



**HAL**  
open science

## Dual Lidar Observations at $10.6 \mu\text{ m}$ and $532 \text{ nm}$ for Retrieving Semitransparent Cirrus Cloud Properties

Marjolaine Chiriaco, H. Chepfer, Vincent Noel, M. Haeffelin, M. Delaval,  
Philippe Drobinski

► **To cite this version:**

Marjolaine Chiriaco, H. Chepfer, Vincent Noel, M. Haeffelin, M. Delaval, et al.. Dual Lidar Observations at  $10.6 \mu\text{ m}$  and  $532 \text{ nm}$  for Retrieving Semitransparent Cirrus Cloud Properties. *Journal of Applied Meteorology and Climatology*, 2006, 45, pp.537-555. 10.1175/JAM2355.1 . hal-00068820

**HAL Id: hal-00068820**

**<https://hal.science/hal-00068820v1>**

Submitted on 17 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Evaluating CMIP5 Model Agreement for Multiple Drought Metrics<sup>✉</sup>

A. M. UKKOLA

*ARC Centre of Excellence for Climate System Science, and Climate Change Research Centre, University of New South Wales, Sydney, New South Wales, Australia*

A. J. PITMAN AND M. G. DE KAUWE

*Climate Change Research Centre, and ARC Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, New South Wales, Australia*

G. ABRAMOWITZ, N. HERGER, J. P. EVANS, AND M. DECKER

*ARC Centre of Excellence for Climate System Science, and Climate Change Research Centre, University of New South Wales, Sydney, New South Wales, Australia*

(Manuscript received 1 June 2017, in final form 29 March 2018)

## ABSTRACT

Global climate models play an important role in quantifying past and projecting future changes in drought. Previous studies have pointed to shortcomings in these models for simulating droughts, but systematic evaluation of their level of agreement has been limited. Here, historical simulations (1950–2004) for 20 models from the latest Coupled Model Intercomparison Project (CMIP5) were analyzed for a variety of drought metrics and thresholds using a standardized drought index. Model agreement was investigated for different types of drought (precipitation, runoff, and soil moisture) and how this varied with drought severity and duration. At the global scale, climate models were shown to agree well on most precipitation drought metrics, but systematically underestimated precipitation drought intensity compared to observations. Conversely, simulated runoff and soil moisture droughts varied significantly across models, particularly for intensity. Differences in precipitation simulations were found to explain model differences in runoff and soil moisture drought metrics over some regions, but predominantly with respect to drought intensity. This suggests it is insufficient to evaluate models for precipitation droughts to increase confidence in model performance for other types of drought. This study shows large but metric-dependent discrepancies in CMIP5 for modeling different types of droughts that relate strongly to the component models (i.e., atmospheric or land surface scheme) used in the coupled modeling systems. Our results point to a need to consider multiple models in drought impact studies to account for high model uncertainties.

## 1. Introduction

Droughts are major natural hazards with widespread impacts on humans and ecosystems. The frequency and magnitude of droughts are expected to change in the coming decades due to climate change, but the regional evolution of future droughts remains highly uncertain (Collins et al. 2013). Climate model projections form an

important basis for understanding and quantifying future changes in drought and ultimately inform policy and adaptation decisions. Many studies have employed climate models to investigate the evolution of past and future droughts, often indicating more severe drought conditions in the future (Dai 2013; Cook et al. 2015; Sheffield and Wood 2008; Taylor et al. 2013; Zhao and Dai 2015; Swann et al. 2016; Orlowsky and Seneviratne 2013). Commonly, the hydrological outputs from climate models are not analyzed directly, but instead the meteorology from the models is used to force offline models to derive estimates of droughts and aridity (Prudhomme et al. 2014; Schewe et al. 2014). There are, however, potential limitations with offline modeling as it cannot represent important

---

<sup>✉</sup> Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JHM-D-17-0099.s1>.

---

Corresponding author: A. M. Ukkola, [a.ukkola@unsw.edu.au](mailto:a.ukkola@unsw.edu.au)

land–atmosphere feedbacks that can modify drought characteristics, such as the coupling between soil moisture and air temperature (Seneviratne et al. 2010; Koster et al. 2006; Yin et al. 2014) associated with the surface energy fluxes (Donat et al. 2017). Common offline modeling approaches also lack relevant land surface–atmosphere processes, such as the effects of elevated CO<sub>2</sub> on plant water use (Milly and Dunne 2016). This has been shown to result in an overestimation of drought effects compared to climate model projections (although a number of studies have also indicated good agreement between offline and online drought metrics for soil moisture; e.g., Cook et al. 2015, Ault et al. 2016). As such, the direct use of coupled climate model outputs, such as runoff, soil moisture, and evapotranspiration, has been encouraged as part of drought assessments (Milly and Dunne 2017, 2016; Swann et al. 2016; Roderick et al. 2015; Decker et al. 2017).

While offline models have been extensively evaluated for drought (Tallaksen and Stahl 2014; Powell et al. 2013; Prudhomme et al. 2011; Whitley et al. 2016; Ukkola et al. 2016), the capability of coupled climate models to capture droughts remains less clear (Huang et al. 2016). Many past studies have concentrated on quantifying model agreement on future changes in drought (Swann et al. 2016; Koirala et al. 2014; Burke and Brown 2008; Cook et al. 2015; Taylor et al. 2013; Zhao and Dai 2015). Arguably, it is equally important to understand model agreement during the historical period to identify limitations in simulated droughts. Climate models should agree on large-scale drought patterns and trends if the relevant processes are represented consistently. Inspecting the historical period also eliminates uncertainties in model projections arising from future greenhouse emissions. Some studies have evaluated climate models for historical drought simulations, most commonly based on precipitation (meteorological drought). Nasrollahi et al. (2015) showed that while climate model simulations from the latest Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012) envelope historical observations of meteorological drought, most models tended to overestimate the area affected by drought. CMIP5 models also commonly disagreed with observations on regional precipitation drought trends over the period 1901–2005. Conversely, Orłowsky and Seneviratne (2013) showed that neither observations nor CMIP5 models showed significant trends in precipitation droughts over most regions, but the time period used in their study was much shorter (1979–2009). A number of studies have also evaluated the number of consecutive dry days (CDD) predicted by climate models (Orłowsky and Seneviratne 2012; Chen et al. 2014; Sillmann et al. 2013; Alexander and Arblaster 2009). While not strictly an indicator of drought, the models capture the spatial pattern in observations of CDD

well, although often underestimating the number of CDD due to the overestimation of rainy days (Chen et al. 2014; Sillmann et al. 2013). Similarly, Rocheta et al. (2014) demonstrated widespread underestimation of precipitation persistence in climate models, which manifests as underestimation in the length of dry periods.

Meteorological droughts only represent one type of drought, and while easier to characterize, they provide limited information about impacts on societies, agriculture, and ecosystems. These are more directly felt through deficits in runoff and soil moisture, commonly termed as hydrological and agricultural drought, respectively. Runoff and soil moisture droughts are governed by precipitation, changes in evaporative demand, and land surface processes, including vegetation behavior. Evaluation of these types of drought has been more limited, but some past studies have shown a large spread in climate model simulations of soil moisture droughts (Sheffield and Wood 2008; Wuebbles et al. 2014; Orłowsky and Seneviratne 2013) but reasonable agreement in historical soil moisture anomalies (Orłowsky and Seneviratne 2013). For runoff droughts, Koirala et al. (2014) reported strong model agreement for future changes in low flows.

Comparing these past studies is in many ways inconclusive, since they employed different methodologies (including the drought metrics and thresholds used), time periods, and drought definitions. As such, the relative model agreement in precipitation, runoff, and soil moisture droughts remains unclear, as does the contribution of precipitation biases to soil moisture and runoff droughts. We address this by evaluating agreement in historical CMIP5 simulations during the period 1950–2004 for different drought metrics and definitions within a single experimental framework. We first evaluate simulations of precipitation drought against observations and examine how model performance varies with drought duration, severity, and metrics. We then investigate agreement in model simulations of runoff and soil moisture droughts and quantify the contribution of precipitation biases to differences in runoff and soil moisture droughts. By analyzing the different hydrological components together, with a consistent methodology, we quantify the level of agreement in climate model simulations for different types of drought and contribute to an improved understanding of the sources of uncertainty in future projections of drought.

## 2. Methods

### a. CMIP5 models

We used monthly outputs from 20 CMIP5 models (Table 1) for the historical simulations over the period

TABLE 1. CMIP5 models analyzed in this study. The information on land and atmospheric schemes was derived from Flato et al. (2013, their Table 9.A.1). Where a model name was not provided (indicated as “Included”), models by the same institution (e.g., GISS) were assumed to use the same scheme.

Model	Institution	Atmospheric scheme	Land surface scheme	Resolution (lon, lat)
BCC_CSM1.1	Beijing Climate Centre	BCC_AGCM2.1	BCC-AVIM1.0	2.8°, 2.8°
BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University	CAM3.5	CoLM+b-NUDGM(C/N)	2.8°, 2.8°
CCSM4	National Center for Atmospheric Research	CAM4	CLM4	1.3°, 0.9°
CESM1(BGC)	Community Earth System Model Contributors	CAM4	CLM4	1.3°, 0.9°
CESM1(CAM5)	Community Earth System Model Contributors	CAM5	CLM4	1.3°, 0.9°
CESM1 (FASTCHEM)	Community Earth System Model Contributors	CAM4-CHEM	CLM4	1.3°, 0.9°
CESM1 (WACCM)	Community Earth System Model Contributors	WACCM4	CLM4	2.5°, 1.9°
CNRM-CM5	Centre National de Recherches Météorologiques/Centre Européen de Recherche et Formation Avancée en Calcul Scientifique	ARPEGE-Climat	SURFEX	1.4°, 1.4°
CNRM-CM5.2	Centre National de Recherches Météorologiques/Centre Européen de Recherche et Formation Avancée en Calcul Scientifique	ARPEGE-Climat	SURFEX	1.4°, 1.4°
FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and CESS, Tsinghua University	GAMIL2	CLM3	2.8°, 3.1°
GISS-E2-H	NASA Goddard Institute for Space Studies	Included	Included	2.5°, 2.0°
GISS-E2-H-CC	NASA Goddard Institute for Space Studies	Included	Included	2.5°, 2.0°
GISS-E2-R	NASA Goddard Institute for Space Studies	Included	Included	2.5°, 2.0°
GISS-E2-R-CC	NASA Goddard Institute for Space Studies	Included	Included	2.5°, 2.0°
HadCM3	Met Office Hadley Centre	HadAM3	Included	3.8°, 2.5°
HadGEM2-ES	Met Office Hadley Centre	HadGAM2	Included	1.9°, 1.3°
INM-CM4.0	Institute for Numerical Mathematics	Included	Included	2.0°, 1.5°
MIROC5	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	CCSR/NIES/FRCGC/AGCM6	MATSIRO	1.4°, 1.4°
NorESM1-M	Norwegian Climate Centre	CAM4-Oslo	CLM4	2.5°, 1.9°
NorESM1-ME	Norwegian Climate Centre	CAM4-Oslo	CLM4	2.5°, 1.9°

1950–2004. One ensemble member [r1i1p1, where r1 denotes the first initial conditions for the first initialization method (i1) using the first set of physics (p1)] was used for each model in the main analysis to avoid difficulties in blending CMIP5 models with different numbers of ensemble members. Additionally, other ensemble members with different initial conditions (r\*i1p1) were analyzed for a subset of models to investigate the effects of internal variability on our conclusions. We analyzed outputs for total precipitation, total runoff, and soil moisture content by layer (variables pr, mrro, and mrlsl in the CMIP5 archive, respectively). We also present results for surface runoff (variable mrros) in the online supplemental

material. The CMIP5 models were chosen based on the availability of all of these outputs. Precipitation and runoff outputs were converted from kilograms per square meter per second to millimeters per month, accounting for the number of days in each month. All outputs were processed in their native grid but were interpolated to the coarsest common grid ( $3.75^\circ \times 2.5^\circ$ ) for plotting using bilinear interpolation.

Previous studies have used both the soil moisture content for the entire soil column and soil moisture standardized to a common depth when comparing CMIP5 models (Berg et al. 2017; Huang et al. 2016; Orlowsky and Seneviratne 2013; Swann et al. 2016). The

(hydrologically active) soil column depths for the models used in this study vary significantly between 3 m for HadCM3 and HadGEM2-ESM, and 14 m for MIROC5. As such, the variability and dynamics of soil moisture in the deeper soil columns would differ significantly from the shallower columns and need not necessarily reflect the same processes (Berg et al. 2017). Therefore, to allow for comparison between models, the soil moisture content was interpolated to the top 3 m (the deepest common depth). This was achieved by summing the soil moisture content ( $\text{kg m}^{-2}$ ) for all soil layers within the top 3 m and (where applicable) any layer partly within 3 m weighted by the fraction of this layer located above 3 m.

### b. Observations

We used two observed precipitation products to evaluate model performance. These were global monthly time series products by 1) the Climatic Research Unit (CRU TS 3.23; Harris et al. 2014) and 2) Global Precipitation Climatology Centre (GPCC, version 7; Schneider et al. 2016). Both products are available at a  $0.5^\circ$  spatial resolution but were regridded to a common grid using the CRU TS 3.23 land mask.

We do not use runoff or soil moisture observations in this study. Global spatially continuous and temporally variable observations of runoff are not available. Global soil moisture observations are only available for near-surface soil layers (as in the case of satellite products) and need not reflect drought, or do not cover a sufficient time period [such as the Gravity Recovery and Climate Experiment (GRACE) product; Tapley et al. 2004]. Rather than directly comparing runoff and soil moisture against observations, we investigate model differences in these variables against the modeled mean and compare this with similar metrics of precipitation. This avoids the problem of the sparseness of relevant observations and enables a consistent methodology to be implemented for the three types of drought considered.

### c. Drought metrics

We analyzed precipitation, runoff, and soil moisture droughts at monthly time scales. We note evapotranspiration is also a key component of the hydrological cycle but is harder to use to quantify droughts because low fluxes can reflect limited water, limited energy, or both (Zhang et al. 2004). We used the methodology underlying the standardized precipitation index (SPI; McKee et al. 1993) to identify drought periods. SPI is a widely used drought metric most commonly employed for meteorological droughts but represents a generic method for transforming time series into standardized anomalies and can be applied to other types of drought

(McKee et al. 1993; Shukla and Wood 2008). It was chosen to provide a standardized measure of droughts, aiding direct comparison between different models and different variables, and can be calculated at different time scales ( $\geq 1$  month). We applied an SPI-based approach to precipitation, runoff (total and surface), and soil moisture time series and refer to this as the standardized drought index (SDI) hereafter.

SPI traditionally uses a gamma distribution to transform the time series as this fits precipitation data well (McKee et al. 1993). We tested the suitability of the gamma distribution for the CMIP5 precipitation, runoff, and soil moisture data using the Kolmogorov–Smirnov test (Fig. S1 in the online supplemental material). We found the distribution to fit precipitation ( $>95\%$  of grid cells for all models) and soil moisture ( $>80\%$  of grid cells for all models except HadCM3) time series well. The fit was lower for runoff ( $>50\%$  of grid cells for all models, but  $>80\%$  for 11 models). The fit was generally lower in drier and colder regions (not shown). Touma et al. (2015) showed that the runoff fit for 15 CMIP5 models could not be improved by using two alternative distributions (log-normal and generalized extreme value) compared to the gamma distribution. As such, we have employed the gamma distribution to calculate SDI for all three variables, but discuss the implications for runoff in section 3b(1).

To calculate SDI, a time series was first constructed for each month by summing the variable over the preceding  $n$  months, where  $n$  is the time scale used to calculate SDI. A two-parameter gamma distribution was then fitted to the  $n$ -monthly sums. The fitted gamma distribution was subsequently transformed into a standard normal distribution such that the resulting SDI time series (in units of standard deviations) had a mean of 0 and standard deviation of 1. We used the standardized precipitation evapotranspiration index (SPEI) R package (<https://cran.r-project.org/web/packages/SPEI/SPEI.pdf>; Vicente-Serrano et al. 2010) to calculate the SDIs. The SDI was calculated separately for each model, observational product, and grid cell so that each model's drought metrics are calculated relative to its own climatology. The SDI was not determined for grid cells with a high occurrence of zero values, and grid cells were only analyzed if the SDI or associated drought metrics could be calculated for at least 66% of the models.

We determined drought events based on 3-month SDI ( $n = 3$ ), as this reflects seasonal-scale (or longer) drought events and incorporates soil moisture “memory” effects (Orth and Seneviratne 2012). This approach has also commonly been employed in previous studies (e.g., Nasrollahi et al. 2015; Orłowsky and Seneviratne 2013). This scale may not be ideal in regions with seasonal snow (due to storage effects) or seasonally dry

climates. As such, we also repeated the analysis using 6-month SPI, but our conclusions were not qualitatively different from 3-month SPI (not shown). We defined three drought severities based on the thresholds below (following McKee et al. 1993; Fig. 1):

$$\text{moderate: } SDI \leq -1, \tag{1}$$

$$\text{severe: } SDI \leq -1.5, \text{ and } \tag{2}$$

$$\text{extreme: } SDI \leq -2. \tag{3}$$

Any months with an SDI value below these thresholds was classified as a drought month. Next, we calculated the duration and intensity separately for each drought event and severity class (illustrated in Fig. 1). Duration  $D$  represents the number of consecutive months below the given severity thresholds. As we use 3-month SPI, the minimum drought duration is by definition 3 months, but can be longer when individual drought months are consecutive (Fig. 1). Intensity is the average monthly departure from the climatological mean during the drought event and can be expressed as absolute  $I_{\text{abs}}$  (mm) or relative  $I_{\text{rel}}$  (%) differences:

$$I_{\text{abs}} = \frac{\sum(\bar{x}_m - x_m)}{D}; \quad m \in [i, j] \quad \text{and} \tag{4}$$

$$I_{\text{rel}} = \left| \frac{\sum(x_m - \bar{x}_m)}{\sum \bar{x}_m} \right| \times 100; \quad m \in [i, j], \tag{5}$$

where  $x$  is the monthly value,  $\bar{x}$  the corresponding monthly mean,  $i$  is the drought start month, and  $j$  the end month (Fig. 1).

The drought events were subsequently classified according to their duration as

$$\text{short term: } 3 \leq D \leq 4, \tag{6}$$

$$\text{medium term: } 5 \leq D \leq 6, \text{ and } \tag{7}$$

$$\text{long term: } 7 \leq D \leq 8. \tag{8}$$

We did not investigate droughts longer than 8 months as they were not sufficiently common in the CMIP5 models to provide a robust sample (<50% of grid cells had droughts with a longer duration across most models and variables). We then calculated the average duration and intensity across all drought events, as well as the frequency of short-, medium-, and long-term droughts (expressed as the average number of events per 10 years). The modeled and observed means were subsequently calculated as the mean of each metric across the 20 models and two observed products, respectively.

We use the relative difference (RD) to quantify the level of agreement across models and against observations:

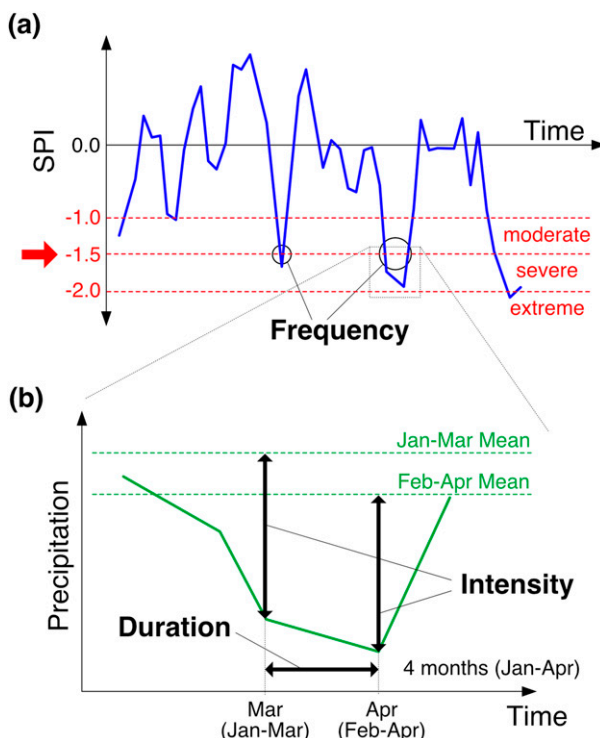


FIG. 1. Illustration of the drought metrics using 3-month SPI. (a) An example monthly SPI time series for precipitation, with the three drought thresholds applied in this study shown in red. (b) Using the “severe” (−1.5 SPI) threshold as an example, a drought event spanning March–April is shown, with each month reflecting 3-monthly totals.

$$RD = (x_{\text{model}} - x_{\text{ref}})/x_{\text{ref}} \times 100, \tag{9}$$

where  $x_{\text{model}}$  is the value for an individual model, and  $x_{\text{ref}}$  is the mean of observations or models. We define three levels of agreement based on the RD:

$$\text{high: } |RD| < 10\%, \tag{10}$$

$$\text{medium: } 10\% < |RD| < 20\%, \text{ and } \tag{11}$$

$$\text{low: } |RD| > 20\%. \tag{12}$$

#### d. Hierarchical clustering

We used hierarchical clustering to explore the similarity of drought metrics across climate models sharing a land surface model (LSM) or atmospheric scheme to better understand sources of model differences. The clustering was performed based on spatial correlations, thus measuring similarity in spatial patterns.

All model simulations and observations were first remapped to a common grid using the coarsest resolution ( $3.75^\circ \times 2.5^\circ$ ). Pairwise area-weighted Pearson correlations  $r$  were then calculated for all combinations of models and observations (where available) using all



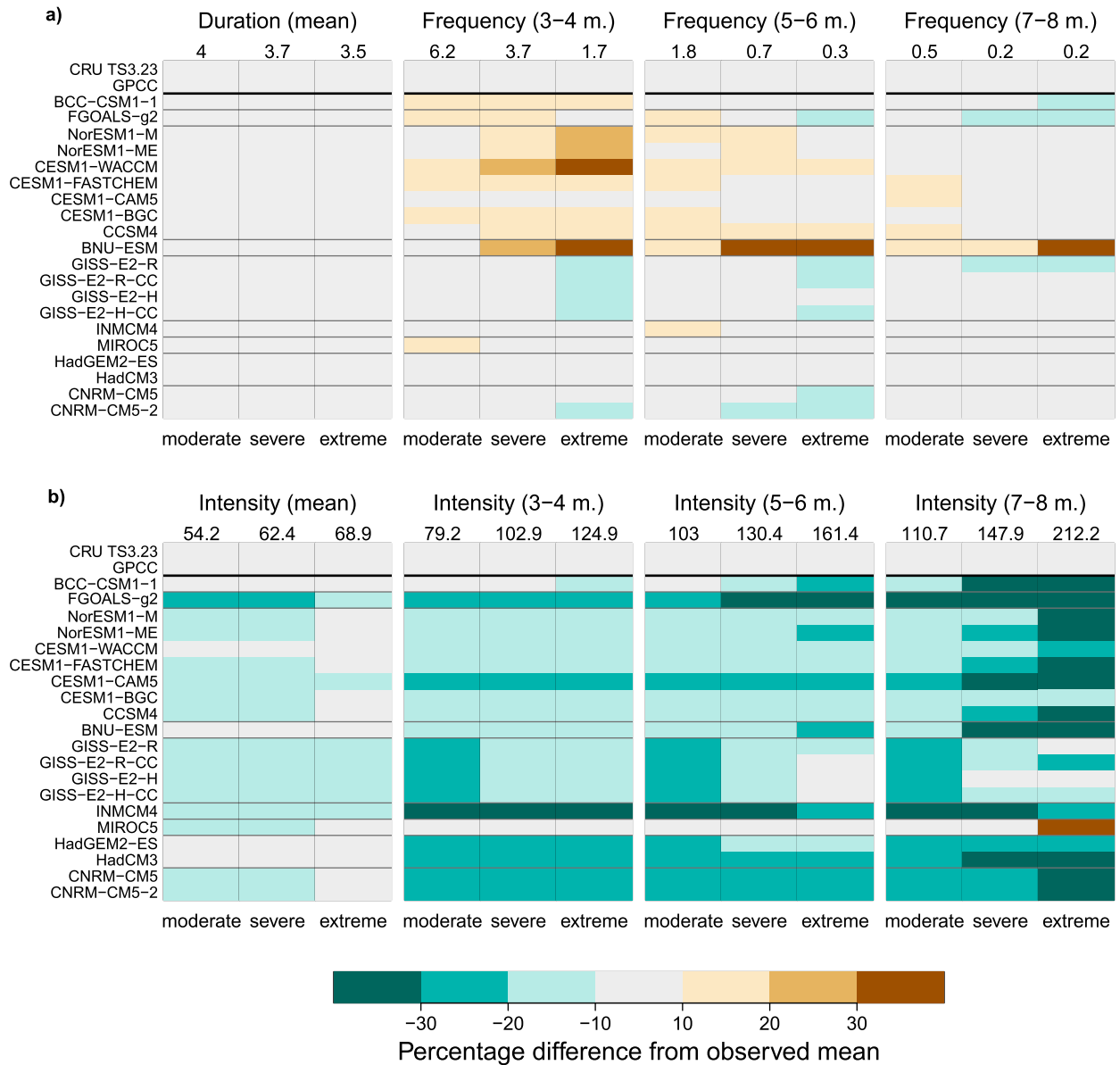


FIG. 2. Globally averaged biases in individual CMIP5 models relative to the observed mean for precipitation drought metrics. (a) Duration (months) and frequency (number of events) for different drought lengths and (b) intensity (mm). The observed mean is noted on top of each column. Metrics were averaged across all drought events (denoted mean) and short- (3-4 months), medium- (5-6 months), and long-term (7-8 months) droughts. The global average was calculated for land areas between 65°N and 65°S, excluding grid cells where no droughts were detected in >33% of the models. The models were arranged by their land surface scheme, with models with different schemes separated by horizontal lines (see Table 1).

available grid cells. A distance matrix was then computed from the correlations as  $1 - r$  and used for the clustering. We used the complete linkage method for clustering following Gibson et al. (2017). The clustering is presented as a dendrogram, where the y axis is a measure of dissimilarity between individual models, with models located on the same branch sharing similarities in their spatial patterns.

### 3. Results

#### a. Evaluation of precipitation droughts against observations

We first analyzed simulations of precipitation droughts against the two observational products globally. Figures 2a and 2b show the agreement in the two observational products and the individual climate model biases

relative to the mean of the observations (with the mean observed values shown above each column). There is “high” agreement (both products within  $\pm 10\%$  of the observed mean for all metrics) among the individual observational products. The models also agreed well with each other and with the observations for mean duration and the frequency of long-term (7–8 month) droughts for the three drought severity classes (Fig. 2a).

Model biases in drought frequency were greater for shorter droughts such that most models (80%–85%) are within  $\pm 10\%$  of the observations for long-term droughts but decreasing to 55%–70% for medium-term and 35%–70% for short-term droughts. Model biases are generally greatest for the frequency of extreme short-term droughts, with BNU-ESM and CESM1(WACCM) showing the largest overestimation (up to  $>30\%$ ) and FGOALS-g2, CNRM, and the GISS family of models the largest underestimation (up to 20%) compared to the observed mean.

Most models show “high” agreement with observations, or are systematically biased in one direction. The direction of bias is consistent among models employing similar LSMs and atmospheric schemes. Models that share an LSM (Fig. 3a) or an atmospheric scheme (Fig. S4) also generally cluster together in the dendrogram, indicating similar spatial patterns in drought duration, although the distances are fairly large ( $>0.4$ ) across models.

We next address drought intensity (Fig. 2b). Despite generally agreeing well on drought duration and frequency, the models systematically underestimate drought intensity compared to the observations. CESM1 (CAM5), FGOALS-g2, and INM-CM4.0 consistently underestimate intensity by  $>10\%$  across all metrics, whereas the magnitude of bias in other models varies with the metric but generally shows medium or low agreement with observations. Overall, the biases are largest for long-term droughts, with most models underestimating the intensity of long-term droughts by  $>20\%$  compared to observations, independent of drought severity. Model agreement is also lowest for long-term drought intensity (Fig. 2b). The models cluster strongly based on their LSMs (Fig. 3a) and atmospheric schemes (Fig. S4), with the models using the same scheme generally having distances  $< 0.2$ .

Next, we explored the spatial patterns in precipitation drought metrics to identify regions with the largest model discrepancies. Figure 4 shows the interdecile range of model estimates (i.e., the difference of the 90th and 10th percentile values of the 20 model estimates) and the relative bias in each grid cell for severe droughts (moderate and extreme droughts are shown in Figs. S4, S5). The models agree well on the mean duration of severe

droughts, varying by  $<0.5$  months across most regions and are in “high” agreement with observations in most grid cells (Fig. 4a). Model differences in absolute intensity are highest in the low latitudes, and in many of these regions the models simulate consistently lower intensities than observed (particularly over the Amazon, Sahel, and India; Fig. 4b). In the Amazon, the models have a tendency to underestimate monthly mean rainfall (Fig. S7), explaining the lower absolute intensities. They also underestimate the coefficient of variation (CV), indicating smaller variations than in the observations relative to the mean precipitation (Fig. S7). Where models overestimate absolute intensity (dry subtropical regions), they also tend to overestimate mean monthly rainfall and the CV relative to observations.

Relative intensities also vary the most in the low latitudes and northern subtropics, but are systematically underestimated by the majority of models across most regions (Fig. 4c), suggesting smaller relative departures from the mean state in the models than observations during drought months. Apart from arid regions, the models are generally in “high” agreement with observations on the frequency of short-term droughts or do not exhibit systematic biases (Fig. 4d). These results are consistent with moderate and extreme droughts (Figs. S5, S6).

#### *b. Evaluation of runoff and soil moisture drought agreement against CMIP5 mean*

We next explored simulations of runoff and soil moisture droughts, evaluating model agreement against the multimodel mean. We did this to quantify the level of model differences across drought metrics. We also explored how the differences vary for similar LSMs. Good model agreement does not imply that the models are simulating these quantities correctly. However, large discrepancies across models indicate unrealistic simulations for at least some of the models, even in the absence of observations, and likely a lack of robustness in future projections of this quantity.

##### 1) RUNOFF

In contrast to precipitation, model agreement in runoff droughts is much lower globally, with the exception of mean drought duration (Fig. 5a). Model estimates of drought frequency vary by  $>30\%$  below and above the model mean, but not consistently across the metrics for individual models. For example, the CESM family of models, BCC\_CSM1.1, and FGOALS-g2 tend to simulate more short-term droughts than the model mean but fewer medium- and long-term droughts, particularly for extreme drought severity. This suggests that the behavior of individual models relative to the model mean is highly dependent on the metric analyzed.



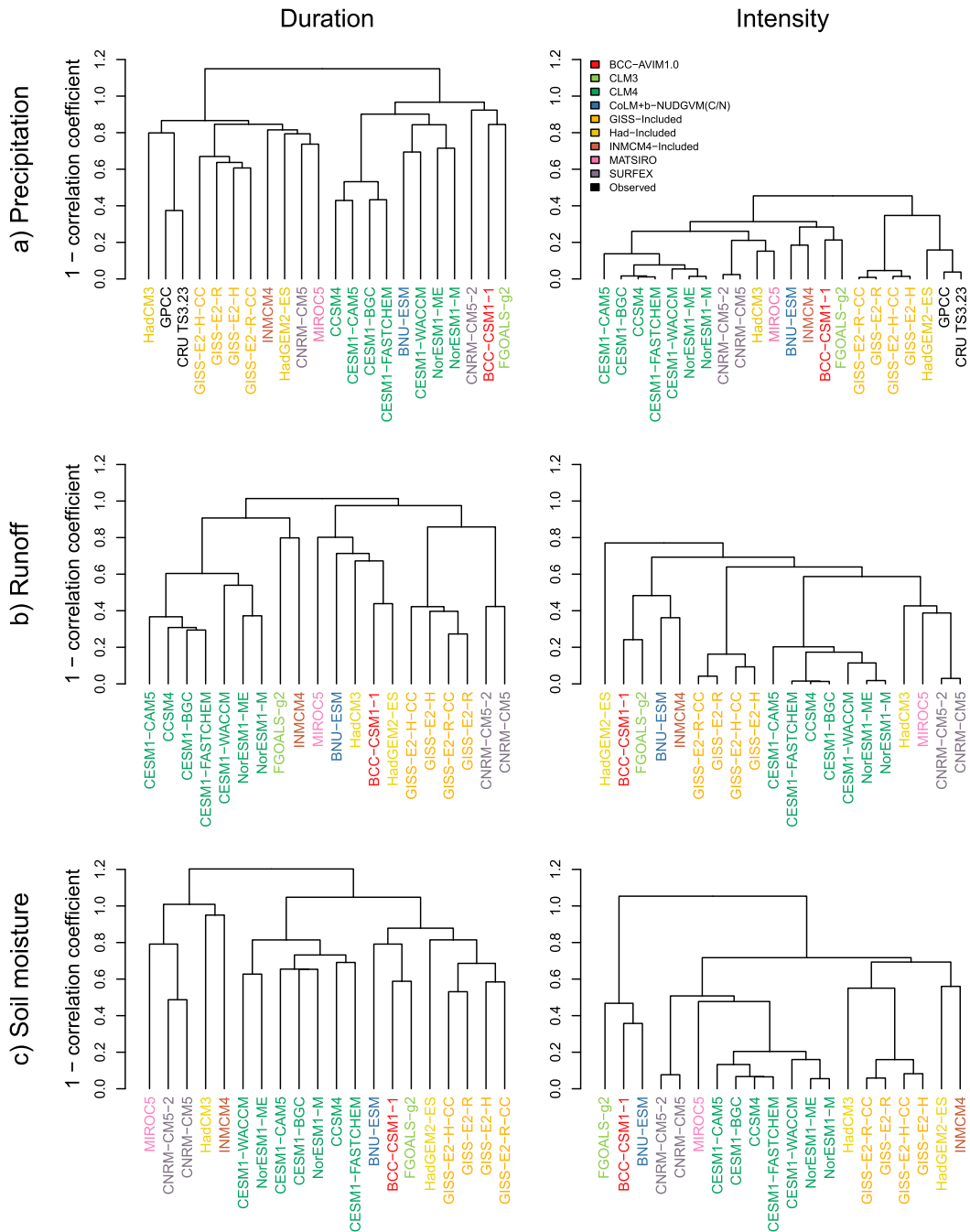


FIG. 3. Dendrograms for (left) mean duration and (right) mean absolute intensity for (a) precipitation, (b) total runoff, and (c) soil moisture. The models were colored according to their land surface schemes (see legend; Table 1). The dendrograms were constructed using hierarchical clustering (see section 2d for details). Severe droughts are shown here, with moderate and extreme droughts shown in Figs. S2 and S3, respectively.

For drought intensity, the range in model estimates is large across all the metrics, but many models show consistent behavior relative to the model mean across the metrics (Fig. 5b). CESM1(CAM5) and the GISS family of models underestimate intensity by >20%

relative to the mean for most metrics, whereas BCC\_CSM1.1, HadCM3, and MIROC5 overestimate intensity by >10%. The models cluster closely according to their LSMs, albeit generally more strongly for intensity than duration (Fig. 3b). Similarly, model

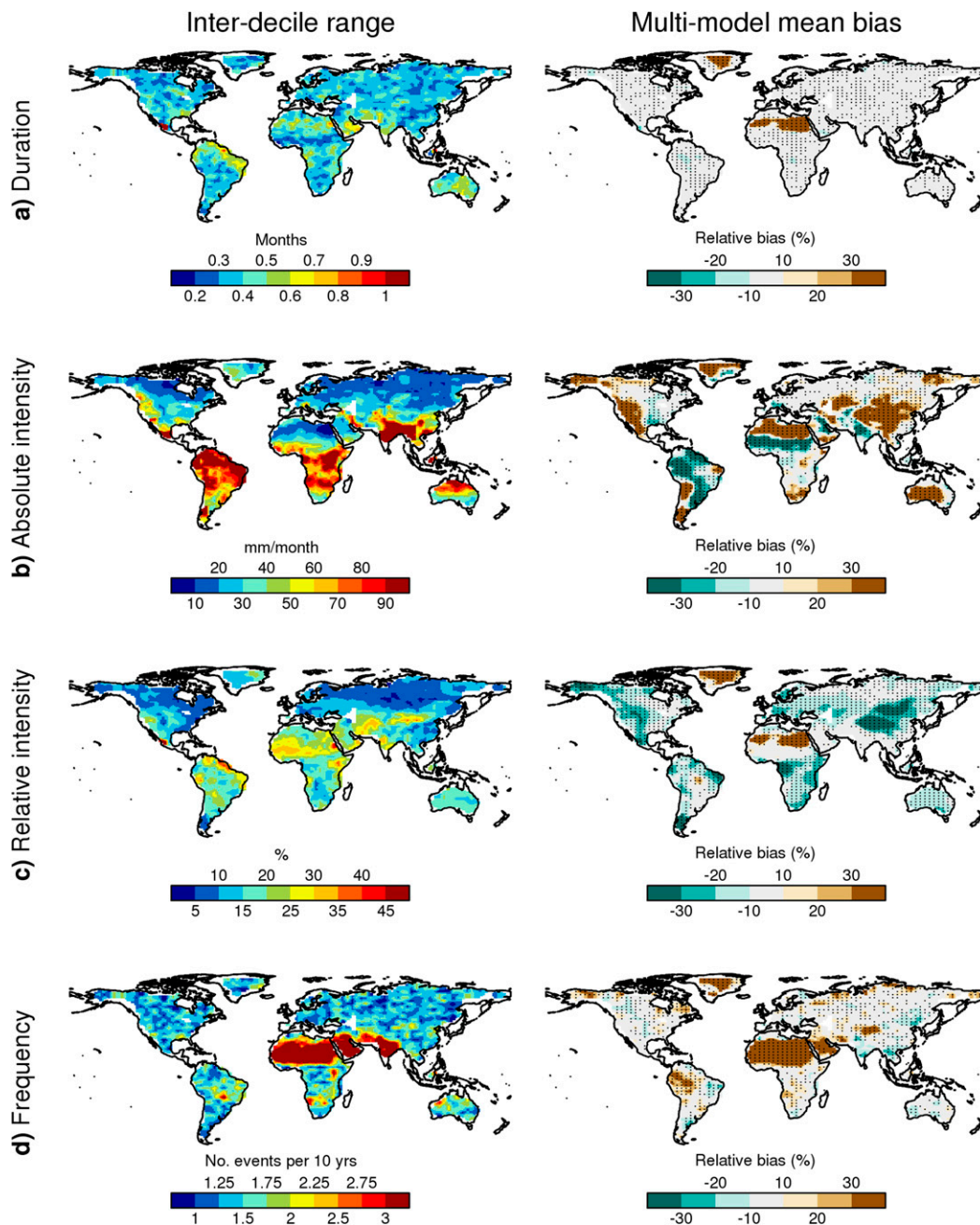


FIG. 4. (left) The interdecile range and (right) multimodel mean relative bias in model estimates for precipitation drought metrics averaged across all severe drought events: (a) the average duration for different drought severities, the average intensity as (b) absolute and (c) relative ranges, and (d) the frequency of short-term droughts. Stippling for relative bias shows where  $\geq 66\%$  of models agree with the ensemble mean on the direction of the bias. For interdecile range, the difference between 10th and 90th percentile estimates across the 20 GCMs is shown for each grid cell. Grid cells where no droughts were detected in  $>33\%$  of the models were masked out. Moderate and extreme droughts are shown in Figs. S5 and S6.

discrepancies are large for surface runoff droughts and highly consistent across the metrics for individual models and between those sharing an LSM (Fig. S8).

Overall, the disagreement in runoff drought frequency and intensity is much larger than for precipitation droughts globally. This may partly be due to the poorer fit of the gamma distribution for runoff

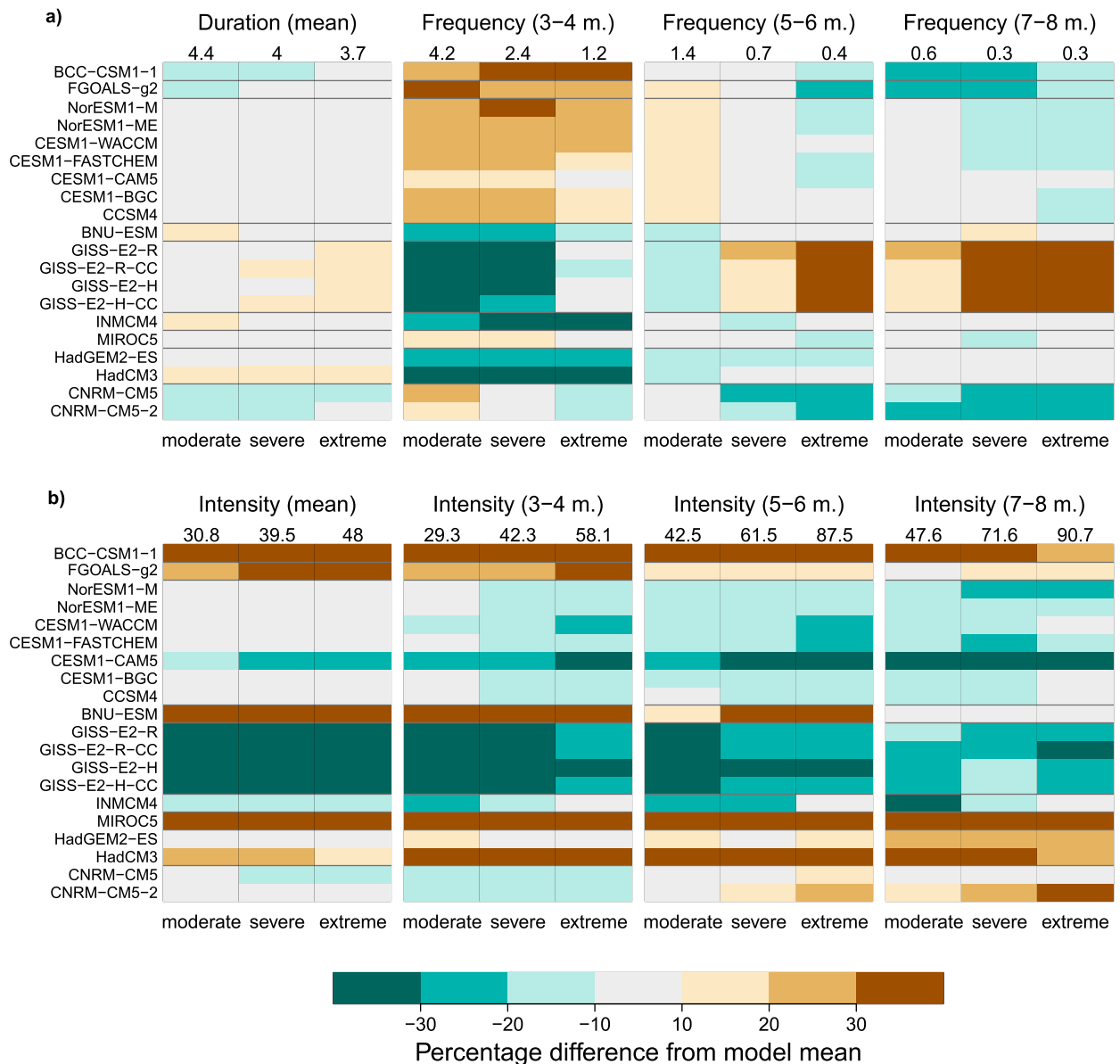


FIG. 5. As in Fig. 2, but for globally averaged differences in individual CMIP5 models relative to the modeled mean for runoff drought metrics.

compared to precipitation (Fig. S1), which can bias the SDI metrics by exaggerating or minimizing drought severity (Stagge et al. 2016). The poorest fit was found for the GISS family of models (53%–54% of grid cells passing the Kolmogorov–Smirnov test annually). These models do indeed generally show more extreme departures from the ensemble mean (Fig. 4). However, models with better fits, including FGOALS-g2 (82%), CESM1(CAM5) (92%), INM-CM4.0 (81%), and CNRM-CM5.2 (82%), also show low agreement with the ensemble mean for individual metrics in either direction. The low level of agreement across the runoff

metrics is thus not merely a result of the models with a lower SDI fit.

Figure 6 shows the interdecile range across the models in runoff drought metrics spatially. Model differences for mean drought duration are higher compared to precipitation droughts across most regions (Fig. 6a). Outside the arid regions, the model range is broadly similar, particularly for moderate and severe droughts, varying by ~1–2 months across the ensemble with the exception of western Russia and parts of North America. Model spread generally decreases for extreme droughts, apart from the arid regions.

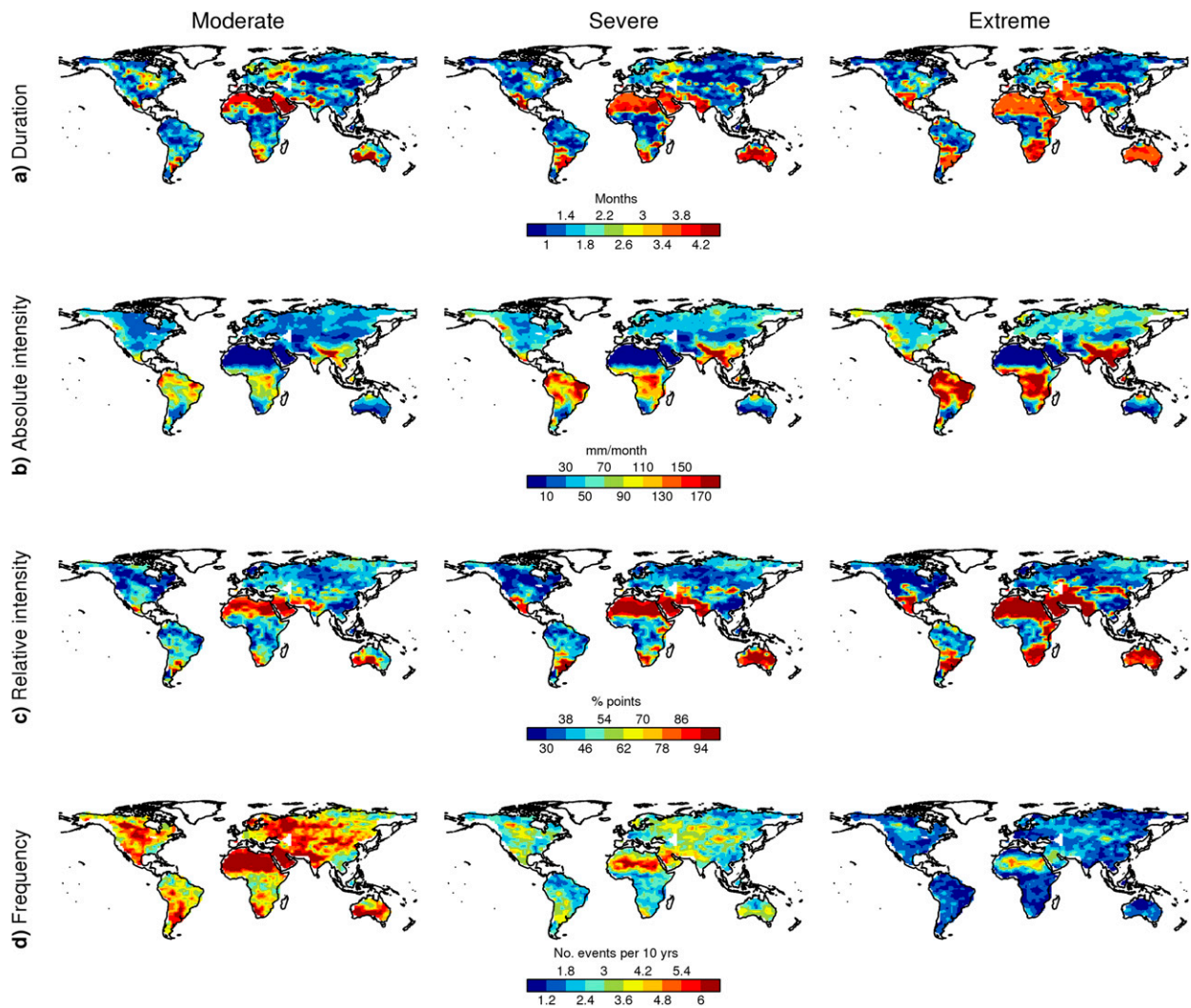


FIG. 6. The interdecile range in model estimates for runoff drought metrics averaged across all drought events: (a) the average duration for different drought severities, the average intensity as (b) absolute and (c) relative ranges, and (d) the frequency of short-term droughts. For each grid cell, the difference between 10th and 90th percentile estimates across the 20 GCMs is shown. Grid cells where no droughts were detected in >33% of the models were masked out.

For moderate droughts, the frequency of short-term events varies by >5 events in most arid regions and parts of Russia and North America, and by >2.5 events elsewhere (averaged over a 10-yr period; Fig. 6d). Model discrepancies in surface runoff correspond well to total runoff but are generally smaller in the Americas (Fig. S9). The regions with large model discrepancies are more widespread in moderate and severe drought frequency for both total and surface runoff compared to precipitation droughts.

For drought intensity, absolute model differences are largest in the tropics and increase with drought severity (Fig. 6b), in agreement with precipitation. Relative model differences are large (>60% points) in many drier regions, in particular Mexico, Australia, the

Sahara, central Asia, and southern Africa and South America (Fig. 6c).

## 2) SOIL MOISTURE

Figure 7 shows the percentage difference of individual models from the model mean for different soil moisture drought metrics globally. The models generally agree well on the mean duration and frequency of soil moisture droughts and are largely in “high” agreement with the model mean (Fig. 7a). Where models disagree more strongly (e.g., BCC\_CSM1.1 and FGOALS-g2), the over- and underestimations relative to the mean are generally inconsistent across models and metrics. Overall, the level of model agreement for soil moisture drought duration and frequency is similar to

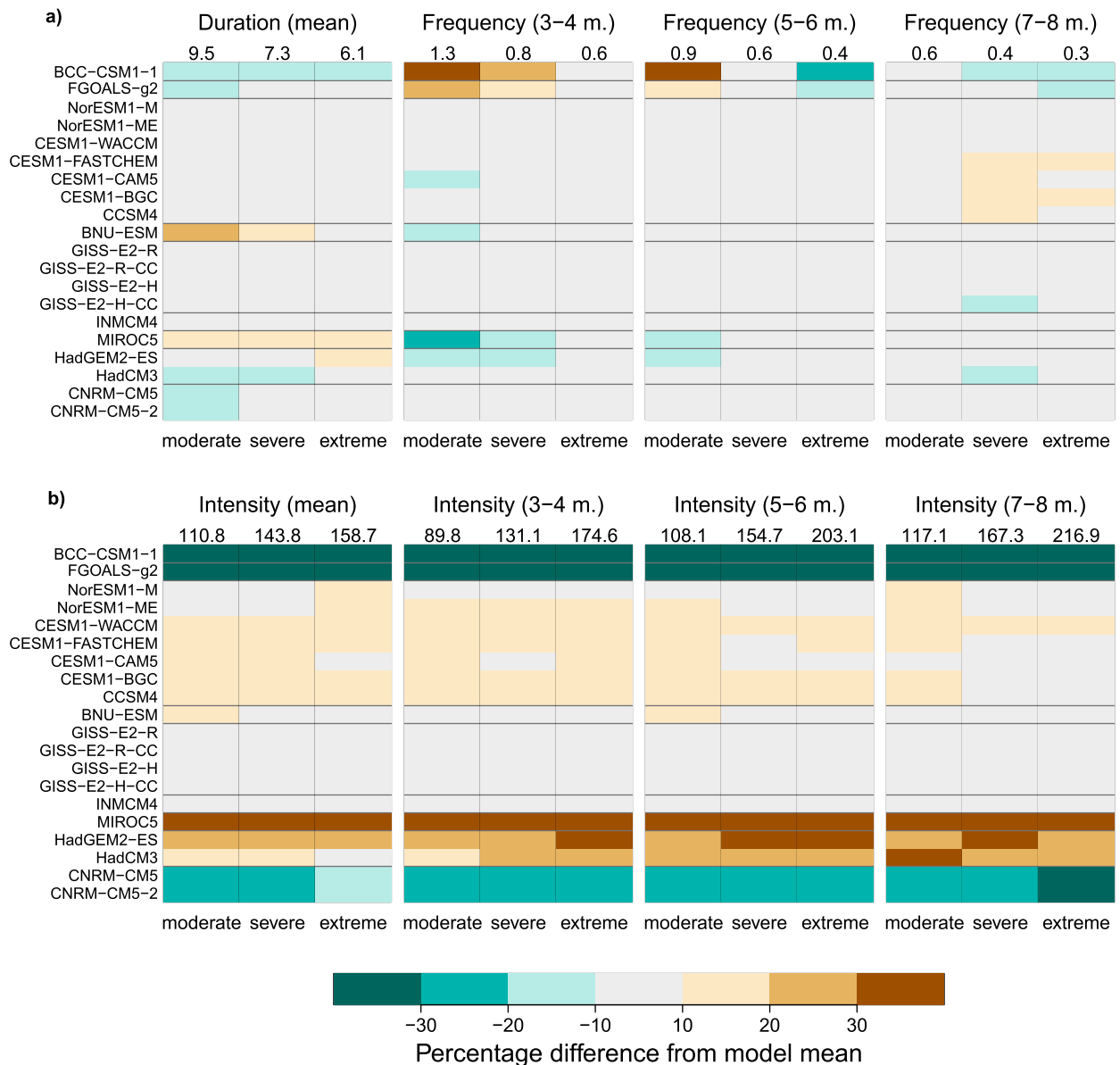


FIG. 7. As in Fig. 2, but for globally averaged differences in individual CMIP5 models relative to the modeled mean for soil moisture drought metrics.

precipitation and better than for runoff. The simulated soil moisture droughts are ~2 times longer in duration than precipitation or runoff droughts on average.

The models, however, disagree more strongly on the intensity of soil moisture droughts, with model estimates varying from underestimations of >30% for BCC\_CSM1.1 and FGOALS-g2 to overestimations of >30% for MIROC5 relative to the model mean (Fig. 7b). Model discrepancies for soil moisture drought intensity are comparable to runoff droughts for individual models and smaller than for precipitation droughts. However, the behavior of individual models is

highly consistent across the different metrics relative to the model mean and agrees strongly across models with the same LSM, both globally averaged (Fig. 7) and spatially (Fig. 3c). The models that tend to be lower than the ensemble mean (BCC\_CSM1.1, FGOALS-g1, and CNRM models) have soil depths varying between 3.4 and 8 m. The models that are generally higher than the ensemble mean (NorESM1 and CESM1 models, CCSM4, MIROC5, HadGEM2-ES, and HadCM3) have soil depths between 3 and 14 m. As such, the differences are not explained merely by different soil column depths that could in principle lead to systematically different



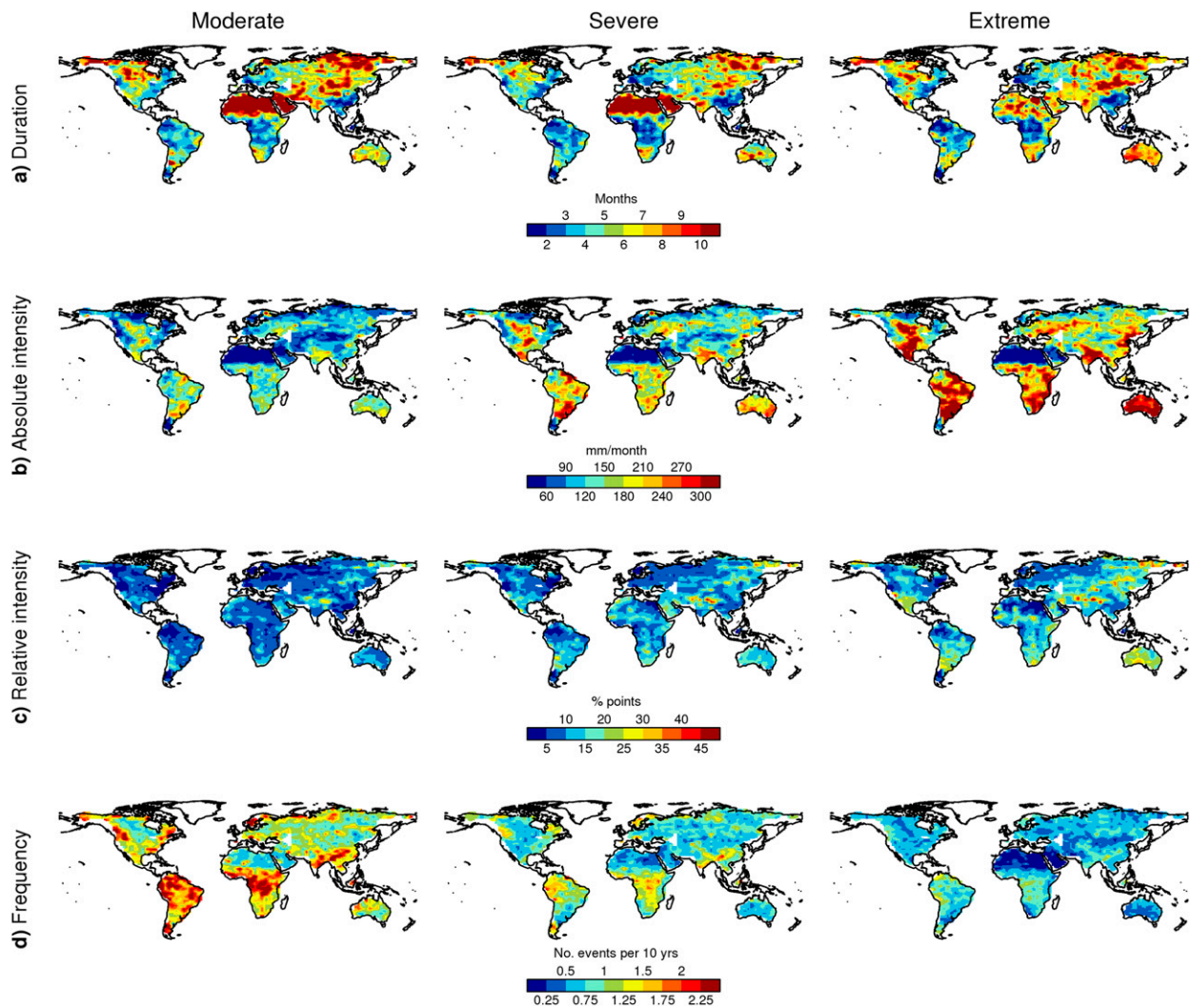


FIG. 8. As in Fig. 6, but for the interdecile range in model estimates for soil moisture drought metrics averaged across all drought events.

drought behavior in shallow versus deep soils [as also illustrated in Koster et al. (2009) for soil moisture time series across LSMs].

Figure 8 shows the interdecile range in soil moisture drought metrics spatially. Model discrepancies in moderate and severe mean drought duration are largest across the Sahara and Middle East, as well as parts of Russia (Fig. 8a), varying by >7 months. Elsewhere, the model range is broadly similar and generally varies between 2 and 5 months. Unlike for precipitation and runoff, model discrepancies do not vary consistently with drought severity. Model discrepancies in the frequency of short-term droughts are highest in the tropics, with model estimates generally varying by >2 events for moderate severity (Fig. 8d). Model agreement increases with drought severity, with the model range generally <6 events for extreme drought frequency due

to the smaller overall occurrence of these events. In agreement with precipitation and runoff droughts, absolute differences in soil moisture drought intensity are large in the tropics (Fig. 8b). However, the range in model estimates is also large in much of the Southern Hemisphere as well as many northern midlatitude regions, such as central North America and parts of Russia. Conversely, relative model differences are largest in the dry regions, such as Australia, southern Africa, and South America, but also across much of Russia, with model discrepancies increasing with drought severity (Fig. 8c).

*c. Interaction between precipitation, runoff, and soil moisture droughts*

Finally, we analyzed the relationship between precipitation, runoff, and soil moisture droughts. While



runoff and soil moisture droughts result from interactions between atmospheric and land surface processes, they are driven at least in part by precipitation deficits (Dai 2011). We have identified large discrepancies in simulated runoff and soil moisture droughts relative to the model mean. It is thus helpful to quantify the contribution of intermodel differences in precipitation droughts to intermodel differences in runoff and soil moisture droughts to better understand the sources of uncertainty in GCM-simulated runoff and soil moisture droughts. We therefore investigated whether models with longer (or more intense) precipitation droughts also simulated longer (more intense) runoff and soil moisture droughts (and vice versa), by correlating estimates from the 20 models for each runoff and soil moisture drought metric with the corresponding precipitation drought metric. We calculated correlations for each grid cell separately to explore the spatial pattern in the correlations. Figure 9 shows the runoff–precipitation correlation, and Fig. 10 shows the soil moisture–precipitation correlation for average drought duration and intensity and the frequency of short-term droughts.

Runoff and precipitation drought duration correlated weakly across most regions, with the exception of parts of the tropics for moderate droughts (Fig. 9a). This implies that models with longer precipitation droughts do not generally have a tendency to simulate longer runoff droughts. This will result partly from strong model agreement for mean duration of precipitation droughts, with only small variations among the models in most regions (Fig. 4a), and points to an important control of the duration by other factors that affect runoff (e.g., precipitation intensity, vegetation use of water, drainage, soil types, and soil moisture stores replenishing subsurface runoff).

Similarly, correlations are weak for moderate short-term drought frequency, but become more evident with increasing drought severity (Fig. 9c) and show strongest spatial coherence over the Sahara, Middle East, India, and parts of Australia and South America. Conversely, runoff and precipitation droughts are significantly correlated across many regions for mean drought intensity, such that models with more intense precipitation droughts also have a tendency to produce more intense runoff droughts. Regions with statistically significant correlations include parts of the northern high to mid-latitudes, South America, sub-Saharan Africa, and Australia, but the spatial coherence overall decreases with increasing drought severity. Regions with low correlations (western Russia, Canada) tend to correspond to regions where biases in precipitation drought intensity are inconsistent across models (Fig. 4b). These

findings are broadly consistent with van Huijgevoort et al. (2013), who showed high correlations between runoff and precipitation droughts in high-runoff regions and low correlations in colder and drier regions in global offline hydrological models (although a different drought definition was used in their study).

Similarly, soil moisture and precipitation droughts correlate weakly for mean drought duration as well as short-term drought frequency across most regions for all drought severities (Figs. 10a,c). Conversely, significant positive correlations between soil moisture and precipitation drought intensity are evident in many parts of the world, but the spatial pattern is less coherent compared to runoff (Fig. 10b). In particular, the correlations are weaker in the tropics and over Europe but stronger over Australia compared to runoff. The correlations for drought intensity become somewhat weaker with increasing drought severity.

#### 4. Discussion

##### *a. Differences across precipitation, runoff, and soil moisture droughts*

Our results show that CMIP5 models have large discrepancies in common drought metrics during the period 1950–2004, particularly at the regional scale. Importantly, the level of agreement varies strongly with the type of drought analyzed. The absolute model differences are lower for precipitation droughts compared to runoff and soil moisture droughts, particularly for average drought duration and intensity. This is likely in part due to the generally shorter duration of precipitation droughts across all models. Despite “high” agreement, the models systematically underestimate precipitation drought intensity compared to observations (Fig. 2b). This points to lower absolute variability in modeled precipitation, particularly in wet regions (Fig. S7), and may lead to an underestimation of drought impacts in studies analyzing precipitation droughts in wet regions in climate models. Absolute model discrepancies in drought intensity were highest in the tropics in both precipitation and runoff (Figs. 4b, 6b). Future droughts in the tropics have received a lot of attention due to their potential implications for the carbon cycle (Lewis et al. 2011; Malhi et al. 2008). Our analysis indicates a need to consider multiple models in studies analyzing drought impacts in the tropics due to a large range of model responses. For soil moisture, large absolute model differences are evident across much of the Southern Hemisphere and parts of North America and Eurasia (Fig. 8b).

While models were shown to agree well on precipitation droughts, runoff and soil moisture are

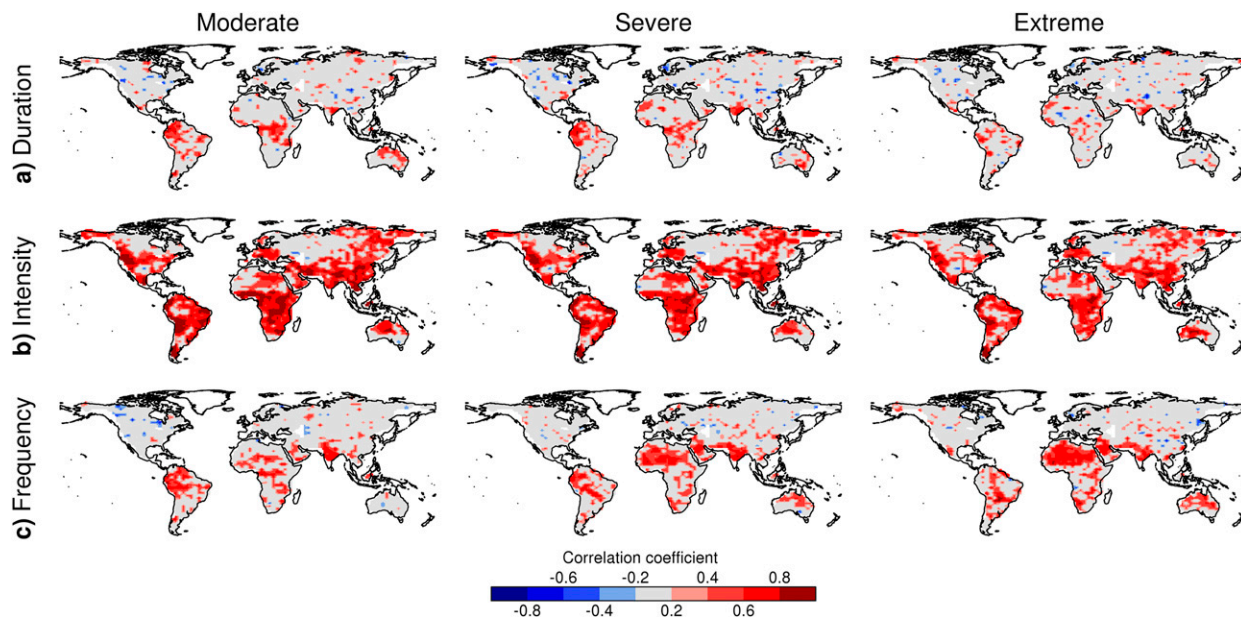


FIG. 9. Correlation between runoff and precipitation drought metrics for (a) mean duration, (b) mean absolute intensity, and (c) frequency of short droughts. The maps show Pearson correlation coefficients for each grid cell with nonsignificant ( $p > 0.1$ ) correlations shown in gray. Grid cells where no droughts were detected in  $>33\%$  of the models were masked out.

arguably more meaningful indicators of drought impacts than precipitation. Our results show that the models disagree on these metrics more strongly than precipitation droughts. Some studies have shown robust changes in climate model projections of runoff and soil moisture droughts (Koirala et al. 2014; Cook et al. 2015), but others have shown highly divergent projections

and historical simulations regionally (Orlowsky and Seneviratne 2013; Burke and Brown 2008; Sheffield and Wood 2008). Even if the models agree on changes in drought, our findings show that they predict vastly different drought lengths and intensities for soil moisture and runoff droughts in some regions (Figs. 6, 8). While we do not compare these quantities to observations and

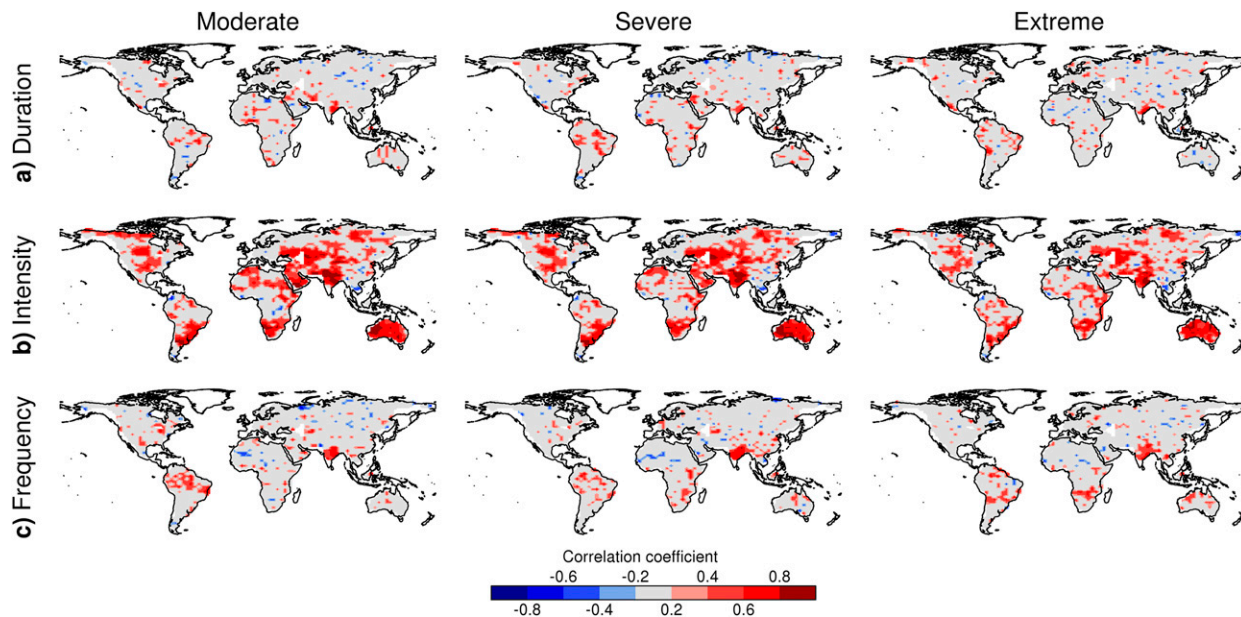


FIG. 10. As in Fig. 9, but for correlation between soil moisture and precipitation drought metrics.

thus do not establish the best performing models, we show that individual models commonly vary  $>30\%$  in either direction from the ensemble mean globally (Figs. 5, 7). This means a spread of  $>60\%$  points between the models at either extreme, pointing to significant challenges in reliably applying these models for assessing drought impacts under future climate.

### *b. Role of internal variability*

We note that our results do not necessarily reflect the full uncertainty in the CMIP5 archive due to not all models reporting the required data. Our study includes approximately half of all the available CMIP5 models and only one ensemble member per model and thus may not fully represent variation among models or the effects of internal variability within model ensembles. We have further quantified the latter to explore the robustness of the model differences to internal variability. In the historical simulations analyzed here, sea surface temperatures (SSTs) are freely evolving and may not reflect the same decadal variability in the 55-yr period analyzed here. We therefore also analyzed the Atmospheric Model Intercomparison Project (AMIP) simulations for precipitation for 12 models where SSTs were prescribed during the available period 1979–2004 (Fig. S10, with historical simulations for the same time period shown in Fig. S11). The model biases are lower in the AMIP runs for drought intensity compared to the historical runs, but overall the two sets of simulations are highly similar, with precipitation drought intensity generally underestimated compared to observations. This points to SST differences as one source of model discrepancies but only partly explains the biases and intermodel differences.

We also analyzed additional ensemble members for those models with at least three ensemble members with different initial conditions. The way models are initialized could also lead to differences in internal variability, thus affecting the drought metrics calculated here. A comparison of drought metrics for the first ensemble member (r1i1p1) used here and other available initial condition ensemble members is shown in Fig. S12 for precipitation and Fig. S13 for runoff. The globally averaged drought metrics are generally highly consistent among different ensemble members from the same model for both runoff and precipitation. This implies that the discrepancies identified across different models are not merely an artifact of our sampling or driven by out-of-sync internal variability during the 55-yr period. This is likely due to the fairly short droughts analyzed here (using 3-month SDI), whereas longer (and rarer) droughts would likely be more strongly affected by mismatches in decadal variability (Ault et al. 2012).

Overall, we show considerable differences among this subset of models that do not arise from internal variability alone, with implications for robustness of future drought projections. Future changes in drought are commonly investigated using an ensemble average and estimating the agreement around it (Collins et al. 2013; Sheffield and Wood 2008; Prudhomme et al. 2014). The choice of models and ensemble members is often subjective and based on the available outputs, with studies considering a vastly different number of models [e.g., 25 used by Berg et al. (2017) and 41 by Nasrollahi et al. (2015)]. Our results imply that the ensemble average and its robustness would be highly dependent on the choice of models, particularly for metrics with large systematic individual model differences (e.g., soil moisture and surface runoff drought intensity; Figs. 7b, S7b), and in regions with large absolute model differences (e.g., drought intensity; Figs. 4b, 6b, 8b). This should be an important consideration, particularly when making policy-relevant statements on future projections based on a small subset of models in order to provide reliable, unbiased estimates of future droughts. Similarly, the level of agreement and magnitude of model spread depended on the metrics and drought severity. This points to a need to consider multiple drought metrics and thresholds in the absence of any standardized drought definition.

### *c. Directions for future evaluation of climate models for drought*

We have analyzed CMIP5 model agreement against the CMIP5 mean (where observations are not available) to ensure a common methodology for each of the types of drought assessed. Clearly, future work should evaluate these models more thoroughly against observations as they become available. This and previous studies have compared CMIP5 simulations of precipitation droughts against global observations (Orlowsky and Seneviratne 2013; Nasrollahi et al. 2015), although previous work has mainly concentrated on drought trends. Fewer studies have attempted to evaluate CMIP5 runoff or soil moisture droughts. Our results indicate that evaluating models for precipitation droughts is insufficient for understanding the performance of climate models for simulating the types of droughts that have the most impact on societies and ecosystems. It is thus critical to evaluate models for the variable of interest directly. Past studies have evaluated climate models for soil moisture from the top few layers ( $<10$  cm; Huang et al. 2016) for which remote sensing observations are available, or against offline models (Sheffield and Wood 2008; Wuebbles et al. 2014). However, moisture from the top few layers offers limited information on soil moisture stress as experienced

by vegetation that have access to deeper water content (Koster et al. 2009). Indeed, Berg et al. (2017) recently showed differing future projections for surface and total soil moisture in CMIP5 models, with changes in opposite directions in some regions (although regional studies have indicated both discrepancies and agreement in top and deep soil moisture; Cheng et al. 2016; Cook et al. 2015). Datasets such as GRACE may provide future opportunities for evaluating simulations of soil moisture droughts as longer data records are gathered. However, without thorough evaluation, only limited confidence can be placed upon future projections of these types of droughts, even where model agreement is high.

We note that the soil moisture outputs used in this study were integrated over the top 3 m, a depth that does not necessarily correspond to vegetation root profile assumed in models. There is no easy solution to this issue when interpreting CMIP5 outputs, and this assumption is commonly used in other studies (Swann et al. 2016; Berg et al. 2017). Setting the depth to the maximum rooting depth would not necessarily have captured more accurate soil moisture dynamics, as different models access water from the root-zone differently based on rooting depth (De Kauwe et al. 2015). Our assumption likely dampens the intensity and frequency of soil moisture droughts compared to a shallower soil depth (Cheng et al. 2016), and it would thus be worthwhile to explore alternative methods and the implications of this on quantifying vegetation water stress in future work. However, we also showed that the model differences in soil moisture drought intensity relative to the ensemble mean could not be merely explained by different soil depths in the underlying LSMs.

We have not explicitly evaluated evapotranspiration (ET), which is also a large component of the hydrological cycle and important in determining water availability on land (Hobbins et al. 2016), because ET droughts cannot be directly quantified using metrics such as SPI due to the dependency of ET on both water and energy availability (see section 2c). Evaluation of offline LSMs has shown large and systematic biases in ET during water-stressed periods, with ET being strongly underestimated across many biomes (grasslands, savannas, and forests) and climate regimes (ranging from arid to humid) compared to observations (Powell et al. 2013; Haverd et al. 2016; Ukkola et al. 2016; De Kauwe et al. 2015; Li et al. 2012; Whitley et al. 2016; Keenan et al. 2009). Mueller and Seneviratne (2014) indicated similar biases in CMIP5 models, although they did not specifically evaluate the models during drought periods. Improving ET biases is likely to be an important step for improving simulations of drought.

Overall, our results showed that the model biases, and agreement with the ensemble mean, were consistent across climate models sharing an LSM or an atmospheric scheme. This result was consistent across global means (Figs. 2, 5, 7) as well as spatial correlations using hierarchical clustering (Figs. 3, S4). Models clustered strongly by their component models for both intensity and duration, but the distances were generally lower for intensity, indicating higher spatial similarity for intensity for a given LSM or atmospheric scheme. This comparison is somewhat superficial due to the lack of detail available on the specific schemes used in each model (e.g., model version or parameters), and we acknowledge that the same schemes may have been run using different settings (e.g., NorESM1-ME simulates the land carbon cycle, whereas NorESM1-M does not; Flato et al. 2013, their Table 9.1). It is also not possible to separate the contributions from the LSMs and atmospheric schemes in the CMIP5 archive, although it is reasonable to assume the LSM has a dominant effect on runoff and soil moisture processes. As such, our results suggest that the LSMs are a significant source of discrepancies in runoff and soil moisture drought simulations despite these limitations. Past offline studies have identified a number of factors driving model differences, including the representation of water stress on evapotranspiration, the treatment of soil hydrology and hydraulics (including groundwater dynamics), root dynamics, and lack of “persistence” (Haverd et al. 2016; Whitley et al. 2016, 2017; Li et al. 2012; Ukkola et al. 2016; De Kauwe et al. 2015; Tallaksen and Stahl 2014; Powell et al. 2013). Clearly, further work is required to better understand the key processes governing runoff and soil moisture droughts in LSMs. However, while the CMIP5 archive is crucial for quantifying future changes in droughts, it offers limited opportunity for understanding such sources of model differences because of the difficulty of disentangling relevant processes in these simulations and a lack of relevant land outputs and parameter fields (e.g., soil texture). Similarly, simulations using offline models are alone likely insufficient for evaluating model behavior due to the lack of important feedbacks with the atmosphere. As such, there is a need for a targeted model intercomparison project (MIP) that allows the separation of land and atmospheric processes and their feedbacks to complement existing online and offline MIPs (e.g., Global Soil Wetness Project Phase 3; <http://hydro.iis.u-tokyo.ac.jp/GSWP3/>). The Land Surface, Snow and Soil Moisture Model Intercomparison Project (LS3MIP; van den Hurk et al. 2016) is likely to provide such a platform for both improving the representation of key processes in the models and better understanding the sources of model differences. The incorporation of climate model simulations in LS3MIP may help better evaluate these models in



terms of their capability to simulate different types of droughts, which would be a crucial step in increasing confidence in future simulations of drought.

## 5. Conclusions

We conclude that CMIP5 models show large, metric-dependent discrepancies in precipitation, runoff, and soil moisture droughts during the historical period (1950–2004). Studies examining only one of precipitation, runoff, or soil moisture drought will likely lead to very different conclusions and levels of confidence (in terms of model agreement) over those that analyze all three simultaneously. We also note that the relationship between these three types of drought, and indeed the duration, frequency, and intensity of each of them, varies depending on which CMIP5 models are examined. This suggests the need to consider multiple models in drought impact studies to account for large model discrepancies.

By analyzing different components of the hydrological cycle together, we showed that the differences in simulated runoff and soil moisture droughts are not driven merely by differences in precipitation. Behavior did in many instances correlate with component models used within the coupled modeling system. Yet, more effort is clearly required to systematically evaluate models for drought to identify the sources of uncertainty in drought-related processes. We suggest this is essential for guiding future model development and reducing uncertainties in future projections of drought.

We note, in the context of future CMIP initiatives, the need for clarity on the LSMs used, the version, and ideally a pointer to a code repository that enables subsequent process-level analysis. We also note the need for detailed metadata to be available with the model simulations (e.g., settings and parameters used). Without this, analyses of the kind we present here are inevitably limited. Given the significance of the question of how drought will change in the future, the benefits of modeling groups clearly identifying the LSMs used, the version, and the associated metadata seem to strongly outweigh any costs.

*Acknowledgments.* This work was supported by the Australian Research Council Centre of Excellence for Climate System Science (CE110001028). A. J. Pitman and M. G. De Kauwe acknowledge support from the Australian Research Council Centre of Excellence for Climate Extremes (CE170100023). We acknowledge the National Computational Infrastructure at the Australian National University and the Earth System Grid Federation for making the CMIP5 model outputs available. We also gratefully acknowledge the World

Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. The GPCP precipitation data were acquired from NOAA/OAR/ESRL PSD, Boulder, Colorado (<http://www.esrl.noaa.gov/psd/>).

## REFERENCES

- Alexander, L. V., and J. M. Arblaster, 2009: Assessing trends in observed and modelled climate extremes over Australia in relation to future projections. *Int. J. Climatol.*, **29**, 417–435, <https://doi.org/10.1002/joc.1730>.
- Ault, T. R., J. E. Cole, and S. St. George, 2012: The amplitude of decadal to multidecadal variability in precipitation simulated by state-of-the-art climate models. *Geophys. Res. Lett.*, **39**, L21705, <https://doi.org/10.1029/2012GL053424>.
- , J. S. Mankin, B. I. Cook, and J. E. Smerdon, 2016: Relative impacts of mitigation, temperature, and precipitation on 21st-century megadrought risk in the American Southwest. *Sci. Adv.*, **2**, e1600873, <https://doi.org/10.1126/sciadv.1600873>.
- Berg, A., J. Sheffield, and P. C. D. Milly, 2017: Divergent surface and total soil moisture projections under global warming. *Geophys. Res. Lett.*, **44**, 236–244, <https://doi.org/10.1002/2016GL071921>.
- Burke, E. J., and S. J. Brown, 2008: Evaluating uncertainties in the projection of future drought. *J. Hydrometeorol.*, **9**, 292–299, <https://doi.org/10.1175/2007JHM929.1>.
- Chen, H., J. Sun, and X. Chen, 2014: Projection and uncertainty analysis of global precipitation-related extremes using CMIP5 models. *Int. J. Climatol.*, **34**, 2730–2748, <https://doi.org/10.1002/joc.3871>.
- Cheng, L., M. Hoerling, A. Aghakouchak, B. Livneh, X. W. Quan, and J. Eischeid, 2016: How has human-induced climate change affected California drought risk? *J. Climate*, **29**, 111–120, <https://doi.org/10.1175/JCLI-D-15-0260.1>.
- Collins, M., and Coauthors, 2013: Long-term climate change: Projections, commitments and irreversibility. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1029–1136.
- Cook, B. I., T. R. Ault, and J. E. Smerdon, 2015: Unprecedented 21st century drought risk in the American Southwest and Central Plains. *Sci. Adv.*, **1**, e1400082, <https://doi.org/10.1126/sciadv.1400082>.
- Dai, A., 2011: Drought under global warming: A review. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 45–65, <https://doi.org/10.1002/wcc.81>.
- , 2013: Increasing drought under global warming in observations and models. *Nat. Climate Change*, **3**, 52–58, <https://doi.org/10.1038/nclimate1633>.
- Decker, M., S. Ma, and A. Pitman, 2017: Local land–atmosphere feedbacks limit irrigation demand. *Environ. Res. Lett.*, **12**, 054003, <https://doi.org/10.1088/1748-9326/aa65a6>.
- De Kauwe, M. G., S.-X. Zhou, B. E. Medlyn, A. J. Pitman, Y. P. Wang, R. A. Duursma, and I. C. Prentice, 2015: Do land

- surface models need to include differential plant species responses to drought? Examining model predictions across a latitudinal gradient in Europe. *Biogeosciences*, **12**, 12349–12393, <https://doi.org/10.5194/bgd-12-12349-2015>.
- Donat, M. G., A. J. Pitman, and S. I. Seneviratne, 2017: Regional warming of hot extremes accelerated by surface energy fluxes. *Geophys. Res. Lett.*, **44**, 7011–7019, <https://doi.org/10.1002/2017GL073733>.
- Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866, <https://doi.org/10.1017/CBO9781107415324>.
- Gibson, P., S. Perkins-Kirkpatrick, L. Alexander, and E. Fischer, 2017: Comparing Australian heat waves in the CMIP5 models through cluster analysis. *J. Geophys. Res. Atmos.*, **122**, 3266–3281, <https://doi.org/10.1002/2016JD025878>.
- Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister, 2014: Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset. *Int. J. Climatol.*, **34**, 623–642, <https://doi.org/10.1002/joc.3711>.
- Haverd, V., M. Cuntz, L. P. Nieradzik, and I. N. Harman, 2016: Improved representations of coupled soil-canopy processes in the CABLE land surface model (subversion revision 3432). *Geosci. Model Dev.*, **9**, 3111–3122, <https://doi.org/10.5194/gmd-9-3111-2016>.
- Hobbins, M., A. Wood, D. McEvoy, J. Huntington, C. Morton, J. Verdin, M. Anderson, and C. Hain, 2016: The Evaporative Demand Drought Index. Part I: Linking drought evolution to variations in evaporative demand. *J. Hydrometeorol.*, **17**, 1745–1761, <https://doi.org/10.1175/JHM-D-15-0121.1>.
- Huang, Y., S. Gerber, T. Huang, and J. W. Lichstein, 2016: Evaluating the drought response of CMIP5 models using global gross primary productivity, leaf area, precipitation and soil moisture data. *Global Biogeochem. Cycles*, **30**, 1827–1846, <https://doi.org/10.1002/2016GB005480>.
- Keenan, T., R. García, D. Friend, S. Zaehle, C. Gracia, and S. Sabate, 2009: Improved understanding of drought controls on seasonal variation in Mediterranean forest canopy CO<sub>2</sub> and water fluxes through combined in situ measurements and ecosystem modelling. *Biogeosciences*, **6**, 1423–1444, <https://doi.org/10.5194/bg-6-1423-2009>.
- Koirala, S., Y. Hirabayashi, R. Mahendran, and S. Kanae, 2014: Global assessment of agreement among streamflow projections using CMIP5 model outputs. *Environ. Res. Lett.*, **9**, 064017, <https://doi.org/10.1088/1748-9326/9/6/064017>.
- Koster, R. D., and Coauthors, 2006: GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview. *J. Hydrometeorol.*, **7**, 590–610, <https://doi.org/10.1175/JHM510.1>.
- , Z. Guo, R. Yang, P. A. Dirmeyer, K. Mitchell, and M. J. Puma, 2009: On the nature of soil moisture in land surface models. *J. Climate*, **22**, 4322–4335, <https://doi.org/10.1175/2009JCLI2832.1>.
- Lewis, S. L., P. M. Brando, O. L. Phillips, G. M. F. Van der Heijden, and D. Nepstad, 2011: The 2010 Amazon drought. *Science*, **331**, 554, <https://doi.org/10.1126/science.1200807>.
- Li, L., Y.-P. Wang, Q. Yu, B. Pak, D. Eamus, J. Yan, E. van Gorsel, and I. T. Baker, 2012: Improving the responses of the Australian community land surface model (CABLE) to seasonal drought. *J. Geophys. Res.*, **117**, G04002, <https://doi.org/10.1029/2012JG002038>.
- Malhi, Y., J. T. Roberts, R. A. Betts, T. J. Killeen, W. Li, and C. A. Nobre, 2008: Climate change, deforestation, and the fate of the Amazon. *Science*, **319**, 169–172, <https://doi.org/10.1126/science.1146961>.
- McKee, T. B., N. J. Doesken, and J. Kleist, 1993: The relationship of drought frequency and duration to time scales. Preprints, *Eighth Conf. on Applied Climatology*, Anaheim, CA, Amer. Meteor. Soc., 179–184.
- Milly, P. C. D., and K. A. Dunne, 2016: Potential evapotranspiration and continental drying. *Nat. Climate Change*, **6**, 946–949, <https://doi.org/10.1038/nclimate3046>.
- , and —, 2017: A hydrologic drying bias in water-resource impact analyses of anthropogenic climate change. *J. Amer. Water Resour. Assoc.*, **53**, 822–838, <https://doi.org/10.1111/1752-1688.12538>.
- Mueller, B., and S. I. Seneviratne, 2014: Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophys. Res. Lett.*, **41**, 128–134, <https://doi.org/10.1002/2013GL058055>.
- Nasrollahi, N., A. AghaKouchak, L. Cheng, L. Damberg, T. Phillips, C. Miao, K. Hsu, and S. Sorooshian, 2015: How well do CMIP5 climate simulations replicate historical trends and patterns of meteorological droughts? *Water Resour. Res.*, **51**, 2847–2864, <https://doi.org/10.1002/2014WR016318>.
- Orlowsky, B., and S. I. Seneviratne, 2012: Global changes in extreme events: Regional and seasonal dimension. *Climatic Change*, **110**, 669–696, <https://doi.org/10.1007/s10584-011-0122-9>.
- , and —, 2013: Elusive drought: Uncertainty in observed trends and short-and long-term CMIP5 projections. *Hydrol. Earth Syst. Sci.*, **17**, 1765–1781, <https://doi.org/10.5194/hess-17-1765-2013>.
- Orth, R., and S. I. Seneviratne, 2012: Analysis of soil moisture memory from observations in Europe. *J. Geophys. Res.*, **117**, D15115, <https://doi.org/10.1029/2011JD017366>.
- Powell, T. L., and Coauthors, 2013: Confronting model predictions of carbon fluxes with measurements of Amazon forests subjected to experimental drought. *New Phytol.*, **200**, 350–365, <https://doi.org/10.1111/nph.12390>.
- Prudhomme, C., S. Parry, J. Hannaford, D. B. Clark, S. Hagemann, and F. Voss, 2011: How well do large-scale models reproduce regional hydrological extremes in Europe? *J. Hydrometeorol.*, **12**, 1181–1204, <https://doi.org/10.1175/2011JHM1387.1>.
- , and Coauthors, 2014: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proc. Natl. Acad. Sci. USA*, **111**, 3262–3267, <https://doi.org/10.1073/pnas.1222473110>.
- Rocheta, E., M. Sugiyanto, F. Johnson, J. Evans, and A. Sharma, 2014: How well do general circulation models represent low-frequency rainfall variability? *Water Resour. Res.*, **50**, 2108–2123, <https://doi.org/10.1002/2012WR013085>.
- Roderick, M. L., P. Greve, and G. D. Farquhar, 2015: On the assessment of aridity with changes in atmospheric CO<sub>2</sub>. *Water Resour. Res.*, **51**, 5450–5463, <https://doi.org/10.1002/2015WR017031>.
- Schewe, J., and Coauthors, 2014: Multimodel assessment of water scarcity under climate change. *Proc. Natl. Acad. Sci. USA*, **111**, 3245–3250, <https://doi.org/10.1073/pnas.1222460110>.
- Schneider, U., M. Ziese, A. Meyer-Christoffer, P. Finger, E. Rustemeier, and A. Becker, 2016: The new portfolio of global precipitation data products of the Global Precipitation Climatology Centre suitable to assess and quantify the global water cycle and resources. *Proc. Int. Assoc. Hydrol. Sci.*, **374**, 29–34, <https://doi.org/10.5194/piahs-374-29-2016>.
- Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, 2010: Investigating soil moisture–climate interactions in a changing climate: A



- review. *Earth Sci. Rev.*, **99**, 125–161, <https://doi.org/10.1016/j.earscirev.2010.02.004>.
- Sheffield, J., and E. F. Wood, 2008: Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations. *Climate Dyn.*, **31**, 79–105, <https://doi.org/10.1007/s00382-007-0340-z>.
- Shukla, S., and A. W. Wood, 2008: Use of a standardized runoff index for characterizing hydrologic drought. *Geophys. Res. Lett.*, **35**, L02405, <https://doi.org/10.1029/2007GL032487>.
- Sillmann, J., V. V. Kharin, F. W. Zwiers, X. Zhang, and D. Bronaugh, 2013: Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *J. Geophys. Res. Atmos.*, **118**, 2473–2493, <https://doi.org/10.1002/jgrd.50188>.
- Stagge, J. H., L. M. Tallaksen, L. Gudmundsson, A. F. Van Loon, and K. Stahl, 2016: Candidate distributions for climatological drought indices (SPI and SPEI). *Int. J. Climatol.*, **36**, 2132–2138, <https://doi.org/10.1002/joc.4564>.
- Swann, A. L. S., F. M. Hoffman, C. D. Koven, and J. T. Randerson, 2016: Plant responses to increasing CO<sub>2</sub> reduce estimates of climate impacts on drought severity. *Proc. Natl. Acad. Sci. USA*, **113**, 10 019–10 024, <https://doi.org/10.1073/pnas.1604581113>.
- Tallaksen, L. M., and K. Stahl, 2014: Spatial and temporal patterns of large-scale droughts in Europe: Model dispersion and performance. *Geophys. Res. Lett.*, **41**, 429–434, <https://doi.org/10.1002/2013GL058573>.
- Tapley, B. D., S. Bettadpur, M. Watkins, and C. Reigber, 2004: The gravity recovery and climate experiment: Mission overview and early results. *Geophys. Res. Lett.*, **31**, L09607, <https://doi.org/10.1029/2004GL019920>.
- Taylor, I. H., E. Burke, L. McColl, P. D. Falloon, G. R. Harris, and D. McNeill, 2013: The impact of climate mitigation on projections of future drought. *Hydrol. Earth Syst. Sci.*, **17**, 2339–2358, <https://doi.org/10.5194/hess-17-2339-2013>.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Touma, D., M. Ashfaq, M. A. Nayak, S. C. Kao, and N. S. Diffenbaugh, 2015: A multi-model and multi-index evaluation of drought characteristics in the 21st century. *J. Hydrol.*, **526**, 196–207, <https://doi.org/10.1016/j.jhydrol.2014.12.011>.
- Ukkola, A. M., M. G. De Kauwe, A. J. Pitman, M. J. Best, G. Abramowitz, V. Haverd, M. Decker, and N. Haughton, 2016: Land surface models systematically overestimate the intensity, duration and magnitude of seasonal-scale evaporative droughts. *Environ. Res. Lett.*, **11**, 104012, <https://doi.org/10.1088/1748-9326/11/10/104012>.
- van den Hurk, B., and Coauthors, 2016: LS3MIP (v1.0) contribution to CMIP6: The Land Surface, Snow and Soil Moisture Model Intercomparison Project—Aims, setup and expected outcome. *Geosci. Model Dev.*, **9**, 2809–2832, <https://doi.org/10.5194/gmd-9-2809-2016>.
- van Huijgevoort, M. H. J., and Coauthors, 2013: Global multimodel analysis of drought in runoff for the second half of the twentieth century. *J. Hydrometeorol.*, **14**, 1535–1552, <https://doi.org/10.1175/JHM-D-12-0186.1>.
- Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno, 2010: A multiscale drought index sensitive to global warming: The Standardized Precipitation Evapotranspiration Index. *J. Climate*, **23**, 1696–1718, <https://doi.org/10.1175/2009JCLI2909.1>.
- Whitley, R., and Coauthors, 2016: A model inter-comparison study to examine limiting factors in modelling Australian tropical savannas. *Biogeosciences*, **13**, 3245–3265, <https://doi.org/10.5194/bg-13-3245-2016>.
- , and Coauthors, 2017: Challenges and opportunities in modelling savanna ecosystems. *Biogeosciences*, **14**, 4711–4732, <https://doi.org/10.5194/bg-14-4711-2017>.
- Wuebbles, D., and Coauthors, 2014: CMIP5 climate model analyses: Climate extremes in the United States. *Bull. Amer. Meteor. Soc.*, **95**, 571–583, <https://doi.org/10.1175/BAMS-D-12-00172.1>.
- Yin, D., M. L. Roderick, G. Leech, F. Sun, and Y. Huang, 2014: The contribution of reduction in evaporative cooling to higher surface air temperatures during drought. *Geophys. Res. Lett.*, **41**, 7891–7897, <https://doi.org/10.1002/2014GL062039>.
- Zhang, L., K. Hickel, W. R. Dawes, F. H. S. Chiew, A. W. Western, and P. R. Briggs, 2004: A rational function approach for estimating mean annual evapotranspiration. *Water Resour. Res.*, **40**, W02502, <https://doi.org/10.1029/2003WR002710>.
- Zhao, T., and A. Dai, 2015: The magnitude and causes of global drought changes in the twenty-first century under a low-moderate emissions scenario. *J. Climate*, **28**, 4490–4512, <https://doi.org/10.1175/JCLI-D-14-00363.1>.