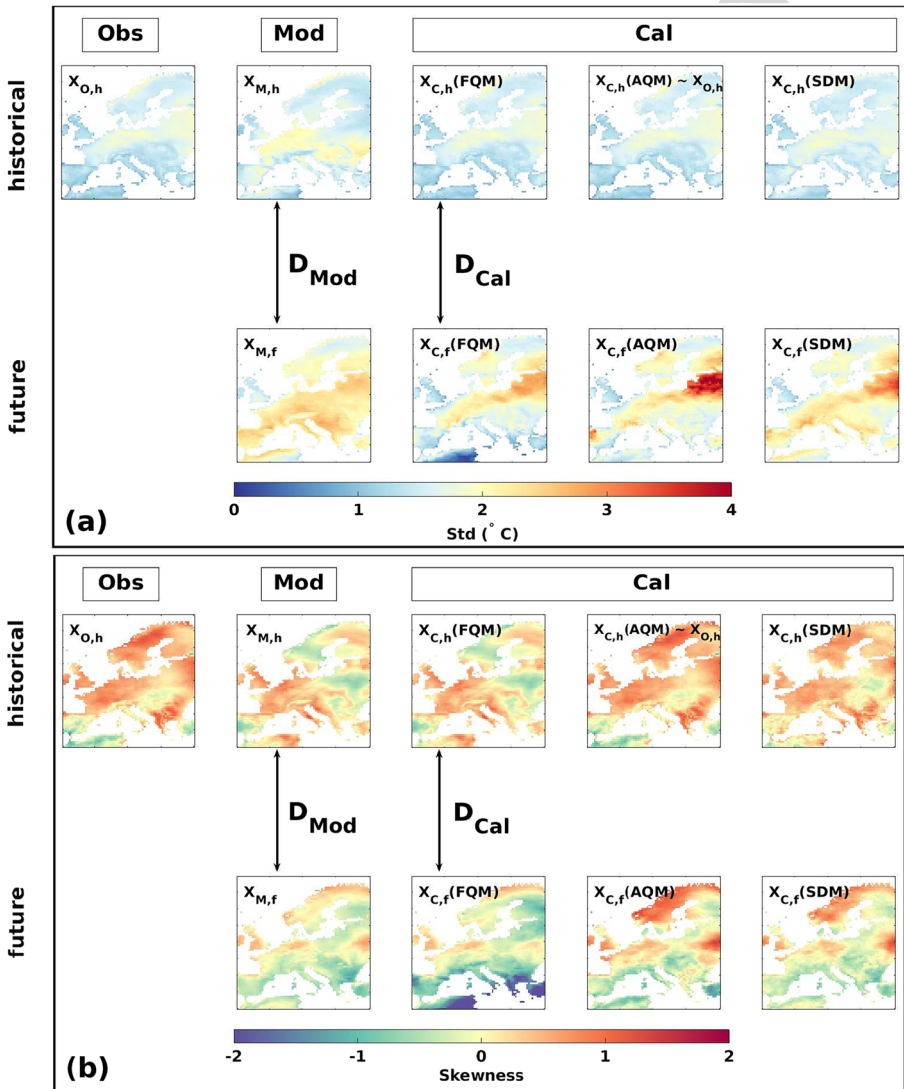


Fig. 4 July temperature statistics over Europe and part of northern Africa, for standard deviation (a) and skewness (b). For both metrics, we show the observed (Obs), raw model (Mod), and calibrated data (Cal) induced from three quantile-based methods (i.e. FQM, AQM, SDM), for the historical (1981–2014) and future (2066–2099) periods

series over France is positive for the historical period (panel of $X_{M,h}$), i.e. right skewness—most values are on the left side of the mean value. However, this value is mostly negative for the future period (panel of $X_{M,f}$). These changes, for instance, in standard deviation and skewness, suggest that a B-C method should better correct higher-order characteristics of the climate distribution rather than just correct the mean (Li et al. 2010).

Impact of B-C methods on climate model outputs For the considered examples in Fig. 4, three quantile-based methods (FQM, AQM, and SDM) are able to correct the standard deviation. For the historical period, FQM calibrates $X_{M,h}$ to be similar to $X_{O,h}$ (i.e. the correlation between $X_{C,h}(FQM)$ and $X_{O,h}$ is 1). AQM, by definition, considers the observations as the baseline data. SDM reduces the variability but not the same as $X_{O,h}$. Similar to the effects in the historical period, the three B-C methods generally reduce the variability for the future period. In particular, AQM captures changes in the climate model outputs and incorporates them into the observations. It becomes evident that the model displays a greater increase in variability over the eastern region (i.e. near Moscow), indicating a correspondingly larger variability increase in these regions in the future than in other regions. SDM slightly improves the model due to its target that focuses on preserving the raw climate model change (more details will be discussed in the next parts). On the other hand, each method deals with other characteristics of the distribution (e.g. skewness) differently. In Fig. 4b, $X_{C,h}(FQM)$ is exactly the same as $X_{M,h}$; and $X_{C,f}(FQM)$ is quite similar to $X_{M,f}$. Although FQM adjusts the whole distribution of the considered variable, this method assumes that the model discrepancies stay constant in time: it is stationary in the skew of the distribution and only changes in the mean and standard deviation. For the delta method, the skewness of $X_{C,f}(AQM)$ tends to retain signals/patterns from $X_{O,h}$, but also accounts for the model change signal. SDM adjusts the climate model data for both historical and future periods.

Impact of B-C methods on climate change signals Figure 5 shows the impact of B-C methods on climate change signals D for July temperature, over Europe and part of northern Africa, between the future and historical periods. Most of the B-C methods (except FQM) preserve well the climate change signals—with correlation scores of about 1—in various characteristics of the distribution (mean, standard deviation, 10th and 90th percentiles, and skewness), in addition to the change in the trend. In contrast, FQM shows much lower correlation scores between D_{Cal} and D_{Mod} for all considered metrics.

On the other hand, for precipitation in Fig. 6, FQM preserves well the climate change signals, which could be explained by the fact that the precipitation distribution fits well the gamma distribution. NewQM focuses on the changes in coefficient of variation rather than in standard deviation; therefore, the obtained results are much better for the coefficient of variation ($COR(D_{Mod}, D_{Cal})=0.91$) than for the standard deviation ($COR(D_{Mod}, D_{Cal})=0.56$).

While Figs. 5 and 6 show particular examples in July, Fig. 7 summarises the temporal evolution of six B-C methods over all pixels shown in Fig. 5, for both temperature and precipitation. Here, we used a B-C model for each month of the year. For temperature, AQM and four combined methods (NewQM, DetQM, EqQM, SDM) outperform the traditional FQM method in preserving the climate change signals for all considered characteristics and the trend (as seen in Fig. 5). For example, FQM shows the $RMSE(D_{Mod}, D_{Cal})$ in the change to the mean of about 1.5 to 2°C while other methods much better minimise this RMSE value (approximately 0°C). Similar behaviours are observed in other characteristics of the distribution (standard deviation, 10th and 90th percentiles) in which FQM shows much higher RMSE values than other methods. SDM performs best with respect to preserving the changes in the trend. This can be explained by the preprocessing step of SDM to adjust the trend of

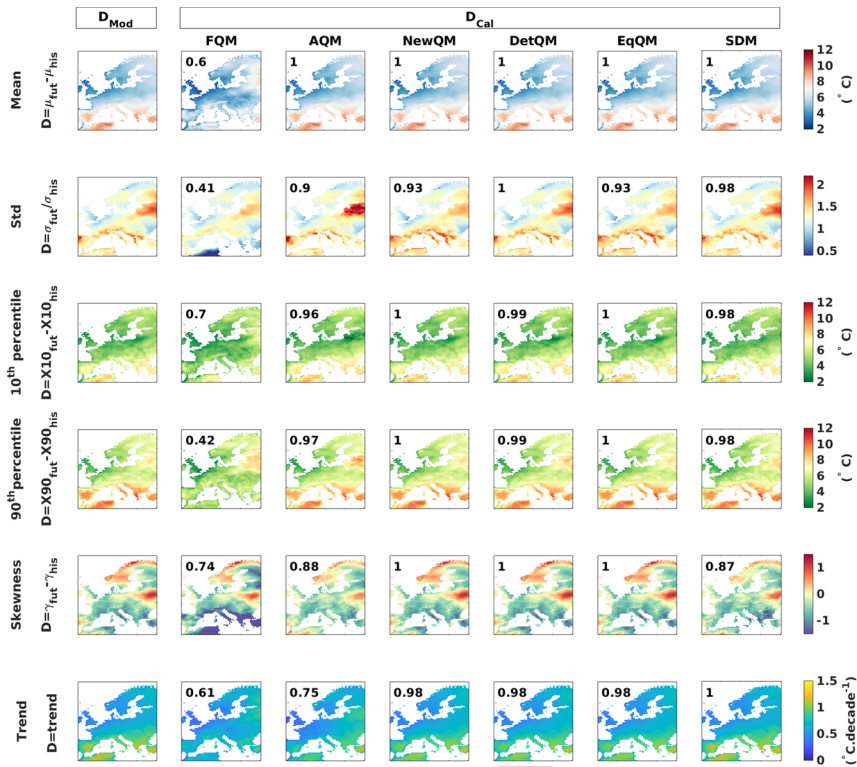


Fig. 5 Impact of B-C methods on climate change signals D for July temperature between the future (2066–2099) and historical (1981–2014) periods over Europe and part of northern Africa. The first column shows the raw model change (D_{Mod}), other columns present the calibrated changes (D_{Cal}) induced from several methods. The figure presents the changes in various characteristics of the distribution (mean μ , standard deviation (std) σ , 10^{th} percentile $X10$, 90^{th} percentile $X90$, skewness γ) in addition to the trend. The spatial correlation between D_{Mod} and D_{Cal} are indicated on each panel

temperature. For precipitation, five quantile-based variants (AQM, NewQM, DetQM, EqQM, and SDM) show improvement in the ability to preserve the raw projected mean change compared to FQM. On the other hand, FQM preserves well the climate change signals for other moments of the distribution. Again, NewQM performs better for the coefficient of variation than standard deviation as its target is to focus on the coefficient of variation correction. There is no significant difference among considered methods for precipitation. This behaviour can be explained by the considered temporal resolution: most methods attempt to handle the non-rain days, while this study focuses on the monthly data.

To sum up, Fig. 7 gives an overview/intercomparison of all methods via several diagnostics. More methods/diagnostics—that are important to the users—can be added easily. This kind of presentation is something simple but very informative. We can readily discern any seasonal dependencies by examining the 12-month results for each method and diagnostic. Also, this comparison helps identify the method that effectively preserves climate change signals, such as changes in the mean or trend. With this comprehensive perspective, users

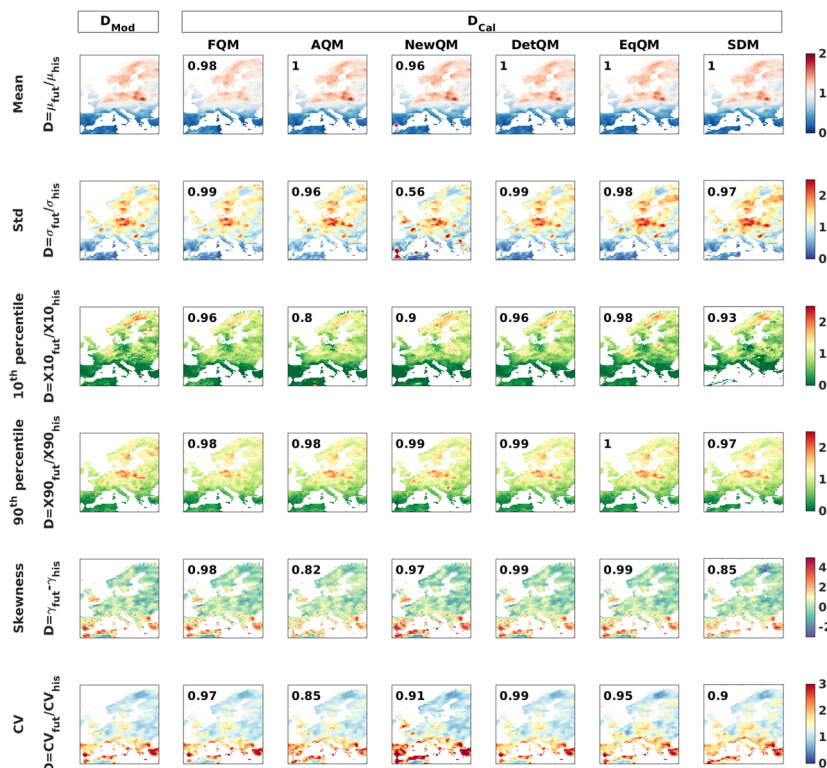


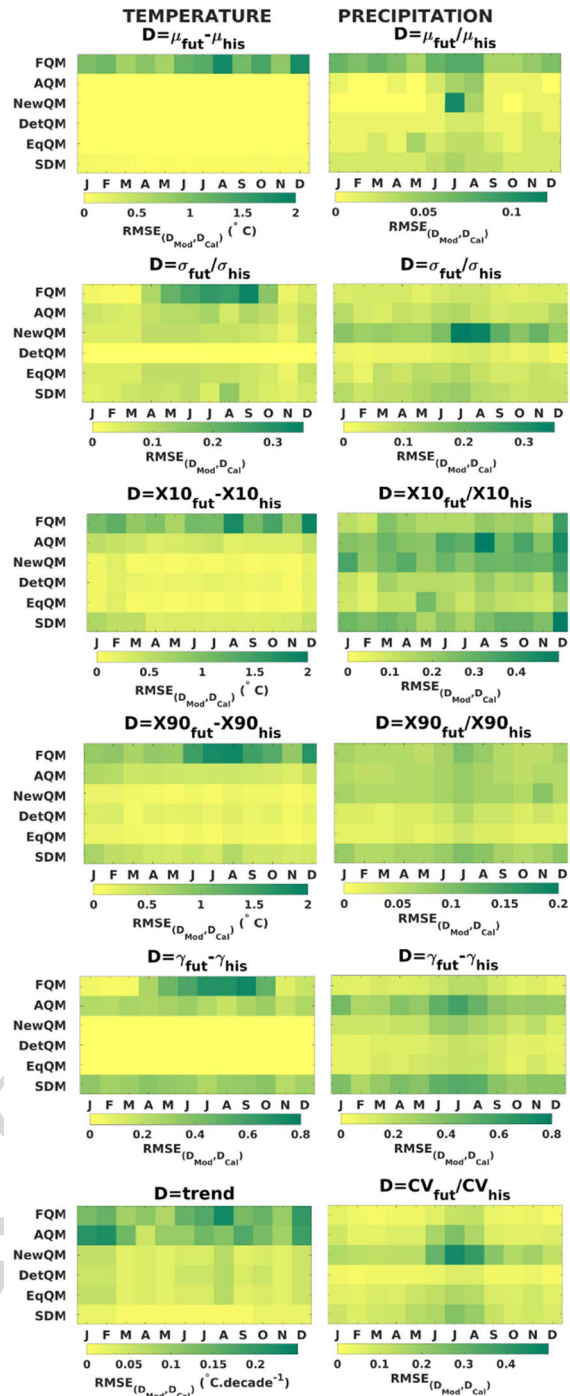
Fig. 6 Same as Fig. 5, but for July precipitation. The coefficient of variation (CV) diagnostic is added, and there is no trend diagnostic for precipitation

can conveniently select the most appropriate method for their specific applications. Thus, we recommend incorporating this synthetic comparison, along with other diagnostics like geographical information, as a vital complement when evaluating B-C methods.

5 Conclusions

B-C techniques become more and more important as they are essential to exploit climate change simulations. In particular, the climate impact studies measuring the socio-economic consequences of climate change require calibrated data in many fields (e.g. energy, agriculture, hydrology, water management). A correction technique often has two goals: (1) to make the climate simulations more realistic by bringing them closer to observations in a historical record; (2) to preserve the climate change signal that is provided by the model simulations. Each technique will make a particular compromise between these two aspects. The compromise will be dictated by the statistical properties that climate scientists want to obtain on the correction data. However, one of the main challenges for climate scientists is determining which methods are likely to yield realistic results in advance of future data availability. This requires a thorough understanding of the underlying reasons for biases (i.e. structural or parametric factors) and the fundamental assumptions of the correction methods most applicable to a given situation (Maraun and Widmann 2018). While certain applications may be

Fig. 7 Performance of six B-C methods on the climate change signals D between the future (2066–2099) and historical (1981–2014) periods, for temperature and precipitation. The Root Mean Square Error (RMSE) between the raw model change (D_{Mod}) and calibrated change (D_{Cal}) for different moments of the distribution in addition to trends are presented here. This score is computed over all pixels shown in Fig. 5 (Europe and part of northern Africa) for monthly adjustment. All the months—January (J), February (F), ..., December (D)—are presented



justifiably addressed by mapping simulated values onto the same corrected values regardless of their temporal context, other applications require deciding whether to preserve or modify the trend (Teutschbein and Seibert 2013; Maraun et al. 2017; Maraun and Widmann 2018). Furthermore, verifying the credibility of any trend modifications made is crucial.

Available B-C methods have become more complex, and it is more difficult for the user to choose and decide which one to use. Therefore, an up-to-date synthetic review of available methods was necessary. This paper tried to do so by presenting in a pedagogic way the available methods, emphasising the main ideas, the mathematical formulas, the pertinent references, and their advantages and drawbacks.

It is not easy to judge if one method is better than another. For instance, it is possible to obtain a method that makes the climate simulations closer to the observations; however, this does not mean that the obtained transformation will be correct in the future, and no direct evaluation can be done. Thus, diagnostics need to check both the proximity of the simulations to the observations and the preservation of climate change information. Several diagnostics were introduced for that reason to measure the impact of using one technique over another, such as climate change signals in the mean, standard deviation, skewness, or trend.

For the three B-C approaches (i.e. direct, delta, and combined), experiments have been conducted for temperature and precipitation (two variables with a very different nature) over Europe. These experiments were considered using monthly data, but a daily scale would be necessary if the important statistical property was related to, for instance, the number of hot days or rain intensity. Our diagnostics can show the limitations and advantages of each approach. Overall, the combined approach seems to be a good candidate when having a credibly simulated trend because it preserves the climate change signals well.

6 Perspectives and discussions

Numerous perspectives can be considered and discussed:

B-C and downscaling In the literature, the B-C of climate model simulations is often associated with a downscaling (from a coarse resolution grid of the model to a higher resolution of the observations) (Volosciuk et al. 2017; Gutiérrez et al. 2019; Widmann et al. 2009; Maraun et al. 2019). Resolutions of 1 km can sometimes be attained. Simple regression models have been traditionally used to downscale model simulations in weather forecasting, and this has been naturally extended to climate simulations. Downscaling, in this context, means that the spatial patterns present in the historical record are projected to the future. This can be a good thing, but it can be dangerous too because high-resolution patterns might be related to, for instance, land use, which can be completely different in the future. Dissociating the downscaling and the calibration can therefore be beneficial.

B-C and model errors B-C can sometimes be used to correct important issues in the climate model. For example, a climate oscillation pattern in the model could not be well located compared to real observations (Flato et al. 2013). Using a B-C model to correct such deficiencies is too large a stretch of what can be asked to simple statistical models (Dosio 2016; Doblas-Reyes et al. 2021). It is illusory to as to a simple regression model, at the pixels level, to fix a complex physical model. Therefore, B-C should not be used to pretend to correct such important climate features.

How to evaluate a B-C? It is not easy to see if a B-C method is of good quality or not. For instance, it can perform well for the historical record but not be adequate for the future that cannot be evaluated (Teutschbein and Seibert 2012). There is even the risk of getting too close to the observations to the detriment of what is done in the future (i.e. this issue is called “over-training” in machine learning theory). The solution proposed here is to use a large set of diagnostics showing the properties of the original and the calibrated data and decide if these diagnostics are satisfactory for the application that has been considered.

B-C of extreme values The B-C of extreme values is a challenging problem for several reasons. First, the definition of extremes is tricky: it can be based on quantile information (highest and lowest values at the 5 or 10% level), a number of (rain/hot) days higher than a threshold, a frequency of occurrence of a particular event, etc. Each extreme definition would require a particular diagnostic tool, a particular transfer function for the B-C. For particular aspects of extreme cases, hand-tailored models need to be designed. Second, a statistical method such as a B-C is mainly driven by the most common events in the database used to set it up, not by the rarest extreme events. Therefore, the applicability of the B-C model to extreme values can be questionable and erroneous. Third, future extreme values in climate simulations can extend beyond the observed ranges in the historical record used to set up the B-C. This out-of-range will jeopardise the applicability of the B-C model for future scenarios. Delta methods utilise relative changes and can overpass this problem, but direct and combined approaches can suffer from such issues. In this case, some form of extrapolation, e.g. proposed by Wood et al. (2004); Themeßl et al. (2012), could help deal with these outside values. Finally, extremes are even more vivid for climate simulations on a daily scale, as the extreme definition is even more complex and comparing observations and model outputs will be very challenging.

B-C of multivariate The B-C methods proposed in this paper deal with marginal distributions, meaning that the adjustment was made independently in each considered variable (e.g. temperature or precipitation), ignoring the influence of all other variables (or locations) (Maraun 2012). This approach risks breaking the link between the climate variables and introducing physical inconsistencies. A good measure of the dependency among climatic variables could help reduce this problem (Dosio and Paruolo 2011). Several studies recently suggested using a multivariate method dealing with the joint distributions of the climate variables (Maraun 2016). Such methods aim at adjusting the joint distributions without breaking the dependency structure between the variables (Vrac and Friederichs 2015; Cannon 2016; Vrac 2018.) However, most multivariate methods are bivariate as they conserve the dependency between two variables only. More variables could theoretically be considered when adjusting joint distributions, but ultimately, all the climate variables should be included. In this case, the statistical model would become a full climate model, and this ambitious task is obviously not possible. This multivariate aspect is a fundamental limitation of the calibration principle: between a practical but very simplistic transformation and a more complex statistical model that is unattainable.

Artificial Intelligence (AI) in B-C AI and, more specifically, machine learning techniques are vastly spreading in the Geoscience discipline (Lary et al. 2016). These types of data-driven techniques could potentially be used to calibrate climate simulations. In fact, AI and traditional techniques are regression models based on a learning database, equivalent to a B-C transfer function set up on the historical record. For instance, Neural Networks (NN) (Bishop 1995) have the advantage of being non-linear models, which could be helpful to represent

quantile-based transfer functions, potentially non-parametric, with no required hypothesis in the distributions. This NN representation richness and flexibility can be an advantage, but it is dangerous too because what has been done in the B-C is less controlled and understood. The non-linearity of AI models also means that the NN transformation would be state-dependent, which could be an exciting feature of the B-C. Furthermore, NN deals extremely well with multivariate relationships, but again this can be risky as inferring multivariate relationships from data can be driven by most representative cases, at the detriment to the rarest cases. This could conduct the B-C to smooth out the climate simulations and simplify multivariate relationships among climate variables.

B-C for impact assessments Understanding the principles and fundamental assumptions of B-C methods is essential; however, it is still very challenging to adjust climate model bias for impact assessment studies for many reasons (Teutschbein and Seibert 2012; Galmarini et al. 2019). B-C techniques, often applied independently on limited variables (Hawkins et al. 2013; Vrac and Friederichs 2015; François et al. 2020), present a considerable obstacle for impact models that are non-linear and sensitive to variable inconsistencies (Asseng et al. 2013), leading to non-realistic solutions. Assessing sensitivity in impact models to various B-C techniques and identifying appropriate methods for integrating climate predictions and projections present significant challenges. Collaboration among climate scientists, impact modellers, statisticians, and regional climate experts is crucial to effectively address these challenges and enhance the accuracy of impact assessments (Galmarini et al. 2019).

Author contribution All authors conceptualised the research and formulated the model. TLAD implemented the experiments in Matlab and analysed the output with FA. All authors contributed to writing the paper.

Funding This work benefits from the French state aid managed by the ANR under the “Investissements d’avenir” programme with the reference ANR-16-CONV-0003, and the Australian Centre for International Agricultural Research and conducted as part of activities for the project “Enhancing smallholder livelihoods in the Central Highlands of Vietnam through improving the sustainability of coffee and black pepper farming systems and value chains” (AGB/2018/175).

Data availability The observed climate data, i.e. ERA5-Land data, can be downloaded from <https://cds.climate.copernicus.eu> (last access: 22 Apr 2021). And the climate mode data are available at <https://esgf-node.ipsl.upmc.fr/projects/cmip6-ipsl/> (last access: 12 Nov 2021).

Declarations

Ethical approval This article does not contain any studies with human or animal participants performed by any of the authors.

Consent to participate Not applicable.

Consent to publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

- Amengual A, Homar V, Romero R, et al (2012) A statistical adjustment of regional climate model outputs to local scales: application to platja de palma, spain. *J Climate* 25(3):939–957. <https://doi.org/10.1175/JCLI-D-10-05024.1>. <https://journals.ametsoc.org/view/journals/clim/25/3/jcli-d-10-05024.1.xml>
- Asseng S, Ewert F, Rosenzweig C, et al (2013) Uncertainty in simulating wheat yields under climate change. *Nature Climate Change* 3(9):827–832. <https://doi.org/10.1038/nclimate1916>
- Berg P, Feldmann H, Panitz HJ (2012) Bias correction of high resolution regional climate model data. *J Hydrol* 448–449:80–92. <https://doi.org/10.1016/j.jhydrol.2012.04.026>. <https://www.sciencedirect.com/science/article/pii/S0022169412003010>
- Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press Inc, USA
- Block PJ, Souza Filho FA, Sun L et al (2009) A streamflow forecasting framework using multiple climate and hydrological models1. *JAWRA J Am Water Resour Assoc* 45(4):828–843. <https://doi.org/10.1111/j.1752-1688.2009.00327.x>. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1752-1688.2009.00327.x>
- Boé J, Terray L, Habets F, et al (2007) Statistical and dynamical downscaling of the seine basin climate for hydro-meteorological studies. *Int J Climatol* 27(12):1643–1655. <https://doi.org/10.1002/joc.1602>. <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.1602>
- Bosshard T, Kotlarski S, Ewen T, et al (2011) Spectral representation of the annual cycle in the climate change signal. *Hydrol Earth Syst Sci* 15(9):2777–2788. <https://doi.org/10.5194/hess-15-2777-2011>. <https://hess.copernicus.org/articles/15/2777/2011/>
- Cannon AJ (2016) Multivariate bias correction of climate model output: matching marginal distributions and intervariable dependence structure. *J Climate* 29(19):7045–7064. <https://doi.org/10.1175/JCLI-D-15-0679.1>. <https://journals.ametsoc.org/view/journals/clim/29/19/jcli-d-15-0679.1.xml>
- Cannon AJ, Sobie SR, Murdock TQ (2015) Bias correction of gcm precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes? *J Climate* 28(17):6938–6959. <https://doi.org/10.1175/JCLI-D-14-00754.1>. <https://journals.ametsoc.org/view/journals/clim/28/17/jcli-d-14-00754.1.xml>
- Casanueva A, Herrera S, Iturbide M, et al (2020) Testing bias adjustment methods for regional climate change applications under observational uncertainty and resolution mismatch. *Atmos Sci Lett* 21(7):e978. <https://doi.org/10.1002/asl.978>. <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/asl.978>
- Chen J, Brissette FP, Leconte R (2011) Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *J Hydrol* 401(3):190–202. <https://doi.org/10.1016/j.jhydrol.2011.02.020>. <https://www.sciencedirect.com/science/article/pii/S0022169411001351>
- Chen J, Brissette FP, Poulin A, et al (2011) Overall uncertainty study of the hydrological impacts of climate change for a canadian watershed. *Water Resources Research* 47(12). <https://doi.org/10.1029/2011WR010602>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011WR010602>
- Coles S (2001) An introduction to statistical modeling of extreme values. Springer
- Dargahi-Noubary GR (1989) On tail estimation: an improved method. *Math Geol* 21(8):829–842. <https://doi.org/10.1007/BF00894450>
- Déqué M (2007) Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values. *Global Planet Chang* 57(1):16–26. <https://doi.org/10.1016/j.gloplacha.2006.11.030>
- Diaz-Nieto J, Wilby RL (2005) A comparison of statistical downscaling and climate change factor methods: impacts on low flows in the river thames, united kingdom. *Climatic Change* 69(2):245–268. <https://doi.org/10.1007/s10584-005-1157-6>
- Doblas-Reyes FJ, Sorensson AA, Almazroui M, et al (2021) Linking global to regional climate change. In: Masson-Delmotte V, Zhai P, Pirani A, et al (eds) *Climate Change 2021: The Physical science basis. Contribution of working group I to the 6th Assessment report of the intergovernmental panel on climate change*. Cambridge University Press, <https://centaur.reading.ac.uk/99896/>
- Dosio A (2016) Projections of climate change indices of temperature and precipitation from an ensemble of bias-adjusted high-resolution euro-cordex regional climate models. *J Geophys Res: Atmos* 121(10):5488–5511. <https://doi.org/10.1002/2015JD024411>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JD024411>
- Dosio A, Paruolo P (2011) Bias correction of the ensembles high-resolution climate change projections for use by impact models: evaluation on the present climate. *Journal of Geophysical Research: Atmospheres* 116(D16). <https://doi.org/10.1029/2011JD015934>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JD015934>
- Eyring V, Bony S, Meehl GA, et al (2016) Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci Model Dev* 9(5):1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>. <https://gmd.copernicus.org/articles/9/1937/2016/>

- Flato G, Marotzke J, Abiodun B, et al (2013) Evaluation of climate models, Cambridge University Press, Cambridge, UK, pp 741–882. <https://doi.org/10.1017/CBO9781107415324.020>
- François B, Vrac M, Cannon AJ, et al (2020) Multivariate bias corrections of climate simulations: which benefits for which losses? *Earth Syst Dyn* 11(2):537–562. <https://doi.org/10.5194/esd-11-537-2020>. <https://esd.copernicus.org/articles/11/537/2020/>
- Galmarini S, Cannon A, Ceglar A, et al (2019) Adjusting climate model bias for agricultural impact assessment: how to cut the mustard. *Climate Serv* 13:65–69. <https://doi.org/10.1016/j.cliser.2019.01.004>. <https://www.sciencedirect.com/science/article/pii/S2405880718300608>
- Graham LP, Andréasson J, Carlsson B (2007a) Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods—a case study on the lule river basin. *Climatic Change* 81(1):293–307. <https://doi.org/10.1007/s10584-006-9215-2>
- Graham LP, Hagemann S, Jaun S, et al (2007b) On interpreting hydrological change from regional climate models. *Climatic Change* 81(1):97–122. <https://doi.org/10.1007/s10584-006-9217-0>
- Gudmundsson L, Bremnes JB, Haugen JE, et al (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>. <https://hess.copernicus.org/articles/16/3383/2012/>
- Gutiérrez JM, Maraun D, Widmann M, et al (2019) An intercomparison of a large ensemble of statistical downscaling methods over europe: results from the value perfect predictor cross-validation experiment. *Int J Climatol* 39(9):3750–3785. <https://doi.org/10.1002/joc.5462>. <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5462>
- Gutjahr O, Heinemann G (2013) Comparing precipitation bias correction methods for high-resolution regional climate simulations using cosmo-clm. *Theor Appl Climatol* 114(3):511–529. <https://doi.org/10.1007/s00704-013-0834-z>
- Haarsma RJ, Roberts MJ, Vidale PL, et al (2016) High resolution model intercomparison project (high-resmip v1.0) for cmip6. *Geosci Model Dev* 9(11):4185–4208. <https://doi.org/10.5194/gmd-9-4185-2016>. <https://gmd.copernicus.org/articles/9/4185/2016/>
- Haerter JO, Hagemann S, Moseley C, et al (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>. <https://hess.copernicus.org/articles/15/1065/2011/>
- Hawkins E, Osborne TM, Ho CK, et al (2013) Calibration and bias correction of climate projections for crop modelling: an idealised case study over Europe. *Agri Forest Meteorol* 170:19–31. <https://doi.org/10.1016/j.agrformet.2012.04.007>
- Hay LE, Wilby RL, Leavesley GH (2000) A comparison of delta change and downscaled gcm scenarios for three mountain basins in the united states. *JAWRA J Am Water Resour Assoc* 36(2):387–397. <https://doi.org/10.1111/j.1752-1688.2000.tb04276.x>. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1752-1688.2000.tb04276.x>
- Hay L, Clark M (2003) Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western united states. *J Hydrol* 282(1):56–75. [https://doi.org/10.1016/S0022-1694\(03\)00252-X](https://doi.org/10.1016/S0022-1694(03)00252-X). <https://www.sciencedirect.com/science/article/pii/S002216940300252X>, mountain Hydrology and Water Resources
- Hempel S, Frieler K, Warszawski L, et al (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>. <https://esd.copernicus.org/articles/4/219/2013/>
- Hersbach H, de Rosnay P, Bell B, et al (2018) Operational global reanalysis: progress, future directions and synergies with nwp (27). <https://doi.org/10.21957/tkic6g3wm>. <https://www.ecmwf.int/node/18765>
- Ho CK, Stephenson DB, Collins M, et al (2012) Calibration strategies a source of additional uncertainty in climate change projections. *Bull Am Meteorol Soc* 93(1):21–26. <https://doi.org/10.1175/2011BAMS3110.1>
- Ines AV, Hansen JW (2006) Bias correction of daily gcm rainfall for crop simulation studies. *Agri Forest Meteorol* 138(1):44–53. <https://doi.org/10.1016/j.agrformet.2006.03.009>. <https://www.sciencedirect.com/science/article/pii/S0168192306000979>
- Kharin VV, Zwiers FW, Zhang X, et al (2007) Changes in temperature and precipitation extremes in the ipcc ensemble of global coupled model simulations. *J Climate* 20(8):1419–1444. <https://doi.org/10.1175/JCLI4066.1>. <https://journals.ametsoc.org/view/journals/clim/20/8/jcli4066.1.xml>
- Kim DI, Kwon HH, Han D (2019) Bias correction of daily precipitation over South Korea from the long-term reanalysis using a composite Gamma-Pareto distribution approach. *Hydrol Res* 50(4):1138–1161. <https://doi.org/10.2166/nh.2019.127>

- Kotlarski S, Keuler K, Christensen OB, et al (2014) Regional climate modeling on european scales: a joint standard evaluation of the euro-cordex rcm ensemble. *Geosci Model Dev* 7(4):1297–1333. <https://doi.org/10.5194/gmd-7-1297-2014>. <https://gmd.copernicus.org/articles/7/1297/2014/>
- Lary DJ, Alavi AH, Gandomi AH, et al (2016) Machine learning in geosciences and remote sensing. *Geosci Frontiers* 7(1):3–10. <https://doi.org/10.1016/j.gsf.2015.07.003>. <https://www.sciencedirect.com/science/article/pii/S1674987115000821>, special Issue: Progress of machine learning in geosciences
- Leander R, Buishand TA (2007) Resampling of regional climate model output for the simulation of extreme river flows. *J Hydrol* 332(3):487–496. <https://doi.org/10.1016/j.jhydrol.2006.08.006>. <http://www.sciencedirect.com/science/article/pii/S0022169406004045>
- Leander R, Buishand TA, van den Hurk BJ, et al (2008) Estimated changes in flood quantiles of the river meuse from resampling of regional climate model output. *J Hydrol* 351(3):331–343. <https://doi.org/10.1016/j.jhydrol.2007.12.020>. <https://www.sciencedirect.com/science/article/pii/S0022169407007743>
- Lenderink G, Buishand A, van Deursen W (2007) Estimates of future discharges of the river rhine using two scenario methodologies: direct versus delta approach. *Hydrol Earth Syst Scie* 11(3):1145–1159 <https://doi.org/10.5194/hess-11-1145-2007>. <https://hess.copernicus.org/articles/11/1145/2007/>
- Li H, Sheffield J, Wood EF (2010) Bias correction of monthly precipitation and temperature fields from intergovernmental panel on climate change ar4 models using equidistant quantile matching. *Journal of Geophysical Research: Atmospheres* 115(D10). <https://doi.org/10.1029/2009JD012882>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009JD012882>
- Maraun D (2012) Nonstationarities of regional climate model biases in european seasonal mean temperature and precipitation sums. *Geophysical Research Letters* 39(6). <https://doi.org/10.1029/2012GL051210>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012GL051210>
- Maraun D (2016) Bias correcting climate change simulations - a critical review. *Curr Climate Change Rep* 2(4):211–220. <https://doi.org/10.1007/s40641-016-0050-x>
- Maraun D, Shepherd TG, Widmann M, et al (2017) Towards process-informed bias correction of climate change simulations. *Nature Climate Change* 7(11):764–773. <https://doi.org/10.1038/nclimate3418>
- Maraun D, Huth R, Gutiérrez JM, et al (2019) The value perfect predictor experiment: evaluation of temporal variability. *Int J Climatol* 39(9):3786–3818. <https://doi.org/10.1002/joc.5222>. <https://rmets.onlinelibrary.wiley.com/doi/abs/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: Atmospheres* 126(6):e2020JD032,824. <https://doi.org/10.1029/2020JD032824>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JD032824>
- Maraun D, Widmann M (2018) *Model output statistics*, Cambridge University Press, pp 170–200. <https://doi.org/10.1017/9781107588783.013>
- Maraun D, Widmann M (2018) *Statistical downscaling and bias correction for climate research*. Cambridge University Press, <https://books.google.fr/books?id=dqpCDwAAQBAJ>
- Michelangeli PA, Vrac M, Loukos H (2009) Probabilistic downscaling approaches: application to wind cumulative distribution functions. *Geophysical Research Letters* 36(11). <https://doi.org/10.1029/2009GL038401>
- Moron V, Robertson AW, Ward MN, et al (2008) Weather types and rainfall over senegal. part ii: downscaling of gcm simulations. *J Climate* 21(2):288–307. <https://doi.org/10.1175/2007JCLI1624.1>
- Muñoz Sabater J (2019) Era5-land monthly averaged data from 1981 to present. <https://doi.org/10.24381/cds.68d2bb3>, Accessed 22 Apr 2021
- Navarro-Racines C, Tarapues J, Thornton P, et al (2020) High-resolution and bias-corrected cmip5 projections for climate change impact assessments. *Sci Data* 7(1):7. <https://doi.org/10.1038/s41597-019-0343-8>
- Olsson J, Berggren K, Olofsson M, et al (2009) Applying climate model precipitation scenarios for urban hydrological assessment: a case study in kalmar city, sweden. *Atmospheric research - ATMOS RES* 92:364–375. <https://doi.org/10.1016/j.atmosres.2009.01.015>
- O'Neill BC, Tebaldi C, van Vuuren DP, et al (2016) The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geosci Model Dev* 9(9):3461–3482. <https://doi.org/10.5194/gmd-9-3461-2016>. <https://gmd.copernicus.org/articles/9/3461/2016/>
- Panofsky HA, Brier GW (1968) *Some applications of statistics to meteorology*. Earth and mineral sciences continuing education, College of earth and mineral sciences
- Piani C, Haerter JO, Coppola E (2009) Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor Appl Climatol* 99(1–2):187–192. <https://doi.org/10.1007/s00704-009-0134-9>
- Pierce DW, Cayan DR, Maurer EP, et al (2015) Improved bias correction techniques for hydrological simulations of climate change. *J Hydrometeorol* 16(6):2421–2442. <https://doi.org/10.1175/JHM-D-14-0236.1>. https://journals.ametsoc.org/view/journals/hydr/16/6/jhm-d-14-0236_1.xml

- Räisänen J, Räty O (2013) Projections of daily mean temperature variability in the future: cross-validation tests with ensembles regional climate simulations. *Climate Dyn* 41(5):1553–1568. <https://doi.org/10.1007/s00382-012-1515-9>
- Riahi K, van Vuuren DP, Kriegler E, et al (2017) The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Global Environ Change* 42:153–168. <https://doi.org/10.1016/j.gloenvcha.2016.05.009>. <https://www.sciencedirect.com/science/article/pii/S0959378016300681>
- Rojas R, Feyen L, Dosio A, et al (2011) Improving pan-european hydrological simulation of extreme events through statistical bias correction of rcm-driven climate simulations. *Hydrol Earth Syst Sci* 15(8):2599–2620. <https://doi.org/10.5194/hess-15-2599-2011>. <https://hess.copernicus.org/articles/15/2599/2011/>
- Schmidli J, Frei C, Vidale PL (2006) Downscaling from gcm precipitation: a benchmark for dynamical and statistical downscaling methods. *Int J Climatol* 26(5):679–689. <https://doi.org/10.1002/joc.1287>
- Smitha P, Narasimhan B, Sudheer K, et al (2018) An improved bias correction method of daily rainfall data using a sliding window technique for climate change impact assessment. *J Hydrol* 556:100–118. <https://doi.org/10.1016/j.jhydrol.2017.11.010>. <https://www.sciencedirect.com/science/article/pii/S0022169417307643>
- Switanek MB, Troch PA, Castro CL, et al (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>. <https://hess.copernicus.org/articles/21/2649/2017/>
- Terink W, Hurkmans RTWL, Torfs PJF, et al (2010) Evaluation of a bias correction method applied to downscaled precipitation and temperature reanalysis data for the rhine basin. *Hydrol Earth Syst Sci* 14(4):687–703. <https://doi.org/10.5194/hess-14-687-2010>. <https://hess.copernicus.org/articles/14/687/2010/>
- 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, Seibert J (2013) Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions. *Hydrol Earth Syst Sci* 17(12):5061–5077. <https://doi.org/10.5194/hess-17-5061-2013>
- Themeßl MJ, Gobiet A, Leuprecht A (2011) Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *Int J Climatol* 31:1530–1544
- Themeßl MJ, Gobiet A, Heinrich G (2012) Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change* 112(2):449–468. <https://doi.org/10.1007/s10584-011-0224-4>
- Thom H (1954) The rational relationship between heating degree days and temperature. *Mon Weather Rev* 82:1–6
- Thom H (1958) A note on the gamma distribution. *Mon Weather Rev* 86:117–122
- van Pelt SC, Kabat P, ter Maat HW, et al (2009) Discharge simulations performed with a hydrological model using bias corrected regional climate model input. *Hydrol Earth Syst Sci* 13(12):2387–2397. <https://doi.org/10.5194/hess-13-2387-2009>. <https://hess.copernicus.org/articles/13/2387/2009/>
- Voldoire A (2019) Cnrm-cerfacs cnrm-cm6-1-hr model output prepared for cmip6 highresmp. <https://doi.org/10.22033/ESGF/CMIP6.1387>
- Voldoire A, Saint-Martin D, Sèrès S, et al (2019) Evaluation of cmip6 deck experiments with cnrm-cm6-1. *J Adv Model Earth Syst* 11(7):2177–2213. <https://doi.org/10.1029/2019MS001683>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001683>
- Volosciuk C, Maraun D, Vrac M, et al (2017) A combined statistical bias correction and stochastic downscaling method for precipitation. *Hydrol Earth Syst Sci* 21(3):1693–1719. <https://doi.org/10.5194/hess-21-1693-2017>. <https://hess.copernicus.org/articles/21/1693/2017/>
- Vrac M (2018) Multivariate bias adjustment of high-dimensional climate simulations: the rank resampling for distributions and dependences (r^2d^2) bias correction. *Hydrol Earth Syst Sci* 22(6):3175–3196. <https://doi.org/10.5194/hess-22-3175-2018>. <https://hess.copernicus.org/articles/22/3175/2018/>
- Vrac M, Friederichs P (2015) Multivariate-intervariable, spatial, and temporal-bias correction. *J Climate* 28(1):218–237. <https://doi.org/10.1175/JCLI-D-14-00059.1>. <https://journals.ametsoc.org/view/journals/clim/28/1/jcli-d-14-00059.1.xml>
- Wang L, Chen W (2014) Equiratio cumulative distribution function matching as an improvement to the equidistant approach in bias correction of precipitation. *Atmos Sci Lett* 15(1):1–6. <https://doi.org/10.1002/asl2.454>. <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/asl2.454>
- Watanabe S, Kanae S, Sato S, et al (2012) Intercomparison of bias-correction methods for monthly temperature and precipitation simulated by multiple climate models. *Journal of Geophysical Research: Atmospheres* 117(D23). <https://doi.org/10.1029/2012JD018192>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012JD018192>

- Watterson IG, Dix MR (2003) Simulated changes due to global warming in daily precipitation means and extremes and their interpretation using the gamma distribution. *Journal of Geophysical Research: Atmospheres* 108(D13). <https://doi.org/10.1029/2002JD002928>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002JD002928>
- Watterson IG (2008) Calculation of probability density functions for temperature and precipitation change under global warming. *Journal of Geophysical Research: Atmospheres* 113(D12). <https://doi.org/10.1029/2007JD009254>. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD009254>
- Widmann M, Bedia J, Gutiérrez JM, et al (2019) Validation of spatial variability in downscaling results from the value perfect predictor experiment. *Int J Climatol* 39(9):3819–3845. <https://doi.org/10.1002/joc.6024>. <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.6024>
- Wood AW, Leung LR, Sridhar V, et al (2004) Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change* 62(1):189–216. <https://doi.org/10.1023/B:CLIM.0000013685.99609.9e>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.