Comparison of statistical downscaling methods for climate change impact analysis on drought

Hossein Tabari[1], Daan Buekenhout[1], Patrick Willems[1,2]

[1]KU Leuven, Department of Civil Engineering, Belgium.
5 [2]Vrije Universiteit Brussel, Department of Hydrology and Hydraulic Engineering, Belgium.

Correspondence to: Hossein Tabari (hossein.tabari@kuleuven.be)

Abstract. General circulation models (GCMs) are the primary tools to evaluate the possible impacts of climate change; however, their results are coarse in temporal and spatial dimensions. In addition, they often show systematic biases compared to observations. Downscaling and bias correction of climate model outputs is thus required for local applications. 10 Besides the computationally intensive strategy of dynamical downscaling, statistical downscaling offers a relatively straightforward solution by establishing relationships between small and large scale variables. This study compares four statistical downscaling methods (SDMs) of bias correction (BC), change factor of mean (CFM), quantile perturbation (QP) and event based weather generator (EBWG) to assess climate change impact on drought by the end of the 21st century (2071-2100) relative to a baseline period of 1971-2000. A set of drought related aspects is analysed: dry day frequency, dry 15 spell duration and total precipitation. The downscaling is applied to a 14-member ensemble of CMIP6 GCMs, each powered by four future scenarios of SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5. A 25-member ensemble of CanESM5 GCM is also used to assess the significance of the climate change signals in comparison to the internal variability of the climate. While all methods show a good agreement on downscaling total precipitation, the CFM method fails to downscale dry day frequency well. The QP method outperforms the others in downscaling dry spells. Using this method, dry day frequency is projected to 20 increase significantly in the summer months, with relative changes of up to 20.4% in the worst-case climate change scenario. At the same time, total precipitation is projected to decrease significantly by up to 41.9% in these months. Lastly, extreme dry spells are projected to increase in length by up to 7.4%.

1 Introduction

Our climate system is changing. Since the mid-20*th* century, global warming has been observed (IPCC, 2014). The 25 atmosphere and oceans have warmed, ice and snow volumes have diminished and the sea level has risen. Climate change is linked to a variety of recent weather extremes worldwide. We entered the current decade with Australia's immense bushfires empowered by severe droughts (Phillips, 2020) and devastating mud slides triggered by extreme precipitation in Brazil (Associated Press, 2020). Nature and human communities all over the world are feeling the impact of global warming.

Projections of how global warming will evolve in the coming decades and centuries would therefore be extremely valuable
to mankind in order to adapt efficiently.

Droughts are natural hazards that have an impact on ecological systems and socioeconomic sectors such agriculture,
drinking water supply, waterborne transport, electricity production (hydropower, cooling water) and recreation (Van Loon,
2015; Xie et al., 2018). Quantification of the evolution of droughts on the local level is thus needed to take adequate
mitigation measures. The hydrological processes behind drought are complex, with varying spatial and temporal scales. One
of the aspects of drought is a lack of precipitation. As the projected increase in total precipitation does not systematically
correspond to a decrease in dry days and longest dry spell length (Tabari and Willems, 2018), besides total precipitation, dry
spells and its building blocks, dry days, should be studied to evaluate the impact of climate change on drought. It is clear that
prolonged periods of consecutive dry days can play an important role, for example in replenishing groundwater levels in
time for the dry summer season (Raymond et al., 2019).

Based on observations of more than 5000 rain gauges in the past six decades, Breinl et al. (2020) assessed the historical
evolution of dry spells in the USA, Europe and Australia. Both trends towards shorter and longer dry spells were found,
depending on the location. For Europe, extreme dry spells have become shorter in the North (Scandinavia and parts of
Germany) and longer in the Netherlands and the central parts of France and Spain. Using climate model data, Raymond et al.
(2018, 2019) found a future evolution towards longer dry spells and a larger spatial extent of extreme dry spells in the
Mediterranean basin. For Belgium, Tabari et al. (2015) studied future water availability by comparing precipitation and
potential evapotranspiration. Precipitation and the number of wet days were found to increase during summer and to
decrease during winter, while evapotranspiration was found to increase for both seasons. This suggests drier summers and
wetter winters.

General circulation models (GCMs) are the primary tools for climate change impact assessment. However, they produce
results at a relatively large temporal and spatial scales, the latter varying between 100 and 300 km, and are often found to
show systematic biases in regards to observed data (Ahmed et al., 2019; Song et al., 2020). The bias is especially introduced
to processes that cannot be captured at the climate model's coarse scales (e.g., convective precipitation). These processes are
therefore simplified by means of parametrization, leading to significant bias and uncertainty in the model (Tabari, 2019). In
order to work with these results on finer scales, which is usually required for hydrological impact studies, a downscaling
approach can be applied. Dynamical downscaling is done by creating regional climate models that use the output of a GCM
as boundary conditions and work at much finer scales (< 50 km). This comes at a large computational cost and does not
necessarily account for bias correction (Maraun et al., 2010). An alternative approach is statistical downscaling which
derives statistical relationships between predictor(s) and predictand, e.g. the large-scale historical GCM output and small-
scale observations from weather stations and use them to downscale GCM results with relative ease to assess future local
climate change impact (Ayar et al., 2016).

To meet the demand of high spatiotemporal results for the hydrological impact analysis of climate change, the use of
statistical downscaling methods (SDMs) has recently increased (e.g., Sunyer et al., 2015; Gooré Bi et al., 2017; Smid and

Costa, 2018; Van Uytven, 2019; De Niel et al., 2019; Hosseinzadehtalaei et al., 2020). The results of SDMs are, nevertheless, often compromised with bias and limitations due to assumptions and approximations made within each method
65 (Trzaska and Schnarr, 2014; Maraun et al., 2015). Some of these assumptions cast doubt on the reliability of downscaled projections and may limit the suitability of downscaling methods for some applications (Hall, 2014). As there is no single best downscaling method, the assumptions that led to the final results for different methods require evaluation. Therefore, end users can select an appropriate method based on their strengths and limitations.

This study evaluates the assumptions, strengths and weaknesses of SDMs by a climate change impact analysis for the end
70 of the 21st century (2071-2100) relative to a baseline period of 1971-2000. The four selected SDMs are a bias correction (BC) method, a change factor of mean (CFM) method, a quantile perturbation (QP) method and an event based weather generator (EBWG). A set of drought related aspects is studied: dry day frequency, dry spell length and total precipitation. The downscaling is applied to a 14-member ensemble of global climate models, each powered by four Coupled Model Intercomparison Project Phase 6 (CMIP6) climate change scenarios: SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5. The
75 CMIP6 scenarios are an update to the CMIP5 scenarios, called Representative Concentration Pathways (RCPs), that only project future greenhouse gas emissions, expressed as a radiative forcing level in the year 2100 (e.g., RCP8.5). The CMIP6 scenarios link these radiative forcing levels to socioeconomic narratives (e.g., demography, land-use, energy use), called Shared Socioeconomic Pathways (SSPs; O'Neill et al., 2015). Historical observations from the Uccle weather station are used for the calibration of the SDMs. A 25-member ensemble of CanESM5 GCM is also used to test the significance of the
80 climate change signals.

2 Data and methodology

2.1 Observed and simulated data

The SDMs in this study use precipitation time series produced by GCMs as sole predictor. The predictand is also a precipitation time series, but at the local point scale (scale of a weather station). The availability of a long and high-quality
85 time series of observations from the Uccle weather station enables us to effectively calibrate this relationship. The Uccle station is the main weather station of Belgium, located at the heart of the country (Lat = 50.80°, Lon = 4.35°), and is run by the Royal Meteorological Institute (RMI). Starting in May 1898, the precipitation is being recorded at 10-min intervals with the same instrument, making it one of the longest high-frequency observation time series in the world (Demarée, 2003). In this study, the 10-min observations are aggregated into daily precipitation values, the same temporal scale as the considered
90 GCMs. The information lost by this aggregation is of low interest for studying drought.

Daily precipitation simulations from 14 CMIP6 GCMs are used in this study (Table 1). The data for the grid cell covering Uccle is selected for every GCM using the nearest neighbour algorithm. To give the GCMs in the ensemble an equal weight in the analysis, the one run per model (1R1M) strategy (Tabari et al., 2019) is applied. Only one run, the first initial condition run (r1), is considered for each individual GCM. For one of the GCMs (CanESM5), 25 runs (r1 – r25) are

95 considered in order to allow for quantification of the internal variability in GCM output. To allow for intercomparison of possible futures, multiple scenarios are selected. The four Tier 1 scenarios in ScenarioMIP (CMIP6) are chosen. This set of scenarios covers a wide range of uncertainties in future greenhouse gas forcings coupled to the corresponding socioeconomic developments (O'Neill et al., 2016). On a practical note, the GCM runs for these four scenarios are widely available since they are a basic requirement for participation in CMIP6.

2.2 Statistical downscaling methods

Four SDMs methods were selected, each with a different take on the downscaling of dry spells. The first method utilizes a direct downscaling strategy by applying the relative change factors directly to the dry spell related research indicators. The other methods opt for an indirect downscaling strategy towards dry spells by integrating the changes in dry days, which are downscaled directly into a coherent time series. For this, SDM2 solely relies on the temporal (precipitation) structure present

105 in the GCM time series. SDM3 on the other hand is expected to actively favor clustering of dry days. Lastly, SDM4 makes use of a probability distribution to sample dry events from.

2.2.1 SDM1: Bias correction (BC) of statistics

The first SDM applies a bias correction to the statistics that describe the precipitation time series. Consequently, this method does not return a precipitation time series, unlike the three other SDMs. This method can be regarded as a BC method

110 applied directly to statistics (indicators) instead of to a daily precipitation time series. The indicators used in this study, to which the BC is applied, are discussed later on in subsection 2.3.

An important assumption of all BC methods is that the climate model precipitation bias is time-invariant, which might not be the case (Leander and Buishand, 2007). Furthermore, BC methods assume the temporal structure of wet and dry days of the scenario-projected precipitation by the climate model is accurate. Successive days are also assumed independent.

2.2.2 SDM2: Change factor of mean (CFM) method

The change factor of mean method is frequently applied in the literature. The same simple rationale of the BCM method can be applied by using a change factor approach instead. Here, no correction is applied to GCM precipitation projections. Instead, the relative change between the historical and projection period runs of the GCM is used to calculate a change factor that can then be applied to the observed time series (Sunyer et al., 2012, 2015). The method applied to the precipitation P of

120 day t in month m can be summarized by Eq. 1.

$$P_{m,t}^{Proj} = a_m . P_{m,t}^{Obs} \tag{1.a}$$

in which $a_m = \dfrac{P_{m,...}^{GCMScen}}{P_{m,...}^{GCMHis}}$ (1.b)

In this notation, the precipitation is given for month m and time step t in the observations. Note that observations are often available at smaller time steps then the typical temporal scale of 1 day for GCMs.

SDM2 does not change the number of dry days (NDD) directly. However, since the change factor is applied to all precipitation in a given month, days in the Uccle time series with precipitation values close to the wet day threshold (dry day: $P < 1.0$ mm, wet day: $P \geq 1.0$ mm) can change state, depending on the change factor a_m. The Uccle precipitation time series has a resolution of 0.1 mm. The wet days nearest to the threshold have a value of 1.0 mm, while the closest dry days have a value of 0.9 mm. Consequently, wet days are changed into dry days for $a_m < 1.0$, while a transformation of dry days into wet days requires $a_m > \frac{1.0}{0.9} = 1.11$. In conclusion, SDM2 is expected to show slight changes in terms of dry days, with a bias towards rising the NDD, and thus the dry spells they compose. The mean monthly total precipitation changes projected in this method can be used as a reference for the other methods.

An important assumption made in all CF methods is that the changes at local (weather station) level are the same as the changes described at the spatial, grid-averaged scale of the climate model. Different from the BC methods, the CF methods assume the temporal structure of the observed time series is preserved. Furthermore, the CFM method assumes that all precipitation in a given period (i.e. month or season) is changed by the same factor, regardless of the time step considered or the precipitation intensity observed. In addition, the method assumes consecutive days are independent.

2.2.3 SDM3: Quantile perturbation (QP) method

Quantile perturbation (QP) methods form a more advanced approach to the application of change factors. The core principle of the methods is that the change factors are calculated and allocated based on the exceedance probability of the precipitation intensities. This is opposed to the idea of applying the same change factor to observed precipitation amounts ranging from zero to the most extreme values, as is done in CFM. The QP version applied by Ntegeka et al. (2014) is used here in which the empirical exceedance probabilities p are estimated by making use of the formula $(\frac{k}{n+1})$ for Weibull plotting positions. This approach can change the exceedance probabilities strongly in comparison to the linear interpolation of the cumulative density function represented by $(\frac{k}{n})$ especially for extreme ranks. In literature, this approach is revealed to be suited best for estimating return periods of extreme events (Makkonen, 2006).

Dry day frequency is perturbed by making use of a two-state (wet = 1, dry = 0) second order Markov chain process. This means the states of the two preceding days are taken into account when determining the chance of conversion of a given day. The transition probabilities for the future climate is estimated by applying the change factor principle to these probabilities directly. First, change factors are calculated for all probabilities calculated by dividing those found in the scenario climate model run by those of the historical run. The change factors are then applied to the transition probabilities of the observed time series. Due to the randomness introduced by the dry day perturbation step, multiple time series are generated. A sensitivity analysis is executed by varying the number of simulations (see Text S2 and Figs. S3-S5). The selection of the 'best' simulation is based on four indicators that can be derived for a precipitation time series: the mean (M), coefficient of variability (CV), skewness (S) and average monthly autocorrelation coefficient for a lag of 1 day (ρ_1).

The CF assumptions remain in place for the QP method, as well as the assumption regarding consecutive days as independent. Unlike the CF method, it is now assumed that extreme and non-extreme precipitation amounts can change with different factors. The temporal structure of the observed time series is not explicitly changed. Furthermore, it is assumed that the highest relative changes are applied to the days with the highest daily precipitation. The method allows for an explicit

160 perturbation of the temporal structure of the observed time series.

2.2.4 SDM4: Event based weather generator (EBWG)

Whereas change factor methods discussed earlier apply the same change factor to all precipitation observed on the same day, event based methods identify precipitation events and apply the same change factor to all precipitation within that event. The fourth selected SDM in this study is the stochastic and event based approach developed by Thorndahl et al. (2017) which is

165 not directly based on change factors but generates stochastic time series instead. Consequently, it belongs to the category of the weather generators.

The method constructs a stochastic time series by alternating wet and dry events. The two-component mixed exponential distribution is first fitted to the historical observations. This process is detailed in the supplementary information (Text S1 and Figs. S1-S2). In the second step, the three parameters (λ_a, λ_b and p_a) of this distribution are converted into stochastic

170 variables (sampled from a uniform distribution) in order to accommodate for climate change. A similar approach is used for extreme precipitation, requiring the sampling of two parameters. In total, five parameters are sampled from uniform distributions for each season.

The stochastic nature of this method requires a large number of simulations. These are evaluated using several target variables and the corresponding change factors, which are calculated using the GCM ensemble. For each climate change

175 scenario, one simulation is picked from the accepted simulations as the 'best' simulation, based on the performance it shows for different target variables. This method requires to make an arbitrary choice on several parameters: the boundaries of the uniform sampling intervals, the number of simulations, target variables and their weights. The sampling boundaries for the dry spell parameters and the number of simulations are the subject of a sensitivity analysis (see Text S2 and Fig. S6). The other parameters are further discussed in detail hereafter.

180 *Parameters for precipitation change factor function*: The 2 parameters (slope α and intercept β) of a linear change factor function, used to alter event precipitation amounts in function of its exceedance probability are sampled from uniform distributions. Thorndahl et al. (2017) specify that the sampling boundaries are empirically selected by executing the method for very broad sampling ranges and iteratively narrowing them down based on the simulations that are accepted. When applying this strategy, a test run comprising 50,000 simulations did, however, not show clear boundaries for these

185 parameters. Instead, sampling ranges are chosen at $0.000 - 0.050$ and $0.80 - 1.20$ for α and β respectively, for all seasons. These values correspond well to the parameter ranges found by Thorndahl et al. (2017) for the accepted runs in their study.

Target variables: The performance of a simulation is evaluated based on a set of target variables. This set is altered in comparison to the original implementation in order to fit the specific needs of this study better. Two target variables related

190 to precipitation with T = 2 years and T = 5 years are removed. Instead, five new target variables are added, assuring the annual and seasonal number of dry days is adequately reproduced in the accepted simulations (Table 2). The weights, attributed to each target variable for calculation of the overall performance, are attributed in favor of the dry days target variables in order to reflect their importance for this study. The largest weights are attributed to the target variables that are expected to undergo the largest changes, which are expected to be the hardest to simulate.

Like the other SDMs, some assumptions are made in the EBWG method. It makes assumptions similar to change factor
195 methods due to the selection procedure. The changes found for climate model grid-averaged spatial scales are treated as targets for the stochastic simulations. Furthermore, this weather generator assumes wet event durations will not change, while dry event durations will. In addition, it is assumed that observed time steps with larger precipitation amounts will have a relatively larger increase in precipitation in comparison to time steps with lower precipitation amounts.

2.3 Research indicators

200 In order to compare climate change scenarios and SDMs, five types of research indicators (RIs) are used in this study (Table 3). The most important indicators for this study are related to dry days, dry spells and total precipitation. A typical threshold used for separating wet and dry days is 0.1 mm (Pérez-Sánchez et al., 2018; Breinl et al., 2020). This value corresponds to the standard resolution used for precipitation observations. However, in recent climate change projection studies this threshold is often chosen higher, at 1 mm (Raymond et al., 2018; Tabari and Willems, 2018; Kendon et al., 2019; Han et al.,
205 2019). This is done to counter the tendency of coarse climate models (GCMs) to overestimate the number of days with low precipitation (Tabari and Willems, 2018), the so-called 'drizzle problem' (Moon et al., 2018).

Following the definition used in the climate change study by Raymond et al. (2018), a dry spell is defined as consecutive dry days with less than 1 mm of precipitation. Furthermore, they define several classes of dry spell lengths (Table 4), based on the percentiles of dry spell length calculated using the historical period of the study. Dry spells are not to be confused with the terms *dry events* (Willems, 2013; Willems and Vrac, 2011) or *inter-events* (Sørup et al., 2017; Thorndahl et al.,
210 2017) used in the SDMs. This is due to the definition of dry spells comprising *consecutive* dry days ($\geq$ 2 days). In the discussed SDM implementations, dry events and inter events respectively have minimum lengths of one day and even shorter than one day.

The number of dry days (NDD) is considered on a monthly basis. To assess changes in dry spell patterns, the
215 classification discussed in the literature review by Raymond et al. (2018) is followed. For each of the five classes based on dry spell lengths, the number of dry spells is calculated. An additional indicator gives more information on the class containing the longest dry spells, *very long dry spells* (VLDS). Here, the median VLDS length is used as an indicator. The indicators related to dry spells are calculated over the entire 30-year period to prevent splitting dry spells up. The last indicator used in this research for drought assessment is the mean monthly precipitation (P^{tot}).

220 An additional precipitation indicator (P^{max}) describes the extreme precipitation in a given month m and allows for a rough comparison in terms of extreme precipitation, which is useful to compare how the different SDMs handle extreme precipitation. P^{max} is defined as the monthly maximum daily precipitation averaged over the 30-year period.

2.4 Significance testing of climate change signals

The projected research indicators found after statistical downscaling can be compared to those found in the observed time
225 series. For research indicator i with value I, this climate change signal (CCS_i) is defined as I_i^{Proj} divided by I_i^{Obs}. Something can be said about the significance of the projected CCS in the GCM ensemble by comparing it with the internal variability of one climate model. A significance test is executed based on the Z-score (Tabari et al., 2019). Here, the stochastic variable X represents CCS_i. The null hypothesis of the Z-test corresponds to a situation without climate change: the mean of CCS_i is equal to 1 ($H_0: \mu = 1$). The standard deviation σ can be estimated by the standard deviation of CCS_i found over the 25
230 CanESM5 runs, denoted $s_{i,25}$. The difference between these GCM runs is that they are initialized using different starting conditions, i.e. points in the pre-industrial control run. The differences in CCS for these 25 runs can thus be attributed to the internal variability of the climate system, which is regarded as 'noise'. Consequently, the CCS is said to be significant if the *signal-to-noise* ratio (S2N), here equal to $|Z|$, is sufficiently large. Similar as in Tabari et al. (2019), the Z-test is applied to the median CCS_i over the 14-member GCM ensemble. For a confidence level of 95%, the null hypothesis is rejected if
235 $|Z| = \Phi\left(1 - \frac{0.05}{2}\right) > 1.96$. 10% and 20% significance levels correspond to $Z = 1.64$ and 1.28, respectively. An important assumption in this approach is that $s_{i,25}$ is a representative description for all climate models within the GCM ensemble.

3 Results

Figure 1 shows the number of dry days per month (NDD). The results are characterized by the median of the CMIP6 GCM ensemble. For SDM1 – 3, each member of the ensemble is downscaled separately. As a consequence, the variation within
240 the downscaled ensemble can be looked at as well. This is not possible for SDM4 since it downscales the ensemble as a whole. The median indicator values for SDM1 – 3 show a similar pattern. Across the four scenarios, the NDD increases between June and September in comparison to the Uccle observations. As expected, the increase becomes larger for higher level scenarios. NDD remains about the same for the other months. Application of SDM4 does not capture the larger number of NDD during the summer months well. Instead, NDD remains relatively unchanged throughout the year. Comparison to
245 the results of the CMIP6 GCM projections (before downscaling) learns the Uccle climate has a more gradual evolution of NDD over the year than the GCM projections. The last observation relates to the variation of the SDM results. SDM1 shows the largest variation of NDD within the ensemble, closely followed by SDM3. SDM2 shows a considerably smaller variation. The results for the CMIP6 GCMs (without downscaling), differ quite largely from the downscaled series during the winter months but show good agreement during summer.

250   To discuss the dry spell related RIs, dry spells are categorized by the quantiles of dry spell lengths in the observed (Uccle) time series. Table 4 gives an overview of the limits for each dry spell class. The results for the NDS indicators (number of dry spells per class, over a 30-year period) are summarized in Fig. 2. Results not only vary strongly between SDMs, but between CMIP6 scenarios too. Most results seem to agree that there is an increase in the medium dry spells (NDS_3) in comparison to the observations. The magnitude of these changes is found to increase for higher level scenarios.

255   Long dry spells (NDS_4) are found to increase in general, although the magnitude of this increase remains more or less constant for the different scenarios. More ambiguity is present for NDS_1, NDS_2 and NDS_5, where the nature of the changes (increase or decrease) depends on the considered scenario. Without downscaling, the CMIP6 GCMs generally show a lower number of dry spells than the SDMs across all classes and a higher value for the DSL indicator.

        Next to NDS5, DSL (median VLDS length) is a characteristic of the most extreme dry spells. Results are given in Fig. 2.

260   In comparison to the historical observations, the general trend is towards an increasing value for DSL. Across the different scenarios, the magnitude of the changes only varies slightly.

        The results for P^{tot} (mean monthly precipitation) are given in Fig. 3. Compared to the historical situation, the clearest changes appear in the summer months (June – September) where P^{tot} decreases according to SDM1 – 3. This decrease becomes larger for higher-end scenarios (SSP3-7.0 and SSP5-8.5). A similar trend is found for SDM4 between June and

265   August, especially for the higher-end scenarios. Between October and May, SDM1 – 3 projections show an increase in P^{tot}, although less pronounced than the decrease in the summer months. In terms of variability, SDM1 – 3 also show similar results. The CMIP6 GCM ensemble without downscaling shows higher values for winter season and lower values for summer season in comparison to the downscaled series.

        The second series of RIs related to precipitation is P^{max}. As mentioned earlier, this research indicator does not attribute

270   towards the drought investigation that is the main objective of this study. Rather, the P^{max} indicator is used to gain further insight in the way the selected statistical downscaling methods work, as many statistical downscaling methods are originally created for extreme precipitation studies.

        P^{max}, the maximum daily precipitation on a monthly basis and averaged over the 30-year period, is given in Fig. 4. An interesting observation is that SDM1 – 3 project a very similar and relatively slight increase during winter season, while for

275   summer season the results vary greatly. The largest changes in comparison to the historical period are given by SDM2, where a considerable decrease is found during the summer months. The results of SDM4 are again less similar to the results of the other SDMs. This SDM seems to show a different pattern of changes for each scenario. Generally speaking, the CMIP6 GCM ensemble before downscaling shows relatively similar results during the winter months, while its results for the summer months are slightly lower than the SDM results.

280   The assessment of the significance of the results is based on the relative changes in comparison to the historical observations. In this study, this relative change is defined as the climate change signal (CCS). The median CCS of the GCM ensemble is given in Tables 5 and 6 for the different scenarios, SDMs and RIs. Based on the variation in CCS within the 25

CanESM5 runs (after downscaling), the significance of the median CCS of the ensemble can also be indicated. This is not possible for SDM4 as it does not downscale each member of the ensemble separately.

285   The monthly indicators NDD and P^{tot} mainly show significance for the medium to high level scenarios during the summer months and to a lesser extent during the winter months. The same observation can be made for P^{max}, although significance is only found for SDM2 during the summer months. Furthermore, some of the dry spell related indicators are significant. Again, most significant results are found for SDM2.

4 Discussion

4.1 SDMs

From the results, it is clear that the SDMs can act quite differently. By uncovering where these differences stem from, the performance of the SDMs for drought research can be quantified. Hence, the results for the four SDMs are discussed and linked to the methods' strengths and weaknesses.

4.1.1 SDM1

295 The first method, SDM1, applies a bias correction directly to the research indicators. This means no underlying time series is created. A first consequence is that not all projections are necessarily compatible with each other if the indicators are interdependent. This is the case for NDS since there are only a limited number of dry days to be distributed over the different classes of dry spells.

     Second, the number of extreme events such as LDS and VLDS is limited. In the 30-year period of observations in Uccle, 300 only 20 and 11 occurred, respectively. Furthermore, the number of LDS and VLDS varies substantially among CMIP6 projections (9 – 34 and 3 – 51 under SSP5-8.5). This leads to very large bias correction factors which in turn lead to (over)spectacular results after downscaling (see Fig. 2; NDS_4 and NDS_5). The same problem holds true for the DSL indicator. An absolute bias correction approach instead of a relative one might be more appropriate. In the same spirit, Raymond et al. (2019), discuss changes in extreme dry spell lengths in absolute terms (days) rather than percentages.

305   Note that these concerns do not take away from this method's ability to qualitatively downscale indicators such as NDD or P^{tot}. These indicators are often projected by making use of relative change factors, as is also the case for the other SDMs.

4.1.2 SDM2

SDM2 does not account directly for changes in NDD. This CFM method applies a change factor to the observed time series in order to match the changes in P^{tot}. For this specific research indicator, the result should consequently be no different than 310 the one obtained using SDM1. The slight differences between these methods in Fig. 3 might be attributed to rounding differences.

The rationale behind the application of this method to assess changes in drought finds its roots in the definition of the dry day threshold at 1 mm. As mentioned earlier, this is done to counter for the so-called 'drizzle problem' many GCMs are affected by, meaning they overestimate the number of days with low numbers of precipitation. Consequently, days with

315 precipitation amounts just below this threshold are classified as 'dry', while they might very well be lifted above this threshold in months where P^{tot} is increased by the SDM. Vice versa for days just over the limit changing into dry days in months with a decreasing P^{tot}. Fig. 3 shows this effect quite clearly for the summer months, where P^{tot} is projected to decrease. The relative change in NDD (+7.8% in August) remains, however, rather small in comparison to SDM1 (+13.8%) or SDM3 (+14.8%), which both account for NDD directly.

320 The most interesting aspect in applying this SDM, however, is the lack of vital assumptions as to how the changing NDD affect the dry spells. All required information is contained within the time series created by the GCM. In this light, the general trends for NDS and DSL indicators as projected by SDM3 are interesting to examine, while keeping in mind the underlying changes in NDD are considerably smaller than one would find through a direct change factor approach.

4.1.3 SDM3

325 An important aspect for drought assessment in SDM3 is in the form of the separate dry day perturbation step. Here, the time series is perturbed to match NDD projections. Consequently, SDM3 should be equal to SDM1 in terms of NDD projections. This is not exactly true, as shown in Fig. 1, but the differences are small enough to attribute them to rounding off the results differently. As dry days are the building blocks of dry spells, a solid downscaling approach towards NDD is vital for downscaling NDS and DSL. Out of the four methods considered in this study, SDM3 is the only method that succeeds in

330 downscaling the NDD in an appropriate manner while coherently downscaling the dry spells as well.

4.1.4 SDM4

In several ways, SDM4 seems to be the odd one out among the considered SDMs. The original implementation of this method (Thorndahl et al., 2017) does, for instance, not downscale each member of the CMIP6 GCM ensemble separately as is the case for the other methods. Instead, SDM4 aims to create one time series that corresponds well to the mean of the

335 ensemble, at least in terms of the selected target variables. In theory, an implementation that downscales each member of the GCM ensemble separately is possible. Tests executed in this direction uncovered a practical problem related to the sampling boundaries for parameters governing the dry event duration distribution. As shown in the sensitivity analysis (see Text S2), the SDM struggles to deal with large changes in NDD, e.g. under SSP5-8.5. While the changes in the sensitivity analysis are averaged out over the GCM ensemble, they are not when downscaling each ensemble member separately. The much larger

340 changes that would have to be tackled by the SDM4 would require much larger sampling boundaries. The largest change found in the GCM ensemble (one of the CanESM5 runs under SSP5-8.5) is a decrease of 40% in the number of dry days. To accommodate for this change, sampling boundaries upwards of 70% are required in theory. It is expected that an even larger sampling range is needed, in combination with large numbers of simulations to generate a comfortable number of accepted

simulations. Testing at 40% and 30,000 simulations showed that for many members in the GCM ensemble, no accepted
345   simulations could be generated. This is especially true for the SSP5-8.5 scenario.

For the monthly indicators, NDD and P^{tot}, SDM1 – 3 more or less match the temporal structure found in the Uccle
observations. This is clearly not the case for SDM4. Two reasons can be identified for this. First, the method is implemented
on a seasonal basis, following the original implementation (Thorndahl et al., 2017). Therefore, the method does not try to
match changes in NDD or P^{tot} for every month but rather for the season as a whole. A comparison between a seasonal and a
350   monthly implementation might be interesting to further investigate this method. A monthly implementation is expected to
require larger numbers of simulations in order to achieve similar numbers of accepted simulations. This is due to the larger
number of research indicators present (monthly instead of seasonal). Second, the downscaled time series do not necessarily
match the mean of the GCM ensemble exactly for each research indicator. On the contrary, the method accepts all simulated
time series that remain within the maximum deviation for each target variable (Table 7). These maximum deviations can be
355   very large, e.g. 21.5% for NDD during winter and 40.3% for P^{tot} in summer (both under SSP5-8.5). Consequently,
simulations that are far from the mean projections for some of the key research indicators (e.g. number of dry days) enter
into the pool of accepted simulations and might be selected as the 'best' simulation due to the high performance of the
simulation for other target variables. This explains the difference of SDM4 for NDD (Fig. 1) and P^{tot} (Fig. 3) in comparison
to the SDMs that accurately downscale these indicators (SDM1 & 2 and SDM1 – 3 respectively), even when grouping the
360   results per season (DJF – MAM – JJA – SON).

The inaccurate simulation of NDD affects the dry spell related RIs. It was concluded earlier that this is also the case for
SDM2. An additional concern for this SDM is that only one data point (best simulation) is available for comparison in Fig. 2,
instead of the 14 data points (size of the ensemble) for the other SDMs. While this concern also holds true for the other RIs,
it is mitigated by using these RIs (or similar) as target variables. In order to prevent the problems encountered with a relative
365   bias correction applied directly to the dry spell indicators (see SDM1), this strategy cannot be followed for dry spell related
RIs.

4.2 Significance of climate change signals

The significance of the results is initially introduced to evaluate how the signal (median CCS) compares to the noise present
in the CMIP6 GCM output, before downscaling. These results are implicitly formulated in Table 5 since they are the same as
370   SDM1's. As discussed earlier, only a limited number of research indicators are found to be significant, even at a relatively
low significance level of 20%. The main takeaway from these results is that the increasing number of dry days in the summer
months (between 13.8% and 20.4%) is found to be significant for the high level climate change scenarios. Besides this, the
number of the shortest dry spell class (NDS_1) significantly decreases (up to 10.1%) while the number of long dry spells
(NDS_4) is found to significantly increase (by 86.7%) under SSP3-7.0.

375   The same methodology is followed to assess the significance of the results after downscaling. From the discussion on the
different SDMs, it is clear that not all RIs are necessarily downscaled accurately. The results should thus be interpreted with

care. As mentioned earlier, the main concern for SDM1 is the direct downscaling of the dry spell related RIs, due to the small sample size and the lack of coherence between the projections for the different dry spell classes. As a consequence, the 86.7% increase for NDS_4 is interpreted as an inaccurate result rather than a significant one. For SDM2, it is observed that only P^{tot} is downscaled accurately. The significant results for P^{max} during the summer months should thus be considered as inaccurate. SDM3 on the other hand shows off some interesting results. This method downscales the monthly indicators (NDD, P^{tot} and P^{max}) accurately. Dry spells are not downscaled directly, but by randomly integrating the NDD changes in the original time series. This assures the dry spell related RIs are coherent. As such, the significant 7.4% increase for DSL under SSP5-8.5 and a significance level of 5% is the most interesting results across all SDMs.

4.3 Research indicators

Five different types of research indicators are selected for this research. This subsection shortly evaluates the value of these indicators for this research.

NDD and P^{tot} are both straightforward indicators that are widely used in literature for drought assessment. Both have proven to be useful to compare SDMs and gain insight in these methods since they often rely directly on them. For example, SDM2 is governed solely by P^{tot} while SDM4 directly considers NDD and SDM3 both through its target variables. In this study, both indicators were structured on a monthly basis. It is believed that a seasonal structure could also form a successful alternative.

As for the dry spell indicators, the NDS RIs offer interesting insights into the changes that occur within the dry spell household. The system introduced by Raymond et al. (2018) offers a straightforward but decent classification. Besides the different dry spell class indicators, the DSL indicator is introduced in order to gain further insight into the longest and most important dry spell class and fulfils this role adequately. An indicator describing the most extreme dry spell within the 30-year period could make for an interesting addition in future research.

Last is the P^{max} indicator, defined as the maximum daily precipitation per month averaged over the 30-year period. This indicator does not capture all nuances of extreme precipitation, but gives a rough impression of extreme precipitation changes. In this research, the P^{max} indicator merely functions as a simple illustration on how the SDMs process extreme precipitation differently. It is not a relevant indicator for drought research.

5 Conclusions and recommendations

Four statistical downscaling methods were applied to the CMIP6 GCM ensemble for climate change impact assessment on drought. The main difference is how they treat the downscaling of dry spells. SDM1 uses a bias correction applied directly to the dry spell research indicators, while SDM2 – 4 approach dry spell downscaling indirectly by changing dry day frequency in the precipitation time series. SDM2 (CFM method) uses the information available in the time series ('drizzle') to convert the state (wet or dry) of days that are just below or over the wet day threshold (1 mm/day). SDM3 (QP method) applies

changes in dry day frequency at random places in the time series. The final method, SDM4 (EBWG), samples dry event lengths from a mixed exponential distribution. Other indicators, NDD and P^{tot} are downscaled directly across all methods, except for SDM2, which only takes P^{tot} into account.

The results for SDM1 mirror the relative changes found in the CMIP6 GCM ensemble. While this seems to be a good approach for NDD and P^{tot}, the dry spell related indicators seem to be inflated due to the relative change applied to indicators with low occurrences, e.g. only 11 dry spells with a length over 25 days are observed in the Uccle precipitation time series. SDM2 fails to project NDD correctly. While this might have been expected (NDD is not taken into account during downscaling), this method is tested to see what dry spell patterns are 'hidden' into the original time series. Due to the poor projections of dry day frequency, this method is not fit to evaluate dry spell changes.

Similar to SDM1, SDM3 downscales NDD directly using the change factors found in the CMIP6 GCM ensemble. By altering the time series at random to match the dry day frequency, the dry spells are altered indirectly. Out of the four SDMs used in this study, SDM3 has the best performance in downscaling dry spells. Lastly, the event based weather generator (SDM4) is a complex but potent method. This method uses the relative changes found in the CMIP6 GCM ensemble as targets for NDD and P^{tot}. A rather large deviation from these projections is, however, allowed. This results in a poor downscaling of the changes in dry day frequency and consequently the projections for dry spells are inaccurate, despite the interesting approach this SDM offers towards dry spells (mixed exponential distribution). Stricter selection criteria and more optimized target variables should improve this method's performance, likely at a larger computational cost.

These conclusions put the results of SDM3 in the spotlight. The most important finding is that the number of dry days is projected to increase significantly in the summer months, from June until September. The magnitude and significance level of this increase depend on the considered climate change scenario: between 7.4% and 10.6% for SSP2-4.5 (significance level $\alpha = 10\%$), between 12.0% and 15% for SSP3-7.0 ($\alpha = 5\%$) and between 13.6% and 20.4% for SSP5-8.5% ($\alpha = 5\%$). Furthermore, the DSL is found to significantly increase by 7.4% for SSP5-8.5 ($\alpha = 5\%$).

The event based weather generator (SDM4) offers ample opportunity for further improvement. The method could be structured per month instead of per season to capture month-to-month variation to match the other methods. Application of the method to each GCM in the ensemble would create more data points, allowing the quantification of the significance of the results found by using this method. Furthermore, alterations could be made to the acceptance criterion in order to lower the allowed deviations from the changes projected by the GCMs. This is especially important for accurate simulations of the number of dry days. With the same goal in mind, the mix of target variables and their corresponding weights could be changed (e.g., only target variables related to dry days). Furthermore, different dry event duration distributions (Weibull, exponential, gamma, generalized Pareto...) can be considered besides the mixed exponential distribution that is used in the original implementation and in this research.

There is also room for new downscaling methods that are optimized to deal with dry spells. For example, a method that uses quantile mapping to assess dry spell changes (similar to precipitation downscaling in the QP method) could make for an

interesting comparison to the other methods. In addition, a method that applies absolute changes to the dry spell indicators could be studied.

Several research indicators can be used to assess the SDMs for the impact analysis of climate change on drought. In combination with P^{tot}, one could consider evapotranspiration to assess water availability. Furthermore, additional indicators can be used to study dry spells. Besides the median length of very long dry spells, the maximum dry spell length over a certain period can also be of interest. Furthermore, the temporal behavior of dry spells could be studied, for example based on their starting, ending or middle day. This might be especially useful to assess the impact of dry spells during the wet season, when water tables have to be replenished in order to bridge the dry summer season.

For climate change impact assessment, this study made use of the CMIP6 GCMs. CMIP6 is an ongoing project, not all participating research institutes have completed and uploaded their GCM runs yet. Future studies should keep an eye on the CMIP6 database (ESGF archive) for additional suitable candidates to enlarge the GCM ensemble. As the size of the used CMIP6 GCM ensemble is large enough to be exempt from the small sample size issue for climate change studies (Hosseinzadehtalaei et al., 2017), similar results are expected to be found using a larger GCM ensemble. However, including more GCMs boosts the confidence to the results and allows a more comprehensive uncertainty analysis.

Data availability. CMIP6 GCM data used in the study are freely available at the ESGF website (https://esgf-index1.ceda.ac.uk). The Uccle historical precipitation time series were provided by the Royal Meteorological Institute (RMI) of Belgium (www.meteo.be)

Author contribution.
All authors collaboratively conceived the idea and conceptualized the methodology. DB carried out the analysis. HT and DB wrote the initial draft of the paper. All authors discussed the results and edited the paper.

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

References

Ahmed, K., Sachindra, D. A., Shahid, S., Demirel, M. C., and Eun-Sung, C.: Selection of multi-model ensemble of general circulation models for the simulation of precipitation and maximum and minimum temperature based on spatial assessment metrics, Hydrol. Earth Syst. Sci., 23(11), 4803-4824, doi:10.5194/hess-23-4803-2019, 2019.

Associated Press. Scores dead as heavy rains bring landslides and evacuations in brazil. Online: https://www.theguardian.com/world/2020/jan/27/heavyrain-landslides-flooding-mass-evacuations-brazil, January 2020. Accessed: 2020-08-14.

Ayar, P. V., Vrac, M., Bastin, S., Carreau, J., Déqué, M., and Gallardo, C.: Intercomparison of statistical and dynamical downscaling models under the EURO-and MED-CORDEX initiative framework: present climate evaluations, Clim. Dyn., 46(3-4), 1301-1329, doi:10.1007/s00382-015-2647-5, 2016.

475   Breinl, K., Di Baldassarre, G., Mazzoleni, M., Lun, D., and Vico, G.: Extreme dry and wet spells face changes in their duration and timing, Environ. Res. Lett., doi:10.1088/1748-9326/ab7d05, 2020.

De Niel, J., Van Uytven, E., and Willems, P.: Uncertainty analysis of climate change impact on river flow extremes based on a large multi-model ensemble, Water Resour. Manage., 33(12), 4319–4333, doi:10.1007/s11269-019-02370-0, 2019.

Demarée, G. R. Le pluviographe centenaire du plateau d'Uccle: Son histoire, ses données et ses applications. Houille
480   Blanche, (4):95–102, 2003.

Gooré Bi, E., Gachon, P., Vrac, M., and Monette, F.: Which downscaled rainfall data for climate change impact studies in urban areas? Review of current approaches and trends, Theor. Appl. Climatol., 127(3-4), 685–699, doi: 10.1007/s00704-015-1656-y, 2017.

Hall, A.: Projecting regional change, Science 346, 1461–1462. doi:10.1126/science.aaa0629, 2014.

485   Han, F., Cook, K. H., and Vizy, E. K.: Changes in intense rainfall events and dry periods across Africa in the twenty-first century, Clim. Dyn., 53(5-6), 2757–2777, doi:10.1007/s00382-019-04653-z, 2019.

Hosseinzadehtalaei, P., Tabari, H., and Willems, P.: Climate change impact on short-duration extreme precipitation and intensity–duration–frequency curves over Europe, J. Hydrol., 590, 125249, doi:10.1016/j.jhydrol.2020.125249, 2020.

Hosseinzadehtalaei, P., Tabari, H., and Willems, P.: Uncertainty assessment for climate change impact on intense
490   precipitation: how many model runs do we need?, Int. J. Climatol., 37, 1105–1117, doi:10.1002/joc.5069, 2017.

IPCC. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]*, pages 1–151. IPCC, Geneva, Switzerland, 2014.

Kendon, E. J., Stratton, R. A., Tucker, S., Marsham, J. H., Berthou, S., Rowell, D. P., and Senior, C. A.: Enhanced future
495   changes in wet and dry extremes over Africa at convection-permitting scale, Nat. Commun., 10(1), 1794, doi: 10.1038/s41467-019-09776-9, 2019.

Leander, R. and Buishand, T. A.: Resampling of regional climate model output for the simulation of extreme river flows, J. Hydrol., 332:487–496, doi:10.1016/j.jhydrol.2006.08.006, 2007.

Makkonen, L.: Plotting positions in extreme value analysis, J. Appl. Meteor. Climatol., 45(2):334–340,
500   doi:10.1175/JAM2349.1, 2006.

Maraun, D., et al.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48(3), RG3003, doi:10.1029/2009RG000314, 2010.

Maraun, D., et al.: VALUE: A framework to validate downscaling approaches for climate change studies, Earth's Future, 3, 1–14, doi:10.1002/2014EF000259, 2015.

505 Moon, H., Gudmundsson, L., and Seneviratne, S. I.: Drought persistence errors in global climate models, J. Geophys. Res. Atmos., 123(7), 3483–3496, doi:10.1002/2017JD027577, 2018.

Ntegeka, V., Baguis, P., Roulin, E., and Willems, P.: Developing tailored climate change scenarios for hydrological impact assessments, J. Hydrol., 508, 307–321, doi:10.1016/j.jhydrol.2013.11.001, 2014.

O'Neill, B. C., et al.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geosci. Model Dev., 9(9),
510   3461–3482, doi:10.5194/gmd-9-3461-2016, 2016.

Pérez-Sánchez, J., and Senent-Aparicio, J.: Analysis of meteorological droughts and dry spells in semiarid regions: a comparative analysis of probability distribution functions in the Segura Basin (SE Spain), Theor. Appli. Climatol., 133(3-4), 1061–1074, doi:10.1007/s00704-017-2239-x, 2018.

Phillips, N.: Climate change made Australia's devastating fire season 30% more likely. Online:
515   https://www.nature.com/articles/d41586-020-00627-y, March 2020. Accessed: 2020-08-14.

Raymond, F., Drobinski, P., Ullmann, A., and Camberlin, P.: Extreme dry spells over the Mediterranean Basin during the wet season: Assessment of HyMeX/Med-CORDEX regional climate simulations (1979–2009), Int. J. Climatol., 38(7), 3090–3105, doi:10.1002/joc.5487, 2018.

Raymond, F., Ullmann, A., Tramblay, Y., Drobinski, P., and Camberlin, P.: Evolution of Mediterranean extreme dry spells
520   during the wet season under climate change, Reg. Environ. Change, 19(8), 2339–2351, doi:10.1007/s10113-019-01526-3, 2019.

Smid, M., and Costa, A. C.: Climate projections and downscaling techniques: a discussion for impact studies in urban systems, Int. J. Urban Sci., 22(3), 277–307, doi:10.1080/12265934.2017.1409132, 2018.

Song, J. H., et al.: Evaluating the performance of climate models in reproducing the hydrological characteristics of rainfall
525   events, Hydrol. Sci. J., 65(9), 1490-1511, doi:10.1080/02626667.2020.1750616, 2020.

Sørup, H. J. D., Georgiadis, S., Bülow Gregersen, I., and Arnbjerg-Nielsen, K.: Formulating and testing a method for perturbing precipitation time series to reflect anticipated climatic changes. Hydrol. Earth Syst. Sci., 21(1), 345–355, doi:10.5194/hess-21-345-2017, 2017.

Sunyer, M. A., et al.: Inter-comparison of statistical downscaling methods for projection of extreme precipitation in Europe,
530   Hydrol. Earth Syst. Sci., 19(4), 1827–1847, doi:10.5194/hess-19-1827-2015, 2015.

Sunyer, M. A., Madsen, H., and Ang, P. H.: A comparison of different regional climate models and statistical downscaling methods for extreme rainfall estimation under climate change, Atmos. Res., 103, 119–128, doi:10.1016/j.atmosres.2011.06.011, 2012.

Tabari, H., and Willems, P.: More prolonged droughts by the end of the century in the Middle East, Environ. Res. Lett.,
535   13(10), doi:10.1088/1748-9326/aae09c, 2018.

Tabari, H., Hosseinzadehtalaei, P., Aghakouchak, A., and Willems, P.: Latitudinal heterogeneity and hotspots of uncertainty in projected extreme precipitation, Environ. Res. Lett., 14(12), doi:10.1088/1748-9326/ab55fd, 2019.

Tabari, H., Taye, M. T., and Willems, P.: Water availability change in central Belgium for the late 21st century, Global Planet. Change, 131, 115–123, doi:10.1016/j.gloplacha.2015.05.012, 2015.

540 Tabari, H.: Statistical analysis and stochastic modelling of hydrological extremes, Water, 11(9), 1–11, doi:10.3390/w11091861, 2019.

Thorndahl, S., Andersen, A. K., and Larsen, A. B.: Event-based stochastic point rainfall resampling for statistical replication and climate projection of historical rainfall series, Hydrol. Earth Syst. Sci., 21, 4433–4448, doi:10.5194/hess-21-4433-2017, 2017.

545 Trzaska, S., and Schnarr, E.: A review of downscaling methods for climate change projections, United States Agency for International Development by Tetra Tech ARD, 1-42, 2014.

Van Loon, A. F.: Hydrological drought explained, WIRES Water, 2(4), 359–392, doi:10.1002/wat2.1085, 2015.

Van Uytven, E.: Evaluation of statistical downscaling methods for climate change impact analysis on hydrological extremes, PhD thesis, KU Leuven, 2019.

550 Willems, P., and Vrac, M.: Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change, J. Hydrol., 402(3-4), 193–205, doi:10.1016/j.jhydrol.2011.02.030, 2011.

Willems, P.: Revision of urban drainage design rules after assessment of climate change impacts on precipitation extremes at Uccle, Belgium, J. Hydrol., 496, 166–177, doi:10.1016/j.jhydrol.2013.05.037, 2013.

Xie, W., Xiong, W., Pan, J., Ali, T., Cui, Q., Guan, D., Meng, J., Mueller, N. D., Lin, E., and Davis, S. J.: Decreases in
555 global beer supply due to extreme drought and heat, Nat. Plants, 4(11), 964–973, doi:10.1038/s41477-018-0263-1, 2018.
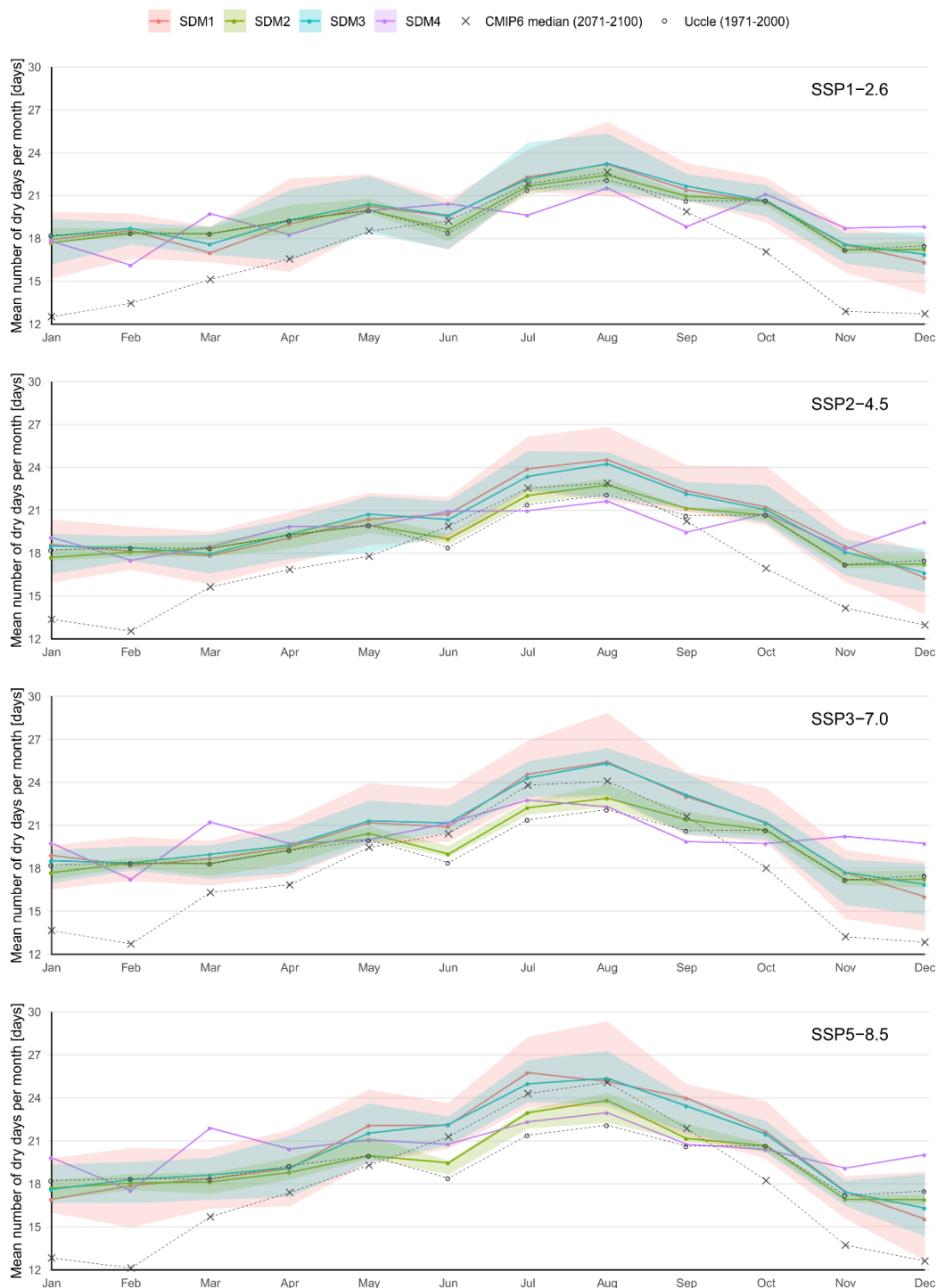
Figure 1. Graphical representation of NDD results. Coloured lines represent median values of the ensemble, shades represent the variation within the ensemble (10% – 90% quantiles). CMIP6 GCM projections (not downscaled) and Uccle observations are given as reference.
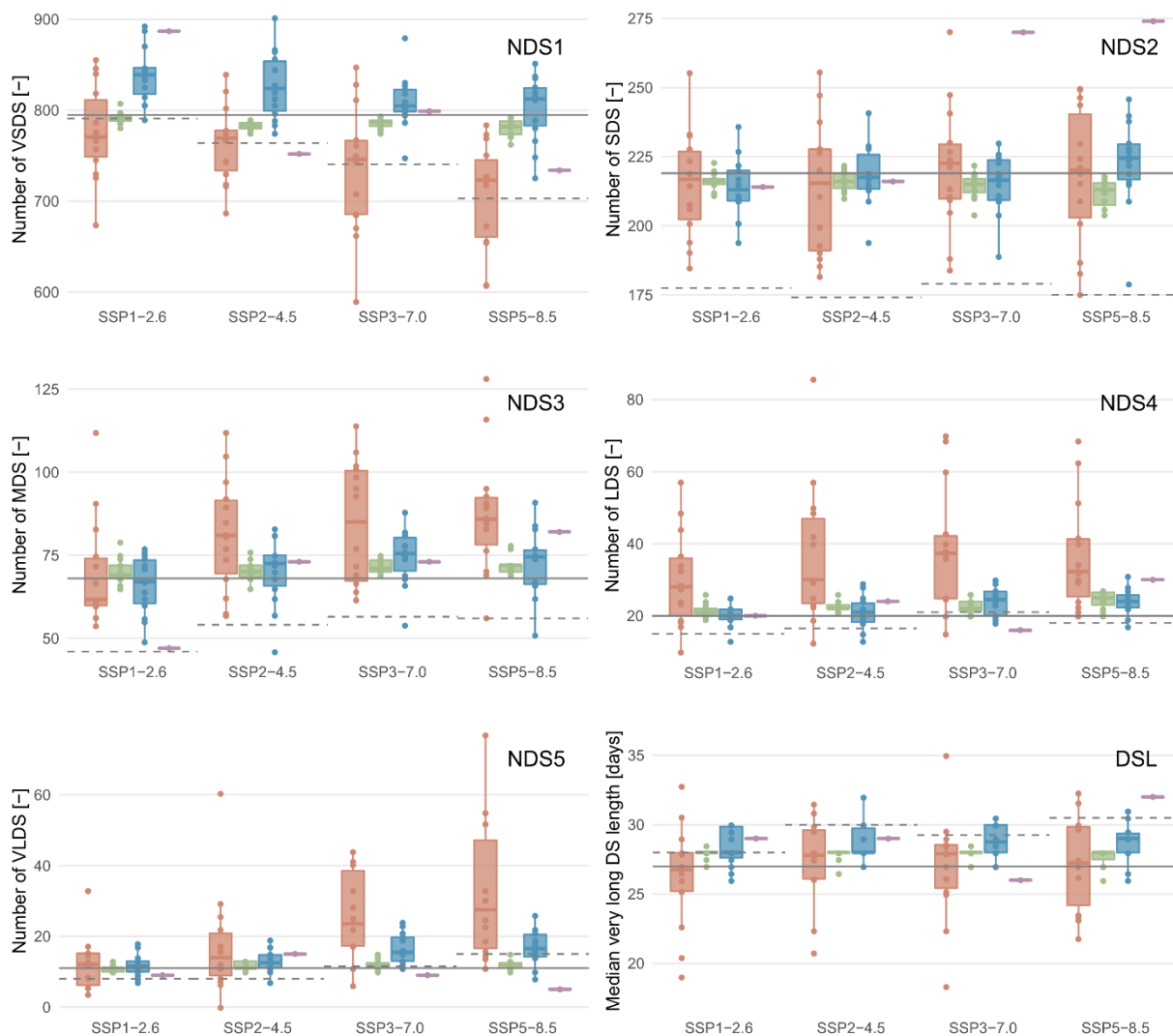
Figure 2. Boxplot representation of NDS and DSL results. SDM4 downscales the ensemble as a whole, resulting in only one data point. CMIP6 GCM projections (not downscaled) and Uccle observations are given as reference.
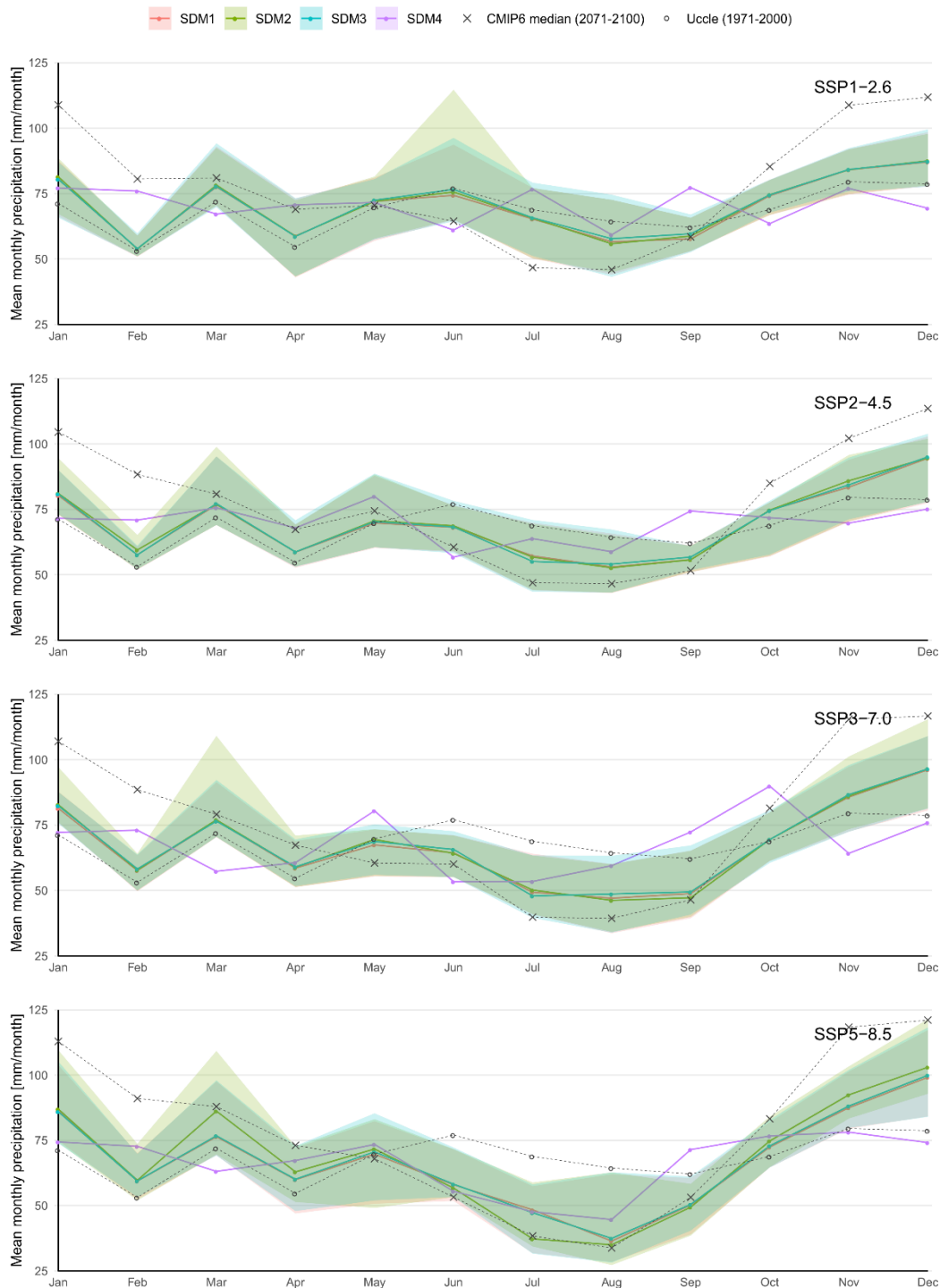
Figure 3. Graphical representation of P^tot results. Coloured lines represent median values of the ensemble, shades represent the variation within the ensemble (10% – 90% quantiles). CMIP6 GCM projections (not downscaled) and Uccle observations are given as reference.
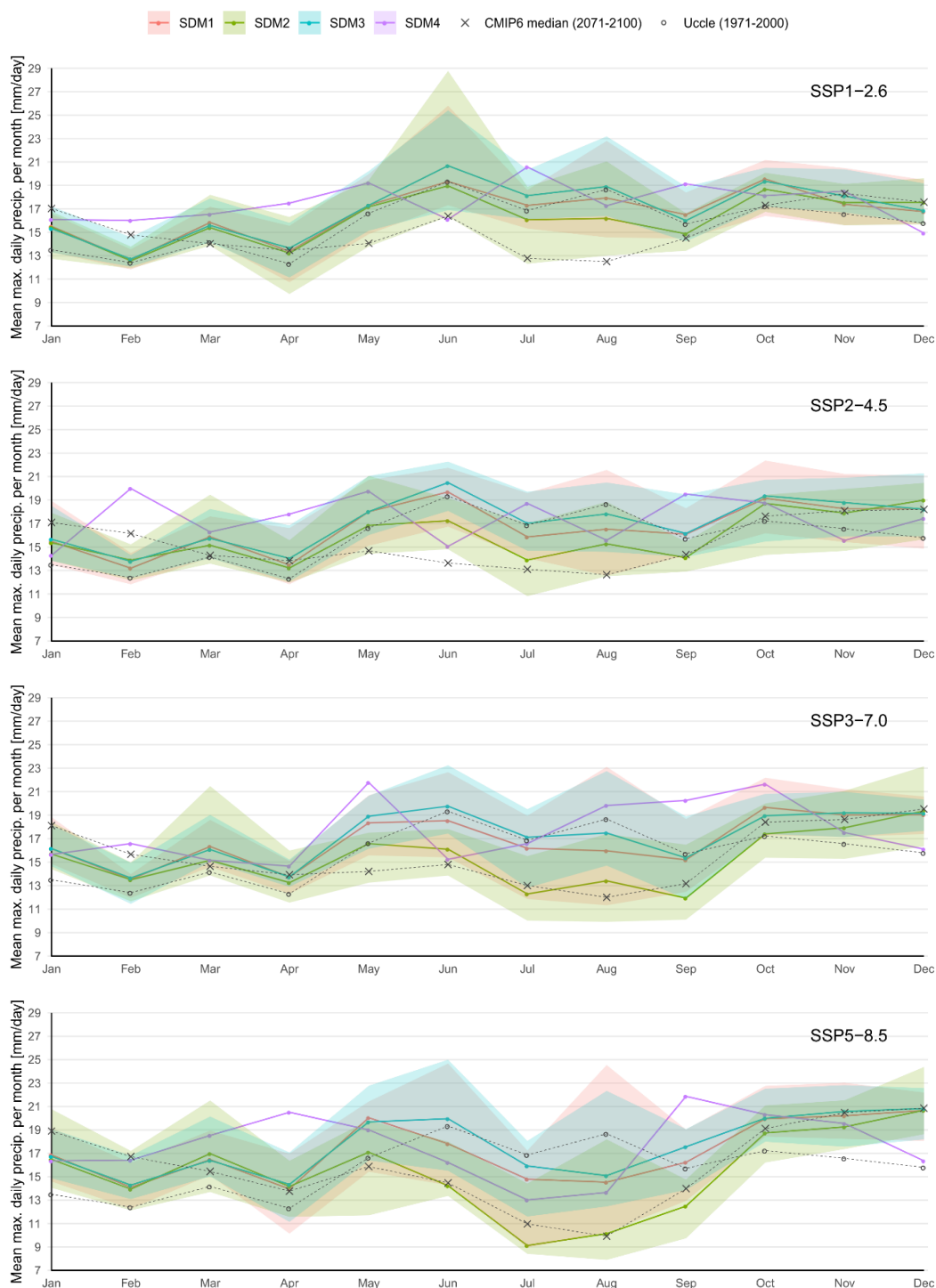
Figure 4. Graphical representation of P^{max} results. Coloured lines represent median values of the ensemble, shades represent the variation within the ensemble (10% – 90% quantiles). CMIP6 GCM projections (not downscaled) and Uccle observations are given as reference.

Table 1. Overview of the CMIP6 GCM ensemble.

Model	Resolution		His.	SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP5-8.5
	Lat [°]	Lon [°]					
ACCESS-ESM1-5	1.3	1.9	1	1	1	1	1
BCC-CSM2-MR	1.1	1.1	1	1	1	1	1
CanESM5	2.8	2.8	25	25	25	25	25
CESM2-WACCM	0.9	1.3	1	1	1	1	1
EC-Earth3-Veg	0.7	0.7	1	1	1	1	1
FGOALS-g3	2.0	2.0	1	1	1	1	1
GFDL-ESM4	1.0	1.3	1	1	1	1	1
INM-CM5-0	1.5	2.0	1	1	1	1	1
IPSL-CM6A-LR	1.3	2.5	1	1	1	1	1
KACE-1-0-G	1.3	1.9	1	1	1	1	1
MIROC6	1.4	1.4	1	1	1	1	1
MPI-ESM1-2-HR	0.9	0.9	1	1	1	1	1
MRI-ESM2-0	1.1	1.1	1	1	1	1	1
NorESM2-MM	0.9	1.3	1	1	1	1	1

Table 2. Target variables used for evaluation of the simulations of SDM4.

Target variable	Abbr.	Unit	Weight
Annual number of dry days	and	–	–
Seasonal number of dry days, winter	sndwi	–	0.10
Seasonal number of dry days, spring	sndsp	–	0.10
Seasonal number of dry days, summer	sndsu	–	0.20
Seasonal number of dry days, autumn	sndau	–	0.10
Annual precipitation	ap	mm	–
Seasonal precipitation, winter	spwi	mm	0.03
Seasonal precipitation, spring	spsp	mm	0.06
Seasonal precipitation, summer	spsu	mm	0.15
Seasonal precipitation, autumn	spau	mm	0.06
Annual number of events above 10 mm per day	n10mm	–	0.10
Annual number of events above 20 mm per day	n20mm	–	0.05
Annual maximum daily precipitation	mdp	mm	0.05

Table 3. Overview of the considered research indicators.

Research indicators	Abbr.	#
Mean monthly number of dry days	NDD_m	12
Number of dry spells per class	NDS_c	5
Median VLDS length	DSL	1
Mean monthly precipitation	P_m^{tot}	12
Maximum monthly precipitation	P_m^{max}	12
		42

Table 4. Classification of dry spells based on their length along with the limits for each class derived from observed time series.

Class name	Abbr.	Research indicator	Percentiles	Limits [days]
Very short dry spell	VSDS	NDS_1	< 20th	[2, 7]
Short dry spell	SDS	NDS_2	20th – 40th	[8, 13]
Medium dry spell	MDS	NDS_3	40th – 60th	[14, 19]
Long dry spell	LDS	NDS_4	60th – 80th	[20, 25]
Very long dry spell	VLDS	NDS_5	> 80th	[26, ∞]

Table 5. CCS and corresponding significance for SDM1 and SDM2. CCS is the change relative to the historical observations (1971-2000). Numbers in italic, bold and bold italic denote significant changes at 20%, 10% and 5% levels, respectively.

Research indicator		Obs.	SDM1				SDM2			
			SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP5-8.5	SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP5-8.5
DSL		27	-0.90%	2.90%	3.40%	0.90%	*3.70%*	**3.70%**	*3.70%*	*3.70%*
NDS$_c$	1	795	-3.10%	-3.20%	-6.20%	***-9.00%***	-0.60%	***-0.80%***	-0.90%	*-0.30%*
	2	219	-1.00%	-1.70%	1.60%	0.50%	-1.40%	-1.40%	-1.80%	***-2.70%***
	3	68	-9.40%	19.00%	25.00%	*26.20%*	1.50%	2.90%	*4.40%*	*5.90%*
	4	20	*40.20%*	50.00%	***86.70%***	*61.30%*	5.00%	**10.00%**	*10.00%*	***25.00%***
	5	11	9.80%	27.20%	114.30%	150.00%	0.00%	0.00%	9.10%	9.10%
NDD$_m$	Jan	18	-1.70%	2.00%	3.70%	*-7.20%*	**-2.90%**	*-2.90%*	**-2.90%**	*-2.90%*
	Feb	18	1.00%	-1.30%	-0.90%	-2.60%	0.00%	-1.60%	0.00%	-1.60%
	Mar	18	-7.30%	-2.70%	1.80%	0.20%	0.00%	0.00%	0.00%	-1.10%
	Apr	19	-1.30%	-1.00%	1.10%	-1.00%	0.00%	0.00%	0.00%	-2.40%
	May	20	1.50%	2.00%	6.10%	**10.60%**	0.00%	0.00%	2.30%	0.00%
	Jun	18	6.20%	***12.70%***	***13.60%***	***20.10%***	1.30%	***3.30%***	3.30%	***5.80%***
	Jul	21	4.20%	***11.70%***	***14.80%***	***20.40%***	1.20%	***3.00%***	***3.90%***	***7.30%***
	Aug	22	5.00%	11.00%	***15.00%***	***13.80%***	1.50%	*3.00%*	*3.60%*	***7.80%***
	Sep	21	3.70%	*8.40%*	**11.50%**	***16.30%***	*1.60%*	**2.40%**	*3.70%*	2.60%
	Oct	21	-0.60%	2.70%	2.60%	4.70%	0.00%	0.00%	0.00%	0.00%
	Nov	17	2.10%	7.40%	2.90%	1.50%	0.00%	0.00%	0.00%	-1.60%
	Dec	18	-6.70%	-6.90%	-8.40%	**-11.10%**	-1.50%	-1.50%	-1.50%	***-3.40%***
P_m^{tot}	Jan	71.2	*13.00%*	12.70%	*14.20%*	**21.10%**	*14.30%*	13.80%	*16.20%*	22.10%
	Feb	53.1	1.60%	8.50%	8.50%	11.80%	1.30%	12.00%	8.90%	12.40%
	Mar	72	7.80%	6.70%	6.60%	6.20%	8.70%	7.00%	7.00%	*19.80%*
	Apr	54.7	7.60%	7.30%	7.10%	9.60%	7.40%	7.40%	7.70%	15.10%
	May	69.7	3.30%	0.10%	-3.30%	0.00%	3.30%	1.30%	-0.30%	2.80%
	Jun	77.1	-3.50%	*-11.60%*	-16.40%	**-24.70%**	-1.90%	-10.80%	-16.70%	**-26.40%**
	Jul	68.9	-5.20%	**-16.90%**	***-28.30%***	***-29.70%***	-4.70%	**-17.60%**	***-27.10%***	***-45.90%***
	Aug	64.4	-12.10%	*-17.70%*	***-26.90%***	***-43.50%***	-13.30%	*-18.20%*	***-28.20%***	***-45.70%***
	Sep	62.1	-7.10%	-10.20%	***-21.40%***	***-19.10%***	-5.40%	-10.30%	***-24.00%***	***-20.50%***
	Oct	68.8	7.90%	8.50%	1.10%	5.30%	8.40%	8.60%	1.00%	8.70%
	Nov	79.6	5.70%	4.90%	7.70%	10.00%	5.80%	8.00%	8.20%	16.20%
	Dec	78.7	**10.60%**	*20.10%*	*22.20%*	*26.00%*	*11.20%*	*20.30%*	*22.60%*	*30.90%*
P_m^{max}	Jan	13.5	**14.90%**	**14.60%**	*19.60%*	*24.80%*	*14.30%*	13.80%	*16.20%*	*22.10%*
	Feb	12.4	1.90%	6.50%	9.20%	13.70%	1.30%	12.00%	8.90%	12.40%
	Mar	14.2	*11.80%*	12.10%	*15.30%*	16.20%	8.70%	7.00%	7.00%	*19.80%*
	Apr	12.3	8.60%	9.50%	11.50%	13.40%	7.40%	7.40%	7.70%	15.10%
	May	16.6	4.10%	8.30%	10.40%	*20.50%*	3.30%	1.30%	-0.30%	2.80%
	Jun	19.3	0.00%	1.90%	-4.10%	-7.80%	-1.90%	-10.80%	-16.70%	**-26.40%**
	Jul	16.9	2.60%	-5.90%	-4.10%	-12.30%	-4.70%	**-17.60%**	***-27.10%***	***-45.90%***
	Aug	18.7	-4.00%	-11.50%	-14.50%	-22.20%	-13.30%	*-18.20%*	***-28.20%***	***-45.70%***
	Sep	15.7	5.00%	2.40%	-3.30%	3.30%	-5.40%	-10.30%	***-24.00%***	***-20.50%***
	Oct	17.2	13.40%	11.20%	14.10%	15.80%	8.40%	8.60%	1.00%	8.70%
	Nov	16.6	5.00%	10.20%	14.80%	*21.90%*	5.80%	8.00%	8.20%	16.20%
	Dec	15.8	6.10%	15.60%	**20.60%**	*30.80%*	*11.20%*	*20.30%*	*22.60%*	*30.90%*

Table 6. CCS for SDM3 and SDM4 and corresponding significance for SDM3. CCS is the change relative to the historical observations (1971-2000). Numbers in italic, bold and bold italic denote significant changes at 20%, 10% and 5% levels, respectively.

Research indicator		Obs.	SDM3				SDM4			
			SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP5-8.5	SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP5-8.5
DSL		27	3.70%	3.70%	*6.50%*	**7.40%**	7.40%	7.40%	-3.70%	18.50%
NDS_c	1	795	**5.50%**	3.60%	1.30%	2.20%	11.60%	-5.40%	0.50%	-7.70%
	2	219	-2.70%	-0.70%	-1.10%	2.50%	-2.30%	-1.40%	23.30%	25.10%
	3	68	-1.50%	6.60%	11.00%	9.60%	-30.90%	7.40%	7.40%	20.60%
	4	20	0.00%	5.00%	22.50%	20.00%	0.00%	20.00%	-20.00%	50.00%
	5	11	4.50%	13.60%	*40.90%*	*50.00%*	-18.20%	36.40%	-18.20%	-54.50%
NDD_m	Jan	18	-0.50%	1.50%	1.60%	-3.40%	-2.40%	4.80%	8.40%	8.80%
	Feb	18	1.90%	0.10%	0.10%	-0.40%	-12.20%	-4.70%	-6.20%	-4.50%
	Mar	18	-4.00%	-2.10%	3.50%	1.50%	7.60%	0.70%	15.80%	19.50%
	Apr	19	0.10%	0.30%	1.80%	-0.70%	-5.20%	3.10%	2.40%	6.10%
	May	20	2.20%	3.80%	6.80%	7.90%	-0.20%	-0.50%	0.20%	5.50%
	Jun	18	6.60%	*10.60%*	*15.00%*	*20.40%*	11.10%	13.80%	15.00%	12.90%
	Jul	21	3.60%	*9.20%*	*13.60%*	*16.70%*	-8.30%	-2.00%	6.40%	4.40%
	Aug	22	*5.20%*	**9.70%**	**14.60%**	**14.80%**	-2.60%	-2.10%	0.90%	3.90%
	Sep	21	*5.00%*	**7.40%**	**12.00%**	**13.60%**	-8.70%	-5.70%	-3.70%	0.60%
	Oct	21	-0.20%	1.80%	2.30%	3.90%	2.10%	0.50%	-4.50%	-1.30%
	Nov	17	2.20%	5.00%	3.00%	1.30%	8.90%	6.20%	17.60%	11.00%
	Dec	18	-3.50%	-5.00%	-3.50%	*-6.80%*	7.60%	15.20%	12.80%	14.50%
P_m^{tot}	Jan	71.2	*13.00%*	13.50%	*15.60%*	**20.70%**	8.30%	0.50%	1.40%	4.40%
	Feb	53.1	1.60%	8.50%	9.70%	12.00%	43.00%	33.50%	37.60%	36.90%
	Mar	72	8.10%	7.30%	6.30%	6.70%	-6.60%	5.00%	-20.40%	-12.30%
	Apr	54.7	7.20%	7.40%	8.00%	10.00%	29.40%	24.20%	10.90%	23.10%
	May	69.7	4.00%	0.80%	-1.40%	1.30%	2.70%	14.60%	15.40%	5.30%
	Jun	77.1	-0.60%	-11.40%	-14.70%	**-24.40%**	-20.70%	-26.40%	-30.80%	-28.00%
	Jul	68.9	-4.50%	**-20.00%**	**-30.40%**	**-31.40%**	11.40%	-7.30%	-22.40%	-30.80%
	Aug	64.4	-10.30%	-16.00%	**-24.40%**	**-41.90%**	-8.00%	-8.70%	-7.60%	-30.70%
	Sep	62.1	-4.00%	-8.60%	**-20.40%**	**-18.80%**	24.40%	19.80%	16.50%	15.00%
	Oct	68.8	8.30%	8.30%	1.00%	6.00%	-7.60%	4.40%	30.60%	11.50%
	Nov	79.6	5.90%	6.00%	8.90%	10.80%	-3.20%	-12.30%	-19.30%	-1.80%
	Dec	78.7	**10.90%**	**20.80%**	**22.60%**	**27.00%**	-11.80%	-4.60%	-3.60%	-5.80%
P_m^{max}	Jan	13.5	**13.50%**	*16.00%*	**19.70%**	**23.90%**	19.20%	5.90%	16.00%	21.10%
	Feb	12.4	2.60%	11.10%	10.10%	15.20%	29.10%	61.20%	33.70%	32.30%
	Mar	14.2	10.00%	11.10%	13.50%	15.70%	16.70%	15.00%	7.00%	30.70%
	Apr	12.3	10.90%	14.00%	11.60%	16.50%	42.10%	44.70%	19.10%	66.60%
	May	16.6	4.10%	8.40%	13.80%	18.30%	15.60%	18.80%	31.00%	14.30%
	Jun	19.3	7.00%	6.00%	2.20%	3.20%	-16.70%	-22.10%	-21.20%	-16.10%
	Jul	16.9	7.50%	1.00%	1.50%	-5.60%	22.00%	11.00%	-1.70%	-22.80%
	Aug	18.7	1.20%	-4.50%	-6.30%	-19.20%	-7.50%	-16.50%	6.20%	-26.80%
	Sep	15.7	1.80%	2.80%	-2.20%	11.70%	21.80%	24.20%	28.90%	39.10%
	Oct	17.2	12.30%	12.40%	10.00%	16.00%	5.20%	8.80%	25.60%	17.80%
	Nov	16.6	9.20%	13.40%	*15.90%*	*24.20%*	11.70%	-6.10%	5.70%	17.90%
	Dec	15.8	6.70%	15.40%	**21.30%**	**31.80%**	-5.50%	10.40%	2.00%	3.50%

Table 7. Maximum deviation relative to the mean change factor projected by the CMIP6 ensemble allowed for acceptance for each target variable in SDM4. These deviations correspond to a 95% confidence interval of the distribution of each target variable projection within the GCM ensemble.

Target variable		Abbr.	Weight	SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP5-8.5
Dry days	Annual	and	0	6.90%	7.60%	7.10%	7.90%
	Winter	sndwi	0.1	11.80%	17.00%	14.50%	21.50%
	Spring	sndsp	0.1	15.20%	11.30%	12.50%	15.20%
	Summer	sndsu	0.2	12.30%	12.50%	13.70%	13.90%
	Autumn	sndau	0.1	7.50%	7.80%	12.10%	10.40%
Total P	Annual	and	0	9.70%	12.60%	12.00%	12.90%
	Winter	sndwi	0.03	12.30%	15.20%	16.70%	19.50%
	Spring	sndsp	0.06	21.10%	20.50%	21.60%	25.20%
	Summer	sndsu	0.15	28.30%	25.80%	28.90%	40.30%
	Autumn	sndau	0.06	14.00%	16.40%	12.80%	11.70%
Extreme P	10 mm	n10mm	0.1	18.30%	22.60%	22.50%	24.10%
	20 mm	n20mm	0.05	35.20%	38.10%	44.10%	41.10%
	Max	mdp	0.05	13.80%	20.60%	21.10%	17.60%