Historical rainfall data in Northern Italy predict larger meteorological drought hazard than climate projections

Rui Guo and Alberto Montanari

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

Correspondence: Rui Guo (rui.guo2@unibo.it)

Abstract. Simulation of daily rainfall for the region of Bologna produced by 13 climate models for the period 1850 - 2100 are compared with the historical series of daily rainfall observed in Bologna for the period 1850 - 2014, and analysed to assess meteorological drought changes up to 2100. In particular, we focus on monthly and annual rainfall data, seasonality and drought events to derive information on the future development of critical events for water resources availability. The results prove that rainfall statistics, including seasonal patterns, are fairly well simulated by models, while the historical sequence of annual rainfall is not satisfactorily reproduced. In terms of show that historical data analysis under the assumption of stationarity provides more precautionary predictions for long term meteorological droughts, we conclude that historical data analysis under the assumption of stationarity may depict a more critical future with respect to climate model simulations, thereby outlining importantthat information integration is key to obtain technical indications.

1 Introduction

Droughts originate one of the most challenging risks for modern society. Indeed, most countries across the globe are exposed to medium/high drought risk (see Fig. 1). Recent events, like the Millenium Drought in Australia (Van Dijk et al., 2013) and droughts in California (Lund et al., 2018) and South Africa (Sousa et al., 2018) pointed out once again that drought events may occasionally last for several years, therefore originating "multiyear droughts", with 5-10 years duration, or "megadroughts", with duration longer than a decade (Cook et al., 2016). The sporadic occurrence of megadroughts across the globe is also confirmed by several paleoclimatic records (Cook et al., 2016; Vance et al., 2015). Multiyear droughts and megadroughts are a reason of concern as they strongly impact ecosystems, water supply, socio-economical assets and, ultimately, public health (Tabari et al., 2021; Ukkola et al., 2020; Tabari, 2020; Stahl et al., 2016). Also, the vulnerability and exposure to these events are much higher than in the past.

Multiyear droughts are rare extreme events, which are related to large scale atmospheric teleconnections whose dynamics are dictated by chaotic behaviors. An implication of the rare occurrence of multiyear droughts is the limited availability of data to decipher their frequency and train prediction models. For the same reason, it is also difficult to predict how climate change may exacerbate drought risk. Indeed, the warmer climate is changing the hydrological cycle and further affects precipitation (Trenberth, 2011), with evidence demonstrating that precipitation is altering in terms of both annual mean (Knutti and Sedláček, 2013), seasonal variation (Polade et al., 2014; Kumar et al., 2014a), and extreme events (Papalexiou and Montanari, 2019;

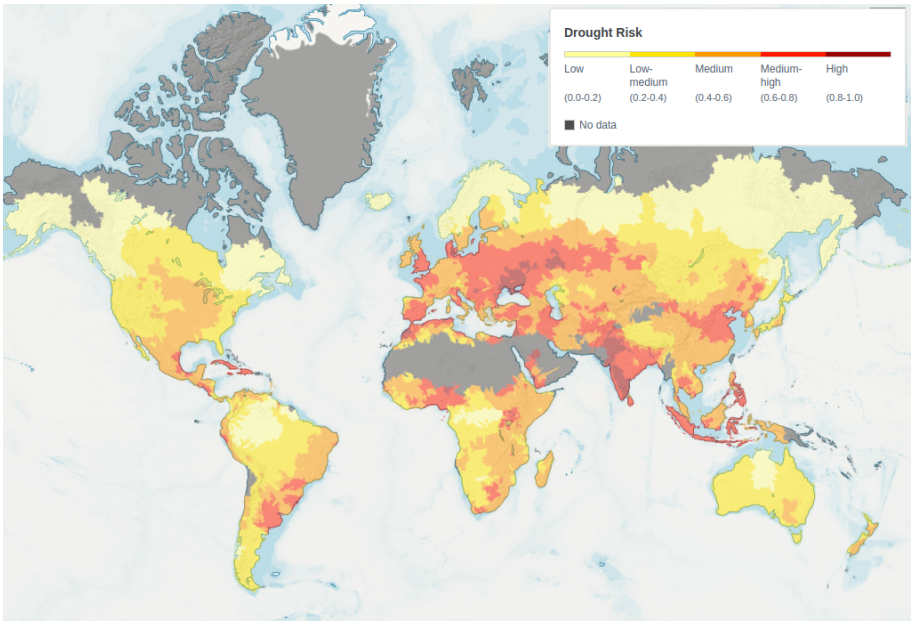

Figure 1. Drought risk index score by countries. It combines information on droughts hazard, exposure, and vulnerability. Higher values indicate higher risk of drought. Source: WRI Aqueduct, accessed on Aug 8, 2022 (aqueduct.wri.org).

Alexander et al., 2006). According to the Intergovernmental Panel on Climate Change Sixth Assessment Report (IPCC AR6), such changing conditions are likely to continue in the 21st century (IPCC, 2021). Therefore, the risk of multiyear droughts is likely to increase in the future, in view also of the increased water demands originated by global warming (Li et al., 2020).

The above reasoning highlights the key role of predictions in mitigating the risk of multiyear droughts and designing adaptation procedures. Future climatic scenarios are usually generated by General Circulation Models (GCMs), which attempt to simulate future climate variables under given scenarios of CO_2 emissions. Recently, the World Climate Research Programme (WCRP) released the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al. (2016)). Simulations from several models are available which deliver ensemble projections of future climate.

The reliability of rainfall simulated by GCMs has been analysed by several studies that focused primarily on the generation of climate models that preceded CMIP6 (CMIP5 and previous models; see Palerme et al., 2017; Koutroulis et al., 2016; Aloysius et al., 2016; Kumar et al., 2014b; Sillmann et al., 2013). Most of these researches compared the climatic scenarios with historical datasets produced by reanalysis such as NCEP or ERA-Interim (Palerme et al., 2017; Kumar et al., 2014b; Sillmann et al., 2013). However, on the one hand, reanalysis may introduce biases in the reproduction of extreme events. On the other hand, the observation period of these datasets covers only the recent past and therefore may undermine a statistical assessment of performances.

The present study aims at inspecting the ability of selected CMIP6 models in simulating regional scale climate by focusing on meteorological drought occurrence. We first compare simulated statistics with those of~~simulations with~~ one of the longest

daily rainfall records globally available: the Bologna rainfall series, whose observation period dates back to 1813. We decided to adopt as baseline an observed record, instead of a reanalysis, to take advantage from an extraordinary long observation period ~~and avoid any bias that may be potentially induced by spatial interpolation~~.

The purpose of our study is twofold: (i) to evaluate the ability of up-to-date GCMs in simulating the statistics of observed precipitation by focusing on multiyear meteorological droughts and (ii) to infer how precipitation and drought risk will change in the future. The paper is organized as follows: Section 2 provides detailed descriptions of the data used in this study. Section 3, 4 and 5 describe the methods for bias correction, reliability testing, and future change assessment, respectively. Section 6 presents the results of the historical evaluation and future projections for both precipitation and drought risk. ~~Section ?? discusses the results.~~ Finally, Section 7 summarises our conclusions.

2 Data

2.1 Rainfall observations~~Observation data~~

Italy is one of the first countries that started to systematically collect meteorological observations. Meteorological instruments and a network of observations were developed by Galileo's scholars and operated in the 17th century already. The rainfall series collected in Padua since 1725 is the longest daily record in the world, and five other rainfall stations have been continuously in operation – with few missing values - since the eighteenth century (Bologna, Milan, Rome, Palermo and Turin). Therefore, a data set of enormous value has been accumulated in Italy over the last three centuries (Brunetti et al., 2006).

Rainfall observation in Bologna at daily time scale dates back to 1714. Since 1813 the series is continuous. Brunetti et al. (2001) provided an interesting description of the history of the time series ~~and the procedures that were applied to detect and resolve apparent inconsistencies~~. The observatory was originally located in the center of the city. The raingauge was changed in 1857 and likely in 1900. After 1978 data were collected by another observatory in a nearby location. Brunetti et al. (2001) proposed monthly correction factors to resolve apparent underestimation of rainfall during 1813-1858, 1900-1928 and 1813-1978. These corrections are valid for monthly rainfall only.

The daily rainfall series observed in Bologna from 1813 to 2021 can be obtained from the European Climate Assessment-Daily Database (ECA & D) (https://climexp.knmi.nl/). ~~It is one of the longest daily rainfall records worldwide publicly available.~~ These are original data as reported in the transcripts of the observatory, without any correction. For the purpose of the present analysis, we assume that the daily series is homogeneous. In view of the time window adopted by CMIP6 GCMs for the historical reconstruction, we limit our analysis of the Bologna time series from 1850 to 2014 (see section 2.2).

Fig. 2 shows the daily time series and the 10-year moving average, which oscillates between a minimum of 1.2 mm and a maximum of 2.5 mm (more than twice the minimum). The above extreme values occurred in the 1820s and the decade ending in 1902, respectively. After 1950, a mild increasing trend is noticed, which mirrors a similar tendency that occurred from ~~about~~ the 1830s~~1830~~ to ~~about~~ the 1890s~~1890~~.

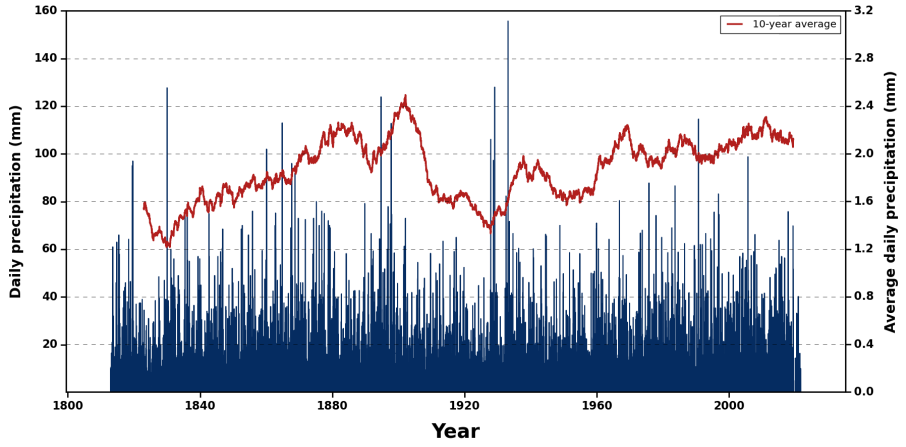

Figure 2. Daily rainfall records in Bologna during 1813-2021, with 10-year moving average (red line)

2.2 General Circulation Model simulation

GCM simulations for both historical (1850-2014) and future (2015-2100) periods are publicly available from the Copernicus database (https://cds.climate.copernicus.eu/). Future projections are obtained under the emission scenarios "Shared Socioeconomic Pathway" (SSP) 1-2.6 (SSP1-2.6), 2-4.5 (SSP2-4.5), and 5-8.5 (SSP5-8.5). These are the future emissions considered by the Scenario Model Intercomparison Project (ScenarioMIP) (O'Neill et al., 2016) of CMIP6.

From each model an ensemble simulation is generated for different initial conditions, initialization methods, physics versions, and forcing datasets. Similarly to Grose et al. (2020) and Kim et al. (2020), we analyse only the first run of each ensemble, that is identified by the acronym "r1i1p1f1", where "r", "i", "p" and "f" indicate initial conditions, initialization method, physics version and forcing data set, respectively (see https://pcmdi.llnl.gov/CMIP6/Guide/dataUsers.html for more details). We assume that analysing a single ensemble member per GCM model is sufficient to sample the ensemble for testing model bias. The assumption above was recently proved to be tenable by Longmate et al. (2022).

Table 1 shows detailed information for 13 GCMs used in this study. We select all the models providing future simulations for the three considered emission scenarios. The multi-model ensemble mean is the arithmetic average value of the 13 GCM outputs. ~~To compare the simulation with observed data, bilinear spatial interpolation is applied to estimate the model prediction for Bologna depending on the four nearest GCM grid points. Given that GCM simulations start from 1850, we adopt the time window 1850-2014 as our "historical period".~~

3 Bias correction

Simulations by GCM are provided at the grid scale. To compare them with observed data, one should take into account the potential bias that may be introduced by subgrid variability. For the considered time scale subgrid variability is expected to be

Table 1. Description of 13 GCMs from CMIP6 used in this study.

Model	Institution	Resolution	Grid (lon × lat)
CMCC-ESM2	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy	$1.25^\circ \times 0.94^\circ$	288×192
NorESM2-MM	Norwegian Climate Centre, Norway	$1.25^\circ \times 0.94^\circ$	288×192
CanESM5	Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change, Victoria, Canada	$2.81^\circ \times 2.81^\circ$	128×64
INM-CM5-0	Institute for Numerical Mathematics, Russian Academy of Science, Russia	$2.00^\circ \times 1.50^\circ$	180×120
INM-CM4-8	Institute for Numerical Mathematics, Russian Academy of Science, Russia	$2.00^\circ \times 1.50^\circ$	180×120
MPI-ESM1-2-LR	Max Planck Institute for Meteorology, Germany	$1.875^\circ \times 1.875^\circ$	192×96
MIROC6	Japan Agency for Marine-Earth Science and Technology, Japan	$1.40^\circ \times 1.40^\circ$	256×128
EC-Earth3-Veg-LR	Consortium of European Research Institution and Researchers, Europe	$1.125^\circ \times 1.125^\circ$	320×160
GFDL-ESM4	Geophysical Fluid Dynamics Laboratory, USA	$1.25^\circ \times 1.00^\circ$	360×180
MRI-ESM2-0	Meteorological Research Institute, Japan	$1.125^\circ \times 1.125^\circ$	320×160
ACCESS-CM2	Commonwealth Scientific and Industrial Research Organization- Australian Research Council Centre of Excellence for Climate System Science, Australia	$1.875^\circ \times 1.250^\circ$	192×144
FGOALS-g3	Institute of Atmospheric Physics, Chinese Academy of Sciences, China	$2.00^\circ \times 2.25^\circ$	180×80
IPSL-CM6A-LR	Institut Pierre Simon Laplace, Paris, France	$2.50^\circ \times 1.25^\circ$	144×143

limited in the region of Bologna. In fact, we focus on monthly and annual rainfall data, which exhibit low spatial variability in the region (see the annual reports of the Regional Agency for Environmental Protection at https://www.arpae.it; see also Antolini et al. (2016) for an analysis of subgrid variability in the considered spatial domain).

To compensate for potential bias, we applied bilinear spatial interpolation to estimate the model prediction for Bologna depending on the four nearest GCM grid points. Moreover, we applied quantile delta mapping (QDM) to correct bias with respect to the observed daily rainfall series.

QDM (Cannon et al., 2015) preserves model-projected relative change in quantiles, while at the same time correcting the systematic biases in quantiles of a model simulation compared to observed values. It is widely adopted for bias correction of GCM output such as precipitation (Xavier et al., 2022; Fauzi et al., 2020). Here, we apply the QDM to the model runs for the historical (1850-2014) and future periods (2015-2100).

First, we compute the empirical frequency of not exceedance $q_f(t)$ of each GCM simulated value $m_f(t)$ during the future (denoted by subscript f) period. Then, the relative change in quantiles $\Delta(t)$ of GCM simulated precipitation over two time periods is given by:

$$\Delta(t) = \frac{F_f^{-1}(q_f(t))}{F_h^{-1}(q_f(t))} = \frac{m_f(t)}{F_h^{-1}(q_f(t))}, \tag{1}$$

where F_f and F_h denote the distribution of empirical frequencies of each GCM simulated value $m_f(t)$ and $m_h(t)$ during the future and historical period, respectively. Then, the bias-corrected value $\hat{o}(t)$ of $q_f(t)$ is computed by the frequency of not exceedance F_o of the observations during the historical period:

$$\hat{o}(t) = F_o^{-1}(q_f(t)) \tag{2}$$

Finally, the bias-corrected GCM simulation $\hat{m}_f(t)$ for the future period is given by:

$$\hat{m}_f(t) = \hat{o}(t) \cdot \Delta(t). \tag{3}$$

Note that bias correction for the historical GCM simulation only needs the application of eq. (2) by plugging in $q_h(t)$ in the right hand side instead of $q_f(t)$.

4 Reliability testing

The reliability of GCM simulations with and without bias correction is evaluated by focusing on different temporal scales ~~and both low and high rainfall,~~ to obtain a comprehensive picture of model performances.

4.1 Monthly and annual rainfall~~Annual rainfall~~

~~To evaluate the performance of each of the considered CMIP6 GCMs in reproducing the behavior of annual rainfall we evaluate their capability to reproduce the observed time series~~~~hyetograph~~ ~~and its statistics during the historical period. The goodness of the time series~~~~hyetograph~~ ~~simulation is evaluated by computing the modified Kling-Gupta efficiency (KGE) (Kling et al., 2012), the mean absolute relative error (MARE), and the Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970). GCM historical simulations are divided into two periods: long-past (1850-1949) and near-past (1950-2014).~~

~~KGE is computed by combining three statistical parameters: correlation coefficient (r), bias ratio (β), and variability ratio (γ)). In eq. **??-??**, obs_i and GCM_i indicate annual rainfall at year i given by historical observations and the considered GCM simulation, respectively, μ_{obs} and μ_{GCM} represent the mean value along the considered period of observed data and GCM, respectively, and σ_{obs} and σ_{GCM} indicate the corresponding standard deviation. MARE and NSE are computed as follows:~~

~~As for the time series~~~~hyetograph~~ ~~statistics, we first make a visual comparison of the sample probability density of observed and simulated annual rainfall. Moreover, we compare the lag-1 autocorrelation coefficient of historical data versus each simulation.~~

To assess the performance of each of the considered CMIP6 GCMs in reproducing the statistics of monthly data during the historical period (1850-2014) we use the "Combined Probability-Probability" (CPP) plot (Koutsoyiannis and Montanari, 2022).

CPP plot compares the probability distributions F_o and F_h of observed and GCM simulated monthly rainfall, respectively, during the historical period. The comparison refers to 5 subsequent 33-year long time windows during 1850-2014 to assess the capability of GCMs to reproduce changes along the time of climate statistics.

To make the CPP plot, first a realisation w from uniform random variable F_w is generated in the range $[0,1]$ and then we compute the empirical probability distribution

$$\hat{F}_w(w) = F_h\left(F_o^{-1}(w)\right). \tag{4}$$

The ability of GCM in reproducing observed statistics is regarded as good if the plot $F_w(w)$ versus w is the equality line, that is, if $F_w(w) = w$. This is possible only if F_h is identical to F_o, which is what we would like to check. In essence, the plot tests whether the two distributions, estimated from the GCM simulation and observation, are identical. Note that a CPP plot lying above (below) the equality line indicates underestimation (overestimation) while an S-shaped CPP plot with the initial part above (below) the equality line and the second part below (above) the equality line indicates an underestimation of low (high) rainfall and overestimation of high (low) rainfall (Koutsoyiannis and Montanari, 2022). We also compare observed and simulated empirical density distribution and lag-1 autocorrelation coefficient of annual rainfall for the whole historical period. Autocorrelation is particularly interesting as indicates the capability of models to simulate long term cycles, such as those occurring during multiyear droughts.

4.2 Mean monthly and seasonal rainfall

The goodness of the simulation of ~~mean~~ monthly and seasonal rainfall averaged over the observation period is evaluated by a graphical comparison with the observed values and the Taylor diagram (Taylor, 2001). The latter is widely applied to summarise how accurately a model simulates an observed record. It integrates three statistical metrics: correlation (R), centered root-mean-square error (CRMSE) and ratio of standard deviation (SD) in a single diagram. The angular coordinate represents R. The CRMSE is measured by the distance from the point of reference (observation). Finally, the radial distance from the origin represents the ratio of SD between model simulation and observation. For perfect model simulation, R and the SD ratio assume unit value and CRMSE = 0. These statistics provide a quick summary of the correspondence between the modeled and observed mean monthly and seasonal rainfall, which is particularly useful in assessing the relative merits and overall performance of climate models (Rivera and Arnould, 2020; Dong and Dong, 2021; Yazdandoost et al., 2021). The Taylor diagram is herein used to assess the goodness of the fit of the (a) mean monthly rainfall, (b) March-April-May (MAM), (c) June-July-August (JJA), (d) September-October-November (SON), (e) December-January-February (DJF) mean rainfall.~~(a) mean monthly rainfall, (b) March-April-May (MAM) and (c) September-October-November (SON) mean rainfall.~~

~~**Extreme rainfall** The climate extreme indexes suggested by the Expert Team on Climate Change Detection and Indices (ETCCDI) (Zhang et al., 2011) have been widely used in the evaluation and projection of climate extremes (Faye and Akinsanola, 2022; Fay Table **??** describes the eight indexes that are herein used to evaluate the change of the rainfall extremes. More detailed information on the indexes can be found on the ETCCDI website (http://etccdi.pacificclimate.org/indices.shtml).~~

~~We compute the relative root-mean-square error ($RMSE'$) (Ge et al., 2019, 2021) of the simulation by each model of each index as follows. First, the root-mean-square error ($RMSE$) is calculated for each GCM in reproducing a given index: where X is a given extreme rainfall index simulated by the considered model and Y is the corresponding observed value. A standardized value $RMSE'$ of the $RMSE$ of each model is then computed as follows: where $RMSE_{Median}$ is the median~~

~~RMSE of all models. A negative $RMSE'$ indicates a better model performance than half of the models. For instance, a value of $RMSE' = -0.4$ means that the model $RMSE$ for the given index is 40% smaller than the median value across all models (Gleckler et al., 2008).~~

4.3 Meteorological droughts

To test GCM's reliability in simulating multiyear meteorological droughts we apply run theory (Yevjevich, 1967) to annual rainfall to characterise drought events in terms of drought frequency, duration, severity and intensity. Run theory is one of the most effective approaches for drought identification and has been applied in several areas worldwide (Mishra et al., 2009; Wu et al., 2020; Ho et al., 2021).

In detail, the long term mean rainfall R_{LT} is adopted as the threshold to identify positive or negative runs (see Fig. 3). If rainfall in a given year is lower than an assigned threshold $T_{lower} < R_{LT}$ a negative run is started which ends in the year when the rainfall is higher than R_{LT}. If the interval between two negative runs is only one year and rainfall in that year is less than a selected threshold $T_{upper} > R_{LT}$, then these two runs are combined into one drought. Finally, only runs which have a duration of no less than 3 years are determined as multiyear drought events. Here, the thresholds T_{upper} and T_{lower} are defined as 20% more and 10% less than R_{LT}, respectively. The threshold values have been identified with a trial and error procedure by verifying that relevant droughts observed in the past have been consistently recognised.

Once a multiyear drought has been identified, drought duration is the time span between the start and the end of the event, and drought severity is computed as the cumulative rainfall deficit with respect to R_{LT} during drought duration divided by the mean rainfall, and drought intensity is computed as the ratio between drought severity and duration. We also estimate the maximum deficit of the drought event, namely, the largest difference between annual rainfall and R_{LT} during the event. Finally, drought frequency is computed as the ratio between the number of drought events that have been identified and the length of the observation period.

5 Future climate change assessment

Statistics of future projections of annual and seasonal rainfall of the 13 considered GCMs under the three considered scenarios are compared with observed and simulated statistics of the historical period to evaluate future climate change in the Bologna region. For a detailed comparison of seasonal rainfall, the future time horizon is divided into near-future (2030-2059) and long-future (2070-2099) time windows. The multi-model median and the 25th–75th percentiles of the projections given by the GCMs are considered for each scenario to represent the associated ensemble uncertainty.

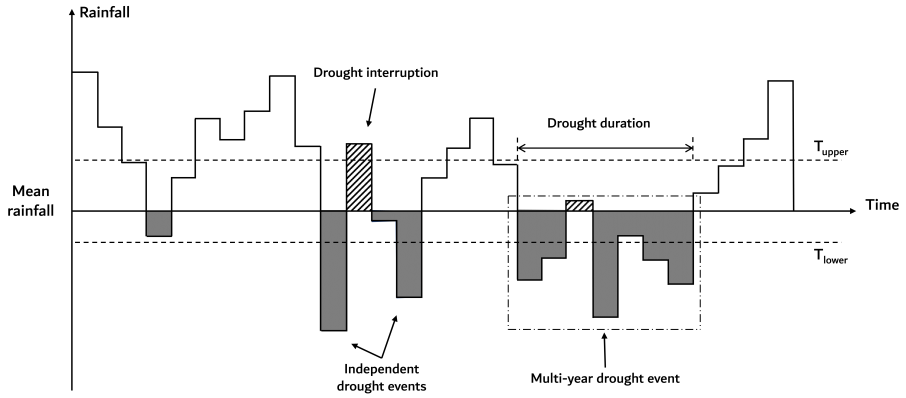

Figure 3. Identification of multiyear drought events and characteristics by using run theory.

6 Results

6.1 Reliability of historical data simulation

6.1.1 Monthly and annual rainfall

Fig. 4 and Fig. 5 show a comparison of simulated versus observed hyetographs for each of the 13 CMIP6 GCMs thus providing a visual assessment of their reliability for long-past and near-past historical period, respectively. Some models (e.g. CanESM5 and MPI-ESM1-2-LR) generally underestimate the annual rainfall for both periods while other models (e.g. ACCESS-CM2 and GFDL-ESM4) end up with a prevailing overestimation. Some other models perform differently in the two considered historical time periods. In INM-CM5-0 and NorESM2-MM, for example, long-past rainfall is relatively better reproduced than near-past rainfall while the opposite result is found for INM-CM4-8, which can better simulate near-past rainfall. The multi-model ensemble (MME), namely the mean value of the 13 GCMs, is characterized by a much lower variability than the historical data and therefore does not provide any improvement with respect to each individual GCM.

Fig. 6 shows the obtained goodness of fit indexes. Good model performances correspond to a high positive correlation coefficient (R), high Kling-Gupta efficiency (KGE), high Nash-Sutcliffe efficiency (NSE), and a small mean absolute relative error (MARE). The metrics are calculated for long-past, near-past, and total historical period 1850-2014.

For both long-past and near-past periods, 4 of the models (GFDL-ESM4, CMCC-ESM2, FGOALS-g3, and INM-CM5-0) and the MME show a positive R while MIROC6, ACCESS-CM2, and NorESM2-MM are characterised by negative R in both periods. Some other models perform differently in each time period. INM-CM4-8 and CanESM5 show a positive R in near-past and negative in long-past. It is interesting to note that INM-CM4-8 obtains the highest R in near-past. An opposite

~~result is obtained for MPI-ESM1-2-LR, EC-Earth3-Veg-LR and MRI-ESM2-0 with the latter showing the highest negative value in near-past. INM-CM5-0 exhibits the highest R in the long-past but shows no significant correlation in the near-past while FGOALS-g3 obtains the highest R in the near-past. In addition, the IPSL-CM6A-LR exhibits the highest negative value in simulating long-past rainfall. In general, the positive correlation is low for all models.~~

~~The MME delivers the lowest MARE in each period and performs similarly for both near-past and long-past. ACCESS-CM2, GFDL-ESM4, and MPI-ESM1-2-LR show a relatively higher MARE for both periods while ACCESS-CM2 shows the highest MARE in each period. The other models exhibit a relatively smaller MARE and perform slightly different in each period.~~

~~As for KGE, 5 of 13 models and the MME show negative KGE values for both long-past and near-past. The ACCESS-CM2 and the MME show the highest negative value for long-past and near-past, respectively. Furthermore, CanESM5 and IPSL-CM6A-LR exhibit a high negative value for long-past while the values are near to zero for the near-past. Only FGOALS-g3, GFDL-ESM4, and CMCC-ESM2 depict relatively higher KGE values for both periods while the INM-CM5-0 and INM-CM4-8 show the highest value for long-past and near-past, respectively.~~

~~All models and the ensemble show negative values of NSE which indicate an overall poor ability in replicating historical annual rainfall. ACCESS-CM2, MPI-ESM2-2-LR, and GFDL-ESM4 depict relatively poorer performancessmaller values for both periods with the ACCESS-CM2 providing the less satisfactory onesshowing the smallest value in each period. The multi-model ensemble and FGOALS-g3 exhibit a higher NSE with a value close to zerovalue near zero, which presents a similar performance to the me~~

~~In general, CMIP6 GCMs do not satisfactorily reproduce the historical sequence of annual rainfall, with the less satisfactory performances exhibited by ACCESS-CM2, MRI-ESM2-0, and EC-Earth3-Veg-LR. FGOALS-g3, GFDL-ESM4 and CMCC-ESM2 perform relatively better for both periods while INM-CM5-0 and INM-CM4-8 show the best ability in simulating long-past and near-past annual rainfall, respectively. Interestingly, all the models show a better performance during one separate period than the total historical period. Compared to every single model, the MME shows no obvious better capability in replicating the long sequence of annual rainfall.~~

Fig. 4 and Fig. 5 show the CPP plots of observed and simulated monthly rainfall time series in the 5 subsequent time windows during the historical period before and after QDM bias correction, respectively. Note that QDM was operated over the whole historical period.

While each individual model displays consistent performance in terms of probability distribution across various time windows with only minor differences, the results do not allow an easy identification of the optimal model for a given time window. Before bias correction, MPI-ESM1-2-LR generally underestimates monthly rainfall for all periods while some other models (ACCESS-CM2, GFDL-ESM4, and MIROC6) end up with a prevailing overestimation. CMCC-ESM2 and CanESM5 relatively well capture the low rainfall while underestimating high rainfall. The remaining models (e.g. FGOALS-g3, IPSL-CM6A-LR, and INM-CM4-8) fit the observed distribution generally well in some time windows while exhibiting slight departures in other periods. The multi-model ensemble satisfactorily simulates the mean value while overestimating and underestimating the low and high rainfall, respectively.

As expected, after bias correction all models show a better performance in reproducing probability distributions in different periods except FGOALS-g3 and INM-CM4-8 which slightly underestimate low rainfall after QDM. The performance of the

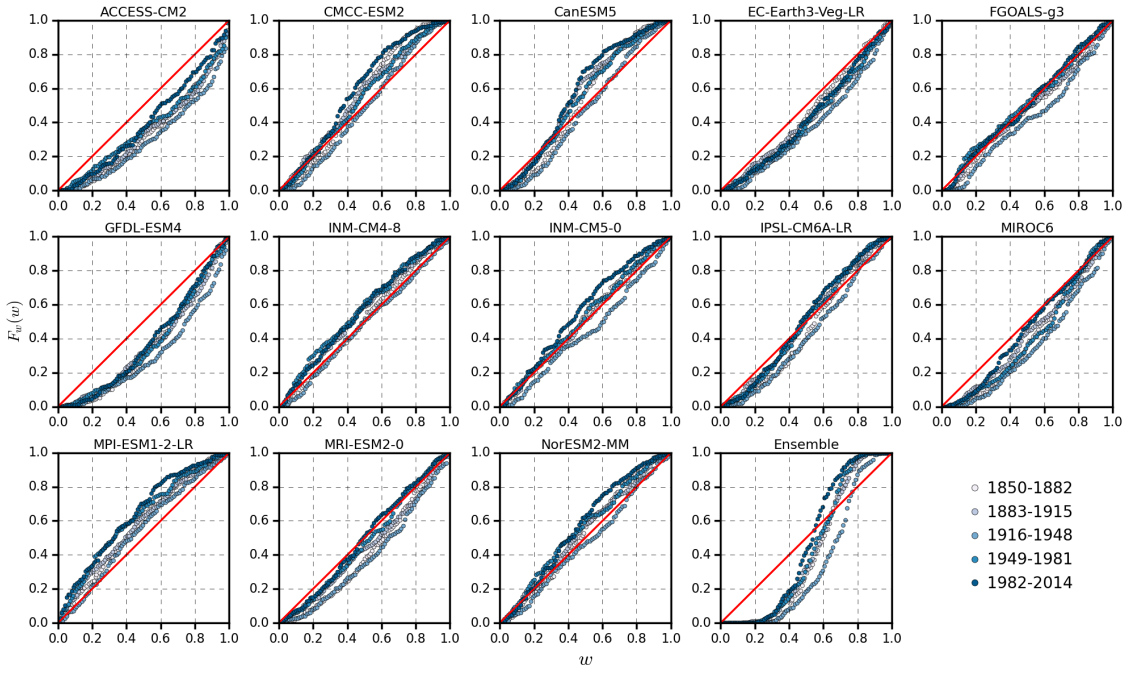

Figure 4. Combined Probability-Probability plot of observed and GCM simulated monthly rainfall in five time windows during the historical period before bias correction.

ensemble remains somewhat consistent, although for some periods one notes an improvement in the fit of the mean value. In general, it is confirmed that bias correction improves the model performances for historical simulation. The obvious limit of QDM is the requirement of an extended data set of historical data.

For the whole historical period, Fig. 6 shows a comparison between observed and simulated empirical density distribution of annual rainfall before and after bias correction~~We further investigate the ability of GCMs in capturing the sample probability density for annual rainfall in the whole historical period~~. Before QDM, the majority of models perform worse in capturing heavier rainfall than lighter rainfall, with the opposite result occurring for EC-Earth3-Veg-LR, MIROC6, and MRI-ESM2-0. Only ACCESS-CM2 and GFDL-ESM4 overestimate both high and low rainfall. After QDM, all the models exhibit improved performances although underestimation of high rainfall is introduced by QDM for 7 models and the ensemble simulation. It is noted that MME exhibits lower variability with respect to observed data both before and after bias correction. This result is coerent with what is shown in Fig. 4 and Fig. 5.

Table 2 displays the lag-1 autocorrelation coefficient for each model simulation before and after QDM bias correction. It is interesting to note that the correlation of observed data is slightly higher than all the models, therefore highlighting possible model weakness in simulating long runs. After QDM, the correlation decreases which indicates that QDM may influence the autocorrelation of raw output of GCM.

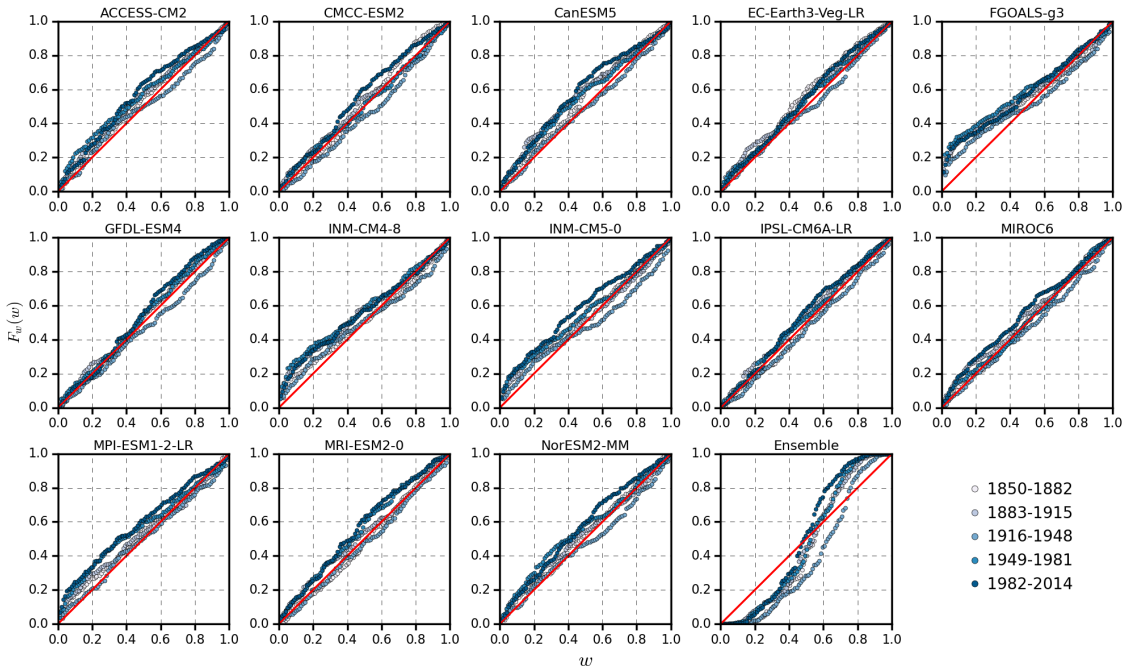

Figure 5. Combined Probability-Probability plot of observed and GCM simulated monthly rainfall in five time windows during the historical period after bias correction.

Table 2. Lag-1 autocorrelation coefficient between observed annual rainfall and each historical simulation before and after bias correction.

Label	OBS	MME	1	2	3	4	5	6	7	8	9	10	11	12	13
RAW	0.226	0.052	−0.012	0.122	0.036	0.173	0.127	0.099	0.085	−0.084	0.018	0.086	0.067	0.081	−0.052
QDM	0.226	0.018	−0.063	0.061	0.006	0.120	0.103	0.080	0.059	−0.079	−0.068	−0.001	0.013	0.013	−0.088

OBS is observation data, MME is multi-model ensemble mean and numbers indicate models: 1. ACCESS-CM2; 2. CMCC-ESM2; 3.CanESM5; 4.EC-Earth3-Veg-LR; 5. FGOALS-g3; 6. GFDL-ESM4; 7. INM-CM4-8; 8. INM-CM5-0; 9. IPSL-CM6A-LR; 10. MIROC6; 11. MPI-ESM1-2-LR; 12. MRI-ESM2-0; 13. NorESM2-MM.

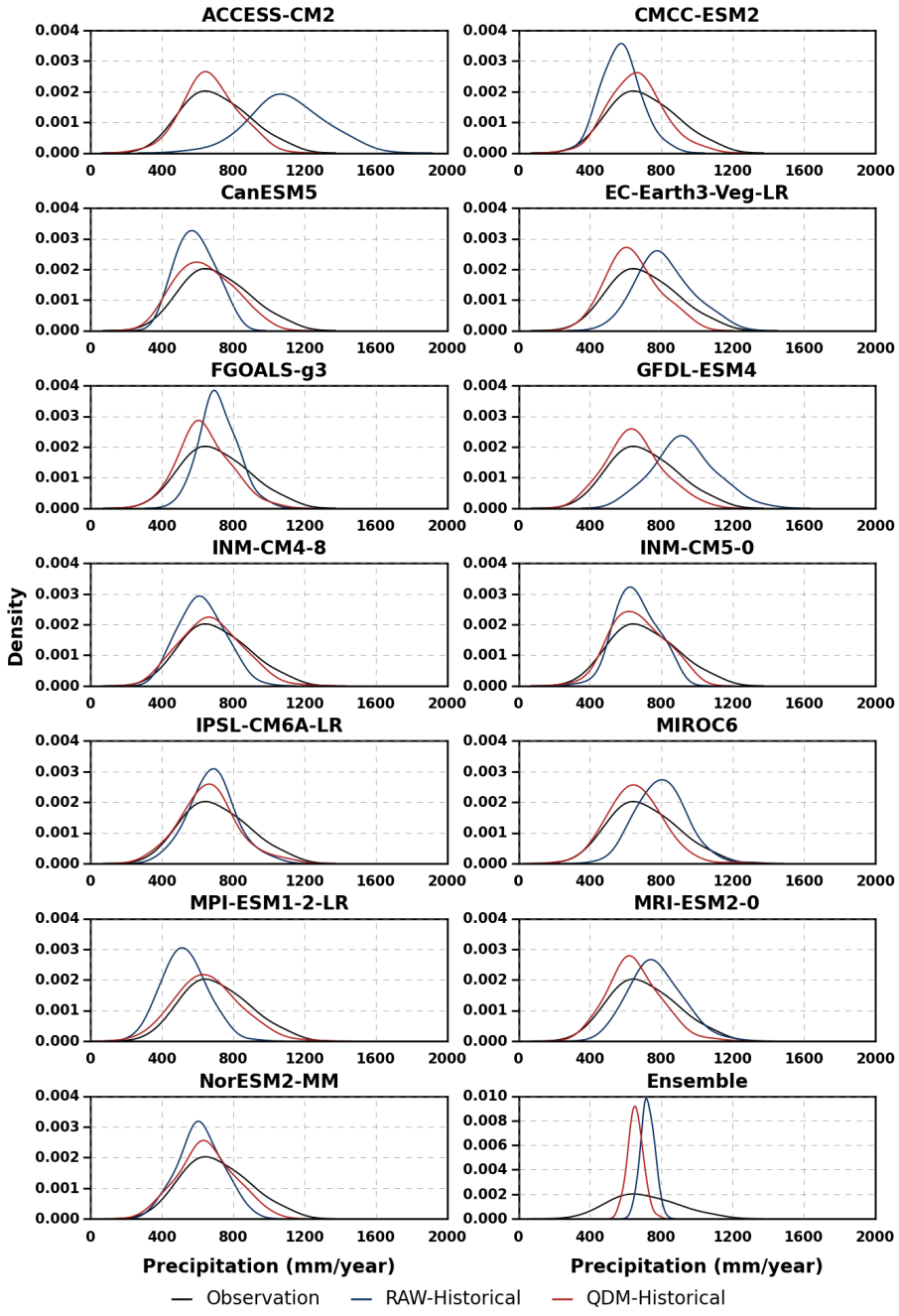

Figure 6. Comparison of sample probability density of annual rainfall for observations and GCM historical simulations.

6.1.2 Mean monthly and seasonal rainfall

Fig. 7 shows the graphical comparison between observed and simulated mean monthly rainfall for the whole historical period before bias correction. Most of the models adequately replicate the seasonal rainfall pattern except INM-CM4-8~~IPSL-CM6A-LR~~ and FGOALS-g3. However, ACCESS-CM2 ~~significantly~~ overestimates rainfall for all months. Moreover, all models but MPI-ESM1-2-LR, CMCC-ESM2, NorESM2-MM, and CanESM5 overestimate rainfall in DJF~~December-January-February~~. Conversely, all models but ACCESS-CM2, GFDL-ESM4, EC-Earth3-Veg-LR, and MRI-ESM2-0 underestimate JJA~~June-July-August~~ and/or SON~~September-October-November~~ rainfall. For MAM~~the March-April-May~~ rainfall, several models exhibit overestimation. In addition, the MME satisfactorily reproduces the annual cycle of rainfall, especially MAM~~the March-April-May~~ rainfall while slightly underestimating JJA~~the June-July-August~~ and SON~~September-October-November~~ and overestimating DJF~~the December-January-February~~ rainfall.

Fig. 8 displays the same graphical comparison after QDM. Although INM-CM4-8 and FGOALS-g3 still fail in reproducing the seasonal pattern, the performance of ACCESS-CM2 has been improved. Other models show improvement in some seasons only (e.g. NorESM2-MM better captures SON rainfall but shows worse performances for MAM rainfall). In general, most of the models underestimate summer JJA rainfall after QDM.

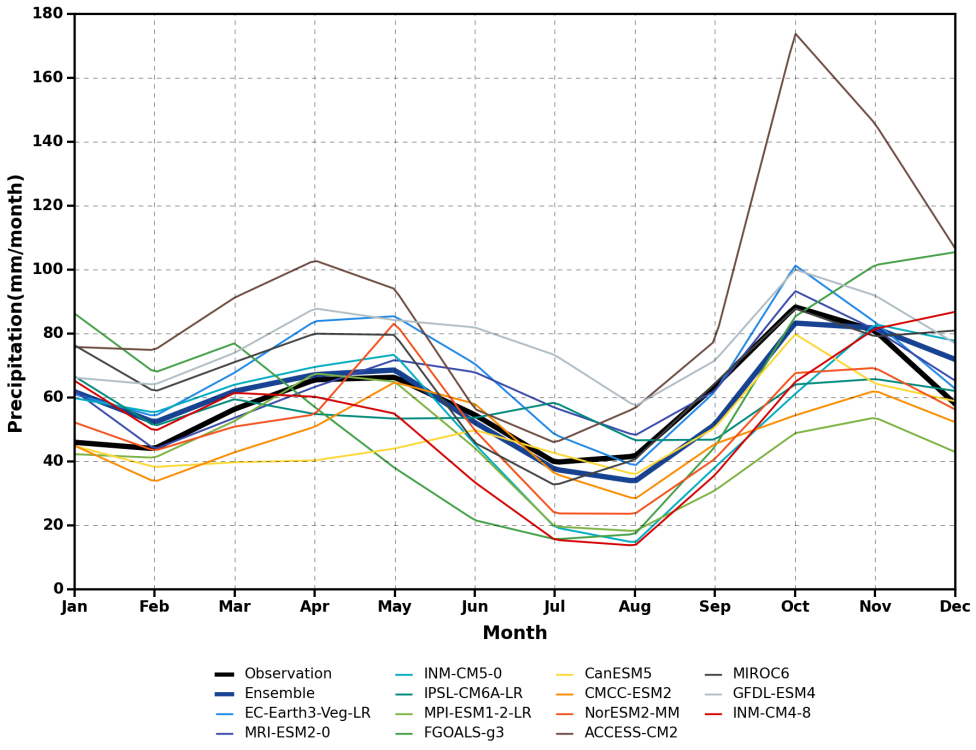

Figure 7. Annual cycle of historical (1850-2014) mean monthly rainfall (mm/month) of MME and CMIP6 models against observation data.

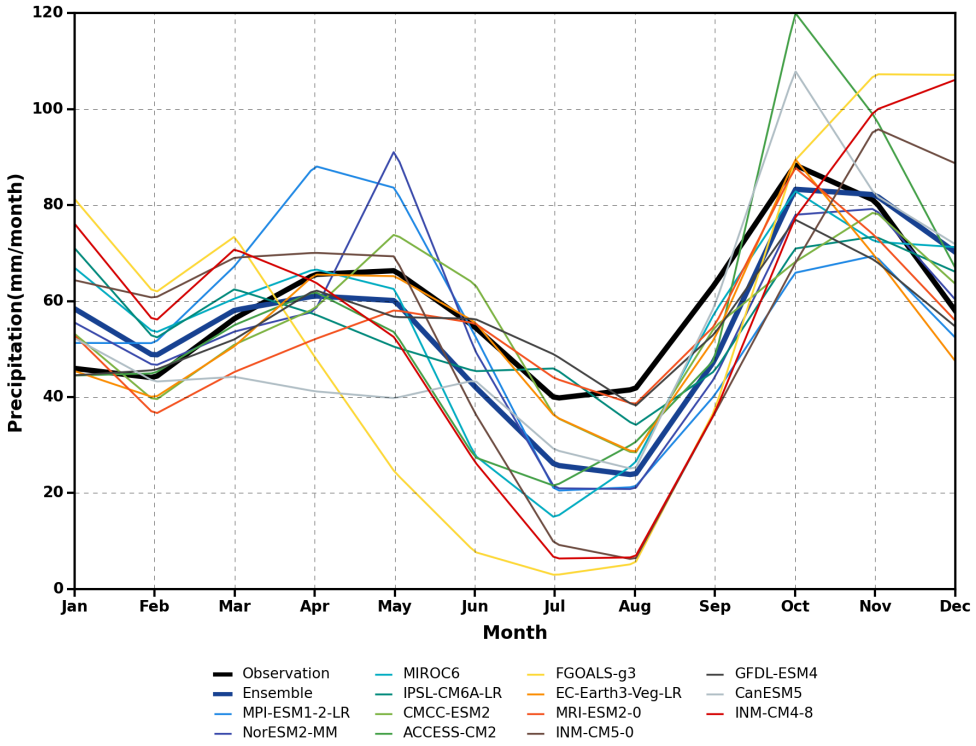

Figure 8. Annual cycle of historical (1850-2014) mean monthly rainfall (mm/month) of MME and CMIP6 models against observation data after bias correction.

Fig. 9 shows the Taylor diagram for each GCM and the MME in simulating the seasonal rainfall pattern. It confirms that 6 of the models and the MME adequately replicate the mean monthly rainfall with relatively high R, low CRMSE, and low SD. In particular, satisfactory results are provided by EC-Earth3-Veg-LR and MRI-ESM2-0. IPSL-CM6A-LR and FGOALS-g3 display relatively poor performance. However, a considerable~~significant~~ improvement is exhibited in reproducing MAM, SON, and DJF~~the March-April-May and September-October-November~~ rainfall for these two models, especially FGOALS-g3 that can simulate SON~~the September-October-November~~ rainfall best. For DJF rainfall, all models show a better performance with respect to other seasons and MME shows the best performance than any individual model. Most of the models depict a worse performance in SON and JJA than MAM and DJF with much larger bias~~a higher R but larger bias in September-October-November than March-April-May~~. In general, the EC-Earth3-Veg-LR provides a satisfactory simulation of both annual cycle and each seasonal~~March-April-May~~ rainfall while GFDL-ESM4, ACCESS-CM2 and FGOALS-g3 can better reproduce the MAM, JJA, and SON~~March-April-May and September-October-November~~ rainfall, respectively. Although R is slightly lower than the EC-Earth3-Veg-LR, the MME performs satisfactorily in simulating the annual cycle of rainfall and best captures DJF rainfall but shows no ~~significant~~ advance over the single models in reproducing MAM, JJA, and SON~~March-April-May and September-October-November~~ rainfall.

Fig. 10 shows the Taylor diagram after QDM. In general several models show higher CRMSE for JJA rainfall and lower SD for SON rainfall while the performance for MAM rainfall is similar before and after QDM. Although QDM displays no evident improvement in the models' ability to capture seasonal rainfall patterns, one notes a slightly enhanced performance for the SON season. Overall, the models still show a better ability to reproduce DJF rainfall compared to other seasons. For the annual cycle, individual models (e.g. GFDL-ESM4 and MRI-ESM2-0), exhibit improved simulation abilities but the MME shows a higher SD after bias correction.

~~Extreme rainfall~~

~~Table **??** shows the median $RMSE$ of each extreme index for the ensemble models. The $RMSE'$ of individual model in simulating rainfall extremes are summarized in Fig. 10. MRI-ESM2-0 performs fairly well with all negative $RMSE'$ values for different indices. Two models, EC-Earth3-Veg-LR and GFDL-ESM4, provide acceptable performance with mainly negative $RMSE'$ values for different indices. The $RMSE'$ of the multi-model mean is shown in the last column of Fig. 10. The ensemble mean performs bett~~

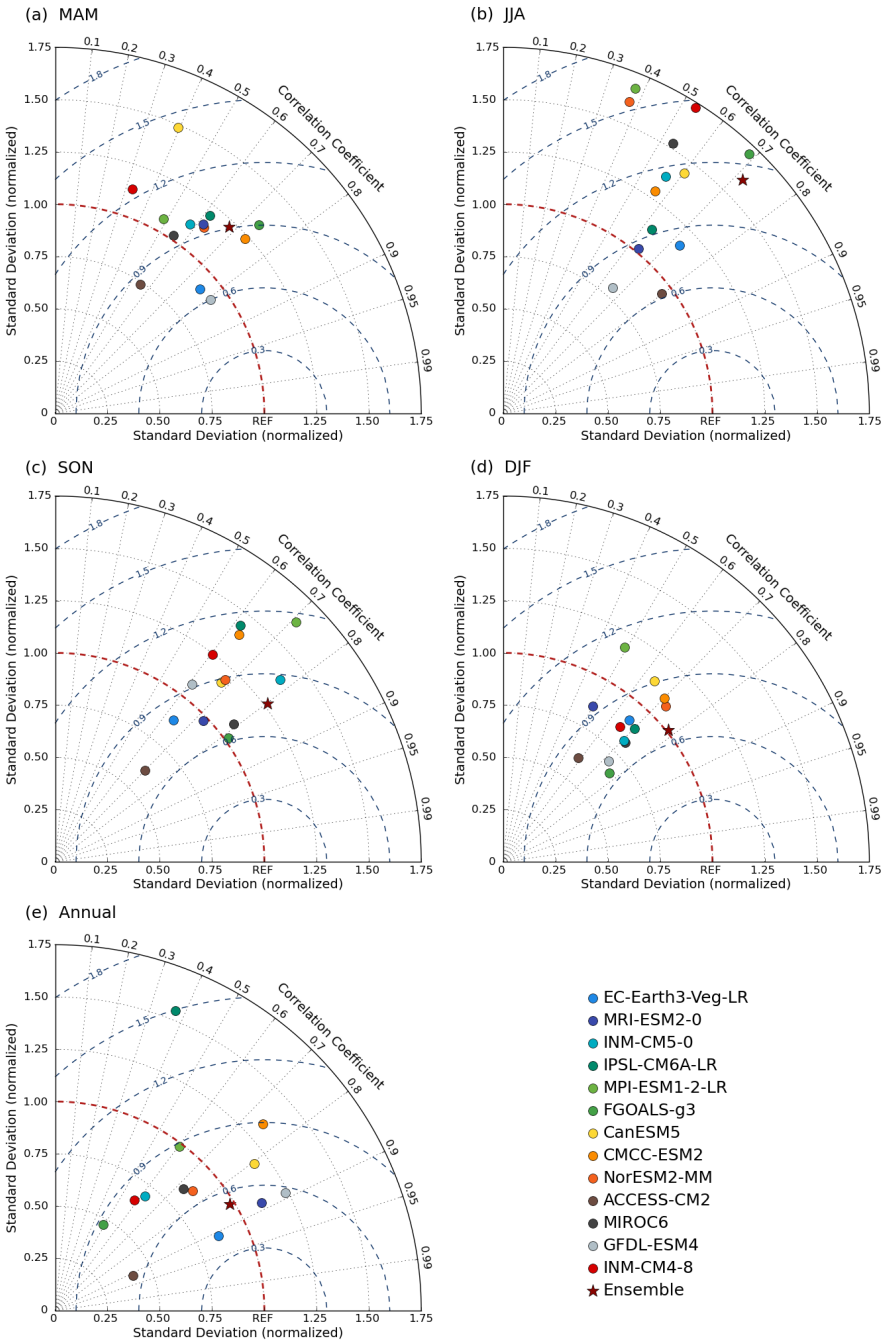

Figure 9. Taylor diagram of (a) March-April-May (MAM), (b) June-July-August (JJA), (c) September-October-November (SON), (d) December-January-February (DJF) and (e) Annual cycle of historical (1850-2014) mean monthly rainfall for the GCM models and their MME before bias correction.

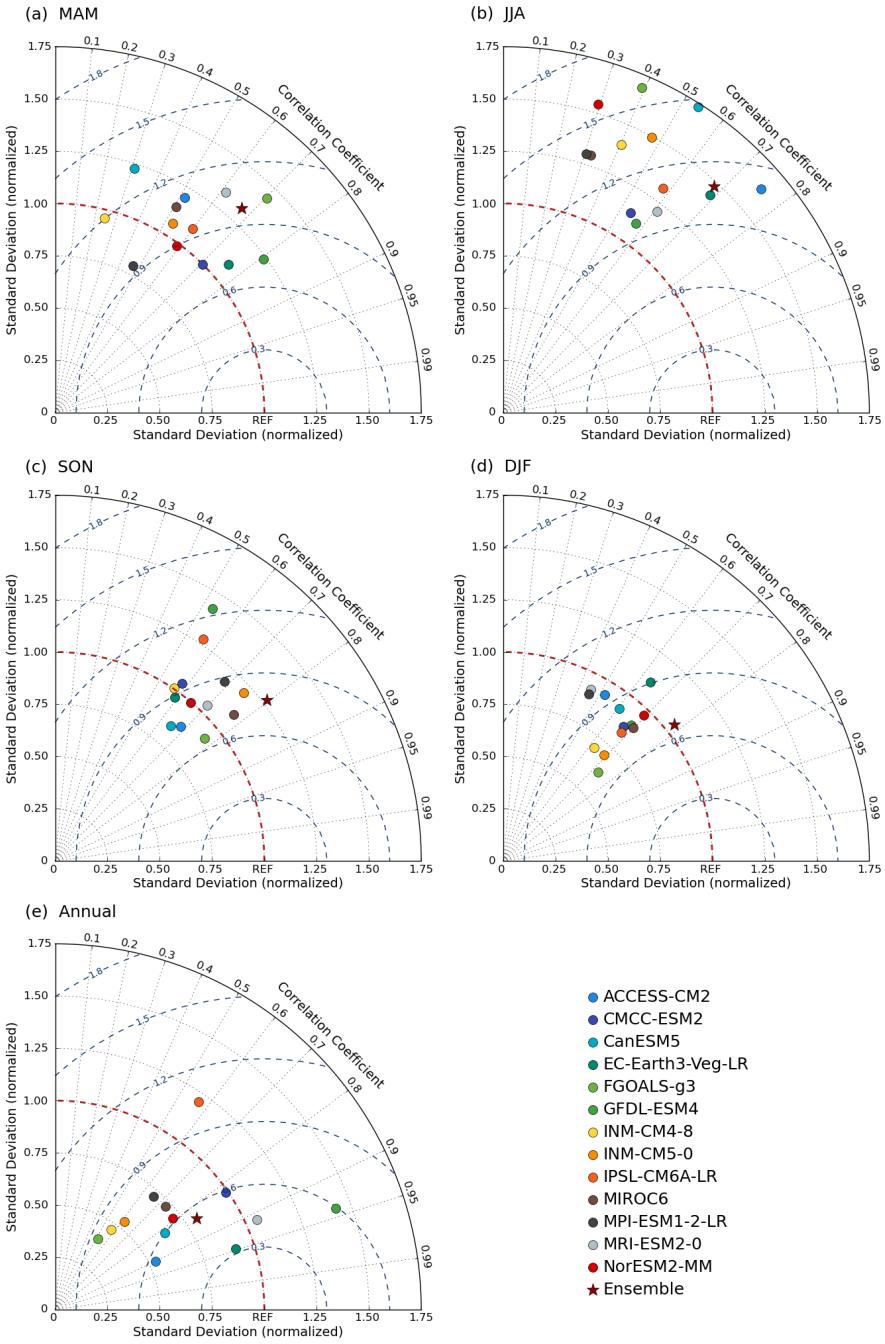

Figure 10. Taylor diagram of (a) March-April-May (MAM), (b) June-July-August (JJA), (c) September-October-November (SON), (d) December-January-February (DJF) and (e) Annual cycle of historical (1850-2014) mean monthly rainfall for the GCM models and their MME after bias correction.

6.1.3 Meteorological droughts

Table 3 and Table 4 show the characteristics of multiyear droughts derived by using run theory for both observations and each model simulation before and after bias correction for both the whole historical and whole future periods, the latter for SSP2.6 and SSP8.5 scenarios. For the observation data, the long-term mean value of annual rainfall, which R_{LT} of 705 (mm/year) is considered as the threshold to identify the observed multiyear drought events.

Table 3. Mean values over the considered period of drought frequency (DF), duration (DD), intensity (DI) and maximum deficit (MD)~~Statistics~~ for multiyear meteorological droughts exhibited by observed data (1850-2014) and reproduced by models before bias correction for the historical (1850-2014) and future (2015-2100) periods under the two considered scenarios.

Models	Drought Frequency			Drought Duration			Drought Intensity			Max Deficit		
	HIS	SSP2.6	SSP8.5	HIS	SSP2.6	SSP8.5	HIS	SSP2.6	SSP8.5	HIS	SSP2.6	SSP8.5
OBS	0.055	-	-	4.80	-	-	0.205	-	-	36.70%	-	-
MME	0.053	0.056	0.050	3.96	4.02	4.75	0.153	0.168	0.184	26.40%	27.87%	32.24%
1	0.030	0.070	0.058	4.20	3.83	4.20	0.134	0.184	0.202	27.05%	33.23%	39.55%
2	0.042	0.035	0.035	3.86	4.00	4.00	0.167	0.154	0.192	23.13%	27.05%	31.84%
3	0.042	0.047	0.058	4.14	3.75	4.20	0.155	0.188	0.210	29.89%	28.32%	36.45%
4	0.079	0.081	0.058	4.23	3.71	4.60	0.140	0.171	0.203	24.60%	28.18%	34.16%
5	0.055	0.035	0.035	4.11	3.00	3.67	0.130	0.110	0.091	22.03%	18.29%	19.21%
6	0.048	0.035	0.047	3.63	6.33	4.50	0.162	0.152	0.173	28.76%	25.12%	26.90%
7	0.048	0.070	0.070	4.25	4.00	3.83	0.167	0.155	0.177	27.27%	23.55%	29.40%
8	0.055	0.058	0.058	4.33	3.60	4.60	0.131	0.189	0.189	21.66%	30.67%	34.00%
9	0.048	0.081	0.023	4.25	3.57	9.00	0.132	0.177	0.246	26.86%	28.29%	48.37%
10	0.055	0.070	0.047	3.44	3.67	5.75	0.163	0.147	0.138	26.55%	24.34%	21.95%
11	0.079	0.047	0.047	3.54	3.75	5.00	0.173	0.226	0.184	29.03%	37.97%	34.04%
12	0.055	0.035	0.058	3.89	5.67	4.40	0.173	0.161	0.191	28.75%	26.34%	28.02%
13	0.055	0.070	0.058	3.56	3.33	4.00	0.158	0.165	0.194	27.63%	30.93%	35.23%

The unit of drought frequency is (times/year) and the unit of drought duration is (years). OBS and HIS are observed data and historical simulation. MME is multi-model ensemble mean and different numbers indicate different models: 1. ACCESS-CM2; 2. CMCC-ESM2; 3.CanESM5; 4.EC-Earth3-Veg-LR; 5. FGOALS-g3; 6. GFDL-ESM4; 7. INM-CM4-8; 8. INM-CM5-0; 9. IPSL-CM6A-LR; 10. MIROC6; 11. MPI-ESM1-2-LR; 12. MRI-ESM2-0; 13. NorESM2-MM.

The results highlight that FGOALS-g3, INM-CM5-0, MIROC6, MPI-ESM2-0, and NorESM2-MM simulate drought frequency (DF) fairly well. Notably, all models fail to replicate the drought duration (DD), drought intensity (DI), and maximum deficit (MD). For instance, IPSL-CM6A-LR and INM-CM4-8 show the best performance in simulating DD, which, however, is underestimated by about 10% by the best simulation. Although MPI-ESM1-2-LR presents the highest value of DI and MD against all the models, marked underestimation with respect to the observations still occurs. MME displays relatively good

Table 4. Mean values over the considered period of drought frequency (DF), duration (DD), intensity (DI) and maximum deficit (MD) for multiyear meteorological droughts exhibited by observed data (1850-2014) and reproduced by models after bias correction for the historical (1850-2014) and future (2015-2100) periods under the two considered scenarios.

Models	Drought Frequency			Drought Duration			Drought Intensity			Max Deficit		
	HIS	SSP2.6	SSP8.5	HIS	SSP2.6	SSP8.5	HIS	SSP2.6	SSP8.5	HIS	SSP2.6	SSP8.5
OBS	0.055	-	-	4.80	-	-	0.205	-	-	36.70%	-	-
MME	0.054	0.056	0.055	4.22	4.05	4.32	0.184	0.199	0.209	32.14%	33.15%	37.02%
1	0.030	0.070	0.081	4.20	3.83	4.43	0.141	0.202	0.203	28.75%	36.10%	38.72%
2	0.030	0.058	0.047	4.40	4.40	4.50	0.198	0.202	0.242	35.35%	30.73%	45.30%
3	0.055	0.047	0.047	4.00	4.00	3.75	0.191	0.213	0.220	28.36%	33.23%	37.02%
4	0.073	0.081	0.047	4.33	3.71	4.75	0.171	0.199	0.271	29.43%	31.78%	44.94%
5	0.061	0.023	0.047	4.60	3.50	4.25	0.191	0.123	0.148	35.21%	21.78%	32.31%
6	0.061	0.035	0.047	4.00	6.00	3.75	0.191	0.217	0.201	32.07%	33.46%	32.54%
7	0.042	0.070	0.058	5.00	4.33	4.00	0.212	0.193	0.232	35.14%	29.17%	38.56%
8	0.042	0.058	0.070	4.29	5.00	4.33	0.179	0.227	0.206	28.46%	42.26%	38.50%
9	0.048	0.058	0.035	4.63	4.00	5.33	0.170	0.219	0.228	33.72%	36.45%	42.97%
10	0.067	0.058	0.058	3.45	3.40	5.00	0.195	0.185	0.168	33.73%	30.12%	26.46%
11	0.079	0.058	0.047	3.69	3.60	3.50	0.194	0.226	0.177	33.03%	38.03%	28.24%
12	0.048	0.047	0.070	4.25	3.00	4.50	0.188	0.211	0.214	33.32%	30.98%	36.38%
13	0.067	0.070	0.058	4.00	3.83	4.00	0.168	0.182	0.204	31.32%	36.83%	39.64%

The unit of drought frequency is (times/year) and the unit of drought duration is (years). OBS and HIS are observed data and historical simulation. MME is multi-model ensemble mean and different numbers indicate different models: 1. ACCESS-CM2; 2. CMCC-ESM2; 3.CanESM5; 4.EC-Earth3-Veg-LR; 5. FGOALS-g3; 6. GFDL-ESM4; 7. INM-CM4-8; 8. INM-CM5-0; 9. IPSL-CM6A-LR; 10. MIROC6; 11. MPI-ESM1-2-LR; 12. MRI-ESM2-0; 13. NorESM2-MM.

performance in terms of DF but still underestimates DD, DI, and MD. In detail, the MME DD is underestimated by about 17%, while DI and MD for observations are even nearly 34% and 39% higher than MME, respectively.

After QDM, a slight improvement is obtained for the simulation of DF. MME confirms its satisfactory performance, although 6 models still end up with underestimation. Notably, all models still fail to replicate the DD, DI and MD, with the only exception of INM-CM4-8 which satisfactorily reproduces drought characteristic. In general, the impact of QDF varies depending on each model and drought characteristic.

QDM slightly mitigates the underestimation by MME of DD, DI, and MD, which however remain 12%, 10% and 12% lower than observations, respectively.

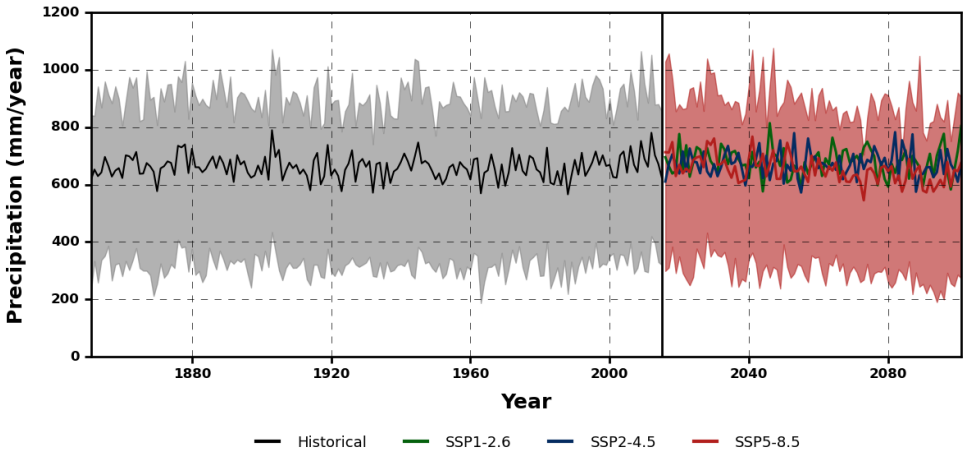

Figure 11. Time series of annual rainfall for both historical simulation and future projections after bias correction (mm/year). Historical (black), 30-year mean of historical (yellow), future MME simulations for SSP1-2.6 (green), SSP2-4.5 (blue), and SSP5-8.5 (red). The uncertainty range for historical simulation (grey shading) and future projection under SSP5-8.5 (red shading) are shown in the 25th and 75th percentiles of the ensemble.

6.2 Future changes

6.2.1 Changes in average rainfall

Fig. 11 shows the future projections of annual rainfall after QDM for three different scenarios (SSP1-2.6, SSP2-4.5 and SSP5-8.5), compared with the MME simulation of the historical period (1850-2014). Future annual rainfall shows only a slight decrease for the SSP5-8.5 scenario while fluctuating with close to stationary variability with respect to the historical period. After applying bias correction, future projections exhibit similarities to the raw output but with a slightly lesser degree of change.

To inspect the temporal progress of changes, the annual cycle of rainfall after bias correction is considered and the future period is divided into near-future (2030-2059) and long-future (2070-2099) related to the present-day simulation (1985-2014). Fig. 12 shows an overall decrease in monthly rainfall under all the scenarios in each future period. Under the SSP2.6 scenario, the MAM rainfall in the long-future is expected to be less than in the near-future period. Conversely, the rainfall during the SON season in the long-future will be higher than in the near-future. Moreover, rainfall may be more likely to be concentrated in November rather than in October in both periods. This change in rainfall pattern also occurs in the near-future period under the SSP4.5 scenario, but the overall monthly rainfall shows no evidentsignificant difference between periods. Under the SSP8.5 scenario, the MAM and JJS rainfall in the long-future will be considerablysignificantly less than in the near-future and the rainfall will be more concentrated in the SON and DJF season. The results after QDM depict an analogous future change of the

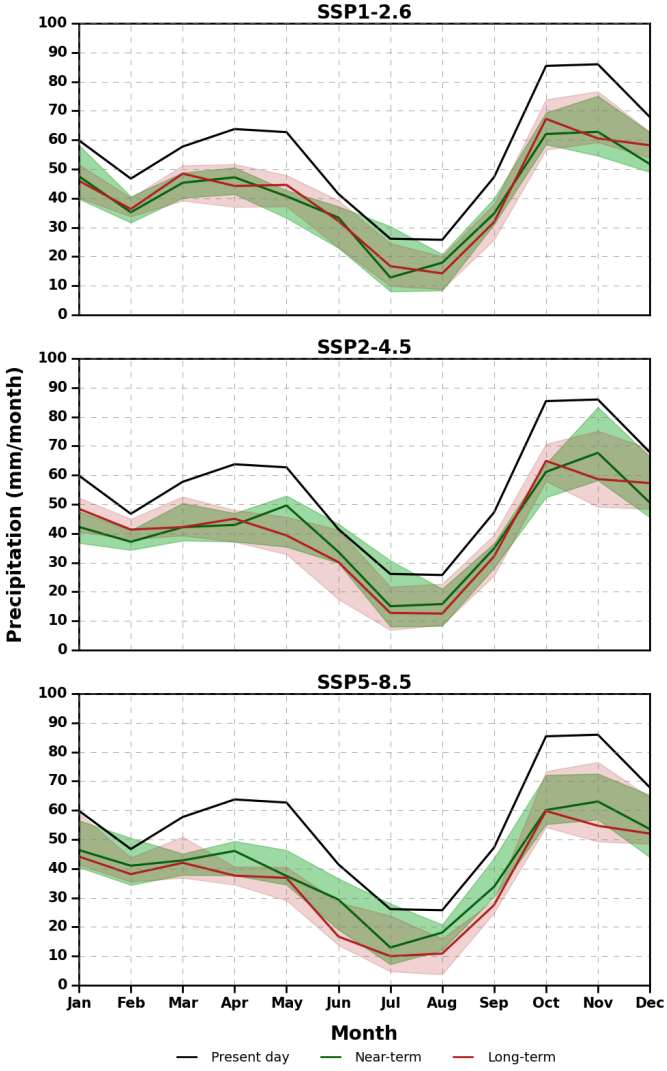

Figure 12. Monthly mean rainfall (mm/month) for multi models in present-day (1985-2014), near-future (2030-2059) and long-future (2070-2099) under SSP1-2.6, SSP 2-4.5 and SSP5-8.5 scenarios. Thick lines indicate multi-model medians while shading indicates the 25th–75th percentile model ranges.

seasonal rainfall pattern. With respect to raw output, however, the JJA season is projected to be drier after QDM which aligns with the historical simulation.

6.2.2 Future meteorological drought changes

Table 3 reports future drought changes for SSP2.6 and SSP8.5 future scenarios~~, which represent low-emission pathway and high-emission pathway, respectively~~. Corresponding statistics for observed data (whole 1850-2014 observations) and model simulations under historical period (1850-2014) are also shown.

~~For drought frequency (DF), t~~The future changes of DF with respect to historical simulations are not remarkable. DD, DF and MD are generally underestimated with respect to historical data.~~Note that this statistic was fairly well reproduced by the simulation of models under historical period with respect to observed DF. The MME also shows no significant change, with DF slightly increasing under SSP2.6 but declining under SSP8.5 compared to historical simulation. The future changes are much more significant for drought duration (DD), drought intensity (DI) and maximum deficit (MD), which, however, were underestimated by models when simulating the historical period.~~ In fact, the values of DD, DI and MD for MME under SSP2.6 and SSP8.5 are both lower with respect to observations. Moreover, DD of all models but GFDL-ESM4 is shorter under SSP2.6 with respect to observed data. Only the MME and models CMCC-ESM2, MIROC6 and MPI-ESM1-2 show a consistent increase of DD when turning from historical simulation to SSP2.6 and SSP8.5. Nearly all models show an increase in DI with respect to historical simulation for at least one future scenario and all models but FGOALS-g3 and MIROC6 show a more intense drought under SSP8.5 than SSP2.6. Changes in MD are similar to DI. The above considerations show that historical data depict a worse future in terms of multiyear droughts with respect to simulations before QDM.~~The DD of most of the models is longer under SSP8.5, while being shorter under SSP2.6, with respect to historical simulation. Nearly all the models show a marked increase in DI for at least one future scenario and the majority of models show a more intense drought under SSP8.5 than SSP2.6 except FGOALS-g3 and MIROC6, which show a continuous decline in both future scenarios. The changes in MD are similar to DI. Additionally, the MME shows a continuous increase for DD, DI and MD with a relatively larger increase under SSP8.5. The DD will only slightly increase under SSP2.6 while resulting 19.9% longer under SSP8.5. Similar changes occur in MD, which will increase by 22.1% under SSP8.5. The MME shows that DI will increase by 9.8% under SSP2.6, while a much larger increase of 20.3% under SSP8.5. We should note that the climate models do not show consistent predictions for future changes in statistics of multiyear droughts, but they depict overall drier conditions in the high-emission scenario compared to both low-emission scenario and historical simulation.~~

Table 4 presents the corresponding future drought changes after QDM. The good performance of MME in terms of DF is confirmed. However, more models exhibit less DF compared to observations under SSP8.5. For DD, QDM results in a general deterioration of performances in terms of underestimation. For DI and MD, similar values to observed data are reached by MME under SSP8.5 only, but with large variability among models. In general, QDM improves MME performances but one still notes large variability among models. Moreover, the expectation of increased drought risk in the future with respect to historical observations is not confirmed even after QDM and the application of the worst emission scenario.

~~In general, few models only predict the worst meteorological drought statistics during 2015-2100 with respect to 1850-2014 observations, and MME does not resolve the problem as it delivers a less conservative prediction with respect to past occurrences of multiyear droughts.~~

~~6 Discussion~~

~~As the results show, the GCMs cannot satisfactorily reproduce the chronological annual rainfall series and the performance of the model will depend on the selected window of the historical period. However, the multi-model ensemble can simulate the monthly rainfall pattern and extreme rainfall fairly well, as well as drought frequency. Conversely, drought duration, intensity, and maximum deficit are underestimated by models.~~

~~Therefore, one may expect that future projections for different emission scenarios provided by GCM models may deliver useful predictions of rainfall statistics to assess the impact of climate change. However, one should note that the projected changes in average rainfall are not remarkable: only for SSP8.5, a decrease is predicted of about 10% with respect to the historical period. The seasonal rainfall pattern is not predicted to change significantly.~~

~~For multiyear meteorological droughts, that is our main focus, we pointed out above that the multi-model ensemble can satisfactorily simulate their frequency while significantly underestimating duration, intensity, and maximum deficit. Additionally, the duration, intensity and maximum deficit predicted by models in the future are generally less critical than what was observed in the past. Such evidence suggests carefully considering the information supporting the technical design of climate change adaptation policies for droughts. For the considered case of the Bologna region, extreme drought estimation carried out by statistical analysis of historical records under the assumption of stationarity delivers a more critical future prediction with respect to climate model simulations.~~

7 Conclusions

The present study refers to the region of Bologna, where the availability of a 209-year-long daily rainfall series allows us to make a unique assessment of GCM reliability and their predicted changes in rainfall and drought risk. ~~Our results confirm the poor capability of GCM to simulate the chronological order of observations, therefore highlighting that their uncertainty should be carefully considered, which is not effectively accounted for by drawing ensemble simulations.~~ The results show that GCMs provide a satisfactory simulation of rainfall seasonality while statistics of rainfall series estimated for the long term historical period exhibit discrepancies among models and limited reliability in some cases. In particular, the correlation of annual rainfall looks underestimated, thereby implying a possible lack of fit in the simulation of long term cycles.

In fact, our focus is concentrated on the statistics of multiyear droughts. We found that the multi-model ensemble can satisfactorily simulate the mean frequency of drought during the historical period. Conversely, the mean duration, intensity, and maximum deficit of multiyear droughts are underestimated.

Bias correction with QDM improves the simulation of the statistics of the monthly and annual series while it does not show consistent enhancements in capturing the correlation of annual rainfall and the distribution of seasonal rainfall. ~~The latter result is consistent with previous studies (Fauzi et al., 2020; Fauzi et al., 2022)~~ The improvement by QDM to simulate drought characteristics is limited. Indeed, future projections by the multimodel ensemble of multiyear droughts depict a similar risk as historical observations, even after bias correction and adopting the most critical emission scenario.

Our results suggest that validation at the local scale of GCM simulations is an essential step to inform downscaling procedures and correction techniques, to make sure that model predictions are consistent with the local features of climate. However, extreme events like multiyear droughts are unfrequent, and therefore validating their predicted statistics is particularly challenging.

Therefore, the identification of future drought risk, which one would expect to be increased under climate change, remains a challenge, especially if we consider that the reliability of bias correction depends on the availability of observed historical data. ~~GCMs perform fairly well in terms of predicted statistics, with uncertainty that is however expected to depend on the considered design variable. GCMs' predictions for the future generally deliver a worse picture with respect to present day simulations, but our results suggest carefully considering the impact of uncertainty when designing climate change adaptation policies.~~ For some situations, classical engineering methods for critical event estimation under the assumption of stationarity, with appropriate integration of the information provided by climate models to account for climate change, may still be the most precautionary approach. Therefore, rigorous use and comprehensive interpretation of the available information are needed to avoid mismanagement, by also taking into account that the impact of multiyear meteorological droughts is likely to be exacerbated by further pressure on water resources due to increasing population and water demand.

Data availability. The historical rainfall series observed in Bologna can be obtained from https://climexp.knmi.nl/. All the CMIP6 GCM outputs are publicly available from the Copernicus Data Store, https://cds.climate.copernicus.eu/.

Author contributions. AM proposed the main research question and supervised the work. RG made the computational analysis, elaborated additional research ideas and prepared the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. RG was supported by the China Scholarship Council (CSC) Scholarship NO.202106060061. AM received partial funding from the RETURN Extended Partnership, financed by the National Recovery and Resilience Plan – NRRP, Mission 4, Component 2, Investment 1.3 – D.D. 1243 2/8/2022, PE0000005. We thank Dr. Zhiqi Yang and Zhanwei Liu for the helpful discussions. We also thank 4 anonymous reviewers and the Editor for their insightful reviews.

References

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Rupa Kumar, K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D. B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M., and Vazquez-Aguirre, J. L.: Global Observed Changes in Daily Climate Extremes of Temperature and Precipitation, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/10.1029/2005JD006290, 2006.

Aloysius, N. R., Sheffield, J., Saiers, J. E., Li, H., and Wood, E. F.: Evaluation of Historical and Future Simulations of Precipitation and Temperature in Central Africa from CMIP5 Climate Models, Journal of Geophysical Research: Atmospheres, 121, 130–152, https://doi.org/10.1002/2015JD023656, 2016.

Antolini, G., Auteri, L., Pavan, V., Tomei, F., Tomozeiu, R., and Marletto, V.: A daily high-resolution gridded climatic data set for Emilia-Romagna, Italy, during 1961–2010, International Journal of Climatology, 36, 1970–1986, 2016.

Brunetti, M., Buffoni, L., Lo Vecchio, G., Maugeri, M., and Nanni, T.: Tre Secoli Di Meteorologia a Bologna, Cooperativa Universitaria Studio e Lavoro Milan Rep, 2001.

Brunetti, M., Maugeri, M., Monti, F., and Nanni, T.: Temperature and Precipitation Variability in Italy in the Last Two Centuries from Homogenised Instrumental Time Series, International Journal of Climatology, 26, 345–381, https://doi.org/10.1002/joc.1251, 2006.

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.

Casanueva, A., Rodríguez-Puebla, C., Frías, M. D., and González-Reviriego, N.: Variability of Extreme Precipitation over Europe and Its Relationships with Teleconnection Patterns, Hydrology and Earth System Sciences, 18, 709–725, https://doi.org/10.5194/hess-18-709-2014, 2014.

Cook, B. I., Cook, E. R., Smerdon, J. E., Seager, R., Williams, A. P., Coats, S., Stahle, D. W., and Díaz, J. V.: North American Megadroughts in the Common Era: Reconstructions and Simulations, WIREs Climate Change, 7, 411–432, https://doi.org/10.1002/wcc.394, 2016.

Dong, T. and Dong, W.: Evaluation of Extreme Precipitation over Asia in CMIP6 Models, Climate Dynamics, 57, 1751–1769, https://doi.org/10.1007/s00382-021-05773-1, 2021.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) Experimental Design and Organization, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Fauzi, F., Kuswanto, H., and Atok, R.: Bias correction and statistical downscaling of earth system models using quantile delta mapping (QDM) and bias correction constructed analogues with quantile mapping reordering (BCCAQ), in: Journal of Physics: Conference Series, vol. 1538, p. 012050, IOP Publishing, 2020.

Faye, A. and Akinsanola, A. A.: Evaluation of Extreme Precipitation Indices over West Africa in CMIP6 Models, Climate Dynamics, 58, 925–939, https://doi.org/10.1007/s00382-021-05942-2, 2022.

Ge, F., Zhu, S., Peng, T., Zhao, Y., Sielmann, F., Fraedrich, K., Zhi, X., Liu, X., Tang, W., and Ji, L.: Risks of Precipitation Extremes over Southeast Asia: Does 1.5 °C or 2 °C Global Warming Make a Difference?, Environmental Research Letters, 14, 044 015, https://doi.org/10.1088/1748-9326/aaff7e, 2019.

Ge, F., Zhu, S., Luo, H., Zhi, X., and Wang, H.: Future Changes in Precipitation Extremes over Southeast Asia: Insights from CMIP6 Multi-Model Ensemble, Environmental Research Letters, 16, 024 013, https://doi.org/10.1088/1748-9326/abd7ad, 2021.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance Metrics for Climate Models, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2007JD008972, 2008.

Grose, M. R., Narsey, S., Delage, F. P., Dowdy, A. J., Bador, M., Boschat, G., Chung, C., Kajtar, J. B., Rauniyar, S., Freund, M. B., Lyu, K., Rashid, H., Zhang, X., Wales, S., Trenham, C., Holbrook, N. J., Cowan, T., Alexander, L., Arblaster, J. M., and Power, S.: Insights From CMIP6 for Australia's Future Climate, Earth's Future, 8, e2019EF001 469, https://doi.org/10.1029/2019EF001469, 2020.

Ho, S., Tian, L., Disse, M., and Tuo, Y.: A New Approach to Quantify Propagation Time from Meteorological to Hydrological Drought, Journal of Hydrology, 603, 127 056, https://doi.org/10.1016/j.jhydrol.2021.127056, 2021.

IPCC: Summary for Policymakers, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., pp. 3–32, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896.001, 2021.

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, 100 269, https://doi.org/10.1016/j.wace.2020.100269, 2020.

Kling, H., Fuchs, M., and Paulin, M.: Runoff Conditions in the Upper Danube Basin under an Ensemble of Climate Change Scenarios, Journal of Hydrology, 424–425, 264–277, https://doi.org/10.1016/j.jhydrol.2012.01.011, 2012.

Knutti, R. and Sedláček, J.: Robustness and Uncertainties in the New CMIP5 Climate Model Projections, Nature Climate Change, 3, 369–373, https://doi.org/10.1038/nclimate1716, 2013.

Koutroulis, A. G., Grillakis, M. G., Tsanis, I. K., and Papadimitriou, L.: Evaluation of Precipitation and Temperature Simulation Performance of the CMIP3 and CMIP5 Historical Experiments, Climate Dynamics, 47, 1881–1898, https://doi.org/10.1007/s00382-015-2938-x, 2016.

Koutsoyiannis, D. and Montanari, A.: Bluecat: A Local Uncertainty Estimator for Deterministic Simulations and Predictions, Water Resources Research, 58, e2021WR031 215, https://doi.org/10.1029/2021WR031215, 2022.

Kumar, D., Kodra, E., and Ganguly, A. R.: Regional and Seasonal Intercomparison of CMIP3 and CMIP5 Climate Model Ensembles for Temperature and Precipitation, Climate Dynamics, 43, 2491–2518, https://doi.org/10.1007/s00382-014-2070-3, 2014a.

Kumar, S., Lawrence, D. M., Dirmeyer, P. A., and Sheffield, J.: Less Reliable Water Availability in the 21st Century Climate Projections, Earth's Future, 2, 152–160, https://doi.org/10.1002/2013EF000159, 2014b.

Li, Z., Fang, G., Chen, Y., Duan, W., and Mukanov, Y.: Agricultural Water Demands in Central Asia under 1.5 °C and 2.0 °C Global Warming, Agricultural Water Management, 231, 106 020, https://doi.org/10.1016/j.agwat.2020.106020, 2020.

Longmate, J. M., Risser, M. D., and Feldman, D. R.: Prioritizing the Selection of CMIP6 Model Ensemble Members for DownscalingProjections of CONUS Temperature and Precipitation, Preprint available at https://www.researchsquare.com/article/rs-1428854/latest.pdf, https://doi.org/10.21203/rs.3.rs-1428854/v1, 2022.

Lund, J., Medellin-Azuara, J., Durand, J., and Stone, K.: Lessons from California's 2012–2016 Drought, Journal of Water Resources Planning and Management, 144, 04018 067, https://doi.org/10.1061/(ASCE)WR.1943-5452.0000984, 2018.

Mishra, A. K., Singh, V. P., and Desai, V. R.: Drought Characterization: A Probabilistic Approach, Stochastic Environmental Research and Risk Assessment, 23, 41–55, https://doi.org/10.1007/s00477-007-0194-2, 2009.

Nash, J. E. and Sutcliffe, J. V.: River Flow Forecasting through Conceptual Models Part I — A Discussion of Principles, Journal of Hydrology, 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geoscientific Model Development, 9, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016, 2016.

Palerme, C., Genthon, C., Claud, C., Kay, J. E., Wood, N. B., and L'Ecuyer, T.: Evaluation of Current and Projected Antarctic Precipitation in CMIP5 Models, Climate Dynamics, 48, 225–239, https://doi.org/10.1007/s00382-016-3071-1, 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.

Polade, S. D., Pierce, D. W., Cayan, D. R., Gershunov, A., and Dettinger, M. D.: The Key Role of Dry Days in Changing Regional Climate and Precipitation Regimes, Scientific Reports, 4, 4364, https://doi.org/10.1038/srep04364, 2014.

Rivera, J. A. and Arnould, G.: Evaluation of the Ability of CMIP6 Models to Simulate Precipitation over Southwestern South America: Climatic Features and Long-Term Trends (1901–2014), Atmospheric Research, 241, 104 953, https://doi.org/10.1016/j.atmosres.2020.104953, 2020.

Schoof, J. T. and Robeson, S. M.: Projecting Changes in Regional Temperature and Precipitation Extremes in the United States, Weather and Climate Extremes, 11, 28–40, https://doi.org/10.1016/j.wace.2015.09.004, 2016.

Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., and Bronaugh, D.: Climate Extremes Indices in the CMIP5 Multimodel Ensemble: Part 1. Model Evaluation in the Present Climate, Journal of Geophysical Research: Atmospheres, 118, 1716–1733, https://doi.org/10.1002/jgrd.50203, 2013.

Sousa, P. M., Blamey, R. C., Reason, C. J. C., Ramos, A. M., and Trigo, R. M.: The 'Day Zero' Cape Town Drought and the Poleward Migration of Moisture Corridors, Environmental Research Letters, 13, 124 025, https://doi.org/10.1088/1748-9326/aaebc7, 2018.

Stahl, K., Kohn, I., Blauhut, V., Urquijo, J., De Stefano, L., Acácio, V., Dias, S., Stagge, J. H., Tallaksen, L. M., Kampragou, E., Van Loon, A. F., Barker, L. J., Melsen, L. A., Bifulco, C., Musolino, D., de Carli, A., Massarutto, A., Assimacopoulos, D., and Van Lanen, H. A. J.: Impacts of European Drought Events: Insights from an International Database of Text-Based Reports, Natural Hazards and Earth System Sciences, 16, 801–819, https://doi.org/10.5194/nhess-16-801-2016, 2016.

Tabari, H.: Climate Change Impact on Flood and Extreme Precipitation Increases with Water Availability, Scientific Reports, 10, 13 768, https://doi.org/10.1038/s41598-020-70816-2, 2020.

Tabari, H., Hosseinzadehtalaei, P., Thiery, W., and Willems, P.: Amplified Drought and Flood Risk Under Future Socioeconomic and Climatic Change, Earth's Future, 9, e2021EF002 295, https://doi.org/10.1029/2021EF002295, 2021.

Taylor, K. E.: Summarizing Multiple Aspects of Model Performance in a Single Diagram, Journal of Geophysical Research: Atmospheres, 106, 7183–7192, https://doi.org/10.1029/2000JD900719, 2001.

Trenberth, K.: Changes in Precipitation with Climate Change, Climate Research, 47, 123–138, https://doi.org/10.3354/cr00953, 2011.

Ukkola, A. M., De Kauwe, M. G., Roderick, M. L., Abramowitz, G., and Pitman, A. J.: Robust Future Changes in Meteorological Drought in CMIP6 Projections Despite Uncertainty in Precipitation, Geophysical Research Letters, 47, e2020GL087 820, https://doi.org/10.1029/2020GL087820, 2020.

Van Dijk, A. I. J. M., Beck, H. E., Crosbie, R. S., de Jeu, R. A. M., Liu, Y. Y., Podger, G. M., Timbal, B., and Viney, N. R.: The Millennium Drought in Southeast Australia (2001–2009): Natural and Human Causes and Implications for Water Resources, Ecosystems, Economy, and Society, Water Resources Research, 49, 1040–1057, https://doi.org/10.1002/wrcr.20123, 2013.

Vance, T. R., Roberts, J. L., Plummer, C. T., Kiem, A. S., and van Ommen, T. D.: Interdecadal Pacific Variability and Eastern Australian Megadroughts over the Last Millennium, Geophysical Research Letters, 42, 129–137, https://doi.org/10.1002/2014GL062447, 2015.

Wu, J., Chen, X., Love, C. A., Yao, H., Chen, X., and AghaKouchak, A.: Determination of Water Required to Recover from Hydrological Drought: Perspective from Drought Propagation and Non-Standardized Indices, Journal of Hydrology, 590, 125 227, https://doi.org/10.1016/j.jhydrol.2020.125227, 2020.

Xavier, A. C. F., Martins, L. L., Rudke, A. P., de Morais, M. V. B., Martins, J. A., and Blain, G. C.: Evaluation of Quantile Delta Mapping as a bias-correction method in maximum rainfall dataset from downscaled models in São Paulo state (Brazil), International Journal of Climatology, 42, 175–190, 2022.

Xu, K., Xu, B., Ju, J., Wu, C., Dai, H., and Hu, B. X.: Projection and Uncertainty of Precipitation Extremes in the CMIP5 Multimodel Ensembles over Nine Major Basins in China, Atmospheric Research, 226, 122–137, https://doi.org/10.1016/j.atmosres.2019.04.018, 2019.

Yazdandoost, F., Moradian, S., Izadi, A., and Aghakouchak, A.: Evaluation of CMIP6 Precipitation Simulations across Different Climatic Zones: Uncertainty and Model Intercomparison, Atmospheric Research, 250, 105 369, https://doi.org/10.1016/j.atmosres.2020.105369, 2021.

Yevjevich, V.: An Objective Approach to Definitions and Investigations of Continental Hydrologic Droughts, Ph.D. thesis, Hydrology Paper23, Colorado State University, Fort Collins, 1967.

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., Trewin, B., and Zwiers, F. W.: Indices for Monitoring Changes in Extremes Based on Daily Temperature and Precipitation Data, WIREs Climate Change, 2, 851–870, https://doi.org/10.1002/wcc.147, 2011.