https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

Climate model ensembles reveal increasing future extreme precipitation in Northern Italy

Yue Lai, Rui Guo, Alberto Montanari

Department of Civil, Chemical, Environmental and Materials Engineering (DICAM), University of Bologna, Bologna, 40136, Italy

Correspondence to: Yue Lai (yue.lai2@unibo.it)

Abstract. Assessing future changes in extreme precipitation remains a critical challenge for climate adaptation and flood risk mitigation. While growing observational evidence indicates an intensification of short-duration extreme precipitation, robust quantification of these changes and how it will change in the future remains uncertain. Using one of the longest continuous daily precipitation records available, dating back to 1 January 1813 in Bologna, we evaluate the ability of 22 bias-corrected CMIP6 climate models to reproduce historical extreme precipitation characteristics based on four standardized extreme precipitation indices. Building on this evaluation, we develop a comprehensive assessment framework that combines bias correction, independent validation and multi-model ensemble (MME) schemes. We present a comparison of Bayesian Model Averaging (BMA) with Simple Model Averaging (SMA) and individual model simulations, and apply both MME schemes to project 21st-century changes in extreme precipitation under different emission scenarios. The results show that MME schemes provide a more stable and robust representation of extreme precipitation than most individual models. Among the projected changes, the intensity indices show robust increasing signals across most scenarios and periods, with more pronounced enhancement under high-emission conditions in the far future. By contrast, the frequency-related indices show more complex responses: the annual count of days where precipitation is greater than or equal to 20mm and 10mm (R20mm and R10mm) exhibits a weak increasing tendency and no sustained increasing signal, respectively, with R10mm projected to decline under SSP5-8.5. This study provides a robust framework for assessing future changes in station-scale extreme precipitation and for better characterizing the associated projection uncertainties.

1 Introduction

Global warming is likely to accelerate the global hydrological cycle, with observed increasing in global and regional extreme rainfall over recent decades (Fowler et al., 2021; IPCC, 2021; Papalexiou and Montanari, 2019). According to the Clausius–Clapeyron relationship, the water-holding capacity of the atmosphere increases by approximately 7% per degree warming (K^{-1}), providing a thermodynamic basis for the intensification of extreme precipitation under global warming (Pall et al., 2017; Risser and Wehner, 2017; Wang et al., 2018). In fact, many regions worldwide are increasingly exposed to rare but high-impact precipitation events, typically associated with return periods of a decade or longer (IPCC, 2021). Over the past five

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

years, a series of extreme precipitation events occurred worldwide has provided compelling observational evidence for both the intensification and increased occurrence of such events. From the record-breaking rainfall in Vietnam in 2025 (World Meteorological Organization, 2025), the “Daniel” storm with exceptional four-day accumulation of 1235 mm in eastern Thessaly, Greece, in 2023 (Argüeso et al., 2024; Qiu et al., 2023), the more than 800mm precipitation within three days during the 2021 Henan event in China (Qin et al., 2022; Xiao et al., 2023), to several similar extreme rainfall events in Spain (Huang et al., 2025) and Italy (Iacomino et al., 2025), these events illustrate an increasing departure of extreme precipitation from their historical variability. Extreme precipitation and the associated flooding not only disrupt ecosystems but also impose substantial socioeconomic losses (Li et al., 2019; Varghese et al., 2025). As the frequency and intensity of these "black swan" events are projected to escalate (IPCC, 2021), quantifying the future changes in the characteristics of extreme precipitation is crucial for robust climate adaptation and flood risk management.

General Circulation Models (GCMs) are widely used to simulate the dynamics and state of the climate system (Chae et al., 2024; Guo and Montanari, 2023; Orlowsky and Seneviratne, 2012; Wang et al., 2017). The Coupled Model Intercomparison Project (CMIP) has progressed through successive phases, with CMIP6 further refining forcing datasets, experimental design, and multi-model evaluation (Chae et al., 2024). The introduction of the Shared Socioeconomic Pathway (SSP) framework further enables climate projections to consistently link socioeconomic development with emission trajectories, strengthening scenario-based impact assessments (Eyring et al., 2016). To facilitate systematic detection and comparison of climate extremes, the Expert Team on Climate Change Detection and Indices (ETCCDI) has developed a standardized suite of precipitation indices, now widely adopted for assessing extreme precipitation across regions and scales (ETCCDI, 2020). Within this framework, GCMs are extensively used to quantify projected changes in extreme precipitation indices (Chae et al., 2024). However, GCM-based simulations continue to exhibit substantial uncertainty arising from systematic model biases, internal climate variability, and structural differences among models, even with ongoing improvements in forcing data and model resolution (Abdelmoaty et al., 2021; Eden et al., 2012; Pathak et al., 2023; Song et al., 2024). To reduce these uncertainties, a range of statistical bias correction methods, including quantile mapping, quantile delta mapping, and cumulative distribution function transformation, has been widely applied to improve the representation of the precipitation extremes (Allard et al., 2024; Padulano et al., 2025; Tong et al., 2021). Nevertheless, substantial inter-model spread persists even after bias correction, underscoring the inherent challenges in robustly qualifying future projections of extreme precipitation (Song et al., 2022).

To narrow the uncertainty range of GCM simulations and enhance the robustness of future projections, the multi-model ensemble (MME) method has been widely adopted (Pavan and Doblas-Reyes, 2000). MME combines simulations from multiple GCMs to mitigate individual model structural limitations (Jiang et al., 2016). The Simple Model Averaging (SMA) is the simplest and most commonly applied method of handling multi-model ensembles (Zhu et al., 2020). Previous studies have shown that SMA can, to some extent, reduce systematic biases and internal variability relative to single-model simulations (Dong et al., 2015; Lei et al., 2023; Raghavan et al., 2018; Yang et al., 2020). However, the inclusion of poorly performing or structurally inconsistent models can undermine the credibility and overall performance of the multi-model ensemble, thereby limiting its effectiveness (Zhu et al., 2023). Consequently, various performance-based ensemble weighting strategies have

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

been developed to address the limitations of SMA, including rank-based weighting, reliability ensemble averaging, weighting schemes that account for both model independence and performance, and Bayesian model averaging (Chen et al., 2011; Knutti et al., 2017; Li et al., 2021; Zhu et al., 2021). These methods typically assign model weights according to their ability to reproduce observed climate characteristics, under the assumption that superior historical performance implies greater credibility for future projections.

Despite these methodological advances, the majority of existing studies on both bias correction and ensemble weighting share a common limitation: the statistical relationships and model weights are estimated over the historical period and evaluated over that same period (Abramowitz et al., 2019; Brunner et al., 2020; Faghih et al., 2022; Tan et al., 2020). When a method is both developed and assessed on the same historical period, apparent performance may simply reflect adaptation to that particular dataset rather than a reliable skill that holds when applied to a different period. The problem is particularly consequential for future climate projections, which cannot be verified against observations, the credibility of such projections therefore depends critically on whether these methods perform reliably when applied to periods outside their training window. Whether bias correction relationships derived from one historical sub-period can be stably transferred to another, and whether ensemble weights estimated over a given period retain their validity for projections beyond the training period, are questions that the existing literature has not systematically addressed.

To mitigate this limitation, this study develops a comprehensive framework for extreme precipitation projection centred on independent validation. Using long-term daily precipitation records as the observational baseline, the historical period is partitioned into distinct training and validation periods: QDM bias correction relationships and MME weights are estimated strictly from the training period and then evaluated on an independent validation period, providing a direct measure of their temporal transferability across historical sub-periods. A sensitivity analysis further examines whether these results depend on the specific choice of training-validation split. Building on this validated framework, the bias correction relationships and ensemble weights derived from the training period are carried forward to project future changes in extreme precipitation under multiple SSP emission scenarios, with the robustness of the projected changes assessed through the directional consistency of individual model simulations.

Accordingly, this study aims to (1) evaluate the ability of the latest-generation of GCMs to reproduce the statistical characteristics of historical extreme precipitation; (2) develop a comprehensive evaluation framework for extreme precipitation based on independent validation and evaluate the temporal transferability of the QDM and MME methods; and (3) assess future changes in extreme precipitation under different emission scenarios and evaluate the robustness of the projected changes.

2 Data

2.1 Rainfall observations

Daily observed precipitation data in Bologna were obtained from the European Climate Assessment & Dataset (ECA&D). ECA&D is a climate observation database jointly established by meteorological and research institutions in multiple European

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

countries, compiling daily observation data from a large number of surface meteorological stations in Europe and the Mediterranean region. The dataset has undergone systematic quality control and comprehensive metadata management, including documentation of station location (latitude, longitude, elevation), instrumentation changes, and station relocations, to support long-term monitoring and analysis of regional climate variability and extremes (https://www.ecad.eu) (Klein Tank et al., 2002). The daily precipitation data provided by ECA&D are based on actual observations, exhibiting good temporal consistency and scientific reproducibility (Klein Tank et al., 2002).

The continuous daily precipitation record at the Bologna meteorological station in Italy dates back to 1813 and extends to the present, representing one of the longest uninterrupted daily rainfall observations in Italy and worldwide. This record provides a unique data basis for examining the long-term evolution of extreme precipitation from the pre-industrial period to modern times. The Bologna daily precipitation series provided in the ECA&D database is compiled from the original manual records, preserving the original daily observations without additional diurnal-scale adjustments. This data has undergone multiple quality checks, with an extremely low missing data rate (less than 1%), and can be considered a nearly homogeneous sequence, suitable for long-term trend analysis and extreme precipitation research.

Figure 1 shows the daily time series and 10-year moving average of Bologna rainfall from 1813 to 2024. The mean fluctuates between a minimum of 1.2 mm and a maximum of 2.5 mm. After reaching a peak around 1902, it began to decline and then slowly increased from 1920 to the present. The black horizontal line in the figure represents the 99th percentile (52.31 mm) based on the historical period (1813-2024). Daily precipitation exceeding this threshold is defined as an extreme rainfall event. The record indicates that extreme rainfall days occurred frequently during the 1830s to 1890s, and have again been prevalent since the 1960s, highlighting distinct periods of elevated extreme-precipitation activity.

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

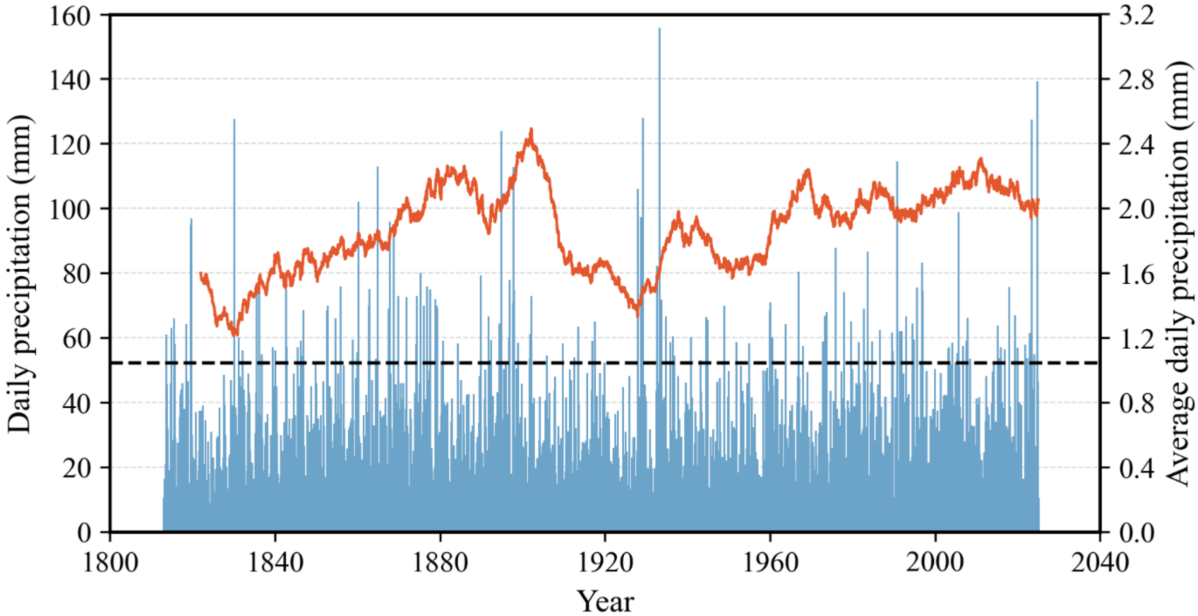


Figure 1: Daily rainfall records for Bologna during 1813-2024, with 10-year moving average (orange line) and 99th percentile based on historical period (black line).

2.2 Global climate models simulation

Global climate simulation data from the Coupled Model Intercomparison Project Phase 6 (CMIP6) are publicly available from the Earth System Grid Federation (ESGF) data portal (https://esgf-node.llnl.gov). The dataset used in this study includes historical simulations (1850-2014) and future projections (2015–2100). The future projections were produced under the emission scenarios "Shared Socioeconomic Pathway" (SSP) SSP1-2.6, SSP2-4.5, and SSP5-8.5.

In this study, only the first realization of each ensemble member, denoted as r1i1p1f1, is analysed. Here, "r," "i," "p," and "f" represent the model run initialization, initialization method, physical parameterization version, and forcing dataset, respectively. We consider all models providing future simulations under three emission scenarios. Table 1 shows details of the 22 GCMs selected in this study.

Table 1. Description of the 22 CMIP6 models used in this study.

NO.	Model	Institution	Realization	Resolution
1	ACCESS-CM2	CSIRO-ARCCSS/Australia	r1i1p1f1	1.875°×1.250°
2	ACCESS-ESM1-5	CSIRO/Australia	r1i1p1f1	1.875°×1.241°
3	BCC-CSM2-MR	BCC/China	r1i1p1f1	1.125°×1.125°
4	CMCC-CM2-SR5	CMCC/Italy	r1i1p1f1	1.250°×0.938°
5	CMCC-ESM2	CMCC/Italy	r1i1p1f1	1.250°×0.938°

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

6	CanESM5	CCCma/Canada	r1i1p1f1	2.813°×2.813°
7	EC-Earth3-Veg-LR	EC-Earth-Consortium/EU	r1i1p1f1	1.125°×1.125°
8	FGOALS-g3	LASG-IAP/China	r1i1p1f1	2.000°×2.250°
9	GFDL-ESM4	NOAA-GFDL/USA	r1i1p1f1	1.250°×1.000°
10	IITM-ESM	CCCR-IITM/India	r1i1p1f1	1.875°×1.915°
11	INM-CM4-8	INM/Russia	r1i1p1f1	2.000°×1.500°
12	INM-CM5-0	INM/Russia	r1i1p1f1	2.000°×1.500°
13	IPSL-CM6A-LR	IPSL/France	r1i1p1f1	2.500°×1.259°
14	KACE-1-0-G	NIMS-KMA/Korea	r1i1p1f1	1.875°×1.250°
15	KIOST-ESM	KIOST/Korea	r1i1p1f1	1.875°×1.875°
16	MIROC6	MIROC/Japan	r1i1p1f1	1.406°×1.406°
17	MPI-ESM1-2-LR	MPI/Germany	r1i1p1f1	1.875°×1.875°
18	MRI-ESM2-0	MRI/Japan	r1i1p1f1	1.125°×1.125°
19	NESM3	NUIST/China	r1i1p1f1	1.875°×1.875°
20	NorESM2-LM	NCC/Norway	r1i1p1f1	2.500°×1.875°
21	NorESM2-MM	NCC/Norway	r1i1p1f1	1.250°×0.938°
22	TaiESM1	RCEC/Taiwan	r1i1p1f1	1.250°×0.938°

3 Methodology

3.1 Framework

In this study, we adopted two distinct temporal definitions for different analytical objectives. The first refers to the training period and independent validation period dedicated to model performance evaluation. The second is a unified climatological baseline period (1985-2014), serving two key purposes: firstly, calculating the percentile thresholds for indicators such as cumulative precipitation during an extremely wet period; secondly, acting as a consistent reference climatology to quantify relative changes in projected future extreme precipitation indices.

The overall methodological framework of this study is illustrated in Figure 2. The central component of the framework is an independent validation strategy: statistical relationships are first estimated over the training period (1850-1949), then evaluated over the independent validation period (1950-2014), and finally transferred directly to future projections (2015-2100). The year 1949 was chosen as the split point because it approximately coincides with the onset of the Great Acceleration, which refers to the rapid intensification of anthropogenic forcing beginning around 1950 (Steffen et al., 2015). This ensures that the validation period operates under a forcing regime distinct from the training period, making it a more demanding test of whether the statistical relationships established during training hold beyond their original conditions. In addition, sensitivity

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

analyses based on alternative training-validation splits were conducted to examine the influence of period selection on the results.

Daily precipitation simulations from 22 CMIP6 models were bilinearly interpolated to Bologna station. The QDM was then trained using observed and simulated daily precipitation data from the training period. The resulting transfer functions were subsequently applied to the independent validation period and future simulations without further modification.

Four extreme precipitation indices, namely Rx1day, R99p, R10mm and R20mm, were then calculated from the bias-corrected daily precipitation series. Rx1day and R99p are maximum daily precipitation and cumulative precipitation during an extremely wet period (see Table 2) while R10mm and R20mm are the annual count of days with rainfall greater than 10 mm and 20 mm, respectively. The 99th percentile thresholds required for R99p calculation were computed once and exclusively within the 1985-2014 climatological baseline period. Separate thresholds were calculated for observational datasets and each individual climate model. These fixed thresholds were consistently applied across all analytical stages without recalculation or revision. Based on these indices, two MME schemes, Simple Model Averaging (SMA) and Bayesian Model Averaging (BMA), were constructed. Both MMEs were evaluated within the independent validation framework, where BMA model weights were estimated during the training period and SMA assigned equal weights to all models. During the validation period, the 22 individual models, SMA and BMA were compared to assess their relative performance. The BMA and SMA model weights were subsequently transferred directly to future projections. Finally, future trends, relative changes, and the robustness of projected changes in extreme precipitation indices were assessed based on the resulting ensemble projections. All relative changes in the indices were quantified relative to the climatological conditions of the 1985-2014 baseline period.

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

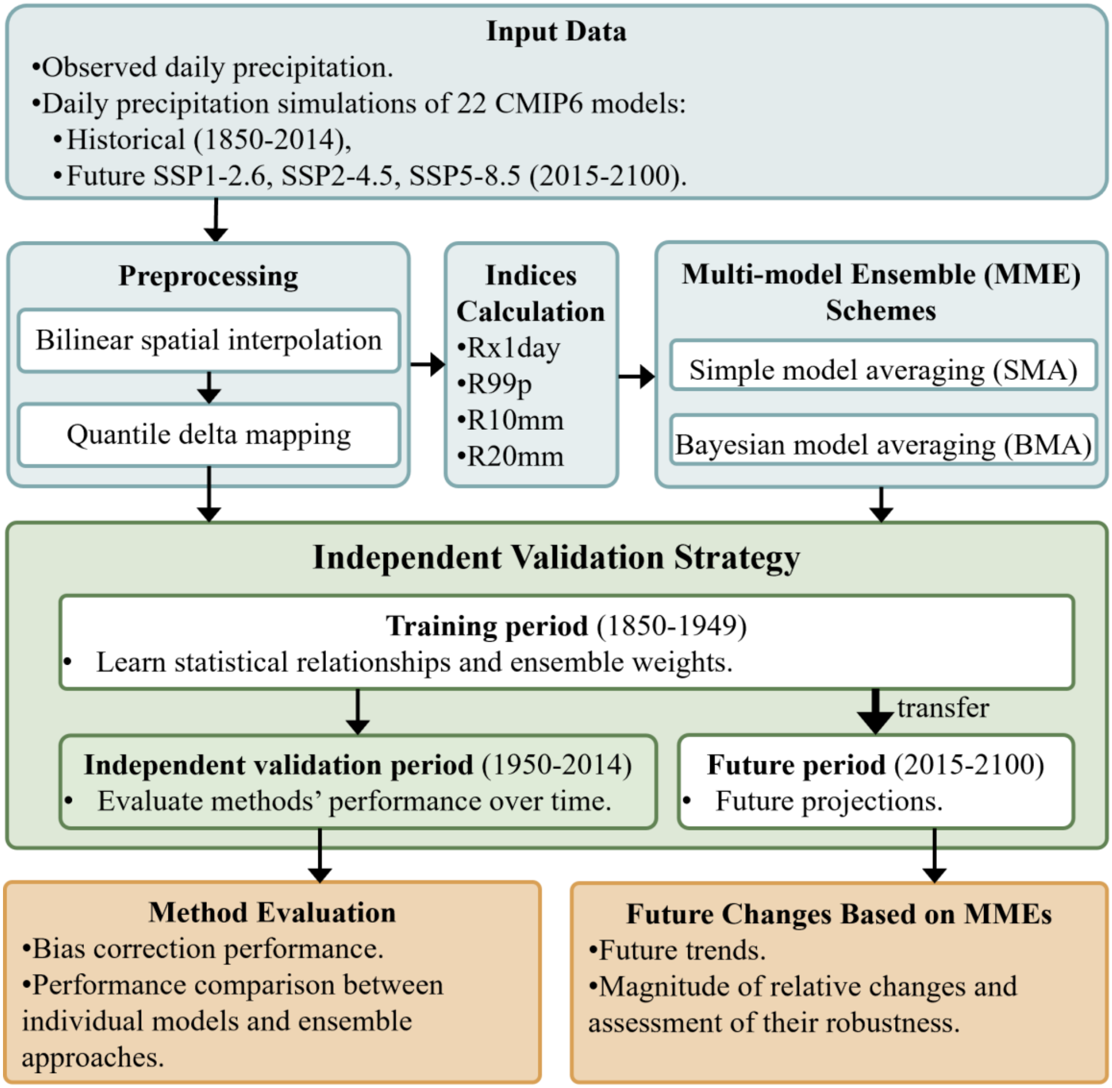


Figure 2: Overall methodological framework of this study.

3.2 Extreme precipitation indices

The four extreme precipitation indices used in this study follow the definitions of the Expert Team on Climate Change Detection and Indices (ETCCDI), with minor adaptations based on the approaches of Tang (Tang et al., 2021) and Kim (Kim et al., 2020) and related studies (Table 2). The selection of indices, encompassing both intensity and frequency, effectively reflects the mean characteristics of precipitation and its extremes and has been widely applied in previous studies (Alexander and Arblaster, 2009; Chen et al., 2025; Marengo et al., 2020; Pellicone and Caloiero, 2025; Thomas et al., 2023). The 99th percentile thresholds for R99p and all relative change calculations were computed using the unified climatological baseline period defined in Section 3.1.

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

170 **Table 2. Definitions of extreme precipitation indices used in this study.**

Index	Description	Definition	Units
Rx1day	Maximum 1-day precipitation	Let RR_{ij} be the daily precipitation amount on day i in period j. The maximum 1-day value for period j are: $Rx1day_j = \max(RR_{ij})$	mm
R99p	Extremely wet precipitation	Total precipitation when daily precipitation exceeds the extreme 99th percentile threshold. Given a study period j. RR_{wj} is the daily precipitation amount on a wet day w (RR≥1.0mm) in period j and w_j is the total wet days in period j. $RR_{wn}99$ is the 99th percentile of precipitation on wet days in the 1985-2014 period. $R99p_j = \sum_{w=1}^{W_j} RR_{wj}$, where $RR_{wj} > RR_{wn}99$	mm
R10mm	Wet days	Annual count of days when daily precipitation greater than or equal to 10mm. Let RR_{ij} be the daily precipitation amount on day i in period j. Count the number of days where: $RR_{ij} \geq 10$mm	d
R20mm	Very wet days	Annual count of days when daily precipitation greater than or equal to 20mm. Let RR_{ij} be the daily precipitation amount on day i in period j. Count the number of days where: $RR_{ij} \geq 20$mm	d

3.3 Bias correction using quantile delta mapping (QDM)

Precipitation outputs from grid-based GCM simulations represent grid-average values at the grid scale, whereas observational data consist of in situ measurements taken at the point (or station) scale. Discrepancies between the two may arise from sub-grid spatial variability that cannot be fully resolved by the models. This mismatch is particularly pronounced at the daily time
175 scale because precipitation processes inherently exhibit strong spatial heterogeneity. At short temporal scales (e.g., daily), such spatial differences cannot be smoothed by temporal averaging, leading to more pronounced contrasts between grid-averaged precipitation estimates and station observations.

To minimize this bias, we first applied bilinear spatial interpolation to estimate model precipitation at the station location using the four grid points closest to the Bologna station. Subsequently, the daily precipitation simulations were bias-corrected

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

using the quantile delta mapping (QDM) method, which has been widely applied in climate data bias correction (Guo and Montanari, 2023; Jahanshahi et al., 2024). Compared with traditional quantile mapping, QDM corrects systematic model biases while preserving the simulated climate change signal. In addition, the method avoids extrapolation of the historical cumulative distribution function when future simulated values exceed the historical range, thereby improving the stability of bias correction (Cannon et al., 2015).

The primary procedure of the QDM method involves determining the relative change between the historical reference period and the projected period. Let P_{oh}, P_{mh}, and P_{mf} denote the daily precipitation series for the observed historical period, the model-simulated historical period, and the model-simulated future period, respectively. For each quantile level p, the corresponding precipitation amounts are represented as $P_{oh}(p)$, $P_{mh}(p)$, and $P_{mf}(p)$. The relative change factor between the model's future and historical distributions is computed as (Qin and Dai, 2022; Tong et al., 2021):

$$\Delta(p) = \frac{P_{mf}(p)}{P_{mh}(p)} . \tag{1}$$

The bias-corrected future quantile precipitation is then obtained by applying this change factor to the observed historical quantile:

$$P_f(p) = \Delta(p) \times P_{oh}(p) = \frac{P_{mf}(p)}{P_{mh}(p)} \times P_{oh}(p) . \tag{2}$$

To generate the bias-corrected daily precipitation series for the future period, each model future daily value $P_{mf}(t)$ is first mapped to its quantile in the model future distribution:

$$p(t) = F_{mf}(P_{mf}(t)) , \tag{3}$$

where $F_{mf}(\cdot)$ is the empirical CDF of P_{mf}. The corrected daily value is then obtained by evaluating the corrected quantile function at this probability level:

$$P_f(t) = P_f\big(p(t)\big) . \tag{4}$$

3.4 Multi-model ensemble methods

We compared the performance of MMEs constructed using two different weighting approaches with that of individual models. The two schemes include the widely used SMA method and the BMA method. In this study, all MME schemes estimate model weights based on annual extreme precipitation index derived from CMIP6 daily precipitation simulations.

3.4.1 Simple model averaging (SMA)

The SMA is one of the widely used ensemble methods in climate simulation studies. In this method, no individual weighting is applied to the outputs of multiple climate models. Instead, the ensemble mean is obtained by taking the arithmetic average of all model simulations, as follows:

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

$$p = \frac{p_1 + \cdots + p_n}{n} \,. \tag{5}$$

where p_i denotes the output of each model, and n represents the total number of models included in the study. Under this weighting scheme, all models are assigned equal weight $1/n$.

3.4.2 Bayesian model averaging (BMA)

BMA is a statistical ensemble model based on Bayesian theory. It integrates inference and prediction by combining the prior probabilities and likelihood functions of the ensemble members to obtain a more robust posterior predictive distribution (Duan et al., 2007). In this study, BMA is used to construct a CMIP6 MME of extreme precipitation indices. The procedure for estimating the optimal BMA weights is described below (Duan et al., 2007; Wei et al., 2023).

During the training period, the predictive probability density function $p(y_t|\mathbf{f}_t)$ of the MME constructed using BMA can be expressed as:

$$p(y_t|\mathbf{f}_t) = \sum_{m=1}^{M} w_m \times g_m\big(y_t|f_{m,t}, \theta\big) \,. \tag{6}$$

here,y_t denotes the observed extreme precipitation index for year t, and $\mathbf{f}_t = \big(f_{1,t}, f_{2,t}, \cdots, f_{M,t}\,\big)$ represents the ensemble of simulated values from all models in year t, where $f_{m,t}$ is the simulated value from the m$^{\text{th}}$ model, M is the total number of models, and w_m is the BMA weight for the m$^{\text{th}}$ model, $g_m\big(y_t|f_{m,t}, \theta\big)$ denotes the conditional probability density (or mass) function given the simulated value of the m$^{\text{th}}$ model, and θ denotes the corresponding distribution parameter.

The BMA weights represent the relative posterior contributions of individual models to the ensemble prediction distribution. The weights are constrained to be non-negative and sum to one:

$$w_m \geq 0, \sum_{m=1}^{M} w_m = 1 \,. \tag{7}$$

The posterior mean of the BMA predictive distribution was used as the deterministic ensemble estimate and is given by:

$$\hat{y}_t^{BMA} = E(Y_t|\boldsymbol{f}_t) = \sum_{m=1}^{M} w_m E_m(Y_t) \,. \tag{8}$$

where $E_m(Y_t)$is the conditional expectation of the m-th model component. For most conditional distributions used in this study, the model simulation $f_{m,t}$represents the component mean, and Eq. (8) reduces to the weighted average of the model simulations.

The BMA weights and distribution parameters are estimated by maximizing the log-likelihood function over the training period:

$$\ell(\boldsymbol{w}, \theta) = \sum_{t=1}^{T} log\big[\sum_{m=1}^{M} w_m\, g_m\big(y_t \,\big|\, f_{m,t}, \theta\big)\big] \,. \tag{9}$$

where $\boldsymbol{w} = (w_1, w_2, \cdots, w_M)$ denotes the set of BMA weights, and $T$is the number of years in the training period. The maximization was performed using the expectation-maximization (EM) algorithm (Jiang et al., 2012; Raftery et al., 2005). The weights were initialized with equal values for all models, $w_m = 1/M$. The EM algorithm alternates between an expectation step and a maximization step to update the posterior contributions of the individual models and the distribution parameters estimates. The iteration was stopped when the change in log-likelihood fell below the prescribed convergence tolerance or

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

when the maximum number of iterations was reached. More detailed formulations of the EM algorithm can be found in (Fraley et al., 2010; McLachlan and Krishnan, 2007).

To account for the different statistical characteristics of the four extreme precipitation indices, index-specific formulation for the distribution $g_m(y_t|f_{m,t}, \theta)$ were used. Rx1day represents the annual maximum daily precipitation and was modelled using a generalized extreme value (GEV) distribution, as justified by extreme value theory (Coles, 2001). R99p was represented using a Hurdle-Gamma distribution, following the discrete-continuous mixture framework established for precipitation by Sloughter et al. (2007). R10mm and R20mm are annual counts of threshold-exceedance days and were modelled using Negative Binomial distributions.

The estimated BMA weights were then applied to the independent validation period (1950-2014) to evaluate model performance on data not used in calibration. After validation, the BMA weights derived from the training period were applied to project future changes under different emission scenarios.

3.5 Independent validation and performance evaluation

All model and ensemble performance evaluations were conducted over the independent validation period (1950–2014).

3.5.1 Evaluation metrics

The performance of QDM bias correction, two MME schemes, and the individual models was independently validated and compared using three statistical metrics. These metrics include bias, mean absolute error (MAE), and root mean square error (RMSE). Specifically, bias is used to measure systematic deviation, MAE and RMSE are used to quantify mean error between the simulated and observed distribution. Details of these metrics are provided in Table 3.

Table 3. The information of the evaluation metrics used in this study.

Evaluation metrics	Equation	Value range	Ideal value
Bias	$Bias = \frac{1}{T}\sum_{t=1}^{T}(\hat{y}_t - y_t)$	(-∞, +∞)	0
MAE	$MAE = \frac{1}{T}\sum_{t=1}^{T}\lvert\hat{y}_t - y_t\rvert$	[0, +∞)	0
RMSE	$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}$	[0, +∞)	0

Note: y_t and $\hat{y}_t$ denote the observed and simulated (or ensemble-estimated) values of a given annual extreme precipitation index in year t, respectively. T is the number of years in the evaluation period. y and $\hat{y}$ represent the corresponding observed and simulated (or ensemble-estimated) annual index series. Bias retains its sign, while $|Bias|$ was used for ranking-based analyses.

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

3.5.2 Independent validation of the QDM method

To evaluate the effectiveness of QDM, the transfer function was established using observed and simulated daily precipitation from the training period and then applied to correct the raw model simulations over the independent validation period.

During the validation period, annual extreme precipitation indices were calculated separately from the raw simulations, QDM-corrected simulations and observations. Their distributions were compared using cumulative distribution functions (CDFs) to assess whether QDM improved the representation of extreme precipitation index distributions. In addition, the correction performance of QDM for each index was quantitatively evaluated using the improvement rate (IR) of three evaluation metrics: Bias, MAE, and RMSE. The improvement rate was defined as:

$$IR(\%) = \frac{|E_{raw}| - |E_{qdm}|}{|E_{raw}|} \times 100 . \quad (10)$$

here, E_{raw} and E_{qdm} denote the evaluation metric values (any of the three metrics: Bias, MAE and RMSE) discrepancy measures between the observed annual extreme precipitation index series and the corresponding series derived from the raw and QDM-corrected simulations, respectively. Positive IR values indicate that QDM-corrected simulations reduced the simulation discrepancy relative to the raw models and improved model performance, whereas negative values indicate that the discrepancy increased after correction.

3.5.3 Independent validation of individual models and MME schemes

To evaluate the performance of individual models and MME schemes, a comprehensive comparison was conducted during the independent validation period based on the three evaluation metrics (Table 3). For each index-metric combination, the raw error values derived from the 22 individual models and MME schemes were normalized utilizing their corresponding min-max error range observed within the validation period:

$$z_{ij} = \frac{d_{j,max} - d_{ij}}{d_{j,max} - d_{j,min}} . \quad (11)$$

where d_{ij} denotes the raw error value of model or ensemble scheme i for the j-th index-metric combination. $d_{j,min}$ and $d_{j,max}$ are the minimum and maximum error values for the j-th index-metric combination, respectively, which are determined collectively across all 22 individual models and the two MME schemes within the validation period. Through this formulation, the raw cost-type errors are linearly mapped onto a [0, 1] scale, where a larger z_{ij} indicates better performance.

For each model or ensemble scheme, the composite performance score was then calculated as the arithmetic mean of the normalized utility values across all extreme precipitation index-metric combinations:

$$S_i = \frac{1}{J} \sum_{j=1}^{J} z_{ij} . \quad (12)$$

where S_i is the composite performance score for model or ensemble scheme i, z_{ij} is the normalized utility of the i-th model or ensemble scheme for the j-th index-metric combination, and J is the total number of index-metric combinations.

3.6 Future climate change assessment of extreme precipitation

To assess the future trends in extreme precipitation in Bologna and their statistical significance, Sen's slope was used to estimate the trend magnitude, and the Mann–Kendall test was applied to evaluate significance. Trend was estimated for both training and validation periods. In addition, relative changes in future extreme precipitation indices were further calculated with respect to the baseline period. 1985–2014 that represents the current climatic state. The future period was divided into three stages: near future (NF, 2015–2040), mid-future (MF, 2041–2070), and far future (FF, 2071–2100). This division was used to evaluate relative changes in extreme precipitation indices across different future stages. For each extreme precipitation index, future emission scenario, future period and ensemble method, the mean value during the future period was compared with the corresponding mean value during the baseline period. This allowed the direction and magnitude of changes in extreme precipitation characteristics to be quantified relative to the baseline climate. Finally, to assess the robustness of the projected relative changes, the BMA and SMA results were compared with the spread of the 22 individual CMIP6 models. Model sign agreement was used to evaluate whether most models supported the same direction of change, while inter-model spread was used to represent uncertainty in the magnitude of change. The difference between BMA and SMA was further examined to assess whether the projected relative changes were sensitive to the choice of ensemble method.

4 Results

4.1 Reliability of historical simulation for the Bologna time series

Figure 3 shows the empirical CDFs of Rx1day from the 22 models during the independent validation period (1950-2014), comparing observations with raw and QDM-corrected model simulations. Before bias correction, except for ACCESS-CM2, the raw simulation curves of most models are shifted to the left of the observed curve, indicating that the uncorrected models generally underestimate the intensity of Rx1day. This underestimation is particularly evident in the upper tail of the distribution, where many raw simulations reach high cumulative probabilities at relatively low Rx1day values. This suggests that the raw simulations have limited ability to represent high-intensity extreme precipitation events. After QDM correction, the distributions of most models move closer to the observations, indicating an improved representation of the Rx1day distribution and a reduction in the systematic underestimation present in the raw simulations. Although some models, including EC-Earth3-Veg-LR and TaiESM1, still exhibit discrepancies from the observations in the upper tail, the QDM-corrected simulations generally outperform the raw simulations.

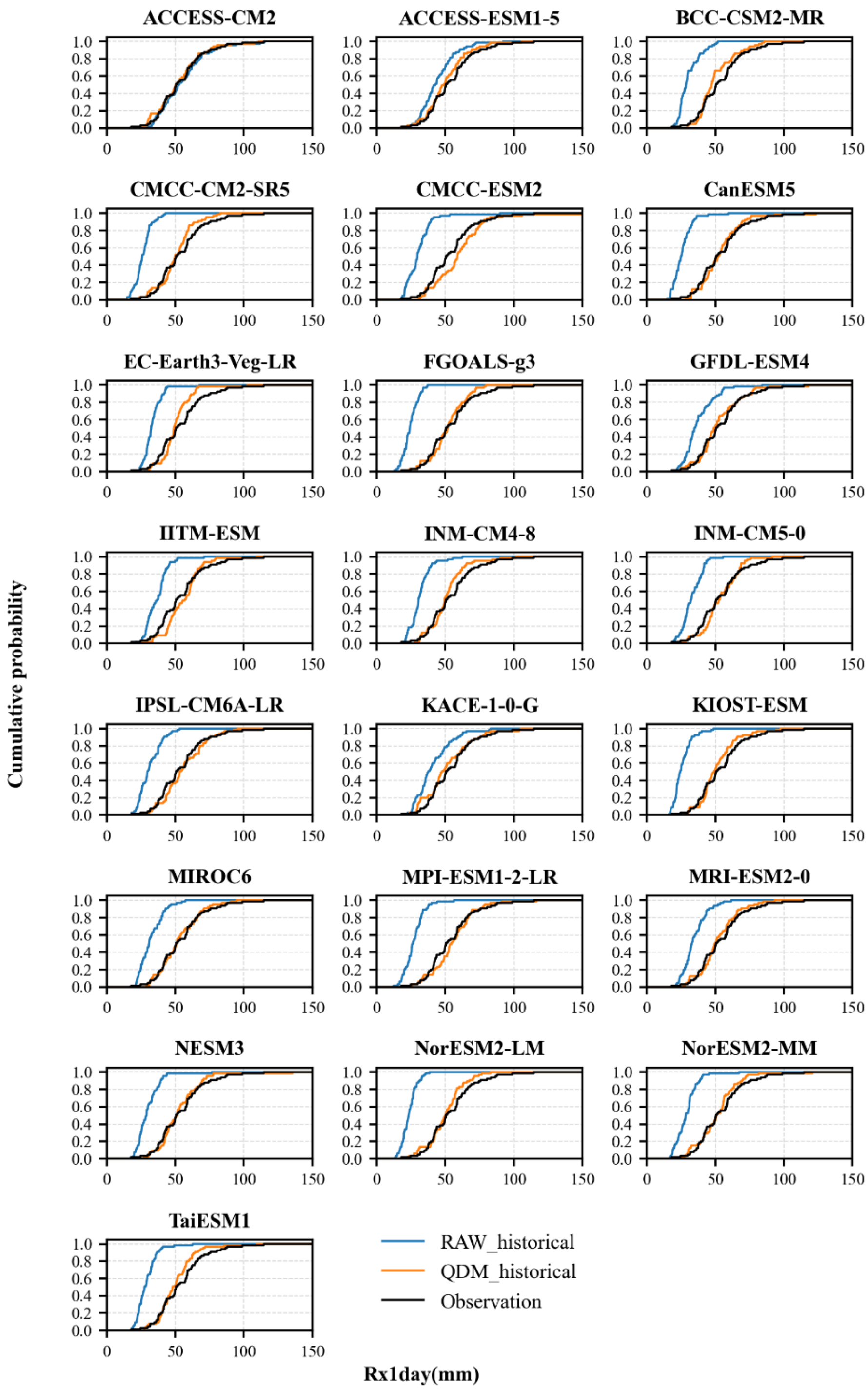
Figures S1-S3 in the supplementary material further present the empirical CDFs of R99p, R10mm and R20mm during the independent validation period. Compared with the raw simulations, the QDM-corrected distributions are generally closer to the observations, indicating that QDM also improves the distributional representation of other extreme precipitation indices. For R99p, some CDFs reach relatively high cumulative probabilities at $x = 0$. This occurs because, in some years, no daily

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

precipitation exceeds the 99th percentile threshold of the reference period, resulting in an annual R99p value of zero. Therefore, the cumulative probability at this point can be interpreted as the proportion of years with R99p equal to zero.

Figure 4 presents the improvement rates of the QDM-corrected simulations relative to the raw simulations across multiple evaluation metrics. The median line in each boxplot represents the typical correction effect across the 22 models, while the box spans the interquartile range (25th - 75th percentile). The scatter points show the performance of individual models, and the half-violin plots illustrate the distributional shape of improvement rates. Therefore, the interpretation mainly focuses on the median improvement rate and the position of the box. For most metrics, the median improvement rates are greater than zero, and the boxes are mainly located in the positive range, indicating that QDM generally reduces the discrepancies between the raw simulations and observations during the independent validation period. The four extreme precipitation indices show clear improvements in Bias, with most median improvement rates close to or greater than 50%. These results suggest that QDM performs well in correcting systematic bias. In contrast, the improvements in MAE and RMSE are relatively weaker. For R99p, the median improvement rates are below zero and the corresponding boxes are mainly located in the negative range, indicating that QDM does not consistently improve the year-to-year error magnitude for this index. For Rx1day, R10mm and R20mm, the median improvement rates for MAE and RMSE are positive but generally below 50%.

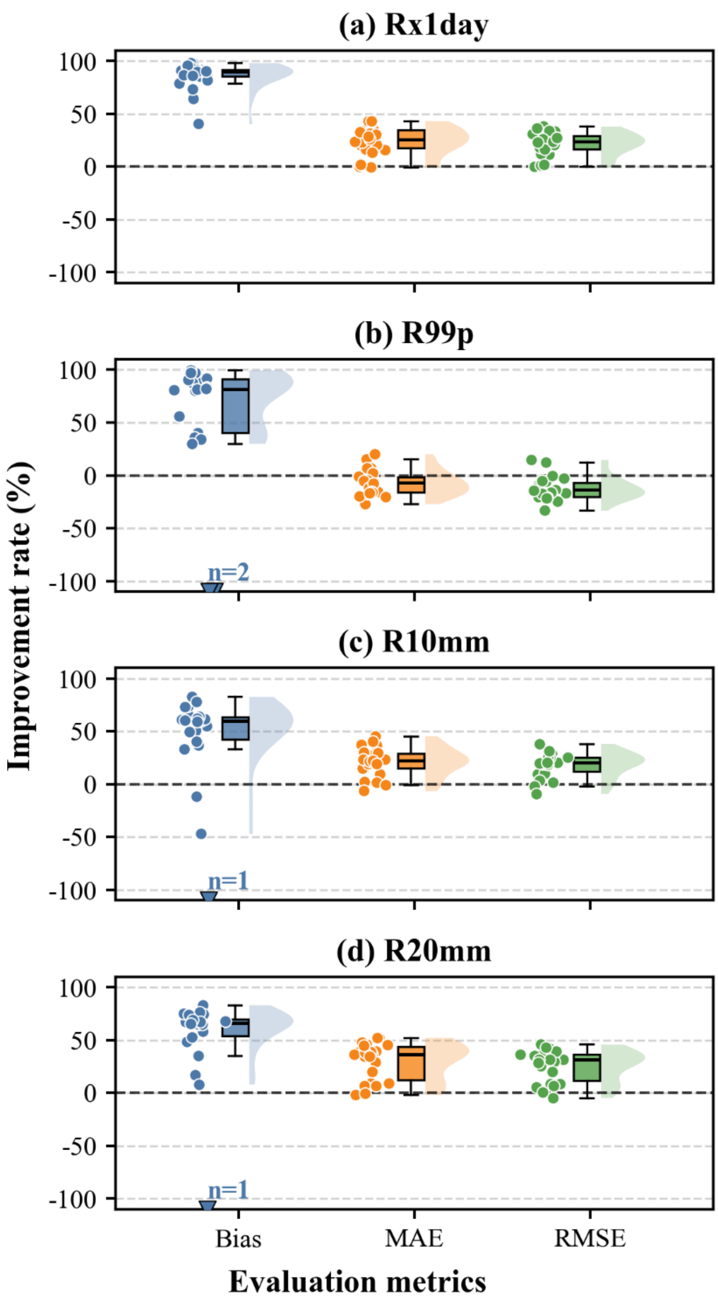
Overall, the independent validation results indicate that QDM improves the ability of CMIP6 models to reproduce the statistical and distributional characteristics of extreme precipitation indices. Although the improvements are not uniform across all indices and metrics, particularly for the year-to-year error magnitude of R99p, QDM substantially reduces systematic bias for most models. Therefore, QDM was deemed appropriate for generating the final bias-corrected historical and future daily precipitation series used in the subsequent MME construction and future projections.



340

Figure 3: Empirical cumulative distribution functions of observed Rx1day and raw and QDM-corrected Rx1day from 22 CMIP6 models during the validation period.

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

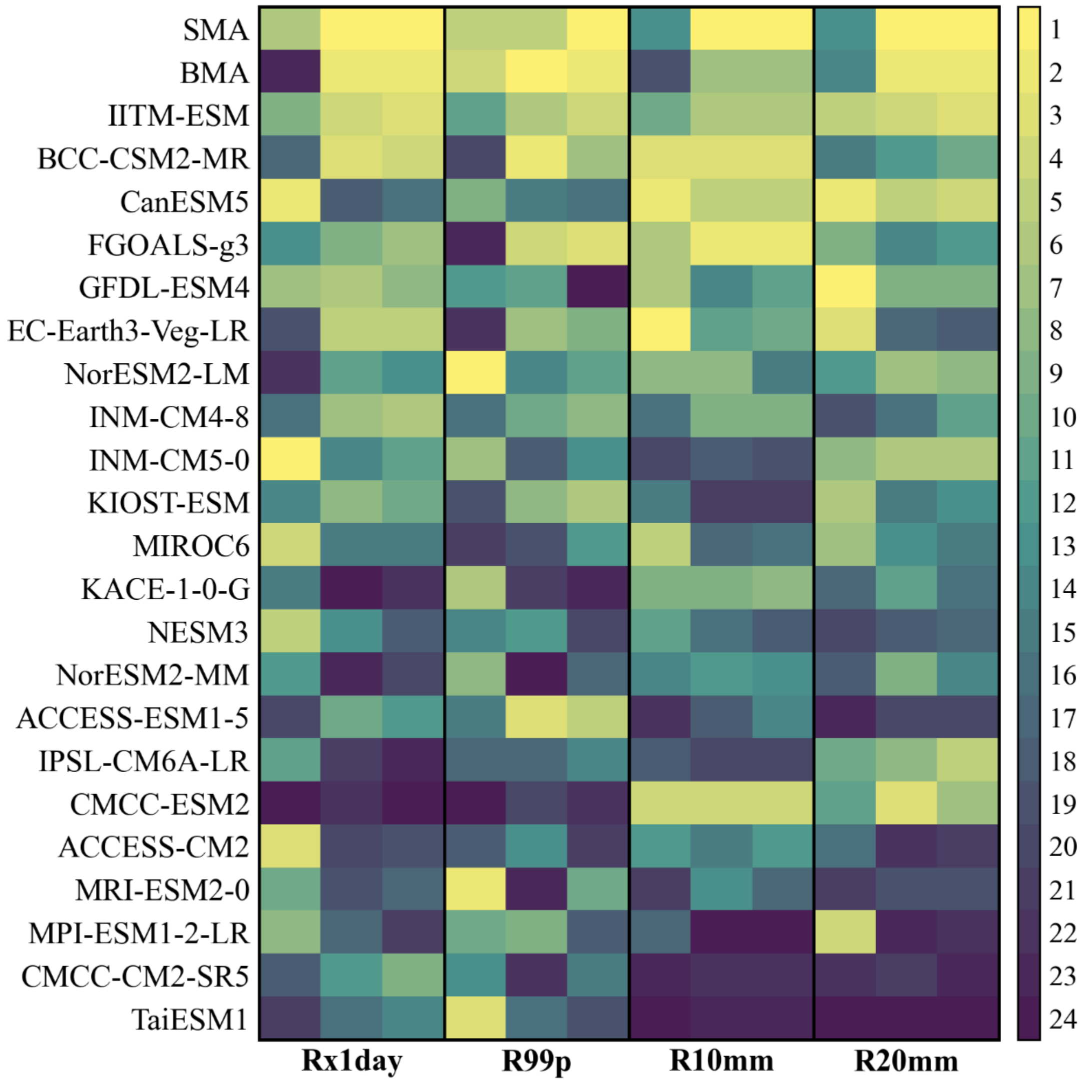


345 **Figure 4: Improvement rates of QDM bias correction for extreme precipitation indices based on three evaluation metrics. Circles represent individual CMIP6 models. Box plots show the median and interquartile range, calculated using all improvement rates including values outside the displayed range. Half-violin plots show kernel density estimates of model improvement rates within the displayed range (-100% to 100%) to avoid distortion by extreme outliers. Downward triangles indicate values below the lower bound of the displayed range (-100%), with n denoting the number of clipped model values.**

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

4.2 Independent validation of individual models and MME schemes

Figure 5 presents the comprehensive performance evaluation of the 22 bias-corrected individual CMIP6 models and the two MME schemes (SMA and BMA) over the independent validation period (1950-2014) across the four extreme precipitation indices. In terms of composite performance score, SMA and BMA rank among the best-performing methods, indicating that the MME schemes provide more stable overall performance than most individual models. However, certain individual models outperform the ensemble schemes for specific metrics or indices. For example, CanESM5 exceeds both the MME schemes in $|Bias|$ for most indices, and INM-CM5-0 achieves the best $|Bias|$ performance for Rx1day. These results are consistent with those of Hagedorn et al. (2005), who found that single models can occasionally outperform MMEs for particular variables. Nevertheless, when evaluated across multiple indices and metrics, the ensemble schemes generally hold an advantage.



https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

Figure 5: Comprehensive performance evaluation of 22 models and multi-model ensembles in Bologna during independent validation period (1950-2014). The metrics are $|Bias|$ (left-hand columns), mean absolute error (center columns) and the root-mean-square error (right-hand columns). Each cell shows the metric-specific rank, with lower ranks indicating better performance.

Table 4 presents $|Bias|$, MAE and RMSE for SMA and BMA during the independent validation period, where smaller values indicate better performance. SMA outperforms BMA for Rx1day, R10mm and R20mm across all three metrics. For R99p, BMA achieves lower $|Bias|$ and MAE, though SMA retains a marginal advantage in RMSE. Overall, SMA shows more consistent performance across most indices and metrics, while BMA's relative advantage is confined to mean bias and absolute error for the most extreme precipitation index (R99p). Given these complementary performance characteristics, both MME schemes are carried forward into the projection analysis, with the spread between them serving as one source of ensemble uncertainty.

Table 4. Performance comparison between SMA and BMA during independent validation period (1950-2014).

Index	MME	$\|Bias\|$	MAE	RMSE
Rx1day	BMA	5.707	14.962	18.623
	SMA	1.368	13.688	17.515
R99p	BMA	0.505	47.327	56.263
	SMA	0.593	49.897	55.389
R10mm	BMA	4.259	6.944	8.840
	SMA	3.293	6.202	7.690
R20mm	BMA	2.095	3.461	4.291
	SMA	2.089	3.303	4.190

To further examine whether the relative performance of SMA and BMA is sensitive to the choice of training-validation split, sensitivity analyses were conducted using different temporal partitioning schemes (Table S1 in the supplementary material). The overall performance patterns remain broadly consistent across splits: SMA outperforms BMA for Rx1day, R10mm and R20mm across all metrics and all split years. For R99p, BMA shows a consistent advantage in MAE across all splits. However, BMA's $|Bias|$ advantage for R99p diminishes with later split years and reverse under the 1959 and 1969 splits, suggesting some dependence on training period length. Overall, the complementary performance characteristics between the two ensemble schemes are largely robust to the choice of split year, with the exception of BMA's $|Bias|$ performance for R99p.

4.3 Future changes

4.3.1 Future extreme precipitation projections

Figure 6 shows the BMA and SMA time series of the four extreme precipitation indices for the historical simulations and future projections under SSP1-2.6, SSP2-4.5 and SSP5-8.5. Corresponding Sen's slope estimates and Mann-Kendall trend test results for the observations and both ensemble schemes are summarized in Table 5.

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

Over the entire historical period (1850-2014), observed trends in all four extreme precipitation indices are weak and statistically insignificant, suggesting that long-term linear changes in historical extreme precipitation in Bologna were limited. In contrast, both ensemble schemes produce statistically significant trends for R99p, with BMA showing a significant negative trend and SMA a significant positive trend. BMA further departs from the observations by producing a statistically significant negative trend in R10mm. When the historical period is divided into sub-periods, the observations reveal significant decreasing trends in R10mm and R20mm during the early industrialization period (EI), whereas neither BMA nor SMA reproduces significant trends for these indices. During the RI, both ensemble schemes show closer overall agreement with the observations in terms of trend direction for most indices. The discrepancies between the ensemble results and the observations across these periods are not unexpected. Multi-model ensemble averaging suppresses the contribution of internal climate variability, such that the ensemble mean primarily reflects the externally forced response, while the observed record represents a single realization of the climate system in which internal variability remains fully expressed (Deser, 2020). Since the effect of internal climate variability on precipitation intensifies at finer spatial scales (Donat et al., 2016; Rampal et al., 2025), discrepancies of this kind are to be expected at the station level.

Projected trends during the future period (2015-2100) differ clearly across emission scenarios. Under SSP1-2.6, all four indices remain statistically insignificant in both BMA and SMA, consistent with the limited forced response expected under a low-emission pathway (IPCC, 2021). Under SSP2-4.5, Rx1day, R99p and R20mm trend positively in both ensemble schemes. In SMA, all three reach statistical significance, whereas the corresponding BMA trends, though positive, do not. The signal strengthens under SSP5-8.5, particularly for intensity-related indices: R99p is significant in both BMA and SMA, and Rx1day reaches high significance in SMA. R10mm, by contrast, trends negatively under this scenario, with SMA reaching statistical significance and BMA showing the same direction but falling short. This suggests a potential, if not yet robust, divergence between intensifying extreme precipitation and the frequency of days exceeding the moderate 10mm threshold.

Across scenarios, BMA and SMA show broadly consistent directional responses. Rx1day and R99p both trend positively under SSP2-4.5 and SSP5-8.5 in either scheme. Differences between the two methods lie mainly in magnitude and statistical significance. SMA tends to produce equal or larger trend magnitudes in most cases and more frequently achieves statistical significance, particularly under the medium-emission and high-emission scenarios. The intensification signal is most robust for R99p under SSP5-8.5, where statistically significant positive trends emerge in both BMA and SMA. Frequency-related indices, by contrast, show weaker and less coherent signals. R10mm trends negatively under SSP5-8.5, reaching statistical significance in SMA but not in BMA, while R20mm shows positive directional agreement between the two schemes, though only SMA reaches significance under SSP2-4.5.

Overall, future changes in extreme precipitation at Bologna are clearly scenario-dependent, with the intensification signal strengthening progressively from SSP1-2.6 to SSP5-8.5. Frequency-related indices, however, respond less coherently across scenarios and ensemble schemes. These results suggest that projected increases in extreme precipitation risk at this station are more likely to manifest through greater event intensity than through a broad rise in the number of precipitation days.

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

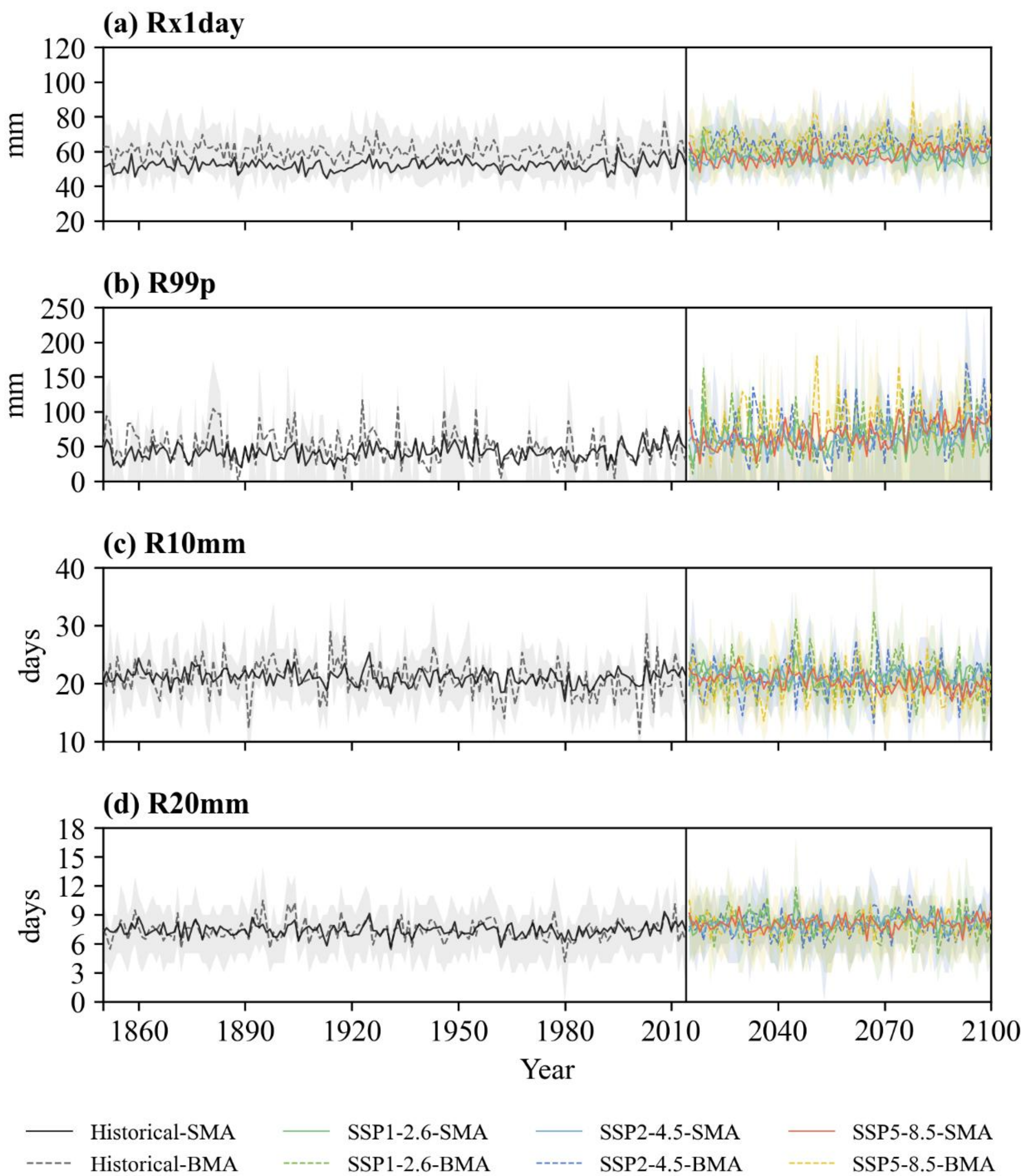


Figure 6: Time series of extreme precipitation indices derived from BMA and SMA for the historical simulation and future projections under SSP1-2.6, SSP2-4.5 and SSP5-8.5 after QDM bias correction. The shaded areas indicate the interquartile range (25th – 75th percentiles) associated with BMA for each period and scenario, representing the uncertainty range. The black vertical line marks 2014, separating the historical and future periods.

Table 5. Historical and future trends of extreme precipitation indices.

Scenario	Period	Scheme	Rx1day	R99p	R10mm	R20mm

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

Historical	Overall	OBS	-0.024	0.000	0.013	0.010
		BMA	-0.007	**-0.084***	**-0.014****	0.000
		SMA	0.010	**0.046***	-0.003	0.000
	EI	OBS	-0.117	0.000	**-0.047***	**-0.027***
		BMA	-0.017	-0.090	0.001	0.002
		SMA	0.014	0.046	-0.003	0.000
	RI	OBS	0.091	0.000	0.100	0.047
		BMA	0.012	-0.002	-0.019	0.007
		SMA	0.025	**0.152***	0.013	0.008
SSP1-2.6	Overall	BMA	-0.028	0.169	-0.020	-0.005
		SMA	-0.003	0.099	-0.008	-0.001
SSP2-4.5	Overall	BMA	0.036	0.223	0.002	0.002
		SMA	**0.049****	**0.181****	-0.005	**0.010*****
SSP5-8.5	Overall	BMA	0.051	**0.358****	-0.020	0.001
		SMA	**0.093*****	**0.423*****	**-0.031*****	0.006

Note: OBS represents observation data, EI denotes the early industrialization period (1850-1949), and RI denotes the rapid industrialization period (1950-2014). Trend significance was tested using the Mann–Kendall test. *, **, and *** denote significance at the 0.05, 0.01, and 0.001 levels, respectively.

4.3.2 Relative changes and robustness of projected signals

Figure 7 shows the projected relative changes in the four extreme precipitation indices across different future periods and emission scenarios, relative to the historical baseline (1985-2014). To assess the robustness of these projections, we analyse the individual-model changes across all 22 models, by considering sign consistency, inter-model spread, and ensemble-method sensitivity (Table 6).

Based on BMA and SMA projections, the four extreme precipitation indices show markedly different future responses. The intensity-related indices, Rx1day and R99p, show more consistent projected increases across scenarios and periods. Rx1day is projected to increase across all period-scenario combinations, with changes generally remaining below 10%. The most pronounced increase occurs in the FF period under SSP5-8.5. R99p also shows positive changes throughout, with its largest projected increase likewise in the FF period under SSP5-8.5. By contrast, the frequency-related indices display more complex responses. R20mm increase in most periods-scenarios combinations, though the magnitudes are generally small and remain below 10% in most cases. R10mm shows a clear dependence on both period and emission scenario. Changes are mostly small and positive under SSP1-2.6 and SSP2-4.5, but the signal weakens under SSP5-8.5 and turns negatively by the FF period. Overall, the intensification signal is more consistent and pronounced for the intensity-related indices than for the frequency-related ones, with the strongest projected changes occurring in the FF period under the high-emission scenario.

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

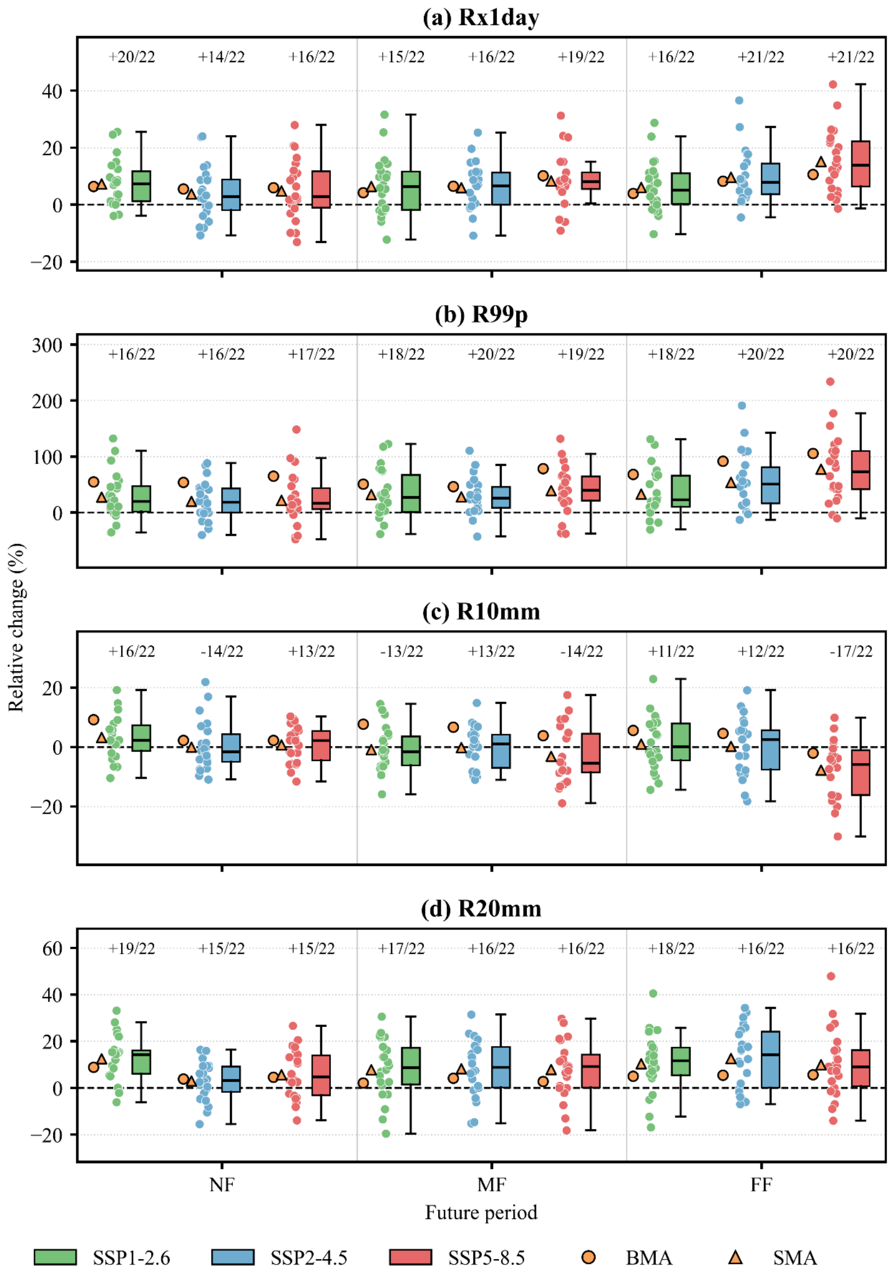
BMA and SMA generally project relative changes in the same direction across most indices, scenarios, and periods, particularly for Rx1day, R99p and R20mm, where both schemes consistently indicate positive changes. Differences arise, however, in the magnitude of those changes. For Rx1day, SMA tends to project somewhat larger increases in a number of period-scenario combinations, while for R99p, BMA yields notably higher estimates than SMA across all combinations. Overall, while the two schemes agree well on the direction of change, the projected magnitudes remain uncertain.

The direction of projected change is generally robust for both intensity-related indices. For Rx1day, the large majority of individual models project increases across all future periods and scenarios, with model agreement reaching +21/22 in the FF period under both SSP2-4.5 and SSP5-8.5. Both BMA and SMA project positive changes across all period-scenario combinations. Together, the high model agreement and the consistent sign of both ensemble projections indicate that the direction of Rx1day change is robust. Furthermore, since the difference in projected magnitude between the two schemes falls within the IQR of the individual model ensemble (Table 6), the magnitude estimates are also relatively insensitive to the choice of ensemble scheme. R99p similarly shows consistent positive projections from both BMA and SMA. Model agreement ranges from +16/22 to +17/22 in the NF period and reaches +18/22 to +20/22 in the MF and FF periods, supporting a robust increasing direction overall. However, ensemble-method sensitivity is relatively high for some NF combinations, particularly under SSP2-4.5 and SSP5-8.5, suggesting that while the direction of change is clear, the specific magnitude of R99p increase in the NF remains more sensitive to the choice of ensemble scheme.

The frequency-related indices display more complex and comparatively weaker responses. R20mm shows positive changes under both MMEs across all period-scenario combinations, though the magnitudes are generally small. In several period-scenario combinations, the IQR of the individual model projections is comparable to or exceeds the median projected change, suggesting that while a weak increasing tendency in R20mm cannot be ruled out, the projected magnitude carries considerable uncertainty. R10mm shows more pronounced dependence on both emission scenario and future period. Under SSP1-2.6, relative changes by both MMEs are small and consistently positive. Under SSP2-4.5, BMA also yields positive values, but SMA gives near-zero or slightly negative estimates, and model agreement is mixed, reaching -14/22 in the NF period, reflecting limited consensus on the direction of change. Under SSP5-8.5, the response evolves markedly across periods: BMA and SMA give opposite signs in the MF period, indicating sensitivity to the choice of ensemble scheme; by the FF period, both schemes project negative changes and model agreement reaches -17/22, pointing to a more consistent decreasing signal under the high-emission scenario. Overall, R10mm shows no sustained increasing signal across future scenarios and periods.

Overall, Figure 7 and Table 6 indicate that intensification of extreme precipitation at Bologna is a relatively robust projected signal, primarily reflected in increases in Rx1day and R99p. By contrast, the frequency-related indices show weaker responses with stronger dependence on emission scenario and future period. R20mm may exhibit a weak increasing tendency, though its projected magnitude remains uncertain, whereas R10mm shows no sustained increasing signal and is projected to decrease in the FF periods under the high-emission scenario. Taken together, future changes in extreme precipitation at Bologna appear more likely to manifest as greater precipitation intensity than as a consistent increase in the frequency of moderate-to-heavy precipitation days.

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026



475 **Figure 7: Relative changes in extreme precipitation indices during the NF, MF and FF periods relative to the historical baseline under the three emission scenarios: SSP1-2.6, SSP2-4.5 and SSP5-8.5. Dots and triangles indicate the relative changes estimates from the BMA and SMA schemes, respectively. The boxplots show the distribution of relative changes across the 22 individual models, with boxes representing the 25th to 75th percentile range and the horizontal line indicating the median value. Scatter points denote the relative change values of individual models. The labels above each group indicate the direction consistency across individual models, where '+n/22' and '-n/22' denote the number of models projecting positive and negative changes, respectively.**

480 **Table 6. Projected relative changes, inter-model spread, and ensemble-method sensitivity across future periods and emission scenarios.**

Index	Period	Scenario	BMA (%)	SMA (%)	Median (%)	IQR (%)	Ensemble-method sensitivity
Rx1day	NF	SSP1-2.6	6.39	7.30	7.30	10.55	0.09
		SSP2-4.5	5.57	3.83	2.85	10.81	0.16
		SSP5-8.5	5.99	4.88	2.82	12.76	0.09
	MF	SSP1-2.6	4.20	6.38	6.42	13.44	0.16
		SSP2-4.5	6.46	6.00	6.58	11.22	0.04
		SSP5-8.5	10.21	8.31	8.07	5.87	0.32
	FF	SSP1-2.6	3.94	6.08	5.17	10.74	0.20
		SSP2-4.5	8.27	9.62	7.82	10.81	0.12
		SSP5-8.5	10.60	15.13	13.89	15.82	0.29
R99p	NF	SSP1-2.6	55.11	27.82	20.33	45.66	0.60
		SSP2-4.5	54.12	20.02	18.64	43.54	0.78
		SSP5-8.5	65.01	22.05	16.43	37.21	1.15
	MF	SSP1-2.6	51.03	32.34	27.19	66.37	0.28
		SSP2-4.5	46.34	28.26	26.14	37.46	0.48
		SSP5-8.5	78.73	39.46	40.04	43.00	0.91
	FF	SSP1-2.6	68.32	33.04	22.70	55.60	0.63
		SSP2-4.5	91.83	54.18	51.09	64.36	0.58
		SSP5-8.5	105.58	77.60	72.64	68.00	0.41
R10mm	NF	SSP1-2.6	9.25	3.24	2.36	8.71	0.69
		SSP2-4.5	2.27	-0.03	-1.61	9.36	0.25
		SSP5-8.5	2.26	0.75	2.19	9.92	0.15
	MF	SSP1-2.6	7.72	-0.85	-1.55	9.72	0.88
		SSP2-4.5	6.69	-0.15	1.11	11.22	0.61
		SSP5-8.5	3.86	-3.14	-5.36	12.98	0.54
	FF	SSP1-2.6	5.62	1.00	0.13	12.41	0.37
		SSP2-4.5	4.61	0.21	2.64	13.27	0.33
		SSP5-8.5	-2.07	-7.75	-5.82	15.18	0.37
R20mm	NF	SSP1-2.6	8.93	12.54	14.34	9.97	0.36

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

	SSP2-4.5	3.87	3.02	3.18	10.92	0.08
	SSP5-8.5	4.64	5.70	4.76	17.15	0.06
MF	SSP1-2.6	2.19	7.77	8.72	15.69	0.36
	SSP2-4.5	4.19	8.25	8.91	17.41	0.23
	SSP5-8.5	2.78	7.93	9.17	14.18	0.36
FF	SSP1-2.6	4.98	10.35	11.62	11.86	0.45
	SSP2-4.5	5.51	12.65	14.24	24.02	0.30
	SSP5-8.5	5.70	9.97	9.13	15.59	0.27

Note: BMA and SMA denote the relative changes projected by the two MME schemes, respectively. Median (%) and IQR (%) are calculated from the relative changes of the 22 individual models, while IQR denotes the interquartile range and is calculated as $Q_{75} - Q_{25}$. Ensemble-method sensitivity is defined as $|BMA - SMA|/IQR$, where lower values indicate weaker sensitivity of the projected magnitude to the choice of ensemble method.

5 Conclusions

This study evaluated the ability of 22 CMIP6 global climate models and two MME schemes, Bayesian Model Averaging (BMA) and Simple Model Averaging (SMA), to simulate and project extreme precipitation at Bologna station. First, the effectiveness of Quantile Delta Mapping (QDM) bias correction in improving station-scale daily precipitation simulations was evaluated against an independent validation period. The ability of the bias-corrected models to reproduce historical extreme precipitation indices was then assessed. On this basis, BMA weights were estimated over the training period, and the performance of the 22 individual models, BMA, and SMA was compared over the independent validation period. For the projection stage, the QDM transferred functions and BMA weights estimated during the training period were transferred directly to future simulations, and both MME schemes were applied under SSP1-2.6, SSP2-4.5, and SSP5-8.5 to assess future changes in extreme precipitation indices and the robustness of the projected signals.

The results show that QDM substantially improves the ability of CMIP6 models to reproduce the distributional characteristics of historical extreme precipitation, with the bias-corrected simulations generally showing closer agreement with observations. Both BMA and SMA outperform most individual models in independent validation period, suggesting that ensemble approaches help mitigate the influence of individual model biases and yield more stable results. The two schemes agree well on the direction of projected future changes, though discrepancies in projected magnitudes remain, indicating that the direction of change is relatively robust while specific magnitudes carry greater uncertainty.

Future projections indicate that changes in extreme precipitation at Bologna are more robustly characterized by an intensification of precipitation intensity than by a consistent increase in frequency. Rx1day and R99p increase across most future periods and emission scenarios, with more pronounced signals in the far future period under high-emission conditions, pointing to a potential increase in the risk of intense extreme precipitation events. By contrast, the frequency-related indices show more complex responses. R20mm exhibits a weak increasing tendency, whereas R10mm shows no sustained increasing signal and is projected to decline in the mid-future and far future periods under SSP5-8.5. These results suggest that future intensification of extreme precipitation at Bologna is more likely to manifest as stronger individual heavy rainfall events than as a broadening in the frequency of moderate-to-heavy precipitation days.

This study demonstrates that combining QDM bias correction, MME methods, and robustness assessment provides a more robust framework for interpreting future changes in station-scale extreme precipitation. Nevertheless, inter-model spread, differences between ensemble methods, and the complex responses of frequency-related indices indicate that considerable uncertainty remains in local extreme precipitation projections. Future research could extend this approach to additional stations and regional scales, and incorporate process-based diagnostics to deepen understanding of the mechanisms driving extreme precipitation changes.

Code and data availability

The historical rainfall series observed in Bologna can be obtained from https://climexp.knmi.nl/ (KNMI and WMO, 2022). All the CMIP6 GCM outputs used in this study are publicly available through the Earth System Grid Federation (ESGF, https://esgf-node.llnl.gov/projects/cmip6/).

Supplement link

The link to the supplement will be included by Copernicus, if applicable.

Author contributions

AM proposed the main research question and supervised the work. YL conducted the analysis, produced the figures, and prepared the paper. RG provided constructive suggestions and contributed to the detailed revision of the manuscript. All authors provided feedback to the paper.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Disclaimer

Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher."

Acknowledgements

We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modelling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. We thank the European Climate Assessment & Dataset (ECA&D) project and the data-providing institutions for providing free access to the daily precipitation observations for Bologna used in this study. During the preparation of this manuscript, the authors used ChatGPT to enhance grammar and syntax and to refine sentence structure. All the content, ideas, and conclusions are the author's own.

Financial support

Y.L. was supported by the China Scholarship Council (CSC) Scholarship, no. 202406510005. A.M. was supported by the Italian Science Fund through the project "Stochastic amplification of climate change into floods and droughts change ($CO_2$2Water)", grant number J53C23003860001.

Review statement

The review statement will be added by Copernicus Publications listing the handling editor as well as all contributing referees according to their status anonymous or identified.

References

Abdelmoaty, H. M., Papalexiou, S. M., Rajulapati, C. R., and AghaKouchak, A.: Biases Beyond the Mean in CMIP6 Extreme Precipitation: A Global Investigation, Earth's Future, 9, e2021EF002196, https://doi.org/10.1029/2021EF002196, 2021.

Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., Lorenz, R., Pincus, R., and Schmidt, G. A.: ESD Reviews: Model dependence in multi-model climate ensembles: weighting, sub-selection and out-of-sample testing, Earth System Dynamics, 10, 91–105, https://doi.org/10.5194/esd-10-91-2019, 2019.

Alexander, L. V. and Arblaster, J. M.: Assessing trends in observed and modelled climate extremes over Australia in relation to future projections, Intl Journal of Climatology, 29, 417–435, https://doi.org/10.1002/joc.1730, 2009.

Allard, D., Vrac, M., François, B., and García De Cortázar-Atauri, I.: Assessing multivariate bias corrections of climate simulations on various impact models under climate change, https://doi.org/10.5194/hess-2024-102, 28 August 2024., 2024.

Argüeso, D., Marcos, M., and Amores, A.: Storm Daniel fueled by anomalously high sea surface temperatures in the Mediterranean, npj Clim Atmos Sci, 7, 307, https://doi.org/10.1038/s41612-024-00872-2, 2024.

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., and Knutti, R.: Reduced global warming from CMIP6 projections when weighting models by performance and independence, Earth System Dynamics, 11, 995–1012, https://doi.org/10.5194/esd-11-995-2020, 2020.

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.

Chae, S. T., Chung, E.-S., and Kim, D.: Evaluation of Optimized Multi-Model Ensembles for Extreme Precipitation Projection Considering Various Objective Functions, Water Resour Manage, 38, 5865–5883, https://doi.org/10.1007/s11269-024-03948-z, 2024.

Chen, S., Men, B., Pang, J., Bian, Z., and Wang, H.: Historical and projected extreme climate changes in the upper Yellow River Basin, China, Sci Rep, 15, 19061, https://doi.org/10.1038/s41598-025-99650-0, 2025.

Chen, W., Jiang, Z., and Li, L.: Probabilistic Projections of Climate Change over China under the SRES A1B Scenario Using 28 AOGCMs, J. Climate, 24, 4741–4756, https://doi.org/10.1175/2011JCLI4102.1, 2011.

Coles, S.: An Introduction to Statistical Modeling of Extreme Values, Springer, London, 2001.

Deser, C.: "Certain Uncertainty: The Role of Internal Climate Variability in Projections of Regional Climate Change and Risk Management," Earth's Future, 8, e2020EF001854, https://doi.org/10.1029/2020EF001854, 2020.

Donat, M. G., Lowry, A. L., Alexander, L. V., O'Gorman, P. A., and Maher, N.: More extreme precipitation in the world's dry and wet regions, Nature Clim Change, 6, 508–513, https://doi.org/10.1038/nclimate2941, 2016.

Dong, S., Xu, Y., Zhou, B., and Shi, Y.: Assessment of indices of temperature extremes simulated by multiple CMIP5 models over China, Adv. Atmos. Sci., 32, 1077–1091, https://doi.org/10.1007/s00376-015-4152-5, 2015.

Duan, Q., Ajami, N. K., Gao, X., and Sorooshian, S.: Multi-model ensemble hydrologic prediction using Bayesian model averaging, Advances in Water Resources, 30, 1371–1386, https://doi.org/10.1016/j.advwatres.2006.11.014, 2007.

Eden, J. M., Widmann, M., Grawe, D., and Rast, S.: Skill, Correction, and Downscaling of GCM-Simulated Precipitation, J. Climate, 25, 3970–3984, https://doi.org/10.1175/JCLI-D-11-00254.1, 2012.

ETCCDI, https://etccdi.pacificclimate.org/, last access: 26 November 2025, 2020., 2020.

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, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Faghih, M., Brissette, F., and Sabeti, P.: Impact of correcting sub-daily climate model biases for hydrological studies, Hydrol. Earth Syst. Sci., 26, 1545–1563, https://doi.org/10.5194/hess-26-1545-2022, 2022.

Fowler, H. J., Lenderink, G., Prein, A. F., Westra, S., Allan, R. P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Do, H. X., Guerreiro, S., Haerter, J. O., Kendon, E. J., Lewis, E., Schaer, C., Sharma, A., Villarini, G., Wasko, C., and Zhang, X.:

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

Anthropogenic intensification of short-duration rainfall extremes, Nat Rev Earth Environ, 2, 107–122, https://doi.org/10.1038/s43017-020-00128-6, 2021.

Fraley, C., Raftery, A. E., and Gneiting, T.: Calibrating Multimodel Forecast Ensembles with Exchangeable and Missing Members Using Bayesian Model Averaging, Monthly Weather Review, 138, 190–202, https://doi.org/10.1175/2009MWR3046.1, 2010.

Guo, R. and Montanari, A.: Historical rainfall data in northern Italy predict larger meteorological drought hazard than climate projections, Hydrology and Earth System Sciences, 27, 2847–2863, https://doi.org/10.5194/hess-27-2847-2023, 2023.

Hagedorn, R., Doblas-Reyes, F. J., and Palmer, T. N.: The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept, Tellus A: Dynamic Meteorology and Oceanography, 57, 219, https://doi.org/10.3402/tellusa.v57i3.14657, 2005.

Huang, T., Fu, S., Li, X., Dong, Y., Zhang, Y., and Sun, J.: Synoptic background conditions and moisture transport for producing the extreme heavy rainfall event in Valencia in 2024, Atmospheric and Oceanic Science Letters, 18, 100666, https://doi.org/10.1016/j.aosl.2025.100666, 2025.

Iacomino, C., Pascale, S., Zappa, G., Iotti, M., Grazzini, F., Ghinassi, P., and Portal, A.: A Classification of High-Risk Atmospheric Circulation Patterns for Italian Precipitation Extremes, https://doi.org/10.1002/joc.70118, 2025.

IPCC: Climate Change 2021: The Physical Science Basis, 1st ed., Cambridge University Press, https://doi.org/10.1017/9781009157896, 2021.

Jahanshahi, A., Booij, M. J., Patil, S. D., and Gupta, H.: An ensemble-based projection of future hydro-climatic extremes in Iran, Journal of Hydrology, 642, 131892, https://doi.org/10.1016/j.jhydrol.2024.131892, 2024.

Jiang, M., Felzer, B. S., and Sahagian, D.: Predictability of Precipitation Over the Conterminous U.S. Based on the CMIP5 Multi-Model Ensemble, Sci Rep, 6, 29962, https://doi.org/10.1038/srep29962, 2016.

Jiang, S., Ren, L., Hong, Y., Yong, B., Yang, X., Yuan, F., and Ma, M.: Comprehensive evaluation of multi-satellite precipitation products with a dense rain gauge network and optimally merging their simulated hydrological flows using the Bayesian model averaging method, Journal of Hydrology, 452–453, 213–225, https://doi.org/10.1016/j.jhydrol.2012.05.055, 2012.

Kim, Y.-H., Min, S.-K., Zhang, X., Sillmann, J., and Sandstad, M.: Evaluation of the CMIP6 multi-model ensemble for climate extreme indices, Weather and Climate Extremes, 29, 100269, https://doi.org/10.1016/j.wace.2020.100269, 2020.

Klein Tank, A. M. G., Wijngaard, J. B., Können, G. P., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pálsdóttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., van Engelen, A. F. V., Forland, E., Mietus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., Antonio López, J., Dahlström, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L. V., and Petrovic, P.: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment, International Journal of Climatology, 22, 1441–1453, https://doi.org/10.1002/joc.773, 2002.

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., and Eyring, V.: A climate model projection weighting scheme accounting for performance and interdependence, Geophysical Research Letters, 44, 1909–1918, https://doi.org/10.1002/2016GL072012, 2017.

Lei, X., Xu, C., Liu, F., Song, L., Cao, L., and Suo, N.: Evaluation of CMIP6 Models and Multi-Model Ensemble for Extreme Precipitation over Arid Central Asia, Remote Sensing, 15, 2376, https://doi.org/10.3390/rs15092376, 2023.

Li, L., Zheng, Z., Biederman, J. A., Xu, C., Xu, Z., Che, R., Wang, Y., Cui, X., and Hao, Y.: Ecological responses to heavy rainfall depend on seasonal timing and multi-year recurrence, New Phytologist, 223, 647–660, https://doi.org/10.1111/nph.15832, 2019.

Li, T., Jiang, Z., Treut, H. L., Li, L., Zhao, L., and Ge, L.: Machine learning to optimize climate projection over China with multi-model ensemble simulations, Environ. Res. Lett., 16, 094028, https://doi.org/10.1088/1748-9326/ac1d0c, 2021.

Marengo, J. A., Alves, L. M., Ambrizzi, T., Young, A., Barreto, N. J. C., and Ramos, A. M.: Trends in extreme rainfall and hydrogeometeorological disasters in the Metropolitan Area of São Paulo: a review, Annals of the New York Academy of Sciences, 1472, 5–20, https://doi.org/10.1111/nyas.14307, 2020.

McLachlan, G. J. and Krishnan, T.: The EM Algorithm and Extensions, Second Edition., John Wiley & Sons, https://doi.org/10.1002/9780470191613, 2007.

Orlowsky, B. and Seneviratne, S. I.: Global changes in extreme events: regional and seasonal dimension, Climatic Change, 110, 669–696, https://doi.org/10.1007/s10584-011-0122-9, 2012.

Padulano, R., Gomez-Mogollon, L. A., Napolitano, L., and Rianna, G.: Quantile-based bias-correction of extreme rainfall: Pros & cons of popular methods for climate signal preservation, Journal of Hydrology, 653, 132814, https://doi.org/10.1016/j.jhydrol.2025.132814, 2025.

Pall, P., Patricola, C. M., Wehner, M. F., Stone, D. A., Paciorek, C. J., and Collins, W. D.: Diagnosing conditional anthropogenic contributions to heavy Colorado rainfall in September 2013, Weather and Climate Extremes, 17, 1–6, https://doi.org/10.1016/j.wace.2017.03.004, 2017.

Papalexiou, S. M. and Montanari, A.: Global and Regional Increase of Precipitation Extremes Under Global Warming, Water Resources Research, 55, 4901–4914, https://doi.org/10.1029/2018WR024067, 2019.

Pathak, R., Dasari, H. P., Ashok, K., and Hoteit, I.: Effects of multi-observations uncertainty and models similarity on climate change projections, npj Clim Atmos Sci, 6, 144, https://doi.org/10.1038/s41612-023-00473-5, 2023.

Pavan, V. and Doblas-Reyes, F. J.: Multi-model seasonal hindcasts over the Euro-Atlantic: skill scores and dynamic features, Climate Dynamics, 16, 611–625, https://doi.org/10.1007/s003820000063, 2000.

Pellicone, G. and Caloiero, T.: Assessment and ranking of CMIP6-global climate models over the Calabria region (southern Italy), Journal of Hydrology, 661, 133655, https://doi.org/10.1016/j.jhydrol.2025.133655, 2025.

Qin, H., Yuan, W., Wang, J., Chen, Y., Dai, P., Sobel, A. H., Meng, Z., and Nie, J.: Climate change attribution of the 2021 Henan extreme precipitation: Impacts of convective organization, Sci. China Earth Sci., 65, 1837–1846, https://doi.org/10.1007/s11430-022-9953-0, 2022.

Qin, X. and Dai, C.: Comparison of different quantile delta mapping schemes in frequency analysis of precipitation extremes over mainland Southeast Asia under climate change, Journal of Hydrology, 606, 127421, https://doi.org/10.1016/j.jhydrol.2021.127421, 2022.

Qiu, J., Zhao, W., Brocca, L., and Tarolli, P.: Storm Daniel revealed the fragility of the Mediterranean region, TIG, 1, 100036, https://doi.org/10.59717/j.xinn-geo.2023.100036, 2023.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M.: Using Bayesian Model Averaging to Calibrate Forecast Ensembles, Monthly Weather Review, 133, 1155–1174, https://doi.org/10.1175/MWR2906.1, 2005.

Raghavan, S. V., Liu, J., Nguyen, N. S., Vu, M. T., and Liong, S.-Y.: Assessment of CMIP5 historical simulations of rainfall over Southeast Asia, Theor Appl Climatol, 132, 989–1002, https://doi.org/10.1007/s00704-017-2111-z, 2018.

Rampal, N., Gibson, P. B., Sherwood, S. C., Queen, L. E., Lewis, H., and Abramowitz, G.: Downscaling with AI reveals the large role of internal variability in fine-scale projections of climate extremes, https://doi.org/10.48550/arXiv.2507.06527, 9 July 2025., 2025.

Risser, M. D. and Wehner, M. F.: Attributable Human-Induced Changes in the Likelihood and Magnitude of the Observed Extreme Precipitation during Hurricane Harvey, Geophysical Research Letters, 44, https://doi.org/10.1002/2017GL075888, 2017.

Sloughter, J. M. L., Raftery, A. E., Gneiting, T., and Fraley, C.: Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging, Monthly Weather Review, 135, 3209–3220, https://doi.org/10.1175/MWR3441.1, 2007.

Song, Y. H., Chung, E.-S., and Shahid, S.: The New Bias Correction Method for Daily Extremes Precipitation over South Korea using CMIP6 GCMs, Water Resour Manage, 36, 5977–5997, https://doi.org/10.1007/s11269-022-03338-3, 2022.

Song, Y. H., Chung, E., and Shahid, S.: Global Future Climate Signal by Latitudes Using CMIP6 GCMs, Earth's Future, 12, e2022EF003183, https://doi.org/10.1029/2022EF003183, 2024.

Steffen, W., Broadgate, W., Deutsch, L., Gaffney, O., and Ludwig, C.: The trajectory of the Anthropocene: The Great Acceleration, The Anthropocene Review, 2, 81–98, https://doi.org/10.1177/2053019614564785, 2015.

Tan, Y., Guzman, S. M., Dong, Z., and Tan, L.: Selection of Effective GCM Bias Correction Methods and Evaluation of Hydrological Response under Future Climate Scenarios, Climate, 8, 108, https://doi.org/10.3390/cli8100108, 2020.

Tang, B., Hu, W., and Duan, A.: Assessment of Extreme Precipitation Indices over Indochina and South China in CMIP6 Models, Journal of Climate, 34, 7507–7524, https://doi.org/10.1175/JCLI-D-20-0948.1, 2021.

Thomas, N. P., Marquardt Collow, A. B., Bosilovich, M. G., and Dezfuli, A.: Effect of Baseline Period on Quantification of Climate Extremes Over the United States, Geophysical Research Letters, 50, e2023GL105204, https://doi.org/10.1029/2023GL105204, 2023.

Tong, Y., Gao, X., Han, Z., Xu, Y., Xu, Y., and Giorgi, F.: Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods, Clim Dyn, 57, 1425–1443, https://doi.org/10.1007/s00382-020-05447-4, 2021.

https://doi.org/10.5194/egusphere-2026-5518
Preprint. Discussion started: 22 September 2026

Varghese, S. J., Pentakota, S., Thadivalasa, P., Podapati, G., and Ashok, K.: Changes in physical characteristics of extreme rainfall events during the Indian summer monsoon based on downscaled and bias-corrected CMIP6 models, Sci Rep, 15, 3679, https://doi.org/10.1038/s41598-025-87949-x, 2025.

Wang, W., Lee, X., Xiao, W., Liu, S., Schultz, N., Wang, Y., Zhang, M., and Zhao, L.: Global lake evaporation accelerated by changes in surface energy allocation in a warmer climate, Nature Geosci, 11, 410–414, https://doi.org/10.1038/s41561-018-0114-8, 2018.

Wang, Y., Zhou, B., Qin, D., Wu, J., Gao, R., and Song, L.: Changes in mean and extreme temperature and precipitation over the arid region of northwestern China: Observation and projection, Adv. Atmos. Sci., 34, 289–305, https://doi.org/10.1007/s00376-016-6160-5, 2017.

Wei, L., Jiang, S., Dong, J., Ren, L., Liu, Y., Zhang, L., Wang, M., and Duan, Z.: Fusion of gauge-based, reanalysis, and satellite precipitation products using Bayesian model averaging approach: Determination of the influence of different input sources, Journal of Hydrology, 618, 129234, https://doi.org/10.1016/j.jhydrol.2023.129234, 2023.

World Meteorological Organization: Devastating rainfall in Asia claims hundreds of lives, https://wmo.int/media/news/devastating-rainfall-asia-claims-hundreds-of-lives, last access: 24 January 2026, 2025., 2025.

Xiao, Y., Yin, S., Yu, B., Fan, C., Wang, W., and Xie, Y.: The most extreme rainfall erosivity event ever recorded in China up to 2022: the 7.20 storm in Henan Province, Hydrol. Earth Syst. Sci., 27, 4563–4577, https://doi.org/10.5194/hess-27-4563-2023, 2023.

Yang, X., Yu, X., Wang, Y., He, X., Pan, M., Zhang, M., Liu, Y., Ren, L., and Sheffield, J.: The Optimal Multimodel Ensemble of Bias-Corrected CMIP5 Climate Models over China, Journal of Hydrometeorology, 21, 845–863, https://doi.org/10.1175/JHM-D-19-0141.1, 2020.

Zhu, H., Jiang, Z., Li, J., Li, W., Sun, C., and Li, L.: Does CMIP6 Inspire More Confidence in Simulating Climate Extremes over China?, Adv. Atmos. Sci., 37, 1119–1132, https://doi.org/10.1007/s00376-020-9289-1, 2020.

Zhu, H., Jiang, Z., and Li, L.: Projection of climate extremes in China, an incremental exercise from CMIP5 to CMIP6, Science Bulletin, 66, 2528–2537, https://doi.org/10.1016/j.scib.2021.07.026, 2021.

Zhu, H., Jiang, Z., Li, L., Li, W., Jiang, S., Zhou, P., Zhao, W., and Li, T.: Intercomparison of multi-model ensemble-processing strategies within a consistent framework for climate projection in China, Sci. China Earth Sci., 66, 2125–2141, https://doi.org/10.1007/s11430-022-1154-7, 2023.