

Model Assessment of Observed Precipitation Trends over Land Regions: Detectable Human Influences and Possible Low Bias in Model Trends

THOMAS R. KNUTSON AND FANRONG ZENG

NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

(Manuscript received 6 October 2017, in final form 1 February 2018)

ABSTRACT

Precipitation trends for 1901–2010, 1951–2010, and 1981–2010 over relatively well-observed global land regions are assessed for detectable anthropogenic influences and for consistency with historical simulations from phase 5 of the Coupled Model Intercomparison Project (CMIP5). The CMIP5 historical all-forcing runs are broadly consistent with the observed trend pattern (1901–2010), but with an apparent low trend bias tendency in the simulations. Despite this bias, observed and modeled trends are statistically consistent over 59% of the analyzed area. Over 20% (9%) of the analyzed area, increased (decreased) precipitation is partly attributable to anthropogenic forcing. These inferred human-induced changes include increases over regions of the north-central United States, southern Canada, Europe, and southern South America and decreases over parts of the Mediterranean region and northern tropical Africa. Trends for the shorter periods (1951–2010 and 1981–2010) do not indicate a prominent low trend bias in the models, as found for the 1901–2010 trends. An atmosphere-only model, forced with observed sea surface temperatures and other climate forcing agents, also underpredicts the observed precipitation increase in the Northern Hemisphere extratropics since 1901. The CMIP5 all-forcing ensemble's low bias in simulated trends since 1901 is a tentative finding that, if borne out in further studies, suggests that precipitation projections using these regions and models could overestimate future drought risk and underestimate future flooding risk, assuming all other factors equal.

1. Introduction

Precipitation changes associated with anthropogenic climate change have the potential for great societal impacts, as precipitation is a key driver of drought and flood risk. Previous studies have documented observed precipitation trends, including regional patterns of trends. For example, in the IPCC Fourth Assessment Report (AR4), trends over 1901–2005 were analyzed, statistically significant trends identified, and time series were shown for a number of key regions (Trenberth et al. 2007; Figs. 3.13, 3.14). In the IPCC AR5, regional trend maps were compared for three different observational datasets and two time periods (1901–2010, 1951–2010) (Hartmann et al. 2013; Fig. 2.29). The observed trend patterns were similar across the different datasets,

although they differed across time periods. However, the ability of historical climate model simulations to reproduce the observed trends in precipitation is less well studied, and this issue is one of the key motivations of our study.

Precipitation projections over the twenty-first and twenty-second centuries under the RCP8.5 climate forcing scenarios are shown in Collins et al. (2013) (Fig. 12.22 of IPCC AR5) and have similar general patterns to those from previous IPCC assessments, particularly outside of the tropics. While this general consistency of modeled precipitation projections across different model generations is suggestive of a robust physical response, more confidence could be gained in such projections if the precipitation trends from climate model historical simulations were also demonstrated to be consistent with the observed trends.

Further motivating our study, future precipitation projections are a critical issue for future societal impacts, including drought risk. For example, recent studies conclude that for historical changes in California (Cheng et al. 2016) or future greenhouse-gas-induced warming (Berg et al. 2017; Roderick et al. 2015), changes in drought risk, as inferred from deep or vertically integrated soil moisture changes (Berg et al. 2017; Cheng et al. 2016) and changes

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-17-0672.s1>.

Corresponding author: Thomas R. Knutson, tom.knutson@noaa.gov

DOI: 10.1175/JCLI-D-17-0672.1

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy \(www.ametsoc.org/PUBSReuseLicenses\)](http://www.ametsoc.org/PUBSReuseLicenses).

in meteorological aridity (Roderick et al. 2015), are closely linked to changes in precipitation.

Previous studies have concluded that some aspects of observed historical trends of mean (seasonal or annual) precipitation or related variables (ocean near-surface salinity) can be attributed to anthropogenic forcing. A recent review of detection and attribution studies for regional precipitation changes (Sarojini et al. 2016) outlines some of the observational and modeling challenges in this research area. Despite these challenges, some examples of mean precipitation metrics with attributable human influence include annual zonal mean precipitation changes over land (Zhang et al. 2007), changes in the tropical marine water cycle (Terray et al. 2012), and midlatitude drying and high-latitude moistening in the extratropical Southern Hemisphere summer (Fyfe et al. 2012). Examples of regional *increases* of precipitation that are attributable in part to human influence include over northern high latitudes (Min et al. 2008; Wan et al. 2015), southern subtropics in summer due to polar ozone depletion (Kang et al. 2011), southeastern South America during summertime due to increased greenhouse gases (Zhang et al. 2016), and the Sahel region since the 1980s, also due to increased greenhouse gases (Dong and Sutton 2015). Examples of regional *decreases* in precipitation that have been attributed to human influence include over southern and southwest Australia (Delworth and Zeng 2014), the Mediterranean region during wintertime (Hoerling et al. 2012), and the Sahel region from the 1950s to the 1980s (Held et al. 2005). Other studies have reported evidence for external influence on precipitation trends, and although these studies did not attribute the changes specifically to anthropogenic influence, in some cases an anthropogenic influence is suggested by the results. Examples of such studies include seasonal zonal-average land precipitation (Noake et al. 2012; Polson et al. 2013), zonal-average (land and ocean) precipitation (Marvel and Bonfils 2013), precipitation increases over parts of the United States (Knutson et al. 2014; Anderson et al. 2015), and possibly the regions with significant linear trends on observed precipitation trend maps (Hartmann et al. 2013). The above studies were generally based on comparison of observations and climate model simulations, with the exception of the statistical trend analyses of Hartmann et al. (2013) and Anderson et al. (2015), which were based on observations alone. In our study, we do not address the issue of changes in extreme precipitation, which has been addressed in a number of previous studies (e.g., Min et al. 2011; Bindoff et al. 2014; Westra et al. 2013; Zhang et al. 2013; Dittus et al. 2016; Fischer and Knutti 2016; Taylor et al. 2017). More physically comprehensive alternative

metrics of land-region wetting or drying, compared to precipitation alone, have been assessed for observations (Greve et al. 2014) but also are not considered here.

Comparatively less has been done to assess the degree to which historical regional precipitation trends are *consistent* with climate model historical forcing simulations. Examples include van Oldenborgh et al. (2013) and Bhend and Whetton (2013), who assessed the consistency of CMIP5 model historical forcing runs versus observed regional precipitation trends since 1951. Van Oldenborgh et al. conclude that the model-simulated precipitation trends are overconfident (i.e., in some sense inconsistent) with regard to observations. Kumar et al. (2013) find relatively low (<0.25) spatial correlations between CMIP5 simulated and observed precipitation trends since 1930. The successful simulation of past or future regional precipitation trends is likely to be particularly challenging for climate models, particularly since precipitation changes will depend on small-scale processes that typically must be parameterized in climate models, and, as a further complication, precipitation changes interact with the atmospheric circulation (Jakob 2010, 2014). Differences between current climate models in responses of large-scale precipitation patterns to even simplified uniform ocean warming can be particularly striking in the tropics, where moist convective parameterizations are particularly important for the coupling of water and circulation (Stevens and Bony 2013). Given these modeling challenges for precipitation and its coupling to the circulation, one should not necessarily expect the modeled responses of regional precipitation patterns to external forcings, such as increased greenhouse gases, to be reliable or robust. Our consistency tests can help objectively assess how well the CMIP5 models perform in reproducing historical regional precipitation trends.

Our study extends both the detection/attribution and model–observation consistency lines of research by systematically comparing, at the gridpoint scale, the observed precipitation trends over land regions, primarily from the GPCC dataset, to an ensemble of climate model historical forcing simulations (see section 2) from the CMIP5 archive (Taylor et al. 2012). We focus on trends on three different trend time scales, up to the centennial scale (1901–2010), and over limited land regions. Through our methodology, we are able to categorize observed trends into classes related to detection, consistency, and attributable anthropogenic influence, as outlined in the results section (section 3). Novel aspects of our work include the detection and model-based attribution of anthropogenic influence on precipitation trends for several additional land regions around the globe compared to previous studies, and an assessment

of the model–observation consistency for precipitation trends in land regions (at the gridpoint scale and for zonal averages) for the three time scales.

We describe our methodology in section 2, present analysis results including trend assessments and sensitivity tests in section 3, consider the possible role of observed data quality issues in section 4, and present discussion and our main conclusions in section 5.

2. Methodology

In this section, we provide details of our methodology, including observed and model datasets used, and some sensitivity tests. In the appendix, we list the CMIP5 model sets used, and provide details of data masking and the methods used for transforming precipitation data into a standardized index form for sensitivity tests. In the study, we use the linear trend as the general metric to describe long-term changes in precipitation (treating observed and modeled data in the same manner). Other types of systematic changes over time, such as changes in means between different epochs, changes in precipitation extremes (Min et al. 2011; Westra et al. 2013; Zhang et al. 2013), or abrupt changes in precipitation (Narisma et al. 2007) could be explored in future extensions.

a. Observational datasets

The primary observational dataset used for the analysis was the Global Precipitation Climatology Centre (GPCC) V6 monthly dataset on a $2.5^\circ \times 2.5^\circ$ grid (Becker et al. 2013), file `precip.mon.total.2.5x2.5.v6.nc` downloaded March 2015 from <https://www.esrl.noaa.gov/psd/data/gridded/data.gpcc.html>. As a sensitivity test for the observations, we also present some trend assessment results using the Climatic Research Unit CRU_ts3.24.01 monthly precipitation dataset on a $0.5^\circ \times 0.5^\circ$ grid (<https://crudata.uea.ac.uk/cru/data/hrg/>; updated from Harris et al. 2014) and the Global Historical Climatology Network (GHCN) version 2 monthly precipitation dataset (Vose et al. 1992; <https://www.ncdc.noaa.gov/temp-and-precip/gHCN-gridded-products/>).

b. CMIP5 models, experiments, and multimodel trend distributions

The CMIP5 multimodel ensemble used for the main anthropogenic influence assessments in our study comprises 10 individual models (see the appendix). The set of individual model runs available to us for each of these 10 models includes 1) a preindustrial control run, 2) at least one all-forcing ensemble member extending to 2010 (with “all-forcing” referring to both anthropogenic

and natural forcing agents; see Knutson et al. 2013), and 3) at least one “natural-forcing-only” ensemble member extending to 2010. The all-forcing runs include both anthropogenic and natural forcing agents and thus incorporate the most comprehensive set of external forcings among the CMIP5 simulations. They represent the most comprehensive effort by the modeling groups to simulate the response of the climate system to a set of known external forcing agents believed to be of primary importance for climate change over the historical period. We estimate the anthropogenic forced response for each year as the difference between the models’ ensemble mean from the all-forcing experiments and the ensemble mean from the natural-forcing experiments. For the case of multimodel ensembles, the ensemble means of the individual models are first created and then averaged to create the multimodel ensemble, which results in an equal weighting for each individual model included in the multimodel ensemble regardless of the number of ensemble members available for each individual model. The natural-forcing runs were forced by estimated past changes in solar irradiance and volcanic aerosol loading. For some of the results presented in this study, we used a larger sample of 36 CMIP5 models (see the appendix), as these particular analyses did not require natural-forcing-only runs extending to 2010, and so we could include the full sample of all-forcing models.

The CMIP5 model archive contains simulations contributed from many different modeling groups. This requires that we adopt a methodology for combining information from different models in our assessment of the observed trends. Our standard or default multimodel assessment combines the information from a set of models into a single distribution of trends for all forcing and another single distribution for the natural forcing, where these single distributions have the *average characteristics* of the group of models they represent, as discussed in detail below. As a sensitivity test, this standard approach is later modified as we assess either individual models or as we combine multiple models in an alternative way (described below).

The multimodel ensemble trend distributions for our standard (default) tests in this study are created by first developing trend distributions—combining ensemble means of forced runs with variability from long control runs—for each of the individual models separately and then averaging those to create a single *average distribution* with the average properties (mean, variance, 5th/95th percentiles) of the trends from the available models. This differs from the approach of Knutson et al. (2013), who analyzed the *aggregate* multimodel distribution of trends obtained by combining the distributions of trends from the individual models into a grand

aggregate distribution and computing the mean and 5th/95th percentiles of that grand multimodel distribution. The aggregate approach generally produces a wider distribution than our standard (average distribution) approach, as it includes some spread due to differences in mean trend behavior across the models as well as due to the internal variability. The aggregate internal variability contribution to the spread also tends to be influenced strongly by the more variable models. Sensitivity tests using the aggregated distribution approach of Knutson et al. (2013) are presented in the online supplemental material.

c. Other sensitivity tests

Among our sensitivity tests, we compute and assess linear trends for a standardized precipitation index (SPI), where modeled and observed monthly precipitation anomalies at each grid point have been transformed into a standardized index having approximately normal distribution and unit variance (described above). This sensitivity test reduces the influence of non-Gaussian behavior and also controls for model biases in simulated variability, since the transformed observed and modeled time series by definition have unit variance at each grid point. Additional sensitivity tests, described briefly in section 3, were performed to examine seasonal (vs. annual) trends and to assess individual models independently rather than combined as a group. We also compare trends across three different observed datasets. As an alternative analysis approach that is more focused on extreme monthly means, we analyze time series of the percent of global analyzed area with monthly precipitation index values above or below certain statistical threshold values for models and observations (see the supplemental material).

3. Results

a. Evaluation of modeled precipitation variability

Simulated internal variability in climate models is an important component of the trend assessment tests in our study. The simulated low-frequency (>10 yr) internal variability of the CMIP5 models, as obtained from long control (unforced) integrations, is compared to an observed internal variability estimate in Fig. 1. The observed internal variability estimate is obtained by subtracting the multimodel ensemble mean of the all-forcing experiments from observations to create an estimated internal variability residual time series. This method assumes that the forced variability signal from the all-forcing experiments can be used as the estimate of the forced variability in the real world. The ensemble average and the average internal variability maps are

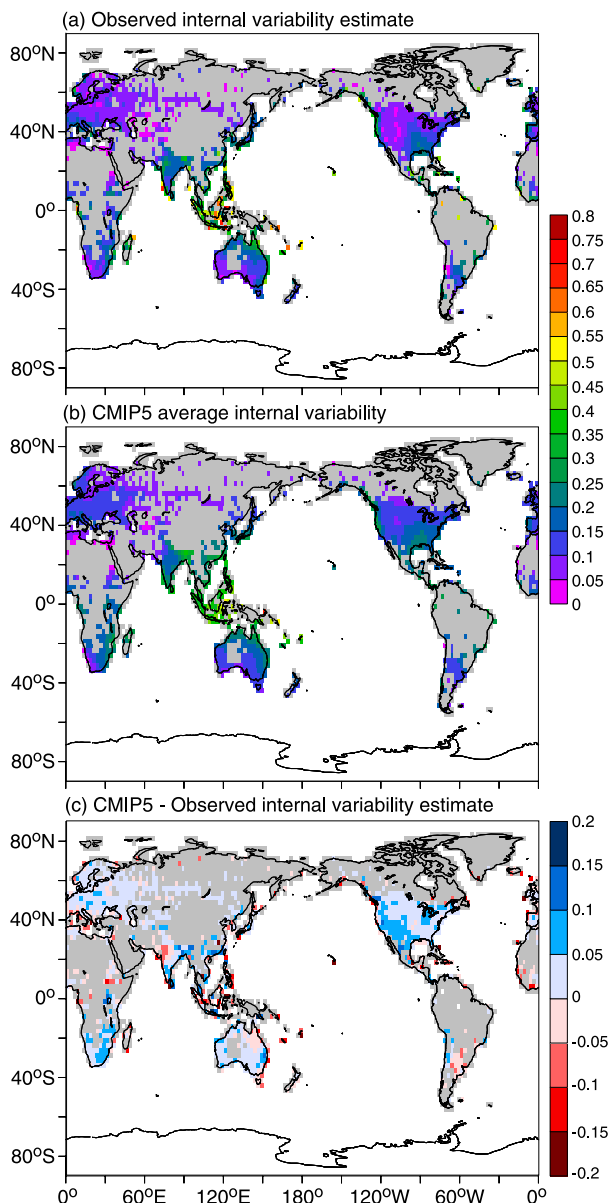


FIG. 1. Comparison of estimated low-frequency internal variability from observations and models. (a) Observed estimate based on the standard deviation of residuals formed by subtracting the CMIP5 ensemble mean all-forcing simulation from observations. (b) Ensemble average of low-frequency standard deviations from the 36 CMIP5 models. (c) CMIP5 ensemble standard deviation minus observed estimate [(b) - (a)]. Low-frequency variations based on low-pass filtered (>10 variability) data. Unit: mm yr^{-1} .

based on all 36 CMIP5 models. Overall, the comparison shows reasonable agreement between the modeled and observed variability maps. According to the multimodel average variability difference map (Fig. 1c), most of the nongray (adequate data) portions of the map have a blue shading, indicating that the modeled internal decadal variability tends to be larger than the observed (residual)

internal variability estimate. In these blue-shaded areas, the apparent model bias would tend to make our trend detection tests overly conservative while our consistency tests (observations versus all-forcing simulations) would be underconservative (too easy to achieve consistency). We judge the simulated internal variability overall to be sufficient for the purposes of the trend assessments in our study. We assess normalized (SPI) data as a sensitivity test later in this report (removing the influence on trends of variance differences between models and observations), and our trend detection results can also be compared with previous studies (e.g., Hartmann et al. 2013) that used different methods. Nonetheless, a caveat of our study is that the true observed level of internally generated low-frequency variability is not perfectly known, as our estimate is based on a limited sample in time and assumes that the models' estimate of the forced response is correct. In addition, discrepancies between the model simulated and observed variability estimates are apparent in Fig. 1. Standard deviation difference maps analogous to that in Fig. 1c but for each individual model are contained in the supplemental material.

b. Trend assessments

Our trend assessment classifies observed trends into a set of categories, defined in this section. The category definitions are applied at each grid point that has sufficient data coverage for trend analysis (see the appendix). We then plot maps showing the assessed category for each grid point for a given trend period of interest. Limited spatial/time coverage of data has been shown to have an important influence on precipitation trends at the zonal mean scale (Sarojini et al. 2012). Therefore, for consistency, we masked out the model data for times/locations where observed data are not available.

The defined categories describe the relationship between observed and modeled trends in our basic assessment figures. The categories are schematically illustrated in Fig. 2 using hypothetical distributions of trends from the all-forcing and natural-forcing experiments. Although more categories are possible, here we use just nine categories ($-4, -3, \dots, +4$). Our nine categories are not designed to cover all possibilities, and a small percent of cases were found to not fit into any of the nine categories and for simplicity are not plotted.

We first consider whether *detectable* trends in regional precipitation are identifiable in the observations (categories $-3, -2, -1, +1, +2, +3$). Detectable trends are defined here as those that are statistically distinct from (i.e., outside the 5th to 95th percentile range of) large samples of simulated trends due to natural variability, with the latter obtained by combining internal variability trends from CMIP5 control runs with the

ensemble mean trends from natural-forcing-only experiments. For detection, we also require that the observed trend have the same sign as the ensemble mean trend from the CMIP5 all-forcing runs. This definition of detection parallels that of detection/attribution studies using regression/pattern-based methods, where detection is claimed when the regression scale factor significantly exceeds zero (i.e., the model has the correct sign of change, among other requirements). Our tests assume that the models' preindustrial control runs realistically simulate internal climate variability, even when applied to the late-twentieth-century climate. Observed trends that are classified as nondetectable thus include both weak trends (category 0) as well as strong observed wetting or drying trends for which the wrong sign of change is simulated in the all-forcing runs (categories $-4, +4$).

Detectable consistent anthropogenic influence (categories $+2, -2$) is inferred where observed changes are both distinct from modeled natural variability (simulated internal variability combined with the response to external natural forcing) and are also consistent with historical all-forcing experiments that include both anthropogenic and natural forcing agents. Note that we estimate the anthropogenic forced response as the difference between the all-forcing and natural-forcing responses. The case of detectable but inconsistent inferred anthropogenic influence (categories $+3, -3$) occurs where a detectable observed change is found, and the anthropogenic forcing response is of the correct sign, but the observed change is nonetheless greater than the 95th percentile (for positive trend) or less than the 5th percentile (negative trend) of the trend distribution from the combined all-forcing and control experiments. Note that for attribution categories ($+3, -3$), we also require that the observed trend be closer to the all-forcing ensemble mean trend than to the natural-forcing ensemble mean trend.

As part of our tests, we assess the *consistency* between observed trends and those from the CMIP5 all-forcing historical runs, apart from the issue of climate change detection; consistent trends are identified by stippling in our assessment maps. We define consistency as simply the case where simulated (all forcing) trends are consistent with observed trends, regardless of the magnitude or statistical significance of the observed (or modeled) trends. It is possible for an observed trend to be consistent with both the all-forcing and the natural-forcing experiments (i.e., where trends are weak and nondetectable but nonetheless consistent with all-forcing experiments). These cases are identified in our assessment maps by gray shading with white stippling.

Among the inconsistent cases, we also identify cases of relatively strong (i.e., distinct from simulated natural

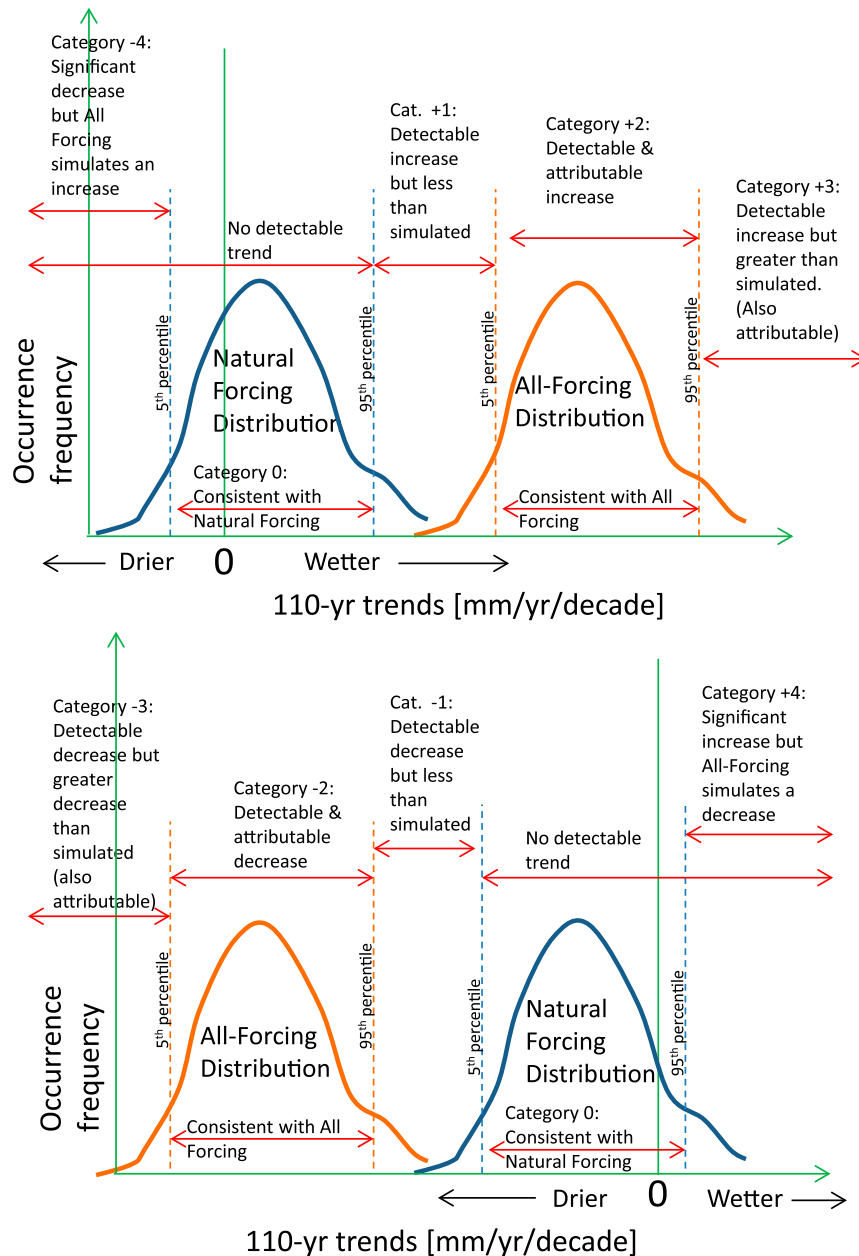


FIG. 2. Schematic diagram illustrating different categories of trend assessment result for hypothetical all-forcing and natural-forcing distributions. The dashed vertical lines are 5th and 95th percentiles of the trend distributions for all-forcing (orange) and natural-forcing (blue) distributions. Axes are green; the zero trend line vertical axis (green) is labeled "0". The category labels indicate the assessment category that would apply if the observed trend fell in the indicated ranges marked by the two-sided red arrows.

variability) trends in observations which are inconsistent with the all-forcing experiments by being either 1) significantly weaker than like-signed changes in the all-forcing runs but still detectable (categories +1, -1), or 2) having the opposite sign of observed trend compared to the ensemble mean of the all-forcing runs (categories +4, -4). Categories +1 and -1 are quite

rare in our analyses, as these cases require, among other things, that the all-forcing trends be relatively well separated from the natural-forcing trend distribution and that the observed trend be much weaker than the all-forcing trend but yet still strong enough to be distinguished from the natural-forcing distribution. Categories -4 and +4, which identify cases of particularly

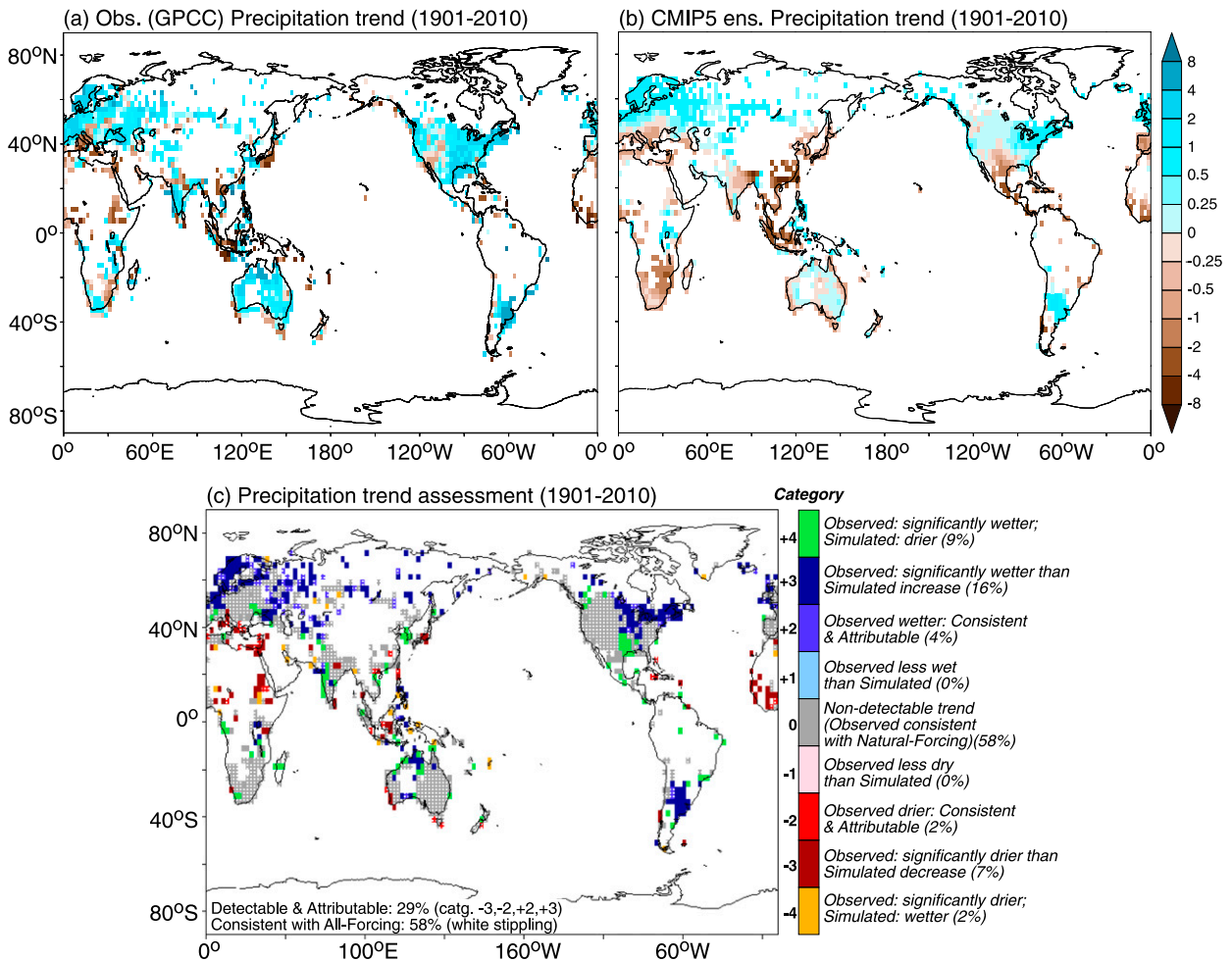


FIG. 3. Assessment of observed (GPCC) precipitation trends over 1901–2010 based on 10 CMIP5 models. (a) Observed and (b) CMIP5 multimodel ensemble precipitation trends in units of $\text{mm yr}^{-1} \text{decade}^{-1}$. (c) Model-based summary assessment of the observed trend at each grid point having sufficient data coverage. Nine assessment categories are defined (see color scale and text for details). The percent of analyzed area classified in each category is listed in parentheses. Grid points in which the observed trend is consistent with (i.e., within the 5th to 95th percentile of) the CMIP5 all-forcing historical run ensemble trend distribution are identified with white stippling. Solid white regions have too sparse data coverage for the trend analysis. Gray regions in (c) have no detectable observed trend. Other color-shaded regions in (c) have various types of significant observed trends (some detectable), which are interpreted as summarized in the category legend.

serious (wrong sign) model–observation inconsistency, are more common in our results.

1) 1901–2010 TRENDS

Observed trend maps and assessments for annual mean precipitation over 1901–2010 (Fig. 3a) show a preponderance of increasing (wetting) trends, especially in the extratropics. While there is some broad-scale resemblance of the CMIP5 all-forcing ensemble mean trend pattern (Fig. 3b) to the observed trend pattern (Fig. 3a), the modeled pattern clearly shows greater coverage and prominence of decreasing (drying) trends than observations. The all-forcing ensemble and observed trends are statistically consistent (Fig. 3c) over

58% of area analyzed (white stippling on various colors). This includes some regions (gray shading with white stippling) where there are only weak observed trends that are nonetheless consistent with the all-forcing runs. The nonstippled color-shaded or nonstippled gray regions all denote significant discrepancies (42%). The summary assessment categories in Fig. 3c, as discussed above, indicate detectable observed trends over 29% of the analyzed global area (categories +1, +2, +3, -1, -2, -3). The areas we identify as having trends outside of the 5th to 95th percentile of the natural forcing and control runs (nongray shaded regions) are similar to those regions identified in earlier studies—but using different methods—as having

statistically significant trends based on observed data alone (e.g., [Hartmann et al. 2013](#)). Observed detectable drying trends (warm colors, categories -1 , -2 , -3 in [Fig. 3c](#)) are found over 9% of the analyzed area, but only 2% of the area has both detectable and consistent drying trends (category -2), compared with the all-forcing ensemble. In 7% of the area, the detectable drying is significantly stronger in the observations than in the models (category -3). We therefore infer that there is attributable anthropogenic drying over 9% (7% + 2%) of the analyzed area, including parts of the Mediterranean region, northern tropical Africa, small regions of Japan, the Maritime Continent, and southwest Australia, plus a few other scattered locations around the globe.

Detectable wetting (increases) trends that are at least partially attributable to anthropogenic forcing are found over 20% of the area, that is, 4% that is consistent (category $+2$) and 16% where the wetting trend is significantly larger than the all-forcing models (category $+3$). These areas with at least some inferred detectable anthropogenic wetting trends (categories $+2$, $+3$) include regions of northern Europe, eastern Europe/western Asia, parts of the central and northern coterminous United States and southern Canada, and parts of southern South America. Some smaller regions of inferred anthropogenic increase are also seen near Iceland, in Australia, and over scattered regions of the tropics, Asia and northern Canada.

The inferred human-induced increases (categories $+2$, $+3$) over southern South America are consistent with [Zhang et al.'s \(2016\)](#) finding for summertime (November–April) precipitation there; our 3-month seasonal attribution maps (supplemental material) suggest that attributable human influence is also present in the December–February (DJF), March–May (MAM), and September–November (SON) seasonal precipitation in that general region. Detectable anthropogenic precipitation decreases since the early 1900s have been previously reported over the Mediterranean region during winter ([Hoerling et al. 2012](#)) and over extreme southwestern Australia during March–August ([Delworth and Zeng 2014](#)), generally consistent with our (category -2 , -3) results. Our analysis also identifies a number of regions (category $+4$, green shading, 9% of area) with strong observed wetting, but where the model ensemble simulates the incorrect (drying) sign of change in the all-forcing runs. These nondetectable regions with striking differences between strong observed wetting and modeled drying trends notably include parts of the south central United States, small regions of India, northern Australia, and the Korean peninsula, as well as scattered locations in southern Africa, Europe, and southwestern Canada.

Seasonal versions of the precipitation assessment map in [Fig. 3c](#) are presented in the supplemental material. Notable features in these include pronounced anthropogenic increasing trends (category $+2$, $+3$) over parts of Europe/western Asia in various seasons excepting June–August (JJA), over parts of southern South America in all seasons except JJA, and over the eastern United States and Great Lakes region in SON.

2) 1951–2010 TRENDS

Annual precipitation trends for 1951–2010 ([Fig. 4](#)) represent a compromise between using a record as long as possible to better detect long-term trends versus concerns about data homogeneity in the earlier part of the record. A smaller percentage area with detectable trends (16%; categories -3 , -2 , $+2$, $+3$) is indicated than for 1901–2010 (29%). The area with observed trends consistent with the models (white stippling) increases to 72% for 1951–2010, compared with 58% for 1901–2010. Detectable anthropogenic increases (10%; categories $+2$, $+3$) are prominently indicated over regions of northern Europe, central Asia, the North American Great Lakes region, northern Australia, and southern South America. The high northern latitude moistening signal is broadly consistent with previous studies ([Min et al. 2008](#); [Wan et al. 2015](#)) attributing northern high-latitude moistening trends over this period partly to anthropogenic influence. There are also some regions of detectable anthropogenic drying (6% in total; categories -2 , -3), including over extreme western Africa and the Mediterranean region. Finally, there is a substantial area (6%; category -4) with significant drying trends, but where the CMIP5 all-forcing ensemble simulates the wrong sign of change (wetting)—notably, parts of tropical Africa/Sahel, but also scattered regions of Asia and other continents. The models simulate an observed drying trend in extreme western Africa, although it is weaker than observed; the observed drying in extreme western Africa is thus attributed in part to anthropogenic forcing. However, the CMIP5 multimodel ensemble simulates the wrong sign of change in northern tropical Africa, including in much of the Sahel region (prominent observed regional drying; category -4), which will be discussed more later. Other problem areas for models include the south central United States (similar to the 1901–2010 trends case) and parts of tropical South America, which are assessed (category $+4$) as having significant wetting trends in the observations, in contrast to simulated drying trend in the models—similar to inconsistencies seen for the 1901–2010 trends. Seasonal assessment maps (supplemental material) indicate that models have difficulty simulating observed drying trends (category -4) during winter

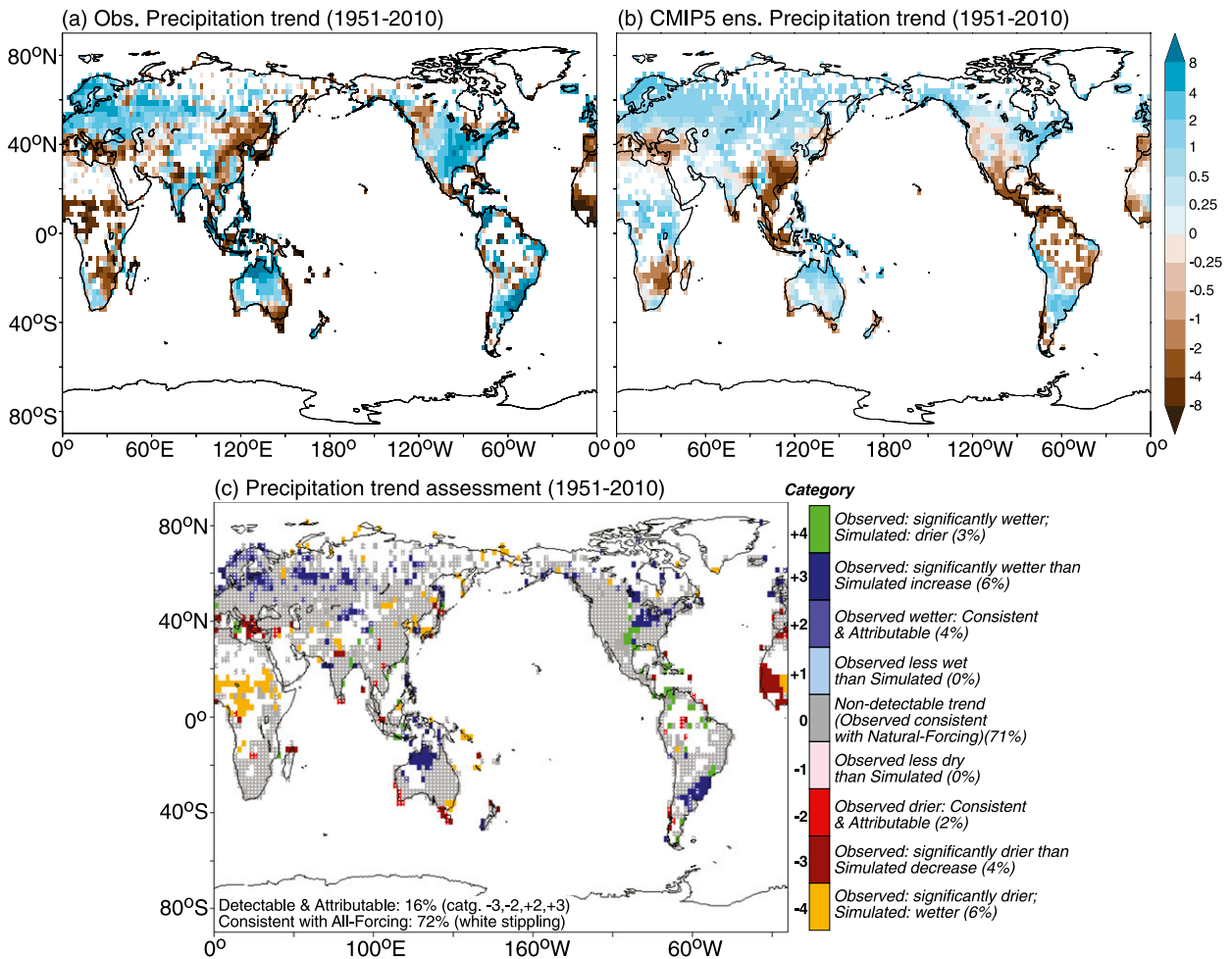


FIG. 4. As in Fig. 3, but for assessment of observed (GPCC) precipitation trends over 1951–2010.

over northeast Asia and parts of Canada, but also indicate attributable anthropogenic drying (categories -2 , -3) over extreme eastern Asia in September through November.

3) 1981–2010 TRENDS

Observed annual precipitation trends over the recent 30-yr period 1981–2010 (Fig. 5) have generally stronger magnitudes (in $\text{mm yr}^{-1} \text{decade}^{-1}$) than observed trends for 1901–2010 or 1951–2010. This is as expected since trends over a shorter period have stronger potential for occurrence of pronounced temporary trends due to internal variability alone. Only 9% of the analyzed area has detectable and attributable anthropogenic influence (categories -3 , -2 , $+2$, $+3$), which represents less than one-third of the fractional areal coverage of such trends for 1901–2010 (29%). Overall, the fraction of global analyzed land area with trends consistent with CMIP5 models (white stippling) is much larger (77%) than that for 1901–2010 (58%). Eight percent of the analyzed area

has attributable anthropogenic *increases* (categories $+2$, $+3$), including parts of eastern Europe, northern Asia, northern Australia/Maritime Continent, and tropical Africa. There are almost no regions with attributable drying trends (categories -2 , -3), but there are some regions with detectable observed drying trend where the CMIP5 ensemble simulates the wrong sign of change (wetting), including over parts of the Mediterranean, the Middle East, South Asia, East Asia, and southern South America (category -4 ; 6% of analyzed area in total). A smaller total area with significant wetting but where the models simulate a drying (category $+4$; 3% of area) occurs mainly over extreme western Africa, the Caribbean, and parts of Amazonia. Two minor regions of possible detectable anthropogenic drying (category -3) were inferred over the southwestern United States/Mexico and central South America; detectable drying was not found in these regions in the longer-term trend analyses. The southwestern U.S. drying detection, while a potentially

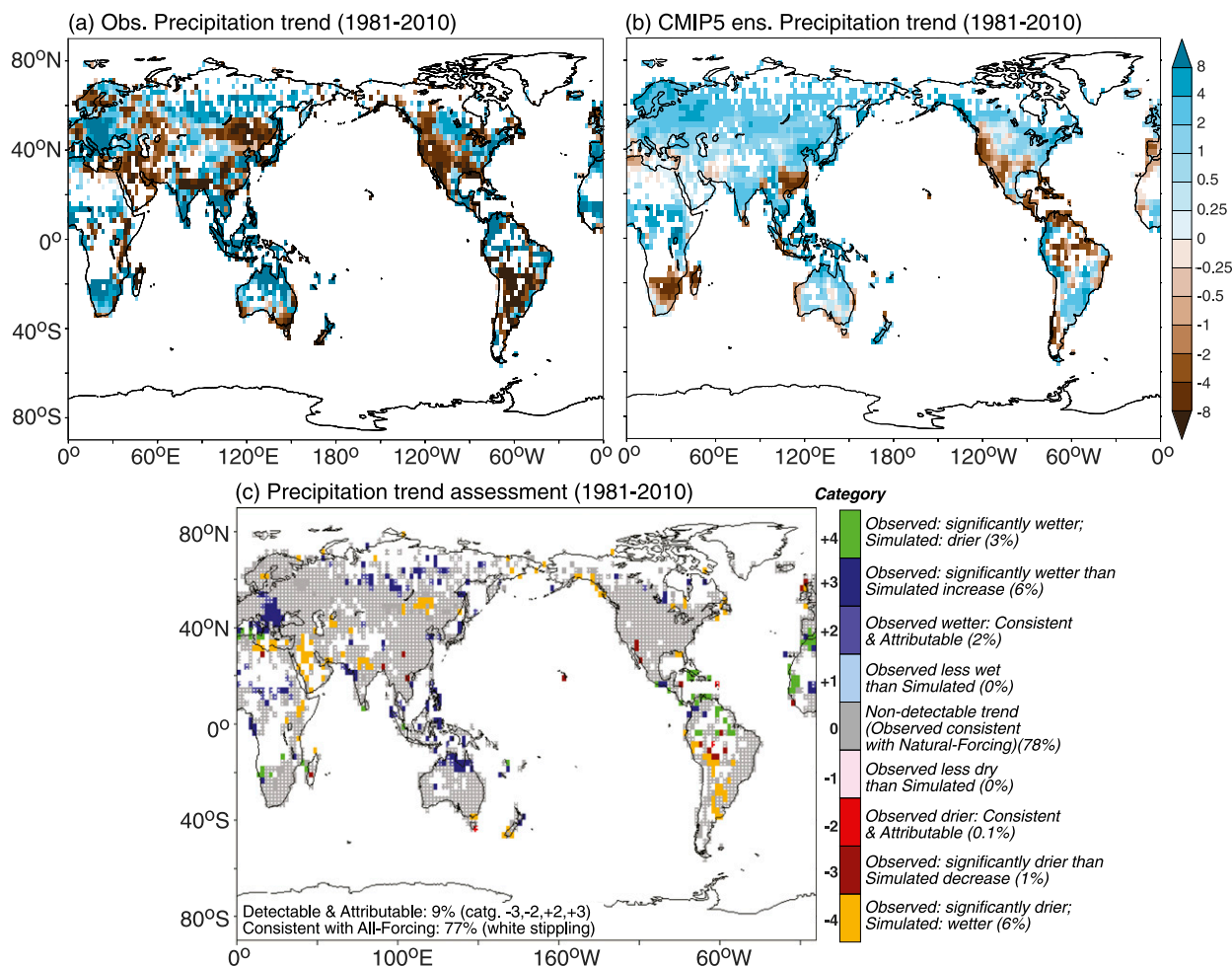


FIG. 5. As in Fig. 3, but for assessment of observed precipitation trends over 1981–2010.

important result, should be treated with caution as the detection occurs for a relatively small and isolated region, surrounded by a large region of nondetection. Seasonal trend assessment maps (supplemental material) suggest attributable anthropogenic wetting (categories +2, +3) over northern Australia and parts of eastern Asia in December–February and some scattered regions/seasons with attributable anthropogenic drying (categories –2, –3) in low latitudes.

In general, we find that for shorter time periods, the detectability of trends decreases while the consistency of observed trends with the all-forcing trend distributions increases. We interpret this general behavior as follows. We expect it to be easier for an observed trend to emerge as significant, compared to background internal variability, for a longer record, as the trend is then extending over multiple “cycles” of the internal variability. The increased consistency for shorter periods occurs as the trend components, which can be a source of discrepancy between models and observations, become

less important relative to the internal variability. Essentially, for very short periods, the models mainly need an adequate simulation of the internal variability in order to be consistent with observations. As discussed previously, excessive simulated internal variability, if present, also leads to a high bias in degree of consistency between historical run trends and observations because the wider the modeled distribution of trends (larger internal variability), the easier it is for an observed trend to lie within the modeled all-forcing distribution (as illustrated by Fig. 2, for example).

4) INDEPENDENT ASSESSMENT OF INDIVIDUAL MODELS

Our trend assessment has some dependence on assumptions in the methods used. The assumption of assessing the 10 models combined together as an ensemble is relaxed for Fig. 6, which depicts, at each grid point, the number of models out of 10 where an attributable anthropogenic increase or decrease is inferred

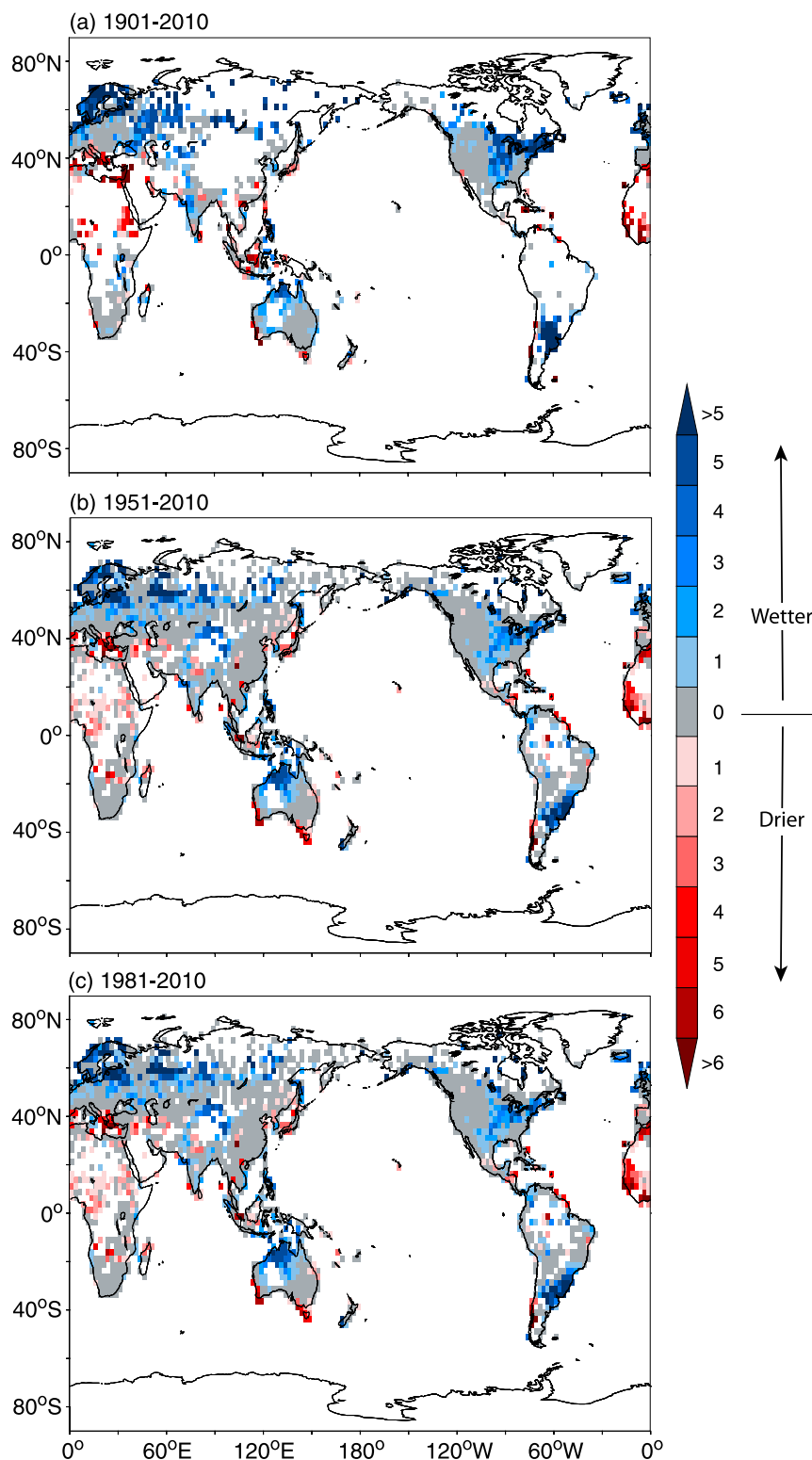


FIG. 6. Number of CMIP5 models out of 10 that are assessed as having a detectable and attributable anthropogenic wetting or drying tendency. The observed (GPCC) precipitation trends for (a) 1901–2010, (b) 1951–2010, and (c) 1981–2010 are analyzed. Detectable and attributable anthropogenic influence refers to categories +2 and +3 for increasing precipitation and categories –2 and –3 for decreasing precipitation (see text).

based on an independent assessment of each individual model. The results overall are generally consistent with our ensemble assessment results (Figs. 3–5), although the 1951–2010 trend assessment for the individual models indicates that some of the problem features identified in the ensemble results (e.g., incorrect sign of simulated precipitation change over parts of northern tropical Africa and the south central United States) are not present in at least some of the 10 individual models analyzed. For the 1901–2010 trends, the summary category assessment (analogous to Fig. 3c) is shown for each of the 10 individual CMIP5 models in Fig. 7. Here one can compare the individual models in how they simulate the observed trends over the central United States, Europe, and tropical Africa, for example. The percent of area with detectable and attributable trends ranges from 18% to 29%, compared with the average model distribution result (Fig. 3c) of 29%. Similarly, the percent of area consistent with observed trends ranges from 45% to 57% compared with the average model distribution result of 58%. Notable results include inconsistencies between some individual models and observations (typically where a model simulates drying while observations show significant wetting) with examples over Europe, the United States, Australia, and India. Trend maps for individual models (1901–2010 and 1951–2010 trend periods) are contained in the supplemental material.

5) SENSITIVITY TESTS

The trend assessment results in Fig. 3 for 1901–2010 trends are based on the GPCC dataset. Here we present an analysis of comparable results using an alternative dataset, the Climatic Research Unit CRU_t3.24.01 monthly precipitation dataset, to explore the robustness of our trend assessment to the use of a different observed dataset. Figure 8 shows the CRU version of the trend assessment shown for GPCC data in Fig. 3. The broad-scale findings are similar, with both datasets having much stronger wetting trends in the extratropics than modeled in the CMIP5 all-forcing runs. Although the areas included in the analysis are similar but not identical, as a first comparison we compare the fraction of areas for various assessment categories. The percent area where modeled trends are consistent with observed trends is 62% for CRU, compared with 58% for GPCC. The CRU data have 23% of area categorized as detectable and attributable wetting trends (categories +2, +3) compared with 20% for GPCC. For detectable and attributable drying trends (categories –2, –3), these percentages are 7% for CRU compared with 9% for GPCC. These results suggest that our main findings are relatively robust to the choice of GPCC versus CRU observed data.

Additional sensitivity tests in the supplemental material related to Figs. 3–5 show the following: 1) Similar detection and attribution results are obtained using 36 models, although in that case we are comparing observed trends with the all-forcing ensemble means and control runs, but not to natural-forcing runs, so the attribution is to external forcing only. 2) Using multimodel distributions of all-forcing and natural-forcing trends obtained by *aggregating* across the 10 models, instead of using distributions having the *average* characteristics of the 10 models, results in wider modeled trend distributions; this then results in less area with detectable changes but also less area with inconsistency between the all-forcing distribution and observations. 3) Assessment results for Standardized Precipitation Index trends are similar to those for precipitation trends, suggesting that effects of non-Gaussian behavior and any model biases in simulated precipitation variability have relatively minor impact on our overall results. The temporal behavior of precipitation monthly extremes is examined by plotting the time evolution of the percent of area where certain low or high SPI monthly thresholds are exceeded (see the supplemental material). That analysis finds little evidence that the monthly SPI dry extremes are becoming more prevalent at the global land scale, although there may be some increasing tendency in SPI wet extreme months.

c. Zonal mean trend comparisons

The zonal average of 1901–2010 trends in annual mean precipitation and SPI (Figs. 9a and 9e, respectively) reflects the model versus observed differences seen in the earlier trend map assessments. The observed zonal means for both precipitation and SPI are generally more positive than the multimodel (36 models; see section 2) ensemble mean, especially outside of the tropics, and are on the outer edge of the 36-model distribution of individual model all-forcing ensemble members. The observed zonal mean trends are positive (wetting trends) for most Southern Hemisphere latitudes and poleward of about 40°N, whereas modeled zonal mean trends over these latitude zones are much more evenly divided between drying trends (from 30°S to 40°N) and wetting trends (poleward of about 40°N). Zonal mean trends from the natural forcing-only CMIP5 experiments have only small magnitudes compared to observations or the all-forcing runs. Seasonal versions of the 1901–2010 zonal average trend plots are presented in the supplemental material. These show that the apparent multimodel bias of underpredicting the observed wetting trends in the extratropics is most pronounced for the fall season (SON) in the Northern Hemisphere and the spring and summer (SON and DJF) in the Southern

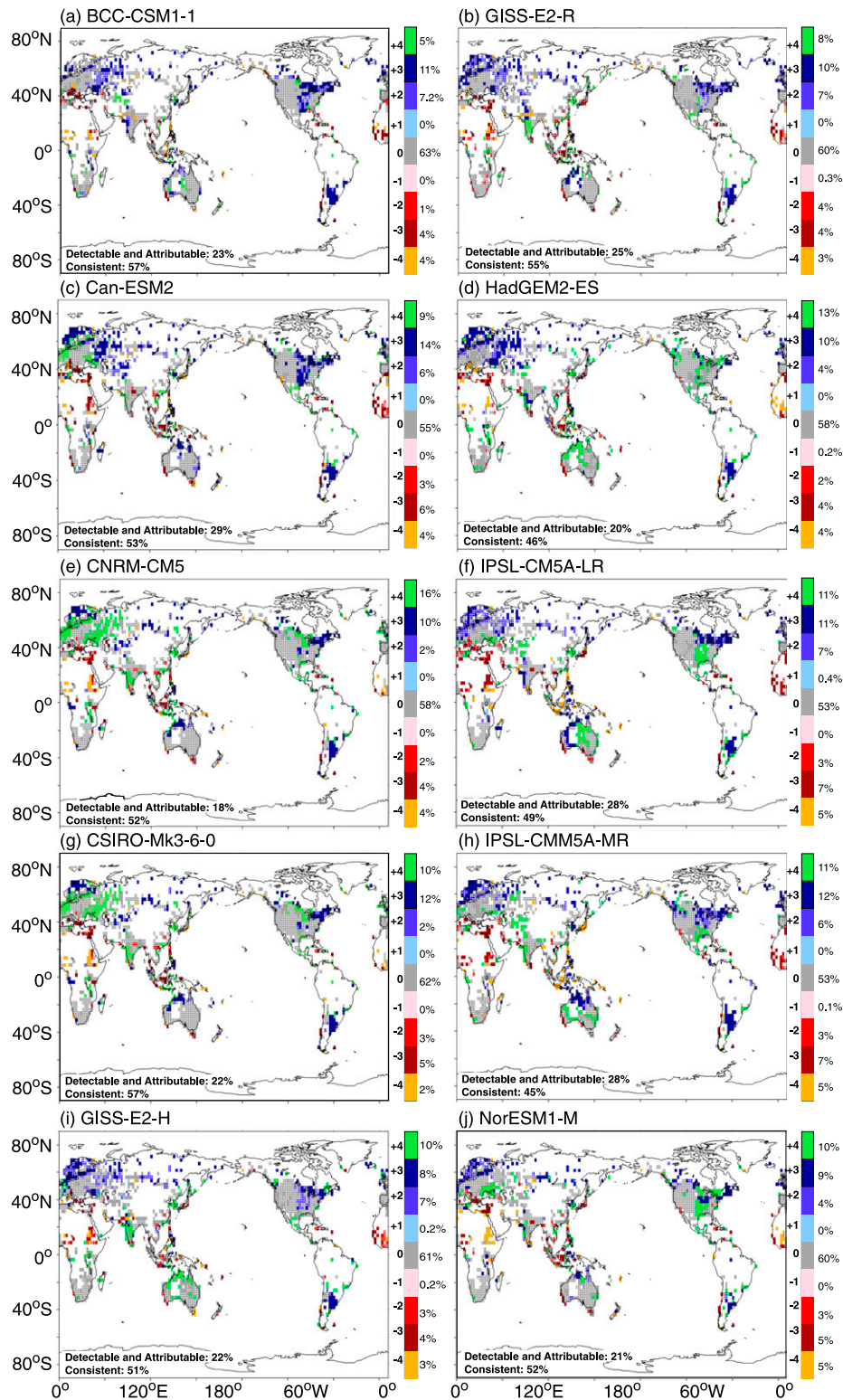


FIG. 7. Summary trend assessment for 1901–2010 precipitation trends as in Fig. 3c, but for each of 10 individual CMIP5 models. See legend of Fig. 3c and text for category (color) definitions.

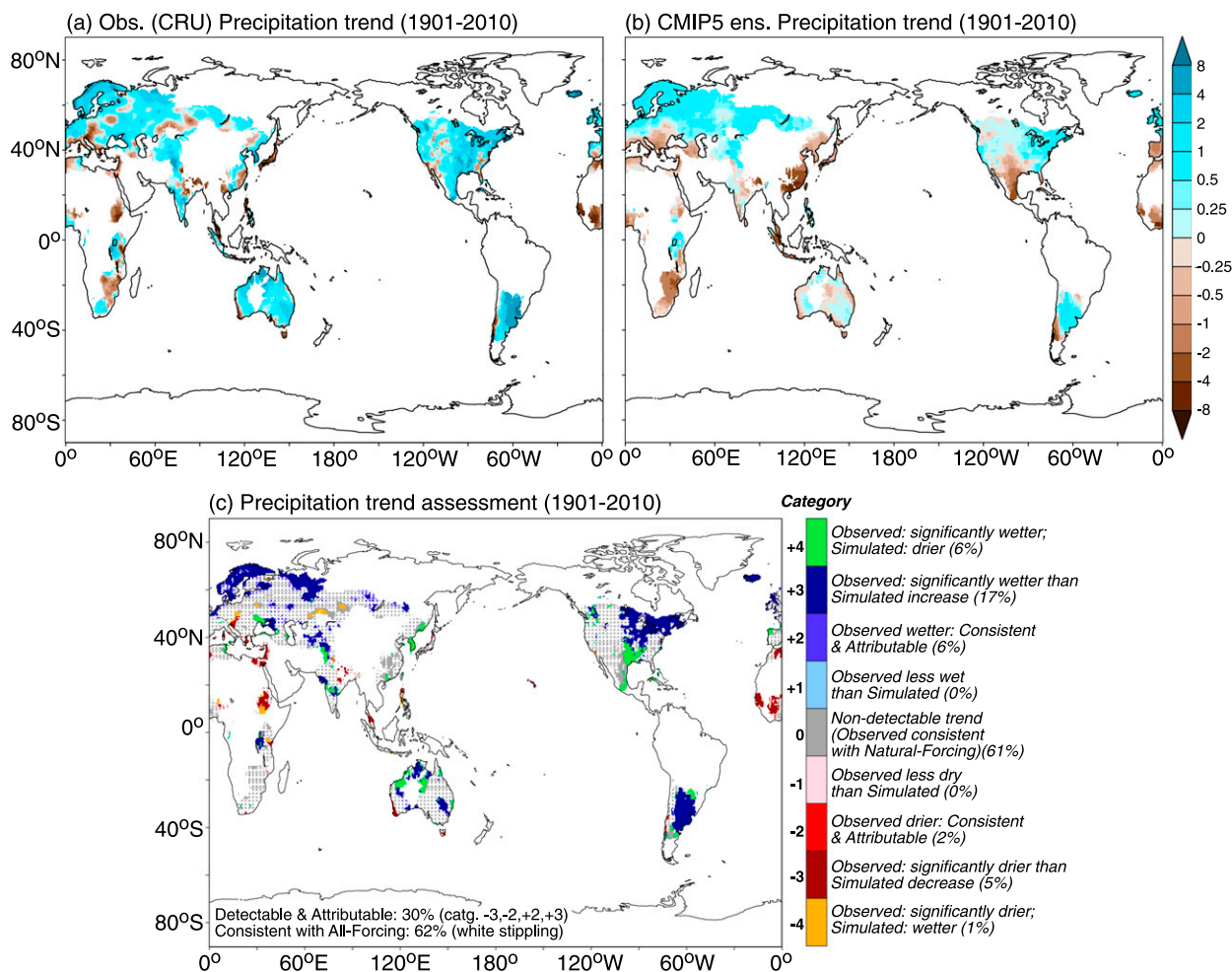


FIG. 8. As in Fig. 3, but assessment of observed precipitation trends (1901–2010) based on CRU observations rather than GPCP observations.

Hemisphere. The seasonal trend biases (1901–2010) are relatively more apparent for SPI trends than for precipitation trends. Zonal-mean annual-mean precipitation trends for 1951–2010 (Fig. 9b) do not indicate as systematic a low bias for models compared to observations as for the 1901–2010 trends, and show a larger spread of trends across the 36 models, as expected for the shorter trend period (note difference in vertical axis scale).

Since a low bias of CMIP5 historical run precipitation and SPI trends over 1901–2010, compared to observed trends, would be an important finding, some further exploration of this tentative finding is warranted. One approach is to attempt to reconcile the model–observation inconsistency by running a more observationally constrained model (e.g., specifying the observed time history of SSTs) to see if such a constrained model can better reproduce the observed precipitation trends. Therefore, we performed such an observationally

constrained set of historical simulations using the GFDL CM3 model’s atmospheric component (AM3) forced with observed SST and sea ice distributions over ocean regions, and also incorporating the same historical climate forcing agents as the all-forcing simulations (i.e., AMIP-type forced experiments). The zonal-mean ensemble-mean precipitation trend from these AM3 model runs (thick blue in Fig. 9c), is based on four ensemble members with each showing similar changes (thin blue curves). These results do not demonstrate appreciably closer agreement between model and observations in the extratropics than the comparable GFDL CM3 all-forcing ensemble mean (red curves). Also, both AM3 and CM3 model experiments notably fail to reproduce observed zonal mean trends over lower latitudes (5°S–30°N). In summary, our more constrained specified SST simulations do not support the idea that limitations in all-forcing simulations of the historical patterns of SST change are the main cause of a low bias

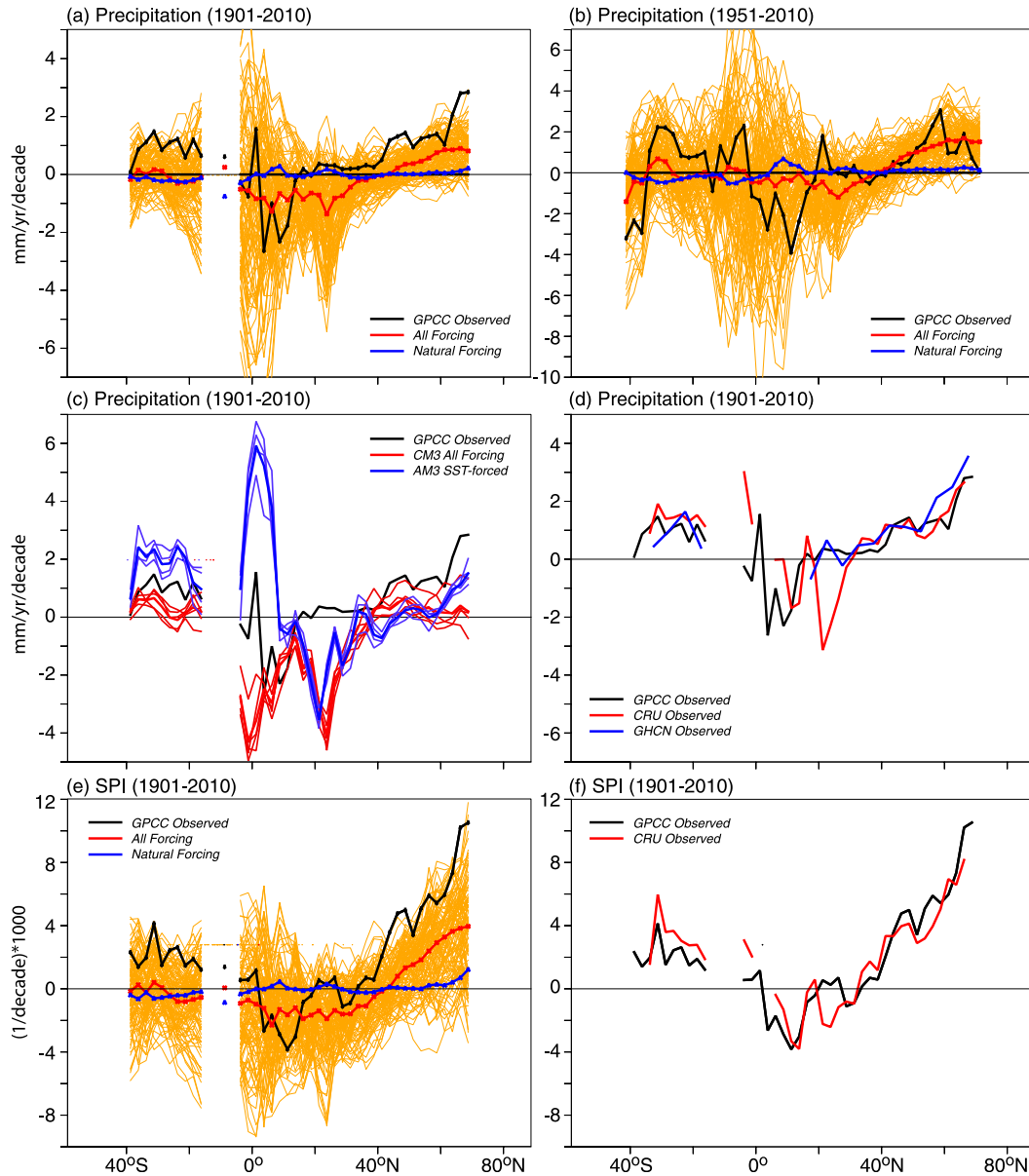


FIG. 9. Zonal averages of trends in observed and modeled precipitation metrics for 1901–2010 [in (a) and (c)–(f)] and 1951–2010 [in (b)]. (a),(b) Precipitation trends averaged over grid points with adequate data coverage for 1901–2010 and 1951–2010, respectively. Black curves are observed trends from GPCCC data; red (blue) curves are CMIP5 multimodel mean all-forcing (natural forcing) trends, computed as the mean across the 36 (10) CMIP5 models of the ensemble means of the individual models (see section 2); and orange curves are individual all-forcing experiment ensemble members. (c) GPCCC observed (black curve) vs GFDL CM3 all-forcing (red curves) and GFDL AM3 specified SST (blue curves) experiment trends; ensemble means are the thicker curves, and thinner curves are the 5 (4) CM3 (AM3) individual ensemble members using the GFDL CM3 coupled model or GFDL AM3 atmospheric model (section 2). (d) Observed precipitation trend comparison (1901–2010) based on GPCCC (black), CRU_ts3.24 (red), or GHCN (blue) data. Unit in (a)–(d) is $\text{mm yr}^{-1} \text{decade}^{-1}$. (e),(f) As in (a),(d), but for standardized precipitation index (SPI) trends (see the appendix). Unit: $\text{decade}^{-1} \times 1000$.

in historical run extratropical precipitation trends over the past century.

Another possible explanation for the apparent low (dry) bias of CMIP5 models in their century-scale

historical trends over land regions is that data problems and/or observational inhomogeneities may have created spurious observed trends; this will be discussed in the following section.

4. Data quality considerations

A principal result from our analysis of century-scale precipitation trends (1901–2010) is that observations have a greater increasing precipitation tendency than CMIP5 historical runs, particularly over a number of extratropical land regions having adequate observational coverage over the period (Figs. 3, 8, 9). This raises the possibility that CMIP5 modeled historical trends are systematically biased dry in these regions, which could have implications for future precipitation projections using these models. However, another possible explanation for the apparent dry trend bias of CMIP5 models is that data problems/observational inhomogeneities may have created spurious observed trends at the century scale. Owing to the potential importance of our biased trend finding, here we present additional discussion of data quality issues for the observed precipitation datasets used in this study.

We first assess whether the strong positive extratropical trends in the GPCC dataset are found in other observed precipitation datasets. While this was already indicated to be the case for the CRU observed data (Fig. 8), in Fig. 9d we compare the observed zonal means of the century-scale precipitation trends (1901–2010) of three observational datasets (GPCC, CRU, and GHCN; see section 2). The comparison shows that there are some discrepancies between datasets in this trend metric, especially in the tropics, where different spatial coverage is available in the different datasets (not shown). However, for the extratropics (poleward of about 30° latitude in both hemispheres) the general features of the zonal mean trends are very similar across all three observed datasets. Comparison of observed SPI trends (1901–2010) for the two observed datasets that have available climatologies (GPCC and CRU), as shown in Fig. 9f, indicates a result with somewhat better agreement between these datasets than for precipitation, even in the tropics.

Another aspect of our results that lends support to the notion of an extratropical dry trend bias in the CMIP5 models is the degree of spatial coherence of the 1901–2010 precipitation trends assessment shown in Figs. 3c and 8c, including widely separated isolated grid boxes that also show similar biases. If data inhomogeneities were the primary cause of model–observation disagreement on trends, we might expect to see a less spatially coherent pattern of these features.

Among the caveats to our finding is that the reliability of precipitation data for trend analysis is generally expected to be higher for more recent periods, implying that trends beginning in the mid-twentieth century or later are expected to be less prone to spurious trend components caused by data inhomogeneities than are

trends beginning in 1901. In that regard, a number of previous precipitation change studies have focused on trends over the second half of the twentieth century. Our zonal mean trend results (cf. Figs. 9a and 9b) indicate that the apparent dry trend bias of the models is more pronounced for the (less reliable) 1901–2010 trends than for the 1951–2010 trends. Using a completely independent data source, Milly et al. (2005) found only a slight tendency (not statistically significant) for observed twentieth-century trends in a related variable (runoff) to be larger than simulated in CMIP3 historical runs.

Becker et al. (2013) discuss the issue of potential inhomogeneities and their impact on trend analysis for the GPCC dataset. They note that stations with “obvious jumps” in their precipitation records were removed from the dataset at a preliminary stage. They also foresee a future version of their dataset based on a more complete assessment and adjustment for inhomogeneities; such as dataset is under development but not available at this writing. The current GPCC dataset includes information on the number of stations available at each grid point and month. We use this information to screen out gridpoint anomalies (from both observations and models) where the number of available stations for observations is zero. The temporal homogeneity of time series at individual grid points will be affected by a varying number of stations over time used to compute anomalies for a given grid box. We have not attempted to assess or adjust the GPCC or CRU dataset time series for discontinuities or changepoints. In the course of our analyses we have noted some aggregate time series metrics that display possible indications of inhomogeneity effects. An example of this is shown in the supplemental material (Fig. S10c), where we document in the GPCC dataset a relatively abrupt shift around 1920 in the time series of percent of analyzed global area having SPI values of less than -2 . Whether this feature is real or an artifact remains to be determined.

We conclude that although inhomogeneities in observed precipitation data may play a role in the observation–model discrepancies in century-scale trends, there is sufficient evidence, based on our preliminary cross-dataset comparisons and the spatial coherence of the trend assessments across large regions, to tentatively suggest that the identified trend bias is real and not simply an artifact of data homogeneity problems. Taking into account the above caveats, as well as the results shown that lend support to the notion of a dry extratropical trend bias since 1901 in the CMIP5 models, we conclude that our finding is, at this stage, a tentative result that merits further investigation—in particular, to determine whether or not it is robust to data quality issues.

5. Discussion and conclusions

As discussed in the previous section, a tentative result from our analysis is that CMIP5 historical all-forcing runs have a dry bias in century-scale (1901–2010) precipitation trends compared to observations, particularly over extratropical land regions having adequate observational coverage. Despite this tentative finding, modeled and observed trends (1901–2010) were assessed as statistically consistent for 58% of the analyzed area in the GPCP data. We found detectable and attributable trends over 29% of the area; for 20% (9%) of the analyzed area, a detectable *increase* (*decrease*) was attributable in part to anthropogenic forcing, according to the models. Similar results were obtained for observed trends based on the CRU precipitation data. Notable examples of areas with attributable anthropogenic increasing trends include northern extratropical land regions and parts of southern South America.

As noted in Knutson et al. (2013), century-scale observed temperature trends (e.g., 1901–2010) have a much more uniform spatial structure than those for the shorter period (1981–2010) and closely resemble the CMIP3 and CMIP5 all-forcing historical run ensemble trend pattern for 1901–2010. This suggests that the 1901–2010 temperature trends may represent the best currently available estimate of the pattern of response to greenhouse gas forcing, since shorter-term trends, such as over 1951–2010 or 1981–2010, have more complicated patterns that may reflect more of a mixture of signals, including internal climate variability, along with responses to volcanic and non-greenhouse gas forcing changes. Similarly by analogy, we speculate that the observed precipitation trends since 1901 (Fig. 2) may provide the best available estimate of the likely pattern of response of mean precipitation to greenhouse gas increases.

Some of the geographic areas identified as having attributable anthropogenic increases or decreases in precipitation are generally consistent with previous studies (e.g., Held et al. 2005; Min et al. 2008; Hoerling et al. 2012; Delworth and Zeng 2014; Wan et al. 2015; Zhang et al. 2016). The attribution of precipitation increases to anthropogenic forcing, as inferred here over some regions of the central and northern United States, southern Canada, and parts of Eurasia and northern Australia, has not been previously reported to our knowledge. We view more cautiously the possible anthropogenic decreases in precipitation inferred over relatively small regions, such as parts of the Maritime Continent, Japan, and the southwestern United States, and we would consider these findings more tentative than those covering larger regions.

A notable shortcoming in the multimodel ensemble simulations for the period 1951–2010 is their failure to

adequately simulate the observed detectable drying trend over this period in the Sahel region (Fig. 3). This feature is related to the observed pronounced decline in Sahel precipitation from the 1950s to 1980s associated with the Sahel drought, which can be reasonably well simulated in at least some climate models (Held et al. 2005). Nonetheless, the cause of the Sahel drought remains uncertain: a number of studies have emphasized an important role for natural variability such as the Atlantic multidecadal variability in causing the observed 1950s to 1980s precipitation decline (e.g., Hoerling et al. 2006; Ting et al. 2009, 2011; Mohino et al. 2011), while other studies point to a possible role for anthropogenic aerosol forcing (Rotstayn and Lohmann 2002; Ackerley et al. 2011) or various combinations of natural and anthropogenic forcings (including greenhouse gases and aerosols) (Held et al. 2005). In addition, some models can simulate the observed precipitation changes in response to SST forcing (Lu and Delworth 2005; Hoerling et al. 2006), although this does not identify the underlying cause for the drought. Some further relevant modeling issues in this region include the difficulties in simulating important aspects of the African monsoon circulation with climate models (e.g., Cook and Vizy 2006), and the widely varying modeled responses of Sahel-region precipitation to uniform SST warming depending on the convective parameterization used in a model (Hill et al. 2017). Given these substantial challenges, it may not be surprising that models have difficulties simulating the observed multidecadal precipitation trends in the region. While model-observed discrepancies related to the Sahel drought identified in the present study for CMIP5 all-forcing runs are significant for 1951–2010 trends, there is much less evidence for a significant discrepancy in this region if one considers instead trends over 1901–2010 or 1981–2010. For example, a century-scale decline, as in observations, is simulated in summertime Sahel rainfall in most CMIP5 model all-forcing runs (Biasutti 2013). Dong and Sutton (2015) conclude that the recent (since the 1980s) partial recovery of Sahel rains in observations can be largely captured by the HadGEM3-A model, primarily as a response to increasing greenhouse gas concentrations.

While our analysis can identify inconsistencies between observed and modeled trends, it cannot determine their causes. Potential causes can include 1) errors in observed trends due to data quality/homogeneity issues (discussed in section 4), 2) errors in specified historical forcings, 3) errors in the model simulated response to the specified forcings, and 4) underestimation of internal variability by the models. The observation–model discrepancies (in terms of precipitation trend maps or zonal mean trends) found for 1901–2010 trends

are not as compelling for trends over more recent periods (e.g., 1951–2010 or 1981–2010). The zonal mean trend discrepancies for 1901–2010 are not remedied by using a more constrained atmosphere-only model forced by observed SSTs and climate forcing agents. The observed zonal mean trends are shown to be fairly consistent across three observational datasets, suggesting that observational biases may not be the primary cause of the model–observation discrepancies.

In viewing the cases where models simulate the wrong sign of the change compared to observations, (e.g., in Figs. 3 and 4, see the orange and green shaded regions for the 1901–2010 and 1951–2010 trends), it is notable that these major discrepancies tend to be found in latitudes between about 30°S and 40°N, whereas at higher latitudes the models and observations tend to be in better agreement at least on the sign of change. The lower-latitude trend simulation problems could be a reflection of the greater difficulty at simulating precipitation responses to forcing in general at lower latitudes as noted by Stevens and Bony (2013). They posit that simulating the coupling between moist processes and circulation, and therefore regional changes, is especially challenging in lower latitudes where these processes are particularly dependent on unresolved, parameterized moist convection and cloud formation. In any case, further work will be needed to better understand the relative contributions of various factors to the model–observation trend discrepancies.

The tendency for models to underpredict the observed precipitation increases in the extratropical latitudes since 1901 and the relatively small area with detectable decreasing trends compared to increasing trends in observations are particularly important tentative findings with possible implications for twenty-first-century projections of drought and flood occurrence from the CMIP5 models. If a similar tendency for a low precipitation trend bias applies also in future model projections in these regions, the CMIP5 models might be expected to overestimate future drought risk, while underestimating future flood risk, all other factors assumed equal, at least over the regions covered by our analysis for 1901–2010. Such a model bias could also affect emergence time estimates for detectable trends (e.g., Mahlstein et al. 2012). Given the potential importance of future drought and flood changes for climate impacts, our tentative findings emphasize the need to better understand whether the (larger than modeled) observed precipitation trends since 1901 represent actual climate changes or data quality artifacts, and to conduct further assessments comparing modeled and observed precipitation trends.

Acknowledgments. We acknowledge PCMDI and the CMIP5 modeling groups for model data, and the Global

Precipitation Climatology Centre, Climatic Research Unit, and NOAA National Centers for Environmental Information for observational data. GPCP precipitation data are provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at <http://www.esrl.noaa.gov/psd/>.

APPENDIX

Further Methodology Details

a. CMIP5 models used

The following are 10 CMIP5 models that were used for the main climate change assessment figures in the study. Listed are the model names, with the number of all-forcing ensemble members, natural-forcing ensemble members, and control run length (in years) used listed in parentheses for each model: BCC-CSM1.1 (3; 1; 500), CanESM2 (5; 5; 995), CNRM-CM5 (10; 6; 850), CSIRO-Mk3.6.0 (10; 5; 500), IPSL-CM5A-LR (6; 3; 1000), IPSL-CM5A-MR (3; 3; 300), HadGEM2-ES (4; 4; 576), GISS-E2-H (6; 5; 540), GISS-E2-R (6; 5; 525), and NorESM1-M (3; 1; 500) (For expansions of model names, see <http://www.ametsoc.org/PubsAcronymList.>)

The 36 CMIP5 models used for the zonal average trend plots (all-forcing runs) and in the external forcing-only assessments in supplemental material are listed here, along with the number of all-forcing ensemble members and control run length used (in years) for each model: BCC-CSM1.1 (3; 500), BNU-ESM (1; 558), CanESM2 (5; 995), CMCC-CM (1; 330), CMCC-CMS (1; 500), CNRM-CM5 (10; 850), ACCESS1.0 (2; 500), ACCESS1.3 (3; 500), CSIRO-Mk3.6.0 (10; 500), FIO-ESM (3; 800), EC-EARTH (3; 450), INM-CM4 (1; 500), IPSL-CM5A-LR (6; 1000), IPSL-CM5A-MR (3; 300), IPSL-CM5B-LR (1; 300), MIROC5 (5; 670), MIROC-ESM-CHEM (1; 255), MIROC-ESM (3; 630), HadGEM2-CC (1; 240), HadGEM2-ES (4; 576), MPI-ESM-LR (3; 1000), MPI-ESM-MR (3; 1000), MRI-CGCM3 (5; 250), GISS-E2-H-CC (1; 250), GISS-E2-H (6; 540), GISS-E2-R-CC (1; 250), GISS-E2-R (6; 525), CCSM4 (6; 1050), NorESM1-ME (1; 250), NorESM1-M (3; 500), HadGEM2-AO (1; 700), GFDL-CM3 (5; 500), GFDL-ESM2G (1; 500), GFDL-ESM2M (1; 500), CESM1-BGC (1; 500), and CESM1-CAM5 (3; 320).

b. Data masking and determining sufficient coverage for trend calculations

Model data are regridded to the observed grid, and locations and times where observed data are not available are set to missing in the model data fields. The following procedure, adopted from Hartmann et al.

(2013), was used to determine if sufficient observational coverage in time exists for a trend analysis at a particular grid point: both the first 10% and the last 10% of an observed time series must have at least 20% data availability, and the entire record must have at least 70% data availability. The missing data summary mask developed for annual means analysis was also applied to the trend results for the individual seasons (see the supplemental material). A monthly anomaly was assumed missing for the GPCC data if there were no stations available in that grid box for that month. For the CRU data, we used a stricter criterion (requiring four stations to be available), which resulted in similar degree of available coverage to facilitate comparison of trend results between the GPCC and CRU datasets for our sensitivity analysis.

c. Transformation of precipitation into a standardized precipitation index

As a sensitivity test for our precipitation trend analyses, the observed and modeled precipitation monthly anomalies are transformed into Standardized Precipitation Index (SPI) anomalies. The SPI is a transformed precipitation metric used for monitoring meteorological drought (Hayes et al. 1999). Some trend category results are presented in the supplemental material using SPI rather than precipitation. Our implementation of SPI uses a two-parameter gamma distribution, and we focus on time scales of 3 months for seasonal and 12 months for annual analysis. Further computational details are found at <http://ccc.atmos.colostate.edu/pub/spi.pdf>, and a sample program for computing SPI is found at <http://drought.unl.edu/monitoringtools/downloadablesipiprogram.aspx>. Each transformed series has an approximately Gaussian distribution with a standard deviation of unity. Thus the trend maps have units of (dimensionless) standard deviations per unit time as opposed to precipitation rates per unit time. This procedure has the advantage of strongly reducing non-Gaussian characteristics of the time series, which can confound statistical tests. Further, the normalization to unit standard deviation of both observed and modeled time series prevents differences in simulated and observed variance from influencing the trend assessments. Along similar lines, Noake et al. (2012) found that discrepancies between observed and simulated precipitation trends were smaller when comparing percent changes in climatology, as opposed to absolute changes. SPI was not computed for the GHCN dataset due to the lack of an available precipitation climatology for that dataset.

REFERENCES

- Ackerley, D., B. B. Booth, S. H. E. Knight, E. J. Highwood, D. J. Frame, M. R. Allen, and D. P. Rowell, 2011: Sensitivity of twentieth-century Sahel rainfall to sulfate aerosol and CO₂ forcing. *J. Climate*, **24**, 4999–5014, <https://doi.org/10.1175/JCLI-D-11-00019.1>.
- Anderson, B. T., D. J. Gianotti, and G. D. Salvucci, 2015: Detectability of historical trends in station-based precipitation characteristics over the continental United States. *J. Geophys. Res. Atmos.*, **120**, 4842–4859, <https://doi.org/10.1002/2014JD022960>.
- Becker, A., P. Finger, A. Meyer-Christoffer, B. Rudolf, K. Schamm, U. Schneider, and M. Ziese, 2013: A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. *Earth Syst. Sci. Data*, **5**, 71–99, <https://doi.org/10.5194/essd-5-71-2013>.
- 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>.
- Bhend, J., and P. Whetton, 2013: Consistency of simulated and observed regional changes in temperature, sea level pressure, and precipitation. *Climatic Change*, **118**, 799–810, <https://doi.org/10.1007/s10584-012-0691-2>.
- Biasutti, M., 2013: Forced Sahel rainfall trends in the CMIP5 archive. *J. Geophys. Res. Atmos.*, **118**, 1613–1623, <https://doi.org/10.1002/jgrd.50206>.
- Bindoff, N. L., and Coauthors, 2014: Detection and attribution of climate change: From global to regional. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 867–952.
- 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, K. H., and E. K. Vizy, 2006: Coupled model simulations of the West African monsoon system: Twentieth- and twenty-first-century simulations. *J. Climate*, **19**, 3681–3703, <https://doi.org/10.1175/JCLI3814.1>.
- Delworth, T. L., and F. Zeng, 2014: Regional rainfall decline in Australia attributed to anthropogenic greenhouse gases and ozone levels. *Nat. Geosci.*, **7**, 583–587, <https://doi.org/10.1038/ngeo2201>.
- Dittus, A. J., D. J. Karoly, S. C. Lewis, L. V. Alexander, and M. G. Donat, 2016: A multiregion model evaluation and attribution study of historical changes in the area affected by temperature and precipitation extremes. *J. Climate*, **29**, 8285–8299, <https://doi.org/10.1175/JCLI-D-16-0164.1>.
- Dong, B., and R. Sutton, 2015: Dominant role of greenhouse-gas forcing in the recovery of Sahel rainfall. *Nat. Climate Change*, **5**, 757–760, <https://doi.org/10.1038/nclimate2664>.
- Fischer, E. M., and R. Knutti, 2016: Observed heavy precipitation increase confirms theory and early models. *Nat. Climate Change*, **6**, 986–992, <https://doi.org/10.1038/nclimate3110>.
- Fyfe, J. C., N. P. Gillett, and G. J. Marshall, 2012: Human influence on extratropical Southern Hemisphere summer precipitation. *Geophys. Res. Lett.*, **39**, L23711, <https://doi.org/10.1029/2012GL054199>.

- Greve, P., B. Orlowsky, B. Mueller, J. Sheffield, M. Reichstein, and S. I. Seneviratne, 2014: Global assessment of trends in wetting and drying over land. *Nat. Geosci.*, **7**, 716–721, <https://doi.org/10.1038/ngeo2247>.
- 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>.
- Hartmann, D. L., and Coauthors, 2013: Observations: Atmosphere and surface. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 159–254.
- Hayes, M. J., M. D. Svoboda, D. A. Wilhite, and O. V. Vanyarko, 1999: Monitoring the 1996 drought using the standardized precipitation index. *Bull. Amer. Meteor. Soc.*, **80**, 429–438, [https://doi.org/10.1175/1520-0477\(1999\)080<0429:MTDUTS>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<0429:MTDUTS>2.0.CO;2).
- Held, I. M., T. L. Delworth, J. Lu, K. L. Findell, and T. R. Knutson, 2005: Simulation of Sahel drought in the 20th and 21st centuries. *Proc. Natl. Acad. Sci. USA*, **102**, 17 891–17 896, <https://doi.org/10.1073/pnas.0509057102>.
- Hill, S. A., Y. Ming, I. M. Held, and M. Zhao, 2017: A moist static energy budget-based analysis of the Sahel rainfall response to uniform oceanic warming. *J. Climate*, **30**, 5637–5660, <https://doi.org/10.1175/JCLI-D-16-0785.1>.
- Hoerling, M., J. Hurrell, J. Eischeid, and A. Phillips, 2006: Detection and attribution of twentieth-century northern and southern African rainfall change. *J. Climate*, **19**, 3989–4008, <https://doi.org/10.1175/JCLI3842.1>.
- , J. Eischeid, J. Perlwitz, X. Quan, T. Zhang, and P. Pegion, 2012: On the increased frequency of Mediterranean drought. *J. Climate*, **25**, 2146–2161, <https://doi.org/10.1175/JCLI-D-11-00296.1>.
- Jakob, C., 2010: Accelerating progress in global atmospheric model development through improved parameterizations: Challenges, opportunities, and strategies. *Bull. Amer. Meteor. Soc.*, **91**, 869–875, <https://doi.org/10.1175/2009BAMS2898.1>.
- , 2014: Going back to basics. *Nat. Climate Change*, **4**, 1042–1045, <https://doi.org/10.1038/nclimate2445>.
- Kang, S. M., L. M. Polvani, J. C. Fyfe, and M. Sigmond, 2011: Impact of polar ozone depletion on subtropical precipitation. *Science*, **332**, 951–954, <https://doi.org/10.1126/science.1202131>.
- Knutson, T. R., F. Zeng, and A. T. Wittenberg, 2013: Multimodel assessment of regional surface temperature trends: CMIP3 and CMIP5 twentieth century simulations. *J. Climate*, **26**, 8709–8743, <https://doi.org/10.1175/JCLI-D-12-00567.1>.
- , —, and —, 2014: Seasonal and annual mean precipitation extremes occurring during 2013: A U.S. focused analysis [in “Explaining Extremes of 2013 from a Climate Perspective”]. *Bull. Amer. Meteor. Soc.*, **95**(9), S19–S23, <https://doi.org/10.1175/1520-0477-95.9.S1.1>.
- Kumar, S., V. Merwade, J. L. Kinter, and D. Niyogi, 2013: Evaluation of temperature and precipitation trends using long-term persistence in CMIP5 twentieth-century climate simulations. *J. Climate*, **26**, 4168–4185, <https://doi.org/10.1175/JCLI-D-12-00259.1>.
- Lu, J., and T. L. Delworth, 2005: Oceanic forcing of the late 20th century Sahel drought. *Geophys. Res. Lett.*, **32**, L22706, <https://doi.org/10.1029/2005GL023316>.
- Mahlstein, I., R. W. Portmann, J. S. Daniel, S. Solomon, and R. Knutti, 2012: Perceptible changes in regional precipitation in a future climate. *Geophys. Res. Lett.*, **39**, L05701, <https://doi.org/10.1029/2011GL050738>.
- Marvel, K., and C. Bonfils, 2013: Identifying external influences on global precipitation. *Proc. Natl. Acad. Sci. USA*, **110**, 19 301–19 306, <https://doi.org/10.1073/pnas.1314382110>.
- Milly, P. C. D., K. A. Dunne, and A. V. Vecchia, 2005: Global pattern of trends in streamflow and water availability in a changing climate. *Nature*, **438**, 347–350, <https://doi.org/10.1038/nature04312>.
- Min, S.-K., X. Zhang, and F. Zwiers, 2008: Human-induced Arctic moistening. *Science*, **320**, 518–520, <https://doi.org/10.1126/science.1153468>.
- , —, F. W. Zwiers, and G. C. Hegerl, 2011: Human contribution to more-intense precipitation extremes. *Nature*, **470**, 378–381, <https://doi.org/10.1038/nature09763>.
- Mohino, E., S. Janicot, and J. Bader, 2011: Sahel rainfall and decadal to multi-decadal sea surface temperature variability. *Climate Dyn.*, **37**, 419–440, <https://doi.org/10.1007/s00382-010-0867-2>.
- Narisma, G., J. A. Foley, R. Licker, and N. Ramakutty, 2007: Abrupt changes in rainfall during the twentieth century. *Geophys. Res. Lett.*, **34**, L06710, <https://doi.org/10.1029/2006GL028628>.
- Noake, K., D. Polson, G. Hegerl, and X. Zhang, 2012: Changes in seasonal land precipitation during the latter twentieth-century. *Geophys. Res. Lett.*, **39**, L03706, <https://doi.org/10.1029/2011GL050405>.
- Polson, D., G. C. Hegerl, X. Zhang, and T. J. Osborn, 2013: Causes of robust seasonal land precipitation changes. *J. Climate*, **26**, 6679–6697, <https://doi.org/10.1175/JCLI-D-12-00474.1>.
- Roderick, M. L., P. Greve, and G. D. Farquhar, 2015: On the assessment of aridity with changes in atmospheric CO₂. *Water Resour. Res.*, **51**, 5450–5463, <https://doi.org/10.1002/2015WR017031>.
- Rotstayn, L., and U. Lohmann, 2002: Tropical rainfall trends and the indirect aerosol effect. *J. Climate*, **15**, 2103–2116, [https://doi.org/10.1175/1520-0442\(2002\)015<2103:TRTATI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<2103:TRTATI>2.0.CO;2).
- Sarojini, B. B., P. A. Stott, E. Black, and D. Polson, 2012: Fingerprints of changes in annual and seasonal precipitation from CMIP5 models over land and ocean. *Geophys. Res. Lett.*, **39**, L21706, <https://doi.org/10.1029/2012GL053373>.
- , —, and —, 2016: Detection and attribution of human influence on regional precipitation. *Nat. Climate Change*, **6**, 669–675, <https://doi.org/10.1038/nclimate2976>.
- Stevens, B., and S. Bony, 2013: What are climate models missing? *Science*, **340**, 1053–1054, <https://doi.org/10.1126/science.1237554>.
- Taylor, C. M., and Coauthors, 2017: Frequency of extreme Sahelian storms tripled since 1982 in satellite observations. *Nature*, **544**, 475–478, <https://doi.org/10.1038/nature22069>.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experimental design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Terray, L., L. Corre, S. Cravatte, T. Delcroix, G. Reverdin, and A. Ribes, 2012: Near-surface salinity as nature’s rain gauge to detect human influence on the tropical water cycle. *J. Climate*, **25**, 958–977, <https://doi.org/10.1175/JCLI-D-10-05025.1>.
- Ting, M., Y. Kushnir, R. Seager, and C. Li, 2009: Forced and internal twentieth-century SST trends in the North Atlantic. *J. Climate*, **22**, 1469–1481, <https://doi.org/10.1175/2008JCLI2561.1>.
- , —, —, and —, 2011: Robust features of Atlantic multi-decadal variability and its climate impacts. *Geophys. Res. Lett.*, **38**, L17705, <https://doi.org/10.1029/2011GL048712>.
- Trenberth, K. E., and Coauthors, 2007: Observations: Surface and atmospheric climate change. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 235–336.
- van Oldenborgh, G. J., F. J. Doblas Reyes, S. S. Drijfhout, and E. Hawkins, 2013: Reliability of regional climate model trends. *Environ. Res. Lett.*, **8**, 014055, <https://doi.org/10.1088/1748-9326/8/1/014055>.

- Vose, R. S., R. L. Schmoyer, T. C. Peterson, P. M. Steurer, R. R. Heim, T. R. Karl, and J. K. Eischeid, 1992: The Global Historical Climatology Network: Long-term monthly temperature, precipitation, sea level pressure, and station pressure data. Oak Ridge National Laboratory Environmental Sciences Division Publ. 3912, 324 pp.
- Wan, H., X. Zhang, F. Zwiers, and S.-K. Min, 2015: Attributing northern high-latitude precipitation change over the period 1966–2005 to human influence. *Climate Dyn.*, **45**, 1713–1726, <https://doi.org/10.1007/s00382-014-2423-y>.
- Westra, S., L. V. Alexander, and F. W. Zwiers, 2013: Global increasing trends in annual maximum daily precipitation. *J. Climate*, **26**, 3904–3918, <https://doi.org/10.1175/JCLI-D-12-00502.1>.
- Zhang, H., T. L. Delworth, F. Zeng, G. Vecchi, K. Paffendorf, and L. Jia, 2016: Detection, attribution, and projection of regional rainfall changes on (multi-)decadal time scales: A focus on southeastern South America. *J. Climate*, **29**, 8515–8534, <https://doi.org/10.1175/JCLI-D-16-0287.1>.
- Zhang, X., F. W. Zwiers, G. C. Hegerl, F. H. Lambert, N. P. Gillett, S. Solomon, P. A. Stott, and T. Nozawa, 2007: Detection of human influence on twentieth-century precipitation trends. *Nature*, **448**, 461–465, <https://doi.org/10.1038/nature06025>.
- , H. Wan, F. W. Zwiers, G. C. Hegerl, and S.-K. Min, 2013: Attributing intensification of precipitation extremes to human influence. *Geophys. Res. Lett.*, **40**, 5252–5257, <https://doi.org/10.1002/grl.51010>.