SUPPLEMENTARY MATERIAL FOR:

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

Thomas R. Knutson* and Fanrong Zeng

Geophysical Fluid Dynamics Laboratory / NOAA, Princeton, NJ 08540

Submitted to: *Journal of Climate*

Revised Version: Jan. 30, 2018 3:10 pm

Email: Tom.Knutson@noaa.gov

Supplementary Note 1. Comparison of low-frequency variability of modeled and observed precipitation. In this supplementary note, we compare the standard deviations of low-frequency (> 10 yr low-pass filtered) precipitation time series between models and observations for individual CMIP5 models based on standard deviation difference maps (Supplementary Figure 1). The observed internal variability is estimated by subtracting the ensemble mean of each model's All-Forcing experiment from observations to create an estimated internal variability residual. In the individual model maps, more regions indicate larger standard deviations in the models than in the observed residuals. Several models simulate less variability than the observed residual estimate over parts of northern Australia and northern India/Himalayas—a result which should be considered in interpreting detection/attribution results from these regions.

Supplementary Note 2. Seasonal analysis of precipitation trends. The main focus of our study is on annual mean trends. In order to include a seasonal perspective, we show seasonal versions of the trend assessments in Supplementary Figures 2-4. The main features found in the annual results are present, although at times in muted form, in the seasonal analyses. For the 1901-2010 trends, comparing the annual results in Fig. 3 with Supplementary Figure 2, there is a greater fraction of area with detectable and attributable trends for the annual means than for any of the individual seasons. Conversely, there is a greater fraction of area with observed trends consistent with the All-Forcing runs for the seasonal results than for the annual means. This same feature is found also for the 1951-2010 and 1981-2010 trend comparisons (annual vs. seasonal). We interpret these result as due to the long-term trends of seasonal means being similar in magnitude to those of annual means, while the interannual variability of seasonal precipitation means is typically larger than that of annual precipitation means (figure not shown).

This tends to make the signal to noise ratio for trends smaller for the case of seasonal means. While this makes climate change detection more difficult, it also makes achieving consistency between modeled and observed easier for seasonal means than for annual means because of the larger relative variability of the simulated seasonal means.

Supplementary Note 3. Assessment of external forcing influence on trends. The main focus of our study is to assess the influence of anthropogenic forcing on observed trends in precipitation. However, a related question is whether an influence of external forcing in general (i.e., anthropogenic plus natural forcing combined) can be detected in observations. One way of addressing this question is to assess observed trends for consistency with trends from model control runs, which by design have no external forcing changes from year to year. Where observed trends are inconsistent with the control run trends, we can infer some external forcing influence.

One advantage of such an external forcing assessment is that we can make use of a much larger set of CMIP5 models than for the anthropogenic forcing assessment. We have obtained data for 36 models having All-Forcing simulations and extended these runs through 2010 as needed using the RCP4.5 forcing runs. The assessment results for 1901-2010, 1951-2010, and 1981-2010 precipitation trends are summarized in Supplementary Figures 5-7. The results indicate that the fraction of area with detectable and attributable trends is slightly larger in the 36-model external forcing assessment than in the 10-model assessment (which compares observed trends to natural forcing distributions). The fractional area with trends consistent with All-Forcing is also slightly higher in the 36 model assessment than in the 10-model assessment. These differences apparently arise from: differences in trend distributions between the 36-model control run sample

vs. the 10-model control run plus natural forcing ensemble mean sample; and differences in the amount of bias between ensemble mean All-Forcing trends vs. observations for the 36-model and 10-model ensemble means.

Supplementary Note 4. Comparison of trend assessments using two alternative methods.

The methodology for our assessments makes certain choices and assumptions, some of which we explore further in this note. One choice which arises is how to create a multi-model distribution of simulated natural variability-caused trends for comparison to observations. The two approaches that we explore in this paper are the *average* model distribution approach and the *aggregate* model distribution approach. For the *average* model distribution approach, which is the main method used in our study, we seek to construct an average model distribution of trends, which we define as a distribution having as its mean value the multi-model All Forcing or Natural Forcing ensemble mean (with each individual model weighted equally in creating this average) and having a 5^{th} to 95^{th} percentile range about this mean, where the 5^{th} and 95^{th} percentiles are based on the ensemble means of the 5^{th} and 95^{th} percentiles of the individual model control runs. The average multi-model distributions (All-Forcing and Natural Forcing) created using this approach are analogous to a single model's distributions but have the average characteristics of the available models.

In contrast, for the *aggregate* distribution approach, which we test in this supplementary note, we combine samples of trends from each of the individual models into one large aggregate distribution and then compute the 5^{th} and 95^{th} percentiles of this aggregate distribution (for All-Forcing runs and separately for Natural-Forcing runs). The All-Forcing sample of trends from

each model is created by combining the ensemble-mean trend from that model's All-Forcing runs with samples of internal variability trends from that model's control run. Similarly the Natural-Forcing sample of trends for each given model is created by combining the ensemble-mean trend from that model's Natural-Forcing runs with samples of internal variability from that model's control run. Note that the aggregate method will generally result in wider Natural Forcing and All Forcing multi-model trend distributions (larger range between 5^{th} and 95^{th} percentiles) as the aggregate distributions contain some spread associated with different mean forced responses of the different models. The internal variability spread can also be larger, as it may reflect the internal variability of higher-variability models within the larger set.

Supplementary Figure 8 compares assessment results using the average and aggregate distribution methods for the 1901-2010 trends. Using the alternative aggregate distribution method (panel b) results in a smaller fraction of grid points with detectable and attributable trends, and a larger fraction with trends that are consistent with All-Forcing runs—at least as compared to the average distribution results (repeated from Fig. 3 c in panel a). This is as expected, since the modeled trend distributions are wider for the aggregate approach.

Further comparing the aggregate distribution results to the average model distribution results, we find some decrease in the extreme discrepancy categories (-4 and +4) in the aggregate distribution results. This is as expected, since these categories, in addition to requiring the All-Forcing simulation to have the wrong sign of change, compared to the observed trend, also requires a significant trend in the observations, yet the area with significant trends has necessarily decreased due to the wider spread in the Natural Forcing-only distribution in the

multi-model aggregate distribution, compared to that in the average model distribution. In summary, the results here illustrate some sensitivity of our results to the choice of distribution construction method, although the overall results are generally similar.

Supplementary Note 5. Comparison of trend assessments for precipitation vs. SPI. Our main focus in this report is on a trend assessment for precipitation. As a sensitivity test, in Supplementary Figure 9 we compare assessment results for 1901-2010 trends using the Standardized Precipitation Index (SPI; Appendix) vs. the original result for precipitation (Fig. 3), as a sensitivity test. The trend patterns are very similar using either method. There is a slightly greater fraction of area with detectable and attributable trends for precipitation than for SPI (29% vs. 27%), but the fraction of area with trends consistent with All-Forcing runs is about the same (58% vs. 59%). The fraction of area with an inferred anthropogenic contribution to precipitation changes is about the same for precipitation and SPI, both for increases (20% vs. 19%) and decreases (9% vs. 8%), here taking categories +2, +3 and -2, -3 for comparison. Therefore the choice of using SPI or precipitation for the trend analysis has relatively minor impact on the overall assessment results.

Supplementary Note 6. Changes in extreme monthly SPI values. In this supplementary note, we explore the temporal behavior of some of the extremes of the distribution of monthly SPI values by examining the time evolution of percent area where certain SPI thresholds are exceeded. Supplementary Figure 10 shows time series of the percent of globally analyzed area with SPI values greater than or less than +/-1 or +/-2. These correspond to dry (-1), very dry (-

2), wet (+1) or very wet (+2) monthly conditions for a given location relative to its climatological behavior. The percent of area with relatively dry (SPI < -1) conditions shows a modest reduction since 1900, with some indication of a shift toward less extensive dry area around 1950. For very dry relative conditions (SPI < -2), again a decrease over time is seen, which appears as an abrupt decrease around 1920. Neither of these observed reduced drying extreme behaviors is evident in the All Forcing multi-model mean series. For relatively wet condition thresholds (SPI > +1) a gradual increase in areal coverage is present in the observations, which is also roughly captured in the All Forcing multimodel ensemble, although the model result shows a temporary dip from the 1960s to the 1980, and a rise after about 1990 to its highest levels in the entire record (1901-2010). Similar but more muted temporal behavior is seen for the very wet threshold (SPI > +2) percent area index.

In summary, the observed time series results indicate some reduced occurrence of dry extreme months along with some increased occurrence of wet extreme months. This is qualitatively consistent with the trends in central tendency for SPI and precipitation discussed in the main text and Supplementary Note 5. A remaining question concerns the robustness of the observed changes considering possible observational dataset limitations, and particularly noting the abrupt transitions in some of these areal coverage time series.

Supplementary Note 7. Seasonal analysis of zonal means of precipitation trends. The zonal averages of observed and modeled SPI and precipitation trends (1901-2010) in the main text were based on annual means. Supplementary Figures 11 and 12 show seasonal versions of these for SPI and precipitation, respectively. The SPI zonal means (Fig. 11) show that the tendency for the models to under-predict the increase in mid- to high-latitude precipitation is present in the

northern hemisphere in all seasons, though least pronounced in summertime. It is also present in the southern hemisphere mainly during the austral spring and summer seasons (SON and DJF). The model bias of under-predicting the extratropical precipitation/SPI increases since 1901 is not as pronounced for precipitation as for SPI, although it is clearly present to some degree for both metrics. The zonal means of seasonal precipitation trends (Fig. 12) show that the multi-model bias of under-predicting the wetting trend in the northern hemisphere extratropics is most pronounced for the fall season (SON). In the southern hemisphere extratropics, the bias is most pronounced in the spring and summer (SON and DJF) seasons.

Supplementary Note 8. Precipitation trend maps (1901-2010) for individual CMIP5 models. Trend maps for annual mean precipitation over the periods 1901-2010 and 1951-2010 are shown in Supplementary Figures 13 and 14, respectively, for the All-Forcing ensemble means of each of the 36 individual CMIP5 used in the study. The 10-model and 36-model ensemble mean maps are similar, and their main features are present for most individual models. However, for many regional trend features in the ensembles, there are also a few models which do not agree with even the sign of the multimodel ensemble mean. In some cases the minority of models have a more simulation closer to the observed trend behavior than the majority of models. For example, for the 1951-2010 trends, pronounced negative (drying) trends in observations over the Sahel region associated with the Sahel drought are captured in only a minority of models. Similarly, on average the models simulate a drying trend (1951-2010) over the south central and southwestern U.S.--whereas no drying trend is apparent in observations—but a minority of models do not display this trend bias feature.

Figure Captions

Supplementary Figure 1. As in Fig. 1 c, but for each of 36 individual CMIP5 models. Unit: mm yr^{-1}. Red values along top of diagrams are the spatial correlations between the modeled and observed low-frequency internal standard deviation fields..

Supplementary Figure 2. Assessment of trends in seasonal-mean precipitation over the period 1901-2010. As in Fig. 3 (c) except based on three-month seasons rather than annual means. The seasons include: a) December-February; b) March-May; c) June-August; and d) September-November. Trend assessment summary categories are denoted by the color shading. See legend of Fig. 3 (c) and text for definitions of categories.

Supplementary Figure 3. Assessment of trends in seasonal-mean precipitation over the period 1951-2010. As in Fig. 4 (c) except based on three-month seasons rather than annual means. The seasons include: a) December-February; b) March-May; c) June-August; and d) September-November. Trend assessment summary categories are denoted by the color shading. See legend of Fig. 3 (c) and text for definitions of categories.

Supplementary Figure 4. Assessment of trends in seasonal-mean precipitation over the period 1981-2010. As in Fig. 5 (c) except based on three-month seasons rather than annual means. The seasons include: a) December-February; b) March-May; c) June-August; and d) September-November. Trend assessment summary categories are denoted by the color shading. See legend of Fig. 3 (c) and text for definitions of categories.

Supplementary Figure 5. Assessment of external forcing influence on trends in annual-mean precipitation over the period 1901-2010. As in Fig. 3 except the attribution is to external forcings in general (natural and anthropogenic) rather than anthropogenic forcing alone. Assessment results are based on a 36-model ensemble of CMIP5 models. Trend assessment summary categories are denoted by the color shading. See color shading legend, Fig. 3 caption and text for details. Unit of trends in (a,b): mm yr^{-1} $decade^{-1}$.

Supplementary Figure 6. Assessment of external forcing influence on trends in annual-mean precipitation over the period 1951-2010. As in Fig. 4 except the attribution is to external forcings in general (natural and anthropogenic) rather than anthropogenic forcing alone. Assessment results are based on a 36-model ensemble of CMIP5 models. Trend assessment summary categories are denoted by the color shading. See color shading legend, Fig. 4 caption and text for details. Unit of trends in (a,b): mm yr^{-1} $decade^{-1}$.

Supplementary Figure 7. Assessment of external forcing influence on trends in annual-mean precipitation over the period 1901-2010. As in Fig. 5 except the attribution is to external forcings in general (natural and anthropogenic) rather than anthropogenic forcing alone. Assessment results are based on a 36-model ensemble of CMIP5 models. Trend assessment summary categories are denoted by the color shading. See color shading legend, Fig. 3 caption and text for details. Unit of trends in (a,b): mm yr^{-1} $decade^{-1}$.

Supplementary Figure 8. Comparison of precipitation trend assessment results using two alternative methods of defining the multi-model trend distributions to compare to observed trends. The modeled trend distribution is based on either: a) the average trend distribution characteristics (mean, 5th percentile, 95th percentile) across the 10 individual CMIP5 models, as in Fig. 1 c), or b) the mean, 5th percentile and 95th percentiles are computed from an aggregate distribution of trends which was created by combining samples of trends from all 10 models into a single distribution. See Methods.

Supplementary Figure 9. Assessment of observed Standardized Precipitation Index (SPI) trends over 1901-2010 based on CMIP5 models. Observed (a) and CMIP5 multi-model ensemble (b) trends in SPI in units of $century^{-1}$. 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), with the percent of analyzed area classified in each category 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 dots. 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 significant observed trends (some detectable) which are assessed as summarized in the category legend.

Supplementary Figure 10. Time series of the percent area with annual SPI values exceeding moderate and extreme dry and wet thresholds. The SPI threshold values used are: a) moderately dry: less than -1; b) moderately wet: greater than +1; c) extremely dry: less than -2; and d) extremely wet: greater than +2. The black curves show the observed percent area coverage of various thresholds over time, using a fixed grid consisting of those points with adequate data coverage for trend analysis from 1901 as shown in Fig. 3. The orange curves are the percent area of coverage for individual CMIP5 model ensemble members, and the dark blue curves are the ensemble averages of the threshold coverage across the CMIP5 models, with each model weighted equally in the average. The light blue curves depict the 5th and 95th percentiles of the percent coverages across the CMIP5 set of individual model runs.

Supplementary Figure 11. Zonal averages of SPI trends over the period 1901-2010 for each three-month season. As in Fig. 9 (e), but for three month seasons defined as: a) December-February; b) March-May; c) June-August; and d) September-November. Unit: $Decade^{-1} * 1000$.

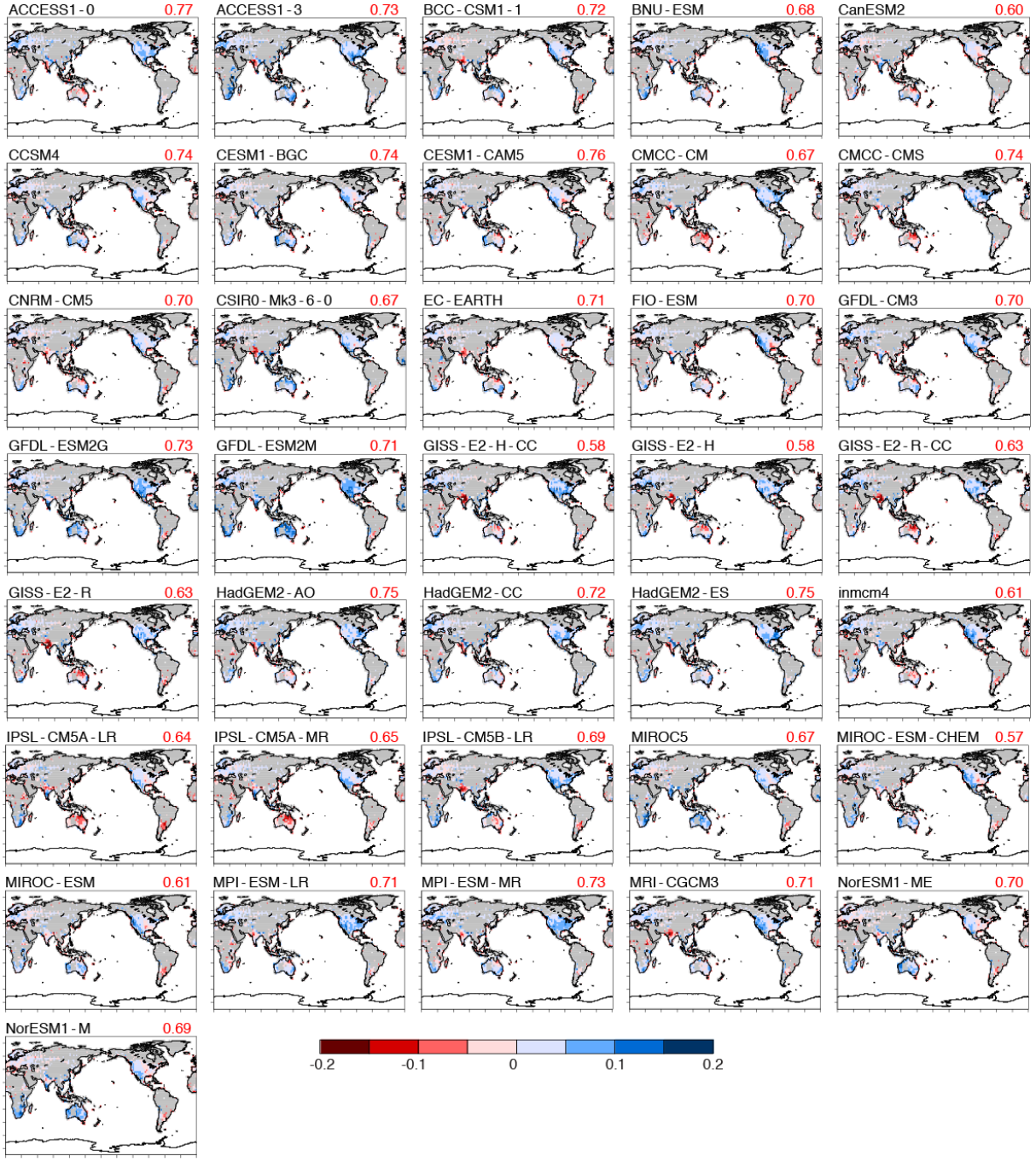
Supplementary Figure 12. Zonal averages of precipitation trends over the period 1901-2010 for each three-month season. As in Fig. 9 (a), but for three month seasons defined as: a) December-February; b) March-May; c) June-August; and d) September-November. Unit: mm yr^{-1} $decade^{-1}$.

Supplementary Figure 13. Annual mean precipitation trends (1901-2010) for (a-jj) CMIP5 individual model All-Forcing runs; kk) observed GPCC trends; (ll) CMIP5 10-model and (mm) 36-model ensemble trends (in units of mm yr^{-1} $decade^{-1}$).

Supplementary Figure 14. Annual mean precipitation trends (1951-2010) for (a-jj) CMIP5 individual model All-Forcing runs; kk) observed GPCC trends; (ll) CMIP5 10-model and (mm) 36-model ensemble trends (in units of mm yr^{-1} $decade^{-1}$).

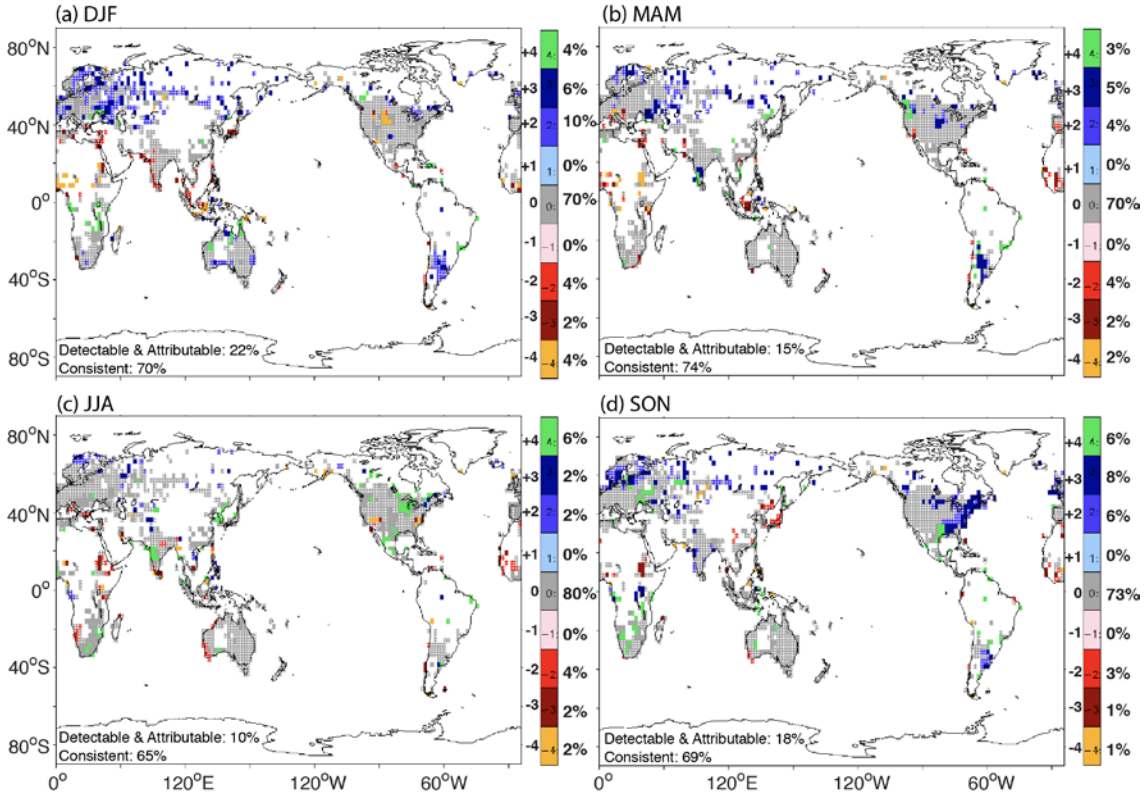
Supplemental Figures

275



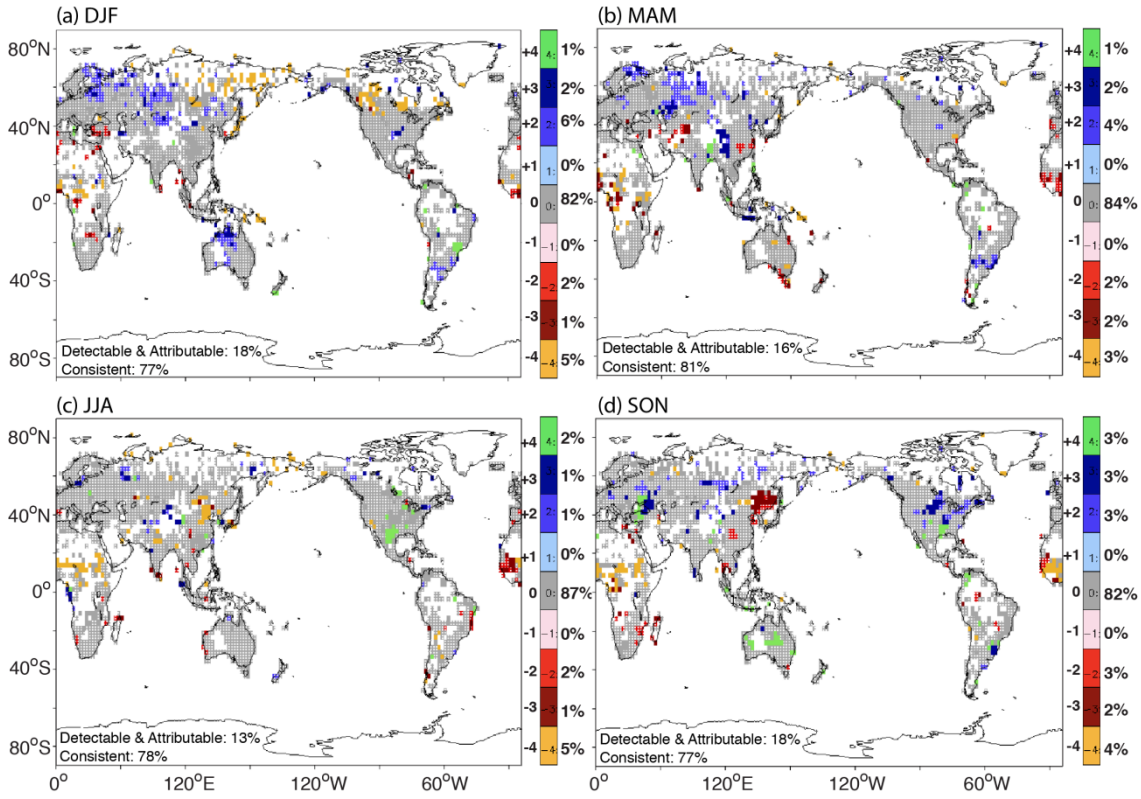
Supplementary Figure 1. As in Fig. 1 c, but for each of 36 individual CMIP5 models. Unit: mm yr^{-1}. Red values along top of diagrams are the spatial correlations between the modeled and observed low-frequency internal standard deviation fields.

276



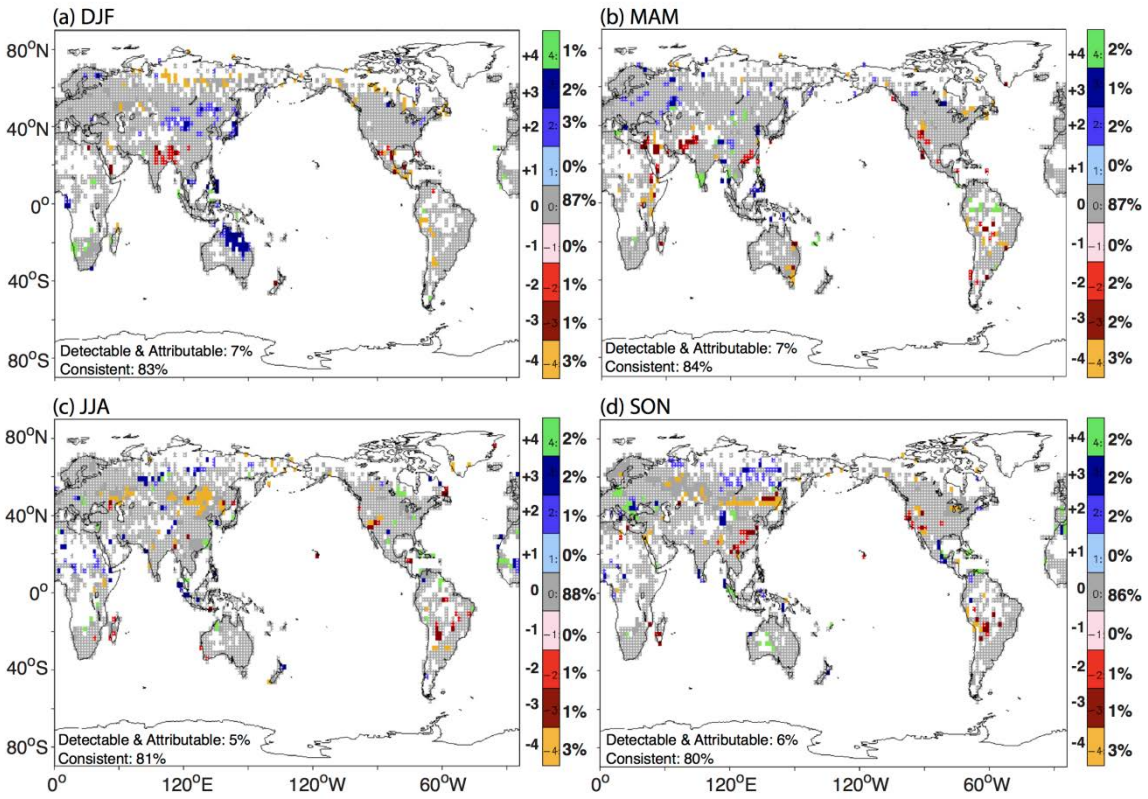
Supplementary Figure 2. Assessment of trends in seasonal-mean precipitation over the period 1901-2010. As in Fig. 3 (c) except based on three-month seasons rather than annual means. The seasons include: a) December-February; b) March-May; c) June-August; and d) September-November. Trend assessment summary categories are denoted by the color shading. See legend of Fig. 3 (c) and text for definitions of categories.

277

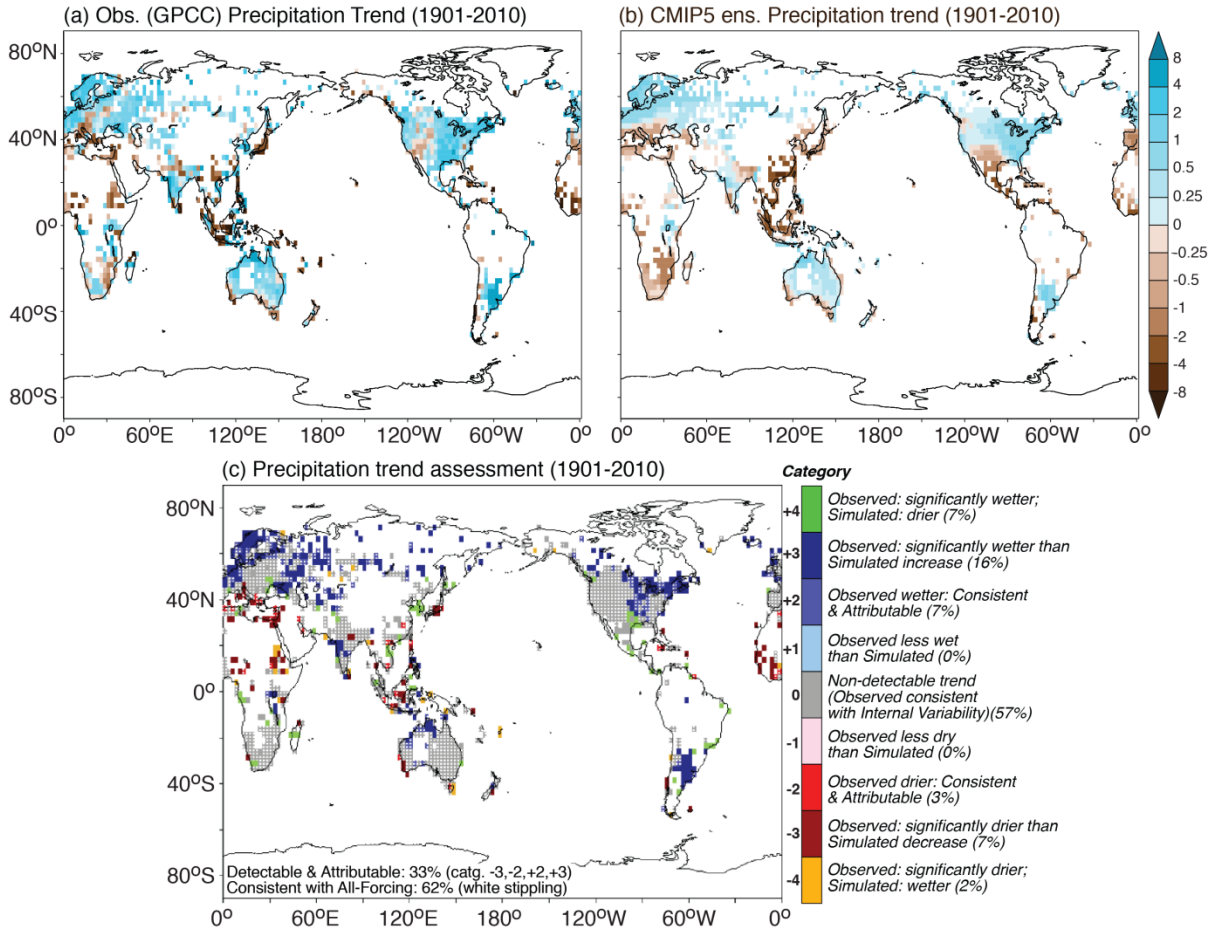


Supplementary Figure 3. Assessment of trends in seasonal-mean precipitation over the period 1951-2010. As in Fig. 4 (c) except based on three-month seasons rather than annual means. The seasons include: a) December-February; b) March-May; c) June-August; and d) September-November. Trend assessment summary categories are denoted by the color shading. See legend of Fig. 3 (c) and text for definitions of categories.

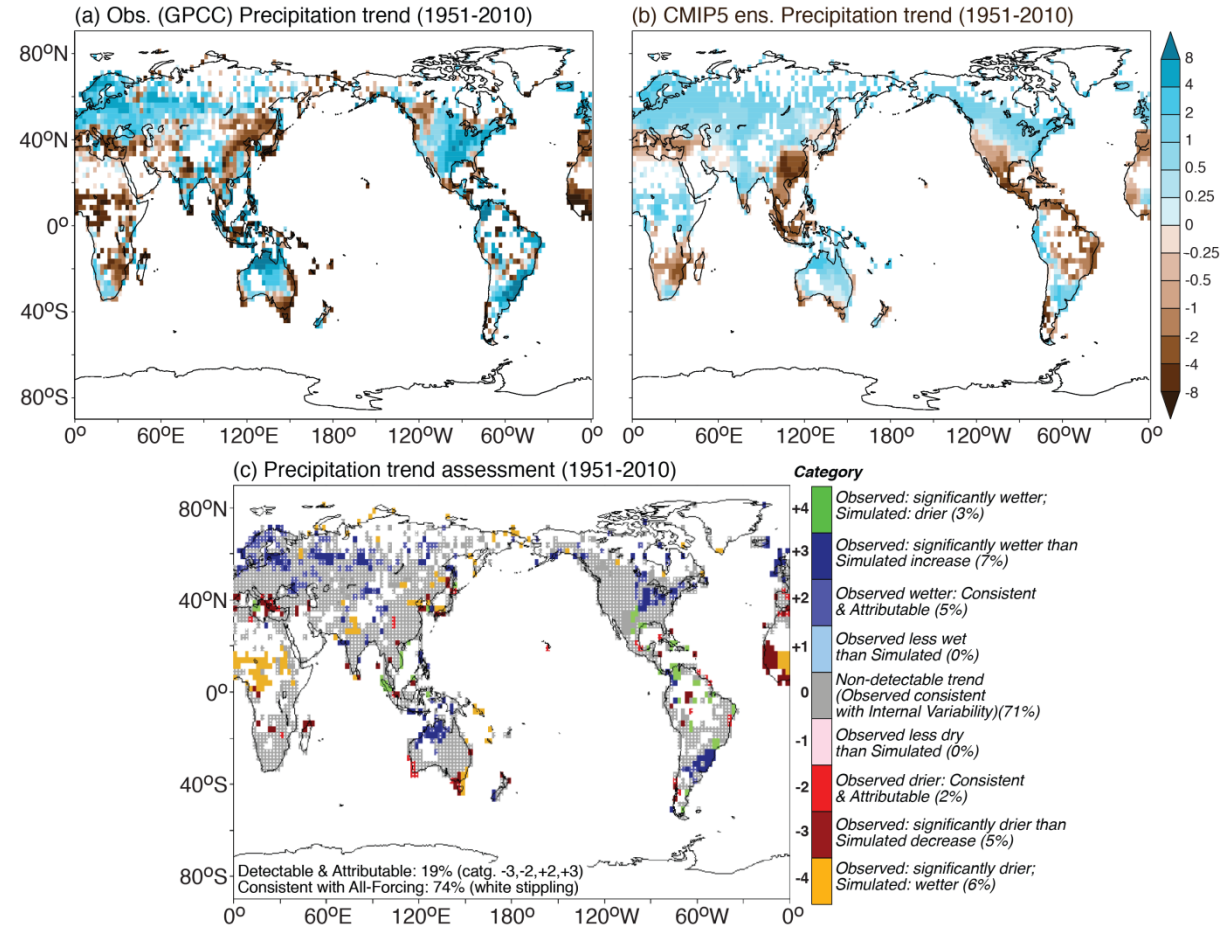
278



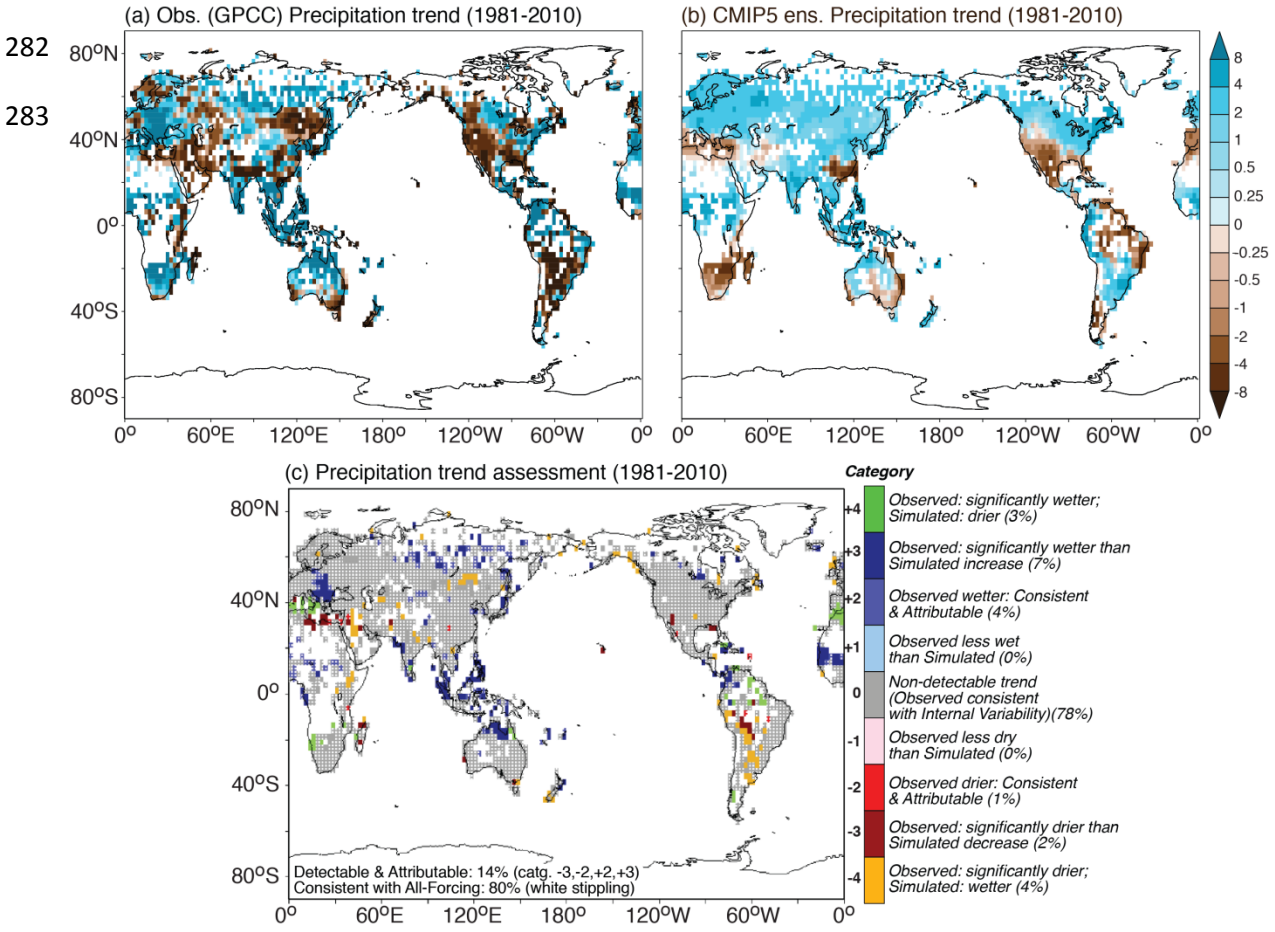
Supplementary Figure 4. Assessment of trends in seasonal-mean precipitation over the period 1981-2010. As in Fig. 5 (c) except based on three-month seasons rather than annual means. The seasons include: a) December-February; b) March-May; c) June-August; and d) September-November. Trend assessment summary categories are denoted by the color shading. See legend of Fig. 3 (c) and text for definitions of categories.



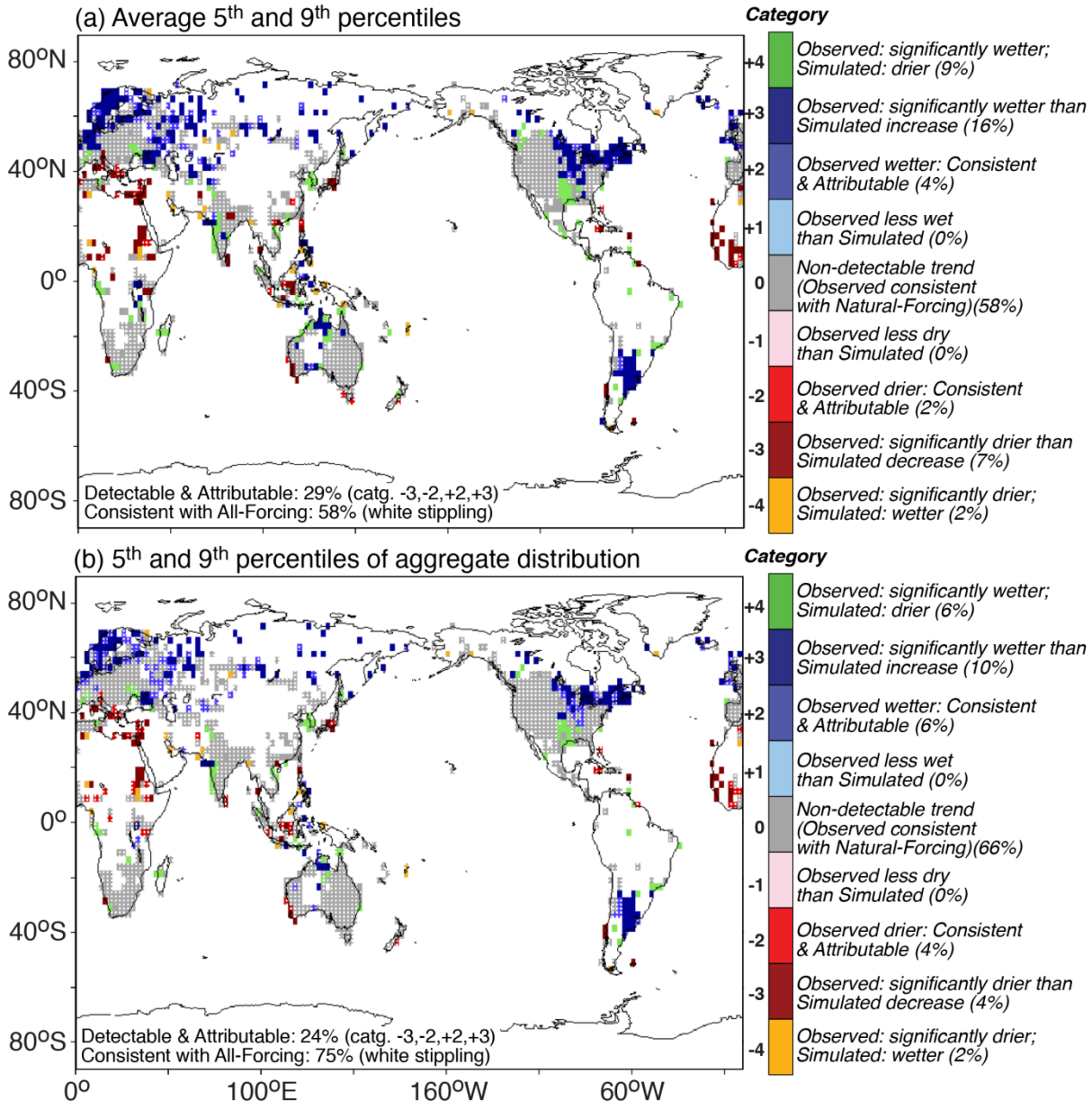
Supplementary Figure 5. Assessment of external forcing influence on trends in annual-mean precipitation over the period 1901-2010. As in Fig. 3 except the attribution is to external forcings in general (natural and anthropogenic) rather than anthropogenic forcing alone. Assessment results are based on a 36-model ensemble of CMIP5 models. Trend assessment summary categories are denoted by the color shading. See color shading legend, Fig. 3 caption and text for details. Unit of trends in (a,b): mm yr^{-1} $decade^{-1}$.



Supplementary Figure 6. Assessment of external forcing influence on trends in annual-mean precipitation over the period 1951-2010. As in Fig. 4 except the attribution is to external forcings in general (natural and anthropogenic) rather than anthropogenic forcing alone. Assessment results are based on a 36-model ensemble of CMIP5 models. Trend assessment summary categories are denoted by the color shading. See color shading legend, Fig. 4 caption and text for details. Unit of trends in (a,b): mm yr^{-1} $decade^{-1}$.



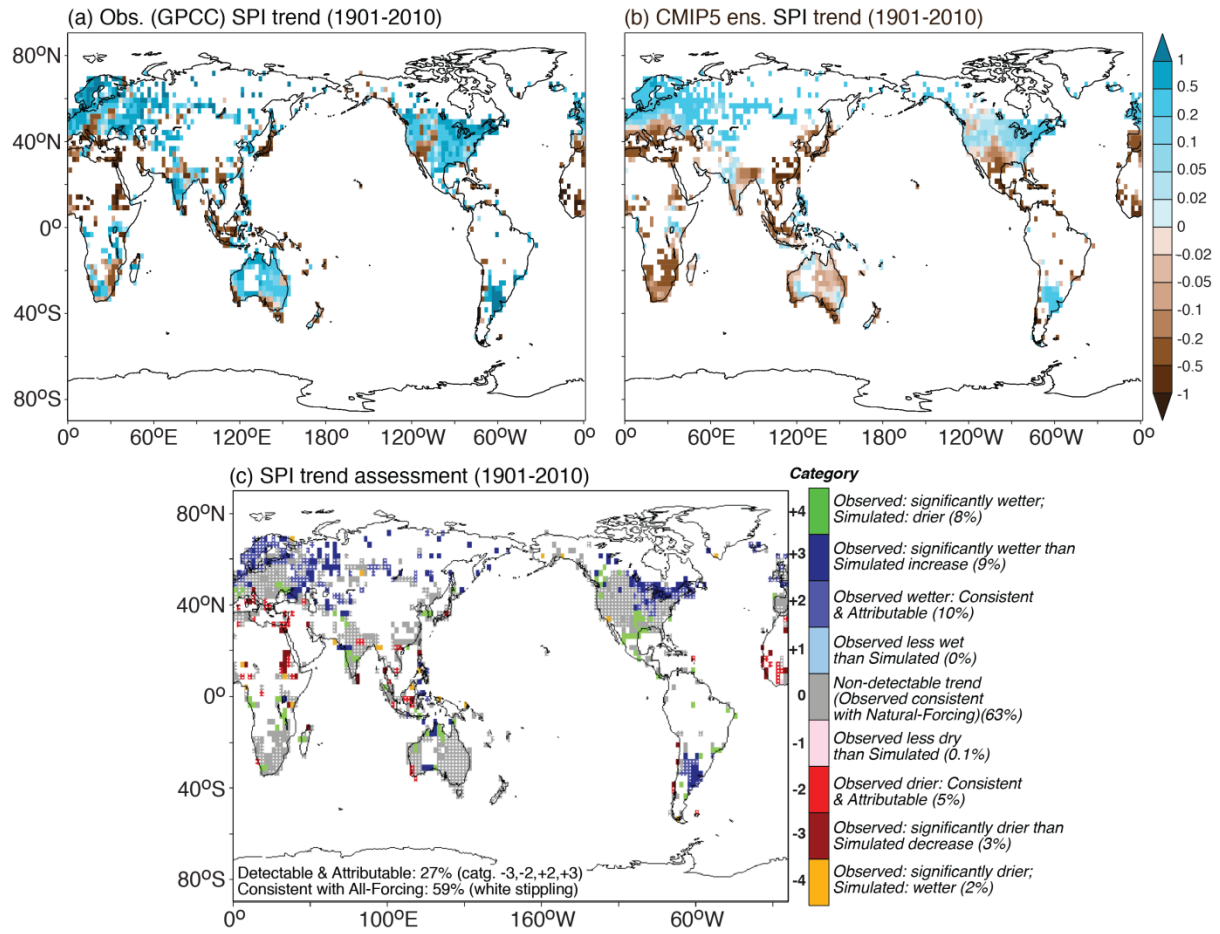
Supplementary Figure 7. Assessment of external forcing influence on trends in annual-mean precipitation over the period 1981-2010. As in Fig. 5 except the attribution is to external forcings in general (natural and anthropogenic) rather than anthropogenic forcing alone. Assessment results are based on a 36-model ensemble of CMIP5 models. Trend assessment summary categories are denoted by the color shading. See color shading legend, Fig. 3 caption and text for details. Unit of trends in (a,b): mm yr^{-1} $decade^{-1}$.



Supplementary Figure 8. Comparison of precipitation trend assessment results using two alternative methods of defining the multi-model trend distributions to compare to observed trends. The modeled trend distribution is based on either: a) the average trend distribution characteristics (mean, 5th percentile, 95th percentile) across the 10 individual CMIP5 models, as in Fig. 1 c), or b) the mean, 5th percentile and 95th percentiles are computed from an aggregate distribution of trends which was created by combining samples of trends from all 10 models into a single distribution. See Methodology section of main report.

286

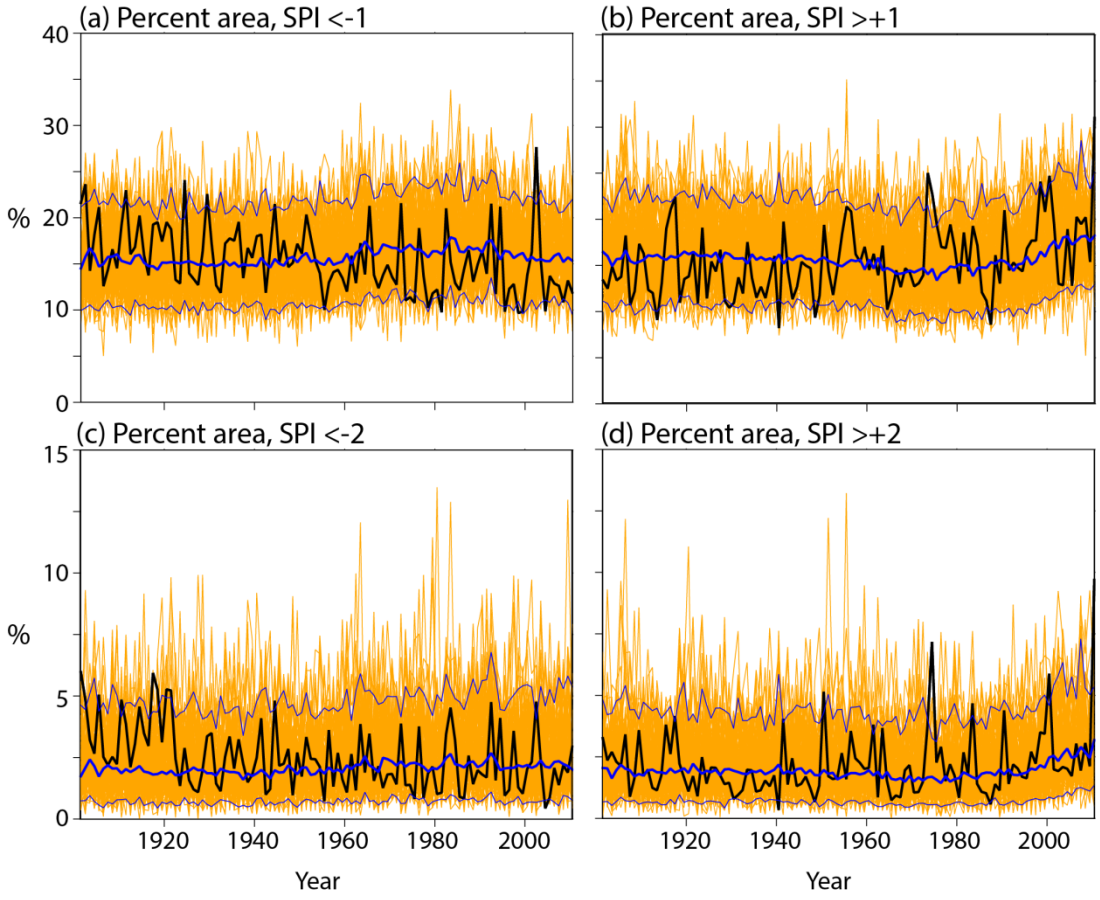
287



Supplementary Figure 9. Assessment of observed Standardized Precipitation Index (SPI) trends over 1901-2010 based on CMIP5 models. Observed (a) and CMIP5 multi-model ensemble (b) trends in SPI in units of century^{-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), with the percent of analyzed area classified in each category 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 dots. 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 significant observed trends (some detectable) which are assessed as summarized in the category legend.

288

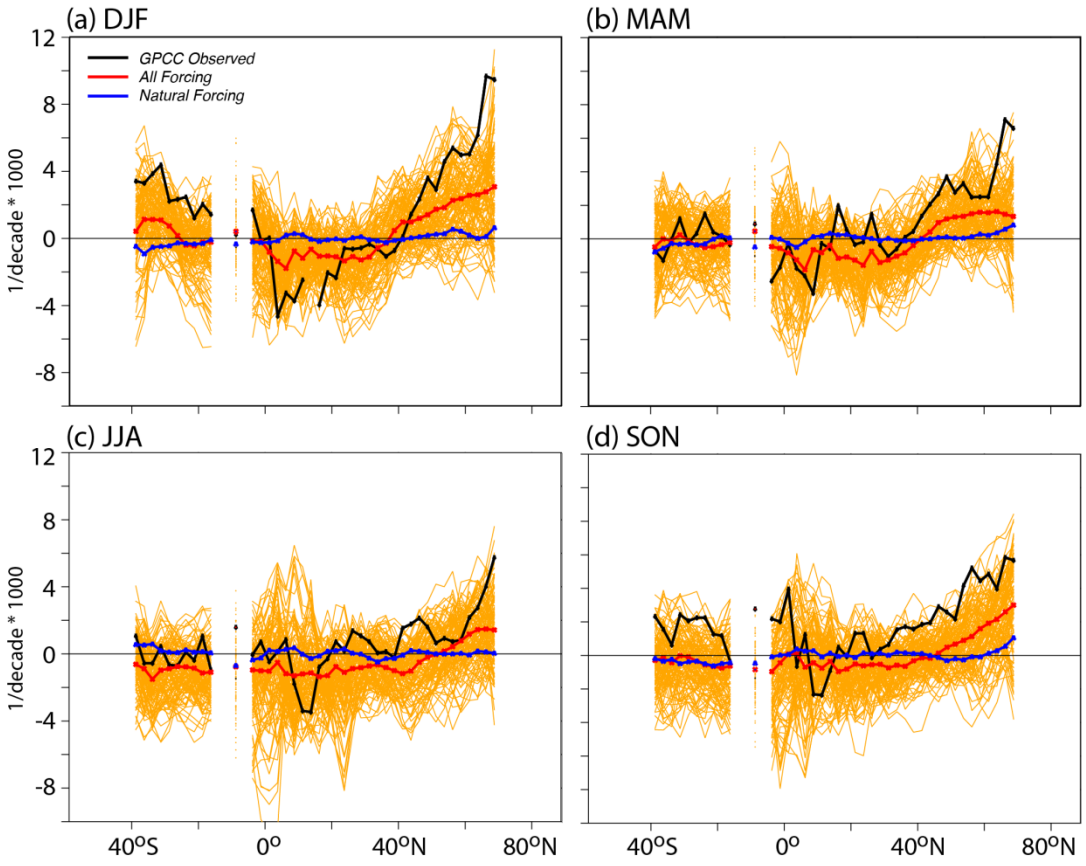
289



Supplementary Figure 10. Time series of the percent area with annual SPI values exceeding moderate and extreme dry and wet thresholds. The SPI threshold values used are: a) moderately dry: less than -1; b) moderately wet: greater than +1; c) extremely dry: less than -2; and d) extremely wet: greater than +2. The black curves show the observed percent area coverage of various thresholds over time, using a fixed grid consisting of those points with adequate data coverage for trend analysis from 1901 as shown in Fig. 3. The orange curves are the percent area of coverage for individual CMIP5 model ensemble members, and the dark blue curves are the ensemble averages of the threshold coverage across the CMIP5 models, with each model weighted equally in the average. The light blue curves depict the 5th and 95th percentiles of the percent coverages across the CMIP5 set of individual model runs.

290

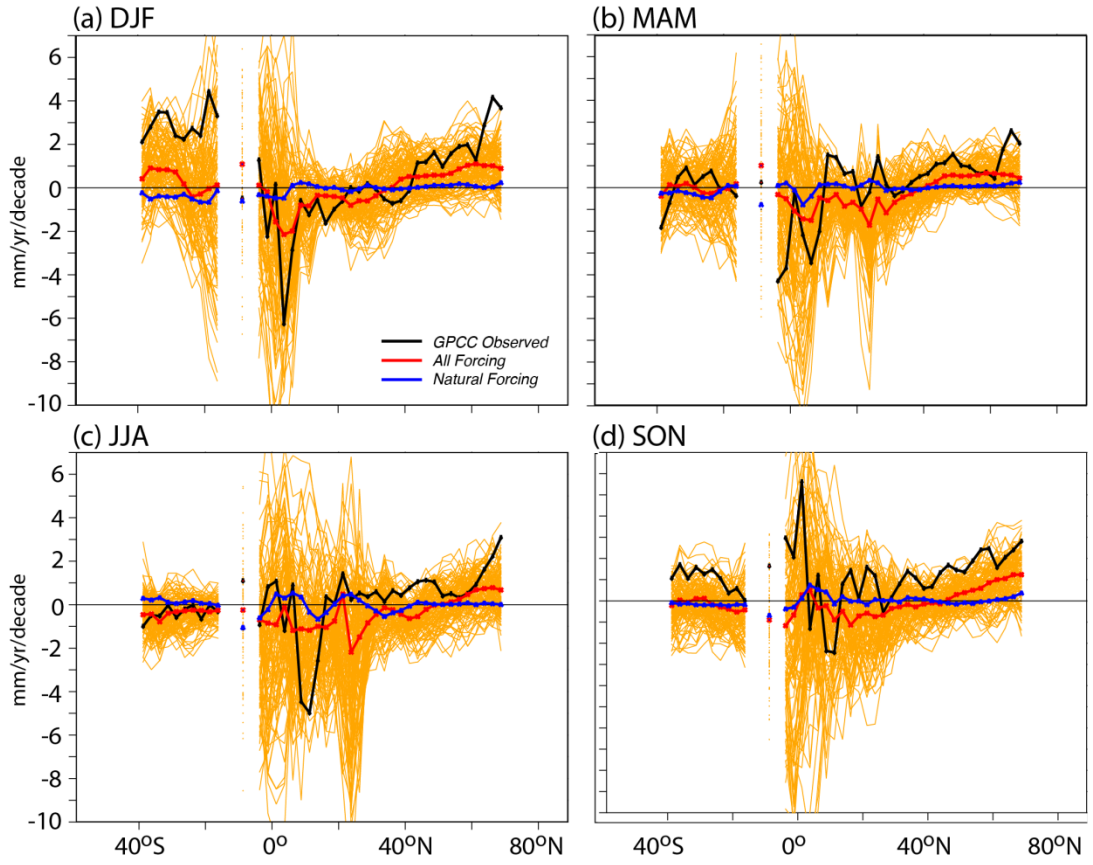
291



Supplementary Figure 11. Zonal averages of SPI trends over the period 1901-2010 for each three-month season. As in Fig. 9 (e), but for three month seasons defined as: a) December-February; b) March-May; c) June-August; and d) September-November. Unit: Decade^{-1} * 1000.

292

293

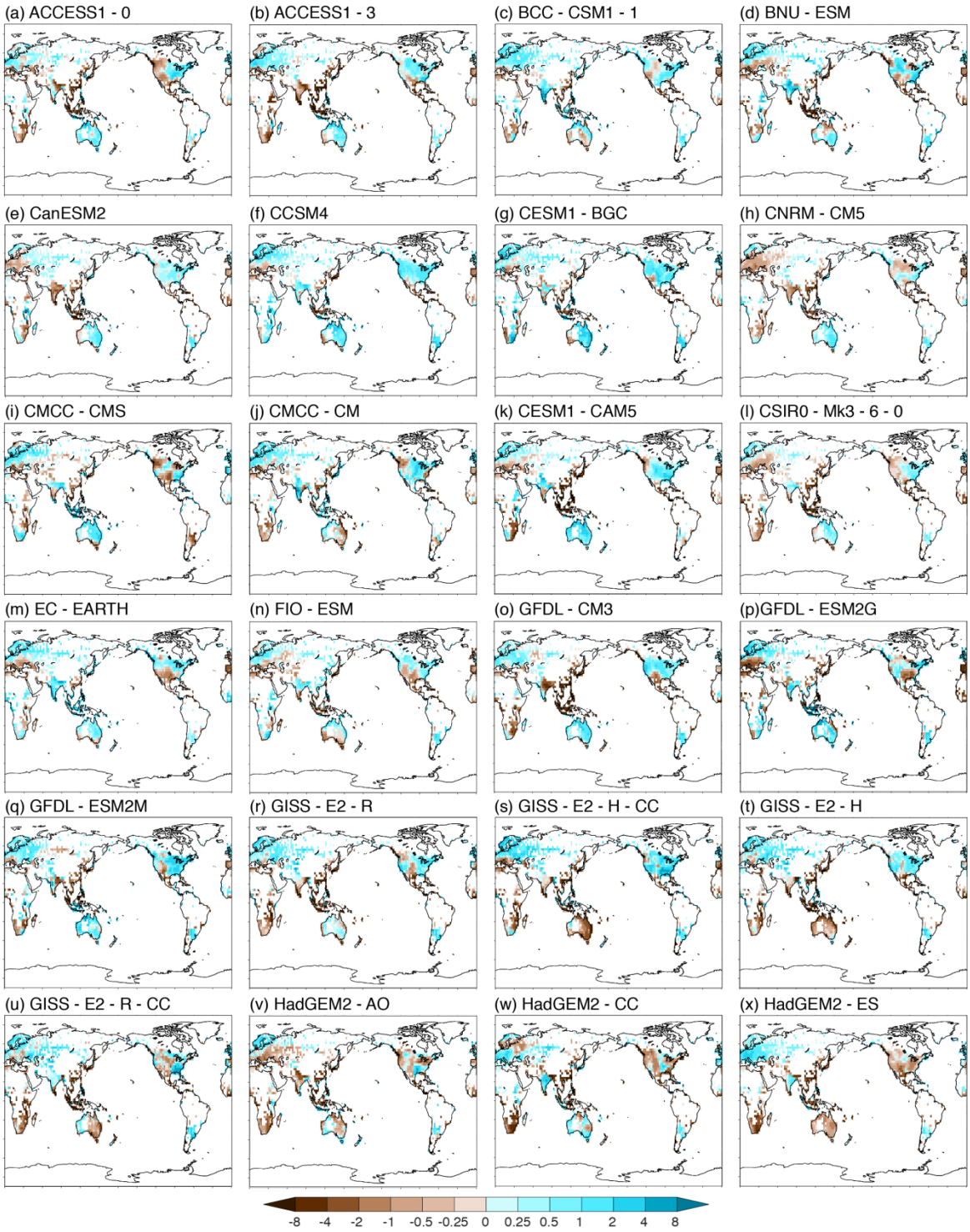


Supplementary Figure 12. Zonal averages of precipitation trends over the period 1901-2010 for each three-month season. As in Fig. 9 (a), but for three month seasons defined as: a) December-February; b) March-May; c) June-August; and d) September-November. Unit: mm yr^{-1} $Decade^{-1}$.

294

295

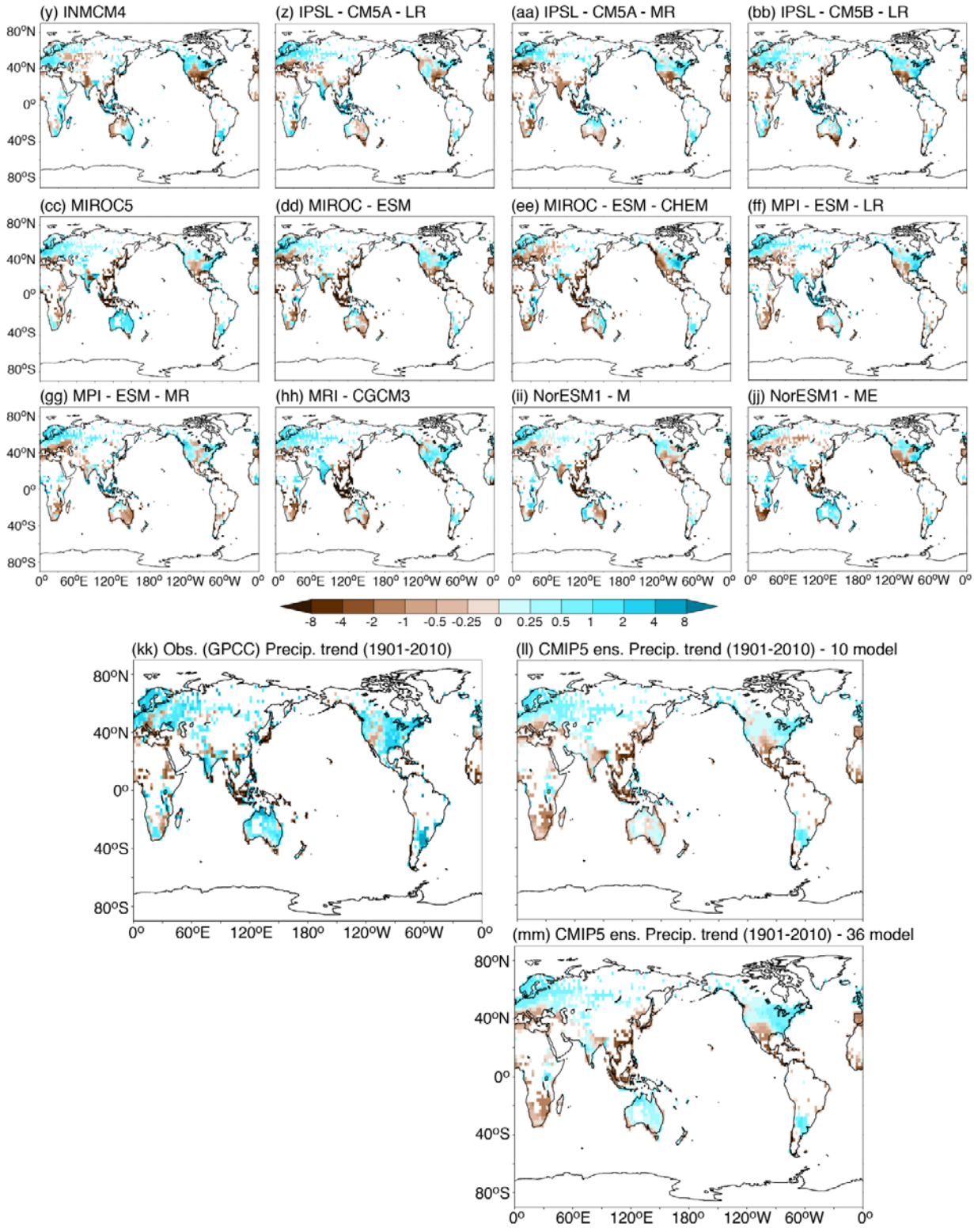
296



Supplementary Figure 13. Annual mean precipitation trends (1901-2010) for (a-jj) CMIP5 individual model All-Forcing runs; kk) observed GPCC trends; (ll) CMIP5 10-model and (mm) 36-model ensemble trends (in units of mm yr^{-1} $decade^{-1}$).

297

298

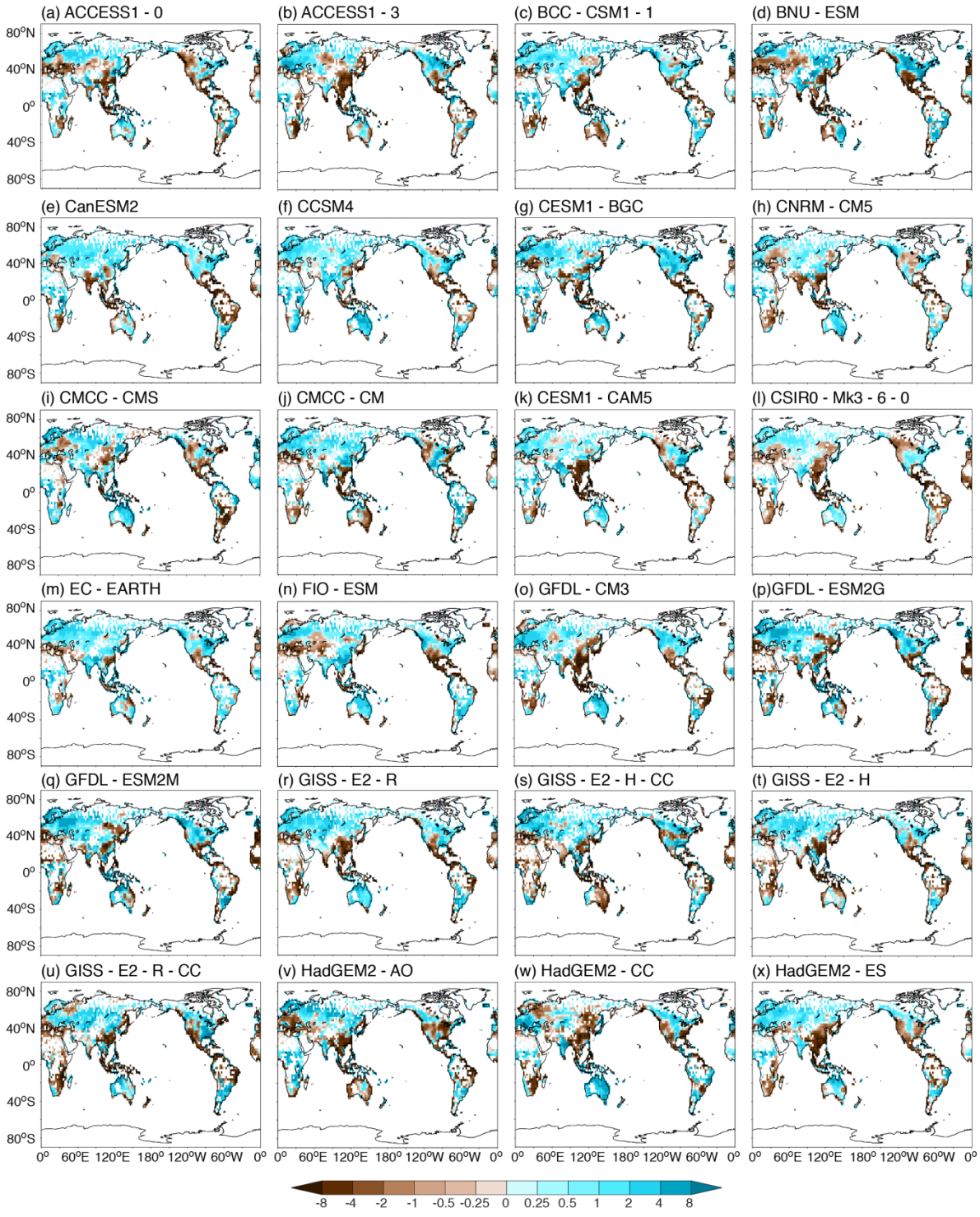


Supplementary Figure 13, contd.

299

300

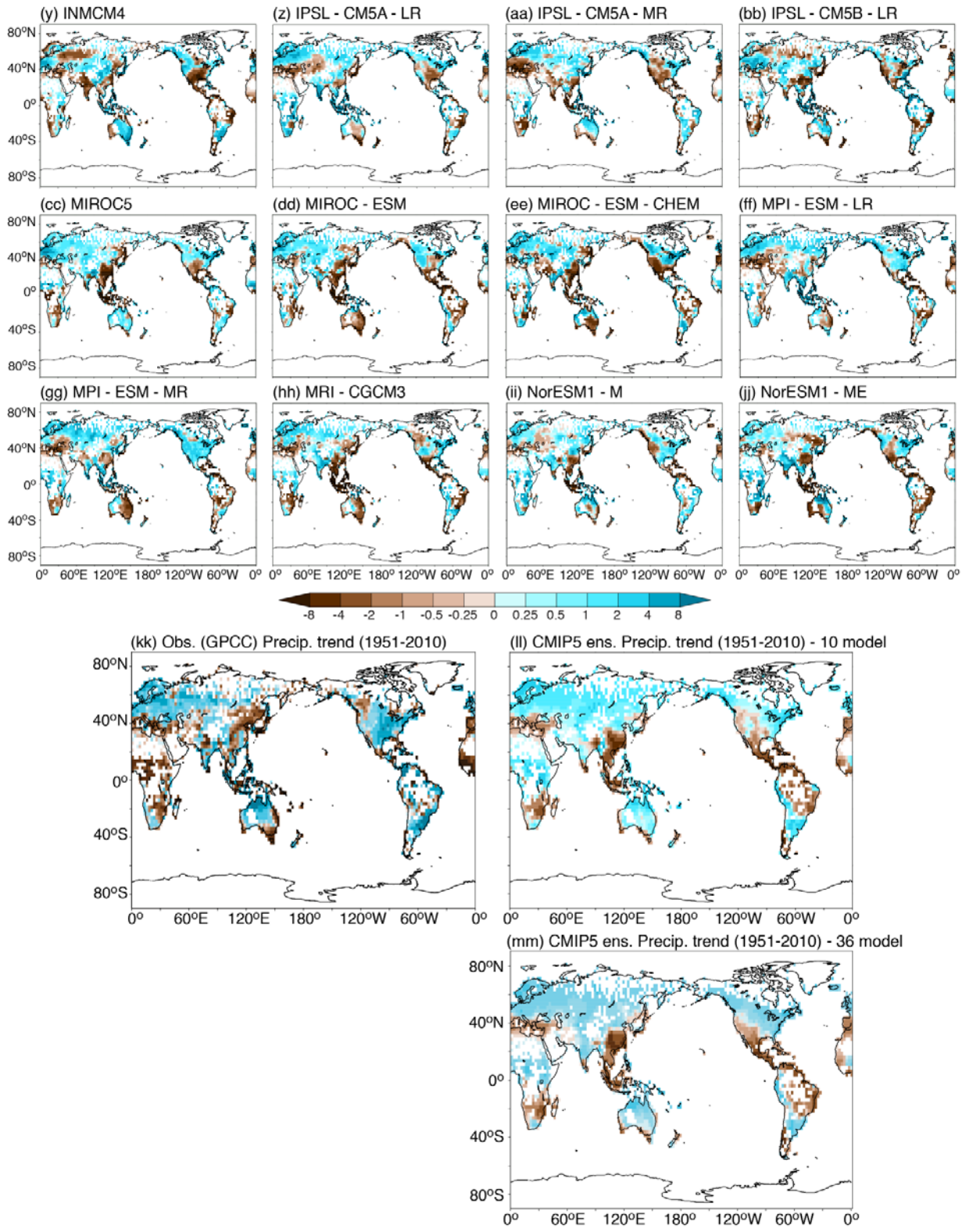
301



Supplementary Figure 14. Annual mean precipitation trends (1951-2010) for (a-jj) CMIP5 individual model All-Forcing runs; kk) observed GPCC trends; (ll) CMIP5 10-model and (mm) 36-model ensemble trends (in units of mm yr^{-1} $decade^{-1}$).

302

303



Supplementary Figure 14, contd.

This page left intentionally blank.