AGU100 ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR022792

Special Section:
Hydrology delivers Earth System Sciences to Society (HESSS4): Improving and Integrating Knowledge across Disciplines on Global Energy, Water and Carbon Cycles

The Effect of Global Warming on Future Water Availability: CMIP5 Synthesis

C. R. Ferguson[1,2], M. Pan[3], and Taikan Oki[4,5]

[1]Atmospheric Sciences Research Center, University at Albany, State University of New York, Albany, NY, USA, [2]Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, Albany, NY, USA, [3]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA, [4]Department of Hydrology and Water Resources Engineering, Institute of Industrial Science, University of Tokyo, Tokyo, Japan, [5]Institute for the Advanced Study of Sustainability, United Nations University, Shibuya, Japan

Key Points:
- CMIP5 projections of significant subseasonal changes in water availability are often overlooked and underestimated
- Carbon-cycling Earth system models project 23% fewer significant shifts in future water availability than general circulation models
- The magnitude of model errors, as quantified through observational comparison, may preclude the credibility of projections for some basins

Supporting Information:
- Supporting Information S1

Correspondence to:
C. R. Ferguson,
crferguson@albany.edu

Citation:
Ferguson, C. R., Pan, M., & Oki, T. (2018). The effect of global warming on future water availability: CMIP5 synthesis. *Water Resources Research*, *54*, 7791–7819. https://doi.org/10.1029/2018WR022792

Received 17 FEB 2018
Accepted 8 SEP 2018
Accepted article online 21 SEP 2018
Published online 16 OCT 2018

©2018. American Geophysical Union. All Rights Reserved.

Abstract Changes in the spatiotemporal dynamics of the global water cycle will constitute some of the greatest challenges to socioeconomic-environmental well-being in a warming world. Large multimodel, multiscenario intercomparisons such as the Coupled Model Intercomparison Project Phase 5 (CMIP5) experiment support our best estimates of projected climate change and associated uncertainty thereof. It is important to continually reevaluate how this information is synthesized and communicated and at what point it becomes actionable. In this study, we demonstrate a systematic and holistic framework for synthesizing multimodel ensemble projections of water availability at large river basin scale—the scale at which water resources are both managed and monitored. We identify statistically significant shifts in mean water availability at annual and monthly scales, its interannual variations, and its relative seasonality, as computed from CMIP5 historical (1976–2005) and Representative Concentration Pathway 8.5 (2070–2099) scenario multimodel ensemble output. Water availability is addressed separately through the lens of meteorologists (precipitation), hydrologists (runoff), and agriculturalists (precipitation minus evapotranspiration). We illustrate limitations in CMIP5 model representativeness through comparisons of atmosphere-only model (Atmospheric Model Intercomparison Project) output against observational best estimates. And we find that warming-induced shifts in water availability projected by CMIP5 carbon-cycling Earth system models are comparatively less substantial than those projected by traditional general circulation models. As we show, knowing the seasonality of both projected changes and of the biased model background climatology onto which they are imposed is paramount to ensuring proper interpretation and ascribing confidence.

1. Introduction

The World Climate Research Programme has identified Water for the Food Baskets of the World as one of seven Grand Challenges for the coming decade. Currently, more than 2 billion people are affected by water stress, a number that will only increase with population growth and continued economic development (United Nations, 2018). Feeding the world's growing and more affluent population will require a 25–70% increase in agricultural production by 2050 and a corresponding increase in agricultural water withdrawals, which already account for 70% of global water demand (Hunter et al., 2017). Climate change will exacerbate the impending water resources management challenge in some regions by reducing total physical water availability and altering the river flow regime (Döll & Schmied, 2012).

Climate models facilitate understanding of how a warming climate may affect the distribution of freshwater globally. The Coupled Model Intercomparison Project Phase 5 (CMIP5; Taylor et al., 2012) constitutes the most extensive database of climate simulations ever compiled and served as the model projection backbone to the Intergovernmental Panel on Climate Change (IPCC, 2013) Fifth Assessment Report. Soon, the CMIP5/AR5 effort will be dwarfed by the CMIP6/AR6 effort (Eyring, Bony, et al., 2016). Given the tremendous level of international investment in these activities, it is meaningful to consider whether in the context of Water for the Food Baskets of the World they are being fully leveraged, or if additional value might be gleaned from them? An unfortunate reality for field campaigns and

model intercomparison projects alike is that resources are disproportionately allocated to data collection and generation as opposed to follow-on in-depth analysis and synthesis. The challenge is to efficiently synthesize the data into a form that balances ease of communication and sufficient detail such that it becomes actionable.

Numerous studies before ours have analyzed CMIP5's simulations from a water resources perspective (e.g., Arnell & Gosling, 2013; Dai, 2013; Koirala et al., 2014; Maloney et al., 2014; Orlowsky & Seneviratne, 2013; Sheffield, Camargo et al., 2013; Sheffield, Barrett et al., 2013). Most often, studies have focused on reporting projected trends and/or long-term change in the annual means in the format of global maps. In the AR5, global maps of projected seasonal mean precipitation (P) and annual mean runoff (Q) and evapotranspiration (E) change were presented (Collins et al., 2013; their Figures 12.22 and 12.24–12.25). Alternatively, spatial averages computed over large zonal, continental, or even global land and global ocean domains have been reported (e.g., Lau et al., 2013; Wills et al., 2016). Such studies have in some cases contributed to broad generalizations about the future climate, such as "dry gets drier, wet gets wetter" (e.g., Chou et al., 2009; Held & Soden, 2006; Liu & Allan, 2013) that are misleading and false (Greve et al., 2014). By contrast, detailed studies have been conducted to assess climate response in a particular region or river basin, or small subset of basins (e.g., Jimenez Cisneros et al., 2014; Masood et al., 2015; Vetter et al., 2015). These studies tend to focus on only a single definition of water availability, namely, in terms of P for meteorologists, Q for hydrologists, and precipitation minus evaporation ($P - E$) for agriculturalists and ecologists, and apply unique statistical and model selection procedures. Thus, all three terms—P, $P - E$, and Q—are generally not analyzed jointly and consistently at a regional scale. Overall, studies of all forms typically fall short with regard to quantifying model uncertainty and result significance.

In this article, we present a rigorous and statistically sound assessment of the impact of projected climate change on the Earth's future water resources for multiple large-scale basins using CMIP5 data. The strength of the approach derives from its comprehensiveness: the inclusion of all available CMIP5 models and their realizations, simultaneous analysis of multiple water budget variables, error characterization, and the vetting of all results by appropriate statistical significance tests. The approach fills gaps in standard studies and may be easily automated for application in diagnostic toolkits such as the Earth system model (ESM) evaluation tool (ESMValTool; Eyring, Righi, et al., 2016). Specifically, the approach covers annual and subseasonal means and their variability, relative seasonality, and model uncertainty for P, E, $P - E$, and Q—all at the scale of large river basins. Large basins, lie at a trade-off scale that is local enough to be actionable to decision makers and water planners yet broad enough to be represented by coarse, 100- to 200-km resolution CMIP5 models. Moreover, basins aggregate climate response across heterogeneous landscapes and enforce a water budget that is monitored in many basins through streamflow and precipitation gauge networks, or remotely with the aid of space-borne sensors (e.g., Gao et al., 2010; Lettenmaier et al., 2015; Pan et al., 2012; Pokhrel et al., 2013; Rodell et al., 2009; Sheffield et al., 2009).

Knowing the seasonality of projected changes to water availability is critical (e.g., Chou & Lan, 2012; Tabari & Willems, 2018). More important than knowing whether a basin will turn drier or wetter on annual mean timescales is knowing the timing of that drying and wetting. Will the dry season become drier and the wet season wetter, or the dry season wetter and the wet season drier? Certainly, in the context of agriculture and the virtual water trade, the value of water is not fixed over the course of the year (e.g., Dalin et al., 2012; Hanasaki et al., 2010; Oki et al., 2017; Oki & Kanae, 2004).

Fitness of the model or model ensemble to the task is also an important consideration. Projected shifts in the seasonality and interannual variability of water availability must be interpreted in the context of the underlying modeled seasonality upon which they are imposed. This concept ties-in to the modeler's responsibility to communicate relative confidence and/or statistical significance of any projected changes in water availability, even if such estimates are based on incomplete information. In our case, CMIP5 ensemble model skill from the Atmospheric Model Intercomparison Project (AMIP; 1979–2008) experiment serves as a surrogate for model fidelity in end-of-the-century (2070–2099) simulations.

We begin in section 2 by describing the CMIP5 model output, observational comparison data, and statistical methods. The results are presented in section 3, and a summary is provided in section 4. We note that there are an extensive number of figures in the supporting information for the interested reader.

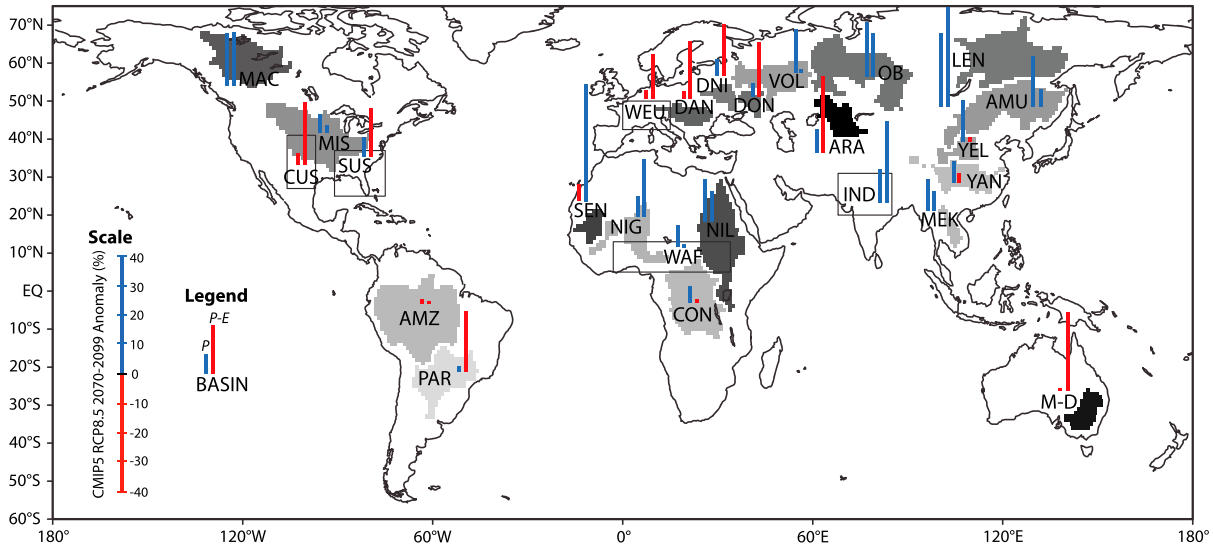

Figure 1. Global distribution of the study basins/regions and CMIP5 RCP8.5 (2070–2099) projected changes to the HIST (1976–2005) annual mean precipitation (P; first bar) and precipitation minus evaporation ($P - E$; second bar). Red and blue coloring indicates projected drying and wetting, respectively. The precise bounding boxes for the three land-atmosphere coupling hotspots (LACH) and two regions of interest are CUS: 27–41°N, 106–97°W; SUS: 25–37°N, 91–75°W; WAF: 5–13°N, 3°W to 34°E; WEU: 42.5–50°N, 0–15°E; IND: 20–31°N, 68–85°E. See Table 1 for a listing of basin abbreviations and corresponding numerical values.

2. Data and Methods

2.1. Study Area

Our study focuses on 20 major river basins distributed globally; three climate-sensitive, semiarid, midcontinental regions; and two additional regions of interest (Figure 1). Basins are natural spatial aggregators of heterogeneous landscape processes and their response to climate; they impose a natural water budget, which facilitates human monitoring and management. If changes in terrestrial water storage are assumed negligible, basin streamflow will equal basin total precipitation minus evapotranspiration on timescales greater than 1 year. This is important for verification of E because spatially distributed measurements of E (i.e., from flux towers) are generally lacking, even when a sufficient network of P and Q gauges are available. The basins span the full spectrum of water availability (Table 1) and size (DON: 440,000 km^2; AMZ: 6,145,000 km^2). See Ferguson et al. (2012; their Table 3) for descriptive hydroclimate and vegetative characteristics. The three climate sensitive regions—the central United States, West Africa, and India—correspond with Global Land-Atmosphere Coupling Experiment-1 AGCM-based land-atmosphere coupling hotspots (Guo et al., 2006; Koster et al., 2004; Koster et al., 2006). The southern U.S. (SUS) and western Europe (WEU) domains were selected to correspond to the areal extent of recent severe droughts in those regions. Namely, the 2006–2008 southeastern U.S. drought and the 2011 European drought. Coincidentally, SUS corresponds with the South Atlantic-Gulf Basin.

2.2. CMIP5 Model Output and Evaluation Periods

We evaluate differences in 30-year climate normals of P, E, $P - E$, and Q for recent (1976–2005) and near-future (2070–2099) periods, as calculated from CMIP5 historical (HIST) and Representative Concentration Pathway 8.5 (RCP8.5) experiments, respectively. The RCP8.5 pathway corresponds to a business-as-usual scenario that does not include any specific climate mitigation target (e.g., Riahi et al., 2011; Taylor et al., 2012). The CMIP standard output variable names of interest are: *pr* (P), *evspsbl* (E), *mrro* (Q), and *sftlf* (land area fraction). Outputs for all models are first bilinearly downscaled to a common 0.125° lat/lon grid. Then the water budget variables are area averaged over the basin fraction for which the model-specific downscaled land fraction is greater than 0.5. The multimodel ensemble mean is calculated by first averaging available realizations for each individual model and then averaging those together (Knutti, Abramowitz, et al., 2010). Realizations constitute model integrations initialized at different times but with identical physics parameterizations.

Table 1
Summary of the Study Basins/Regions, Their Abbreviations, and Their Historical (1976–2005) Water Availability and RCP8.5 (2070–2099) Projected Anomalies According to the CMIP5 Multimodel Ensemble

		P		*P* − *E*		*Q*	
Basin/Region	Abbr.	HIST (mm/year)	RCP8.5 Anomaly (%)	HIST (mm/year)	RCP8.5 Anomaly (%)	HIST (mm/year)	RCP8.5 Anomaly (%)
Senegal	SEN	264	-6	26	40	3	9
Aral	ARA	332	8	62	−26	5	−21
Lena	LEN	499	25	215	34	16	34
Mackenzie	MAC	536	18	206	18	15	21
Ob	OB	559	19	194	15	14	15
Don	DON	567	5	113	−19	8	−16
Nile	NIL	582	14	117	11	8	38
Amur	AMU	586	17	192	6	12	22
Murray-Darling	M-D	587	−1	52	−27	2	30
Niger	NIG	627	7	127	20	11	8
India LACH	IND	649	12	155	28	15	22
Volga	VOL	668	15	252	1	17	14
Dnieper	DNI	702	6	211	−18	14	−5
Central U.S. LACH	CUS	771	−4	83	−21	6	−30
Danube	DAN	799	−3	199	−20	15	−22
Yellow	YEL	832	14	326	−2	22	11
Mississippi	MIS	916	6	199	3	14	−1
Western Europe	WEU	1030	−3	380	−15	32	−13
Parana	PAR	1130	2	255	−21	15	7
West Africa LACH	WAF	1187	7	305	1	20	18
Southeastern U.S.	SUS	1286	7	341	−16	25	8
Yangtze	YAN	1420	8	694	−3	50	5
Mekong	MEK	1438	11	463	7	30	34
Amazon	AMZ	1766	−1	518	−1	41	−4
Congo	CON	1821	6	645	−1	45	13

Note. See Table S1 for participating model details. The acronym LACH stands for Land-Atmosphere Coupling Hotspot. CMIP5 = Coupled Model Intercomparison Project Phase 5; RCP8.5 = Representative Concentration Pathway 8.5.

Output from additional CMIP5 experiments is analyzed to lend greater context to the HIST-RCP8.5 results. Specifically, we test whether coupled ESMs produce significantly different projections of water availability. The CMIP5 ESMHIST and ESMRCP8.5 experiments include 15 ESMs with prognostic carbon-cycle capabilities. Such capabilities were "switched off" when these same ESMs were run as part of the HIST experiment, and they were forced by prescribed historical time series of atmospheric greenhouse-gas emissions, as distinct from other CMIP5 models that were forced by the historical time series of greenhouse-gas concentrations (Taylor et al., 2012).

For the mean shift analyses (section 3.1), ensemble members were limited to only those 29 models for which *P*, *E*, and *Q* were available from both HIST and RCP8.5 experiments. In this way, physical consistency as well as proper interpretation of Figures 2–6 in the context of the water budget is assured. Furthermore, the bcc-csm1-1 and bcc-csm1-1-m models were excluded from the mean shift analysis due to concerns arising from large negative (unphysical) *Q* in MEK for all experiments.

For the ensuing analyses (sections 3.2–3.5), all available model estimates were included, regardless of whether all fields—*P*, *E*, and *Q*—were simultaneously available. For example, EC-EARTH total runoff estimates (mrro) are not available for HIST or RCP8.5 experiments (Table S1). Therefore, EC-EARTH output is not included in the mean shift analysis (this section); however, EC-EARTH *P* and *E* estimates are used in the variability and seasonality analyses (section 3.3–4).

In section 3.5, we evaluate CMIP5 participant model skill against global observational data sets using output from the AMIP experiment spanning 1979–2008. AMIP model simulations are constrained by observed monthly SST and sea ice concentration time series, so they correspond directly to the observational record (e.g., Scaife et al., 2009). By contrast, free-running (from 1,850) HIST simulations, constrained only by a

Table 2
CMIP5 Models and Their Reference Numbers as Used in Figures 11–14

Model name	No.
ACCESS1-0	1
ACCESS1-3	2
bcc-csm1-1	3
bcc-csm1-1-m	4
BNU-ESM	5
CanESM2	6
CCSM4	7
CESM1-BGC	8
CESM1-CAM5	9
CESM1-FASTCHEM	10
CESM1-WACCM	11
CMCC-CESM	12
CMCC-CM	13
CMCC-CMS	14
CNRM-CM5	15
CSIRO-Mk3-6-0	16
EC-EARTH	17
FGOALS-g2	18
FGOALS-s2	19
FIO-ESM	20
GFDL-CM2p1	21
GFDL-CM3	22
GFDL-ESM 2G	23
GFDL-ESM 2M	24
GISS-E2-H	25
GISS-E2-R	26
HadCM3	27
HadGEM2-AO	28
HadGEM2-CC	29
HadGEM2-ES	30
inmcm4	31
IPSL-CM5A-LR	32
IPSL-CM5A-MR	33
IPSL-CM5B-LR	34
MIROC4h	35
MIROC5	36
MIROC-ESM-CHEM	37
MIROC-ESM	38
MPI-ESM-LR	39
MPI-ESM-MR	40
MPI-ESM-P	41
MRI-CGCM3	42
MRI-ESM 1	43
NorESM1-ME	44
NorESM1-M	45

Note. For full details regarding the models and available realizations per CMIP experiment, see Table S1. CMIP5 = Coupled Model Intercomparison Project Phase 5; RCP8.5 = Representative Concentration Pathway 8.5.

prescribed historical CO_2 concentration time series, constitute only one possible integration of the historical period among many in the context of internal atmospheric variability (e.g., Deser et al., 2012; Deser et al., 2016; Williams et al., 2017). A table of all available models and realizations considered in this study is available in Table S1.

2.3. Observational Comparison Data

For each of the water budget variables, we selected a single observationally based data set to establish baseline model performance during the AMIP period. It is common practice in statistical downscaling (e.g., Ho et al., 2012; Watanabe et al., 2012) to assume that historical model biases in magnitude and seasonality will remain constant in near future projections. Such an assumption is imperfect given that it presumes stationarity but nonetheless reasonable in the context of monthly large basin-scale terms and the uncertainties of accounting for nonstationarity (e.g., Hurst, 1957; Luke et al., 2017; Serinaldi & Kilsby, 2015).

For *P*, we use the Multi-Source Weighted-Ensemble Precipitation version 1 data set (MSWEPv1.0; Beck, van Dijk, et al., 2017), which is produced through an optimal merging of available gauge, satellite, and reanalysis data. Satellite and reanalysis data tend to be more heavily weighted in the tropics and mid-to-high latitudes, respectively (Beck, van Dijk, et al., 2017; their Figure 6). For example, reanalysis data are almost exclusively used in MAC and LEN. Relative weighting of gauge-based *P* is highest for the CUS, SUS, MIS, WEU, DAN, ARA, and M-D. In comparisons against other state-of-the-art global *P* products, MSWEP has been shown to be of comparatively good skill for all climate types and considering a basket of performance metrics (Beck, Vergopolan, et al., 2017).

For *E*, we use the Global Land Evaporation Amsterdam Model version 3.1a data set (GLEAMv3.1a; Martens et al., 2017; Miralles et al., 2011), which has been well validated against eddy covariance flux tower and long-term catchment water balance-inferred estimates (McCabe et al., 2016; Miralles et al., 2016). GLEAM is a simplified land surface model with four surface tiles: bare soil, low vegetation, tall vegetation, and surface water, and three soil layers (0–10, 10–100, and 100–250 cm). Intercepted canopy evaporation is estimated using the analytical model of Gash (1979) as refined by Valente et al. (1997). Remaining evaporation components—snow sublimation, open-water evaporation, bare soil evaporation, and transpiration—are solved according to the Priestley and Taylor (1972) formulation for potential evaporation (PET). For transpiration estimates, PET is scaled by a stress factor that is a function of both soil moisture and satellite-based vegetation optical depth (Liu et al., 2011; Liu et al., 2013). For soil evaporation, the stress factor is a function of soil moisture alone. Notably, satellite-based soil moisture (Dorigo et al., 2017; Liu et al., 2012) is used to update the surface layer soil moisture using a Newtonian nudging scheme (Martens et al., 2017). The version "a" data set used herein relies on European Reanalysis-Interim (Dee et al., 2011) for surface air temperature and radiation inputs and MSWEP v1.0 for *P*.

For *Q*, we use the monthly basin-scale Global Runoff Data Centre (GRDC)-based estimates from Pan et al. (2012), available for the period from 1984–2006. For basin-months lacking GRDC *Q* estimates, Pan et al. (2012) filled the record with estimates from an integration of the Variable Infiltration Capacity (VIC) land surface model with Princeton Global Forcing data (Sheffield et al., 2006). While the extent of filling by VIC *Q* was typically minimal, it was necessary to fill greater than 40% of the record for Aral, Dnieper, Don, Nile, and Volga basins (Table S2).

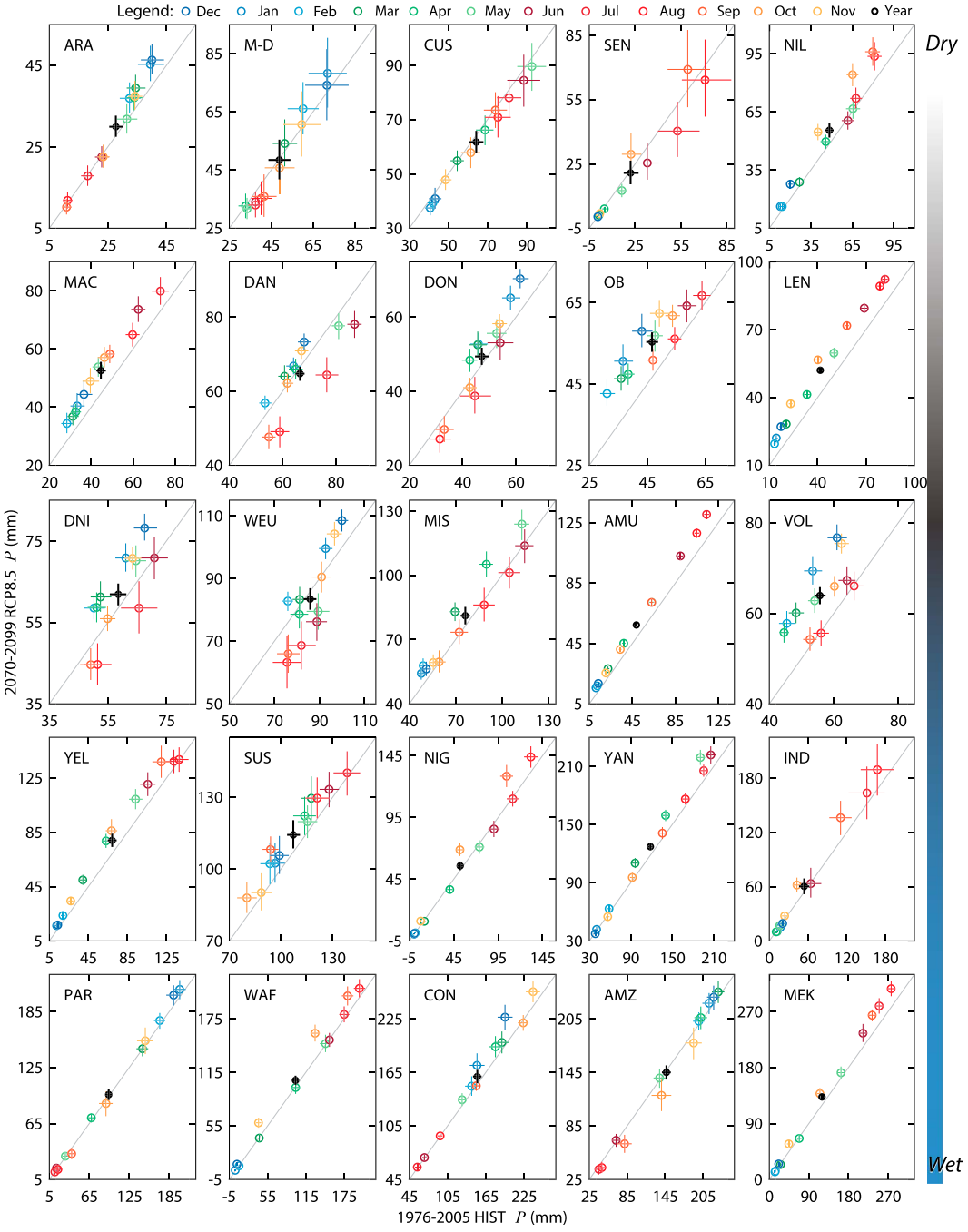

Figure 2. For each basin, scatter plots of the CMIP5 RCP8.5 ensemble mean P versus the CMIP5 HIST ensemble mean P (circles). The vertical and horizontal crosshairs denote the 95% bootstrap confidence intervals for RCP8.5 and HIST ensemble means, respectively. The annual mean (in mm/month) marker is black; other colors correspond to a particular calendar month (see figure legend). If neither crosshair intersects the 1:1 line, then the difference is statistically significant at the $\alpha = 0.05$ level. Note that the basins are arranged according to increasing P (i.e., maximum mean P for a single calendar month) from top to bottom.

Because GLEAMv3.1 E is computed with MSWEPv1.0 P, there is physical consistency enforced between the two products and we can compute $P - E$. Three-hourly MSWEPv1.0 P and daily GLEAMv3.1 E are available from 1979 to present and from 1980 to present, respectively. Both products were spatially downscaled from their native resolutions of 0.25° to 0.125° and temporally aggregated to monthly time step, consistent with the CMIP5 postprocessing (section 2.2).

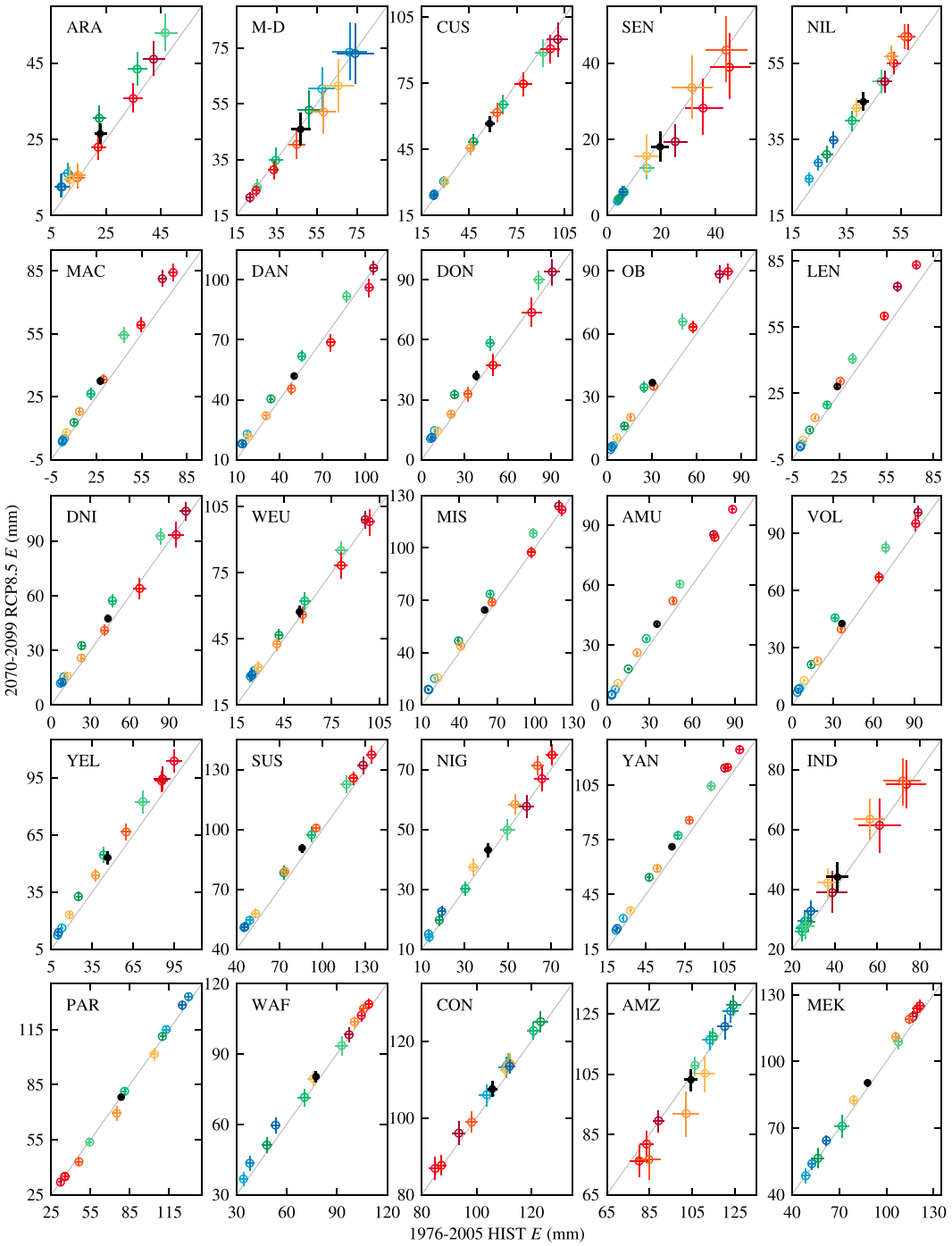

Figure 3. As in Figure 2, but comparing CMIP5 RCP8.5 ensemble mean *E* versus the CMIP5 HIST ensemble mean *E*.

2.4. Statistical Methods

2.4.1. Effective Sample Size

The basin-averaged monthly and yearly time series generated for our analysis (section 2.2) do not carry equal confidence due to large differences in basin area and regional spatiotemporal autocorrelation characteristics. Statistically, we may account for this information by computing for each variable/basin pairing an effective sample size that can in turn be used to scale the associated confidence interval (CI; e.g., Bretherton et al., 1999).

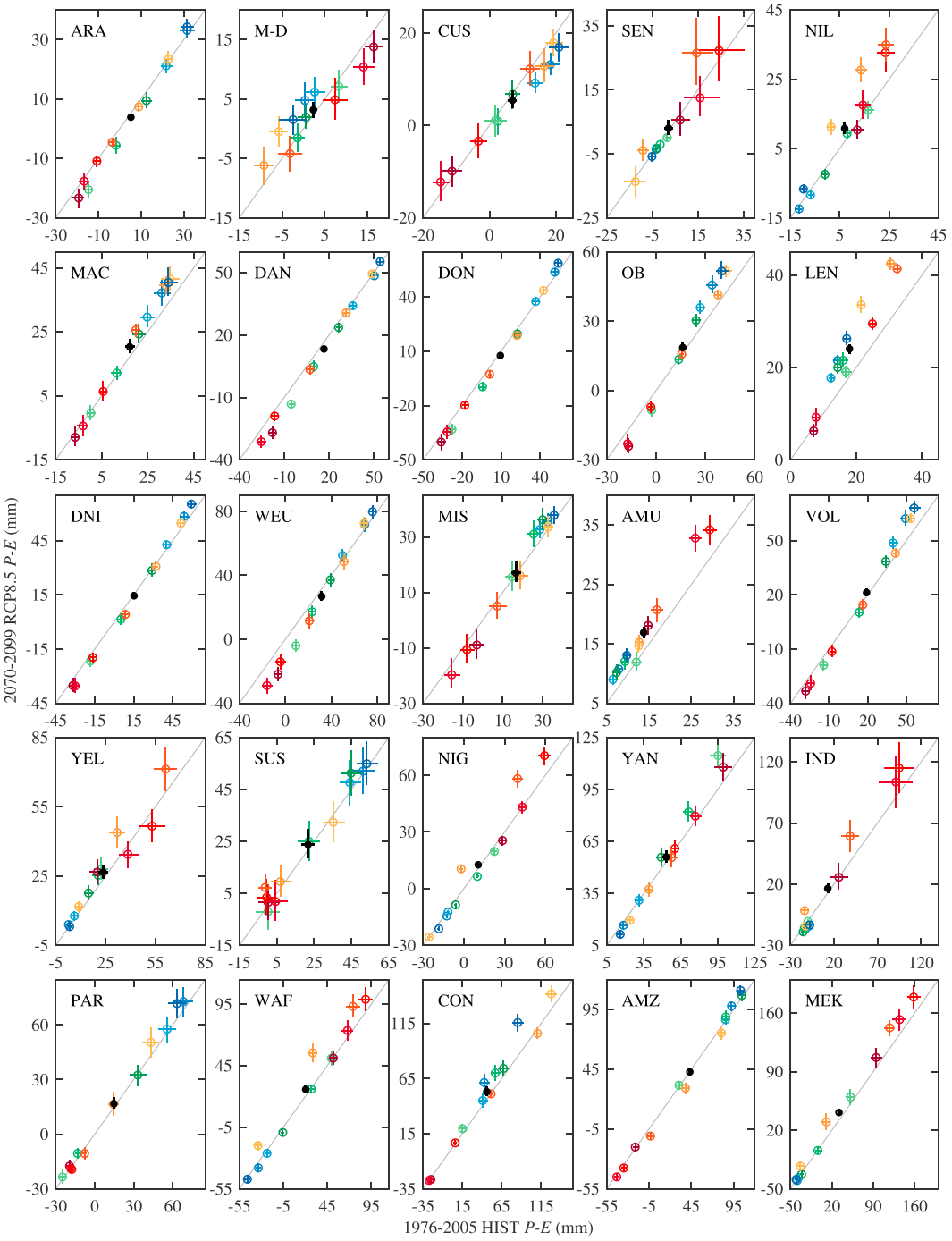

Figure 4. As in Figures 2, 3, but comparing CMIP5 RCP8.5 ensemble mean $P - E$ versus the CMIP5 HIST ensemble mean $P - E$.

In our case, we apply the following brute force approach to estimate effective sample size. First, we loop over each 0.125° grid in a given basin and compute a Pearson's r for each distance in increments of 0.375° from 0.125° to the maximum within basin grid separation distance, taking care not to double count any two grid combination. To be clear, we compute the correlation between the 30-year time series at the central grid and the series' from all other grids at the specified distance. Second, the set of Pearson's r values computed for each distance are transformed to Fisher's z values, averaged, and back-transformed to yield a single Pearson's r per distance. Third, a search for the correlation length, or

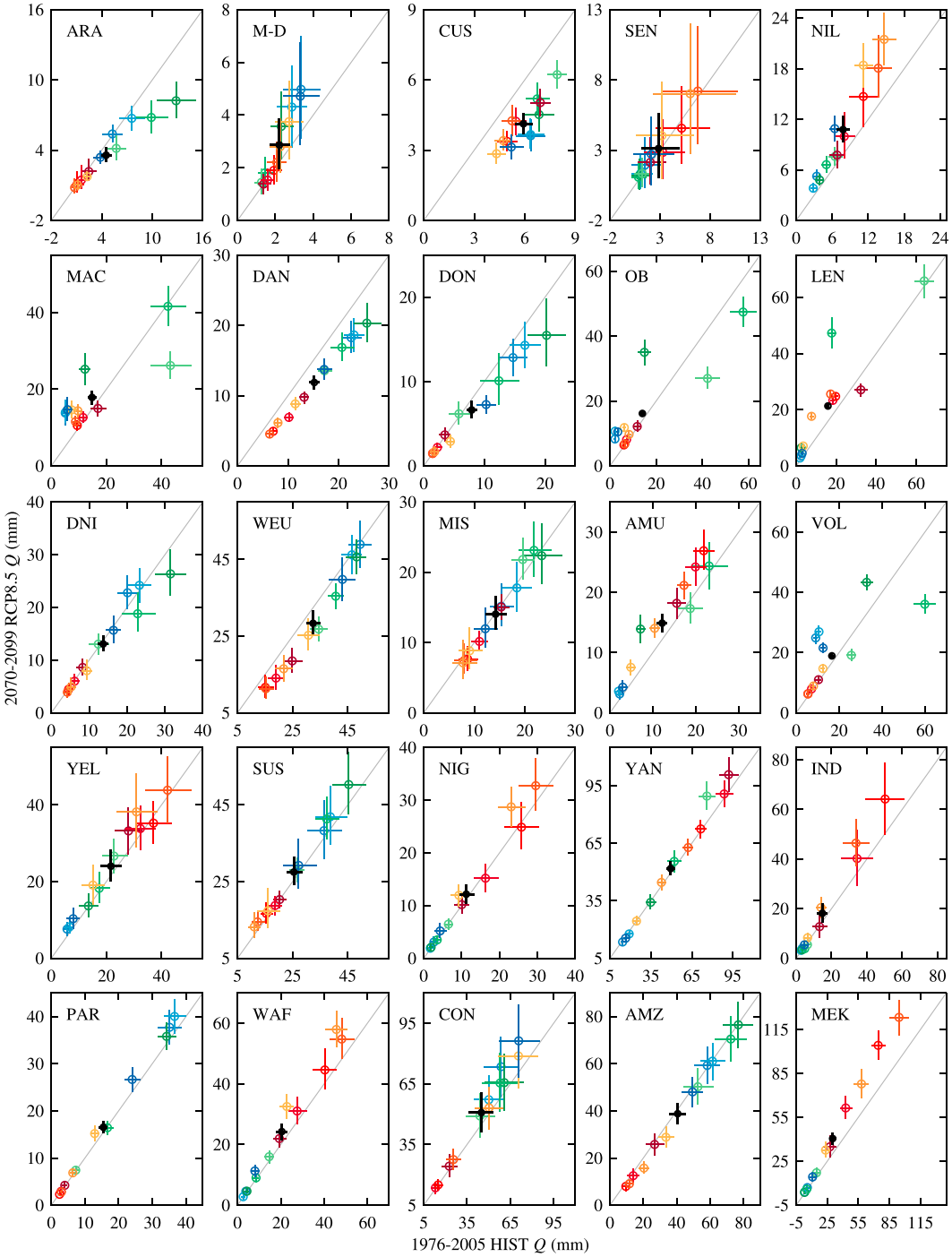

Figure 5. As in Figures 2–4, but comparing CMIP5 RCP8.5 ensemble mean *Q* versus the CMIP5 HIST ensemble mean *Q*.

minimum distance at which Pearson's *r* is less than the 95% critical value is carried out. Assuming a cylindrical equal distance projection, the effective spatial resolution may be approximated by the product of the correlation length and the square root of two. Finally, expanding radially from the basin's centroid, we determine the number of grids at the effective resolution needed to span all 0.125° grids within the basin. This grid count is the effective sample size, which we calculate for each variable/basin (Table S3).

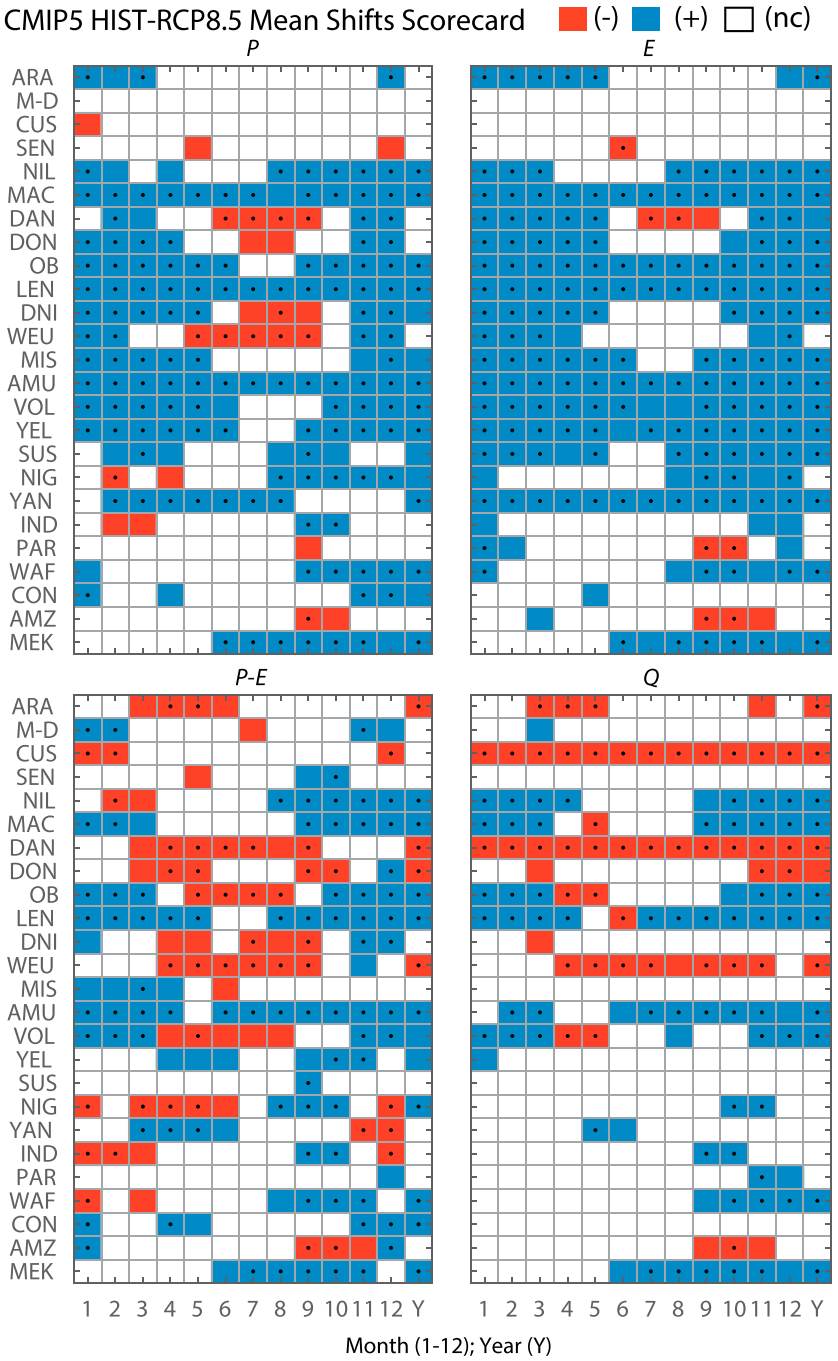

Figure 6. A scorecard summary for Figures 2–5, which highlights shifts in the projected CMIP5 RCP8.5 2070–2099 ensemble mean P, E, $P - E$, and Q that are statistically significant at the $\alpha = 0.05$ level. If a grid is neither red (negative shift) nor blue (positive shift), then no statistically significant change (nc) was detected. Black dots indicate shifts that meet the same significance level according to a Student's t test, which assumes data normality.

Effective sample size was calculated from the MIROC5 RCP8.5 0.125° monthly gridded output for 2070–2099 and in this study, we apply the minimum effective sample size computed for all months. The effective sample size is found to range from one to eight, with a mean of two (Table S3). Accordingly, a conservative one-sample assumption would have been fairly accurate. And we estimate the impact on the ensemble mean 30-year CIs would be less than 0.3 mm/year. The variance analysis that is conducted on a model basis is

relatively more sensitive to effective sample size accounting, with an estimated additional 5–40% of model shifts deemed significant.

2.4.2. Pettitt Test for Statistical Breakpoint Detection

A meaningful comparison of climate-related statistics from two time periods presupposes that (1) the climate during both periods is stationary, or free from significant changes in the mean (incl. trends) or variability and that (2) the time periods are long enough to include natural low-frequency variability. A 30-year average is generally accepted to capture sufficient natural variability and coincidentally, corresponds with the World Meteorological Organization standard climate normal (e.g., Arguez & Vose, 2011). Whether stationarity over a 30-year period during the present Anthropocene is a valid assumption has been brought into question (e.g., Milly et al., 2008).

Using the Pettitt (1979) breakpoint test, we test for statistically significant shifts in the ensemble mean over this study's 30-year averaging periods (1976–2005 and 2070–2099). The Pettitt test is designed to detect a single abrupt change in the mean of the distribution of the variable of interest at an unspecified point in time (for a review, see Mallakpour & Villarini, 2016). It has been widely used with hydroclimatological data (e.g., Ferguson & Villarini, 2012, 2014) and is most sensitive to breaks that occur in the middle of the series. Because the presence of a monotonic trend could lead to false detection of a breakpoint by the Pettitt test (Ferguson & Mocko, 2017), we jointly test for trends using the nonparametric rank-based Mann-Kendall test (Kendall, 1975; Mann, 1945). A nonparametric approach is necessary because less than 20% of all model P, E, $P - E$, and Q basin-averaged monthly subseries and annual time series conform to a normal distribution, according to a Lilliefors (1967) test conducted at the 5% significance level. The impact of serial correlation on the Mann-Kendall test was mitigated through trend free pre-whitening of the original time series following Yue et al. (2002), as also detailed in Ferguson and Mocko (2017).

2.4.3. Student's *t* Test for Significantly Different Means

The 95% CI associated with the multimodel ensemble means were calculated from the Student's t distribution (e.g., Gillett & Fyfe, 2013; their Figure 2). To find the margin of error, the sample standard deviation was first divided by the square root of the number of models in the ensemble scaled by the effective sample size and then multiplied by the 5% cutoff value for a two-sided Student's t distribution with degrees of freedom equal to one less than the number of models scaled by the effective sample size. By adding and subtracting the margin of error from the ensemble mean, 2.5% and 97.5% confidence limits were found. When the 95% CIs of two ensemble means do not overlap, their difference is deemed statistically significant.

As it turns out, in 28%, 40%, 55%, and 32% of P, E, $P - E$, and Q cases, respectively, the distribution of 30-year HIST and RCP8.5 model basin averages violates the t tests' normality assumption, according to a Lilliefors (1967) test conducted at the 5% significance level. Furthermore, in these nonnormal cases, the central limit theorem does not appear to hold for the limited sample size involved. Consequently, the t test results in overly conservative estimates of the 95% CIs and thus, fewer significant shifts. Our remedy is to quantify significance using 95% bootstrap CIs, following Pryor and Schoof (2008). Specifically, the distribution of 30-year model means are resampled with replacement to generate 10,000 realizations the size of the model count scaled by the effective sample size. The sample means are computed and the 500th and 9,500th ranked values constitute the 95% bootstrap CI. Given the prevalence of the t test in CMIP5 analyses, we elect to include t test statistics where feasible.

2.4.4. Bootstrap Test for Significantly Different Variances

Due to the aforementioned nonnormality of the time series (Section 2.4.2), we apply a bootstrap approach to evaluate differences in the variance between the HIST (1976–2005) and RCP8.5 (2070–2099) periods. The test is conducted on a model-specific basis because the results from a comparison of two groups of variances (i.e., HIST and RCP8.5) can be difficult to interpret. Specifically, basin-averaged monthly subseries and annual time series are resampled with replacement to generate 10,000 realizations the size of the record length ($\underline{n} = 30$) scaled by the effective sample size. Then, the sample variances are computed and the 500th and 9,500th ranked values define the 95% CI for the given model/variable/basin/period. If the HIST and RCP8.5 95% bootstrap CIs do not overlap, then the difference in variance between the two periods is deemed significant.

Note that for several models, multiple realizations are available. In such cases, exact pooled variances were calculated from the bootstrapped 5th and 95th percentile variances. The exact pooled variance is calculated by summing the mean of the variances and the variance of the means (of the component realizations). As detailed in Rudmin (2010), taking the mean of the realization variances as an estimate of the exact pooled

variance is ill-advised for many reasons, foremost among those being it will always lead to underestimation of the exact pooled variance.

2.4.5. Seasonality Index

In some cases, information on the precise timing (i.e., calendar month) of shifts in the P seasonal cycle may not be needed. Shifts in the relative seasonality, or variability in monthly P throughout the year, may be quantified in terms of the Walsh and Lawler (1981) seasonality index (SI):

$$SI_i = \frac{1}{Ri}\Sigma_{n=1}^{12}\left|X_{in} - \frac{R_i}{12}\right| \quad (1)$$

where X_{in} is the P for month n of year i, R_i is the annual P accumulation in year i, and SI_i is SI for year i. Similarly, the index may be applied to E, $P - E$, and Q records. To account for negative monthly $P - E$ values in the SI calculation (equation (1)), we replaced R_i with the sum of absolute monthly $P - E$ values in year i.

We calculated SI_i for each year of the 30-year analysis period (e.g., HIST: 1976–2005) and computed period mean values to test shift significance between periods. When more than one realization was available for a given model (see Table S1) we computed SI for all realizations and took the median. The SI can vary from zero (with uniform monthly rainfall) to 1.83 (if all the rainfall arrives in a single month), and its expected distribution is poorly defined. Thus, significance of differences in SI was quantified using bootstrap CIs, consistent with the approach followed for means and variances (sections 2.4.3–4). Time series of the SI_i values are resampled with replacement to generate 10,000 realizations of the mean SI value during the 30-year period. SI's from two analysis periods (e.g., 1976–2005 and 2070–2099) that lie within the middle 9,500 values in an ordered sequence of the distribution of 10,000 realizations are deemed to be insignificantly different at the 95% confidence level.

3. Results

3.1. Stationarity and Water Conservation

We analyzed the basin annual mean time series of P, E, and Q for all available models and their realizations (Table S1) over the selected 30-year HIST (1976–2005) and RCP8.5 (2070–2099) time periods. The results from both periods/experiments were fairly consistent. Most breakpoints were detected in the E time series. Of all the basin series analyzed, 12–13% had significant breakpoints in E, compared to 5–7% in Q, and 4–5% in P (Figures S1 and S2). Greater than 98.7% of these breakpoints (considering all variables and both periods) were found to be associated with statistically significant trends. Multimodel agreement on historical E trend significance exceeds 30% on positive trends over VOL, WEU, OB, and DAN (Figure S1). Similarly, more than 30% of models project significant positive end-of-century E trends over CON, AMU, LEN, OB, and MAC (Figure S2). The only basin/period for which simultaneous trends in P and Q are simulated by a substantial fraction of models is LEN during the RCP8.5 period—both trends are positive (Figure S2). Among models, significant trends were more (e.g., BNU-ESM, MIROC-ESM-CHEM, MRI-CGCM3, and MRI-ESM 1) or less widespread (e.g., CESM1-WACCM, GFDL-CM3, and MPI-ESM-MR) and could vary substantially between HIST and RCP8.5 periods (Table S3). The fact that relatively more breakpoints/trends were detected in E is more likely a function of the detectability limits of the Pettitt test (Mallakpour & Villarini, 2016; their Figure 2) than a true physical climate signal. The median coefficient of variation (CV) is 0.06 for E, 0.11 for P, and 0.27 for Q, considering all basins and model output from both HIST and RCP8.5 experiments.

Given the modest level of nonstationarity found, we hold that the simplifying assumption of stationarity over the 30-year comparison periods is acceptable. Moreover, we note that significant trends were generally found to be small in magnitude (not shown) and thus, in most cases would not have a substantial impact on the 30-year mean. By leaving the trends in-tact, we are in effect calculating a more conservative estimate of the significance of any mean climate shift (i.e., detections of significance are more likely).

Because this study is conducted in the context of the water budget it is pertinent to examine the models' mass conservation or numerical stability. Most land surface schemes enforce a free drainage lower boundary condition, but this water is transferred to rivers rather than to long-term storage reservoirs such as aquifers. Accordingly, the models do not explicitly represent changes in aquifer storage or aboveground glacial mass.

We found that for most basins and models, the amount of nonconservation ($P - E - Q$) is negligible, or on the order of 1% of the annual P. However, there are a few exceptions such as GISS-E2-H, GISS-E2-R, CMCC-CESM, and IPSL-CM5B-LR that do generate larger than normal terrestrial water storage gains/losses across several basins (Figures S3–S5). Further analysis of model-specific mass conservation (or terrestrial storage loss) is beyond the scope of this study. An in-depth analysis would need to be carried out on the model native grids to avoid any unphysical spatial interpolation artifacts.

3.2. Projected Shifts in Mean Water Availability

The scatter plots of Figures 2–5 illustrate the difference between HIST 1976–2005 and RCP8.5 2070–2099 ensemble mean P, E, $P - E$, and Q at monthly (colors) and annual (black) timescales. Markers lying above (below) the 1:1 line indicate projected increased (decreased) water availability. The vertical and horizontal crosshairs mark the 95% bootstrap CIs for the RCP8.5 and HIST ensemble means, respectively. Markers and associated CIs that do not intersect the 1:1 line are statistically significant at the $\alpha = 0.05$ level. The basins are arranged in order of increasing P (i.e., maximum mean P for a single calendar month) from top to bottom. It is possible to summarize the significance and sign of projected shifts using a simple scorecard (Figure 6) without magnitude information. In general, RCP8.5 CIs (ensemble spread) are larger than HIST CIs and CIs are larger in both experiments for the warm season (Figures 2–5).

Projected shifts in annual mean P, E, $P - E$, and Q are statistically significant for 15 (P), 16 (E), 15 ($P - E$), and 13 (Q) basins, respectively. Of these shifts, all those in P and E are in the direction of wetting. Although, eight basins are projected to experience significant declines in P at subannual timescales. For DAN, PAR, and AMZ these projected declines in P coincide with projected declines in E. The combined analysis of $P - E$ and Q corroborate significant wetting at annual timescale in eight basins (NIL, MAC, OB, LEN, AMU, VOL, WAF, and MEK) and drying in four basins (ARA, DAN, DON, and WEU). Independently, the $P - E$ analysis yields significant wetting at annual mean timescale for YEL, NIG, and CON, whereas the Q analysis indicates future drying in CUS. Overall, the bulk of the statistically significant shifts occur in the midrange of basin wetness (i.e., from MAC to VOL). Fifteen percent of significant P, E, and Q shifts and one quarter of all significant $P - E$ shifts are insignificant according to the parametric t test (Figure 6).

Importantly, Figures 2–6 enable attribution of significant projected changes in annual mean water availability to the signals seasonal source. For example, annual mean increases in P for YAN can be traced to increased P during the February–August period. Figures 2–6 also serve to highlight several basins for which significant subseasonal shifts in P, E, $P - E$, and/or Q are projected but the projected annual mean change is statistically insignificant (at $\alpha = 0.05$). For some basins the subseasonal response reinforces underlying seasonality, as manifested in the "S" pattern formed by the monthly markers in the case of WEU P (Figure 2). The cold season (November–February) is projected to receive greater P, whereas the warm season (May–September) is projected to receive significantly less P. A similar pattern emerges in Figure 2 for DAN and DNI—other European basins, as well as DON in western Russia.

While the mean annual shift results for $P - E$ and Q are more or less consistent at annual timescale as expected, the results do differ substantially at the monthly timescale (Figures 4–6). For example, Q is projected to decline significantly in CUS for all months but only during December–February for $P - E$. For MIS, significant increases in $P - E$ are projected for January–April, a period when Q is projected to remain unchanged. For YEL, significant increases in $P - E$ are projected for April–June and September–November, but only for January in terms of Q. And for IND, decreases in $P - E$ are projected for December–March for which Q is projected to vary insignificantly.

Recall that for this comparison, the ensemble was composed only of models for which P, E, and Q were simultaneously available (section 2.1). To test the sensitivity of our results, we performed a parallel set of analyses using all available models/realizations for each variable, thus allowing for ensembles of different sizes for each variable. As may have been expected, most differences were limited to the $P - E$ field because mass was no longer being conserved. In summary, fewer basins are projected to undergo significant shifts in $P - E$. With respect to the other terms, the following contrasts with the conserved ensemble analysis are noteworthy: significant declines in June–September P and August–October E for M-D, an enhanced seasonal drying signal in P for SEN (December–January and April–July) and NIG (January–June), and significant annual mean shifts in P for ARA (+), DON (+), and DAN (−; Figure S6).

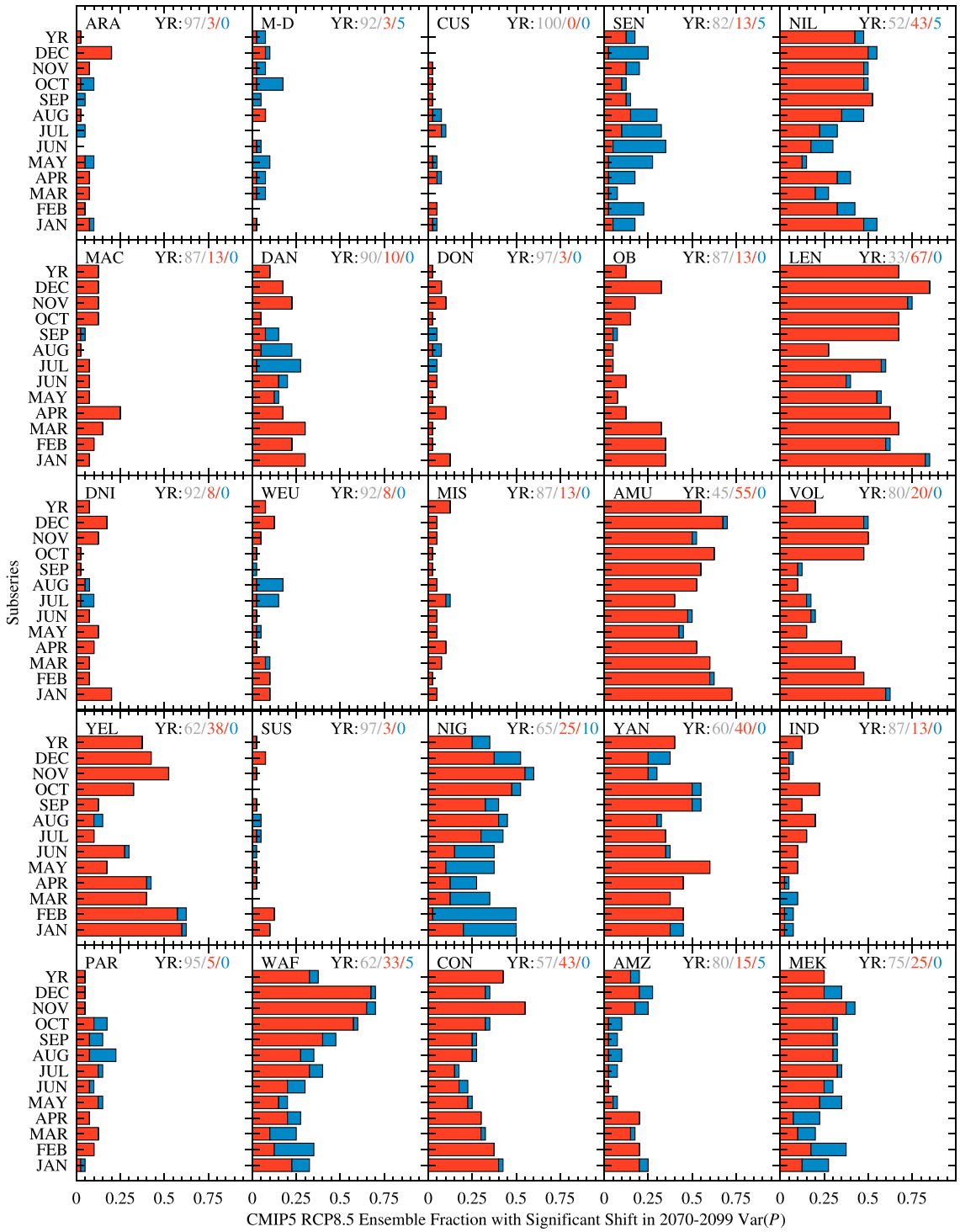

Figure 7. Summary of Figure S22; the fraction of CMIP5 models for which statistically significant shifts (α = 0.05) are detected in the RCP8.5 2070–2099 variance of *P* relative to the HIST 1976–2005 variance of *P* for each month and for the annual mean. Red and blue shading denotes the ensemble fraction for which RCP8.5 projected variance is significantly greater or less than HIST variance, respectively. The percent breakdown of ensemble response (no change/increased variance/decreased variance) for interannual variance (topmost bar) is annotated on the top of each regional panel. Forty models comprise the *P* ensemble (see Table S1).

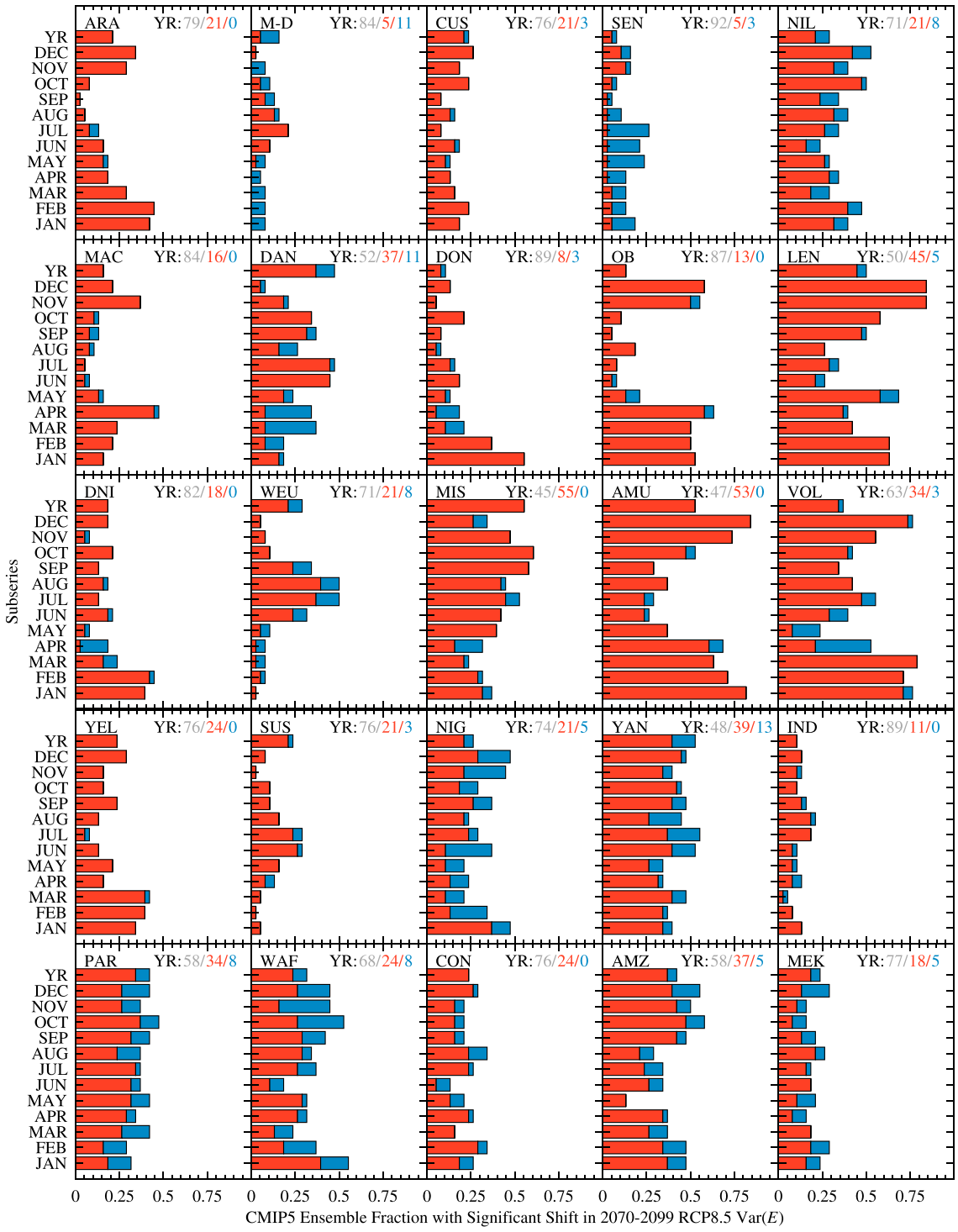

Figure 8. Summary of Figure S23; as in Figure 7 but for *E*. Thirty-eight models comprise the *E* ensemble (see Table S1).

Results from the ESMHIST versus ESMRCP8.5 analogue scorecard were similar but with 23% fewer significant shifts overall and 30% fewer basin/months projected to experience increased water availability. In fact, the ESM results include deeper and significant declines in the annual mean *P* for DAN and WEU and $P - E$ for CUS (Figures S7–S11). Towards attribution of these projection differences, we also compiled HIST-ESMHIST and RCP8.5-ESMRCP8.5 scatter plots and scorecards (Figures S12–S21). All statistically

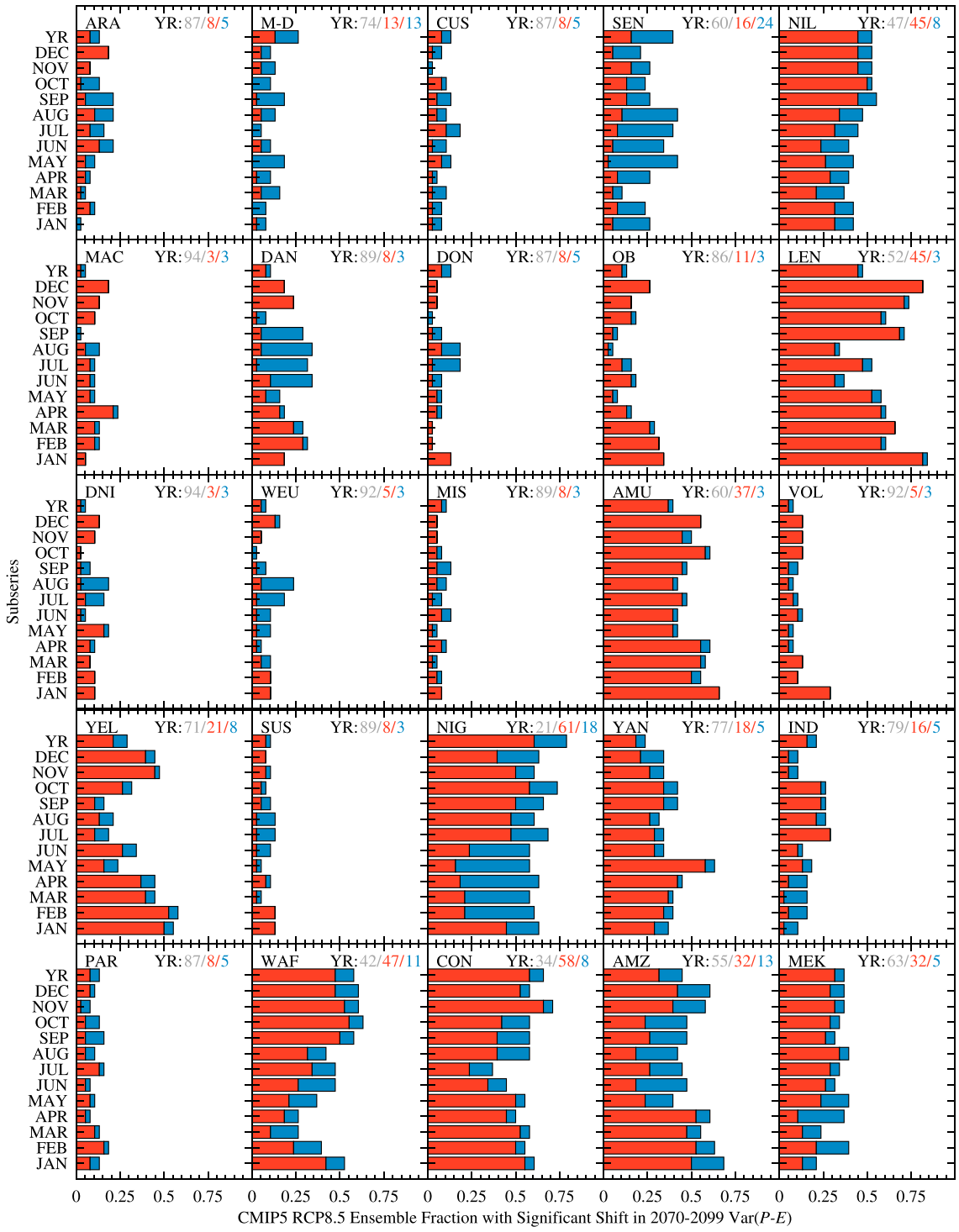

Figure 9. Summary of Figure S24; as in Figures 7, 8 but for $P - E$. Thirty-eight models comprise the $P - E$ ensemble (see Table S1).

significant annual mean differences are detailed in Table S5. In summary, ESMHIST and ESMRCP8.5 generate greater water availability in LEN—by all measures, greater Q in VOL and MEK, and less $P - E$ in SUS. Differences in SUS $P - E$ constitute the largest GCM-ESM percent differences for both periods (HIST: 41%; RCP8.5: 24%).

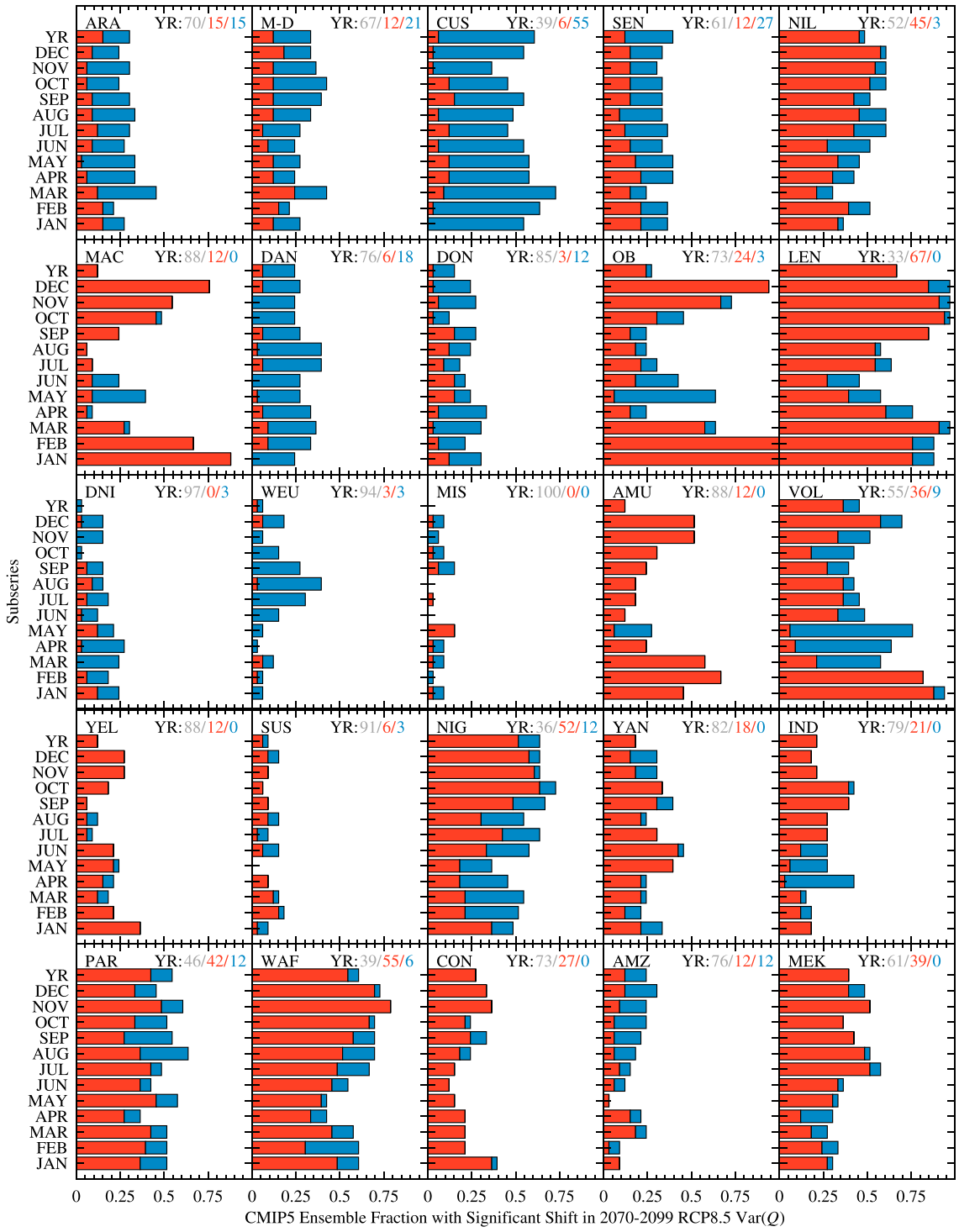

Figure 10. Summary of Figure S25; as in Figures 7–9 but for *Q*. Thirty-four models comprise the *Q* ensemble (see Table S1).

3.3. Projected Shifts in the Variability of Water Availability

Figure 7 illustrates the fraction of CMIP5 RCP8.5 models that project significantly more (red), significantly less (blue), or statistically equivalent (empty bar) variability in *P* during the 2070–2099 period for each monthly subseries and for the annual mean relative to the 1976–2005 HIST baseline. The ensemble's tendencies at

annual mean scale are annotated in terms of percentages at the top of each subpanel. Figures 8–10 present results in similar format for E, $P - E$, and Q.

For the majority of basins and variables most models project insignificant changes in future variability. In fact, a projection of no significant change in interannual variability is the most common model response in 89 (89%) of basin/variable comparisons. Of the remaining 11 cases, variability is projected to increase in all but one case. The model majority points to increased variability for: P in LEN and AMU; E in MIS and AMU; $P - E$ in NIG, WAF, and CON; and Q in LEN, NIG, and WAF. On the other hand, a significant reduction in variability is projected for Q in CUS (Figures 7–10).

A similar summary may be given for shifts in monthly subseries variability. In cases where models project significant change in future variability, it is overwhelmingly in the direction of increased variability (i.e., red bars). Interestingly, signs of seasonality do emerge for select basins/variables, as in the mean shift analysis. For example, a larger fraction of models project significant increases in E variability during the OB, LEN, and AMU cold season (November–April) than in other months of the year (Figure 8). Seasonal shifts in variability like these in E, as well as those in Q for MAC, OB, AMU, and VOL, tend to go undetected at annual timescale. Nonetheless, they constitute important information. Declining warm season variability in $P - E$ as projected by some models for DAN and WEU, combined with the findings our mean shift analysis (section 3.2), points to consistently drier warm season conditions for these basins (Figures 7–10).

Overall, a larger fraction of the ensemble projected significant shifts in the future variability of Q than for P, E, and $P - E$ and the number of models projecting changes in E is greater than for P or $P - E$. Indeed, the results for $P - E$ (Figure 9) appear to better match the response obtained for P (Figure 7) than for E (Figure 8). Notably, there are basins for which most models agree there will be little change in future variability—at neither monthly nor annual timescales. For P, these basins include ARA, M-D, CUS, DAN, DON, DNI, WEU, SUS, and PAR (Figures 7–10).

3.4. Projected Shifts in the Relative Seasonality of Water Availability

An alternative to evaluating projected calendar month changes in water availability is to use a SI (Walsh & Lawler, 1981). Figure 11 illustrates the RCP8.5 2070–2099 P SI for all basins and models. If a model's projected SI is significantly greater or less than its HIST 1976–2005 baseline SI, then the SI is colored accordingly (red: increase; blue: decrease), and a line trace to its baseline SI is plotted. For each basin (subpanel of Figure 11), the most popular model future SI projection (increase, decrease, or no change) is annotated in terms of its ensemble percentage. For example, the most common projection is increased P SI for 14 basins and for the other basins the popular model response is split between diminished P SI (six basins) and statistically equal P SI (five basins; Figure 11). The largest shifts in P SI by magnitude can be seen for MAC (decrease), DAN (increase), DON (increase), LEN (decrease), DNI (increase), WEU (increase), and AMU (decrease). Consistent with the signature S pattern in projected shifts in monthly mean P for WEU, 88% of the models show a significant increase in P SI for that basin. Similar plots for the projected SI of other water budget variables—E, $P - E$, and Q—are provided in Figures 12–14.

The results indicate declines in E SI over most basins, except M-D, PAR, AMZ, and MEK. In terms of $P - E$ SI, the dominant model signal is no change (13 basins), with increased SI in nine basins and decreased SI in three basins. Lastly, in terms of Q SI, increased SI is projected for a majority of basins ($n = 14$), decreased SI for nine basins, and statistically equal SI for two basins.

It is interesting to note that the results for $P - E$ SI and Q SI highlight basins for which SI shifts are of opposite sign. For $P - E$ SI, the largest shifts by magnitude occur for WEU, YAN, and AMZ for which SI is projected to increase substantially. For Q SI, the largest shifts by magnitude occur for MAC, OB, and VOL for which substantial decreases in SI are projected. Our takeaway from these two findings is that changing high latitude cold/shoulder season processes (incl. terrestrial water storage changes) are behind large shifts in Q SI whereas dry season shifts in P and/or constraints on E drive future $P - E$ SI shifts.

3.5. AMIP Model Performance

Figure 15 illustrates the CMIP5 AMIP 1979–2008 multimodel mean seasonal P and E cycles in contrast to MSWEPv1.0 P and GLEAMv3.1 E. In Figure 15, gray- and blue-filled areas denote periods when AMIP P is less than or greater than MSWEPv1.0 P, respectively. Similar to Figures 2–5, Figure 15 is useful because it illustrates

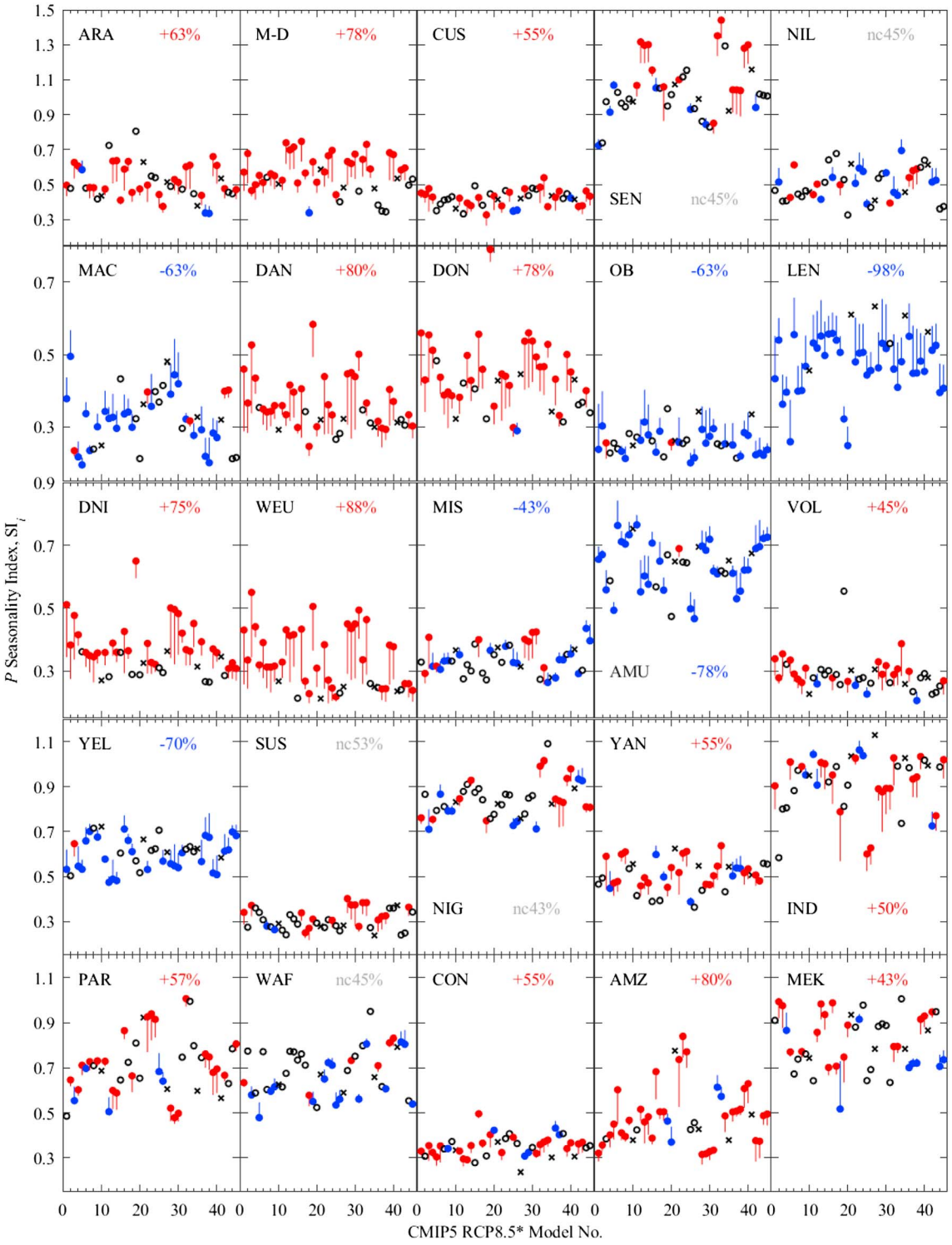

Figure 11. Shifts in the *P* seasonality index (SI) between CMIP5 RCP8.5 and HIST for all available models (see Table 2 for model number assignments). The filled circle marks the RCP8.5 SI and the stem traces back to the HIST SI. For models with more than one realization available, the median SI was used. Statistical significance at the $\alpha = 0.05$ level is indicated by red (projected increase) and blue (projected decrease). Insignificant shifts are indicated with empty circles and x's denote models for which the RCP8.5 output was unavailable. At the top of each panel, the dominant ensemble response (%) and its sign—positive (red), negative (blue), or neutral (gray)—is annotated.

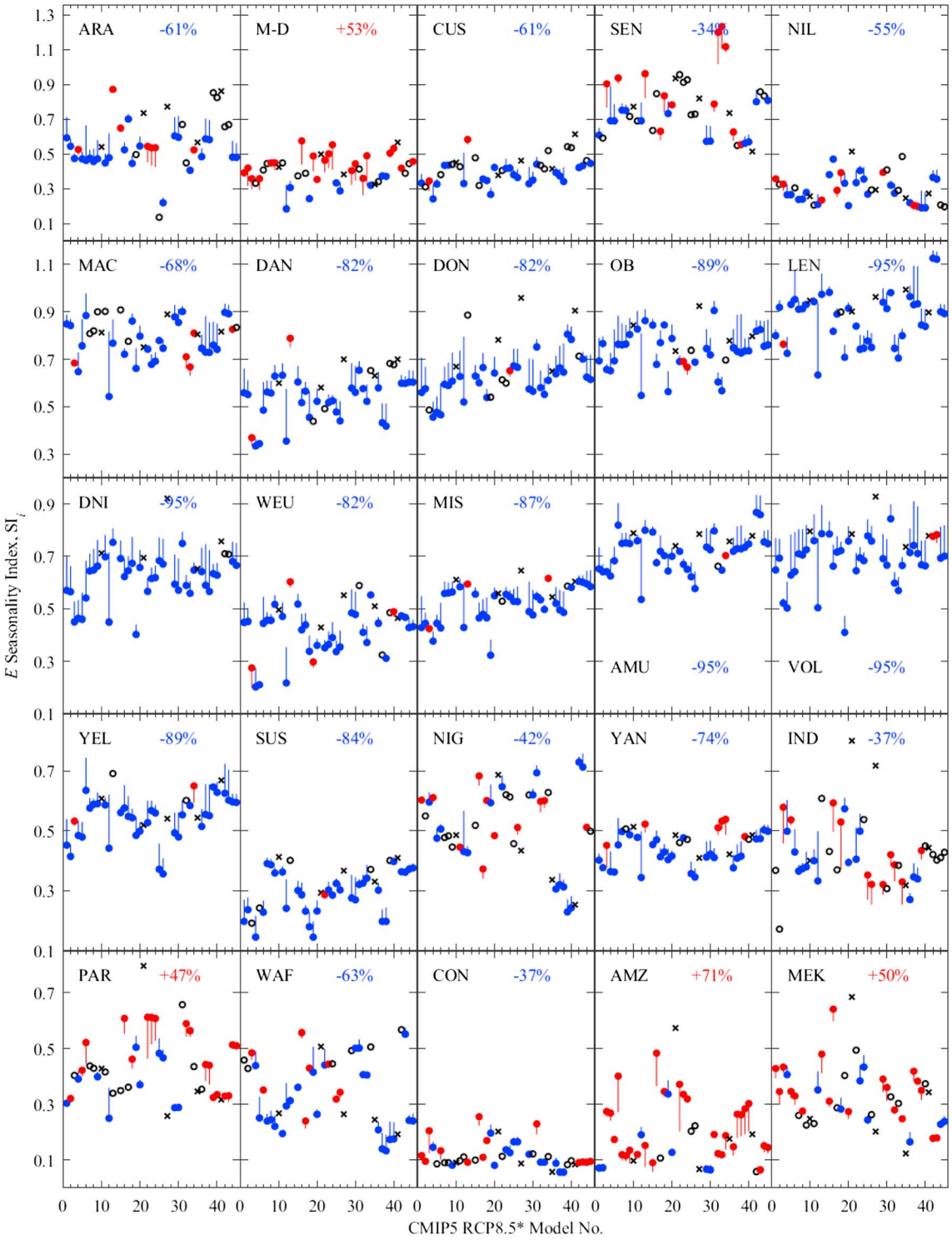

Figure 12. As in Figure 11, but for shifts in the *E* SI.

the magnitude of the differences, as well as the 95% bootstrap CI of the AMIP ensemble. For example, the CI is remarkably wide for June–October *P* in IND and MEK.

Overall, the AMIP ensemble significantly underestimates warm season *P* for CUS, DAN, DON, OB, DNI, WEU, MIS, and VOL and significantly overestimates cold season *P* for many of these same basins (Figure 15).

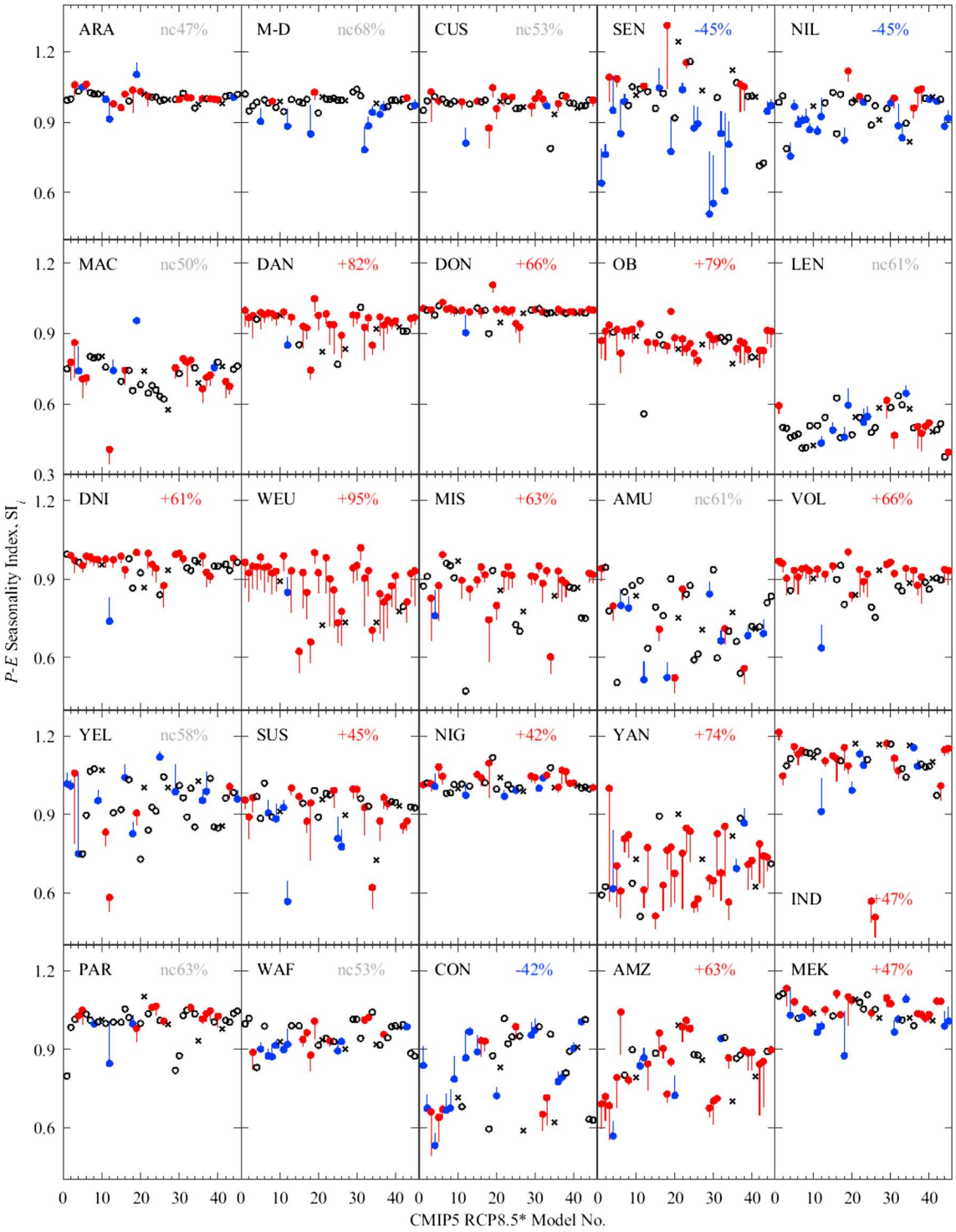

Figure 13. As in Figures 11, 12, but for shifts in the $P - E$ SI.

When the information in Figure 15 is reduced to a scorecard summary, this seasonal tendency stands out (Figure 16). In other words, the AMIP *P* SI is too large. In terms of *E*, the AMIP ensemble is biased high for nearly all basin months, with the exception of a few months for CUS, MAC, DON, OB, LEN, DNI, MIS, AMU, VOL, SUS, PAR, and AMZ.

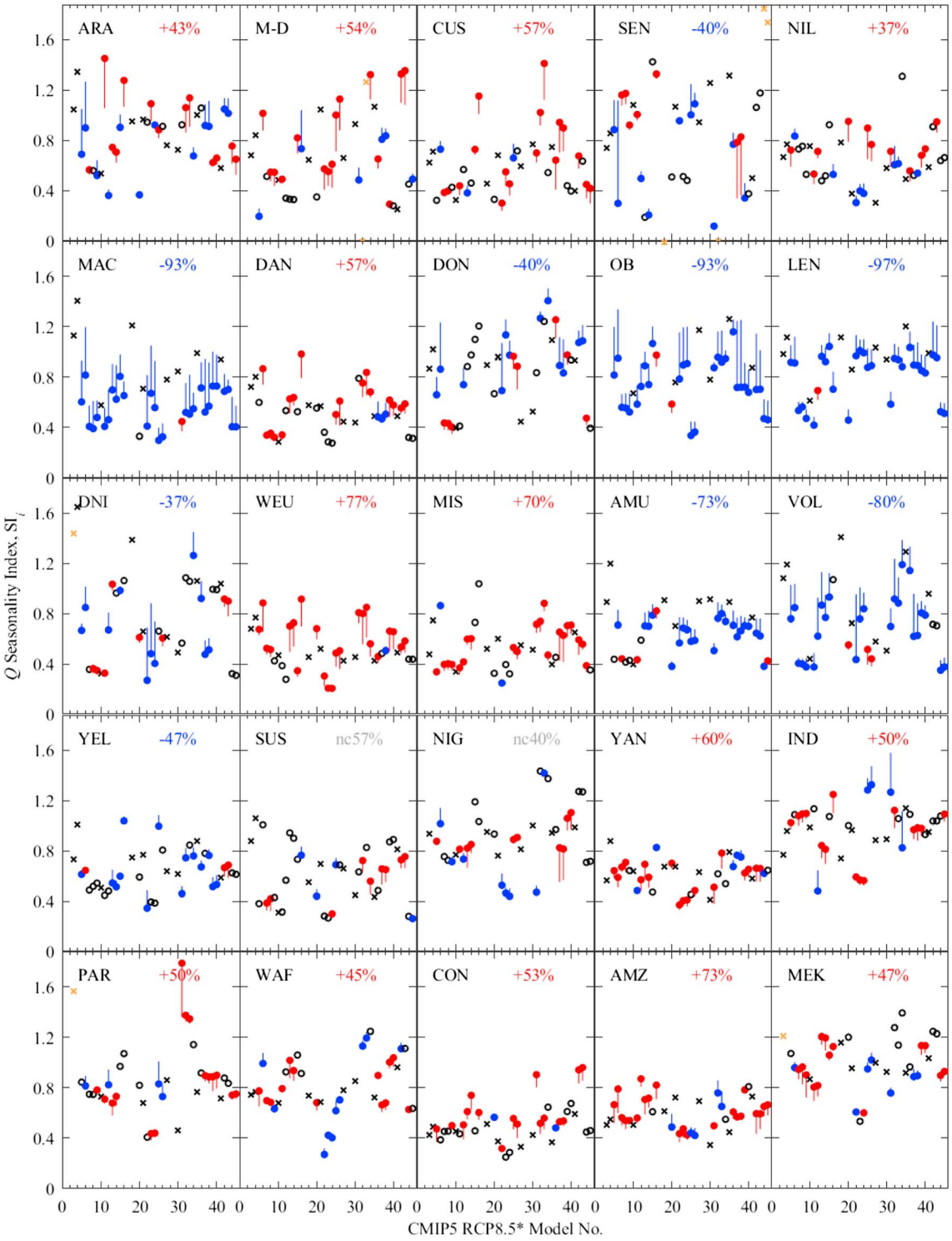

Figure 14. As in Figures 11–13, but for shifts in the *Q* SI.

The scorecard summary in Figure 16 also includes AMIP comparison results for $P - E$, calculated from MSWEPv1.0 *P* and GLEAMv3.1 *E*, and *Q* from the GRDC-based estimates of Pan et al. (2012). The AMIP ensemble $P - E$ is generally biased low with the exception of some cold season months in mid-to-high latitude basins, in which case it is biased high. For the most part, biases in AMIP *Q* are consistent with those for $P - E$ (Figure 16).

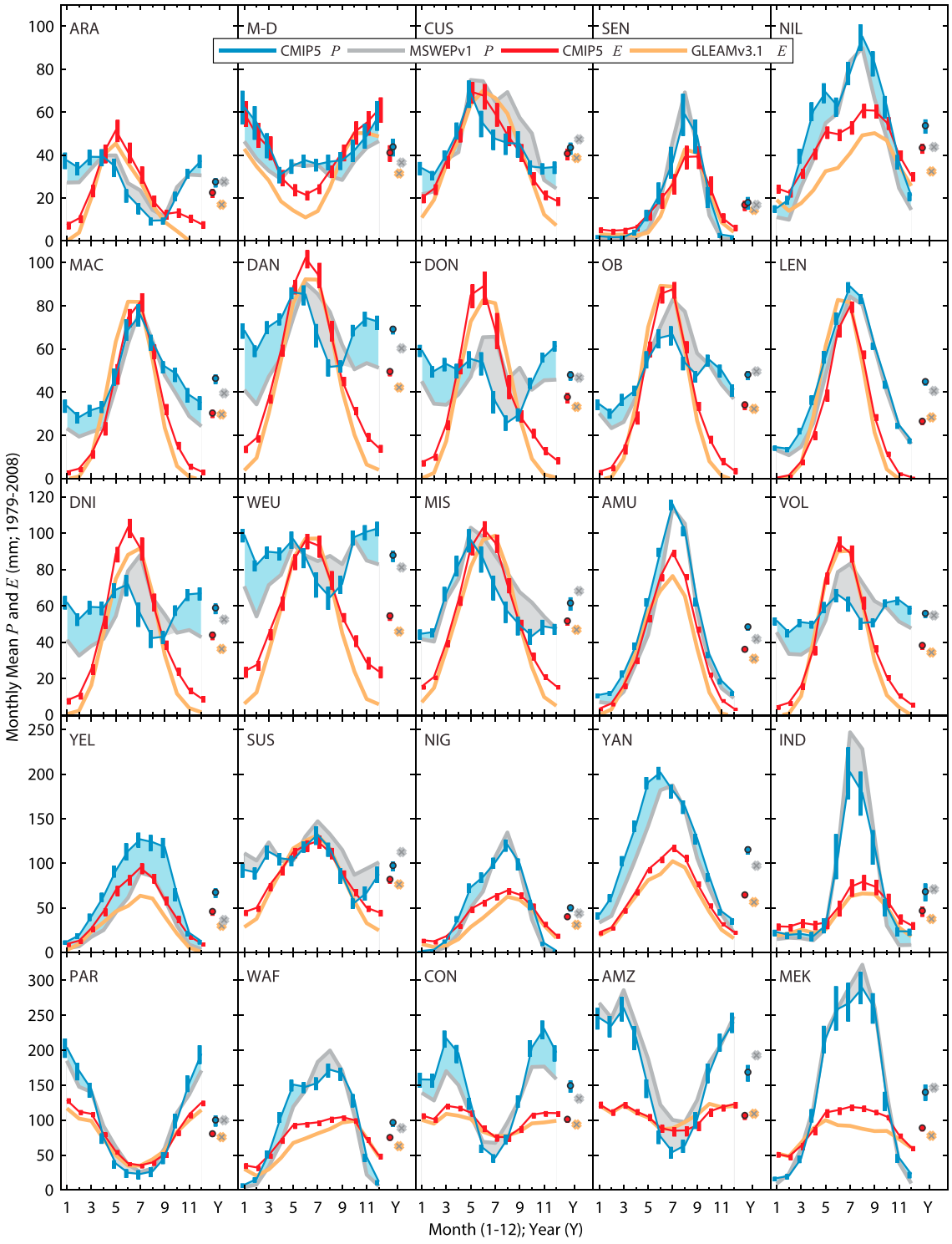

Figure 15. Regional CMIP5 AMIP 1979–2008 monthly (lines) and annual (filled circles) mean *P* and *E* relative to 1979–2008 MSWEPv1.0 *P* and 1980–2009 GLEAMv3.1 *E*. Vertical bars constitute the 95% bootstrap confidence intervals on the CMIP5 AMIP ensemble mean. Light blue and gray fill areas highlight months for which the CMIP5 AMIP ensemble mean is higher or lower than the observational counterpart, respectively. Twenty-nine models comprise both the CMIP5 AMIP *P* and *E* ensembles (see Table S1).

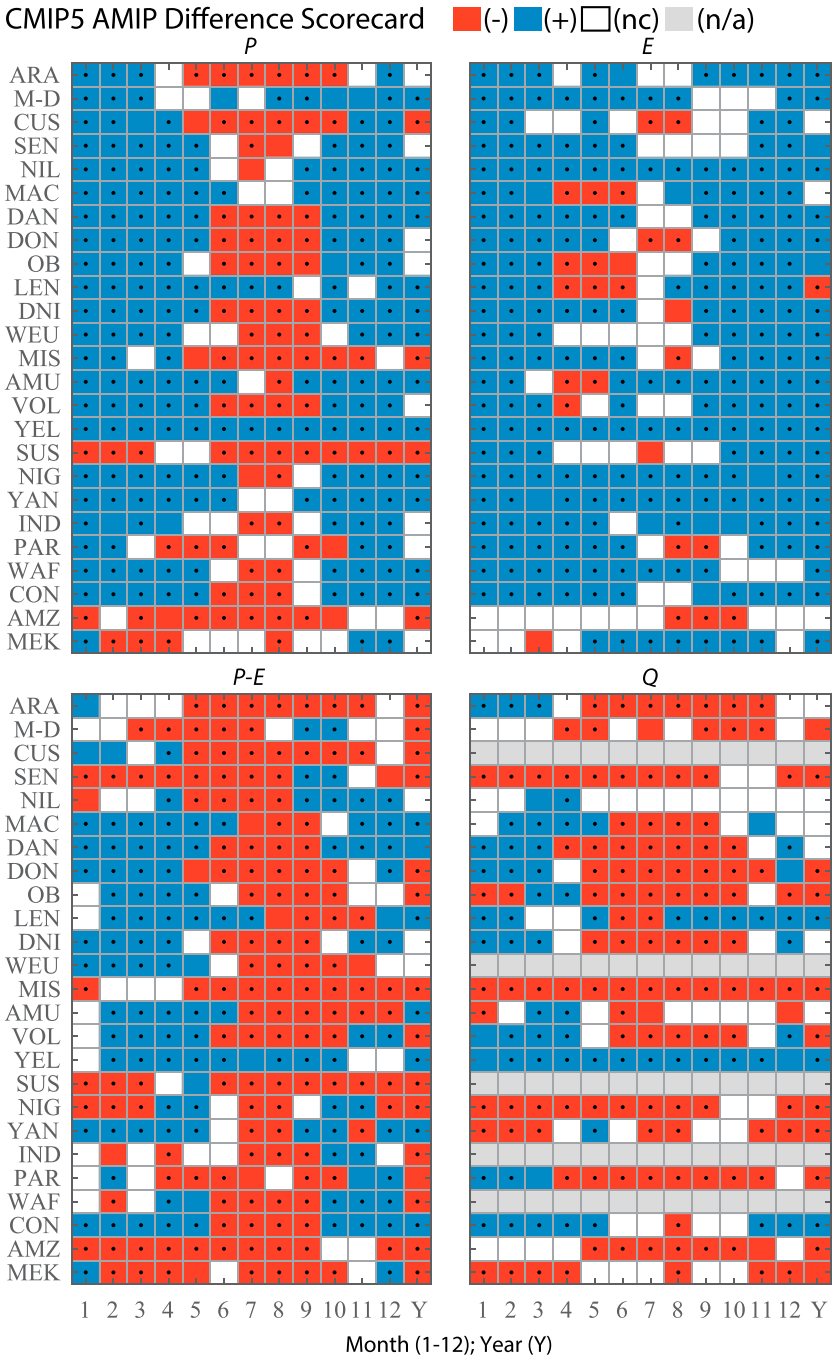

Figure 16. As in Figure 6, a scorecard summary for CMIP5 AMIP (1979–2008) ensemble mean monthly and mean annual differences from observational estimates. MSWEPv1.0 and GLEAMv3.1 were used for *P* and *E*, respectively. For *Q*, mean AMIP values for the period from 1984–2006 were compared with GRDC-based estimates taken from Pan et al. (2012). Twenty-nine models comprise the CMIP5 AMIP *P*, *E*, and $P - E$ ensembles and 20 models comprise the *Q* ensemble (see Table S1). Non-river basins for which there is no GRDC-based estimate of *Q* are shaded in gray. Corresponding scatter plots like Figures 2–5 are provided in Figures S26–S29.

The number of significant differences between the AMIP ensemble means and observationally based estimates underscores limits to the spatiotemporal representativeness of the modeled water cycle. However, the fact that relatively more AMIP versus observation differences than HIST versus RCP8.5 differences are significant is due in part to our treatment of the observationally based products as truth (i.e., CI = 0), and

to a lesser extent, the relatively smaller ensemble size (i.e., AMIP P: n = 28; RCP8.5 P: n = 38) with less dispersion. All RCP8.5 projections should be interpreted in the context of known model systematic biases. For example, the finding that WEU $P - E$ is projected to decline in the future (Figure 6) is consistent with the finding that the models may be inherently predisposed to low $P - E$ in that basin (Figure 16). While space constraints preclude detailed discussion here, we have included assessments of AMIP variability and SI in Figures S30–S37).

4. Summary and Conclusion

Using CMIP5 as our case study, we have demonstrated a systematic approach to synthesizing large multimodel ensemble climate projections in the context of basin-scale water availability, as defined for meteorological (P), hydrological (Q), and agricultural/ecological ($P - E$) purposes. We assert that these basin-scale summaries constitute valuable and actionable information to regional water planners and decision makers. Our approach could be easily automated and implemented in a diagnostic toolkit, such as ESMValTools (Eyring, Righi, et al., 2016), and the results included as a compulsory component of future IPCC assessment reports. The comprehensiveness of the approach, which entails full inclusion of all available model output; identification and handling of statistical breakpoints, trends, nonnormality, and effective sample size; analysis of change and variability on multiple time scales; error characterization; and summary across 25 large-scale basins, distinguishes it from prior efforts. It is clear that inappropriate use of the t test in prior CMIP5 studies has led to the underestimation of statistically significant changes in future water availability by 15–25%.

As we have stressed throughout our study, knowing the seasonality of both projected changes and of the biased model background climatology onto which they are imposed is critical to ensuring proper interpretation and ascribing confidence. It is not enough to know whether a region will undergo drying or wetting over the long term. It is the seasonal timing of the projected change in water availability that is key. A prime example of this ideology is the reinforced seasonality, or signature S pattern manifested in the HIST-to-RCP8.5 1:1 scatter plots for WEU P.

Our AMIP comparisons against observational data sets exposed shortcomings in model representativeness. In our view, the best first step towards addressing these issues is to quantify uncertainty and face it head-on with follow-on observing system improvements and error attribution studies. We concede that the observational data sets are subject to shortcomings of their own, including limited station counts and varying network density over time (e.g., Ferguson & Villarini, 2012). These are challenging issues.

Importantly, comprehensive syntheses such as ours can serve to motivate, justify, and steer regional hydroclimate projects (RHPs), such as those supported programmatically through the Global Energy and Water Cycle Exchanges project of the World Climate Research Programme (e.g., Debeer et al., 2015; Drobinski et al., 2014; Groisman et al., 2009; van Dijk et al., 2013). RHPs fill important gaps in observations and concentrate the efforts of international, multidisciplinary teams towards improving understanding and prediction of a region's weather and hydroclimate. In this sense, basins in our study for which the model dispersion (CI) was shown to be relatively large may constitute candidate RHP targets.

Across all regions, it will be important to further assess the robustness of the smaller ESMRCP8.5 ensemble projections of fewer significant shifts in water availability relative to the full RCP8.5 ensemble. Specifically, to what processes (e.g., vegetation dynamics) may these differences be attributed and what is our current confidence in the corresponding parameterizations (e.g., Prudhomme et al., 2014)? The large number of parameters included in vegetation-carbon schemes and the myriad of pathways for feedbacks underscores the need to exercise great care in interpreting the results of such schemes (e.g., Kauwe et al., 2017). While attribution hypothesis testing is supported to a certain extent by monthly standard CMIP5 outputs (e.g., leaf area index, cloud properties, and radiation budget), the only path to irrefutable attribution lies through model perturbation "(sensitivity)" experiments conducted at, or in close coordination with the modeling centers themselves.

As the resolution of climate models increases, orographic-induced features such as the Great Plains low-level jet (e.g., Weaver & Nigam, 2008) and related mesoscale convective systems (MCSs; e.g., Fritsch et al., 1986; Jirak & Cotton, 2007) will be better represented and along with it the regional hydroclimate. Human land and water management represents yet another significant source of uncertainty for the local climate and regional water cycle that is not accounted for in CMIP5. Irrigation and reservoir operation rules are slowly

being integrated into land and coupled ESMs (e.g., Hanasaki et al., 2010; Neverre & Dumas, 2016; Pokhrel et al., 2016; Polcher et al., 2016).

In closing, we acknowledge limitations to our analysis framework as well as the probable existence of other equally valid frameworks in which projected water availability may be quantified and communicated. A variation on our approach in which ensemble members are not equally weighted but weighted according to regional (AMIP) skill would be meaningful (e.g., Knutti, Furrer, et al., 2010) because model performance does vary regionally. The same weighting convention could be applied to an ensemble of observationally based comparison data, which also has varying (perhaps complimentary) skill over time and space (e.g., Beck, van Dijk, et al., 2017). Finally, internal (atmospheric) variability, which constitutes irreducible uncertainty in the climate projections deserves further consideration (e.g., Hawkins et al., 2016; Thompson et al., 2015; Williams et al., 2017). The CMIP5 multimodel ensemble spread itself covers only a modest degree of all potential climate pathways (Deser et al., 2017). Special experiments may be designed to quantify the relative forced and unforced response (e.g., Kay et al., 2015; McKinnon et al., 2017). Communicating the inherent uncertainty in climate projections to decision makers will present an ongoing challenge.

Acknowledgments

C. R. F. designed the study, performed all analyses and wrote the manuscript. M. P. and T. O. contributed to the interpretation of results and reviewed and edited the manuscript. C. R. F. was supported by NSF award AGS-1638936. MSWEP data were obtained from Hylde Beck (http://www.gloh2o.org/), whose work was completed with IPA support from the U.S. Army Corps of Engineers' International Center for Integrated Water Resources Management (ICIWaRM), under the auspices of UNESCO. GLEAM data were obtained from Diego G. Miralles (https://www.gleam.eu/). CMIP5 data were obtained through the Earth System Grid Federation (ESGF) Peer-to-Peer (P2P) enterprise system. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table S1) 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 (GO-ESSP). C. R. F. also thanks Sergey Kivalov for helpful discussions on the matter of effective sample size.

References

Arguez, A., & Vose, R. S. (2011). The definition of the standard WMO climate normal the key to deriving alternative climate normals. *Bulletin of the American Meteorological Society*, *92*(6), 699–704. https://doi.org/10.1175/2010BAMS2955.1

Arnell, N. W., & Gosling, S. N. (2013). The impacts of climate change on river flow regimes at the global scale. *Journal of Hydrology*, *486*, 351–364. https://doi.org/10.1016/j.jhydrol.2013.02.010

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & de Roo, A. (2017). MSWEP: 3-hourly 0.25 degrees global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, *21*(1), 589–615. https://doi.org/10.5194/hess-21-589-2017

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., et al. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, *21*(12), 6201–6217. https://doi.org/10.5194/hess-21-6201-2017

Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M., & Bladé, I. (1999). The effective number of spatial degrees of freedom of a time-varying field. *Journal of Climate*, *12*(7), 1990–2009. https://doi.org/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2

Chou, C., & Lan, C.-W. (2012). Changes in the annual range of precipitation under global warming. *Journal of Climate*, *25*(1), 222–235. https://doi.org/10.1175/JCLI-D-11-00097.1

Chou, C., Neelin, J. D., Chen, C. A., & Tu, J. Y. (2009). Evaluating the "rich-get-richer" mechanism in tropical precipitation change under global warming. *Journal of Climate*, *22*(8), 1982–2005. https://doi.org/10.1175/2008JCLI2471.1

Collins, M., Knutti, R. A., Dufresne, J. M., Fichefet, J.-L., Thierry Wehner, M., et al. (2013). Long-term climate change: Projections, commitments and irreversibility. Climate change 2013: The physical science basis. In T. F. Stocker, et al. (Eds.), *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1029–1136). New York: Cambridge University Press.

Dai, A. G. (2013). Increasing drought under global warming in observations and models. *Nature Climate Change*, *3*(1), 52–58. https://doi.org/10.1038/nclimate1633

Dalin, C., Konar, M., Hanasaki, N., Rinaldo, A., & Rodriguez-Iturbe, I. (2012). Evolution of the global virtual water trade network. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(16), 5989–5994. https://doi.org/10.1073/pnas.1203176109

Debeer, C. M., Wheater, H. S., Quinton, W. L., Carey, S. K., Stewart, R. E., MacKay, M. D., & Marsh, P. (2015). The changing cold regions network: Observation, diagnosis and prediction of environmental change in the Saskatchewan and Mackenzie River basins, Canada. *Science China Earth Sciences*, *58*(1), 46–60. https://doi.org/10.1007/s11430-014-5001-6

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: The role of internal variability. *Climate Dynamics*, *38*(3-4), 527–546. https://doi.org/10.1007/s00382-010-0977-x

Deser, C., Simpson, I. R., McKinnon, K. A., & Phillips, A. S. (2017). The northern hemisphere extratropical atmospheric circulation response to ENSO: How well do we know it and how do we evaluate models accordingly? *Journal of Climate*, *30*(13), 5059–5082. https://doi.org/10.1175/JCLI-D-16-0844.1

Deser, C., Terray, L., & Phillips, A. S. (2016). Forced and internal components of winter air temperature trends over North America during the past 50 years: Mechanisms and implications. *Journal of Climate*, *29*(6), 2237–2258. https://doi.org/10.1175/JCLI-D-15-0304.1

Döll, P., & Schmied, H. M. (2012). How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis. *Environmental Research Letters*, *7*(1), 014037. https://doi.org/10.1088/1748-9326/7/1/014037

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., et al. (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*, *203*, 185–215. https://doi.org/10.1016/j.rse.2017.07.001

Drobinski, P., Ducrocq, V., Alpert, P., Anagnostou, E., Béranger, K., Borga, M., et al. (2014). HyMeX: A 10-year multidisciplinary program on the Mediterranean water cycle. *Bulletin of the American Meteorological Society*, *95*(7), 1063–1082. https://doi.org/10.1175/BAMS-D-12-00242.1

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Eyring, V., Righi, M., Lauer, A., Evaldsson, M., Wenzel, S., Jones, C., et al. (2016). ESMValTool (v1.0)—A community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geoscientific Model Development*, *9*(5), 1747–1802. https://doi.org/10.5194/gmd-9-1747-2016

Ferguson, C. R., & Mocko, D. M. (2017). Diagnosing an artificial trend in NLDAS-2 afternoon precipitation. *Journal of Hydrometeorology*, *18*(4), 1051–1070. https://doi.org/10.1175/JHM-D-16-0251.1

Ferguson, C. R., & Villarini, G. (2012). Detecting inhomogeneities in the twentieth century reanalysis over the central United States. *Journal of Geophysical Research*, *117*, D05123. https://doi.org/10.1029/2011JD016988

Ferguson, C. R., & Villarini, G. (2014). An evaluation of the statistical homogeneity of the twentieth century reanalysis. *Climate Dynamics*, *42*(11-12), 2841–2866. https://doi.org/10.1007/s00382-013-1996-1

Ferguson, C. R., Wood, E. F., & Vinukollu, R. K. (2012). A global intercomparison of modeled and observed land-atmosphere coupling. *Journal of Hydrometeorology*, *13*(3), 749–784. https://doi.org/10.1175/JHM-D-11-0119.1

Fritsch, J. M., Kane, R. J., & Chelius, C. R. (1986). The contribution of mesoscale convective weather systems to the warm-season precipitation in the United States. *Journal of Climate and Applied Meteorology*, *25*(10), 1333–1345. https://doi.org/10.1175/1520-0450(1986)025<1333:TCOMCW>2.0.CO;2

Gao, H., Tang, Q., Ferguson, C. R., Wood, E. F., & Lettenmaier, D. P. (2010). Estimating the water budget of major US river basins via remote sensing. *International Journal of Remote Sensing*, *31*(14), 3955–3978. https://doi.org/10.1080/01431161.2010.483488

Gash, J. H. C. (1979). An analytical model of rainfall interception by forests. *Quarterly Journal of the Royal Meteorological Society*, *105*(443), 43–55. https://doi.org/10.1002/qj.49710544304

Gillett, N. P., & Fyfe, J. C. (2013). Annular mode changes in the CMIP5 simulations. *Geophysical Research Letters*, *40*, 1189–1193. https://doi.org/10.1002/grl.50249

Greve, P., Orlowsky, B., Mueller, B., Sheffield, J., Reichstein, M., & Seneviratne, S. I. (2014). Global assessment of trends in wetting and drying over land. *Nature Geoscience*, *7*(10), 716–721. https://doi.org/10.1038/ngeo2247

Groisman, P. Y., Clark, E. A., Kattsov, V. M., Lettenmaier, D. P., Sokolik, I. N., Aizen, V. B., et al. (2009). The Northern Eurasia earth science partnership: An example of science applied to societal needs. *Bulletin of the American Meteorological Society*, *90*(5), 671–688. https://doi.org/10.1175/2008BAMS2556.1

Guo, Z. C., Dirmeyer, P. A., Koster, R. D., Sud, Y. C., Bonan, G., Oleson, K. W., et al. (2006). GLACE: The global land-atmosphere coupling experiment. Part II: Analysis. *Journal of Hydrometeorology*, *7*(4), 611–625. https://doi.org/10.1175/JHM511.1

Hanasaki, N., Inuzuka, T., Kanae, S., & Oki, T. (2010). An estimation of global virtual water flow and sources of water withdrawal for major crops and livestock products using a global hydrological model. *Journal of Hydrology*, *384*(3-4), 232–244. https://doi.org/10.1016/j.jhydrol.2009.09.028

Hawkins, E., Smith, R. S., Gregory, J. M., & Stainforth, D. A. (2016). Irreducible uncertainty in near-term climate projections. *Climate Dynamics*, *46*(11-12), 3807–3819. https://doi.org/10.1007/s00382-015-2806-8

Held, I. M., & Soden, B. J. (2006). Robust responses of the hydrological cycle to global warming. *Journal of Climate*, *19*(21), 5686–5699. https://doi.org/10.1175/JCLI3990.1

Ho, C. K., Stephenson, D. B., Collins, M., Ferro, C. A. T., & Brown, S. J. (2012). Calibration strategies: A source of additional uncertainty in climate change projections. *Bulletin of the American Meteorological Society*, *93*(1), 21–26. https://doi.org/10.1175/2011BAMS3110.1

Hunter, M. C., Smith, R. G., Schipanski, M. E., Atwood, L. W., & Mortensen, D. A. (2017). Agriculture in 2050: Recalibrating targets for sustainable intensification. *Bioscience*, *67*(4), 386–391. https://doi.org/10.1093/biosci/bix010

Hurst, H. E. (1957). A suggested statistical model of some time series which occur in nature. *Nature*, *180*, 494.

IPCC (2013). Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, et al. (Eds.) (1535 pp.). Cambridge, UK and New York: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324

Jiménez Cisneros, B. E., Oki, T., Arnell, N. W., Benito, G., Cogley, J. G., Döll, P., et al. (2014). Freshwater resources. In C. B. Field, V. R. Barros, D. J. Dokken, K. J. Mach, M. D. Mastrandrea, T. E. Bilir, et al. (Eds.), *Climate Change 2014: Impacts, adaptation, and vulnerability. Part A: Global and sectoral aspects. Contribution of working group II to the fifth assessment report of the intergovernmental panel on climate change* (pp. 229–269). Cambridge, UK and New York: Cambridge University Press.

Jirak, I. L., & Cotton, W. R. (2007). Observational analysis of the predictability of mesoscale convective systems. *Weather and Forecasting*, *22*(4), 813–838. https://doi.org/10.1175/WAF1012.1

Kauwe, M. G. D., Medlyn, B. E., Walker, A. P., Zaehle, S., Asao, S., Guenet, B., et al. (2017). Challenging terrestrial biosphere models with data from the long-term multifactor prairie heating and CO2 enrichment experiment. *Global Change Biology*, *23*(9), 3623–3645. https://doi.org/10.1111/gcb.13643

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The community earth system model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349. https://doi.org/10.1175/BAMS-D-13-00255.1

Kendall, M. G. (1975). *Rank correlation methods* (4th ed.). London: Charles Griffin.

Knutti, R., Abramowitz, G., Collins, M., Eyring, V., Gleckler, P., Hewitson, B., & Mearns, L. (2010). IPCC good practice guidance paper on assessing and combining multi model climate projections. In T. F. Stocker, et al. (Eds.), *Meeting Report of the IPCC Expert Meeting on Assessing and Combining Multi Model Climate Projections* (pp. 1–13). IPCC Working Group I Tech. Supp. Unit, University of Bern, Bern.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., & Meehl, G. A. (2010). Challenges in combining projections from multiple climate models. *Journal of Climate*, *23*(10), 2739–2758. https://doi.org/10.1175/2009JCLI3361.1

Koirala, S., Hirabayashi, Y., Mahendran, R., & Kanae, S. (2014). Global assessment of agreement among streamflow projections using CMIP5 model outputs. *Environmental Research Letters*, *9*(6), 064017. https://doi.org/10.1088/1748-9326/9/6/064017

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., & GLACE Team (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, *305*(5687), 1138–1140. https://doi.org/10.1126/science.1100217

Koster, R. D., Sud, Y. C., Guo, Z., Dirmeyer, P. A., Bonan, G., Oleson, K. W., et al. (2006). GLACE: The global land-atmosphere coupling experiment. Part I: Overview. *Journal of Hydrometeorology*, *7*(4), 590–610. https://doi.org/10.1175/JHM510.1

Lau, W. K. M., Wu, H. T., & Kim, K. M. (2013). A canonical response of precipitation characteristics to global warming from CMIP5 models. *Geophysical Research Letters*, *40*, 3163–3169. https://doi.org/10.1002/grl.50420

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., & Wood, E. F. (2015). Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research*, *51*, 7309–7342. https://doi.org/10.1002/2015WR017616

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, *62*(318), 399–402. https://doi.org/10.1080/01621459.1967.10482916

Liu, C. L., & Allan, R. P. (2013). Observed and simulated precipitation responses in wet and dry regions 1850-2100. *Environmental Research Letters*, *8*, 034002.

Liu, Y. Y., de Jeu, R. A. M., McCabe, M. F., Evans, J. P., & van Dijk, A. I. J. M. (2011). Global long-term passive microwave satellite-based retrievals of vegetation optical depth. *Geophysical Research Letters*, *38*, L18402. https://doi.org/10.1029/2011GL048684

Liu, Y. Y., Dorigo, W. A., Parinussa, R. M., de Jeu, R. A. M., Wagner, W., McCabe, M. F., et al. (2012). Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sensing of Environment*, *123*, 280–297. https://doi.org/10.1016/j.rse.2012.03.014

Liu, Y. Y., van Dijk, A. I. J. M., McCabe, M. F., Evans, J. P., & de Jeu, R. A. M. (2013). Global vegetation biomass change (1988–2008) and attribution to environmental and human drivers. *Global Ecology and Biogeography*, *22*(6), 692–705. https://doi.org/10.1111/geb.12024

Luke, A., Vrugt, J. A., AghaKouchak, A., Matthew, R., & Sanders, B. F. (2017). Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. *Water Resources Research*, *53*, 5469–5494. https://doi.org/10.1002/2016WR019676

Mallakpour, I., & Villarini, G. (2016). A simulation study to examine the sensitivity of the Pettitt test to detect abrupt changes in mean. *Hydrological Sciences Journal*, *61*(2), 245–254. https://doi.org/10.1080/02626667.2015.1008482

Maloney, E. D., Camargo, S. J., Chang, E., Colle, B., Fu, R., Geil, K. L., et al. (2014). North American climate in CMIP5 experiments: Part III: Assessment of twenty-first-century projections. *Journal of Climate*, *27*(6), 2230–2270. https://doi.org/10.1175/JCLI-D-13-00273.1

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*, *13*(3), 245. https://doi.org/10.2307/1907187

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., et al. (2017). GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, *10*(5), 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017

Masood, M., Yeh, P. J. F., Hanasaki, N., & Takeuchi, K. (2015). Model study of the impacts of future climate change on the hydrology of Ganges-Brahmaputra-Meghna basin. *Hydrology and Earth System Sciences*, *19*(2), 747–770. https://doi.org/10.5194/hess-19-747-2015

McCabe, M. F., Ershadi, A., Jimenez, C., Miralles, D. G., Michel, D., & Wood, E. F. (2016). The GEWEX LandFlux project: Evaluation of model evaporation using tower-based and globally gridded forcing data. *Geoscientific Model Development*, *9*(1), 283–305. https://doi.org/10.5194/gmd-9-283-2016

McKinnon, K. A., Poppick, A., Dunn-Sigouin, E., & Deser, C. (2017). An "observational large ensemble" to compare observed and modeled temperature trend uncertainty due to internal variability. *Journal of Climate*, *30*(19), 7585–7598. https://doi.org/10.1175/JCLI-D-16-0905.1

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Climate change—Stationarity is dead: Whither water management? *Science*, *319*(5863), 573–574. https://doi.org/10.1126/science.1151915

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., & Dolman, A. J. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences*, *15*(2), 453–469. https://doi.org/10.5194/hess-15-453-2011

Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M. F., et al. (2016). The WACMOS-ET project—Part 2: Evaluation of global terrestrial evaporation data sets. *Hydrology and Earth System Sciences*, *20*(2), 823–842. https://doi.org/10.5194/hess-20-823-2016

Neverre, N., & Dumas, P. (2016). Projecting basin-scale distributed irrigation and domestic water demands and values: A generic method for large-scale modeling. *Water Economics and Policy*, *02*(04), 1650023. https://doi.org/10.1142/S2382624X16500235

Oki, T., & Kanae, S. (2004). Virtual water trade and world water resources. *Water Science and Technology*, *49*(7), 203–209. https://doi.org/10.2166/wst.2004.0456

Oki, T., Yano, S., & Hanasaki, N. (2017). Economic aspects of virtual water trade. *Environmental Research Letters*, *12*(4), 044002. https://doi.org/10.1088/1748-9326/aa625f

Orlowsky, B., & Seneviratne, S. I. (2013). Elusive drought: Uncertainty in observed trends and short- and long-term CMIP5 projections. *Hydrology and Earth System Sciences*, *17*(5), 1765–1781. https://doi.org/10.5194/hess-17-1765-2013

Pan, M., Sahoo, A. K., Troy, T. J., Vinukollu, R. K., Sheffield, J., & Wood, E. F. (2012). Multisource estimation of long-term terrestrial water budget for major Global River basins. *Journal of Climate*, *25*(9), 3191–3206. https://doi.org/10.1175/JCLI-D-11-00300.1

Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *Applied Statistics*, *28*(2), 126–135. https://doi.org/10.2307/2346729

Pokhrel, Y. N., Fan, Y., Miguez-Macho, G., Yeh, P. J. F., & Han, S. C. (2013). The role of groundwater in the Amazon water cycle: 3. Influence on terrestrial water storage computations and comparison with GRACE. *Journal of Geophysical Research: Atmospheres*, *118*, 3233–3244. https://doi.org/10.1002/jgrd.50335

Pokhrel, Y. N., Hanasaki, N., Wada, Y., & Kim, H. (2016). Recent progresses in incorporating human land–water management into global land surface models toward their integration into Earth system models. *Wiley Interdisciplinary Reviews Water*, *3*(4), 548–574. https://doi.org/10.1002/wat2.1150

Polcher, J., Harding, R. J., Boone, A., Florke, M., Oki, T., & Quintana Segui, P. (2016). A new GHP/GLASS crosscutting project: Human regulation of the water cycle (HRWC). *GEWEX Newsletter*, *26*, 3.

Priestley, C. H. B., & Taylor, R. J. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, *100*(2), 81–92. https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., et al. (2014). Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proceedings of the National Academy of Sciences*, *111*(9), 3262–3267. https://doi.org/10.1073/pnas.1222473110

Pryor, S. C., & Schoof, J. T. (2008). Changes in the seasonality of precipitation over the contiguous USA. *Journal of Geophysical Research*, *113*, D21108. https://doi.org/10.1029/2008JD010251

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., et al. (2011). RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Climatic Change*, *109*(1-2), 33–57. https://doi.org/10.1007/s10584-011-0149-y

Rodell, M., Velicogna, I., & Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. *Nature*, *460*(7258), 999–1002. https://doi.org/10.1038/nature08238

Rudmin, J. W. (2010). *Calculating the exact pooled variance* (p. 4). Harrisonburg, VA: Physics Department James Madison University. Retrieved from https://arxiv.org/ftp/arxiv/papers/1007/1007.1012.pdf

Scaife, A. A., Kucharski, F., Folland, C. K., Kinter, J., Brönnimann, S., Fereday, D., et al. (2009). The CLIVAR C20C project: Selected twentieth century climate events. *Climate Dynamics*, *33*(5), 603–614. https://doi.org/10.1007/s00382-008-0451-1

Serinaldi, F., & Kilsby, C. G. (2015). Stationarity is undead: Uncertainty dominates the distribution of extremes. *Advances in Water Resources*, *77*, 17–36. https://doi.org/10.1016/j.advwatres.2014.12.013

Sheffield, J., Barrett, A. P., Colle, B., Nelun Fernando, D., Fu, R., Geil, K. L., et al. (2013). North American climate in CMIP5 experiments. Part I: Evaluation of historical simulations of continental and regional climatology. *Journal of Climate*, *26*(23), 9209–9245. https://doi.org/10.1175/JCLI-D-12-00592.1

Sheffield, J., Camargo, S. J., Fu, R., Hu, Q., Jiang, X., Johnson, N., et al. (2013). North American climate in CMIP5 experiments. Part II: Evaluation of historical simulations of Intraseasonal to decadal variability. *Journal of Climate*, *26*(23), 9247–9290. https://doi.org/10.1175/JCLI-D-12-00593.1

Sheffield, J., Ferguson, C. R., Troy, T. J., Wood, E. F., & McCabe, M. F. (2009). Closing the terrestrial water budget from satellite remote sensing. *Geophysical Research Letters*, *36*, L07403. https://doi.org/10.1029/2009GL037338

Sheffield, J., Goteti, G., & Wood, E. F. (2006). Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate*, *19*(13), 3088–3111. https://doi.org/10.1175/JCLI3790.1

Tabari, H., & Willems, P. (2018). Seasonally varying footprint of climate change on precipitation in the Middle East. *Scientific Reports*, *8*(1), 4435. https://doi.org/10.1038/s41598-018-22795-8

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Thompson, D. W. J., Barnes, E. A., Deser, C., Foust, W. E., & Phillips, A. S. (2015). Quantifying the role of internal climate variability in future climate trends. *Journal of Climate*, *28*(16), 6443–6456. https://doi.org/10.1175/JCLI-D-14-00830.1

United Nations (2018). The sustainable development goals report 2018. United Nations Statistics Division.

Valente, F., David, J. S., & Gash, J. H. C. (1997). Modelling interception loss for two sparse eucalypt and pine forests in central Portugal using reformulated Rutter and Gash analytical models. *Journal of Hydrology*, *190*(1-2), 141–162. https://doi.org/10.1016/S0022-1694(96)03066-1

van Dijk, A. I. J. M., Abramowitz, G., Evans, B., Gouweleeuw, B., Guerschman, J. P., Johnson, F., et al. (2013). OzEWEX: The Australian energy and water exchange initiative. *GEWEX Newsletter*, *23*, 4.

Vetter, T., Huang, S., Aich, V., Yang, T., Wang, X., Krysanova, V., & Hattermann, F. (2015). Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents. *Earth System Dynamics*, *6*(1), 17–43. https://doi.org/10.5194/esd-6-17-2015

Walsh, R. P. D., & Lawler, D. M. (1981). Rainfall seasonality: Description, spatial patterns and change through time. *Weather*, *36*(7), 201–208. https://doi.org/10.1002/j.1477-8696.1981.tb05400.x

Watanabe, S., Kanae, S., Seto, S., Yeh, P. J. F., Hirabayashi, Y., & Oki, T. (2012). Intercomparison of bias-correction methods for monthly temperature and precipitation simulated by multiple climate models. *Journal of Geophysical Research*, *117*, D23114. https://doi.org/10.1029/2012JD018192

Weaver, S. J., & Nigam, S. (2008). Variability of the Great Plains low-level jet: Large-scale circulation context and hydroclimate impacts. *Journal of Climate*, *21*(7), 1532–1551. https://doi.org/10.1175/2007JCLI1586.1

Williams, P. D., Alexander, M. J., Barnes, E. A., Butler, A. H., Davies, H. C., Garfinkel, C. I., et al. (2017). A census of atmospheric variability from seconds to decades. *Geophysical Research Letters*, *44*, 11,201–11,211. https://doi.org/10.1002/2017GL075483

Wills, R. C., Byrne, M. P., & Schneider, T. (2016). Thermodynamic and dynamic controls on changes in the zonally anomalous hydrological cycle. *Geophysical Research Letters*, *43*, 4640–4649. https://doi.org/10.1002/2016GL068418

Yue, S., Pilon, P., Phinney, B., & Cavadias, G. (2002). The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes*, *16*(9), 1807–1829. https://doi.org/10.1002/hyp.1095