Attribution of Extreme Precipitation with Updated Observations and CMIP6 Simulations

SIYAN DONG,[a] YING SUN,[a,b] CHAO LI,[c] XUEBIN ZHANG,[d] SEUNG-KI MIN,[e] AND YEON-HEE KIM[e]

[a] *National Climate Center, Laboratory for Climate Studies, China Meteorological Administration, Beijing, China*
[b] *Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, China*
[c] *Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences, East China Normal University, Shanghai, China*
[d] *Climate Research Division, Environment and Climate Change Canada, Toronto, Ontario, Canada*
[e] *Division of Environmental Science and Engineering, Pohang University of Science and Technology, Pohang, South Korea*

(Manuscript received 31 December 2019, in final form 30 October 2020)

ABSTRACT: While the IPCC Fifth Assessment Working Group I report assessed observed changes in extreme precipitation on the basis of both absolute and percentile-based extreme indices, human influence on extreme precipitation has rarely been evaluated on the basis of percentile-based extreme indices. Here we conduct a formal detection and attribution analysis on changes in four percentile-based precipitation extreme indices. The indices include annual precipitation totals from days with precipitation exceeding the 99th and 95th percentiles of wet-day precipitation in 1961–90 (R99p and R95p) and their contributions to annual total precipitation (R99pTOT and R95pTOT). We compare these indices from a set of newly compiled observations during 1951–2014 with simulations from models participating in phase 6 of the Coupled Model Intercomparison Project (CMIP6). We show that most land areas with observations experienced increases in these extreme indices with global warming during the historical period 1951–2014. The new CMIP6 models are able to reproduce these overall increases, although with considerable over- or underestimations in some regions. An optimal fingerprinting analysis reveals detectable anthropogenic signals in the observations of these indices averaged over the globe and over most continents. Furthermore, signals of greenhouse gases can be separately detected, taking other forcing into account, over the globe and over Asia in these indices except for R95p. In contrast, signals of anthropogenic aerosols and natural forcings cannot be detected in any of these indices at either global or continental scales.

KEYWORDS: Rainfall; Anthropogenic effects; Climate change; Climate variability; Climate models

1. Introduction

Changes in extreme precipitation are attracting attention because of the disastrous effects of extreme precipitation on human life, social economy, agriculture, and ecosystems (IPCC 2013). Increases in the frequency and intensity of extreme precipitation have been observed at continental to global scales (Donat et al. 2013; Hartmann et al. 2013; Min et al. 2011; Westra et al. 2013). Globally, precipitation extremes have increased in more regions than have decreased (Alexander 2016; Rhein et al. 2013). At the continental scale, there exists an overall increase in the intensity and frequency of heavy precipitation in Europe, Asia, and North and South America (de los Milagros Skansi et al. 2013; Donat et al. 2013; Huang et al. 2017).

It is important to understand the causes of the observed changes in extreme precipitation. Some studies have used simulations conducted with the models participating in phases 3 and 5 of the Coupled Model Intercomparison Project (CMIP3 and CMIP5, respectively) to conduct the detection and attribution analysis of extreme precipitation. Min et al. (2011) attributed the observed increases in annual maximum 1-day precipitation (Rx1day) in the Northern Hemispheric land region to anthropogenic influence on the climate. Using CMIP5 simulations and observations with improved spatial and temporal coverage, Zhang et al. (2013) reported that anthropogenic influence on Rx1day and annual maximum 5-day precipitation (Rx5day) was detected in regions such as North America and Europe. Wan et al. (2014) showed that the anthropogenic influence in high-latitude precipitation was detectable. Kirchmeier-Young and Zhang (2020) provided robust evidence that human influence has intensified extreme precipitation in North America. Attribution of changes in extreme precipitation at regional or country scales is less robust. For example, whereas Chen and Sun (2017) showed that anthropogenic signal could be detected in extreme precipitation over China, Li et al. (2018) indicated that the detection was not robust. While there is evidence of human influence on extreme precipitation, model simulations suggest important differences in the roles that greenhouse gases and anthropogenic aerosols

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-19-1017.s1.

Corresponding author: Ying Sun, sunying@cma.gov.cn

DOI: 10.1175/JCLI-D-19-1017.1

play in the changes of mean and extreme precipitation. Wu et al. (2013) showed that the lack of discernable trend in global mean precipitation expected from global warming is in part due to the counteracting effects of anthropogenic greenhouse gases and aerosols. Pendergrass et al. (2015) showed that the influence of greenhouse gases and aerosols on extreme precipitation can differ substantially from that on mean precipitation. Studies with climate model simulations have suggested a substantial influence of aerosols on extreme precipitation in East Asia and in northwest Australia (Dey et al. 2019; Lin et al. 2016; Sillmann et al. 2017; Wang 2015). Yet the individual roles of greenhouses gases and anthropogenic aerosols in the observed changes in precipitation extremes have not been studied using a formal detection and attribution framework.

Percentile-based extreme precipitation indices (Zhang et al. 2011) including R99p and R95p and their fractions in the annual total precipitation (R99pTOT and R95pTOT) have been widely used in the studies of past and future changes in extreme precipitation (e.g., Rhein et al. 2013; Dunn et al. 2020) and assessment of changes in extreme precipitation (e.g., Hartman et al. 2013). The R99p and R95p are defined as annual total precipitation accumulations from days with precipitation exceeding respectively the 99th or 95th percentiles of the probability distribution of wet-day (defined as days with precipitation amount no less than 1 mm) precipitation during a base period (e.g., 1961–90). The R99pTOT and R95pTOT are defined by the ratio of these values to annual total precipitation during wet days (PRCPTOT). These indices provide a different angle of view than that of Rx1day (Zhang et al. 2011) although it can be difficult to compare these indices with Rx1day (Schär et al. 2016). Rx1day, defined as the maximum amount of daily precipitation that occurred in the year, has an advantage in that it occurs once every year regardless of climate and that it is widely used in many engineering applications. The number of events included in the calculation of R99p and R95p is generally different from one and varies from one region to the other depending on climates. For example, the exceedance of the 99th percentile of wet-day amounts may happen more frequently than once per year in wet climates with more than 100 wet days (precipitation $>$ 1 mm) per year, or less frequently than once per year in dry climates. The exceedance of the 95th percentile of wet-day amounts would happen more frequently than once per year in many regions. As a result, the intensity of such events is of different level of rarity than Rx1day, and less rare for R95p in most regions. Different levels of rarity of extreme precipitation may respond to warming differently (e.g., Li et al. 2019). The calculation of R95p in most regions, and of R99p in wet regions, involves the use of information about precipitation on more than one day in a year; this may allow the extraction of more information from available data than that from Rx1day, in a way similar to that of the peak-over-threshold method for the analysis of extreme value (e.g., Coles 2001). Previous detection and attribution studies on extreme precipitation have focused mainly on Rx1day and/or Rx5day (e.g., Min et al. 2011; Zhang et al. 2013; Paik et al. 2020). An exception might be Dong et al. (2020), who conducted detection and attribution analysis on percentile-based extreme precipitation indices for Asia.

Here we examine possible human influence on these percentile-based precipitation indices at both the global and continental scales, and paying particular attention to the possibility of separating the influence of anthropogenic greenhouse gases from other forcings such as aerosols as the availability of simulations from the Detection and Attribution Model Intercomparison Project (DAMIP; Gillett et al. 2016) component of phase 6 of the Coupled Model Intercomparison Project (CMIP6) offers a unique opportunity to do so. We consider R99p and R95p. We also consider contributions of annual precipitation from extremely wet days (R99pTOT) and from very wet days (R95pTOT), defined as ratios between R99p or R95p to PRCPTOT (Dunn et al. 2020). This is because the heterogeneous nature of extreme precipitation across space and the uneven distribution of available station data both in time and space present a unique challenge when computing mean values for large regions (also see section 2). The use of ratios may reduce such difficulty to some extent. The paper is structured as follows. Section 2 describes observational and CMIP6 model data used in this study, along with an introduction about the data processing and attribution analysis method. We then present the results in section 3 and conclude the study in section 4.

2. Data and methods

a. Regions

We examine human influence on extreme precipitation at global and continental scales. While the IPCC Sixth Assessment Working Group I report attempts to use a common set of reference regions (Iturbide et al. 2020), there has not yet been an attempt to assign these reference regions with continental boundaries. For this reason, we follow the geographic demarcation of the continents used by Jones et al. (2013) (see Fig. S1 in the online supplemental material). Here, we consider six continents including Asia (ASI), Europe (EUR), North America (NAM), South America (SAM), Australia (AUS), and Africa (AFR). Among them, observational data coverage in SAM, AUS, and AFR is very limited (Fig. S2 in the online supplemental material).

b. Observations

HadEX2 (Donat et al. 2013) has been a popular global dataset for the analysis of changes in weather and climate extremes including precipitation extremes. The last year in this dataset was 2010. At the time of writing, an update to HadEX2, HadEX3, was not available. We therefore decided to compile a global dataset using HadEX2 as a model. Here we use the daily precipitation dataset compiled by Sun et al. (2021), who examined historical changes in annual maximum 1-day and 5-day precipitation. It extends the HadEX2 collection in time to include more recent years, and also adds more stations in several countries including Australia, Argentina, Canada, China, and Russia. These additional data were obtained from relevant national meteorological services in respective countries. In total, it contains observations from 14 796 land-based observing stations with more than 30 years of record across the global land area during 1900–2018. Compared with HadEX3 (Dunn et al. 2020), which is now available, this dataset has a better spatial coverage in some regions especially in Asia, North

America, and Africa (Fig. S2 in the online supplemental material). More details about this dataset can be found in Sun et al. (2021).

We computed R99p and R95p as annual totals of wet-day precipitation amounts that exceed the 99th and 95th percentiles during base period 1961–90 as was defined in Zhang et al. (2011). We also computed the annual total wet-day precipitation (PRCPTOT). Here after, we refer the ratio between R99p and PRCPTOT as R99pTOT and that between R95p and PRCPTOT as R95pTOT. A wet day is defined as a day with precipitation amount no less than 1 mm day^{-1}. We only considered long-term stations in our analyses, using stations with at least 70% of years (45 yr) during 1951–2014. In total, approximately 8000 stations over the global land area were retained for analysis.

For two reasons, our analyses are conducted for the period 1951–2014: historical simulations became less available after 2014, and the spatial coverage of observational data is poor prior to 1950 and after 2014 (Fig. S3 in the online supplemental material). We conduct our analyses on global and continental averages of the indices. To obtain these averages, we first gridded the station anomalies (relative to the 1961–90 average) of the four indices (R99p, R95p, R99pTOT, and R95pTOT) onto 1.875° × 1.25° longitude–latitude grid boxes by averaging all station anomalies within each grid cell if that grid cell has at least one station. This grid size is consistent with that used in HadEX3 (Dunn et al. 2020). It should be noted that the gridded values represent grid box averages of point estimation of extremes, rather than extremes of grid box mean precipitation.

We find that the global and continental mean series from our data are quite similar to those of HadEX3 for the globe and the Northern Hemispheric continents, with clear differences in the Southern Hemispheric continents (Fig. S4 in the online supplemental material). The global and continental mean series for Northern Hemispheric continents are very consistent between our series and HadEX3, indicating that the two datasets are consistent at such spatial scale despite differences in the underlying station coverage and in the methods of data processing. We also found visible and sometimes sizeable difference in the continental mean series for SAM, AUS, and AFR, and for Africa in particular between the two datasets. Given that the spatial coverage in the Southern Hemisphere including SAM, AUS, and AFR is very limited, with the number of grid boxes less than 100 (not shown), and there exist differences between the two datasets, we conclude that the data coverage for the Southern Hemispheric continents is not sufficiently good for conducting detection and attribution analysis. We thus conduct detection and attribution analyses on global and continental averages in the Northern Hemisphere.

When computing global and continental averages, difficult choices need to be made regarding station density and spatial coverage and representativeness. On one hand, the use of a coarser grid without increasing the minimum number of stations per grid for gridding will result in larger areas with data coverage but at the cost of higher uncertainty in individual grid box values. On the other hand, if higher station density is required, which will reduce uncertainty in the estimates for individual boxes, there will be a reduction in the number of boxes with data, thereby reducing the fractional area covered by grid boxes containing observations. Our choice of using the resolution of 1.875° × 1.25° should be considered as more conservative in representing values at the grid box level when compared with HadEX3 data since we do not use an interpolation scheme that enables the infilling of data for grid boxes without any observations. To explore the sensitivity of detection and attribution result to the choices made, we also gridded station data onto 3.75° × 2.5° longitude–latitude grid boxes, again requiring at least one station observation in a box. We find that this difference in gridding methods can have impact in quantitative results, but the qualitative conclusions about whether different forcing signal can be detected are not affected. We choose to present the results that are based on the 1.875° × 1.25°grid boxes in the main text and provide the results that are based on the 3.75° × 2.5°grid boxes in the online supplemental material (Figs. S5–S7).

c. *Model data*

We use simulations of daily precipitation for the period 1951–2014 from historical (hist), historical natural forcing only (hist-nat), historical greenhouse gases only (hist-GHG), and historical anthropogenic aerosol only (hist-aer) experiments collected in CMIP6's DAMIP (Gillett et al. 2016). Simulations in these experiments are forced respectively with all-known historical (ALL) forcing, historical natural (NAT) forcing (e.g., solar and volcanic forcings), historical greenhouse gas (GHG) forcing, and historical aerosol (AER) forcing. For hist-aer simulations, there are two experimental designs depending on whether a model includes a complete representation of atmospheric chemistry (Gillett et al. 2016). In this study, we use 7 models that have at least three ensemble members for each forcing experiment and 16 models that provide the preindustrial control experiments. Table 1 provides details on the climate models and simulations used in this study. In total, there are 32 runs for hist, hist-nat, hist-GHG, and hist-aer and 111 chunks from preindustrial control experiments (CTL). Extreme indices are calculated at each grid cell of each model's native grid. All of the indices are converted to anomalies relative to 1961–90 and then bilinearly interpolated to the 1.875° × 1.25° grid. We mask the gridded anomalies by the availability of gridded observations to ensure a fair comparison between observations and models.

While detailed evaluation of models performance in simulating extreme precipitation indices used in this study is beyond the scope of this study, we will provide some basic comparisons between observation and the simulations to give a flavor of model skill. Note that detection and attribution analysis, which examines how well a model-simulated signal fits the observation, is in itself a stringent form of model evaluation. Figure S8 in the online supplemental material shows the spatial patterns of observed and CMIP6 model-simulated extreme indices during 1971–2000. Models produce the spatial patterns of the climatologies of all indices well, though the magnitudes in simulated indices are slightly higher than that observed in western North America. The simulated R99pTOT or R95pTOT values are smaller than the observations in central and eastern Asia, Europe, and eastern North America. We also produced a Taylor

TABLE 1. List of multimodel simulations used in this study. Numbers represent the ALL, GHG, AER, and NAT forcing simulation ensemble sizes or the number of 64-yr chunks for the CTL simulations. Expansions of model names can be found online (https://www.ametsoc.org/PubsAcronymList).

	Model name	ALL	GHG	AER	NAT	CTL
1	BCC_CSM2(m)	3	3	3	3	
2	CanESM5	10	10	10	10	11
3	CNRM-CM6.1	3	3	3	3	7
4	HadGEM3-GC31-LL	4	4	4	4	7
5	IPSL-CM6A-LR	6	6	6	6	17
6	MRI-ESM2.0	3	3	3	3	2
7	NorESM3-LM	3	3	3	3	4
8	CNRM-ESM2.1					7
9	EC-EARTH3					7
10	EC-EARTH3-Veg					5
11	GFDL CM4					7
12	GFDL-ESM4					7
13	INM-CM4.8					7
14	MIROC-ES2L					7
15	MIROC6					7
16	MPI-ESM1.2-HR					7
17	UKESM1.0-LL					2
	Sum (models) 1951–2014	32 (7)	32 (7)	32 (7)	32 (7)	111 (16)

diagram (Taylor 2001) with the spatial correlation and normalized standard deviation for mean annual indices during 1951–2014 from observations and simulations over the global and three continental regions (Fig. S9 in the online supplemental material). In general, spatial correlation between observation and simulation is high, around 0.9, and the normalized standard deviations in the observation and simulations are comparable and the ensemble mean is closer to observation than individual model simulations. All these indicate that the CMIP6 simulations reproduce the basic features of changes in extreme precipitation, giving confidence that they can be used for the purpose of this study. Dittus et al. (2016) also found that many CMIP5 models can reproduce the observed trends in R95pTOT in many Northern Hemisphere regions.

d. Detection and attribution

Detection and attribution analysis is conducted with a generalized linear regression (Ribes et al. 2013), which regresses the observations $\mathbf{Y}$ onto multimodel-simulated signal patterns $\mathbf{X}$, plus internal climate variability $\boldsymbol{\varepsilon}$: $\mathbf{Y} = (\mathbf{X} - \mathbf{v})\beta + \boldsymbol{\varepsilon}$, where $\mathbf{v}$ represents the influence of natural internal variability in the modeled signal patterns and the regression coefficients β are unknown scaling factors to be estimated. If the 90% confidence interval of β falls above 0, the corresponding modeled signals are considered to be detected in the observations. In these cases, if the 90% confidence interval of β also includes 1, then the modeled responses are considered to be consistent with the observations (Allen and Stott 2003; Ribes et al. 2013).

For all of the observations and model simulations, the gridded anomalies (relative to 1961–90) of extreme precipitation indices during 1951–2014 at grid boxes are averaged over the globe and in three continental regions. In the detection analyses, the temporal and spatial dimensions are both reduced. For the temporal dimensions, the regional averaged time series is reduced by taking nonoverlapping 5-yr means, producing 13 temporal points for a given region (i.e., the globe and three continents). Since the historical simulations end in the year 2014, the last element of the 5-yr mean anomaly time series represent effectively the 4-yr (2011–14) mean anomaly. The application of 5-yr-mean filtering smooths out the response of extreme indices to short-term forcing fluctuations, such as large volcanic events that potentially affect the detectability of the natural forcing signal. On the other hand, the spatial dimension is reduced by the use of global and continental averages. We thus only have one spatial dimension for the global and regional detection analyses. Spatial averaging reduces regional features of responses to aerosol forcing, which may make it more difficult to separate this forcing response from that of other forcings. Allowing higher dimensionality, on the other hand, increases the challenges involved in estimating the covariance structure of the noise terms in the total least squares regression model.

To estimate and test the scaling factors in the detection analyses, we obtain 111 total 64-yr nonoverlapping chunks of preindustrial control simulations from 16 CMIP6 climate models to estimate internal variability in the extreme indices (Table 1). These 64-yr chunks emulate the 1951–2014 observations, with the first year representing the observations of 1951. These chunks are used to estimate the scaling factor and its 5%–95% confidence interval and also to carry out a standard residual consistency test (Allen and Stott 2003). The confidence intervals for the scaling factors are obtained following Ribes et al. (2013). Also, we calculate the signal-to-noise ratios of changes in these indices in historical simulations. The signal is estimated as the linear trend during 1951–2014 in the ensemble mean historical simulations of an extreme index in a given model, while the noise is estimated as the standard deviation of corresponding extreme indices in preindustrial control simulations.

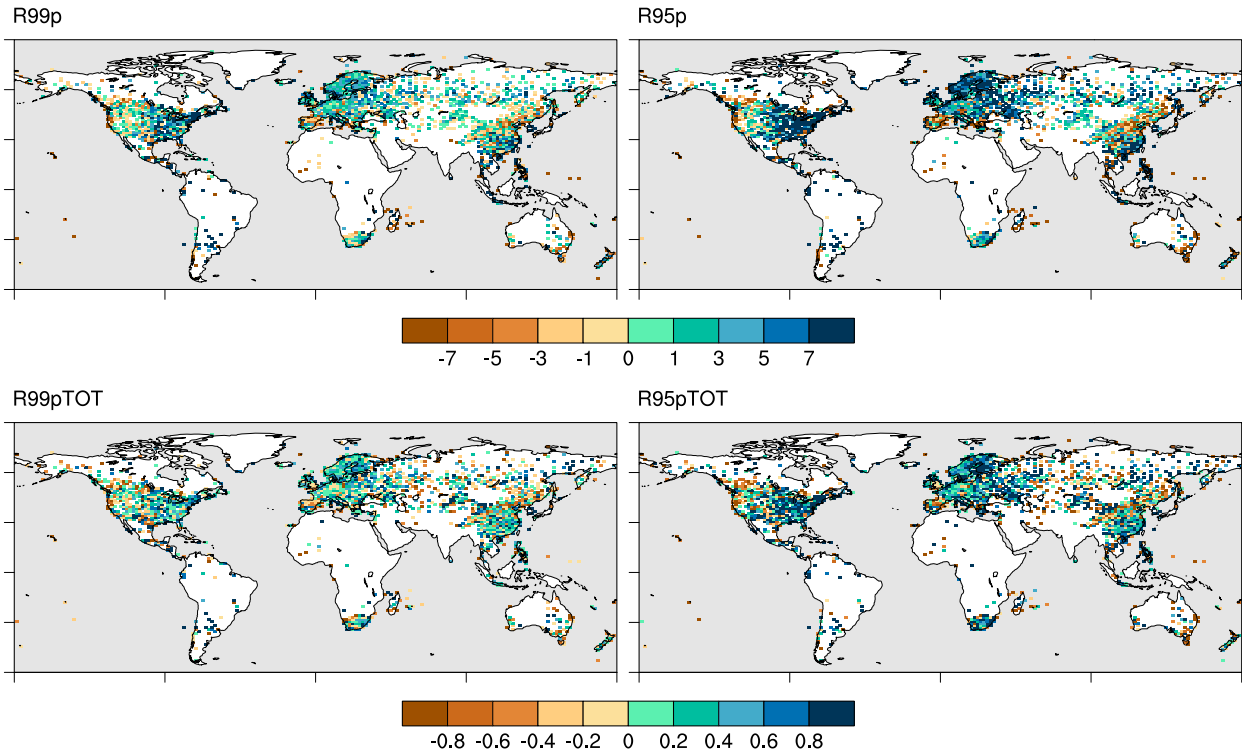

FIG. 1. Observed linear trends in the four extreme precipitation indices—(top left) R99p (mm decade^{-1}), (top right) R95p (mm decade^{-1}), (bottom left) R99pTOT (% decade^{-1}), and (bottom right) R95pTOT (% decade^{-1})—based on the Sun et al. (2021) dataset during 1951–2014. The trends were computed at grid cells with at least 45 years of data available.

Single-signal, two-signal, and three-signal analyses are performed in our study. For single-signal analysis, observed changes are regressed onto the forced response to historical ALL forcing. For two-signal analysis, observed changes are regressed onto responses to anthropogenic forcing (ANT, in which ANT forcing equals to ALL forcing minus natural forcing) and natural (NAT) forcing simultaneously so as to distinguish the individual role of ANT and NAT forcings in driving the observed changes. For three-signal analysis, responses to GHG forcing, AER forcing, and NAT forcing are involved simultaneously in the regression. A residual consistency test is applied to test whether the modeled internal variability is consistent with that observed as represented by regression residuals (Ribes et al. 2013). In these analyses, we assume that the modeled responses of these extreme indices to anthropogenic forcings including GHG and AER forcings and NAT forcing are additive. The basic idea for the assumption is that all of the forcings over the historical period presumably produced small perturbations to the global radiative balance, thus allowing linear decomposition of the combined response. A few studies (e.g., Shiogama et al. 2013; Marvel et al. 2015) have shown that the response to GHG forcing may be different from that due to the AER and NAT forcings. In the former, forcing is via a change in longwave radiation, while for the latter two cases forcing occurs dominantly via changes in shortwave radiation. Despite these differences, we find for our analyzed precipitation indices that the linear decompositions are generally reasonable since the response to ALL forcing is very similar to the sum of those from the GHG, AER, and NAT forcings (not shown).

3. Results

a. *Observed and modeled changes*

Figure 1 presents maps of the observed trends in R99p, R95p, R99pTOT, and R95pTOT during 1951–2014. Although changes in these indices are spatially complex, particularly at small spatial scales, they display increasing trends over most land areas covered by observations. Negative trends are found mainly in a southwest–northeast belt from Southwest China to Northeast China, in western North America, and in southeastern Australia, with the most negative trends occurring to R95p. The spatial patterns of trends in R99pTOT and R95pTOT appear quite consistent with those of percentile-based indices. However, in some areas of East Asia for which R99p and R95p showed a decrease, these two ratio indices show positive values, indicating the increased contribution from heavy precipitation while the total precipitation amount decreases (not shown). This overall picture is consistent with that based on the HadEX2 and HadEX3 observations (Donat et al. 2013; Dunn et al. 2020).

Figure 2 displays the trends in multimodel responses under the ALL, GHG, AER, and NAT forcings for 1951–2014. For R99p and R95p, the spatial patterns of multimodel mean trends in historical ALL forcing simulations are

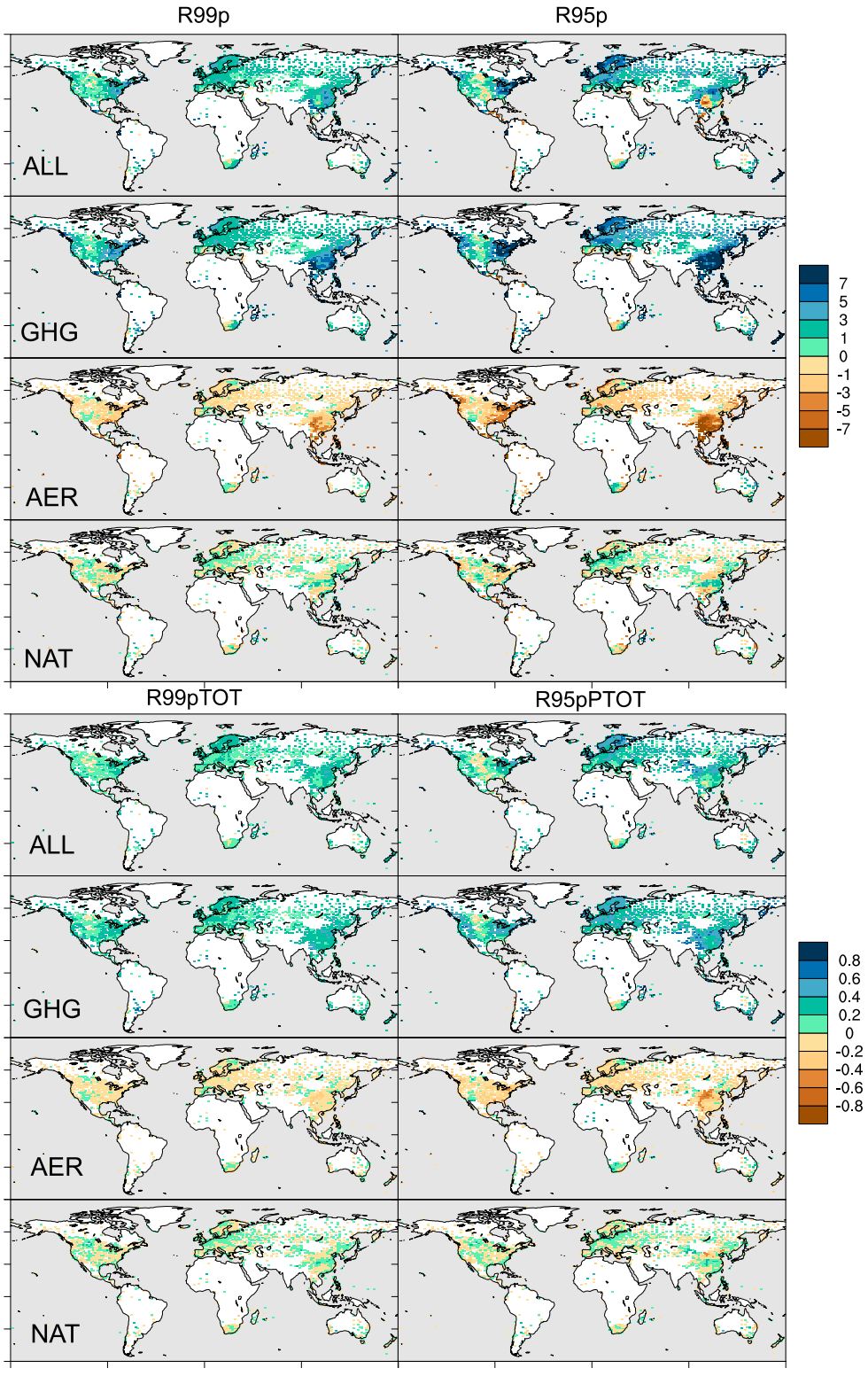

FIG. 2. Model-simulated multimodel mean linear trends under ALL, GHG, AER, and NAT forcings (as labeled) for the four extreme precipitation indices: trends are shown for (top left) R99p (mm decade^{-1}), (top right) R95p (mm decade^{-1}), (bottom left) R99pTOT (% decade^{-1}), and (bottom right) R95pTOT (% decade^{-1}). All simulations are masked by the availability of observations.

generally similar to those seen in the observations, similarly showing increasing trends over most land area covered with the observations. Modeled trends generally show less spatial variability than observed, mainly because the multimodel and multimember ensemble averaging filters out a large part of the internal variability in these extreme indices. The observed decreasing trends in Asia and western North America do not appear in the multimodel ensemble mean trends. Possible reasons include the possibility that the observed decreasing trends are induced by internal climate variability, which would have been smoothed out in the ensemble mean trends, or that they may be associated with regional forcings that are not adequately represented in the CMIP6 forcings. The hist-GHG simulations show similar spatial patterns of trends seen in the historical simulations, but generally with much larger trend magnitudes. In contrast, the hist-aer simulations show decreasing trends almost everywhere, with particularly strong decreases in R95p in Asia. The counteracting effects of greenhouse gases and aerosols on extreme precipitation produce the weaker increasing trends in historical simulations than in hist-GHG simulations in most continents. In Europe, the GHG response is smaller than the ALL response in recent decades because of small aerosol response and positive contribution from NAT forcing. We also observe remarkably larger decreases in these two percentile-based indices, especially R95p, in hist-aer simulations in Southeast Asia and the eastern United States, suggesting an important role of aerosols in extreme precipitation in these regions. The hist-nat simulations exhibit mixed decreasing and increasing trends of small magnitudes, suggesting a limited role of natural forcing in driving changes in the percentile-based precipitation extreme indices. Similar findings are seen in the modeled trends in the two fractional indices R99pTOT and R95pTOT, but with more uniform spatial patterns, resulting from the reduced effect of local precipitation climatology differences after normalizing the percentile-based indices by local annual total precipitation and also from better comparability between observations and models.

Figure 3 shows the 5-yr mean anomaly time series for the observed and modeled precipitation extreme indices averaged over the globe and the continents. It is found that multimodel mean responses to ALL forcings reproduce reasonably well the observed upward trends for all studied indices (red lines), with the observed trends (black lines) falling within the central 90% ranges of the simulated responses of individual model runs to ALL forcings in all regions (light pink envelopes). For all indices, the multimodel mean responses to GHG forcings also simulate upward trends, but the magnitudes tend to be larger than the observed trends in many cases (blue lines). The multimodel mean responses to AER forcings show decreasing trends (brown lines). The NAT forcings responses do not exhibit statistically significant long-term trends in these indices, but they do show weak decadal variations that may reflect the effects of volcanic forcing (green lines).

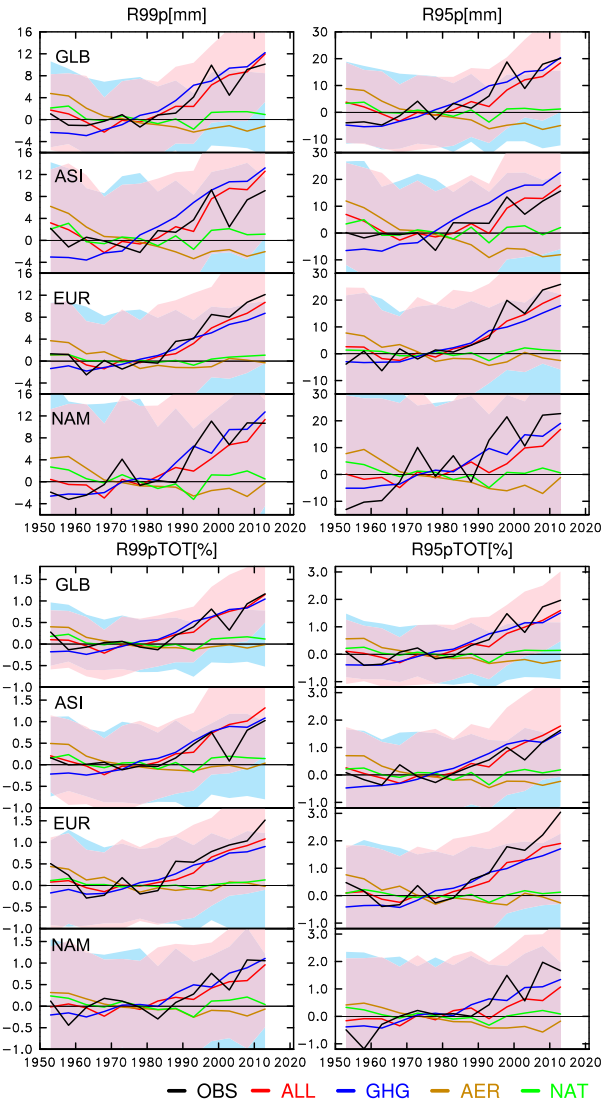

FIG. 3. Time series of 5-yr mean anomalies (relative to the 1961–90 average) for (top left) R99p, (top right) R95p, (bottom left) R99pTOT, and (bottom right) R95pTOT from the Sun et al. (2021) dataset (black lines) and the multimodel response to the ALL (red lines), GHG (blue lines), AER (brown lines), and NAT (green lines) forcings, respectively. The 5%–95% ranges of the individual models under ALL and NAT are given by the light-pink and light-blue shadings, respectively. Results are shown for the globe (GLB), Asia (ASI), Europe (EUR), and North America (NAM), as labeled.

b. Detection results for extreme precipitation changes

Figure 4 shows estimates of the scaling factors from single-signal analyses of the four extreme indices during 1951–2014. For all indices, the best estimates of the scaling factors for the globe and most continents are greater than zero, indicating that the ALL signals are robustly detected at the 5% significance level in the globe and the continents of Northern Hemisphere. In Asia, the best estimates of scaling factors are slightly smaller than unity, indicating that the available CMIP6 models overestimate changes in the extreme precipitation changes, which is in line with the results reported in Dong et al. (2020), who used a shorter period (1958–2012) and CMIP5 simulations. In EUR and NAM, the best estimates of scaling factors are greater than unity and thus indicate an underestimate of model-simulated

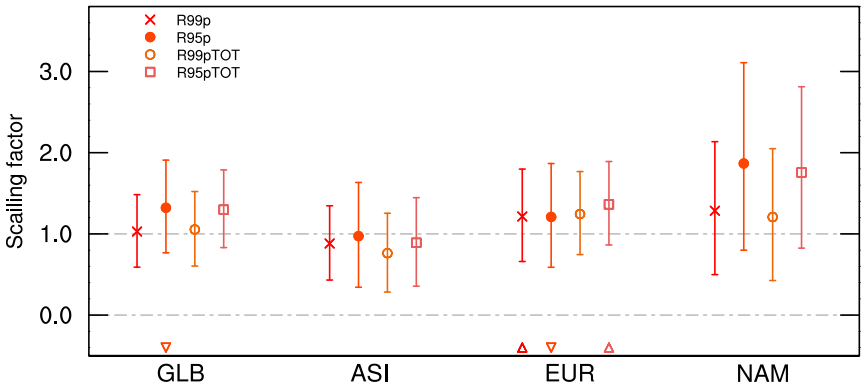

FIG. 4. Best estimates of scaling factors and their 5%–95% confidence intervals from the single-signal analyses for the GLB, ASI, EUR, and NAM series. The observations are regressed onto the model-simulated response to ALL forcing during 1951–2014. The internal variability here is calculated on the basis of the standard deviation of 64-yr chunks of preindustrial control experiments, which are used for the residual consistency test. According to the residual consistency tests, a downward triangle indicates that the model-simulated variability is too low and an upward triangle indicates that the model-simulated variability is too high.

changes when compared with observations. For most cases, the residual consistency test passed except in EUR. In EUR, the model-simulated variability and the regression residual are not consistent for most indices. The triangles at the bottom of Fig. 4 show that the model-simulated internal variabilities (based on the preindustrial control experiments) in R99p and R95pTOT are larger than the observations while those in R95p are smaller than the observations.

Figure 5 shows estimates of the ANT and NAT scaling factors from two-signal analyses. For the two percentile-based indices, the ANT signals can be detected globally and in Northern Hemispheric continental regions, while the NAT signals cannot be detected. The best estimates of ANT scaling factor in GLB and three continental regions are very close to or slightly larger than unity, indicating that the ANT responses from analyzed CMIP6 models are consistent with observations or tend to slightly underestimate the observed changes in these regions. In contrast, the NAT signal cannot be detected in any of the cases. The residual consistency tests for cases where the ANT signals are detected are passed with exceptions for two indices in EUR and for R99p in NAM. Similar findings exist in the two fractional indices; the ANT signals can be detected globally and in Northern Hemispheric continental regions, while the NAT signals cannot be detected. Nevertheless, the confidence intervals of the scaling factors are generally narrower than the two percentile-based indices, especially for the

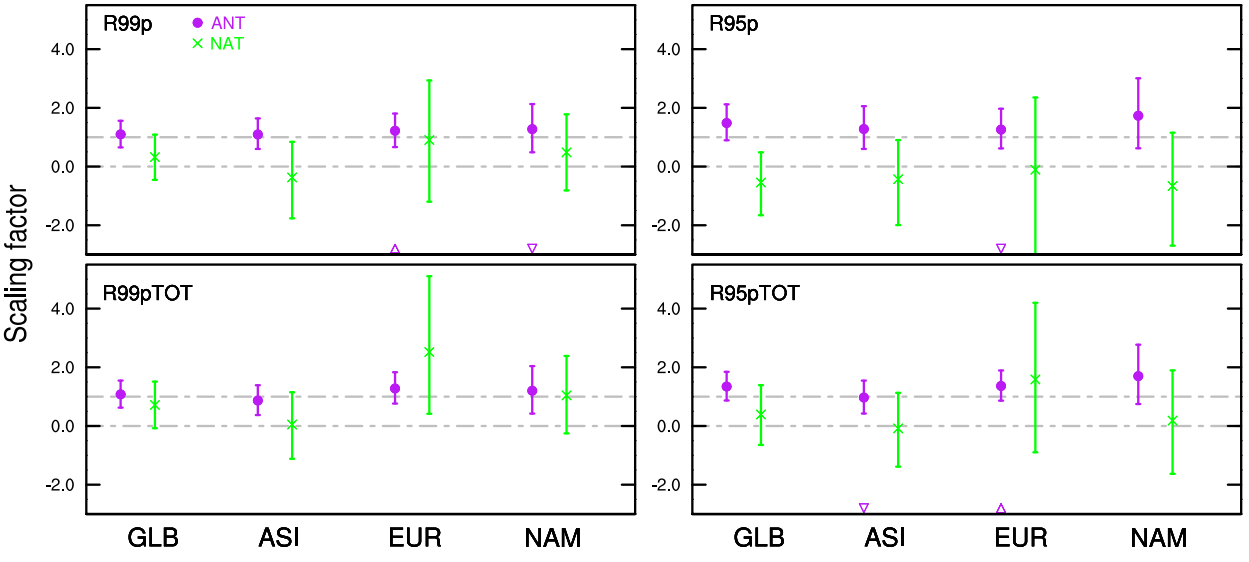

FIG. 5. As in Fig. 4, but for the scaling factors and 5%–95% confidence intervals from the two-signal analyses that involve ANT (purple) and NAT (green).

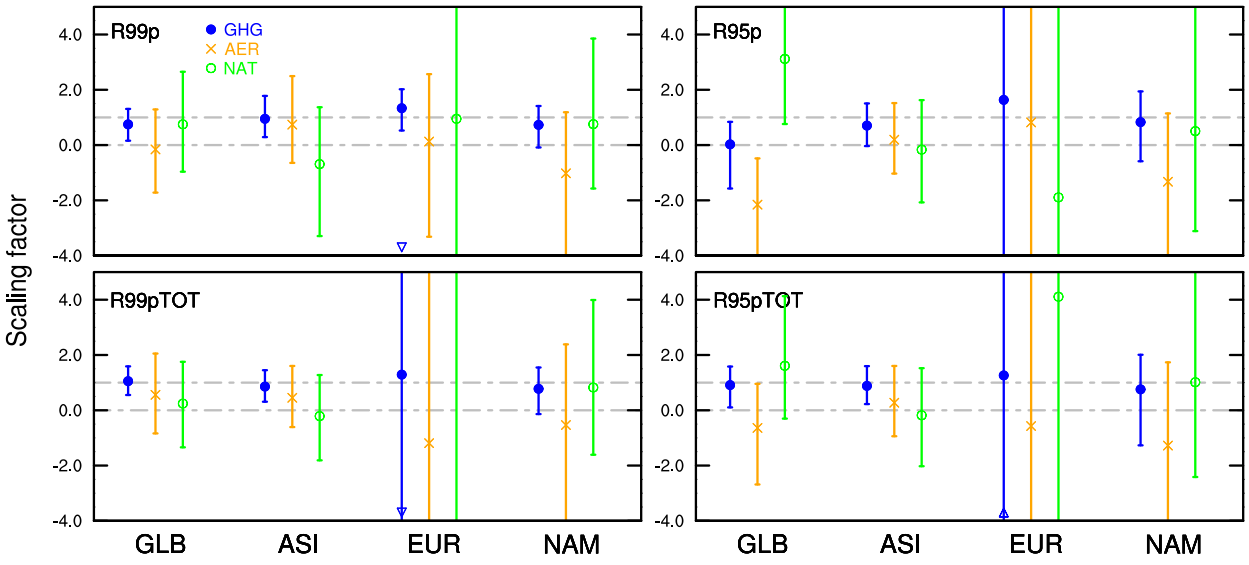

FIG. 6. As in Fig. 4, but for scaling factors and 5%–95% confidence intervals from the three-signal analyses that involve GHG (blue), AER (brown), and NAT (green).

global analysis and in Asia. This suggests that human influence is more detectable in the ratio indices, consistent with our previous findings in Asia (Dong et al. 2020). The best estimates of the ANT scaling factor for R99p and R99pTOT in GLB and three continental regions are quite close to unity, indicating a decent performance of the available CMIP6 in simulating the ANT signals presented in observations.

Estimates of scaling factors for the three-signal analysis are presented in Fig. 6. When the signals of GHG, AER, and NAT are considered simultaneously in the detection analyses, only the GHG signal can be detected in several cases. For R99p, the best estimates of the scaling factors for the GHG signal are generally greater than zero for GLB and the Northern Hemispheric continental regions. For R95p, the GHG signal can be detected only marginally in Asia. For both R99pTOT and R95pTOT, the GHG signals are detected only for GLB and Asia. The AER and NAT signals cannot be detected for most indices in all the regions although the NAT signal seems to be detected at global scale for R95p but with large confidence intervals. Also, to test the influence of the collinearity between the GHG and AER changes on the detection results, we conduct the global detection analyses based on three space dimensions from the 5-yr series in three continents. We find similar results but with the residual consistency test not passed, which may reflect the difficulties of separating the response of extreme precipitation to different forcings against the noise when the GHG, AER, and NAT forcings are all taken into consideration.

4. Conclusions and discussion

In this study, we present a formal detection and attribution analysis of the observed changes in four precipitation extreme indices using the updated observations and the newly available CMIP6 simulations. The four extreme indices are R99p, R95p, R99pTOT, and R95pTOT, representing precipitation totals from heavy precipitation days and the contributions of these extreme precipitation days to annual precipitation totals. We find that the influence of anthropogenic forcing can be detected in these extreme indices globally and in the Northern Hemispheric continental regions. The net influence of anthropogenic greenhouse gases is detectable globally and in some Northern Hemispheric continental regions such as Asia and Europe. In general, the CMIP6 models forced with all-known historical forcings reproduce the observed changes in all four extreme indices, while they may over- or underestimate the observed changes in some continental regions. For example, models substantially underestimate changes in the fractional indices in Europe and overestimate them in Asia.

Analyzing long-term changes in the percentile-based indices such as those used here provides different yet complementary picture to that of annual maximum 1-day or 5-day precipitation. In particular, it reflects the combined effect in the changes in the frequency and intensity of heavy precipitation events. Unfortunately, this advantage comes with an important caveat: it is difficult to separate the effects of changes in the frequency and from those intensity. As different mechanisms may come into play in frequency change and in that intensity change, it can be harder to understand the physics behind the changes in these percentile indices. For example, it becomes difficult to establish a relatively simple relationship between changes in these indices and global warming levels (Li et al. 2020). While changes in Rx1day closely follow the Clausius–Clapeyron relationship both in the observations (Westra et al. 2013; Sun et al. 2021) and in the model simulations (Kharin et al. 2013), the rates of change in R95p per 1°C global warming could be very different from that in R99p for the same region and across regions of different climates.

Acknowledgments. We thank Francis Zwiers and one anonymous reviewer for their very helpful suggestions. Authors Dong,

Sun, and Li are supported by the National Key R&D Program of China (2018YFA0605604, 2018YFC1507702) and the National Science Foundation of China (42025503, 41675074). Authors Min and Kim are supported by the Korea Meteorological Administration Research and Development Program under Grant KMI 2018-03610. We acknowledge the Program for Climate Model Diagnosis and Intercomparison and the Working Group on Coupled Modeling of the World Climate Research Programme (WCRP) for their roles in making the WCRP CMIP multimodel datasets available.

REFERENCES

Alexander, L. V., 2016: Global observed long-term changes in temperature and precipitation extremes: A review of progress and limitations in IPCC assessments and beyond. *Wea. Climate Extremes*, **11**, 4–16, https://doi.org/10.1016/j.wace.2015.10.007.

Allen, M. R., and P. A. Stott, 2003: Estimating signal amplitudes in optimal fingerprinting, Part I: Theory. *Climate Dyn.*, **21**, 477–491, https://doi.org/10.1007/s00382-003-0313-9.

Chen, H., and J. Sun, 2017: Contribution of human influence to increased daily precipitation extremes over China. *Geophys. Res. Lett.*, **44**, 2436–2444, https://doi.org/10.1002/2016GL072439.

Coles, S., 2001: *An Introduction to Statistical Modelling of Extreme Values*. Springer, 209 pp.

de los Milagros Skansi, M., and Coauthors, 2013: Warming and wetting signals emerging from analysis of changes in climate extreme indices over South America. *Global Planet. Change*, **100**, 295–307, https://doi.org/10.1016/j.gloplacha.2012.11.004.

Dey, R., S. C. Lewis, and N. J. Abram, 2019: Investigating observed northwest Australian rainfall trends in Coupled Model Intercomparison Project phase 5 detection and attribution experiments. *Int. J. Climatol.*, **39**, 112–127, https://doi.org/10.1002/joc.5788.

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.

Donat, M. G., and Coauthors, 2013: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *J. Geophys. Res. Atmos.*, **118**, 2098–2118, https://doi.org/10.1002/jgrd.50150.

Dong, S., Y. Sun, and C. Li, 2020: Detection of human influence on precipitation extremes in Asia. *J. Climate*, **33**, 5293–5304, https://doi.org/10.1175/JCLI-D-19-0371.1.

Dunn, R. J., and Coauthors, 2020: Development of an updated global land in situ–based data set of temperature and precipitation extremes: HadEX3. *J. Geophys. Res. Atmos.*, **125**, e2019JD032263, https://doi.org/10.1029/2019JD032263.

Gillett, N., and Coauthors, 2016: The Detection and Attribution Model Intercomparison Project (DAMIP v1.0) contribution to CMIP6. *Geosci. Model Dev.*, **9**, 3685–3697, https://doi.org/10.5194/gmd-9-3685-2016.

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.

Huang, H., J. M. Winter, E. C. Osterberg, R. M. Horton, and B. Beckage, 2017: Total and extreme precipitation changes over the northeastern United States. *J. Hydrometeor.*, **18**, 1783–1798, https://doi.org/10.1175/JHM-D-16-0195.1.

IPCC, 2013: *Climate Change 2013: The Physical Science Basis*. T. F. Stocker et al., Eds., Cambridge University Press, 1535 pp.

Iturbide, M., and Coauthors, 2020: An update of IPCC climate reference regions for subcontinental analysis of climate model data: Definition and aggregated datasets. *Earth Sys. Sci. Data*, **12**, 2959–2970, https://doi.org/10.5194/essd-12-2959-2020.

Jones, G. S., P. A. Stott, and N. Christidis, 2013: Attribution of observed historical near-surface temperature variations to anthropogenic and natural causes using CMIP5 simulations. *J. Geophys. Res. Atmos.*, **118**, 4001–4024, https://doi.org/10.1002/jgrd.50239.

Kharin, V. V., F. W. Zwiers, X. Zhang, and M. Wehner, 2013: Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, **119**, 345–357, https://doi.org/10.1007/s10584-013-0705-8.

Kirchmeier-Young, M. C., and X. Zhang, 2020: Human influence has intensified extreme precipitation in North America. *Proc. Natl. Acad. Sci. USA*, **117**, 13 308–13 313, https://doi.org/10.1073/pnas.1921628117.

Li, C., F. Zwiers, X. Zhang, and G. Li, 2019: How much information is required to well constrain local estimates of future precipitation extremes? *Earth's Future*, **7**, 11–24, https://doi.org/10.1029/2018EF001001.

——, Y. Sun, F. Zwiers, H. Wu, G. Chen, X. Zhang, and D. Wang, 2020: Rapid warming in summer wet bulb globe temperature in China with human-induced climate change. *J. Climate*, **33**, 5697–5711, https://doi.org/10.1175/JCLI-D-19-0492.1.

Li, W., Z. Jiang, X. Zhang, and L. Li, 2018: On the emergence of anthropogenic signal in extreme precipitation change over China. *Geophys. Res. Lett.*, **45**, 9179–9185, https://doi.org/10.1029/2018GL079133.

Lin, L., Z. Wang, Y. Xu, and Q. Fu, 2016: Sensitivity of precipitation extremes to radiative forcing of greenhouse gases and aerosols. *Geophys. Res. Lett.*, **43**, 9860–9868, https://doi.org/10.1002/2016GL070869.

Marvel, K., G. A. Schmidt, D. Shindell, C. Bonfils, A. N. LeGrande, L. Nazarenko, and K. Tsigaridis, 2015: Do responses to different anthropogenic forcings add linearly in climate models? *Environ. Res. Lett.*, **10**, 104010, https://doi.org/10.1088/1748-9326/10/10/104010.

Min, S. K., X. Zhang, 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.

Paik, S., S. K. Min, X. Zhang, M. G. Donat, A. D. King, and Q. Sun, 2020: Determining the anthropogenic greenhouse gas contribution to the observed intensification of extreme precipitation. *Geophys. Res. Lett.*, **4**, e2019GL086875, https://doi.org/10.1029/2019GL086875.

Pendergrass, A. G., F. Lehner, B. M. Sanderson, and Y. Xu, 2015: Does extreme precipitation intensity depend on the emissions scenario? *Geophys. Res. Lett.*, **42**, 8767–8774, https://doi.org/10.1002/2015GL065854.

Rhein, M., and Coauthors, 2013: Observations: Ocean. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 255–315.

Ribes, A., S. Planton, and L. Terray, 2013: Application of regularised optimal fingerprinting to attribution. Part I: Method, properties and idealised analysis. *Climate Dyn.*, **41**, 2817–2836, https://doi.org/10.1007/s00382-013-1735-7.

Schär, C., and Coauthors, 2016: Percentile indices for assessing changes in heavy precipitation events. *Climatic Change*, **137**, 201–216, https://doi.org/10.1007/s10584-016-1669-2.

Shiogama, H., D. A. Stone, T. Nagashima, T. Nozawa, and S. Emori, 2013: On the linear additivity of climate forcing–response relationships at global and continental scales. *Int. J. Climatol.*, **33**, 2542–2550, https://doi.org/10.1002/joc.3607.

Sillmann, J., C. W. Stjern, G. Myhre, and P. M. Forster, 2017: Slow and fast responses of mean and extreme precipitation to different forcing in CMIP5 simulations. *Geophys. Res. Lett.*, **44**, 6383–6390, https://doi.org/10.1002/2017GL073229.

Sun, Q., X. Zhang, F. Zwiers, S. Westra, and L. V. Alexander, 2021: A global, continental, and regional analysis of changes in extreme precipitation., *J. Climate*, **34**, 243–258, https://doi.org/10.1175/JCLI-D-19-0892.1.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192, https://doi.org/10.1029/2000JD900719.

Wan, H., X. Zhang, F. Zwiers, and S.-K. Min, 2014: 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.

Wang, Y., 2015: Air pollution or global warming: Attribution of extreme precipitation changes in eastern China—Comments on ''Trends of extreme precipitation in Eastern China and their possible causes.'' *Adv. Atmos. Sci.*, **32**, 1444–1446, https://doi.org/10.1007/s00376-015-5109-4.

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.

Wu, P., N. Christidis, and P. Stott, 2013: Anthropogenic impact on Earth's hydrological cycle. *Nat. Climate Change*, **3**, 807–810, https://doi.org/10.1038/nclimate1932.

Zhang, X., and Coauthors, 2011: Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 851–870, https://doi.org/10.1002/WCC.147.

——, 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.