# Bias-Adjustment Methods for Future Subdaily Precipitation Extremes Consistent Across Durations

Van de Vyver, Hans; Van Schaeybroeck, Bert; De Cruz, Lesley; Hamdi, Rafiq; Termonia, Piet

*Published in:*
earth and space science

*DOI:*
10.1029/2022EA002798

*Publication date:*
2023

AGU ADVANCING EARTH AND SPACE SCIENCE

Earth and Space Science

RESEARCH ARTICLE
10.1029/2022EA002798

# Bias-Adjustment Methods for Future Subdaily Precipitation Extremes Consistent Across Durations

**Hans Van de Vyver[1], Bert Van Schaeybroeck[1], Lesley De Cruz[1,2], Rafiq Hamdi[1], and Piet Termonia[1,3]**

[1]Department of Meteorological and Climatological Research, Royal Meteorological Institute, Uccle, Belgium, [2]Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussels, Belgium, [3]Department of Physics and Astronomy, Ghent University (UGent), Ghent, Belgium

**Key Points:**

- New bias-adjustment methods for subdaily extreme precipitation consistent across different precipitation durations
- Pseudo-reality cross-validation shows higher skill than existing bias-adjustment methods for subdaily extremes
- Application to regional climate model projections over Belgium shows, despite large uncertainties, consistent future increase in subdaily precipitation extremes

**Supporting Information:**

Supporting Information may be found in the online version of this article.

**Correspondence to:**

H. Van de Vyver,
hvijver@meteo.be

**Citation:**

Van de Vyver, H., Van Schaeybroeck, B., De Cruz, L., Hamdi, R., & Termonia, P. (2023). Bias-adjustment methods for future subdaily precipitation extremes consistent across durations. *Earth and Space Science*, *10*, e2022EA002798. https://doi.org/10.1029/2022EA002798

Received 14 DEC 2022
Accepted 3 FEB 2023

**Author Contributions:**

**Conceptualization:** Hans Van de Vyver
**Data curation:** Bert Van Schaeybroeck, Lesley De Cruz
**Formal analysis:** Hans Van de Vyver
**Funding acquisition:** Bert Van Schaeybroeck, Rafiq Hamdi, Piet Termonia
**Investigation:** Hans Van de Vyver, Bert Van Schaeybroeck



**Abstract** Model output from climate projections often requires bias-adjustment to compensate for systematic model errors. A bias-adjustment method for extreme precipitation intensity is proposed that preserves the scaling equation for different accumulation levels from hourly to daily, using intensity-duration-frequency (IDF) modeling. A validation is performed within a pseudo-reality setting, based on hourly precipitation from 28 regional climate model projections of the EURO-CORDEX ensemble over Belgium. The scaling-based adjustment methods improve upon previous methods, an optimal method is identified, and, analytical quantile mapping methods must be avoided due to three identified problems. The ensemble mean of the adjusted extreme precipitation intensity obeys the above-mentioned scale-invariance property, which is consistent with observed extreme intensities. We thus show that IDF modeling provides added value in the context of bias-adjustment, and, that the particular IDF model proposed balances well between accuracy and the preservation of desired properties such as scale invariance and consistency among rainfall durations.

**Plain Language Summary** Heavy precipitation is known to have major impacts on society, including crop damage, soil erosion, landslides, and increased flood risk. In recent years, there is a growing body of evidence that human-induced climate change has intensified extreme precipitation on the global scale. Both regional and global climate models predict that this will continue as the climate warms. Nevertheless, trend analysis or climate-change attribution of local rainfall extremes remains particularly difficult for many reasons including high variability in space and time. Therefore it is essential to use statistical methods that make use of all available information. Climate model simulations are known to exhibit systematic biases. Here, for the first time, we use a well-established technique in hydrology (intensity-duration-frequency modeling) to adjust extreme precipitation intensity as a function of the rainfall duration and event probability. Validation results with a large multi-model ensemble illustrate that the new method improves upon existing approaches. As an illustration, we present a bias-adjusted multi-model climate ensemble for future projections of heavy rainfall statistics over Belgium.

## 1. Introduction

Rainfall extremes are responsible for serious damage to various aspects of the human society and ecosystems. Short-duration extremes in particular affect people in urban areas due to flash floods (Fadhel et al., 2018; Hettiarachchi et al., 2018; Pall et al., 2011). The severe 2021 floods in Western Europe, for instance, resulted in 197 deaths in Germany and 39 in Belgium, and caused enormous damage to infrastructure (Kreienkamp et al., 2021). Through an intensification of the water cycle, climate change has been observed to intensify daily rainfall extremes (Fischer & Knutti, 2016) and model simulations project that this will continue (O'Gorman, 2015; Rajczak & Schär, 2017). Recent investigations have pointed out that also subdaily rainfall extremes will increase in a warmer climate (Fowler et al., 2021; Guerreiro et al., 2018; Westra et al., 2014). Still, robust analysis of climate trends or attribution of rainfall extremes remains a major challenge due to the intermittency of rainfall and the lack of long time series. Robust signals can be obtained using large climate-model ensembles or spatial averaging in the statistical analysis (Kreienkamp et al., 2021; Philip et al., 2020), but these methods generally do not exploit relationships between rainfall extremes and duration.

High resolution regional climate models (RCMs) are commonly assumed to better simulate rainfall extremes than global models as convective processes are better represented by either parameterization schemes (10–50 km),

**Methodology:** Hans Van de Vyver
**Project Administration:** Bert Van Schaeybroeck, Rafiq Hamdi, Piet Termonia
**Software:** Hans Van de Vyver
**Supervision:** Rafiq Hamdi, Piet Termonia
**Validation:** Hans Van de Vyver
**Visualization:** Hans Van de Vyver, Bert Van Schaeybroeck
**Writing – original draft:** Hans Van de Vyver, Bert Van Schaeybroeck, Lesley De Cruz
**Writing – review & editing:** Hans Van de Vyver, Bert Van Schaeybroeck, Lesley De Cruz

or by explicitly resolving at km-scale (Lucas-Picher et al., 2021). Large ensembles of RCM projections are for instance generated within the Coordinated Regional Climate Downscaling Experiment (CORDEX) which includes runs at a resolution of 0.11° (12.5 km) over Europe (Berg et al., 2019; Hosseinzadehtalaei et al., 2021; Schmith et al., 2021).

A large part of the model errors is systematic in nature and can be adjusted (Maraun & Widmann, 2018; Vrac & Friederichs, 2015). Bias-correction methodologies usually map the statistical properties (such as means, quantiles) or the full distribution from the model onto the observations. The delta-change (DC) approach, on the other hand, establishes a mapping from historical to future simulations and thereby inherently avoids the problem that the bias to be adjusted could be non-stationary in time (Cannon et al., 2015; Maraun, 2016).

Recently Schmith et al. (2021) identified bias-adjustment methods for extreme precipitation and applied them to high return levels of hourly and 24-hourly rainfall from EURO-CORDEX ensemble projections at 0.11° resolution. In addition, they proposed an alternative approach based on quantile mapping of extreme value distributions. Taking this work as a starting point, we extend and improve in a number of aspects. While they considered hourly and 24-hourly extremes, we incorporate information for various predefined precipitation durations, using the so-called "scale invariance" property (Burlando & Rosso, 1996; Gupta & Waymire, 1990; Innocenti et al., 2017, 2019). Using a pseudo-reality approach, we show that this leads to improved adjustments. Additionally we identify technical complications with analytical quantile-based methods, associated to bounded probability distributions.

Intensity-duration-frequency (IDF) curves mainly serve to provide extreme precipitation intensities for various durations and return periods, and are commonly used in water resources and hydraulic engineering (Koutsoyiannis et al., 1998). Assuming scale invariance, IDF curves can be determined using only a few parameters (Blanchet et al., 2016; Burlando & Rosso, 1996; Van de Vyver, 2015, 2018). The difference between IDF-based and conventional (or reference) methodologies for calculating return levels, is schematically represented in Figure 1. While return levels for different rainfall durations (e.g., 1, 2 hr, …) are calculated independently by conventional methodologies, IDF-based methods combine the information of all durations into a parsimonious statistical model, by the addition of an extra scaling parameter ($\eta$). The methodological benefit of such procedure is: (a) consistency, that is, no crossing return levels across durations, (b) the improved fitting, (c) the potential extension to rainfall durations within the range of durations used for fitting, and (d) the correction of systematic errors in the extremes due to the available temporal resolution of the RCM data (e.g., hourly).

Our main assumption in the construction of bias-adjusted IDF-values is that they must be scale invariant by nature, in accordance with the observations. We identify only one such method that meets this requirement, namely a multiplicative correction under the so-called "simple-scaling" hypothesis. Apart from these methodological advances, we estimate ensemble future extreme precipitation quantiles by adjusting station-based subdaily and daily precipitation extremes, using a multi-model ensemble of 28 EURO-CORDEX RCMs. We investigate to what extent the ensemble mean satisfies the scale-invariant IDF-relationship.

The paper is organized as follows. Section 2 gives a brief overview of the data used (observations and simulations). Section 3 describes the extreme value methodology and the resulting bias-adjustment methods. In Section 4 we point out some potential problems that can arise with analytical quantile mapping methods. Validation and application is performed in Section 5. Finally, in Section 6, conclusions are drawn.

## 2. Data

### 2.1. Observations

We consider 18 Belgian pluviograph series with 10-min precipitation (1967–2004), obtained from the hydro-meteorological network of the Royal Meteorological Institute of Belgium (Figure S1 in Supporting Information S1). Methods for handling missing values are described in Van de Vyver (2018) and Van de Vyver et al. (2021). We extracted annual maxima from $d$-hourly accumulations using a sliding window over 10-min precipitation amounts, with $d = 1, 2, 6, 12, 24, 48$, and 72 hr, chosen in exponentially increasing order since IDF characteristics are usually described in logarithmic scale.

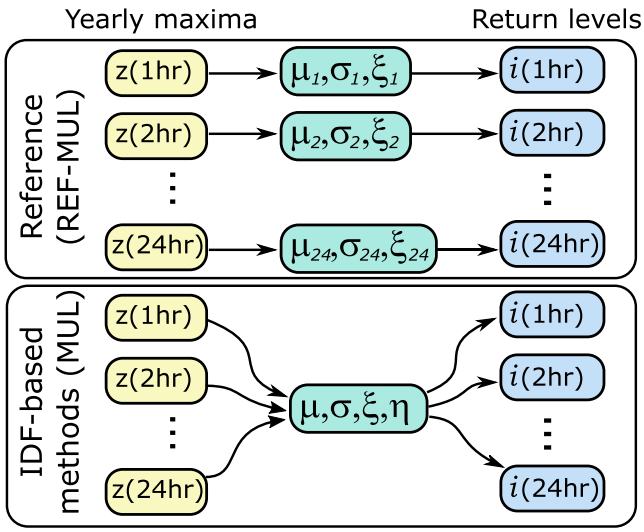


**Figure 1.** Schematic representation of the main methodological differences between the existing reference and the newly proposed return levels bias-adjustment techniques.

### 2.2. EURO-CORDEX Data

As model data the EURO-CORDEX climate projections (Jacob et al., 2014, 2020) at spatial resolution of 0.11° (12.5 km) are taken. More specifically, hourly precipitation from 28 different RCM projections (see Table S1 in Supporting Information S1) for the 18 station locations over Belgium are selected, using the nearest-grid point approach. For the bias-adjustment experiments, we selected two 30-year time windows: 1971–2000 from the historical run, and 2071–2100 from the high-emission RCP8.5 run. Similar to the observations, $d$-hourly accumulation series were calculated. However, the annual maxima were extracted using a sliding window of 1 hr as opposed to 10-min used for the observations.

## 3. Methods

### 3.1. Extreme Value Analysis and IDF Relationships

For a sequence of independent and identically distributed random variables $X_1, \ldots, X_n$, the block maximum $Z = \max\{X_1, \ldots, X_n\}$ is commonly modeled with the *generalized extreme value* (*GEV*) distribution (Beirlant et al., 2004; Coles, 2001; Leadbetter et al., 1983):

$$G(z;\mu,\sigma,\xi) = \exp\left[-\left(1+\xi\,\frac{z-\mu}{\sigma}\right)_+^{-1/\xi}\right], \qquad \text{with } v_+ = \max\{0,v\}, \quad (1)$$

where $\mu$ is the location parameter, $\sigma > 0$ is the scale parameter, and $\xi$ is the shape parameter. The GEV distribution has been commonly applied for modeling heavy rainfall, floods, high wind gusts, and many other maxima in environmental processes (Davison & Gholamrezaee, 2011; Katz et al., 2002; Naveau et al., 2005). The parameters of the GEV distribution are easily estimated with probability-weighted methods or maximum likelihood estimation (MLE) (Hosking, 1990; Martins & Stedinger, 2000). Once a GEV distribution is fitted to a block maximum series, return levels or return periods can be easily calculated.

The annual maximum precipitation intensity $I(d)$, defined as the annual maximum $d$-hourly rainfall amount divided by the duration $d$, follows the GEV distribution: $I(d) \sim \text{GEV}[\mu(d), \sigma(d), \xi]$. The shape parameter $\xi$ is kept constant over $d$, as suggested in Nadarajah et al. (1998). The specific form of $\mu(d)$ and $\sigma(d)$ can be determined from the scaling invariance of maximum intensity. "*Simple-scaling*" implies that the non-central moments of the maximum intensity follow a power-law as a function $d$, that is, $E[I^l(d)] \propto d^{-l\eta}$ (Burlando & Rosso, 1996; Gupta & Waymire, 1990), thereby implying (Menabde et al., 1999; Nguyen et al., 1998):

$$\mu(d) = \mu\, d^{-\eta}, \qquad \text{and} \qquad \sigma(d) = \sigma\, d^{-\eta} \qquad (0 < \eta < 1), \qquad \text{for duration } 1 \text{ hr} \le d \le 72 \text{ hr}, \quad (2)$$

giving rise to a four-parameter family of $d$-dependent simple-scaling GEV distributions, with parameter vector ($\mu$, $\sigma$, $\xi$, $\eta$). The MLE, $(\hat{\mu}, \hat{\sigma}, \hat{\xi}, \hat{\eta})$, is obtained by maximizing the associated likelihood-function of the GEV-model (see Section S1B in Supporting Information S1 for details).

Intensity-duration-frequency (IDF) curves are graphical representations of the return levels of maximum intensity, $i_T(d)$, as a function of $d$, for a range of return periods $T$. We refer to $i_T(d)$ as "IDF values." The associated IDF-relationship of the simple-scaling GEV-model, Equation 2, is readily obtained as

$$i_T(d) = \left(\mu - \frac{\sigma}{\xi}\left\{1 - \left[-\log\left(1-\frac{1}{T}\right)\right]^{-\xi}\right\}\right) d^{-\eta}, \qquad \text{with} \qquad 0 < \eta < 1. \quad (3)$$

Although the simple-scaling assumption is generally accepted as a sufficiently good working hypothesis (Blanchet et al., 2016; Bougadis & Adamowski, 2006; Innocenti et al., 2017, 2019; Mascaro, 2020; Mélèse et al., 2018; Menabde et al., 1999; Nguyen et al., 1998; Panthou et al., 2014), there is still room for improvement. Gupta and Waymire (1990) and Burlando and Rosso (1996) showed that the non-central moments of extreme intensity can scale with different exponents, a property that is referred to as "*multiscaling*." Van de Vyver (2018) illustrated that the multiscaling assumption implies that the associated GEV-parameters scale with different exponents:

$$\mu(d) = \mu\, d^{-\eta_1}, \qquad \text{and} \qquad \sigma(d) = \sigma\, d^{-\eta_2} \qquad (0 < \eta_1 \le \eta_2 < 1), \text{ for duration } 1 \text{ hr} \le d \le 72 \text{ hr}. \quad (4)$$

The resulting multiscaling IDF-relationship is given by:

$$i_T(d) = \mu\, d^{-\eta_1} - \frac{\sigma\, d^{-\eta_2}}{\xi}\left\{1 - \left[-\log\left(1 - \frac{1}{T}\right)\right]^{-\xi}\right\} \qquad \text{with} \qquad 0 < \eta_1 \le \eta_2 < 1, \tag{5}$$

which degenerates to the simple-scaling IDF-model, Equation 3, if $\eta_1 = \eta_2$. In Van de Vyver (2018) the difference between the scaling exponents ($\eta_{1,2}$) was found to be statistically significant for in situ sub-daily rain gauge data of Belgium. Overeem et al. (2008) and Courty et al. (2019) also noted (without significance testing) that the GEV location- and scale-parameter scale with different exponents, for observations in the Netherlands and the UK, respectively. On the global scale, Courty et al. (2019) also found a different scaling of the two GEV-parameters for the ERA5 reanalysis.

With regard to RCM output, Van de Vyver et al. (2021) showed that the multiscaling property can be found at convection-permitting scale (≤4 km) and coarser scale (12 km) over Belgium, in agreement with the empirical observations. In a different climate, Cannon and Innocenti (2019) applied the simple-scaling hypothesis to convection-permitting simulations (4 km) over North-America to obtain future IDF curves, but the possibility of further exploring the multiscaling approach was not considered.

As an example, we show in Figure S2 in Supporting Information S1 a double logarithmic plot of the GEV location- and scale-parameter against duration, for past and future simulations of a particular EURO-CORDEX member. The log-log linearity of both GEV-parameters is clearly seen within the range 3–72 hr, and each scales with a different slope. Within the range 1–2 hr, however, we find an underestimation of the scaling law. On the contrary, maxima based on continuous observations (i.e., "sliding" maxima) show no significant deviation of the scaling law (Overeem et al., 2008; Van de Vyver, 2018). After all, the multiscaling theory assumes that rainfall accumulation is a continuous process in time (Burlando & Rosso, 1996). Additional factors that may explain this underestimation are the spatial mismatch between gridded model output and point observations, and the systematic model errors due to the parameterization of convective processes. Consequently, the scaling IDF models offer the extra advantage of being able to correct this underestimation, in particular for hourly extremes.

Finally, for all the EURO-CORDEX simulations (56 in total) and locations (18 in total), we performed a goodness-of-fit of the scaling law with the coefficient of determination $r^2$. Within the range of durations 3–72 hr, in 95% of the series $r^2$ exceeds 0.99 for the location and 0.80 for the scale. Considering a range with a higher minimum value, 6–72 hr, we found a better fit for the scale parameter, where $r^2$ exceeds 0.89 in 95% of the series, thus we may consider the scaling law as a robust property of the RCMs. We then used AIC to make a selection between the simple-scaling and multiscaling model, and we found that 50% of the series show a significant multiscaling pattern.

### 3.2. Bias-Adjustment Methods for Return Levels

Two types of bias-adjustment methods are proposed: the first type is the most common and directly adjusts extreme quantiles estimated from observations (Section 3.2.1), while the second adjusts the full distribution of maximum intensity (Sections 3.2.2 and 3.2.3). The latter was studied by Schmith et al. (2021) for 1- and 24-hr duration extremes, using peaks-over-threshold (POT) instead of annual maxima. This, however, does not affect the generality of our results since there is a strong theoretical connection between both extreme value models (Coles, 2001).

One of the main arguments of our framework is that bias-adjusted IDF-values should fall within a family of scaling IDF models (preferably exact). However, it is not obvious to comply with this, as we will see for analytical quantile mapping (Section 4), but also the use of the multiscaling IDF model gives rise to inconsistent adjustments (Section 3.2.1).

We introduce the following adjustment methods:

- MUL: multiplicative adjustment of the simple-scaling IDF-model (see Section 3.2.1).
- REF-MUL: same as MUL, but we use individually-estimated extreme precipitation quantiles for each rainfall duration, without introducing an IDF-model (Schmith et al., 2021; Tabari et al., 2016). See Section S1A in Supporting Information S1 for details on MLE.
- DC: analytical quantile mapping delta-change, using the simple-scaling GEV distribution (see Section 3.2.2).
- BC: analytical quantile mapping bias-correction, using the simple-scaling GEV distribution (see Section 3.2.3).

As reference method, we consider the raw future run (RAW) to which we fit the simple-scaling IDF-model, Equation 3.

#### 3.2.1. Return Level Adjustment Using a Multiplicative Correction (MUL)

First, denote by $i_T^o(d)$, $i_T^c(d)$ and $i_T^s(d)$, the $T$-year return levels based on Equation 3 using the station observations, control run, and future scenario run, respectively. Our first adjustment method is called MUL, since it scales the return levels by a multiplicative term:

$$i_T^{mul}(d) = i_T^o(d).\, i_T^s(d)/i_T^c(d). \tag{6}$$

The great advantage of this adjustment is that the simple-scaling property is preserved, as one can easily obtain that:

$$i_T^{mul}(d) \propto d^{-(\eta^o-\eta^c+\eta^s)}. \tag{7}$$

The scaling exponent $\eta$, however, must always satisfy $0 < \eta < 1$ which constrains $\eta^o - \eta^c + \eta^s$, but experiments show that this is always met in practice.

However, when considering the multiscaling IDF-model, Equation 5, the adjusted values $i_T^{mul}(d)$ cannot be traced back to multiscaling behavior. This provides a valid argument to use the (sufficiently good) simple-scaling model in the adjustment, even though the multiscaling model could perform slightly better on its own.

#### 3.2.2. Delta-Change (DC) Approach

The adjusted future return levels $i_T^{dc}(d)$ are obtained by mapping the return levels estimated from observations through the following transformation, which accounts for the relative trend in climate models:

$$i_T^{dc}(d) = G_s^{-1}\left(G_c\left(i_T^o(d)\right)\right) \text{ for all } T > 1 \text{ year}, \tag{8}$$

and $G$ is the GEV function defined in Equations 1 and 2. After calculation, we obtain the exact solution:

$$i_T^{dc}(d) = \left(\mu^s + \frac{\sigma^s}{\xi^s}\left[Y_+^{\xi^s/\xi^c} - 1\right]\right) d^{-\eta^s}, \tag{9}$$

where $Y = 1 + \xi^c\left[i_T^o(d)\, d^{\eta^c} - \mu^c\right]/\sigma^c$.

#### 3.2.3. Bias-Correction (BC) Approach

Assuming a stationary bias, bias-correction is based on the mapping of the raw future return levels through the transformation, analogously to DC:

$$i_T^{bc}(d) = G_o^{-1}\left(G_c\left(i_T^s(d)\right)\right). \tag{10}$$

An exact solution, similar to Equation 9, can be easily derived:

$$i_T^{bc}(d) = \left(\mu^o + \frac{\sigma^o}{\xi^o}\left[Y_+^{\xi^o/\xi^c} - 1\right]\right) d^{-\eta^o}, \tag{11}$$

where $Y = 1 + \xi^c\left[i_T^s(d)\, d^{\eta^c} - \mu^c\right]/\sigma^c$.

### 3.3. Validation Approach: Pseudo-Reality and Scores

As in Maraun (2012), Van Schaeybroeck and Vannitsem (2016), and Schmith et al. (2021), we perform inter-model cross-validation using a *pseudo-reality* setting. Thereby a particular ensemble member is considered as the (pseudo-)observation while the other members as the models that need adjustment. The benefit of this approach is the availability of future "observations," and scores are calculated by iteratively considering each member as the pseudo-observation. See Section S2 in Supporting Information S1 for a full motivation of this approach. The validation is robust because it generates a large amount of data (28 EURO-CORDEX members and 18 locations result in $28 \times 27 \times 18$ bias-adjusted IDF-values, for each duration and return period).

As pseudo-observations, we consider the IDF-values obtained by fitting the multiscaling IDF model, Equation 5, to the pseudo-observed maxima. The idea of using IDF models as a reference was proposed in Berg et al. (2019) where EURO-CORDEX simulations were evaluated against gauge-based depth-duration-frequency (DDF) curves. This may lead to a more robust validation and, moreover, the multiscaling theory corrects for the clock-hourly accumulated maxima. Note that the reference and validation data use different IDF models (multiscaling and simple-scaling, respectively), which needs not be a contradiction. On the one hand, the choice for the multiscaling approach as reference is justified since it provides a more general framework than simple-scaling. On the other hand, the adjusted return levels are simple-scaling which balances accuracy and consistency (see Section 3.2.1).

Next, we validate the bias-adjustment methods by comparing the future bias-adjusted return levels with the future reference in the pseudo-reality approach. We analyze the distribution of the relative error (RE), Equation A1, and its absolute value using the median and the 5th and 95th percentiles. A measure of the total accuracy is the relative bias (R-BIAS), obtained by averaging the RE over all member combinations and over all station locations. Additionally, we use the mean squared error (MSE) skill score (MSESS), which expresses the reduction of the MSE compared to the unadjusted return levels (RAW). For a detailed description of the scores, we refer to Appendix A.

## 4. Complications of GEV Quantile Mapping

Prior to validating the new bias-adjustment methods, three potential problems that may be caused by GEV quantile mapping approaches (DC and BC) are discussed, although these issues also arise when using individual-duration distributions (e.g., Schmith et al. (2021)).

- *Problem 1*. A technical complication occurs when an observed return level, $i_T^o(d)$, lies outside the support of the GEV distribution of the control run (i.e., $Y < 0$ in Equation 9).
- *Problem 2*. The adjusted distribution is not a GEV distribution, however, if we consider the transformed values as being the annual maxima from the future climate, we would expect them to follow a GEV distribution.
- *Problem 3*. Similar to Problem 2, there is the expectation of scale invariance of the bias-adjusted distribution. It can readily be seen that the DC method does not meet this requirement since $Y$ in Equation 9 is generally not independent of $d$.

We explore these issues by applying DC adjustment to all the RCM simulations at the nearest grid point of each station. In order to address Problems 2 and 3, the IDF-model is fitted to the DC-adjusted return levels, by minimizing the total squared sum of the difference between the adjusted return levels and Equation 3. Visual tests showed that the adjusted return levels deviate from log-log linearity when the residual error of the least square fit (ResE) exceeds the, subjectively picked, value of 0.1. As such, the IDF curves are straight parallel lines in the double-logarithmic graph, where the slope of the lines corresponds to minus the scaling exponent, $\eta$.

In 94% of the cases investigated, no issues were found and the IDF-relationship fits well (example in Figure 2a). Problem 1 appeared in 3% of the cases, making DC adjustment impossible (Figure 2c). In another 3% of the cases, the goodness-of-fit criterion ResE $< 0.1$ was violated amounting to the appearance of both Problems 2 and 3 (Figure 2b). The considerable loss of scale invariance could be traced back to the combined effect of small positive $Y$-values and $\hat{\xi}^s/\hat{\xi}^c < 0$ in Equation 9. In addition, the small $Y$-values are also responsible for the unrealistically high return levels.

## 5. Results

### 5.1. Validation

Here the accuracy of the adjustment methods of Section 3.2 is examined within the pseudo-reality framework as described in Section 3.3. Thereby the end-of-the-century rainfall extremes from the EURO-CORDEX ensemble are bias-adjusted at different locations over Belgium.

The three problems as described in Section 4, appear in 6.8% and 7.9% of the cases for DC and BC, respectively. In order to maintain a fair comparison, all the concerned data points have been removed from the evaluation. Figure 3 shows the relative bias (R-BIAS, top row) and the mean squared error skill score (MSESS, bottom row)

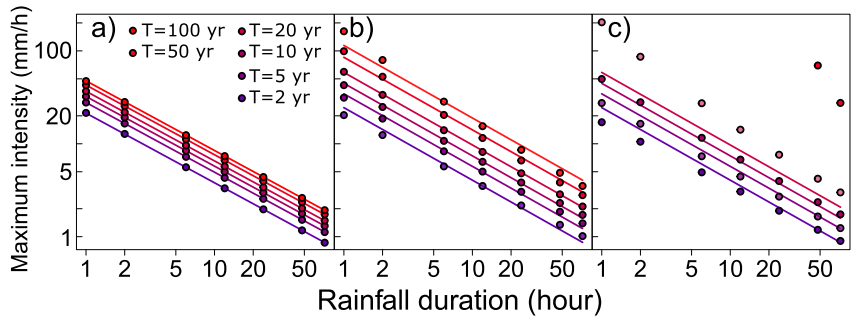


**Figure 2.** Examples showing potential problems encountered for the adjusted return levels with the delta-change approach using pairs of station observations and regional climate model data. Dots: adjusted return levels $\hat{i}_T^{dc}(d)$, with $T = 2$, 5, 10, 20, 50, and 100 years. The lines indicate the intensity-duration-frequency relationship, Equation 3, fitted to the dots with least squares. The three situations are: (a) Successful adjustment. (b) Adjusted return levels deviate from log-log linearity. (c) Same as (b), but there is a more pronounced deviation here. In addition, there are infinitely high (undefined) return levels for $T > 20$ years.

of the future return levels for different subdaily durations and return periods of 2, 5, and 10 years (columns). The scores have not been supplemented with confidence intervals because they are so narrow (due to the large amount of data) that they are not informative.

While the unadjusted RAW model overestimates the return levels, bias-adjustment leads to an underestimation for durations of 1 and 2 hr (Figures 3a–3c). Also, the relative biases are reduced in magnitude upon increase of the rainfall duration for all methods. On the other hand, the skill score increases and decreases for return periods $T = 2$ and $T = 10$ years, respectively (Figures 3d–3f). Three important results arise from the validation:

1. MUL is consistently the best method for all rainfall durations and all return periods, with a positive skill score between 20% and 50% (Figures 3d–3f).
2. For $T = 2$ years, the skill of DC and BC is comparable to MUL (Figure 3d). However, in almost all other cases it is significantly lower (Figures 3e and 3f), despite their comparable bias (Figures 3a–3c).
3. REF-MUL is strongly biased for 1- and 2-hr durations. For 10 years-return levels and for hourly precipitation (all return levels), this non-scaling-based adjustment even loses skill compared to RAW.

For larger return periods ($T = 20$–100 years), MUL remains the most skillful bias-adjustment method (see Figure S4 in Supporting Information S1) but the added value of bias-adjustment is reduced upon increase of the rainfall duration and the return period. More specifically, while skillful for $T = 20$ years for all durations except

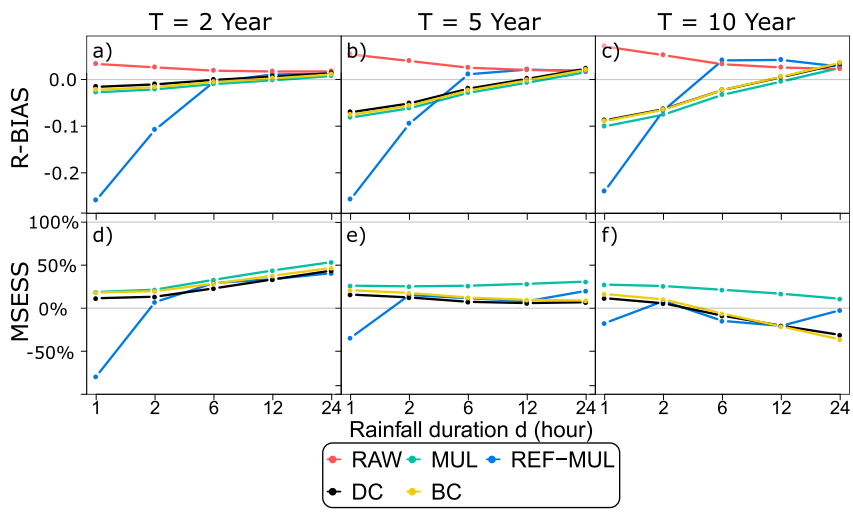


**Figure 3.** Validation of the bias-adjustment methods for extreme intensity, as a function of rainfall duration $d$ and for different return periods $T$. First row: the relative bias of the adjustment methods (negative bias indicates underestimation, and vice versa). Second row: the mean squared error skill score.

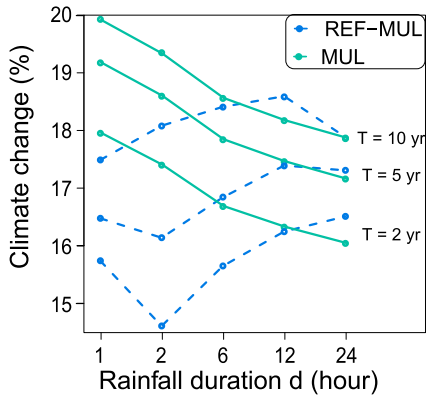


**Figure 4.** Relative climate changes of the IDF-values (i.e., "future scenario" vs. "control run": $i_T^s(d)/i_T^c(d)$) at the end of the century following the RCP8.5 scenario. The results are averaged over the 28 EURO-CORDEX members and 18 station locations.

for $d$ = 24 hr, MUL is not skillful for $T$ = 50 years ($d$ = 12–24 hr) and $T$ = 100 years ($d$ = 6–24 hr). In other words, for these cases, it is better not to adjust the model extremes (i.e., use RAW). The lack of predictive skill for return periods exceeding 20 years is understandable given the considered projection time period of 30 years.

Figures S5–S9 in Supporting Information S1 reveal that the distribution of the RE is highly positively skewed. Especially for DC and BC it can be clearly seen that, despite the small biases, there appear large outlying errors, resulting in a strong reduction of MSESS. As discussed in Section 4, these arise when small positive $Y$-values and $\hat{\xi}^s/\hat{\xi}^c < 0$ occur simultaneously in Equation 9.

Figures S10–S14 in Supporting Information S1 indicate that the 5th–95th percentile intervals of the RE of MUL, BC, and DC are similar, but the interval length of MUL is slightly shorter. For the specific return periods and durations for which MUL is not skillful (see above), RAW's interval length is similar or even shorter. In addition, if skillful, MUL has the lowest 95th percentile of the RE, in absolute value (|RE|).

The large REF-MUL biases for 1- and 2-hr durations are due to the fact that this method takes into account clock-hourly aggregated maxima, which severely underestimate the continuous "sliding" maxima. The scaling methods, on the other hand, account for the log-log linearity of the GEV-parameters with rainfall duration, which allows to correct the error caused by the hourly temporal resolution (see Section S1B in Supporting Information S1), but also ensures the return levels to be smoothed over $d$. This smoothing reduces the error variability compared to REF-MUL, and improves the overall skill.

Finally, since a multi-model ensemble mean generally provides a better estimate than each constituent ensemble member (Gleckler et al., 2008; Reichler & Kim, 2008; Tebaldi & Knutti, 2007), the validation effort was repeated in an ensemble framework by comparing the pseudo future IDF-values with the mean of the member-by-member adjusted IDF-values. Again, MUL came up as the best model, and, the ensemble mean was better than all individual models for nearly all locations, rainfall durations and return periods (not shown).

### 5.2. Application: Bias-Adjustment of Simulated IDF-Values in an Ensemble Framework

After selecting the most appropriate bias-adjustment method (MUL), we will apply it here to calculate future projections of IDF-values in the EURO-CORDEX ensemble. Prior to this, we need to assess the model agreement on the sign of the climate change. Following (Tebaldi et al., 2011) to distinguish the lack of climate change signal for precipitation from possible model disagreement, we perform the two-sided $t$-test to check the null hypothesis that there is no difference between the mean of the historical and future extremes, at the 5% significance level. The percentage of simulations (out of the 28 RCMs and 18 locations) with a significant change is 55% ($d$ = 1 hr), 53% ($d$ = 2 hr), 56% ($d$ = 6 hr), 59% ($d$ = 12 hr), and 63% ($d$ = 24 hr). Figure S15 in Supporting Information S1 shows no clear spatial structures in the acceptance rates for all rainfall durations. Next, we established that almost every simulation with significant change features an increasing trend.

To illustrate the climate-change signal for precipitation extremes, we show in Figure 4 the spatially-averaged ensemble mean of the relative climate change of the IDF-values. It is clear that the scaling-based IDF-method provides smooth climate-change signals across the durations, while the non-scaling-based method (REF-MUL) fails to do so. The MUL-based changes are consistently larger for shorter durations.

End-of-century ensemble predictions of 10 years-return levels of hourly precipitation are shown in Figure 5, in the form of a histogram and kernel density of the member-by-member MUL-adjustments. The ensemble spread is quite high which, globally, across all stations, varies between 15 and 55 mm. As previously shown, the climate-change signal is positive and robust (around 16%–20%), and does not vary much across the station locations.

Finally, in Figure S16 in Supporting Information S1, we display for each pair ($d$, $T$) the ensemble mean of the MUL-adjusted IDF-values in a log-log plot, and we added the simple-scaling IDF curves, $i_T(d) \propto d^{-\eta}$, fitted

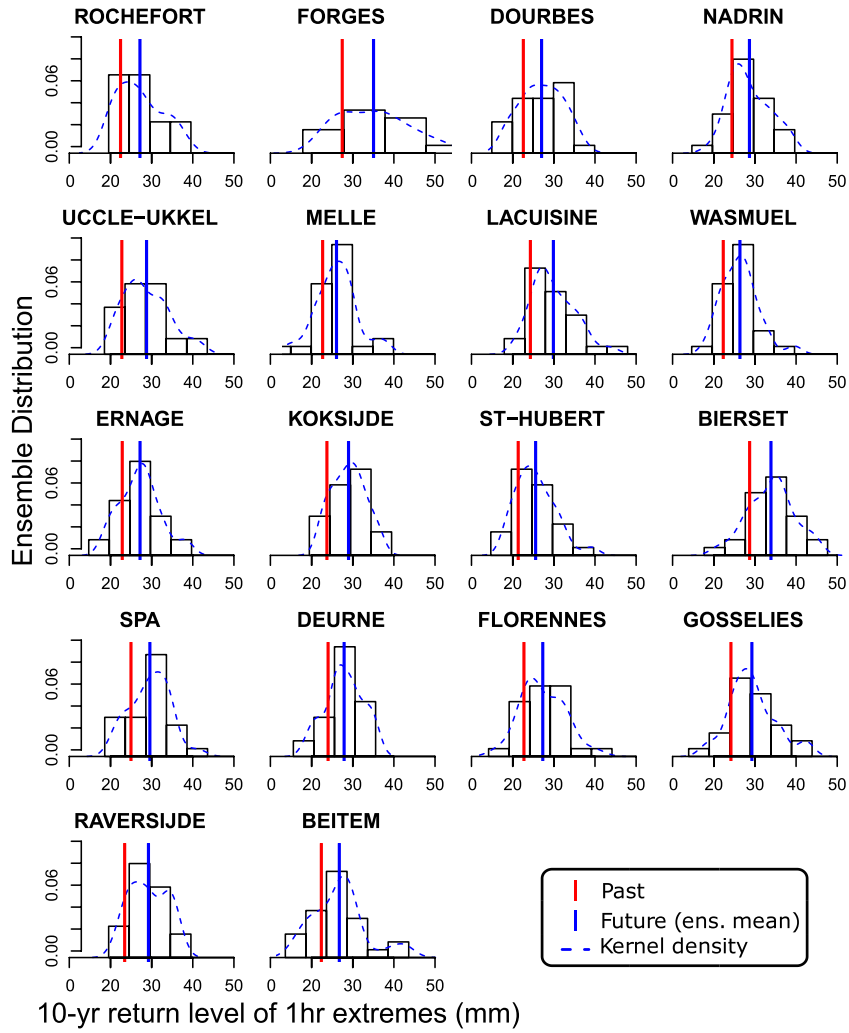


**Figure 5.** Histogram and kernel density plot of the EURO-CORDEX ensemble of bias-adjusted 10 years-return levels of future hourly precipitation (method: MUL). Added are the return levels estimated with observations ("Past," 1967–2004), and the ensemble mean ("Future," 2071–2100).

to these values. Interestingly, the ensemble averaged IDF-values satisfy to a very good approximation the IDF-relationship. The slope $\eta$ (in absolute value) of the IDF curves ranges from approximately 0.64 to 0.75, and corresponds well with the observed one.

## 6. Conclusions and Outlook

New bias-adjustment methods are proposed and validated that estimate future return levels that are consistent over a range of rainfall durations by accounting for scaling invariance of extreme rainfall intensity. This includes two main adjustment approaches: (a) considering the relative climate change for the return levels (MUL), based on the scale-invariant IDF-model and (b) quantile mapping, based on the scale-invariant GEV distribution (DC and BC). MUL is found to always improve upon the non-scale invariant methods for all durations and return periods. With respect to the unadjusted model, MUL features skill scores ranging from 20% to 50%. The use of GEV quantile mapping based adjustments (DC and BC) is not recommended due to the high potential for technical issues.

The best method (MUL) builds on previously developed approaches (Schmith et al., 2021; Tabari et al., 2016), but significantly reduces the variability in the high-return level estimates for two reasons. First, without the temporal-scaling assumption, the variability in the observed return level estimates is higher and can be physically

inconsistent across durations, and second, this effect is strengthened because the relative climate change (i.e., a ratio of IDF-values) is then also subject to larger estimation variability. Additionally, in the case that model data is available at hourly time resolution, our scaling-based method allows to transform the return levels of the clock-hourly aggregated maximum to those of the (correct) sliding maximum.

MUL preserves the temporal-scale invariance property of extreme intensity as modeled with a simple-scaling model, cfr. Equation 7. More accurate multiscaling IDF-relationships (Van de Vyver, 2018) have been tested for the adjustment methods, but the multiscaling properties are found not to be preserved by MUL. Therefore the combination of MUL with the simple-scaling IDF model constitutes a balance between a sufficiently accurate IDF-model, and, the preservation of desired properties (i.e., scale invariance and consistency).

MUL can be easily integrated in a multi-model framework as the ensemble-average of the scaling-based bias-adjusted models is itself found to be scale invariant. Ensembles of climate simulations are required in order to identify robust climate changes for rainfall extremes, as the single-model significance of a change is generally too low due to the small signal-to-noise ratio, associated to atmospheric circulation (Shepherd, 2014). Ensemble projections can be used to estimate the probability distribution of climate change under different emission scenarios.

While bias adjustment has been shown to be useful, it can not be a solid alternative to entirely replace poorly understood physical processes. Convection-permitting models (CPMs) are more recommended for projecting subdaily precipitation extremes (Ban et al., 2021; Lucas-Picher et al., 2021; Van de Vyver et al., 2021) but their availability is mostly limited to single-model efforts, short simulation periods or specific domains due to their high computational cost. The number of papers on future projections of IDF curves in CPMs is therefore still very limited (Cannon & Innocenti, 2019; Tabari et al., 2016), and more research is needed to reach more robust conclusions. Recently, an ensemble of (short-duration) CPMs simulations were produced over a number of specific domains in Europe (Coppola et al., 2020) and it would be of interest to apply our bias-adjustment methodology on these. In particular, future research could assess more reliable projections on changes in IDF characteristics such as, for example, the steepening of IDF curves under global warming.

## Appendix A: Validation Scores

Denote by $\hat{i}_{T,m}(d)$, the pseudo future intensity-duration-frequency (IDF)-values based on the $m$-th regional climate model (RCM), and $\hat{i}^{adj}_{T,m'}(d)$ the adjusted IDF-values of the $m'$-th RCM ($m' \neq m$).

### A1. Relative Bias (R-BIAS)

First, the relative error (RE) is given by:

$$\mathrm{RE}_{m,m'} = \frac{\hat{i}^{adj}_{T,m'}(d) - \hat{i}_{T,m}(d)}{\hat{i}_{T,m}(d)}. \tag{A1}$$

Next, for an $M$-member ensemble, the relative bias (R-BIAS) is given by:

$$\text{R-BIAS} = \frac{1}{M\,(M-1)} \sum_{m=1}^{M} \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \mathrm{RE}_{m,m'} \tag{A2}$$

A negative (positive) bias indicates an underestimation (overestimation).

### A2. Mean Squared Error Skill Score (MSESS)

The squared error (SE) is defined as:

$$\mathrm{SE}_{m,m'} = \left(\hat{i}^{adj}_{T,m'}(d) - \hat{i}_{T,m}(d)\right)^2, \tag{A3}$$

and the mean squared error (MSE) is defined as the average over all the pseudo observation/model-pairs ($m$, $m'$):

$$\mathrm{MSE} = \frac{1}{M\,(M-1)} \sum_{m=1}^{M} \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \mathrm{SE}_{m,m'}. \tag{A4}$$

The improvement of a particular adjustment method over a reference prediction (for which we considered the unadjusted future runs, RAW) is measured with the MSE skill score (MSESS):

$$\mathrm{MSESS} = 1 - \frac{\mathrm{MSE}^{adj}}{\mathrm{MSE}^{raw}}. \tag{A5}$$

The skill score MSESS is positive (negative) when the bias-adjustment performs better (worse) than the reference (RAW). Moreover, a perfect adjustment ($\mathrm{MSE}^{\mathrm{adj}} = 0$) results in MSESS = 1, and $\mathrm{MSE}^{\mathrm{adj}} = \mathrm{MSE}^{\mathrm{raw}}$ results in MSESS = 0.

Finally, to get an overall view, the validation scores are averaged over all the locations and then (for MSE) used to calculate the skill scores.

## Data Availability Statement

The EURO-CORDEX data can be accessed on the Earth System Grid Federation infrastructure (https://esgf.llnl.gov/). The observed subdaily precipitation extremes for Belgium are openly available from Zenodo (http://doi.org/10.5281/zenodo.4741178).

**Acknowledgments**
This work is supported by URCLIM and has received funding from EU's H2020 Research and Innovation Program under Grant Agreement 690462. B.V.S. acknowledges support by the Belgian Science Policy (BELSPO) within the REGE+ (B2/212/P1/REGE+) project. L.D.C. acknowledges support from the Belgian Science Policy Office (BELSPO) through the FED-tWIN programme (Prf-2020-017). We acknowledge the World Climate Research Programme's Working Group on Regional Climate, and the Working Group on Coupled Modeling, former coordinating body of CORDEX and responsible panel for CMIP5. We also thank the climate modeling groups (listed in Table S1 in Supporting Information S1) for producing and making available their model output. We also acknowledge the Earth System Grid Federation infrastructure, an international effort led by the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison, the European Network for Earth System Modeling and other partners in the Global Organisation for Earth System Science Portals (GO-ESSP).

## References

Ban, N., Caillaud, C., Coppola, E., Pichelli, E., Sobolowski, S., Adinolfi, M., et al. (2021). The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, Part I: Evaluation of precipitation. *Climate Dynamics*, *57*(1), 275–302. https://doi.org/10.1007/s00382-021-05708-w

Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. (2004). *Statistics of extremes*. John Wiley & Sons, Ltd.

Berg, P., Christensen, O. B., Klehmet, K., Lenderink, G., Olsson, J., Teichmann, C., & Yang, W. (2019). Summertime precipitation extremes in a EURO-CORDEX 0.11° ensemble at an hourly resolution. *Natural Hazards and Earth System Sciences*, *19*(4), 957–971. https://doi.org/10.5194/nhess-19-957-2019

Blanchet, J., Ceresetti, D., Molinié, G., & Creutin, J.-D. (2016). A regional GEV scale-invariant framework for intensity-duration-frequency analysis. *Journal of Hydrology*, *540*, 82–95. https://doi.org/10.1016/j.jhydrol.2016.06.007

Bougadis, J., & Adamowski, K. (2006). Scaling model of a rainfall intensity-duration-frequency relationship. *Hydrological Processes*, *20*(17), 3747–3757. https://doi.org/10.1002/hyp.6386

Burlando, P., & Rosso, R. (1996). Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation. *Journal of Hydrology*, *187*(1–2), 45–64. https://doi.org/10.1016/S0022-1694(96)03086-7

Cannon, A. J., & Innocenti, S. (2019). Projected intensification of sub-daily and daily rainfall extremes in convection-permitting climate model simulations over North America: Implications for future intensity–duration–frequency curves. *Natural Hazards and Earth System Sciences*, *19*(2), 421–440. https://doi.org/10.5194/nhess-19-421-2019

Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, *28*(17), 6938–6959. https://doi.org/10.1175/JCLI-D-14-00754.1

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer-Verlag, London Ltd.

Coppola, E., Sobolowski, S., Pichelli, E., Raffaele, F., Ahrens, B., Anders, I., et al. (2020). A first-of-its-kind multi-model convection permitting ensemble for investigating convective phenomena over Europe and the Mediterranean. *Climate Dynamics*, *55*(1), 3–34. https://doi.org/10.1007/s00382-018-4521-8

Courty, L. G., Wilby, R. L., Hillier, J. K., & Slater, L. J. (2019). Intensity-duration-frequency curves at the global scale. *Environmental Research Letters*, *14*(8), 084045. https://doi.org/10.1088/1748-9326/ab370a

Davison, A. C., & Gholamrezaee, M. M. (2011). Geostatistics of extremes. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *468*(2138), 581–608. https://doi.org/10.1098/rspa.2011.0412

Fadhel, S., Rico-Ramireza, M. A., & Han, D. (2018). Sensitivity of peak flow to the change of rainfall temporal pattern due to warmer climate. *Journal of Hydrology*, *560*, 546–559. https://doi.org/10.1016/j.jhydrol.2018.03.041

Fischer, E. M., & Knutti, R. (2016). Observed heavy precipitation increase confirms theory and early models. *Nature Climate Change*, *6*(11), 986–991. https://doi.org/10.1038/nclimate3110

Fowler, H. J., Lenderink, G., Prein, A. F., Westra, S., Allan, R. P., Ban, N., et al. (2021). Anthropogenic intensification of short-duration rainfall extremes. *Nature Reviews Earth & Environment*, *2*(2), 107–122. https://doi.org/10.1038/s43017-020-00128-6

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research*, *113*(D6), D06104. https://doi.org/10.1029/2007JD008972

Guerreiro, S. B., Fowler, H. J., Barbero, R., Westra, S., Lenderink, G., Blenkinsop, S., et al. (2018). Detection of continental-scale intensification of hourly rainfall extremes. *Nature Climate Change*, *8*(9), 803–807. https://doi.org/10.1038/s41558-018-0245-3

Gupta, V. K., & Waymire, E. (1990). Multiscaling properties of spatial rainfall and river flow distributions. *Journal of Geophysical Research*, *95*(D3), 1999–2009. https://doi.org/10.1029/JD095iD03p01999

Hettiarachchi, S., Wasko, C., & Sharma, A. (2018). Increase in flood risk resulting from climate change in a developed urban watershed - The role of storm temporal patterns. *Hydrology and Earth System Sciences*, *22*(3), 2041–2056. https://doi.org/10.5194/hess-22-2041-2018

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society B*, *52*(1), 105–124. https://doi.org/10.1111/j.2517-6161.1990.tb01775.x

Hosseinzadehtalaei, P., Ishadi, N. K., Tabari, H., & Willems, P. (2021). Climate change impact assessment on pluvial flooding using a distribution-based bias correction of regional climate model simulations. *Journal of Hydrology*, *598*, 126239. https://doi.org/10.1016/j.jhydrol.2021.126239

Innocenti, S., Mailhot, A., & Frigon, A. (2017). Simple scaling of extreme precipitation in North America. *Hydrology and Earth System Sciences*, *21*(11), 5823–5846. https://doi.org/10.5194/hess-21-5823-2017

Innocenti, S., Mailhot, A., Frigon, A., Cannon, A. J., & Leduc, M. (2019). Observed and simulated precipitation over Northeastern North America: How do daily and subdaily extremes scale in space and time? *Journal of Climate*, *32*(24), 8563–8582. https://doi.org/10.1175/JCLI-D-19-0021.1

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., et al. (2014). EURO-CORDEX: New high-resolution climate change projections for European impact research. *Regional Environmental Change*, *14*(2), 563–578. https://doi.org/10.1007/s10113-013-0499-2

Jacob, D., Teichmann, C., Sobolowski, S., Katragkou, E., Anders, I., Belda, M., et al. (2020). Regional climate downscaling over Europe: Perspectives from the EURO-CORDEX community. *Regional Environmental Change*, *20*(2), 51. https://doi.org/10.1007/s10113-020-01606-9

Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, *25*(8), 1287–1304. https://doi.org/10.1016/S0309-1708(02)00056-8

Koutsoyiannis, D., Kozonis, D., & Manetas, A. (1998). A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, *206*(1), 118–135. https://doi.org/10.1016/S0022-1694(98)00097-3

Kreienkamp, F., Philip, S. Y., Tradowsky, J. S., Kew, S. F., Lorenz, P., Arrighi, J., et al. (2021). Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021. Retrieved from https://www.worldweatherattribution.org/wp-content/uploads/Scientific-report-Western-Europe-floods-2021-attribution.pdf

Leadbetter, M., Lindgren, G., & Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer.

Lucas-Picher, P., Argüeso, D., Brisson, E., Tramblay, Y., Berg, P., Lemonsu, A., et al. (2021). Convection-permitting modeling with regional climate models: Latest developments and next steps. *WIREs Climate Change*, *12*(6), e731. https://doi.org/10.1002/wcc.731

Maraun, D. (2012). Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. *Geophysical Research Letters*, *39*(6), L06706. https://doi.org/10.1029/2012GL051210

Maraun, D. (2016). Bias correcting climate change simulations - A critical review. *Current Climate Change Reports*, 2(4), 211–220. https://doi.org/10.1007/s40641-016-0050-x

Maraun, D., & Widmann, M. (2018). *Statistical downscaling and bias correction for climate research*. Cambridge University Press. https://doi.org/10.1017/9781107588783

Martins, E. S., & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, *36*(3), 737–744. https://doi.org/10.1029/1999WR900330

Mascaro, G. (2020). Comparison of local, regional, and scaling models for rainfall intensity-duration-frequency analysis. *Journal of Applied Meteorology and Climatology*, *59*(9), 1519–1536. https://doi.org/10.1175/JAMC-D-20-0094.1

Mélèse, V., Blanchet, J., & Molinié, G. (2018). Uncertainty estimation of intensity-duration-frequency relationships: A regional analysis. *Journal of Hydrology*, *558*, 579–591. https://doi.org/10.1016/j.jhydrol.2017.07.054

Menabde, M., Seed, A., & Pegram, G. (1999). A simple scaling model for extreme rainfall. *Water Resources Research*, *35*(1), 335–339. https://doi.org/10.1029/1998WR900012

Nadarajah, S., Anderson, C. W., & Tawn, J. A. (1998). Ordered multivariate extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *60*(2), 473–496. https://doi.org/10.1111/1467-9868.00136

Naveau, P., Nogaj, M., Ammann, C., Yiou, P., Cooley, D., & Jomelli, V. (2005). Statistical methods for the analysis of climate extremes. *Comptes Rendus Geoscience*, *337*(10), 1013–1022. https://doi.org/10.1016/j.crte.2005.04.015

Nguyen, V., Nguyen, T., & Wang, H. (1998). Regional estimation of short duration rainfall extremes. *Water Science and Technology*, *37*(11), 15–19. https://doi.org/10.1016/S0273-1223(98)00311-4

O'Gorman, P. A. (2015). Precipitation extremes under climate change. *Current Climate Change Reports*, *1*(2), 49–59. https://doi.org/10.1007/s40641-015-0009-3

Overeem, A., Buishand, A., & Holleman, I. (2008). Rainfall depth-duration-frequency curves and their uncertainties. *Journal of Hydrology*, *348*(1), 124–134. https://doi.org/10.1016/j.jhydrol.2007.09.044

Pall, P., Aina, T., Stone, D. A., Stott, P. A., Nozawa, T., Hilberts, A. G., et al. (2011). Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature*, *470*(7334), 382–385. https://doi.org/10.1038/nature09762

Panthou, G., Vischel, T., Lebel, T., Quantin, G., & Molinié, G. (2014). Characterising the space–time structure of rainfall in the Sahel with a view to estimating IDAF curves. *Hydrology and Earth System Sciences*, *18*(12), 5093–5107. https://doi.org/10.5194/hess-18-5093-2014

Philip, S., Kew, S., van Oldenborgh, G. J., Otto, F., Vautard, R., van der Wiel, K., et al. (2020). A protocol for probabilistic extreme event attribution analyses. *Advances in Statistical Climatology, Meteorology and Oceanography*, *6*(2), 177–203. https://doi.org/10.5194/ascmo-6-177-2020

Rajczak, J., & Schär, C. (2017). Projections of future precipitation extremes over Europe: A multimodel assessment of climate simulations. *Journal of Geophysical Research: Atmospheres*, *122*(20), 10773–10800. https://doi.org/10.1002/2017JD027176

Reichler, T., & Kim, J. (2008). How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society*, *89*(3), 303–312. https://doi.org/10.1175/BAMS-89-3-303

Schmith, T., Thejll, P., Berg, P., Boberg, F., Christensen, O. B., Christiansen, B., et al. (2021). Identifying robust bias adjustment methods for European extreme precipitation in a multi-model pseudo-reality setting. *Hydrology and Earth System Sciences*, *25*(1), 273–290. https://doi.org/10.5194/hess-25-273-2021

Shepherd, T. G. (2014). Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geoscience*, *25*(10), 703–708. https://doi.org/10.1038/ngeo2253

Tabari, H., Hosseinzadehtalaei, P., Willems, P., Saeed, S., Brisson, E., & Van Lipzig, N. (2016). How will be future rainfall IDF curves in the context of climate change? In *Sustainable hydraulics in the era of global change: Proceedings of the 4th IAHR Europe congress* (pp. 796–803).

Tebaldi, C., Arblaster, J. M., & Knutti, R. (2011). Mapping model agreement on future climate projections. *Geophysical Research Letters*, *38*(23), L23701. https://doi.org/10.1029/2011GL049863

Tebaldi, C., & Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *365*(1857), 2053–2075. https://doi.org/10.1098/rsta.2007.2076

Van de Vyver, H. (2015). Bayesian estimation of rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, *529*(3), 1451–1463. https://doi.org/10.1016/j.jhydrol.2015.08.036

Van de Vyver, H. (2018). A multiscaling-based intensity-duration-frequency model for extreme precipitation. *Hydrological Processes*, *32*(11), 1635–1647. https://doi.org/10.1002/hyp.11516

Van de Vyver, H., Van Schaeybroeck, B., De Troch, R., De Cruz, L., Hamdi, R., Villanueva-Birriel, C., et al. (2021). Evaluation framework for subdaily rainfall extremes simulated by regional climate models. *Journal of Applied Meteorology and Climatology*, *60*(10), 1423–1442. https://doi.org/10.1175/JAMC-D-21-0004.1

Van Schaeybroeck, B., & Vannitsem, S. (2016). Assessment of calibration assumptions under strong climate changes. *Geophysical Research Letters*, *43*(3), 1314–1322. https://doi.org/10.1002/2016GL067721

Vrac, M., & Friederichs, P. (2015). Multivariate–intervariable, spatial, and temporal–bias correction. *Journal of Climat*e, *28*(1), 218–237. https://doi.org/10.1175/JCLI-D-14-00059.1

Westra, S., Fowler, H., Evans, J., Alexander, L., Berg, P., Johnson, F., et al. (2014). Future changes to the intensity and frequency of short-duration extreme rainfall. *Reviews of Geophysics*, *52*(3), 522–555. https://doi.org/10.1002/2014RG000464

## References From the Supporting Information

Bellprat, O., Kotlarski, S., Lüthi, D., & Schär, C. (2013). Physical constraints for temperature biases in climate models. *Geophysical Research Letters*, *40*(15), 4042–4047. https://doi.org/10.1002/grl.50737

Charles, S. P., Bates, B. C., Whetton, P. H., & Hughes, J. P. (1999). Validation of downscaling models for changed climate conditions: Case study of southwestern Australia. *Climate Research*, *12*(1), 1–14. https://doi.org/10.3354/cr012001

Frías, M. D., Zorita, E., Fernández, J., & Rodríguez-Puebla, C. (2006). Testing statistical downscaling methods in simulated climates. *Geophysical Research Letters*, *33*(19), L19807. https://doi.org/10.1029/2006GL027453

Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., et al. (2015). VALUE: A framework to validate downscaling approaches for climate change studies. *Earth's Future*, *3*(1), 1–14. https://doi.org/10.1002/2014EF000259

Räisänen, J., & Räty, O. (2013). Projections of daily mean temperature variability in the future: Cross-validation tests with ENSEMBLES regional climate simulations. *Climate Dynamics*, *41*(5), 1553–1568. https://doi.org/10.1007/s00382-012-1515-9

Räty, O., Räisänen, J., Bosshard, T., & Donnelly, C. (2018). Intercomparison of univariate and joint bias correction methods in changing climate from a hydrological perspective. *Climate*, *6*(2), 33. https://doi.org/10.3390/cli6020033

Räty, O., Räisänen, J., & Ylhäisi, J. S. (2014). Evaluation of delta change and bias correction methods for future daily precipitation: Intermodel cross-validation using ENSEMBLES simulations. *Climate Dynamics*, *42*(9), 2287–2303. https://doi.org/10.1007/s00382-014-2130-8

Van de Velde, J., Demuzere, M., De Baets, B., & Verhoest, N. E. C. (2022). Impact of bias nonstationarity on the performance of uni- and multivariate bias-adjusting methods: A case study on data from Uccle, Belgium. *Hydrology and Earth System Sciences*, *26*(9), 2319–2344. https://doi.org/10.5194/hess-26-2319-2022

Velázquez, J. A., Troin, M., Caya, D., & Brissette, F. (2015). Evaluating the time-invariance hypothesis of climate model bias correction: Implications for hydrological impact studies. *Journal of Hydrometeorology*, *16*(5), 2013–2026. https://doi.org/10.1175/JHM-D-14-0159.1

Vrac, M., Stein, M. L., Hayhoe, K., & Liang, X.-Z. (2007). A general method for validating statistical downscaling methods under future climate change. *Geophysical Research Letters*, *34*(18), L18701. https://doi.org/10.1029/2007GL030295