



# Comparison of occurrence-bias-adjusting methods for hydrological impact modelling

Jorn Van de Velde<sup>1,2</sup>, Bernard De Baets<sup>2</sup>, Matthias Demuzere<sup>3,1</sup>, and Niko E. C. Verhoest<sup>1</sup>

<sup>1</sup>Hydro-Climatic Extremes Lab, Ghent University, Ghent, Belgium

<sup>2</sup>KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium

<sup>3</sup>Department of Geography, Ruhr-University Bochum, Bochum, Germany

**Correspondence:** Jorn Van de Velde (jorn.vandavelde@ugent.be)

**Abstract.** Over the past decade, various methods for bias adjustment of precipitation occurrence or intensity have been proposed. However, the performance of combined methods has not yet been thoroughly evaluated, especially in a hydrological and climate change context. In this study, four occurrence-bias-adjusting methods are combined with one univariate and one multivariate intensity-bias-adjusting method. The occurrence-bias-adjusting methods include thresholding, Stochastic Singularity Removal, Triangular Distribution Adjustment, and are compared with the intensity-bias-adjusting methods without specific adjustment as a baseline. These combined methods are compared with respect to precipitation amount, precipitation occurrence and discharge. This comparison, summarized in terms of the residual bias relative to both the observations and the model bias, shows significant differences in performance. Occurrence-bias-adjusting methods that add stochasticity perform worse, an effect that is reinforced by multivariate intensity-bias-adjusting methods. The use of simpler methods is thus advised until the uncertainty caused by combining methods is better understood.

## 1 Introduction

Climate change is one of the largest threats currently faced by society, with impacts on many (eco)systems. These impacts are caused by the increase in naturally occurring hazards, such as droughts, wildfires, hurricanes and floods (IPCC, 2012). To assess how these hazards are influenced by climate change conditions, a modeling chain consisting of Global Circulation Models (GCMs), Regional Climate Models (RCMs) and local impact models is commonly used (Maraun et al., 2010). The GCMs are applied to generate global future time series (Eyring et al., 2016). However, their scale is too coarse to be used directly in local impact models. This is especially true in hydrology, where local meteorological variables such as precipitation can differ substantially within a watershed. Therefore, RCMs are used to downscale the global data to a local scale. This is conducted by physically simulating the local climate, using the GCM's output as boundary conditions. RCMs with a grid resolution of 10 km are commonly used, enhancing the amount of available information (Jacob et al., 2014). Nevertheless, some systematic errors still occur in these models, due to imperfect parametrization, discretization and spatial averaging (Teutschbein



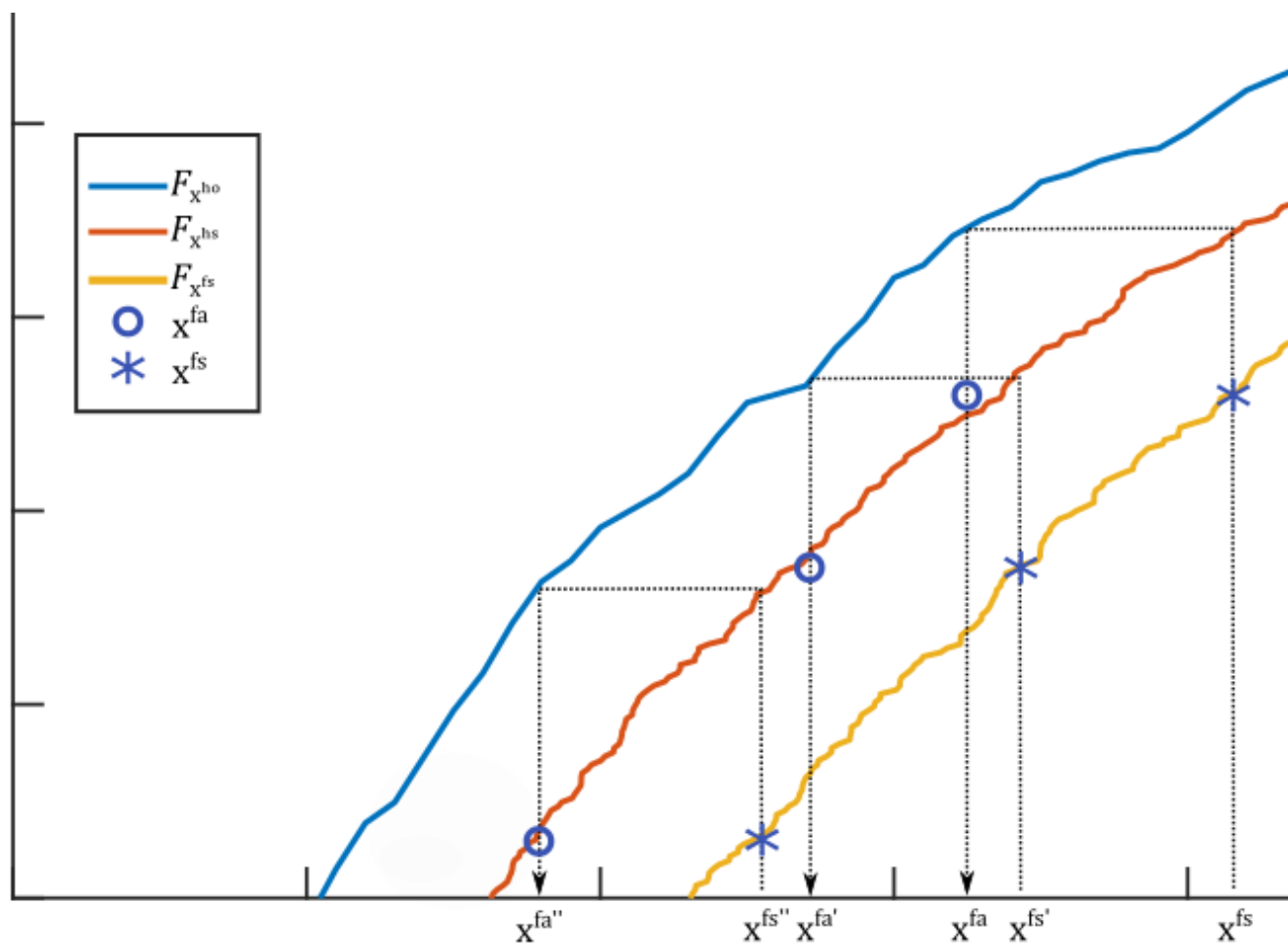
and Seibert, 2012). These errors lead to the presence of biases, most often in precipitation (Kotlarski et al., 2014). Recently, convection-permitting models, which have a grid resolution of up to 1 km have become available (Prein et al., 2015; Kendon et al., 2017; Helsen et al., 2019). However, these models have a high computational demand and their implementation is therefore no common practice yet in impact studies. The biases in coarser RCMs are generally described as *"a systematic difference between a simulated climate statistic and the corresponding real-world climate statistic"* (Maraun, 2016), though there is some debate on the use of the term ‘bias’ (Ehret et al., 2012). Biases increase the uncertainty in the last step of the modeling chain, i.e. the local impact modelling, thus entailing the necessity to adjust them (Fowler et al., 2007; Christensen et al., 2008; Maraun, 2016). Until recently, the term ‘bias correction’ was prominently used, but to emphasize that most of these methods cannot completely remove the bias, the term ‘bias adjustment’ has gained popularity. In this paper, we follow this recent trend (Vrac et al., 2016; Maraun et al., 2017; Vrac, 2018; Rätty et al., 2018; Zscheischler et al., 2019) and thus refer to ‘bias adjustment’.

During the last 15 years, many methods have been developed to overcome the above-mentioned bias problem (see Teutschbein and Seibert (2012); Gutiérrez et al. (2019) for recent overviews). These methods, which adjust the mean, variance and/or the full distribution of the variable under consideration, are well-studied for hydrological impact studies (e.g. Addor and Seibert (2014); Rätty et al. (2018)) and use a ‘transfer function’ to transfer the information from the observations to the simulations. However, an overlooked aspect in many of these methods is the adjustment of rainfall occurrence. Climate models often simulate too many rainy days, the so-called drizzle effect (Gutowski et al., 2003; Argüeso et al., 2013) and sometimes too many dry days (Thiemeßl et al., 2012). This has since long been acknowledged to be a problem, especially when the intermittence of rainfall is important (Ines and Hansen, 2006), as is the case for hydrological impact assessment. A rainy day may cause an increase in soil moisture. As a consequence, the infiltration capacity decreases, leading to more runoff and riverflow, and possibly leading to flooding. Though most bias-adjusting methods implement some basic form of frequency correction before adjusting the bias in intensity, there is limited research on the effect of these methods. One of the earliest studies on the performance of bias adjustment with respect to transition probabilities was published by Rajczak et al. (2016), while Vrac et al. (2016) published one of the first studies on the comparison of occurrence-bias-adjusting methods. The latter proposed a new occurrence-bias-adjusting method, named Stochastic Singularity Removal (SSR). This method was shown to perform well, but has only been compared with few other methods. As various other occurrence-bias-adjusting methods are also available and seem to perform well, the goal of this paper is to compare a selection of these methods and study how they interact with the subsequent intensity-bias-adjusting methods. Most of the common bias-adjusting methods, and therefore also the methods used in this comparison, stem from the quantile-mapping-method family (Panofsky et al., 1958). Quantile mapping was found to be the best-performing method, especially with respect to the simulation of hydrological time series (Rojas et al., 2011; Gudmundsson et al., 2012). This method adjusts the full distribution of the variable’s future simulations on the basis of cumulative distribution functions (CDFs) of the historical simulations (hs) and the historical observations (ho). A transfer function is composed from these CDFs, resulting in an adjusted future simulation (fa) for a day  $i$  in the time series:

$$x_i^{\text{fa}} = F_{x^{\text{ho}}}^{-1} \left( F_{x^{\text{hs}}} \left( x_i^{\text{fs}} \right) \right), \quad (1)$$



with  $f_s$  the index of the uncorrected future simulations,  $F$  a CDF, and  $F_x$  the corresponding CDF value for a certain quantile value  $x$ . This process is shown in Fig. 1.



**Figure 1.** Conceptualization of the quantile mapping method. Every  $x^{fs}$  (star) is replaced by a  $x^{fa}$  (circle) following Eq. (1), nudging the future simulations closer to the supposed future observations.

Various extensions and variants of the standard quantile mapping have been proposed. A recent variant is Quantile Delta Mapping (QDM) (Li et al., 2010; Wang and Chen, 2014; Cannon et al., 2015). It was introduced to better retain the trends of the climate model and though being criticized (Switanek et al., 2017), it is increasingly used (Mehrotra and Sharma, 2016; Nguyen et al., 2016; Cannon, 2018). Older quantile mapping methods are also still in use, such as CDF-t (Michelangeli et al. (2009), used in e.g. Vrac (2018)), and standard empirical quantile mapping, used in e.g. Rätty et al. (2018), Zscheischler et al. (2019). Nevertheless, given its generally good performance, QDM was chosen as the intensity-bias-adjusting method in this

60  
 65 comparison study.



During the last few years, multivariate intensity-bias-adjusting methods have received considerable attention. The development of these methods became important after some authors began to question the inability of univariate methods such as QDM to adjust biases in the inter-variable correlation of climate models (Hagemann et al., 2011; Wilcke et al., 2013; Hewitson et al., 2014). Subsequently, many multivariate methods have been proposed (e.g. Li et al. (2014); Mao et al. (2015);  
70 Vrac and Friederichs (2015); Mehrotra and Sharma (2016); Dekens et al. (2017); Cannon (2018); Nguyen et al. (2018)). As these methods have increased the complexity of bias adjustment, there is a growing need to study the combined effect of an occurrence-bias-adjusting method and a subsequent multivariate intensity-bias-adjusting method. To the authors' knowledge, the combination of occurrence-bias-adjusting and multivariate intensity-bias-adjusting methods has not yet been investigated. To make an initial assessment, one multivariate method is chosen to be combined with the different occurrence-bias-adjusting  
75 techniques, namely the Multivariate Bias Correction in  $n$  dimensions (MBCn) method (Cannon, 2018). This method is chosen as it is one of the more commonly implemented multivariate bias-adjusting methods (Ráty et al., 2018; Meyer et al., 2019; Zscheischler et al., 2019). Additionally, it only differs from the QDM method in the final step, where a rank-based shuffling is applied. All differences in performance may thus be linked to this step.

The use of both a univariate and a multivariate intensity-bias-adjusting method will facilitate a more complete comparison  
80 of the occurrence-bias-adjusting methods and their interaction with other methods. The methods compared in this paper include the combination of three occurrence-bias-adjusting methods with respectively a classic univariate intensity-bias-adjusting method, i.e. QDM, and a multivariate intensity-bias-adjusting method, i.e. MBCn. The three occurrence-bias-adjusting methods include an extended thresholding method, a stochastic thresholding method using the triangular distribution, and the SSR method. These can all be considered more advanced methods. Simpler methods, such as the 'positive correction' (Mao et al.,  
85 2015) and the 'threshold adaptation' method (Schmidli et al., 2006) were proven to be outperformed by SSR (Vrac et al., 2016). To enable a baseline comparison, the intensity-bias-adjusting methods are also applied without any specific occurrence-bias-adjustment, which corresponds to the direct approach as also implemented in Vrac et al. (2016). The intensity-bias-adjusting methods used here inherently apply a basic occurrence-bias-adjustment: if the observations have more dry days than the simulations, the days with the lowest precipitation amounts are automatically mapped to zero precipitation due to the quantile  
90 mapping. It also works the other way around: if the observations have less dry days, then days with a low precipitation depth are mapped to a higher precipitation amount.

The comparison in this paper is performed by calibrating on historical time series and validating on recent past time series. In order to perform a robust calibration and validation, the time series used has to be long enough (Berg et al., 2012). In Belgium, the Royal Meteorological Institute collects precipitation measurements in Uccle since the late 19th century. This time series  
95 has been used in multiple studies (Verhoest et al., 1997; De Jongh et al., 2006; Verstraeten et al., 2006; Vandenberghe et al., 2011; Willems, 2013) and will also be used in this study. All the combined methods will be compared using this rainfall time series as observations (Demarée, 2003) to adjust EURO-CORDEX RCM simulations (Jacob et al., 2014). The time series is used for the period 1970-1989 for calibration and for the period 1998-2017 for validation. This allows for the assessment of the effect of climate change trends on bias adjustment, as there is a visible impact in the validation time window (IPCC, 2013).  
100 If the data are used in this way, it is possible to gain understanding of how the bias-adjusting methods will perform under



real future conditions, e.g. from 2070-2100, a period that is often used to study future conditions. To evaluate the precipitation adjustment thoroughly, indices have been chosen for the precipitation amount, its occurrence and the resulting discharge. The use of these indices permits a clear comparison of the combined methods and their performance.

To summarize, the goal of this study is to compare the performance and hydrological impact of eight different combinations of occurrence- and intensity-bias-adjusting methods in a climate change context. This paper is structured as follows. In Section 2, the observational data and climate model simulations are addressed. In Section 3, the occurrence- and intensity-bias-adjusting methods are explained together with the evaluation and calculation set-up. In Section 4, a novel visualisation method, based on residual biases is explained and used to present the results, with a discussion in Section 5 and a conclusion in Section 6.

## 110 2 Data

### 2.1 Observations

The observational data used were obtained from the Belgian Royal Meteorological Institute (RMI) Uccle observatory. The most important time series used is the 10-min precipitation amount, gauged with a Hellmann-Fuess pluviograph, from 1898 to 2018. An earlier version of this precipitation dataset was described by Demarée (2003) and analyzed in De Jongh et al. (2006). The 10-min precipitation time series was aggregated to daily level to be comparable with the other time series used. As the last complete year was 2017, the data were used from 1901 to 2017, amounting to 117 years of daily data.

For the multivariate method, the precipitation time series was combined with a 2 meter air temperature and potential evaporation time series. The daily potential evaporation was calculated by the RMI from 1901 to 2019, using the Penman formula for a grass reference surface (Penman, 1948) with variables measured at the Uccle observatory. Daily average temperatures were obtained using measurements from 1901 to 2019. Limited by the length of the precipitation time series, the same 117 years of data as for the univariate method were used.

### 2.2 Climate simulations

For the simulations, data from the EURO-CORDEX project (Jacob et al., 2014) were used. The Rosby Centre regional climate model RCA4 was used (Strandberg et al., 2015) as it is one of the few RCMs with potential evaporation as an output variable. This RCM was forced with boundary conditions from the MPI-ESM-LR GCM (Popke et al., 2013). Historical data and scenario data for the grid cell comprising Uccle were respectively obtained for 1970-2005 and 2006-2100. The former time frame is limited by the earliest available data from the RCM. The latter time frame was only used until 2017, in accordance with the observational data. As climate change scenario, an RCP4.5 forcing was used in this paper (Van Vuuren et al., 2011). Since only 'near future' (from the model point of view) data were used, the choice of forcing does not have a large impact. However, when studying scenarios in a time frame further away from the present, using an ensemble of forcings is more relevant to be aware of the uncertainty regarding future climate change impact. Evaluations of the RCA4 model have shown that there is a bias in



precipitation, especially in winter (Strandberg et al., 2015), but this bias is in line with the biases from other EURO-CORDEX models (Kotlarski et al., 2014).

135 As Uccle (near Brussels) is situated in a region with small topographic differences, it is assumed that the conditions in Uccle can be applied anywhere within the climate model grid cell and the variance within this cell is about the same. This assumption can be made as long as the resulting adjusted data are not used for extremely localized studies, such as urban hydrology impact assessment.

### 3 Bias-adjusting methods

#### 3.1 Thresholding

140 Thresholding is one of the most common occurrence-bias-adjusting methods and has been in use for many years (e.g. Hay and Clark (2003); Schmidli et al. (2006); Ines and Hansen (2006)). This method assumes more wet days in the simulations than in the observations (Vrac et al., 2016; Themeßl et al., 2012), an idea based on the drizzle effect of RCMs (Gutowski et al., 2003). For Belgium, this assumption holds: the drizzle effect generally occurs in RCMs in north-western Europe (Themeßl et al., 2012). The simplest form of thresholding transforms all values in the climate model simulations below a certain threshold  
145 to 0. A more advanced method, used e.g. in Switanek et al. (2017), ensures that the number of days below the threshold in the observed and simulated time series is the same.

To adjust the number of wet and dry days, the frequency of dry days in the observations and in the simulations is calculated. The difference between the two frequencies,  $\Delta N$ , is the number of days of the simulation time series that has to be adapted. This adaption is done by sorting the wet days in increasing order of precipitation amount and thus only changing the  $\Delta N$   
150 lowest days of the simulation time series by setting them to 0.

In this advanced version of thresholding, some considerations are made. First, a day is considered wet if its simulated precipitation amount is above 0.1 mm, to account for measurement errors in the observations. Second, the adjustment is done on a monthly basis, to withhold a realistic temporal structure. This implies that to correct the number of wet days in month  $m$ , all days of month  $m$  of the time series are selected. Third, both historical and future simulations are adjusted at the same  
155 moment, to ensure a sound comparison during the intensity phase of the adjustment. If only either the historical or future time series would have been adjusted, the assumption that the bias can be transferred from the historical to the future time period would be impaired. The thresholding method is summarized in Algorithm 1.



---

**Algorithm 1** Thresholding

---

**Input:**

Historical observations  $X^{\text{ho}}$

Historical simulations  $X^{\text{hs}}$

Future simulations  $X^{\text{fs}}$

**Output:**

Adjusted historical  $X_{\text{out}}^{\text{hs}}$  and future simulations  $X_{\text{out}}^{\text{fs}}$

Initialization

**for**  $m = 1 : 12$  **do**

Select data for month  $m$ :  $X_m^{\text{ho}}$ ,  $X_m^{\text{hs}}$  and  $X_m^{\text{fs}}$  {Loop over months}

Calculate the percentage of dry days in month  $m$  for the historical observations

Calculate the percentage of dry days in month  $m$  for the historical simulations

Calculate  $\Delta N$  for month  $m$

{Adjustment of historical simulated time series}

Select and sort the wet days

Set the  $\Delta N$  wet days with the lowest precipitation amount to 0

Restore the original order of the wet days of month  $m$

Restore the full historical time series for month  $m$

{Adjustment of future simulated time series}

Select and sort the wet days

Set the  $\Delta N$  wet days with the lowest precipitation amount to 0

Restore the original order of the wet days of month  $m$

Restore the full future time series for month  $m$

{Reconstruction}

Replace the data in  $X_{\text{out}}^{\text{hs}}$  with the adjusted data for month  $m$

Replace the data in  $X_{\text{out}}^{\text{fs}}$  with the adjusted data for month  $m$

**end for**

---

### 3.2 Stochastic Singularity Removal

To overcome the assumption that the simulated time series has to have more wet days than the observations, Vrac et al. (2016) proposed the Stochastic Singularity Removal (SSR) method. This method temporarily removes the zeroes from all time series and reintroduces the zeroes after the bias adjustment is done, based on the frequency method mentioned in Cannon et al. (2015) and a method used to alter temperature data in Zhang et al. (2009).



The removal of zeroes is done by calculating the lowest amount of rain  $P_{min}$  in any of the time series used, including the historical observations, historical simulations and future simulations. If days with a precipitation amount below 0.1 mm are considered to be dry days (based on the measurement error, see Section 3.1),  $P_{min}$  is the lowest amount of precipitation above 0.1 mm. As Vrac et al. (2016) mention, the use of a precipitation threshold for dry days does not have any implications, but is implemented here to be in line with other implementations of the measurement error. For all days with rainfall amounts below  $P_{min}$ , a new rainfall depth  $P_{new}$  is randomly drawn from a uniform distribution on  $[0.1, P_{min}]$ . The random simulation also adds a stochastic element to this form of bias adjustment, which some authors have noted to be necessary (Maraun, 2013; Mao et al., 2015; Vrac, 2018). By using this method, the subsequent intensity-bias adjustment will be applied on all days, as if they were all wet days. There are thus no longer dry days, or singularities, in the time series, hence the method's name. After the application of the intensity-bias adjustment, a last post-processing step is applied: all days of  $X^{fs}$  with a rainfall amount below  $P_{min}$  are set to zero. The full procedure is summarized in Algorithm 2.

---

**Algorithm 2 SSR**

---

**Input:**

Historical observations  $X^{ho}$

Historical simulations  $X^{hs}$

Future simulations  $X^{fs}$

**Output:**

Adjusted historical  $X_{out}^{hs}$  and future simulations  $X_{out}^{fs}$

{Before the intensity-bias adjustment}

Determine the length of the time series  $N_{days}$

Determine  $P_{min}$  based on  $X^{ho}$ ,  $X^{hs}$  and  $X^{fs}$

**for**  $i = 1 : N_{days}$  **do**

**if**  $X^{ho}(i) < P_{min}$  **then**

    Simulate a new value

**end if**

**if**  $X^{hs}(i) < P_{min}$  **then**

    Simulate a new value

**end if**

**if**  $X^{fs}(i) < P_{min}$  **then**

    Simulate a new value

**end if**

**end for**

{After the intensity-bias adjustment}

Set all  $X^{fs} < P_{min}$  to 0

---





### 3.3 Triangular Distribution Adjustment

175 The Triangular Distribution Adjustment method (TDA), as developed by Pham (2016), is a more complex method. In the first step, the number of days to be removed from or added to the future time series is calculated using the ratio of the historical observation and simulation dry day frequency respectively  $f^{ho}$  and  $f^{hs}$ . This ratio is assumed to be the same for the future scenarios such that it can be used to calculate a corrected future dry day frequency,  $f^{fa}$ , which is in turn used to calculate the number of days  $\Delta N$  to be removed or added.

180 To make sure that the removal or addition of dry days does not change the extremes and adds stochasticity (noted to be necessary by Maraun (2013); Mao et al. (2015); Vrac (2018)), the triangular distribution, hence the method's name, is used. The general idea of this method is given in Figure 2. The CDF of the triangular distribution is given by

$$T(x) = \begin{cases} 1 - \frac{(b-x)^2}{b^2} & \text{if } 0 \leq x < b \\ 1 & \text{if } b \leq x \end{cases} \quad (2)$$

The single parameter of this triangular distribution,  $b$ , is based on a threshold value  $x_{thr}$ , to ensure that extreme values are not removed, and is calculated as follows:

$$b = F_{x^{fs}}(x_{thr} \mid x > 0.1), \quad (3)$$

with  $F_{x^{fs}}(x \mid x > 0.1)$  the CDF of the precipitation of wet days.

Dry days are added as follows:

1. For every wet day to be removed, chose a day  $t$  (with precipitation  $x_t > 0.1$ ) randomly from the wet day time series.  
 190 This day has a corresponding cumulative probability of  $\xi = F_{x^{fs}}(x_t \mid x > 0.1)$ .
2. Sample  $k$  from a uniform distribution on  $[0, 1]$ . If  $T(\xi) < k$ , then draw  $x_t$  from a uniform distribution on  $[0, 0.1]$  to replace a rainy day with a dry day.  $x$  is drawn randomly, to take into account that model simulations hardly have zero values.
3. If  $T(\xi) > k$ , then repeat from step 1.

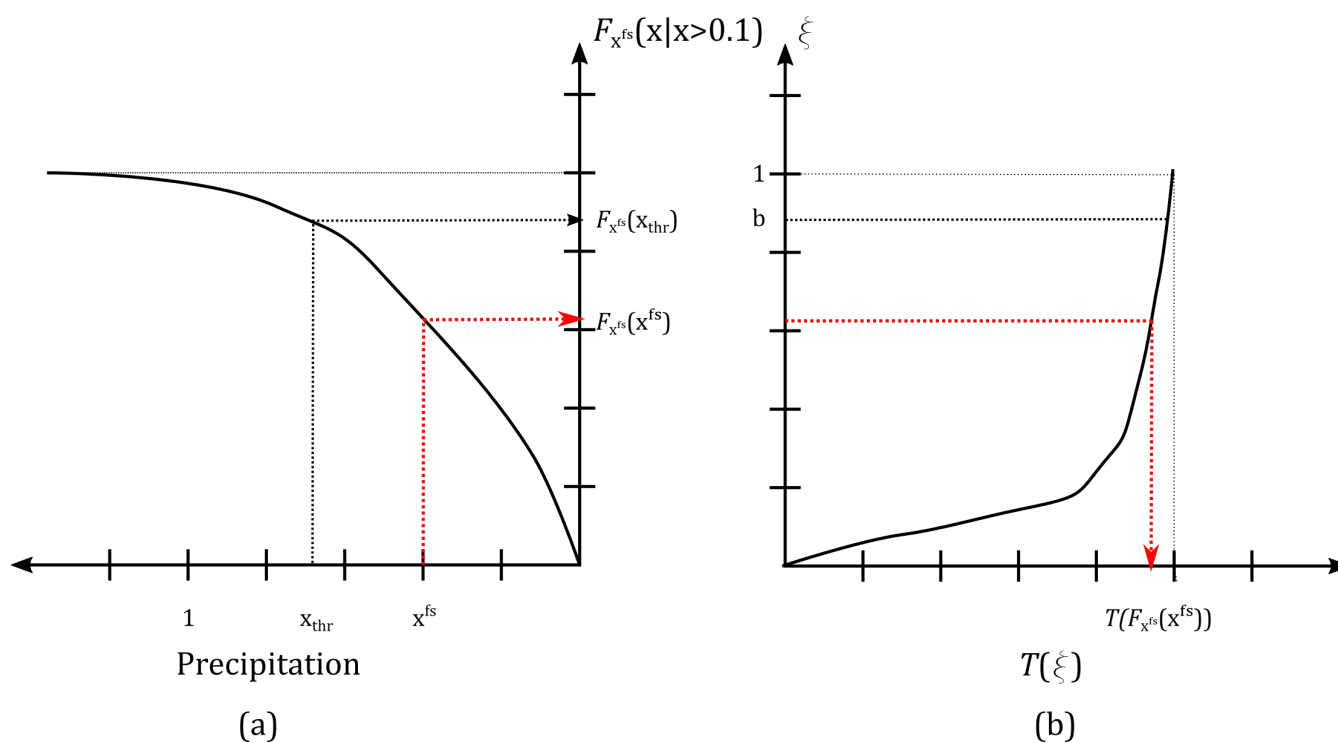
195 If dry days need to be removed, this is done using the value of  $\xi$ , which is calculated using Eq. (2):

$$T(\xi) = 1 - \frac{(b-\xi)^2}{b^2} \text{ from which } \xi = b \left( 1 - \sqrt{1 - T(\xi)} \right). \quad (4)$$

In this case, there is no restriction on the choice of days, except that they have to be dry. Thus  $\Delta N$  dry days will be randomly selected from the time series and removed, without considering the temporal structure of the time series. The process of dry day removal is given as follows:



- 200
1. For every dry day to be removed, choose a dry day  $t$  randomly from the dry day time series (with  $x_t < 0.1$ ) and sample  $k$  from a uniform distribution on  $[0, 1]$ .
  2.  $T(\xi) = k$  and thus  $\xi = b(1 - \sqrt{1 - k})$ .
  3. Then set  $x_t = F_{x^{fs}}^{-1}(\xi) = F_{x^{fs}}^{-1}(b(1 - \sqrt{1 - k}) \mid x > 0.1)$ .



**Figure 2.** Overview of the distributions used in TDA: (a) CDF of the precipitation of the wet days in the future simulation, (b) CDF of the triangular distribution. Adapted from Pham (2016).

As in thresholding, this method is applied on a monthly basis and for both historical and future simulations, which results in

205 Algorithm 3.



---

**Algorithm 3** Triangular Distribution Adjustment

---

**Input:**

Historical observations  $X^{\text{ho}}$

Historical simulations  $X^{\text{hs}}$

Future simulations  $X^{\text{fs}}$

**Output:**

Adjusted future simulations  $X_{\text{out}}^{\text{fs}}$

Initialization

**for**  $m = 1 : 12$  **do**

Select data for month  $m$ :  $X_m^{\text{ho}}$ ,  $X_m^{\text{hs}}$  and  $X_m^{\text{fs}}$  {Loop over months}

Calculate  $f^{\text{hs}}$ ,  $f^{\text{ho}}$  and  $f^{\text{fs}}$  {Begins dry day frequency calculation}

Calculate  $f^{\text{fa}} = f^{\text{fs}} \cdot f^{\text{ho}} / f^{\text{hs}}$

Calculate the difference in dry days  $\Delta N$

Select the wet days for month  $m$  {Empirical CDF}

Select the dry days for month  $m$

Calculate the empirical CDF  $F_{x^{\text{fs}}}$  for the wet days

{Adjustment}

**if**  $\Delta N > 0$  **then**

{Begins addition of dry days} **while** counter  $\leq \Delta N$  **do**

Randomly select a day  $t$

Calculate  $\xi$  using  $x_t$

Select a random value  $k$

**if**  $T(\xi) < k$  **then**

Randomly replace the value of day  $t$  with a value on  $[0,0.1]$

Add 1 to the counter

**end if**

**end while**

**else if**  $\Delta N < 0$  **then**

{Begins Removal of dry days} Set counter to 1 **while** counter  $\leq |\Delta N|$  **do**

Randomly select a day  $t$

Calculate  $\xi$

Calculate  $x_t$

Add 1 to the counter

**end while**

**end if**

Recombine the wet and dry day time series in the original order for the month  $m$

Reintroduce the adjusted data for month  $m$  in the original time series

**end for**

---



Note that Algorithm 3 is easily adapted for historical simulations by calculating  $\Delta N$  using only  $f^{\text{hs}}$  and  $f^{\text{ho}}$ , in the same way as in Algorithm 1. The parameter  $b$  was set to 0.9, as this ensures that the highest extremes are never changed to dry days, while not restricting the choice of days.

### 3.4 Quantile Delta Mapping

210 The Quantile Delta Mapping (QDM) method was first proposed by Li et al. (2010), with the idea of preserving the climate simulation trends and taking some non-stationarity into account, and was extended by Wang and Chen (2014) to be able to handle precipitation adjustment better. A comparison with other quantile mapping methods by Cannon et al. (2015) proved this method to outperform others with respect to the preservation of trends. Cannon et al. (2015) bundled both the methods by Li et al. (2010) (Equidistant CDF-matching) and Wang and Chen (2014) (Equiratio CDF-matching) under the name ‘Quantile  
215 Delta Mapping’. This name was chosen because of the link with other delta methods, as presented in Olsson et al. (2009), Willems and Vrac (2011) and Rätty et al. (2014). Consequently, the name QDM will also be used in this paper.

Mathematically, this method can be written as

$$x_i^{\text{fa}} = x_i^{\text{fs}} + F_{x^{\text{ho}}}^{-1} (F_{x^{\text{fs}}} (x_i^{\text{fs}})) - F_{x^{\text{hs}}}^{-1} (F_{x^{\text{fs}}} (x_i^{\text{fs}})) \quad (5)$$

in the additive case and

$$220 \quad x_i^{\text{fa}} = x_i^{\text{fs}} \frac{F_{x^{\text{ho}}}^{-1} (F_{x^{\text{fs}}} (x_i^{\text{fs}}))}{F_{x^{\text{hs}}}^{-1} (F_{x^{\text{fs}}} (x_i^{\text{fs}}))} \quad (6)$$

in the ratio or multiplicative case. In this paper, the additive version is used for temperature time series and the multiplicative for precipitation time series. This choice is based on the work of Wang and Chen (2014), who have shown that using the additive version for precipitation results in unrealistic precipitation values and introduced a multiplicative version.

To ensure the consistency of the time series, Themeßl et al. (2011) implemented a 61-day moving window. Here, a 91-day  
225 moving window is opted for, as suggested by Rajczak et al. (2016) and Reiter et al. (2018). This enables the adjustment of each day based on  $91 \text{ days/year} \cdot 20 \text{ years} = 1820 \text{ days}$ . These days were used to build an empirical CDF (as in Gudmundsson et al. (2012); Gutjahr and Heinemann (2013), among others), because of the ease of application. It is also important to note that for precipitation and evaporation, the adjustment was calculated and applied only on the days considered wet, e.g. with a precipitation higher than 0.1 mm.

### 230 3.5 Multivariate Bias Correction in $n$ dimensions

About 10 years ago, it became clear that multivariate correlation adjustment was necessary, as different studies started pointing out that univariate methods could never directly adjust the correlation bias in climate models (Hagemann et al., 2011; Wilcke et al., 2013; Hewitson et al., 2014). This problem encouraged the publication of many studies describing multivariate bias-adjusting methods (e.g. Li et al. (2014); Mao et al. (2015); Vrac and Friederichs (2015); Mehrotra and Sharma (2016); Dekens



235 et al. (2017); Cannon (2018); Nguyen et al. (2018)). These methods all use different approaches to describe the dependence  
 between variables. However, Cannon (2018) recently pointed out that many of these approaches are either too limited to adjust  
 dependence, as in Mehrotra and Sharma (2016) or Cannon (2016), or that they make strong stationarity assumptions, as in  
 Vrac and Friederichs (2015). Consequently, Cannon (2018) proposed a novel method, based on the N-dimensional probability  
 density function transform (N-pdft), from the field of image processing and computer vision (Pitié et al., 2005, 2007). Instead  
 240 of adjusting pixels with an RGB colour combination, this method is now used for adjusting time steps with meteorological  
 variables. Combining this with the trend-preserving QDM, Cannon named this multivariate bias-adjusting method ‘Multivariate  
 Bias Correction in  $n$  dimension’ (MBCn). Though this paper follows the recent trend to use ‘adjust’ instead of ‘correct’, the  
 name ‘MBCn’ will be used for clarity and clear reference to the original method.

In brief, the method consists of three steps that are applied iteratively until convergence is reached and a fourth step to  
 245 restore the trends. These steps are applied on the different types of data: historical observations of temperature, evaporation  
 and precipitation (combined in the matrix  $\mathbf{X}^{\text{ho}}$ ), historical climate model simulations of the three variables (the matrix  $\mathbf{X}^{\text{hs}}$ )  
 and climate model projections of the three variables (the matrix  $\mathbf{X}^{\text{fs}}$ ), which have to be adjusted. All these datasets are of  
 size  $T \times N$ , with  $T$  the number of time steps and  $N$  the number of variables. As there were no prior examples of evaporation  
 bias-adjustment, it was assumed that this variable could be adjusted in the same manner as precipitation, i.e. a multiplicative  
 250 adjustment was used and only ‘wet days’ (evaporation  $> 0.1$  mm) were adjusted.

Considering the  $j$ -th iteration of the method, denoted by the subscript  $[j]$ , the first step consists of rotating the data sets  
 using an  $N \times N$  randomly generated orthogonal rotation matrix  $\mathbf{R}_{[j]}$ . This orthogonal rotation matrix was created using the  
 algorithm by Mezzadri (2007, pg. 597). This rotation is formulated as

$$\begin{aligned}\tilde{\mathbf{X}}_{[j]}^{\text{hs}} &= \mathbf{X}_{[j]}^{\text{hs}} \mathbf{R}_{[j]} \\ \tilde{\mathbf{X}}_{[j]}^{\text{fs}} &= \mathbf{X}_{[j]}^{\text{fs}} \mathbf{R}_{[j]} \\ \tilde{\mathbf{X}}_{[j]}^{\text{ho}} &= \mathbf{X}_{[j]}^{\text{ho}} \mathbf{R}_{[j]}\end{aligned}\tag{7}$$

255 with  $\tilde{\mathbf{X}}_{[j]}$  the resulting rotated matrix. In the next step, quantile delta mapping is applied to each variable in  $\tilde{\mathbf{X}}_{[j]}^{\text{hs}}$  and  $\tilde{\mathbf{X}}_{[j]}^{\text{fs}}$ ,  
 using the corresponding variable in  $\tilde{\mathbf{X}}_{[j]}^{\text{ho}}$  as the target. The resulting matrices  $\mathbf{X}_{[j]}^{\text{hc}}$  and  $\mathbf{X}_{[j]}^{\text{fa}}$  are rotated back:

$$\begin{aligned}\mathbf{X}_{[j+1]}^{\text{hs}} &= \mathbf{X}_{[j]}^{\text{hc}} \mathbf{R}_{[j]}^{-1} \\ \mathbf{X}_{[j+1]}^{\text{fs}} &= \mathbf{X}_{[j]}^{\text{fa}} \mathbf{R}_{[j]}^{-1} \\ \mathbf{X}_{[j+1]}^{\text{ho}} &= \mathbf{X}_{[j]}^{\text{ho}}\end{aligned}\tag{8}$$

These steps have to be repeated until the multivariate distribution of  $\mathbf{X}_{[j+1]}^{\text{hs}}$  matches  $\mathbf{X}^{\text{ho}}$ , or the similarity between the two  
 matrices cannot be increased further. This is measured using the (squared) energy distance (Székely and Rizzo, 2004, 2013;



260 Rizzo and Székely, 2016), a measure of statistical discrepancy between two multivariate distributions. For two  $N$ -dimensional independent random vectors  $\mathbf{x}$  and  $\mathbf{y}$  with respective CDFs  $F$  and  $G$ , this measure is given by:

$$D^2(F, G) = 2E[\|\mathbf{x} - \mathbf{y}\|] - E[\|\mathbf{x} - \mathbf{x}'\|] - E[\|\mathbf{y} - \mathbf{y}'\|] \geq 0 \quad (9)$$

with  $E$  the expected value,  $\|\cdot\|$  the Euclidean norm and  $\mathbf{x}'$  and  $\mathbf{y}'$  and  $\mathbf{x}$  and  $\mathbf{y}$  i.i.d.. A practical way of calculating this measure is as follows (Székely and Rizzo, 2013), with  $\mathbf{X} = (X_1, \dots, X_{n_1})$  and  $\mathbf{Y} = (Y_1, \dots, Y_{n_2})$ :

$$D(\mathbf{X}, \mathbf{Y}) = \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{m=1}^{n_2} |X_i - Y_m|$$

$$- \frac{1}{n_1^2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} |X_i - X_j|$$

$$- \frac{1}{n_2^2} \sum_{l=1}^{n_2} \sum_{m=1}^{n_2} |Y_l - Y_m|, \quad (10)$$

with  $i, j, l$  and  $m$  denoting the time steps. In this study, a tolerance of 0.0001 was used to ensure that the computation would not take too long.

As a last step, the preservation of trends of QDM has to be combined with the restoration of the multivariate ranks by the transformations. To do this, first either the equidistant or equiratio version of quantile delta mapping (depending on the variable) has to be applied to each variable of the original data set  $\mathbf{X}^{\text{fs}}$ , using  $\mathbf{X}^{\text{hs}}$  and  $\mathbf{X}^{\text{ho}}$  as historical baseline data. As a final step, the elements of each column of  $\mathbf{X}^{\text{fa}}$  are reordered following a method known as the ‘Schaake Shuffle’ (Clark et al., 2004; Vrac and Friederichs, 2015). In this method, the ranks of each variable of  $\mathbf{X}^{\text{fa}}$  are swapped with the ranks of the corresponding variables of  $\mathbf{X}^{\text{fs}}_{[j+1]}$ , thus reordering the time series’ structure according to the ranks of the observations. The Schaake Shuffle can be mathematically formulated as follows (Clark et al., 2004). Let  $\mathbf{X}$  be a vector of  $n$  time steps of a variable, and  $\chi$  be the sorted vector of  $\mathbf{X}$ , that is:

$$\mathbf{X} = (x_1, x_2, \dots, x_n), \text{ and} \quad (11)$$

$$\chi = (x_{(1)}, x_{(2)}, \dots, x_{(n)}), \quad x_{(1)} \leq x_{(2)} \dots \leq x_{(n)} \quad (12)$$

Let  $\mathbf{Y}$  be a vector of  $n$  historical observations and  $\gamma$  be the sorted vector of  $\mathbf{Y}$ :

$$\mathbf{Y} = (y_1, y_2, \dots, y_n), \text{ and} \quad (13)$$

$$280 \quad \gamma = (y_{(1)}, y_{(2)}, \dots, y_{(n)}), \quad y_{(1)} \leq y_{(2)} \dots \leq y_{(n)}. \quad (14)$$

$\mathbf{B}$  is then the vector of indices describing the original observation number that corresponds to the values in the ordered vector  $\gamma$ . The main step of the Schaake Shuffle is to reconstruct the reordered vector  $\mathbf{X}^{\text{ss}}$ :

$$\mathbf{X}^{\text{ss}} = (x_1^{\text{ss}}, x_2^{\text{ss}}, \dots, x_n^{\text{ss}}) \quad (15)$$



where

$$285 \quad x_q^{ss} = x_{(r)}, \quad q = \mathbf{B}[r], \text{ and } r = 1, \dots, n \quad (16)$$

This method is essentially the only difference with the univariate QDM, implying that differences in performance can be related to it.

The MBCn method was shown by Cannon (2018) to outperform many earlier multivariate bias-adjusting methods, such as the EC-BC (Vrac and Friederichs, 2015), the JBC (Li et al., 2014), MBCr and MBCp methods (Cannon, 2016). In contrast, Cannon (2018) also pointed out some problems. Depending on the number of variables, the computational cost can get too high and convergence speed too low, or overfitting might become an issue. The first problem can be tackled by implementing sufficient time steps when having a lot of variables. Second, to address the convergence speed, Pitié et al. (2007) suggested using a deterministic selection of rotation matrices that maximizes the distance between rotation axis sets instead of randomly generating them. It is also suggested by Cannon (2018) to use the most efficient form of quantile mapping and to limit the use of an advanced quantile mapping technique to the last step. Third, to avoid overfitting (and to reduce the computational cost), early stopping is suggested (e.g. Prechelt (1998)).

### 3.6 Evaluation

To compare and evaluate different bias-adjusting methods, a logical evaluation structure is required. One of the goals of this study is to compare the bias-adjusting methods in a climate change context. To do this, 1970-1989 was chosen as the control or 'historical' time period and 1998-2017 as the validation or 'future' period. Choosing these time frames allowed for a comparison with observations in the validation period for a time period that was already affected by climate change (IPCC, 2013). For a robust calculation of the bias adjustment, 30 years of data are advised (Berg et al., 2012; Reiter et al., 2018). This decreases the effect of internal variability of the climate model and therefore decreases the uncertainty of the bias adjustment results (Maraun, 2012). However, 30 years of data would in this specific case have resulted in overlapping time series, as no earlier data are available. This would have consequently decreased the differences between the time series. The 20 years chosen as an alternative were used in a set-up called a 'pseudo-projection' (e.g. Li et al. (2010)). This evaluation set-up resembles the Differential Split-Sample Testing (DSST) implemented by Teutschbein and Seibert (2013), which is based on the work of Klemeš (1986). In DSST, the calibration and validation time series are chosen to make the difference between both as large as possible, which allows to study how robust a model is to changes.

Besides the choice of calibration and validation years, it is important to define the indices used for the evaluation (Table 1). As discussed by Maraun et al. (2015) and Maraun and Widmann (2018), it is important to use indices that are both directly and indirectly affected by the bias adjustment. They argue that only using directly adjusted indices can possibly be misleading, especially when using cross-validation. To account for this, indices based on the discharge  $Q$  were used (indices 2 and 3). These indices allowed for the assessment of the impact of the occurrence-bias-adjusting methods on simulated river flow. As a final important group of indices, occurrence indices were used, in order to assess how the combined methods differ in changing the precipitation occurrence of the time series (indices 4 to 7). For the precipitation amount, the quantiles of the empirical



**Table 1.** Overview of the indices used

Nr	Index	Name
1	$P_x$	Precipitation amount quantiles, with $x$ the quantile considered
2	$Q_x$	Discharge quantiles, with $x$ the quantile considered
3	$Q_{T20}$	20-year return period value of discharge
4	$P_{P00}$	Precipitation transition probability from a dry to a dry day
5	$P_{P10}$	Precipitation transition probability from a wet to a dry day
6	$N_{dry}$	Number of dry days
7	$P_{lag1}$	Precipitation lag-1 autocorrelation

distributions were considered. Expert Team on Climate Change Detection and Indices (ETCCDI) precipitation indices (Zhang et al., 2011) have also been considered and calculated. However, these are not included in this paper, as the ETCCDI indices did not allow to discern between the different combined methods. All indices were calculated taking all days into account, instead of only calculating them on wet days.

Similar to Pham et al. (2018), we use the Probability Distributed Model (PDM, Moore (2007)), a lumped conceptual rainfall-runoff model to calculate the discharge for the Grote Nete watershed in Belgium. The PDM as used here was calibrated by Cabus (2008) using the Particle Swarm Optimization algorithm (PSO, Eberhart and Kennedy (1995)). As in Pham et al. (2018), it was assumed that the differences between meteorological conditions in the Grote Nete-watershed and Uccle are negligible, and that thus the adjusted data for the Uccle grid cell can be used as a forcing for the PDM. Furthermore, the goal is not to make predictions, but to assess the impact of different post-processing chains on the discharge values. To calculate the bias on the indices, both the raw and adjusted RCM time series were used as forcing for this model.

### 3.6.1 Calculation set-up

As some of the bias-adjusting methods include randomness, it is important to have enough computer simulations to account for variability. Yet, to ensure the availability of sufficient computing capacity, just 20 calculations were carried out for each stochastic step. Only 1 calculation is used in the case of a deterministic method, as more simulations will always result in the same outcome. All methods used here, except for SSR and TDA, are deterministic, thus only the combinations of these methods with respectively QDM and MBCn were simulated 20 times. After these multiple calculations were carried out, the value of each index was first calculated on the basis of every simulation. Then, the values were averaged over the simulations; these averaged values were used for the comparison of the methods. Biases on the indices are always calculated as simulations minus observations, indicating a positive bias if the simulations are larger than the observations and vice versa.





## 4 Results

The results are shown and discussed according to the used uni- or multivariate intensity methods. For each intensity method, the results are shown for three coherent groups of indices used: precipitation amount, precipitation occurrence and discharge.

340 To summarize and compare the performance of the methods for each of the index groups, the residual biases for the indices of each combination of occurrence-bias-adjusting and intensity-bias-adjusting methods were calculated, based on the ‘added value’ concept (discussed in Di Luca et al. (2015)). This residual bias can be calculated relative to the model bias or the observations. This enables a comparison based on how well the methods perform in removing the bias and the size of the bias removal in comparison with the original value for the corresponding index for the observation time series. The residual bias  
345 relative to the observations  $RB_O$  for an index  $k$  is calculated as follows

$$RB_{O_k} = 1 - \frac{|\text{bias}_{\text{raw},k}| - |\text{bias}_{\text{adj},k}|}{|\text{obs}_k|} \quad (17)$$

wit  $\text{obs}_k$ ,  $\text{bias}_{\text{raw},k}$  and  $\text{bias}_{\text{adj},k}$  respectively the value of the observations for index  $k$ , the bias of the raw climate model simulations and of the adjusted climate simulations for index  $k$ .

The residual bias relative to the model bias  $RB_{MB}$  for an index  $k$  is calculated as follows

350

$$RB_{MB_k} = 1 - \frac{|\text{bias}_{\text{raw},k}| - |\text{bias}_{\text{adj},k}|}{|\text{bias}_{\text{raw},k}|} \quad (18)$$

The absolute values are used for calculation to indicate the distance between the raw and adjusted values, thus neglecting if the sign of the bias is changed. If the values of these residual biases are lower than 1 for an index, the method performs better than the raw RCM for this index. For the  $RB_{MB}$  it is impossible to have values lower than 0. However, for  $RB_O$ , this is possible, if the value after bias adjustment is exceptionally small in comparison to the value of the observations. The best methods have  
355 low scores on both residual biases for their indices.

The original data used for these residual biases, i.e. the observational data, the raw climate simulations and the effective biases of the adjusted simulations, can be found in Appendix A.

### 4.1 Univariate bias adjustment

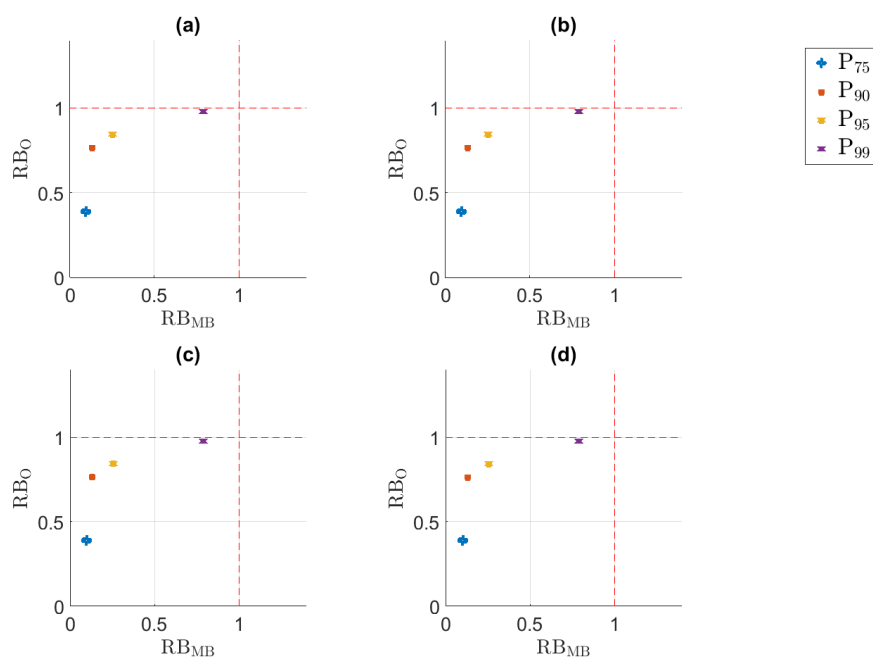
For the univariate case, QDM was used without a specific occurrence correction, and in combination with thresholding, SSR  
360 and TDA. This yields four time series, which will be compared on basis of the precipitation amount, precipitation occurrence and discharge indices.

#### 4.1.1 Precipitation amount

In Fig. 3, it can be seen that for all occurrence-bias-adjusting methods, the adjusted simulations perform better than the raw climate simulations, for most quantiles, with values ranging from 0.39 to 0.98 for  $RB_O$  and from 0.10 to 0.79 for  $RB_{MB}$ . Only



365 the highest quantile (99.5) performs worse than the climate simulations and can thus not be plotted. This quantile has for all  
 occurrence methods a value of 1.09 for  $RB_O$  and of 2.69 for  $RB_{MB}$ . This is as expected: the QDM method has been proven in  
 other papers to generally perform well (Cannon et al., 2015). However, in contrast to many other quantile mapping methods  
 (e.g. Li et al. (2010)), quantile delta mapping does not apply extrapolation for the highest quantiles. This extrapolation is often  
 advised for these quantiles, as their simulated values might be larger than the largest values of the observations. Applying no  
 370 extrapolation causes underestimation, which is demonstrated here. This implies that QDM might perform even better if this  
 extrapolation or a parametric approach were implemented, as was the case in the original version of Li et al. (2010). This effect  
 is also visible in the other quantiles: the values for both  $RB_O$  and  $RB_{MB}$  rise with the higher quantiles. This implies that it is  
 harder to correctly adjust the bias for the higher quantiles and that the 90th quantile and higher quantiles are also influenced by  
 this extrapolation problem. Three other quantiles cannot be plotted besides the 99.5th: the 5th, the 25th and 50th quantile. The  
 375 value for the observations of the former two is 0 mm, which causes the  $RB_O$  to be impossible to calculate. The latter is very  
 small for  $RB_O$ , caused by a decrease in bias that is larger than the observed value.



**Figure 3.**  $RB_{MB}$  versus  $RB_O$  for the univariate precipitation amount quantiles. (a) without specific adjustment, (b) thresholding, (c) SSR, (d) TDA.

As for the comparison of the occurrence-bias-adjusting methods, all methods perform similarly. This was to be expected, as the difference in dry days between the methods is relatively small compared to the total number of dry days and its influence is thus negligible.



#### 380 4.1.2 Precipitation occurrence

For the precipitation occurrence, the indices indicate that all bias-adjusted simulations perform better than the climate simulations (Fig. 4). The values for  $RB_0$  range from 0.54 to 0.88 and from 0 to 0.46 for  $RB_{MB}$ . The  $RB_{MB}$  values indicate that the bias is almost entirely removed, although the  $RB_0$  values imply that this bias removal does not add much value in comparison to the observations.

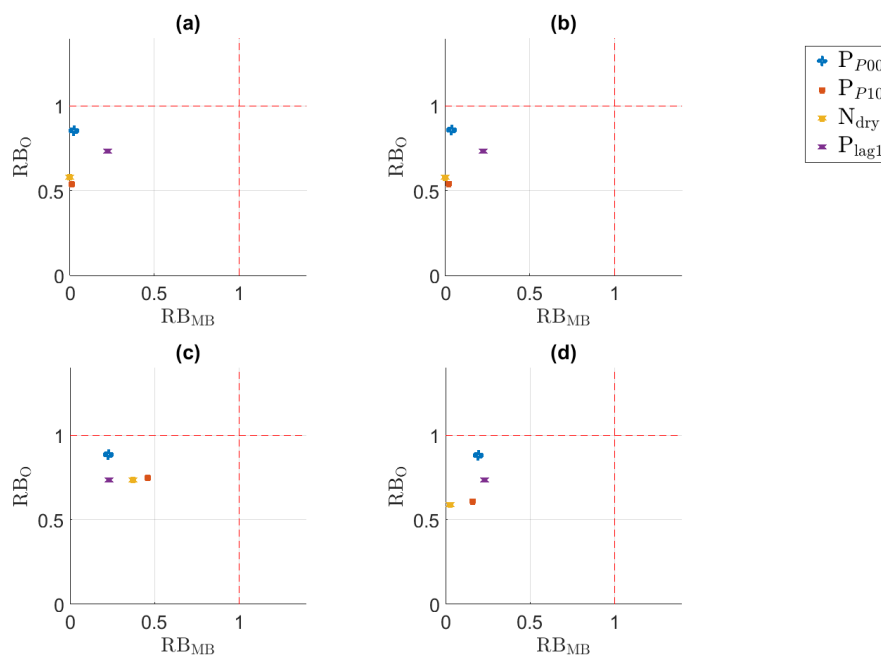
385 In contrast to the precipitation amount indices, the performances of the transition probabilities depend on the occurrence-bias-adjusting method used. The best values are obtained without a specific occurrence-bias-adjusting method or when the thresholding method is used. The SSR method performs the worst of all methods considered, while the TDA has an intermediate performance. An effect by the occurrence-bias-adjusting methods on the transition probabilities was expected by construction. However, the differences noticed here are not in line with the expectations. It is often assumed that additional randomness,  
390 which is introduced by the TDA and SSR methods, brings the simulations closer to the observations. Our results seem to contradict this assumption, as TDA and SSR have a worse performance than thresholding or the baseline without any specific occurrence-bias-adjustment. This indicates that the climate models simulated the wet and dry days at the right positions in the time series. In that case, the additional randomness induced on wet days by TDA and SSR might change the simulated time series too much. This is especially striking for the number of dry days, which SSR adjusts worst. Whereas the other methods  
395 completely remove the bias in number of dry days, the simulations are still biased after occurrence-bias-adjustment by SSR. This implies that the final step of SSR, the reintroduction of dry days, reintroduces too few dry days.

#### 4.1.3 Discharge

For the discharge, the quantiles and the 20-year return period value all indicate a better performance of the adjusted climate simulations than the original climate simulations, as summarized in Fig. 5. The values for  $RB_0$  range from 0.04 to 0.74 and  
400 from 0.01 to 0.34 for  $RB_{MB}$ . This is in line with the precipitation amount quantile results mentioned above and implies that adjusted simulations are almost bias-free. The range in values for  $RB_0$  is mostly related to the small absolute range of the bias values, ranging from  $-0.03 \text{ m}^3/\text{s}$  to  $0.97 \text{ m}^3/\text{s}$ .

There are some small differences between the occurrence-bias-adjusting methods, indicating that the combination of the non-linear PDM and the occurrence-bias-adjusting methods increases the effect of the latter. However, these differences are subtle  
405 and it is hard to discern the best method. Only the 99th quantile for the TDA method indicates a slightly better performance than the other methods. However, this difference is small: TDA has a  $RB_{MB}$  value of 0.03 for this quantile whereas the other methods have values of 0.08 or 0.09.

The small, almost indiscernible differences seem to imply that for the combinations of QDM and occurrence-bias-adjusting methods, the more advanced methods do not bring any added value.



**Figure 4.**  $RB_{MB}$  versus  $RB_0$  for the univariate precipitation occurrence indices. (a) without specific adjustment, (b) thresholding, (c) SSR, (d) TDA.

## 410 4.2 Multivariate bias adjustment

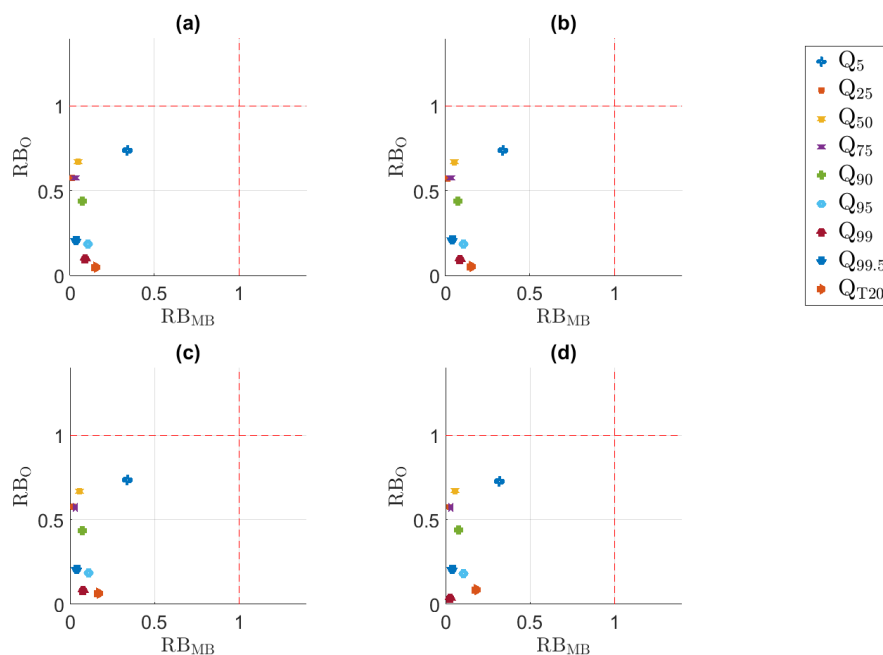
For the multivariate case,  $MBC_n$  was used in combination with no specific occurrence correction, thresholding, SSR and TDA. As in the univariate case, this yields four resulting time series, which will be compared on the basis of the precipitation amount, precipitation occurrence and discharge indices.

### 4.2.1 Precipitation amount

415 Figure 6 illustrates that the results for the precipitation amount indices of the combinations of occurrence-bias-adjusting methods and  $MBC_n$  are essentially the same as those for QDM. This follows from the construction of  $MBC_n$ : in the last steps, QDM is applied and the resulting time series is changed on the basis of the ranks of the previous steps of  $MBC_n$ , using the Schaake Shuffle. A change in ranks has no influence on the CDF and corresponding quantiles of precipitation amount, which explains that the plots are the same. More information on the specific quantiles can be found above, in Subsection 4.1.1.

### 420 4.2.2 Precipitation occurrence

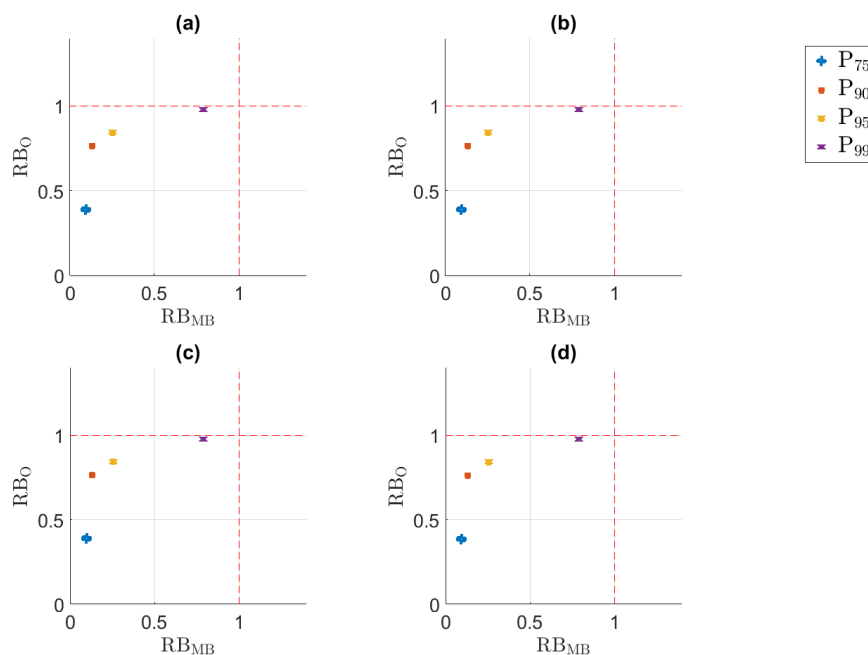
Unlike all other index groups, for the combinations with both univariate and multivariate intensity-bias-adjusting methods, the precipitation occurrence indices for the combination of  $MBC_n$  and occurrence-bias-adjusting methods indicate that the



**Figure 5.**  $RB_{MB}$  versus  $RB_0$  for the univariate discharge quantiles and the 20 year return period value. (a) without specific adjustment, (b) thresholding, (c) SSR, (d) TDA.

adjusted climate simulations generally perform worse than the raw climate simulations (Fig. 7). The wet-to-dry transition probability and precipitation lag-1 autocorrelation are close to the 1-1 line in the plots, but the dry-to-dry transition probability has  $RB_{MB}$  values above 1.5 for every combined method, as can be seen in Table 2. The only two exceptions to the generally poor performance are the wet-to-dry transition probability for SSR and the number of dry days. The former is contrasting with all other index groups, in which the SSR method often performed the worst. However, SSR performed worst for the dry-to-dry transition probability, with values of 1.23 and 2.56 for respectively  $RB_0$  and  $RB_{MB}$ . The number of dry days for the different methods is the same as in the combinations of QDM and the occurrence-bias-adjusting methods. These are unaffected by the multivariate step of MBCn, as was the case for the precipitation intensity in Section 4.2.1.

Except for SSR, the influence of the occurrence-bias-adjusting methods on precipitation occurrence is negligible, with e.g. values for  $P_{P10}$  ranging from 1.03 to 1.04 for  $RB_0$  and ranging from 1.07 to 1.09 for  $RB_{MB}$ . This implies that the performance is largely influenced by the intensity-bias-adjusting method. The only difference between the multivariate and the univariate intensity-bias-adjusting methods is the rank-based Schaake Shuffle, as previously mentioned, which implies that this shuffling might be responsible for the performance. As the shuffle changes the time series structure, it is a plausible cause for the poor performance of occurrence-related indices. The post-processing done by the SSR method might then balance this effect by reintroducing some dry days, thus increasing the wet-to-dry transition probability. This is in contrast with the other



**Figure 6.**  $RB_{MB}$  versus  $RB_O$  for the multivariate precipitation amount quantiles. (a) without specific adjustment, (b) thresholding, (c) SSR, (d) TDA.

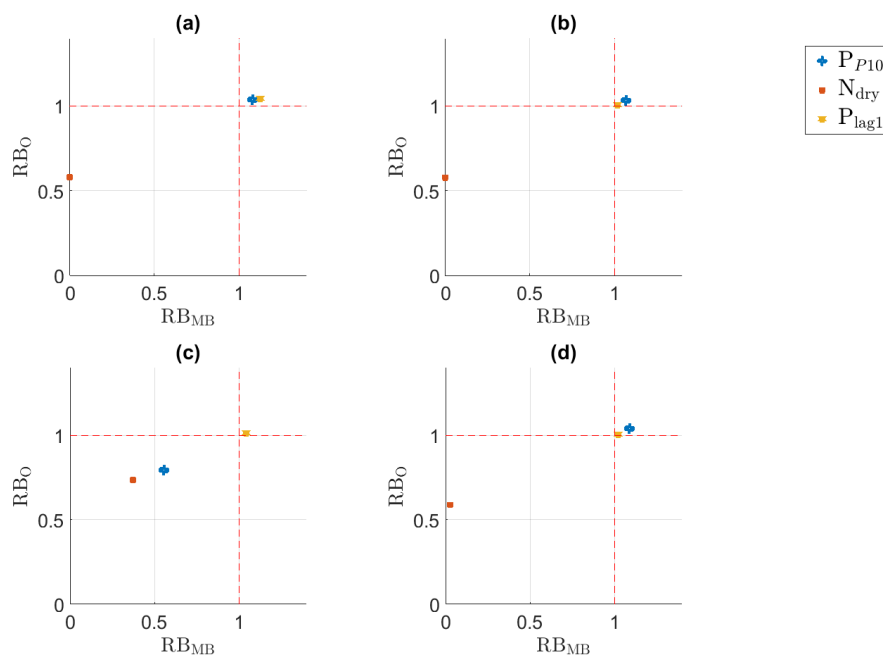
**Table 2.**  $RB_{MB}$  and  $RB_O$  values of  $P_{P00}$  for the combinations of MBCn and the occurrence-bias-adjusting methods

Occurrence-bias-adjusting method	$RB_{MB}$	$RB_O$
Without specific adjustment	1.84	1.12
Thresholding	1.81	1.12
SSR	2.56	1.23
TDA	1.73	1.11

occurrence-bias-adjusting methods, which do not apply post-processing after the intensity correction step. However, this would also explain the poor performance for  $P_{P00}$  for SSR, as the post-processing also changes the dry-to-dry transition probability by reintroducing dry days. This changes the structure of dry day sequences and thus implies an extra change in the temporal structure on top of the effect of the Schaake Shuffle.

### 4.2.3 Discharge

The results of the adjusted climate simulations for discharge (Fig. 8) are all better than the raw climate simulations. In this case, the values for  $RB_O$  range from 0.09 to 0.97 and from 0.03 to 0.91 for  $RB_{MB}$ . Both indices imply that the added value of the bias

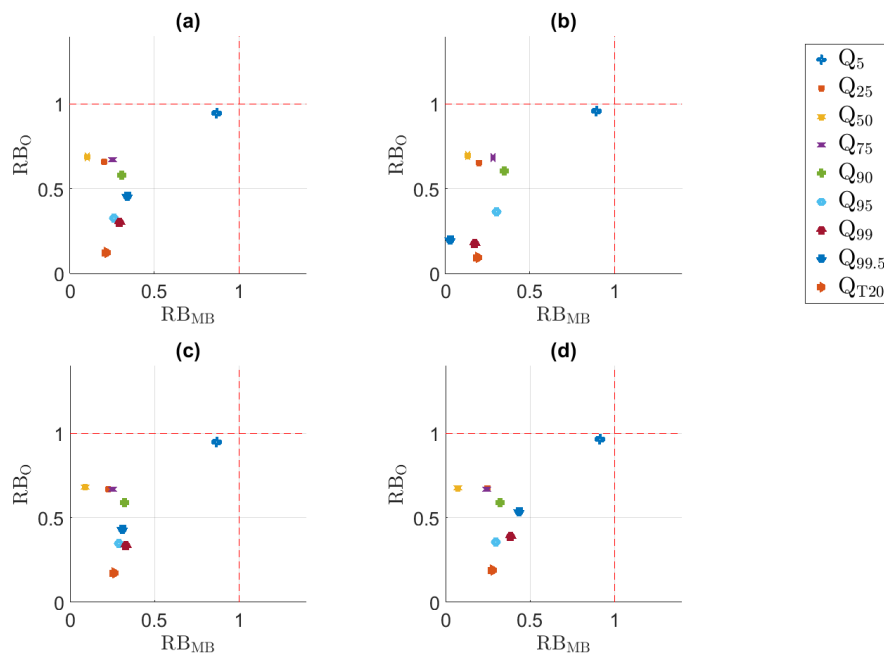


**Figure 7.**  $RB_{MB}$  versus  $RB_0$  for the multivariate precipitation occurrence indices. (a) without specific adjustment, (b) thresholding, (c) SSR, (d) TDA.

445 adjustment depends strongly on the quantile under consideration. For the lowest (5th) discharge quantile, the performance of the adjusted climate simulations is only marginally better than for the raw climate simulations. With values ranging from only 0.95 to 0.97 and from 0.97 to 0.91 for respectively  $RB_0$  and  $RB_{MB}$ , it can be assumed that the influence of the occurrence-bias-adjusting methods on the lowest discharge quantiles is negligible. However, the performance of MBCn combined with the occurrence-bias-adjusting methods for the lowest quantile shown here is worse than for the same quantile in Fig. 5. As

450 mentioned in Section 3.5, the reordering by the Schaake Shuffle is the only difference with QDM. Therefore, it can be assumed that changing the time series structure by the rank-based reordering breaks sequences of low precipitation amounts, which results in less or different low flow sequences. In contrast to the 5th quantile, the resulting performance for the combinations under consideration is better for the highest quantiles, for both  $RB_0$  and  $RB_{MB}$ . This indicates that the climate model bias was large and most of this bias has been removed by the bias-adjustment.

455 When comparing the occurrence-bias-adjusting methods, variability in their performance can be noticed for the indices under consideration, unlike in Section 4.2.1. However, as can be seen in Table 3, the differences are small for most quantiles. This implies that there is no single method that performs considerably better than any other. Thus for this case, the assumption that randomized steps nudge the adjusted climate simulations closer to the observations does not hold, as was the case for the univariate discharge quantiles in Section 4.1.3.



**Figure 8.**  $RB_{MB}$  versus  $RB_O$  for the multivariate discharge quantiles and 20 year return period value. (a) without specific adjustment, (b) thresholding, (c) SSR, (d) TDA.

**Table 3.** Maximal difference among the discharge indices for the combinations with MBCn for  $RB_{MB}$  and  $RB_O$

Index	$RB_{MB}$	$RB_O$
Q <sub>5</sub>	0.05	0.02
Q <sub>25</sub>	0.05	0.02
Q <sub>50</sub>	0.06	0.02
Q <sub>75</sub>	0.04	0.02
Q <sub>90</sub>	0.04	0.03
Q <sub>95</sub>	0.04	0.04
Q <sub>99</sub>	0.21	0.21
Q <sub>99.5</sub>	0.41	0.34
Q <sub>T20</sub>	0.09	0.10

460 The range of values is much larger for the highest quantiles, i.e. the 99th and 99.5th quantiles, which contribute to the most extreme events. These quantiles have ranges of respectively 0.21 and 0.34 for  $RB_O$  and 0.21 and 0.41 for  $RB_{MB}$ . It is unexpected that these are influenced by the occurrence-bias-adjusting methods, as these methods are supposed to only change the low precipitation amount values. However, the combination of changes in low precipitation amount values with





the rank-based Schaake Shuffle and the non-linearity of the rainfall-runoff model might cause this effect. This becomes more  
465 clear in comparison with Fig. 5, which includes the same steps, except the Schaake Shuffle based on the preceding rotations.  
As mentioned above, the highest quantiles are the most sensitive to these structural changes (Switanek et al., 2017). When  
comparing the CDFs of the combinations under consideration (figure not shown), this sensitivity of the quantiles is visible.  
This seems to imply that the small effect of the occurrence-bias-adjusting methods (which can be seen in Fig. 4) is reinforced  
by the multivariate MBCn method. This difference in effect of methods becomes clearer when comparing the highest quantiles  
470 for thresholding with the other occurrence-bias-adjusting methods: the 99.5th quantile performs much better for this method  
than for the other methods, with values for  $RB_O$  and  $RB_{MB}$  close to zero. As in Section 4.1.2, the best performing method is  
not one of the advanced methods, but the simpler thresholding method.

Table A2 shows what the poor performance of the highest quantiles means in practice: the bias in the climate simulation is  
an overestimation, and whereas all methods correct this overestimation, only the thresholding method does correct too much,  
475 implying that its temporal structure is the closest to the observations and that the temporal structure of the other methods does  
not allow the precipitation to build up the antecedent conditions for extreme floods. When comparing the time series of the  
adjusted climate simulations (figures not shown), it can be seen that the extreme floods in time series adjusted by combinations  
of QDM and occurrence-bias-adjusting methods are situated in periods with strong antecedent conditions, i.e. more than one  
week of daily rainfall and can be caused by moderate rainfall peaks on top of these antecedent conditions. In contrast, in time  
480 series adjusted by combinations of MBCn and occurrence-bias-adjusting methods, the antecedent conditions are less strong,  
with only a few days of rainfall or even dry days shortly before the flood event: the floods are caused by extreme rainfall peaks.  
The combination of MBCn with the thresholding occurrence-bias-adjusting method is situated somewhat in between, causing  
extreme floods in periods of moderate-to-strong antecedent conditions and extreme rainfall peaks.

## 5 Discussion

485 When all methods are compared, the performance of the adjusted climate simulations is for the indices considered here generally  
better than the climate simulations. As a recent time period, 1998-2017, is used for validation purposes, the good performance  
shows that so far, for precipitation bias adjustment, the stationarity assumption holds and that the adjustments are  
meaningful in a climate change context. However, this does not give any indication of the time for which this assumption will  
continue to hold. Only for the precipitation occurrence indices of the combinations of MBCn and the occurrence-bias-adjusting  
490 methods (Fig. 7), the performance is worse than the climate simulations.

When comparing the univariate and multivariate intensity-bias-adjusting methods, some major differences can be noticed.  
For QDM, the effect of the occurrence-bias-adjusting methods was only seen on the precipitation occurrence indices, which  
was to be expected. In contrast, for the combinations of the multivariate intensity-bias-adjusting methods and the occurrence-  
bias-adjusting methods, there were differences in both the discharge and precipitation occurrence indices. These differences  
495 were most pronounced for the methods that apply randomization: they performed unexpectedly worse (Figs. 4 and 8). This  
is in contrast with the findings of Vrac et al. (2016), who demonstrated the superior performance of SSR in comparison with



other methods. However, Vrac et al. (2016) used a different combination of occurrence and intensity-bias-adjusting methods. First, the occurrence-bias-adjusting methods that were proven to be outperformed by SSR by Vrac et al. (2016), the positive correction and threshold adaptation, were not used in this paper. Second, another univariate method, CDF-t (Michelangeli et al.,  
500 2009) was used. Third, no multivariate method was used in the comparison of Vrac et al. (2016). Besides, the geographical region was different, though comparable, as Vrac et al. (2016) used the complete EURO-CORDEX region.

Nonetheless, in the univariate case, which is most comparable to the study by Vrac et al. (2016), both randomness-inducing methods perform worst. The combinations of QDM without a specific adjustment method and with a simpler method such as thresholding, perform best in the univariate precipitation occurrence case (Fig. 4). In the multivariate discharge case (Fig. 8)  
505 thresholding performed especially good. This implies that other multivariate methods, which apply different algorithms to restore the multivariate correlation structure, might have a varying impact on the occurrence correction. This has to be studied in more detail.

The effect of using a multivariate intensity-bias-adjusting methods as such is also interesting. When assessing the precipitation occurrence for the combinations with MBCn, it was noticed that the poor performance seemed to be more related to  
510 the use of the rank-based Schaake Shuffle alone, than to the use of occurrence-bias-adjusting methods, as they mostly had a similar performance. This effect could also be noticed for the discharge indices for the combinations with MBCn (Fig. 8), though in those combinations the effect of the occurrence-bias-adjusting methods is slightly larger. However, there is not a lot of information available yet on the effect of multivariate intensity-bias-adjusting methods on precipitation occurrence, though this paper partially fills the gap.

## 515 6 Conclusions

In this paper, different occurrence-bias-adjusting methods were compared in combination with univariate and multivariate intensity-bias-adjusting methods. The occurrence-bias-adjusting methods compared were thresholding, Stochastic Singularity Removal (SSR) and Triangular Distribution Adjustment (TDA). The intensity-bias-adjusting methods were also run once without a specific adjustment method. The intensity-bias-adjusting methods used were QDM (univariate) and MBCn (multivariate).  
520 The combinations of occurrence-bias-adjusting and intensity-bias-adjusting methods were compared for precipitation amount, precipitation occurrence and discharge indices, using the residual bias relative to the model bias and the residual bias relative to the observations.

These comparisons have shown that the performance is very specific depending on the combination of bias-adjusting methods. Two main observations could be made. First, in both the univariate and multivariate combinations the randomness-inducing  
525 methods, SSR and TDA, performed generally worse than the more traditional methods such as thresholding. They performed even worse than the baseline where the intensity-bias-adjusting method was applied. Second, the multivariate method has a large influence on the performance of the occurrence-bias-adjusting methods, by changing the time series structure and possibly reinforcing the effect of the occurrence-bias-adjusting methods, worsening the performance of the resulting adjusted climate simulations. In the multivariate case, small changes made by the occurrence-bias-adjusting methods thus seem to have



530 an unexpectedly large influence. This implies that more research should be done on the interaction of multivariate intensity-  
bias-adjusting methods and occurrence-bias-adjusting methods, to clarify how combined methods influence the assessment of  
climate change impact. The effect of multivariate methods themselves on the time series structure is also uncertain, and will  
be investigated in further research. This is especially important for the assessment of hazards that depend on the time series  
structure, as is the case in hydrology and agriculture. This also has to be assessed for other geographic/climatic regions, where  
535 precipitation characteristics and in turn the interaction between subsequent bias-adjusting methods may be different. To con-  
clude, as long as the effect of the interactions between bias-adjusting methods is unclear, the use of simpler methods seems  
recommended to reduce the uncertainty as much as possible when assessing and communicating global change impacts. In the  
framework of this study, the simpler method would be the thresholding occurrence-bias-adjusting method.

*Code and data availability.* The code for the calculations is publicly available at <https://doi.org/10.5281/zenodo.3675305> (Van de Velde,  
540 2020). The RCA4 data are downloaded and are available from the Earth System Grid Federation data repository. The local observations were  
obtained from RMI in Belgium, and cannot be shared with third parties.

*Author contributions.* JV, BDB and NV designed the experiments. JV developed the code and performed the simulations. JV, BDB and  
NV designed the visualization. JV prepared the manuscript with contributions from MD, BDB and NV. All co-authors contributed to the  
interpretation of the results.

545 *Competing interests.* The authors declare that they have no conflict of interests.

*Acknowledgements.* The authors are grateful to the RMI for allowing the use of 117-year Uccle dataset. This work was funded by FWO,  
grant number G.0039.18N.



## Appendix A: Appendix A

**Table A1.** Observed values, and biases for the raw climate simulations, and the combinations of QDM and occurrence-bias-adjusting methods.

Index	Observed value	Bias				
		Raw climate simulations	Without specific adjustment	SSR	TDA	Thresholding
$Q_5$ (m <sup>3</sup> /s)	2.30	0.92	-0.31	-0.31	-0.29	-0.31
$Q_{25}$ (m <sup>3</sup> /s)	3.36	1.45	0.02	0.03	0.03	0.02
$Q_{50}$ (m <sup>3</sup> /s)	4.39	1.53	0.08	0.08	0.09	0.08
$Q_{75}$ (m <sup>3</sup> /s)	5.72	2.52	-0.08	-0.08	-0.08	-0.08
$Q_{90}$ (m <sup>3</sup> /s)	7.83	4.76	-0.36	-0.35	-0.37	-0.36
$Q_{95}$ (m <sup>3</sup> /s)	10.09	9.22	-1.01	-1.00	-0.97	-1.00
$Q_{99}$ (m <sup>3</sup> /s)	18.71	18.58	-1.67	-1.41	-0.53	-1.65
$Q_{99.5}$ (m <sup>3</sup> /s)	23.90	19.70	-0.77	0.78	0.74	0.84
$Q_{T20}$ (m <sup>3</sup> /s)	48.69	54.61	8.30	9.13	9.99	8.36
$P_5$ (mm)	0.00	0.00	0.00	0.00	0.00	0.00
$P_{25}$ (mm)	0.00	0.08	0.00	0.00	0.02	0.00
$P_{50}$ (mm)	0.10	1.01	0.05	0.06	0.04	0.05
$P_{75}$ (mm)	2.70	1.83	-0.18	-0.18	-0.19	-0.18
$P_{90}$ (mm)	7.40	1.99	-0.26	-0.26	-0.26	-0.26
$P_{95}$ (mm)	11.42	2.38	-0.61	-0.61	-0.61	-0.61
$P_{99}$ (mm)	21.80	2.36	-1.86	-1.86	-1.86	-1.86
$P_{99.5}$ (mm)	29.09	1.56	-4.20	-4.20	-4.20	-4.20
$P_{P00}$	0.65	-0.10	0.00	-0.02	-0.02	0.00
$P_{P10}$	0.32	-0.15	0.00	-0.07	0.02	0.00
$N_{dry}$	3470.00	-1466.00	0.00	-544.70	38.00	0.00
$P_{lag1}$	0.33	0.11	0.03	0.03	0.03	0.03



**Table A2.** Observed values, and biases for the raw climate simulations, and the combinations of MBCn and occurrence-bias-adjusting methods.

Index	Observed values	Bias				
		Raw climate simulations	Without specific adjustment	SSR	TDA	Thresholding
$Q_5$ (m <sup>3</sup> /s)	2.30	0.92	0.79	0.80	0.84	0.82
$Q_{25}$ (m <sup>3</sup> /s)	3.36	1.45	0.29	0.33	0.36	0.29
$Q_{50}$ (m <sup>3</sup> /s)	4.39	1.53	-0.16	-0.14	-0.11	-0.20
$Q_{75}$ (m <sup>3</sup> /s)	5.72	2.52	-0.64	-0.63	-0.62	-0.72
$Q_{90}$ (m <sup>3</sup> /s)	7.83	4.76	-1.46	-1.54	-1.55	-1.66
$Q_{95}$ (m <sup>3</sup> /s)	10.09	9.22	-2.42	-2.66	-2.74	-2.78
$Q_{99}$ (m <sup>3</sup> /s)	18.71	18.58	-5.51	-6.17	-7.17	-3.21
$Q_{99.5}$ (m <sup>3</sup> /s)	23.90	19.70	-6.72	-6.15	-8.58	-0.57
$Q_{T20}$ (m <sup>3</sup> /s)	48.69	54.61	-11.85	-14.26	-15.16	-10.40
$P_5$ (mm)	0.00	0.00	0.00	0.00	0.00	0.00
$P_{25}$ (mm)	0.00	0.08	0.00	0.00	0.02	0.00
$P_{50}$ (mm)	0.10	1.01	0.05	0.06	0.04	0.05
$P_{75}$ (mm)	2.70	1.83	-0.18	-0.18	-0.19	-0.18
$P_{90}$ (mm)	7.40	1.99	-0.26	-0.26	-0.26	-0.26
$P_{95}$ (mm)	11.42	2.38	-0.61	-0.61	-0.61	-0.61
$P_{99}$ (mm)	21.80	2.36	-1.86	-1.86	-1.86	-1.86
$P_{99.5}$ (mm)	29.09	1.56	-4.20	-4.20	-4.20	-4.20
$P_{P00}$	0.65	-0.10	-0.18	-0.25	-0.17	-0.17
$P_{P10}$	0.32	-0.15	0.16	0.08	0.16	0.16
$N_{dry}$	3470.00	-1466.00	0.00	-544.70	41.15	0.00
$P_{lag1}$	0.33	0.11	-0.13	-0.12	-0.12	-0.12



## References

- 550 Addor, N. and Seibert, J.: Bias correction for hydrological impact studies—beyond the daily perspective, *Hydrological Processes*, 28, 4823–4828, <https://doi.org/10.1002/hyp.10238>, 2014.
- Argüeso, D., Evans, J. P., and Fita, L.: Precipitation bias correction of very high resolution regional climate models, *Hydrology and Earth System Sciences*, 17, 4379, <https://doi.org/10.5194/hess-17-4379-2013>, 2013.
- Berg, P., Feldmann, H., and Panitz, H.-J.: Bias correction of high resolution regional climate model data, *Journal of Hydrology*, 448, 80–92, <https://doi.org/10.1016/j.jhydrol.2012.04.026>, 2012.
- 555 Cabus, P.: River flow prediction through rainfall–runoff modelling with a probability-distributed model (PDM) in Flanders, Belgium, *Agricultural Water Management*, 95, 859–868, <https://doi.org/10.1016/j.agwat.2008.02.013>, 2008.
- Cannon, A. J.: Multivariate bias correction of climate model output: Matching marginal distributions and intervariable dependence structure, *Journal of Climate*, 29, 7045–7064, <https://doi.org/10.1175/JCLI-D-15-0679.1>, 2016.
- 560 Cannon, A. J.: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables, *Climate Dynamics*, 50, 31–49, <https://doi.org/10.1007/s00382-017-3580-6>, 2018.
- Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes?, *Journal of Climate*, 28, 6938–6959, <https://doi.org/10.1175/JCLI-D-14-00754.1>, 2015.
- Christensen, J. H., Boberg, F., Christensen, O. B., and Lucas-Picher, P.: On the need for bias correction of regional climate change projections of temperature and precipitation, *Geophysical Research Letters*, 35, <https://doi.org/10.1029/2008GL035694>, 2008.
- 565 Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., and Wilby, R.: The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields, *Journal of Hydrometeorology*, 5, 243–262, [https://doi.org/10.1175/1525-7541\(2004\)005<0243:TSSAMF>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2), 2004.
- De Jongh, I. L. M., Verhoest, N. E. C., and De Troch, F. P.: Analysis of a 105-year time series of precipitation observed at Uccle, Belgium, *International Journal of Climatology*, 26, 2023–2039, <https://doi.org/10.1002/joc.1352>, 2006.
- 570 Dekens, L., Parey, S., Grandjacques, M., and Dacunha-Castelle, D.: Multivariate distribution correction of climate model outputs: A generalization of quantile mapping approaches, *Environmetrics*, 28, e2454, <https://doi.org/10.1002/env.2454>, 2017.
- Demarée, G. R.: The centennial recording raingauge of the Uccle Plateau: Its history, its data and its applications, *Houille Blanche*, 4, 95–102, 2003.
- 575 Di Luca, A., de Elía, R., and Laprise, R.: Challenges in the quest for added value of regional climate dynamical downscaling, *Current Climate Change Reports*, 1, 10–21, <https://doi.org/10.1007/s40641-015-0003-9>, 2015.
- Eberhart, R. and Kennedy, J.: A new optimizer using Particle Swarm Theory, in: *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, 1995. MHS'95., pp. 39–43, IEEE, 1995.
- Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions" Should we apply bias correction to global and regional climate model data?", *Hydrology and Earth System Sciences*, 16, 3391–3404, <https://doi.org/10.5194/hess-16-3391-2012>, 2012.
- 580 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geoscientific Model Development*, 9, 1937–1958, <https://doi.org/10.5194/gmd-9-1937-2016>, 2016.
- Fowler, H. J., Blenkinsop, S., and Tebaldi, C.: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, *International journal of climatology*, 27, 1547–1578, <https://doi.org/10.1002/joc.1556>, 2007.
- 585



- Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T.: Downscaling RCM precipitation to the station scale using statistical transformations—a comparison of methods, *Hydrology and Earth System Sciences*, 16, 3383–3390, <https://doi.org/10.5194/hess-16-3383-2012>, 2012.
- Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., Roessler, O., Wibig, J., Wilcke, R., Kotlarski, S., et al.: An intercomparison of a large ensemble of statistical downscaling methods over Europe: results from the VALUE perfect predictor cross-validation experiment, *International journal of climatology*, 39, 3750–3785, <https://doi.org/10.1002/joc.5462>, 2019.
- Gutjahr, O. and Heinemann, G.: Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM, *Theoretical and applied climatology*, 114, 511–529, <https://doi.org/10.1007/s00704-013-0834-z>, 2013.
- Gutowski, W. J., Decker, S. G., Donavon, R. A., Pan, Z., Arritt, R. W., and Takle, E. S.: Temporal–spatial scales of observed and simulated precipitation in central US climate, *Journal of Climate*, 16, 3841–3847, [https://doi.org/10.1175/1520-0442\(2003\)016<3841:TSEOAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<3841:TSEOAS>2.0.CO;2), 2003.
- Hagemann, S., Chen, C., Haerter, J. O., Heinke, J., Gerten, D., and Piani, C.: Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models, *Journal of Hydrometeorology*, 12, 556–578, <https://doi.org/10.1175/2011JHM1336.1>, 2011.
- Hay, L. E. and Clark, M. P.: Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States, *Journal of Hydrology*, 282, 56–75, [https://doi.org/10.1016/S0022-1694\(03\)00252-X](https://doi.org/10.1016/S0022-1694(03)00252-X), 2003.
- Helsen, S., van Lipzig, N. P. M., Demuzere, M., Vanden Broucke, S., Caluwaerts, S., De Cruz, L., De Troch, R., Hamdi, R., Termonia, P., Van Schaeybroeck, B., et al.: Consistent scale-dependency of future increases in hoClimate Dynamics, pp. 1–14, <https://doi.org/10.1007/s00382-019-05056-w>, 2019.
- Hewitson, B. C., Daron, J., Crane, R. G., Zermoglio, M. F., and Jack, C.: Interrogating empirical-statistical downscaling, *Climatic change*, 122, 539–554, <https://doi.org/10.1007/s10584-013-1021-z>, 2014.
- Ines, A. V. M. and Hansen, J. W.: Bias correction of daily GCM rainfall for crop simulation studies, *Agricultural and Forest Meteorology*, 138, 44–53, <https://doi.org/10.1016/j.agrformet.2006.03.009>, 2006.
- IPCC: Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change, Cambridge University Press, 2012.
- IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2013.
- Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, *Regional Environmental Change*, 14, 563–578, <https://doi.org/10.1007/s10113-013-0499-2>, 2014.
- Kendon, E. J., Ban, N., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., Evans, J. P., Fosser, G., and Wilkinson, J. M.: Do convection-permitting regional climate models improve projections of future precipitation change?, *Bulletin of the American Meteorological Society*, 98, 79–93, <https://doi.org/10.1175/BAMS-D-15-0004.1>, 2017.



- Klemeš, V.: Operational testing of hydrological simulation models, *Hydrological Sciences Journal*, 31, 13–24, <https://doi.org/10.1080/02626668609491024>, 1986.
- 625 Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., Van Meijgaard, E., et al.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, *Geoscientific Model Development*, 7, 1297–1333, <https://doi.org/10.5194/gmd-7-1297-2014>, 2014.
- Li, C., Sinha, E., Horton, D. E., Diffenbaugh, N. S., and Michalak, A. M.: Joint bias correction of temperature and precipitation in climate model simulations, *Journal of Geophysical Research: Atmospheres*, 119, 13–153, <https://doi.org/10.1002/2014JD022514>, 2014.
- 630 Li, H., Sheffield, J., and Wood, E. F.: 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, <https://doi.org/10.1029/2009JD012882>, 2010.
- Mao, G., Vogl, S., Laux, P., Wagner, S., and Kunstmann, H.: Stochastic bias correction of dynamically downscaled precipitation fields for Germany through Copula-based integration of gridded observation data, *Hydrology and Earth System Sciences*, 19, 1787–1806, <https://doi.org/10.5194/hess-19-1787-2015>, 2015.
- 635 Maraun, D.: Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums, *Geophysical Research Letters*, 39, <https://doi.org/https://doi.org/10.1029/2012GL051210>, 2012.
- Maraun, D.: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue, *Journal of Climate*, 26, 2137–2143, <https://doi.org/10.1175/JCLI-D-12-00821.1>, 2013.
- 640 Maraun, D.: Bias correcting climate change simulations—a critical review, *Current Climate Change Reports*, 2, 211–220, <https://doi.org/10.1007/s40641-016-0050-x>, 2016.
- Maraun, D. and Widmann, M.: Cross-validation of bias-corrected climate simulations is misleading, *Hydrology and Earth System Sciences*, 22, 4867–4873, <https://doi.org/10.5194/hess-22-4867-2018>, 2018.
- Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., <https://doi.org/10.1029/2009RG000314>, 2010.
- 645 et al.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Reviews of Geophysics*, 48, <https://doi.org/10.1029/2009RG000314>, 2010.
- Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R., and Wilcke, R. A. I.: VALUE: A framework to validate downscaling approaches for climate change studies, *Earth’s Future*, 3, 1–14, <https://doi.org/10.1002/2014EF000259>, 2015.
- 650 Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutierrez, J. M., Hagemann, S., Richter, I., Soares, P. M. M., Hall, A., et al.: Towards process-informed bias correction of climate change simulations, *Nature Climate Change*, 7, 764, <https://doi.org/10.1038/nclimate3418>, 2017.
- Mehrotra, R. and Sharma, A.: A multivariate quantile-matching bias correction approach with auto- and cross-dependence across multiple time scales: Implications for downscaling, *Journal of Climate*, 29, 3519–3539, <https://doi.org/10.1175/jcli-d-15-0356.1>, 2016.
- 655 Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J., and Cannon, A. J.: Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments, *Hydrology and Earth System Sciences*, 23, 1339–1354, <https://doi.org/10.5194/hess-23-1339-2019>, 2019.
- Mezzadri, F.: How to generate random matrices from the classical compact groups, *Notices of the American Mathematical Society*, 54, 592–604, 2007.





- 660 Michelangeli, P.-A., Vrac, M., and Loukos, H.: Probabilistic downscaling approaches: Application to wind cumulative distribution functions, *Geophysical Research Letters*, 36, L11 708, <https://doi.org/10.1029/2009GL038401>, 2009.
- Moore, R. J.: The PDM rainfall-runoff model, *Hydrology and Earth System Sciences*, 11, 483–499, <https://doi.org/10.5194/hess-11-483-2007>, 2007.
- Nguyen, H., Mehrotra, R., and Sharma, A.: Correcting for systematic biases in GCM simulations in the frequency domain, *Journal of Hydrology*, 538, 117–126, <https://doi.org/10.1016/j.jhydrol.2016.04.018>, 2016.
- 665 Nguyen, H., Mehrotra, R., and Sharma, A.: Correcting systematic biases across multiple atmospheric variables in the frequency domain, *Climate Dynamics*, pp. 1–16, <https://doi.org/10.1007/s00382-018-4191-6>, 2018.
- Olsson, J., Berggren, K., Olofsson, M., and Viklander, M.: Applying climate model precipitation scenarios for urban hydrological assessment: A case study in Kalmar City, Sweden, *Atmospheric Research*, 92, 364–375, <https://doi.org/10.1016/j.atmosres.2009.01.015>, 2009.
- 670 Panofsky, H. A., Brier, G. W., and Best, W. H.: Some application of statistics to meteorology, *Earth and Mineral Sciences Continuing Education*, College of Earth and Mineral Sciences, Pennsylvania State University, 1958.
- Penman, H. L.: Natural evaporation from open water, bare soil and grass, *Proc. R. Soc. Lond. A*, 193, 120–145, 1948.
- Pham, M. T.: Copulas-based stochastic modelling of evapotranspiration time series conditioned on rainfall as design tool in water resources management, Ph.D. thesis, Department of Forest and Water management, Ghent University, 2016.
- 675 Pham, M. T., Vernieuwe, H., De Baets, B., and Verhoest, N.: A coupled stochastic rainfall–evapotranspiration model for hydrological impact analysis, *Hydrology and Earth System Sciences*, 22, 1263–1283, [https://doi.org/10.5194/hess-22-1263\\_2018](https://doi.org/10.5194/hess-22-1263_2018), 2018.
- Pitié, F., Kokaram, A. C., and Dahyot, R.: N-dimensional probability density function transfer and its application to colour transfer, in: *ICCV 2005. Tenth IEEE International Conference on Computer Vision, 2005.*, vol. 2, pp. 1434–1439, IEEE, <https://doi.org/10.1109/ICCV.2005.166>, 2005.
- 680 Pitié, F., Kokaram, A. C., and Dahyot, R.: Automated colour grading using colour distribution transfer, *Computer Vision and Image Understanding*, 107, 123–137, <https://doi.org/10.1016/j.cviu.2006.11.011>, 2007.
- Popke, D., Stevens, B., and Voigt, A.: Climate and climate change in a radiative-convective equilibrium version of ECHAM6, *Journal of Advances in Modeling Earth Systems*, 5, 1–14, <https://doi.org/10.1029/2012MS000191>, 2013.
- Prechelt, L.: Automatic early stopping using cross validation: quantifying the criteria, *Neural Networks*, 11, 761–767, [https://doi.org/10.1016/S0893-6080\(98\)00010-0](https://doi.org/10.1016/S0893-6080(98)00010-0), 1998.
- 685 Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., Lipzig, N. P. M., and Leung, R.: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges, *Reviews of Geophysics*, 53, 323–361, <https://doi.org/10.1002/2014RG000475>, 2015.
- Rajczak, J., Kotlarski, S., and Schär, C.: Does quantile mapping of simulated precipitation correct for biases in transition probabilities and spell lengths?, *Journal of Climate*, 29, 1605–1615, <https://doi.org/10.1175/JCLI-D-15-0162.1>, 2016.
- 690 Rätty, O., Räisänen, J., and Ylhäisi, J. S.: Evaluation of delta change and bias correction methods for future daily precipitation: intermodel cross-validation using ENSEMBLES simulations, *Climate dynamics*, 42, 2287–2303, <https://doi.org/10.1007/s00382-014-2130-8>, 2014.
- Rätty, O., Räisänen, J., Bosshard, T., and Donnelly, C.: Intercomparison of univariate and joint bias correction methods in changing climate from a hydrological perspective, *Climate*, 6, 33, <https://doi.org/10.3390/cli6020033>, 2018.
- 695 Reiter, P., Gutjahr, O., Schefczyk, L., Heinemann, G., and Casper, M.: Does applying quantile mapping to subsamples improve the bias correction of daily precipitation?, *International Journal of Climatology*, 38, 1623–1633, <https://doi.org/10.1002/joc.5283>, 2018.



- Rizzo, M. L. and Székely, G. J.: Energy distance, *Wiley Interdisciplinary Reviews: Computational Statistics*, 8, 27–38, <https://doi.org/10.1002/wics.1375>, 2016.
- Rojas, R., Feyen, L., Dosio, A., and Bavera, D.: Improving pan-European hydrological simulation of extreme events through statistical bias correction of RCM-driven climate simulations., *Hydrology & Earth System Sciences*, 15, <https://doi.org/10.5194/hess-15-2599-2011>, 2011.
- Schmidli, J., Frei, C., and Vidale, P. L.: Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods, *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 26, 679–689, <https://doi.org/10.1002/joc.1287>, 2006.
- 705 Strandberg, G., Bärring, L., Hansson, U., Jansson, C., Jones, C., Kjellström, E., Kupiainen, M., Nikulin, G., Samuelsson, P., and Ullerstig, A.: CORDEX scenarios for Europe from the Rossby Centre regional climate model RCA4, SMHI, 2015.
- Switanek, M. B., Troch, P. A., Castro, C. L., Leuprecht, A., Chang, H. I., Mukherjee, R., and Demaria, E. M. C.: Scaled distribution mapping: A bias correction method that preserves raw climate model projected changes, *Hydrology and Earth System Sciences*, 21, 2649–2666, <https://doi.org/10.5194/hess-21-2649-2017>, 2017.
- 710 Székely, G. J. and Rizzo, M. L.: Testing for equal distributions in high dimension, *InterStat*, 5, 1249–1272, 2004.
- Székely, G. J. and Rizzo, M. L.: Energy statistics: A class of statistics based on distances, *Journal of statistical planning and inference*, 143, 1249–1272, <https://doi.org/10.1016/j.jspi.2013.03.018>, 2013.
- Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, *Journal of Hydrology*, 456, 12–29, <https://doi.org/10.1016/j.jhydrol.2012.05.052>, 2012.
- 715 Teutschbein, C. and Seibert, J.: Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions?, *Hydrology and Earth System Sciences*, 17, 5061–5077, <https://doi.org/10.5194/hess-17-5061-2013>, 2013, 2013.
- Themeßl, M. J., Gobiet, A., and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, *International Journal of Climatology*, 31, 1530–1544, <https://doi.org/10.1002/joc.2168>, 2011.
- Themeßl, M. J., Gobiet, A., and Heinrich, G.: Empirical-statistical downscaling and error correction of regional climate models and its impact 720 on the climate change signal, *Climatic Change*, 112, 449–468, <https://doi.org/10.1007/s10584-011-0224-4>, 2012.
- Van de Velde, J.: *Jornvdtv/OccMethods: Occurrence Methods: Calculations*, [https://doi.org/DOI: 10.5281/zenodo.3675305](https://doi.org/DOI:10.5281/zenodo.3675305), 2020.
- Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., et al.: The representative concentration pathways: an overview, *Climatic Change*, 109, 5, <https://doi.org/10.1007/s10584-011-0148-z>, 2011.
- Vandenbergh, S., Verhoest, N. E. C., Onof, C., and De Baets, B.: A comparative copula-based bivariate frequency analysis of observed and 725 simulated storm events: A case study on Bartlett-Lewis modeled rainfall, *Water Resources Research*, 47, W07 529, 2011.
- Verhoest, N., Troch, P. A., and De Troch, F. P.: On the applicability of Bartlett–Lewis rectangular pulses models in the modeling of design storms at a point, *Journal of Hydrology*, 202, 108–120, 1997.
- Verstraeten, G., Poesen, J., Demarée, G., and Salles, C.: Long-term (105 years) variability in rain erosivity as derived from 10-min rainfall depth data for Ukkel (Brussels, Belgium): Implications for assessing soil erosion rates, *Journal of Geophysical Research*, 111, D22 109, <https://doi.org/10.1029/2006jd007169>, 2006.
- 730 Vrac, M.: Multivariate bias adjustment of high-dimensional climate simulations: the Rank Resampling for Distributions and Dependences (R 2 D 2) bias correction, *Hydrology and Earth System Sciences*, 22, 3175, <https://doi.org/10.5194/hess-22-3175-2018>, 2018.
- Vrac, M. and Friederichs, P.: Multivariate—intervariable, spatial, and temporal—bias correction, *Journal of Climate*, 28, 218–237, <https://doi.org/10.1175/JCLI-D-14-00059.1>, 2015.



- 735 Vrac, M., Noël, T., and Vautard, R.: Bias correction of precipitation through Singularity Stochastic Removal: Because occurrences matter, *Journal of Geophysical Research: Atmospheres*, 121, 5237–5258, <https://doi.org/10.1002/2015JD024511>, 2016.
- Wang, L. and Chen, W.: Equiratio cumulative distribution function matching as an improvement to the equidistant approach in bias correction of precipitation, *Atmospheric Science Letters*, 15, 1–6, <https://doi.org/10.1002/asl2.454>, 2014.
- Wilcke, R. A. I., Mendlik, T., and Gobiet, A.: Multi-variable error correction of regional climate models, *Climatic Change*, 120, 871–887, <https://doi.org/10.1007/s10584-013-0845-x>, 2013.
- 740 Willems, P.: Revision of urban drainage design rules after assessment of climate change impacts on precipitation extremes at Uccle, Belgium, *Journal of Hydrology*, 496, 166–177, <https://doi.org/10.1016/j.jhydrol.2013.05.037>, 2013.
- Willems, P. and Vrac, M.: Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change, *Journal of Hydrology*, 402, 193–205, <https://doi.org/10.1016/j.jhydrol.2011.02.030>, 2011.
- 745 Zhang, X., Zwiers, F. W., and Hegerl, G.: The influences of data precision on the calculation of temperature percentile indices, *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 29, 321–327, <https://doi.org/10.1002/joc.1738>, 2009.
- Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., Trewin, B., and Zwiers, F. W.: Indices for monitoring changes in extremes based on daily temperature and precipitation data, *Wiley Interdisciplinary Reviews: Climate Change*, 2, 851–870, <https://doi.org/10.1002/wcc.147>, 2011.
- 750 Zscheischler, J., Fischer, E. M., and Lange, S.: The effect of univariate bias adjustment on multivariate hazard estimates, *Earth System Dynamics*, 10, 31–43, <https://doi.org/10.5194/esd-10-31-2019>, 2019.