Role of Natural Climate Variability in the Detection of Anthropogenic Climate Change Signal for Mean and Extreme Precipitation at Local and Regional Scales

JEAN-LUC MARTEL

École de Technologie Supérieure, Université du Québec, Montreal, Quebec, Canada

ALAIN MAILHOT

Eau Terre Environnement Research Centre, Institut National de la Recherche Scientifique, Quebec City, Quebec, Canada

FRANÇOIS BRISSETTE AND DANIEL CAYA

École de technologie supérieure, Université du Québec, Montreal, Quebec, Canada

(Manuscript received 28 April 2017, in final form 7 February 2018)

ABSTRACT

Climate change will impact both mean and extreme precipitation, having potentially significant consequences on water resources. The implementation of efficient adaptation measures must rely on the development of reliable projections of future precipitation and on the assessment of their related uncertainty. Natural climate variability is a key uncertainty component, which can result in apparent decadal trends that may be greater or lower than the long-term underlying anthropogenic climate change trend. The goal of the present study is to assess how natural climate variability affects the ability to detect the climate change signal for mean and extreme precipitation. Annual and seasonal total precipitation are used as indicators of the mean, whereas annual and seasonal maximum daily precipitation are used as indicators of extremes. This is done using the CanESM2 50-member and CESM1 40-member large ensembles of simulations over the 1950–2100 period. At the local scale, results indicate that natural climate variability will dominate the uncertainty for annual and seasonal extreme precipitation going up to the end of the century in many parts of the world. The climate change signal can, however, be reliably detected much earlier at the regional scale for extreme precipitation. In the case of annual and seasonal total precipitation, the climate change signal can be reliably detected at the local scale without resorting to a regional analysis. Nonetheless, natural climate variability can impede the detection of the anthropogenic climate change signal until the middle to late century in many parts of the world for mean and extreme precipitation.

1. Introduction

Research conducted in the past decades has emphasized human influence on the climate system through anthropogenic emissions of greenhouse gases (IPCC 2013). It is also expected that global climate warming will induce significant changes in many parts of the world in the distribution of extremes such as extreme precipitation events, droughts, and floods. To ensure public safety, the most important infrastructures are typically designed based on an estimate of the recurrence likelihood of a specific extreme precipitation event (e.g., the 100-yr storm). This estimate is itself usually based on available historical annual daily maxima data. Since such infrastructures often have typical lifespans exceeding 75 years, the potential impact of the anthropogenic climate change signal (referred to as the *climate change signal* hereafter) on extreme precipitation events has important implications for design practice and public safety.

While the climate change signal needs to be accounted for in design practice, consideration also needs to be given to the inherent chaotic nature of the climate system (i.e., the unforced variability that naturally appears in the climate system, and which will be hereafter referred to as *natural variability*). There are many indications that natural variability may mask the climate change signal for short- and long-term precipitation at both the local and regional scales (Deser et al. 2012a,b, 2014; Fischer and Knutti 2014; Fischer et al. 2013, 2014; Giorgi and Bi 2009; Hawkins and Sutton 2011, 2012; King et al. 2015; Maraun 2013; Mora et al. 2013; Thompson et al. 2015; Sanderson et al. 2018). A good example of how natural variability can conceal the climate change signal at the decadal scale is the hiatus in the rise of the global mean surface temperature observed between 1998 and 2012 (Hawkins et al. 2014; IPCC 2013).

To convince policy makers of the importance of adapting infrastructures to climate change, it is crucial to better understand and explain the influence of natural variability on the climate system. However, the ability to assess natural variability is strongly hampered by the short length of available historical records for key weather variables. An alternative approach is to study it through simulations of a general circulation model (GCM) or an Earth system model (ESM). Most published studies use many GCMs and/or ESMs [e.g., models from phase 5 of the Coupled Model Intercomparison Project (CMIP5); Taylor et al. 2012] to gather a large enough ensemble of models to perform such analyses (Fischer et al. 2014; Giorgi and Bi 2009; Hawkins and Sutton 2012; IPCC 2013; King et al. 2015; Maraun 2013; Mora et al. 2013). In many such studies, the concept of time of emergence (TOE) is defined to assess the moment when the climate change signal emerges from natural variability (Giorgi and Bi 2009; Hawkins and Sutton 2012; IPCC 2013; King et al. 2015; Maraun 2013). Generally, it is defined through a signal-to-noise (S/N) ratio based on a measure of the anthropogenic climate change signal (S) and some measure of natural variability (i.e., noise; N). The TOE is then estimated for each simulation (either from an individual model or from different models), and then some measure of the TOE distribution over all simulations (e.g., mean or median TOE) is used.

Most of these studies look at mean climate variables, and few analyze precipitation extremes under such a framework (Fischer et al. 2014; King et al. 2015; Maraun 2013). Most of them, though, share a common limitation in their ability to separate natural variability from intermodel variability (uncertainties) since they combine simulations from various models. To correctly assess the sole impact of natural variability, one must first disentangle the intermodel uncertainties from natural variability (Fischer et al. 2013; Kay et al. 2015).

This can be done using a large ensemble of climate simulations from a single GCM or ESM to assess the simulated natural variability (Fischer et al. 2013; Kay et al. 2015). To date, quite a few studies of this kind using large ensembles have been conducted on mean precipitation (as well as other mean climate variables; Deser et al. 2012a,b, 2014; Fischer et al. 2014; Kay et al. 2015; Thompson et al. 2015; Sanderson et al. 2018). These studies showed that natural variability has a substantial influence over mean precipitation trends at the local and regional scales.

A relatively limited number of studies have been conducted on the influence of natural variability on the detection of climate change signals for precipitation extremes, based on large ensembles of climate simulations from a single model (Fischer and Knutti 2014; Fischer et al. 2013, 2014). One of the key findings in these studies is that the signal for precipitation extremes is more robust than that for mean precipitation, indicating a potential earlier emergence of the climate change signal from natural variability in many regions. However, the impact of natural variability on the probability of detecting a climate change signal at the local and regional scales remains a complex problem.

Accordingly, the main objective of the present study is to look at how natural variability could impair the detection of the climate change signal for both precipitation means and extremes at the local and regional scales. This is addressed using two large ensembles of 150-yr climate simulations. The models and methods used are developed in section 2. A comparison of model data against observations, as well as results for both mean and extreme precipitation, is presented in section 3 and discussed in section 4. Concluding remarks are presented in section 5.

2. Datasets and methods

a. The CanESM2 and CESM1 large ensembles

The first large ensemble used in the present study is composed of 50 climate simulations with a 2.8° resolution, derived from the Canadian Centre for Climate Modeling and Analysis (CCCma) second-generation Canadian Earth System Model (CanESM2; Arora et al. 2011; Sigmond and Fyfe 2016). Five simulations covering the 1850–1950 historical period were performed to generate five different states of the ocean in 1950. Then, 10 coupled ocean–atmospheric simulations were run from each of these five historical simulations using randomly perturbed initial conditions (in 1950),

for a total of fifty 150-yr simulations over the 1950–2100 period. Because of the chaotic nature of the climate system, small perturbations in the initial 1950 conditions quickly resulted in different atmospheric states after a few days following the perturbation (Deser et al. 2012a; IPCC 2013). The simulations were conducted from 1950 to 2006 using historical greenhouse gas concentrations data. From 2006 on, the representative concentration pathway scenario resulting in an $8.5\,W\,m^{-2}$ increase in the atmospheric radiative forcing in 2100 (i.e., RCP8.5) was used (IPCC 2013).

The second large ensemble is made up of 40 climate simulations with a 1° resolution, derived from the Community Earth System Model version 1 (CESM1) coupled with CAM5.2 for the atmospheric component (Kay et al. 2015). The covered period ranges from 1920 to 2100, but only the 1950–2100 period was analyzed in this study to allow a direct comparison with the CanESM2 large ensemble. The same RCP8.5 scenario was considered from 2006 until the end of the simulation period. Aside from the model structure, the main differences between the two ensemble simulations lie in the spatial resolutions (2.8° for CanESM2 vs 1° for CESM1) and the initial ocean conditions (five different ocean states for CanESM2 vs a single ocean state for CESM1).

b. Precipitation indices

Two precipitation indices were used in this study: 1) the total wet-day precipitation (PRCPTOT) from days ≥ 1 mm and 2) the max 1-day precipitation amount (RX1day). Both indices were analyzed at the annual and seasonal scales for winter [December–February (DJF)] and summer [June–August (JJA)].

These two indices are recommended by the Expert Team on Climate Change Detection and Indices (ETCCDI; Klein Tank et al. 2009; Sillmann et al. 2013a,b; Zhang et al. 2011). Using the same indices allows a comparison and further discussion of the results obtained here with observed datasets (Donat et al. 2013a,b) and with other climate change studies (e.g., Fischer and Knutti 2014; Fischer et al. 2013, 2014; IPCC 2013). Having both mean and extreme indices furthers our understanding of the role of natural variability in the climate change signal.

c. Probability of detecting the climate change signal at the local scale

Eleven periods (1950–2000, 1950–2010, 1950–2020, . . . , 1950–2100) were considered to investigate annual and seasonal time series of PRCPTOT and RX1day indices at each grid point of both ensembles. The nonparametric Theil–Sen estimator (Sen 1968), which corresponds to the median of the slopes over all pairs of sample points, was used to estimate the slope of a linear trend over each period for all 50 members. This estimator was mainly used to compare observed trends with the simulated trends of both ensembles (see section 3a). The local trend significance of each grid point was estimated using the nonparametric Mann–Kendall test (Kendall 1975) at a 95% confidence level:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{m} \text{sign}(x_i - x_j), \qquad (1)$$

where x is the index value (i.e., PRCPTOT or RX1day) at time i and j, with sign() being equal to $+1$ if x_i is greater than x_j and -1 if x_i is smaller than x_j. Also, S represents the number of times x_i is greater than x_j minus the number of times x_i is smaller than x_j. The sign of S also indicates the sign of the trend.

The Mann–Kendall test was used to characterize the climate change signal at the local scale (i.e., over a given grid point without considering regional spatial correlations) over the corresponding periods. The probability of detecting the climate change signal for a given period was then defined by the percentage of members with a significant trend of a given sign (positive or negative) at the 95% confidence level. The 11 predefined periods allowed the investigation of the probability of locally detecting the climate change signal over the 11 periods.

An advantage of using these two tests is that they do not make assumptions about the distribution of the analyzed variable and they can be applied to both observed and simulated series. When dealing with recorded series, the Theil–Sen estimator and Mann–Kendall test are often used to detect the nonstationarity associated with the climate change signal (Donat et al. 2013a; Lins and Slack 1999; Westra et al. 2013).

The 90% detection decade (90%DD) was defined as the decade ending the first period (e.g., decade 2060–70 of the 1950–2070 period), where at least 45 out of 50 members for CanESM2 or 36 out of 40 members for CESM1 (therefore, a 90% probability of detecting the trend among the various simulations) had a significant trend (95% confidence level) of the same sign (either positive or negative) over that period and over all subsequent periods up to the 1950–2100 period (in our example the trend must remain over the 1950–2080, 1950–2090, and 1950–2100 periods). The 45 members of the CanESM2 (36 members for CESM1) thresholds were chosen such that the probability of having 5 members (4 members) with a nonsignificant trend due to type II errors (false negatives) was less than 5%. The 90%DD was estimated using the annual index series and the seasonal index series. The 90%DD is, to some extent, related to the time of
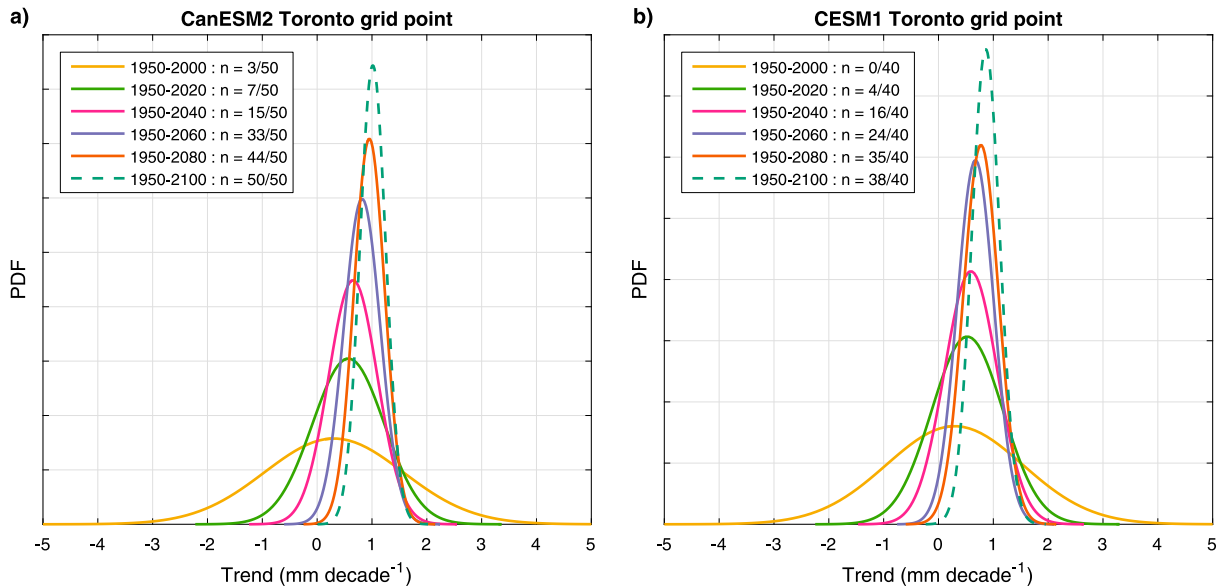
FIG. 1. Gaussian distribution of the nonparametric Theil–Sen trend estimators for the grid point corresponding to the city of Toronto for the RX1day index estimated from the (a) 50-member CanESM2 ensemble and (b) 40-member CESM1 ensemble. Continuous lines correspond to the distributions for periods with fewer than 45 members with a 95% significant trend and dashed lines correspond to periods with 45 or more members with a 95% significant trend of the same sign. The number of members with a significant trend (*n*) is shown in the legend for each period.

emergence used in previous studies (Giorgi and Bi 2009; Hawkins and Sutton 2012; IPCC 2013; King et al. 2015; Maraun 2013). Results shown hereafter, based on the local trends analysis, are referred to as the "local scale."

An example of an estimated 90%DD is shown in Fig. 1 for the land grid point containing the city of Toronto (Ontario, Canada) for the RX1day index. In this example, the 90%DD is the 2090–2100 decade. Two main features can be observed in Fig. 1 for both ensembles (CanESM2 in Fig. 1a and CESM1 in Fig. 1b) as the length of data increases: 1) the distribution becomes narrower, and 2) there is a shift in the central value of the distribution. This suggests that, when using a smaller number of decades, natural variability has a greater influence on the detected trend resulting in a wider distribution. However, when a greater number of decades is used, the distribution becomes narrower as the signal increases and the influence of natural variability on the trends decreases. Moreover, as the climate change signal becomes stronger, the central value of the distribution shifts to the right.

There was a possibility of inaccurate results being obtained when the estimated 90% probability of detecting the climate change signal was reached near the end of the 1950–2100 period, since it could theoretically have fallen below the 90% threshold in the decades after 2100. This situation was investigated by looking at the probability of a grid point that had reached the 90% probability threshold before 2100 dropping back below

the threshold of 45/50 members for CanESM2 or 36/40 members for CESM1 in any subsequent periods. The probability of occurrence of such cases was estimated to average 0.0103 for PRCPTOT and 0.0039 for RX1day over all land grid points (for both ensembles and for annual and seasonal scales). It would therefore be very unlikely that grid points with a reported 90%DD before 2100 would be changed beyond 2100.

d. Probability of detecting the climate change signal at the regional scale

The methodology described in section 2c does not take into consideration a possible spatial correlation between neighboring grid points. It is expected that if gridpoint values are spatially correlated, this could result in earlier 90%DD than expected at the local scale.

To investigate regional trends, a field significance test combined with a resampling approach by bootstrap is performed over each grid point of both ensembles. The method proposed for assessing the regional trend significance is also described in Douglas et al. (2000), Kiktev et al. (2003), and Westra et al. (2013). Figure 2 describes the method through an example using the grid point containing the city of Toronto for the RX1day index with the CanESM2 ensemble.

Regions were defined in CanESM2 by using the nearest neighboring grid points for each grid point ($3 \times 3 = 9$ total grid points). In CESM1, a relatively similar surface
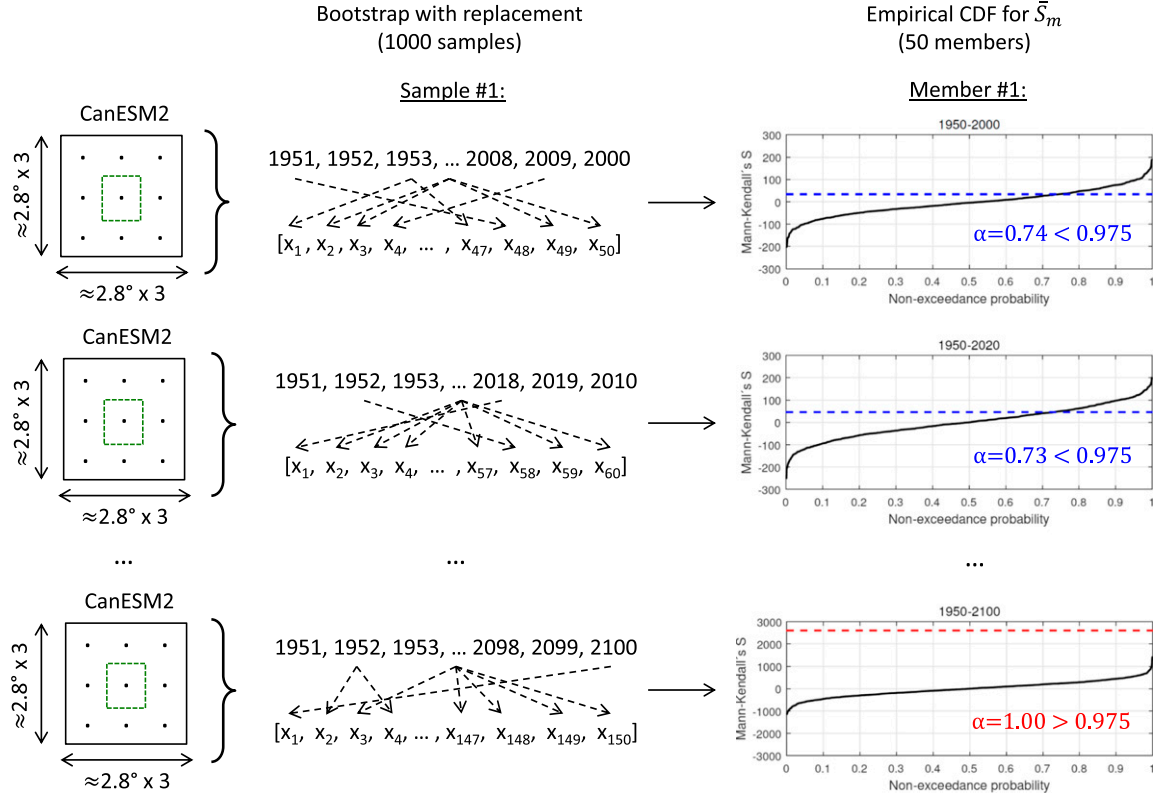
FIG. 2. Summary of the methodology used to assess regional trend (see section 2d). The grid point containing the city of Toronto in CanESM2 is used in this example.

area was selected to allow a fair comparison with CanESM2 results. A total of 81 grid points ($9 \times 9 = 81$ grid points) were taken for each region. The result of the test was associated to the middle grid point of each region.

The regional average Mann–Kendall's S ($\overline{S}_m$) is then computed as the average of the local trend values from each grid point within the region:

$$\overline{S}_m = \frac{1}{m} \sum_{k=1}^{m} S_k, \qquad (2)$$

where S_k is the Mann–Kendall S [see Eq. (1)] for the kth grid point in a region of m grid points ($m = 9$ for CanESM2 and $m = 81$ for CESM1).

To determine whether or not the regional trend is significant, a bootstrap resampling approach was performed (Douglas et al. 2000). For each bootstrap sample, a sample of years with replacement corresponding to the period analyzed (i.e., 1950–2000, 1950–2010, ... 1950–2100) was randomly generated (Fig. 2). The same sample of years was then used for each grid points of the region to compute the Mann–Kendall's S metric [Eq. (1)]. Using the same years allows us to keep track of the spatial correlation between neighboring grid points. The regional average Mann–Kendall $\overline{S}_m$ is then

computed using Eq. (2). This procedure is repeated 1000 times and sorted in ascending order of S assigning a nonexceedance probability based on the Weibull plotting position formula:

$$P = \frac{r}{B + 1}, \qquad (3)$$

where r is the rank of each sample and $B = 1000$ (1000 samples). The 95% confidence level of the empirical CDF obtained is then defined as the Mann–Kendall $\overline{S}_m$ associated with the 25th rank ($\alpha = 0.025$; negative significant trend) and the 975th rank ($\alpha = 0.975$; positive significant trend).

This methodology is then repeated using all available members of both ensembles for each of the 11 periods. As for the local trend analysis described in section 2c, the 90%DD is defined as the decade ending the first period where at least 45 out of 50 members for CanESM2 (36 out of 40 members for CESM1) had a significant trend at the 95% confidence level of the same sign over that period and over all subsequent periods. Finally, the methodology is reproduced over all grid points using the same sample of years for the bootstrap. Results shown hereafter,

TABLE 1. List of the 21 regions used in this study. The number of land grid points available in CanESM2, CESM1, HadEX2, and GHCNDEX gridded datasets for the PRCPTOT and RX1day indices over the 1950–2010 period are also shown and will be further discussed in section 3. Only HadEX2 and GHCNDEX grid points with at least 40 years of data available over the 1950–2010 period were considered.

| | | | | Number of land grid points | | | | | |
| | | | | CanESM2 | CESM1 | HadEX2 | | GHCNDEX | |
Name	Acronym	Latitude (°)	Longitude (°)	Both indices	Both indices	PRCPTOT	RX1day	PRCPTOT	RX1day
Australia	AUS	45°–11°S	110°–155°E	89	587	42	18	108	86
Amazon basin	AMZ	20°S–122°N	82°–34°W	127	856	32	9	39	15
Southern South America	SSA	56°–20°S	76°–40°W	70	462	50	23	36	17
Central America	CAM	10°–30°N	116°–83°W	32	183	29	17	35	14
Western North America	WNA	30°–60°N	30°–103°W	89	553	75	59	109	84
Central North America	CNA	30°–50°N	103°–85°W	41	301	43	42	68	85
Eastern North America	ENA	25°–50°N	85°–60°W	34	215	29	27	52	47
Alaska	ALA	60°–72°N	170°–103°W	68	503	71	7	80	10
Greenland	GRL	50°–85°N	103°–10W	186	1272	88	14	84	15
Mediterranean Basin	MED	30°–48°N	10°W–40°E	64	460	55	35	76	45
Northern Europe	NEU	48°–75°N	10°W–40°E	85	569	67	67	110	103
Western Africa	WAF	12°S–18°N	20°W–22°E	71	551	10	1	0	0
Eastern Africa	EAF	12°S–18°N	22°–52°E	82	571	5	1	0	0
Southern Africa	SAF	35°–12°S	10°W–52°E	66	417	35	19	47	22
Sahara	SAH	18°–30°N	20°W–65°E	127	743	11	1	7	1
Southeast Asia	SEA	11°S–20°N	95°–155E	41	293	22	13	21	8
East Asia	EAS	20°–50°N	100°–145°E	103	693	95	67	140	90
South Asia	SAS	5°–30°N	65°–100°E	58	391	46	25	21	6
Central Asia	CAS	30°–50°N	40°–75°E	77	557	78	18	95	26
Tibet	TIB	30°–50°N	75°–100°E	63	438	62	22	81	12
Northern Asia	NAS	50°–70°N	40°–180°E	296	1995	277	119	420	203
Land grid points	LGP	—	—	1869	12610	1222	604	1629	889

based on the regional trends described in this section, are referred to as the "regional scale."

The proposed regional trend analysis is based on the hypothesis that PRCPTOT and RX1day annual series are temporally uncorrelated. The median value of the lag-1 autocorrelation coefficient across all land grid points over the 1950–2100 period for annual values (similar values for DJF and JJA) was equal to 0.011 for the PRCPTOT index and −0.010 for the RX1day index in the CanESM2 ensemble and 0.046 and −0.004 for the CESM1 ensemble. Autocorrelations were computed on the residuals from a linear regression. These small values suggest that, on average, the hypothesis of temporal independence is valid for both indices. Nonetheless, the field significance resampling approach was also performed using a moving-block bootstrap method to account for autocorrelations (Wilks 1997, 2011). A moving block of 2 years was used in the bootstrapping (which was above the median obtained for both indices and both ensembles). The results were consistent with that obtained under the hypothesis of temporal independence (not shown for conciseness).

e. Global region analysis

The analyses described in the previous two sections were performed globally and then using the 21 geographical

regions listed in Table 1 and shown in Fig. 3. These 21 geographical regions were also used by Giorgi and Francisco (2000), Sanderson et al. 2018, and Sillmann et al. 2013a,b. An analysis of the combined land grid points from these 21 regions (LGP; excluding Antarctica) is also shown.

3. Results

a. Representation of natural variability in CanESM2 and CESM1

Since the representation of natural variability in CanESM2 (resolution of 2.8° latitude × 2.8° longitude) and CESM1 (resolution of 1° latitude × 1° longitude) is a key element of the present study, variability in trends in both ensembles is compared to corresponding values in the observed values from the climate extremes gridded datasets of the Hadley Centre (HadEX2; Donat et al. 2013a; resolution of 2.5° latitude × 3.75° longitude) and the Global Historical Climatology Network (GHCNDEX; Donat et al. 2013b; resolution of 2.5° latitude × 2.5° longitude). These two datasets have different spatial and temporal coverage due to the different data sources used and quality control performed (Dittus et al. 2015). There is also a larger number of grid points available for the
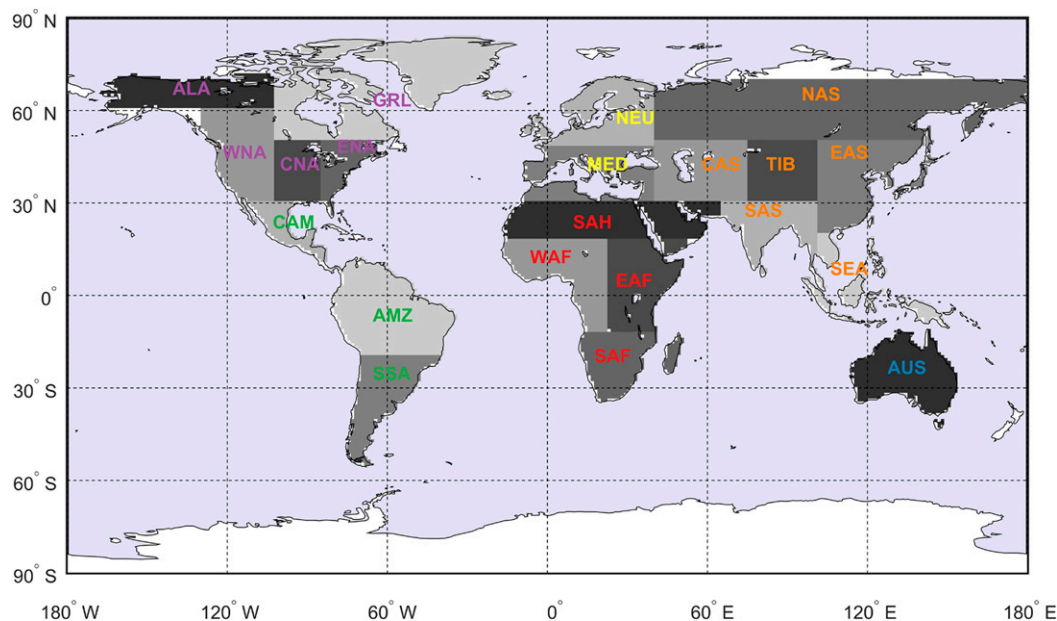
FIG. 3. Map of the 21 geographical regions used in this study. Label colors represent the respective continents: Oceania (blue), South America (green), North America (purple), Europe (yellow), Africa (red), and Asia (orange). Only land grid points were considered in the regional analysis (see Table 1) (Giorgi and Francisco 2000; Sillmann et al. 2013a,b). Grid points correspond to CESM1 grid points.

PRCPTOT index compared to the RX1day index in both the HadEX2 and GHCNDEX datasets due to the interpolation technique used to create these datasets (Donat et al. 2013a, 2013b).

Only grid points with at least 40 (out of 60) years over the 1950–2010 period were considered for the observed datasets (resulting in a total of 1222 grid points and 1629 grid points for the PRCPTOT index and 604 grid points and 889 grid points for the RX1day index for HadEX2 and GHCNDEX respectively). The numbers of land grid points within each of the 21 analyzed regions analyzed are shown in Table 1.

The performance of the CanESM2 and CESM1 ensembles is first assessed through the comparison of the 60-yr annual mean and annual standard deviation (1950–2010) of the PRCPTOT and RX1day indices with the HadEX2 and GHCNDEX datasets (Figs. 4 and 5). For both ensembles, the median of the distribution of annual mean and annual standard deviation values (i.e., one value for each member over the 1950–2010 period) at each grid point was considered.

As shown in maps on the left-hand side of Fig. 4, the spatial distribution of the annual mean PRCPTOT values is globally well reproduced by both ensembles when compared to the HadEX2 and GHCNDEX datasets. Similarly, both ensembles capture relatively well the observed spatial pattern of annual standard deviation as shown by the maps on the right-hand side of Fig. 4.

Mean annual values of the RX1day index (maps on the left-hand side of Fig. 5) are generally underestimated by both ensembles when compared to the observed datasets. Such results were expected, however, because of the spatial mismatch between the ensembles resolution and the smoothed grid point estimates constructed in the HadEX2 and GHCNDEX datasets (Sillmann et al. 2013a). However, the interannual variability as estimated by the annual standard deviation is well captured by both CanESM2 and CESM1 ensembles (maps on the right-hand side of Fig. 5).

Trends estimated by the Theil–Sen estimator and the Mann–Kendall test for the PRCPTOT and RX1day annual time series over the 1950–2010 period were also compared. Figure 6 (PRCPTOT) and Fig. 7 (RX1day) show maps of land grid points comparing local linear trend values from HadEX2 and GHCNDEX datasets to the member with the smallest, median, and largest global trend (defined as the median of the distribution of trends over all grid points) for both ensembles.

As seen for the PRCPTOT index (Fig. 6) and RX1day index (Fig. 7), a larger number of grid points displayed a significant trend for PRCPTOT than for RX1day (44.5% vs 16.3% for HadEX2 and 45.8% vs 14.8% for GHCNDEX). A similar behavior was observed for individual members of both ensembles. However, as shown in Figs. 6 and 7, there is a much smaller fraction of grid points with a significant trend in the different

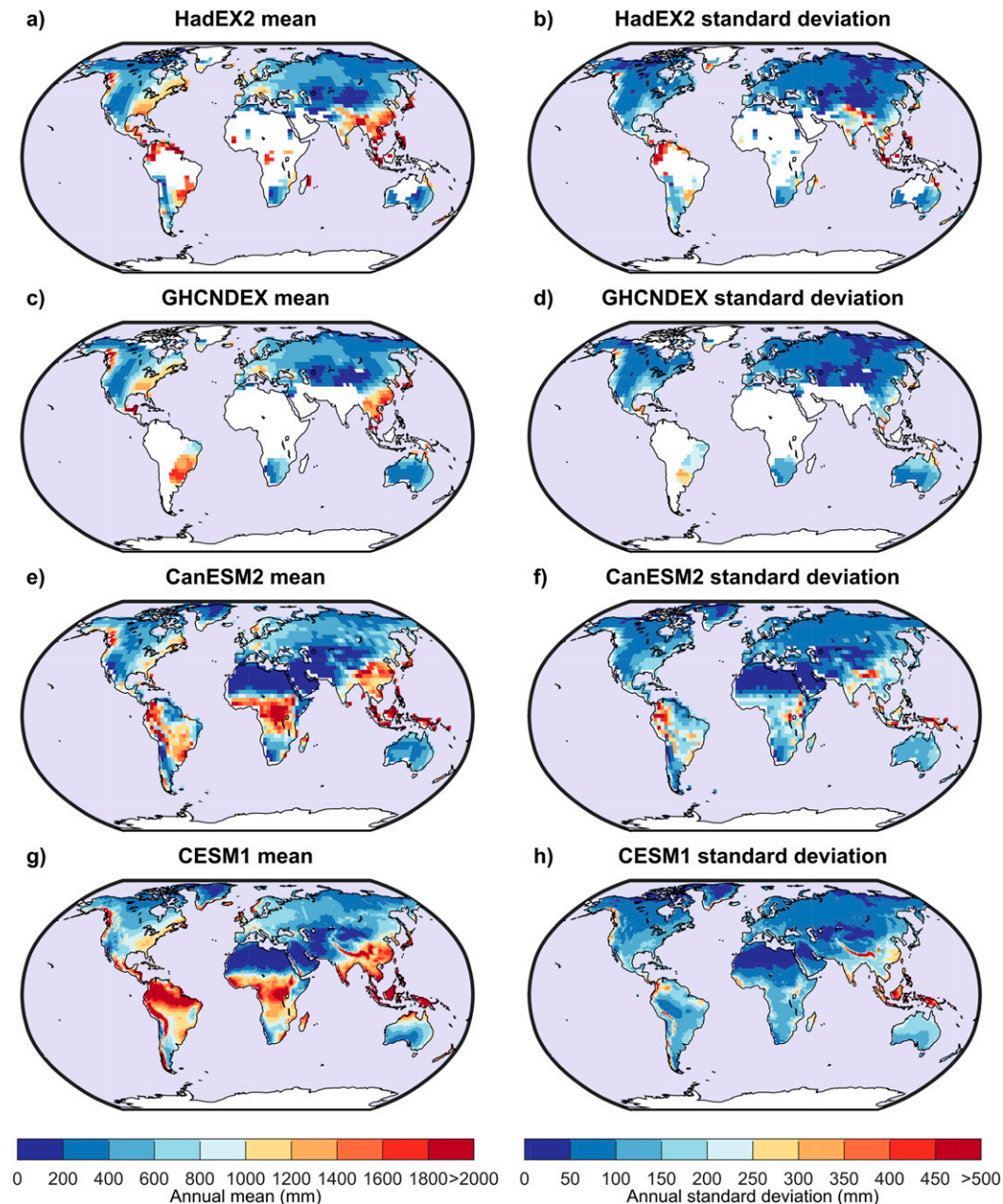## Mean and standard deviation in annual total wet-day precipitation (PRCPTOT)

FIG. 4. Comparison of the (left) mean and (right) standard deviation of annual total wet-day precipitation (PRCPTOT) over the 1950–2010 period. Shown are the observed (a),(b) HadEX2 and (c),(d) GHCNDEX datasets, and the median value over the (e),(f) CanESM2 50-member ensemble and (g),(h) CESM1 40-member ensemble.

members of both ensembles as compared with observations when comparing the same areas. These results outline the stronger influence of natural variability at the local scale for the RX1day index and the ability of the two ensembles to reproduce this behavior. The selected individual members also highlight the large range of possible local trends (individual grid points). This

range is due to the uncertainty related to natural variability, which can even span negative and positive trends at a given grid point for various members.

b. PRCPTOT index

An analysis of the 90%DD for annual and seasonal total precipitation (using the PRCPTOT index) allows
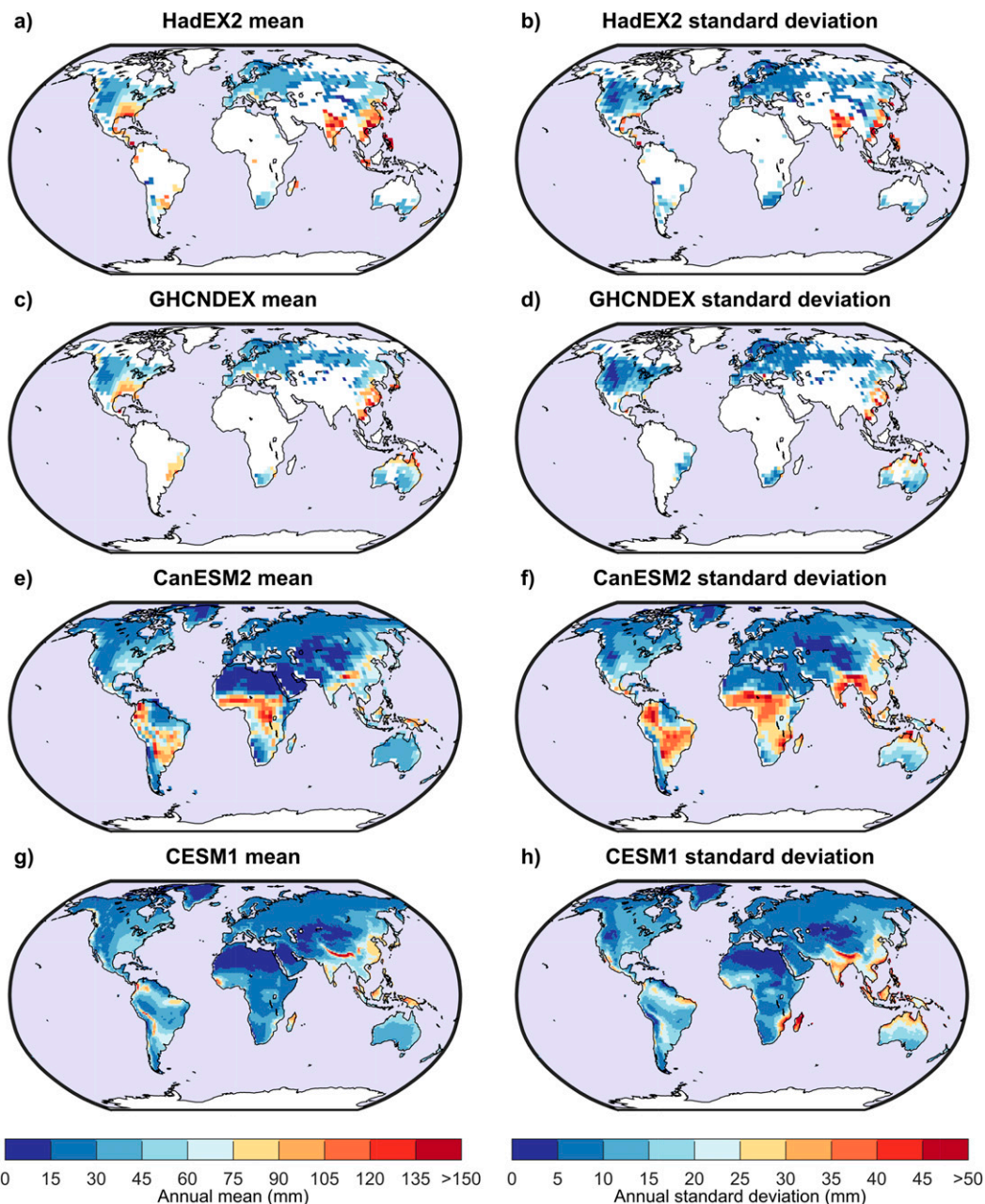
FIG. 5. As in Fig. 4, but for RX1day.

an overview of how natural variability affects the detection of the climate change signal in both the CanESM2 and CESM1 ensembles. Figure 8 (local scale) and Fig. 9 (regional scale) show maps of the decade in which the PRCPTOT index reaches 90%DD. Figure 10 (local scale) and Fig. 11 (regional scale) give a more detailed analysis of these results over the 21 geographical regions listed in Table 1.

A global comparison between Figs. 8 and 9 suggests that there is a relatively good agreement between both ensembles for both the annual (Y) and seasonal scales (DJF, JJA). Figure 8 indicate that the PRCPTOT 90% DD based on local trends occurs before the end of the century over large fractions of ocean and land surface areas, especially at higher latitudes and over the tropics. The seasonal analysis of 90%DD for DJF and

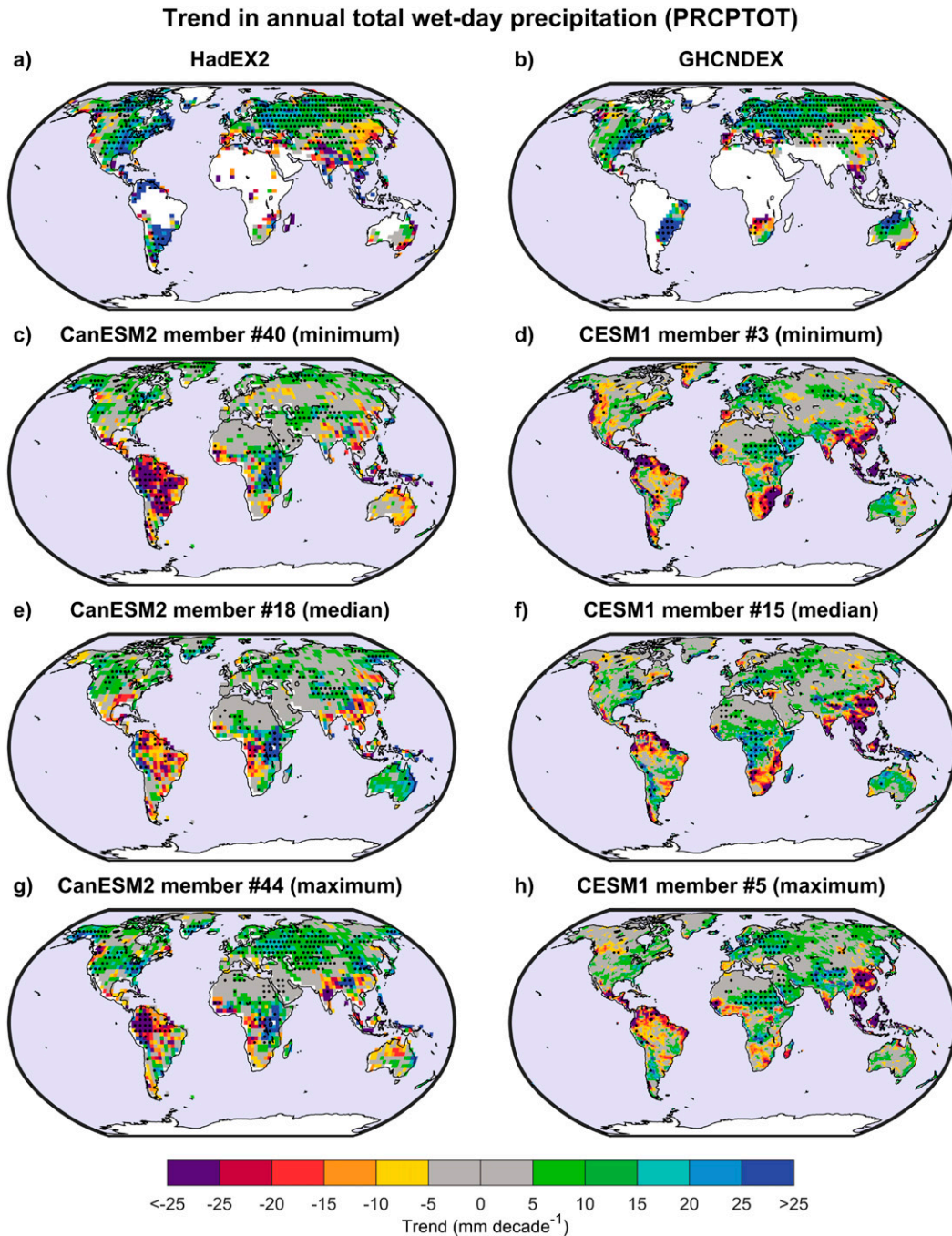## Trend in annual total wet-day precipitation (PRCPTOT)

FIG. 6. Linear trends in annual total wet-day precipitation (PRCPTOT) over the 1950–2010 period as defined by the Theil–Sen estimator. The first row represents observed (a) HadEX2 and (b) GHCNDEX dataset decadal trends. Grid points where fewer than 40 out of 60 years of data were available over the 1950–2010 period are shown in white. The remaining rows represent decadal trends for two individual members corresponding to the (c),(d) smallest, (e),(f) median, and (g),(h) largest median global trend value in the CanESM2 50-member ensemble and CESM1 40-member ensemble, respectively.

JJA shows a later detection than in the annual case. These results show that the likelihood of detecting a significant signal (stippled regions) is greater at the annual scale than at the seasonal scale for most regions. Figure 9 shows very similar results for the regional trends analysis based on field significance resampling approach. Overall, the 90%DD is reached somewhat earlier (slightly darker colors) and there is

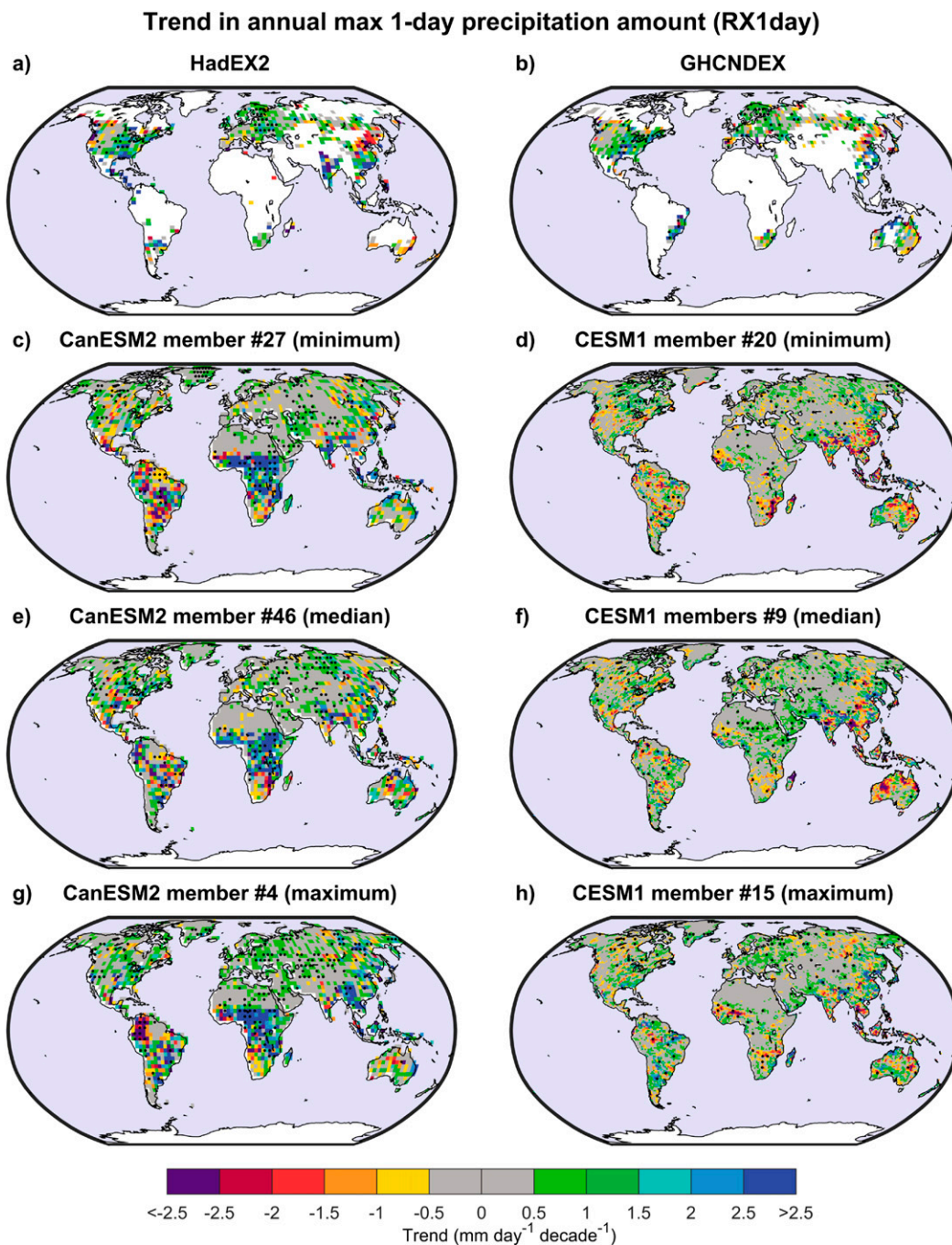## Trend in annual max 1-day precipitation amount (RX1day)

FIG. 7. As in Fig. 6, but for RX1day.

less noise in the maps as compared to the results obtained at the local scale.

The spatial patterns of average trend signs tend to be similar over both the annual scale and DJF, but differ in JJA. For instance, average trends are of different sign over most parts of Europe and North America, where more negative trends are observed for JJA as compared

to the annual scale and DJF (Figs. 8 and Figs. 9e,f). Overall, for CanESM2 (CESM1), there are 75.8% (76.4%) of all grid points with a positive trend, 70.8% (74.2%) for DJF and 68.5% (68.7%) for JJA.

As shown in the left-hand side panels of Figs. 10 and 11 for CanESM2, 17 (18) regions out of 21 have 50% of their land grid points with 90%DD occurring prior to

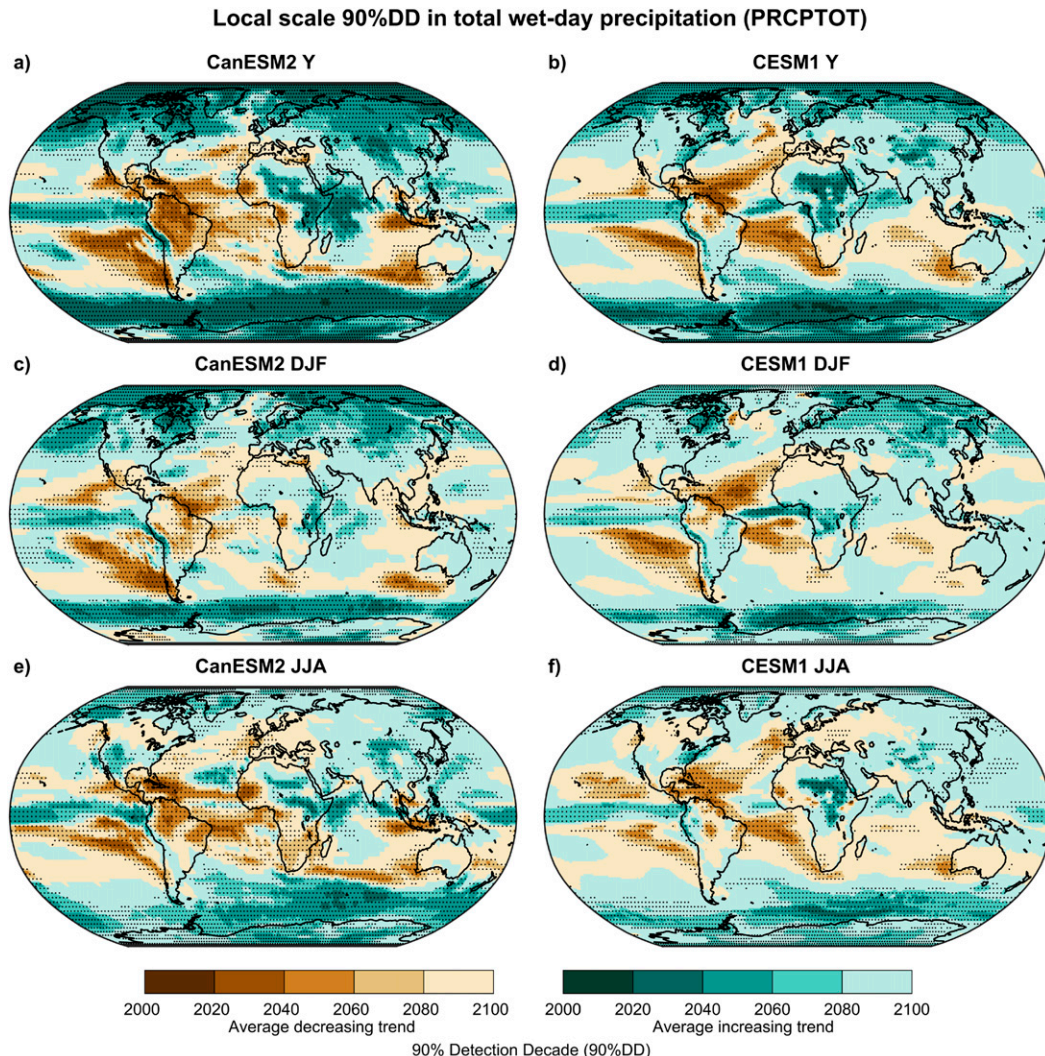**Local scale 90%DD in total wet-day precipitation (PRCPTOT)**

FIG. 8. Global local trends analysis (i.e., corresponding to each grid point) of the estimated 90%DD for the PRCPTOT index, showing results based on (left) the CanESM2 ensemble and (right) the CESM1 ensemble for the (top)–(bottom) annual (Y), DJF, and JJA scales. The brown colors represent an average decreasing trend while the blue/teal colors indicate an average increasing trend. The stippled patterns identify grid points where the estimated 90%DD occurred before 2100.

2100 at the annual scale, with 13 (16) regions for DJF and 9 (11) regions for JJA based on the local (regional) scale. Not a single region crosses this 50% of land grid points threshold before 2040 at the local scale and 2030 at the regional scale (and for most regions this will only occur a few decades later) at both the annual and seasonal scales.

For CESM1, the 90%DD is reached later than for CanESM2 as shown in the right-hand side panels of Figs. 10 and 11. A total of 12 (15) regions out of 21 have 50% of their land grid points reach their 90%DD prior to 2100 at the annual scale, with 9 (13) regions for DJF and 3 (8) regions for JJA at the local (regional) spatial

scale. Not a single region crosses the threshold before 2060 for the local and 2050 for regional trends [except for Tibet (TIB)] at both the annual and seasonal scales (two decades later than for CanESM2). On average, the threshold where 50% of the regions' land grid points reach their 90%DD in CESM1 is 1.6 decades later than for CanESM2 for the annual scale (1.6 for DJF and 0.9 for JJA at the regional scale).

Despite CESM1 having a later 90%DD than CanESM2, as well as some differences in their spatial patterns (see Figs. 8 and 9), both ensembles agree in many respects. The 90%DD is reached earlier at the regional scale for all 21 geographical regions and at the global land scale.

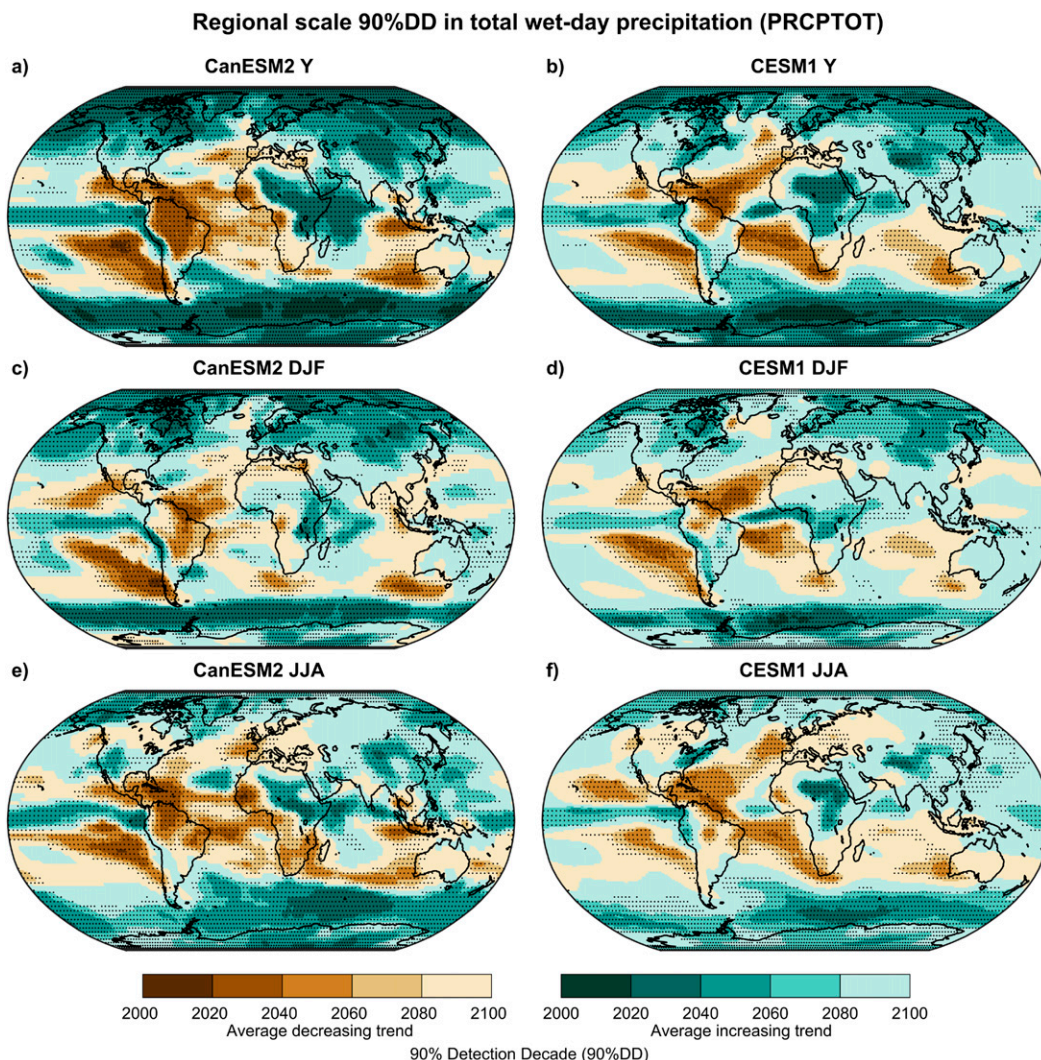## Regional scale 90%DD in total wet-day precipitation (PRCPTOT)

FIG. 9. Global regional trends analysis (i.e., computed with the field significance resampling approach) of the estimated 90%DD for the PRCPTOT index, showing results based on (left) the CanESM2 ensemble (region size of 3×3 grid points) and (right) the CESM1 ensemble (region size of 9×9 grid points), for the (top)–(bottom) annual (Y), DJF, and JJA scales. The brown colors represent an average decreasing trend while the blue/teal colors indicate an average increasing trend. The stippled patterns identify grid points where the estimated 90%DD occurred before 2100.

The regions with the earliest 90%DD are TIB, the tropical zones [the Amazon basin (AMZ; except for CESM1), eastern Africa (EAF), and western Africa (WAF)] and high-latitude zones above the 50th parallel [Alaska (ALA), Greenland (GRL), and northern Asia (NAS)]; see Table 1 for all regions and abbreviations. Eastern North America (ENA) is also one of the regions with the earliest 90%DD with CESM1, but this is not as clear for CanESM2. At the annual scale, a clear climate change signal emerges worldwide for the PRCPTOT index, except for the Australia (AUS), the Mediterranean Basin (MED), southern Africa (SAF), and South Asia (SAS) regions. When looking at DJF

and JJA, the climate change signal emerges later. By the end of the century, the climate change signal will most likely be detected in many regions of the world at the local or regional scales for this index.

c. RX1day index

A 90%DD analysis was also realized for precipitation extremes (using the RX1day index). Figure 12 (local scale) and Fig. 13 (regional scale) show maps of the 90% DD, while Fig. 14 (local scale) and Fig. 15 (regional scale) show the results for the 21 geographical regions.

The comparison between Figs. 12 and 13 indicates that differences between both ensembles is much smaller

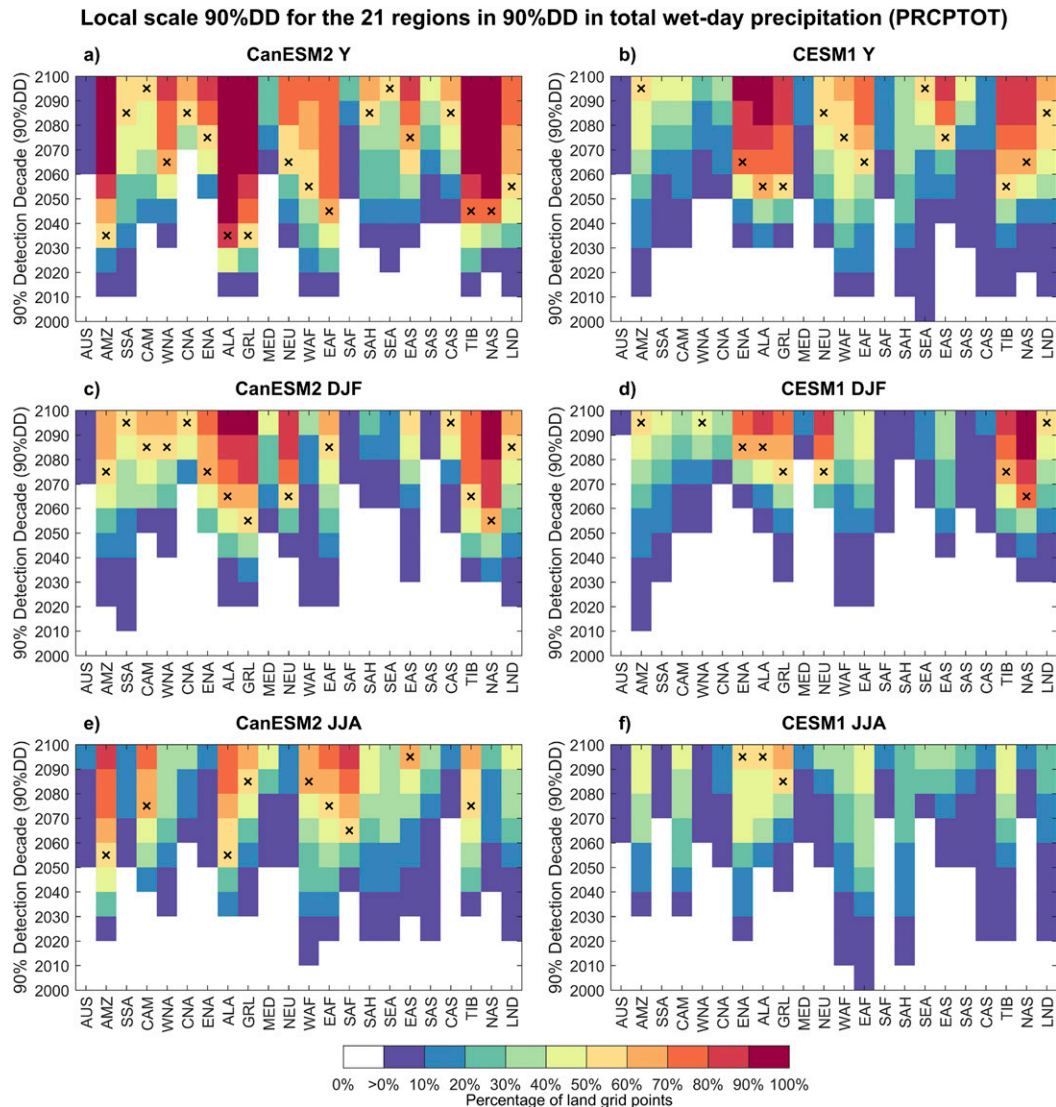Local scale 90%DD for the 21 regions in 90%DD in total wet-day precipitation (PRCPTOT)

FIG. 10. Cumulative percentage of land grid points within each of the 21 regions (listed in Table 1) with the local trend analysis (i.e., corresponding to each grid point) estimated 90%DD occurring in each decade for the PRCPTOT index, showing results based on (left) the CanESM2 ensemble (region size of 3 × 3 grid points) and (right) the CESM1 ensemble (region size of 9 × 9 grid points), for the (top)–(bottom) annual (Y), DJF, and JJA scales. The black "x" indicates the decade when more than 50% of the region's land grid points reached the 90% probability of detecting the climate change signal. The white boxes correspond to regions that had no 90%DD at any grid point in (and prior to) that decade.

than for the PRCPTOT index. Figure 12 shows that the local-scale results have a much larger fraction of both oceans and land surface areas that do not reach the 90% DD by the end of the simulation (nonstippled areas) for both annual and seasonal scales. However, Fig. 13 shows that the 90%DD occur earlier at the regional scale.

One distinctive feature here is that a larger number of average positive trends is observed for RX1day than for PRCPTOT. The percentage of all grid points showing a positive trend for CanESM2 (CESM1) is 86.4% (90.3%) at

the annual scale, with 80.5% (83.3%) for DJF and 77.1% (77.6%) for JJA. As was the case for PRCPTOT, the spatial patterns are similar for the annual scale and DJF, but notable differences are seen for JJA. Negative trends are observed across large parts of Europe and North America at the JJA scale for RX1day (Figs. 12 and Figs. 13e,f).

As shown in the left-hand side panels of Figs. 14 and 15 for CanESM2, 11 (21) regions out of 21 have 50% of their land grid points with a 90% probability of detecting the climate change signal before the end of the simulation
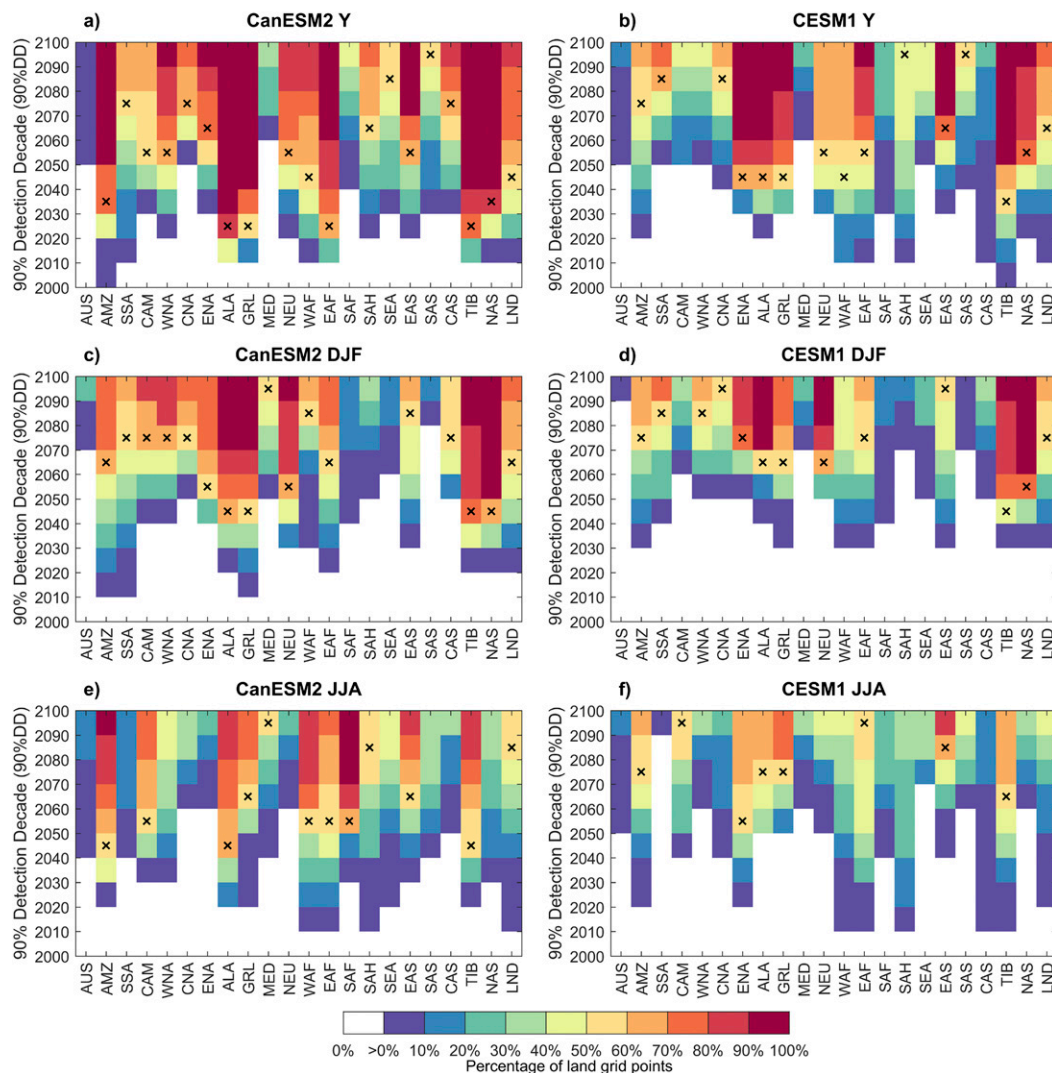
FIG. 11. Cumulative percentage of land grid points within each of the 21 regions (listed in Table 1) with the regional trend analysis (i.e., computed with the field significance resampling approach) estimated 90%DD occurring in each decade for the PRCPTOT index, showing results based on (left) the CanESM2 ensemble (region size of 3 × 3 grid points) and (right) the CESM1 ensemble (region size of 9 × 9 grid points), for the (top)–(bottom) annual (Y), DJF, and JJA scales. The black ''x'' indicates the decade when more than 50% of the region's land grid points reached the 90% probability of detecting the climate change signal. The white boxes correspond to regions that had no 90%DD at any grid point in (and prior to) that decade.

at the annual scale, with 8 (13) regions for DJF, and 4 (12) for JJA at the local (regional) scale. The threshold of 50% of land grid points was not achieved for any of the 21 geographical regions before 2050 (and most regions beyond that decade) at the local scale, and 2030 at the regional scale (with the exception for EAF at the annual scale, where it reached as early as 2010).

For the CESM1 ensemble, the 90%DD was also reached slightly later than for CanESM2, as shown in the right-hand side panels of Figs. 14 and 15. Overall, 8

(18) regions out of 21 reached the same threshold at the annual scale, 6 (14) regions for DJF and 2 (8) regions for JJA. For this ensemble, the regions that have 50% of their land grid points reaching the 90%DD the earliest were two high-latitude regions beyond 50°N: GRL (2070 for annual) and NAS (2070 for DJF) at the local scale. For the regional scale, the earliest was 2040 for five regions at the annual scale (ENA, GRL, WAF, EAF, and TIB; for TIB only for DJF and 2050 in the TIB region also for JJA). On average, the threshold where 50% of the

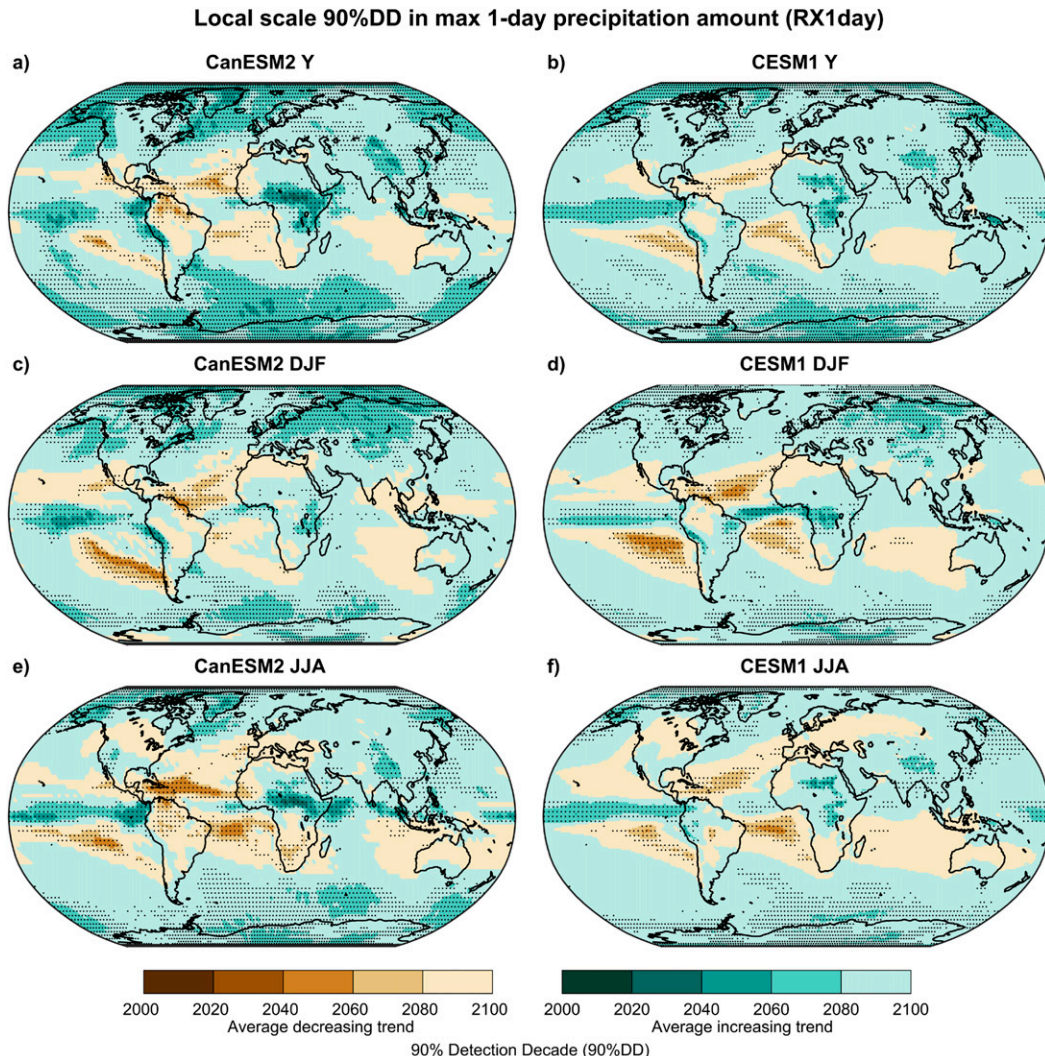**Local scale 90%DD in max 1-day precipitation amount (RX1day)**

FIG. 12. As in Fig. 8, but for the RX1day index.

land grid points reached their 90%DD in CESM1 is 0.4 decades later than for CanESM2 at the annual scale, and 0.3 for DJF and 0.8 for JJA at the regional scale.

Similarly to the PRCPTOT index, the geographical regions with the earliest 90%DD are also consistent for both ensembles when looking at the regional scale. This is also reflected in the combined land grid points, where we see a similar percentage of grid points reaching 90%DD globally. The regions with the earliest 90%DD for the RX1day index are the tropical zones (EAF and WAF), high-latitude zones above the 50th parallel (ALA, GRL, and NAS), regions affected by monsoons [SAS, East Asia (EAS), and TIB], and finally ENA, which is affected by hurricanes. These regions share in common the fact that an increase in warming will likely result in a robust climate change signal for RX1day. A later 90%DD is expected at the seasonal scale.

4. Discussion

a. Validation of both ensembles

The comparison with observations suggest that the spatial patterns of interannual variability and mean PRCPTOT index values and, to a lesser extent, of the RX1day index, as simulated by both CanESM2 and CESM1 ensembles are globally in agreement with corresponding patterns of the observed HadEX2 and GHCNDEX datasets. Differences can be partly explained by natural variability, as the distribution of annual mean and standard deviation over the various members can be quite dispersed, especially for grid points displaying large interannual variability (see Fig. S1 for PRCPTOT and Fig. S2 for RX1day in the online supplemental material). These discrepancies can also be due to biases in both ensembles and also from

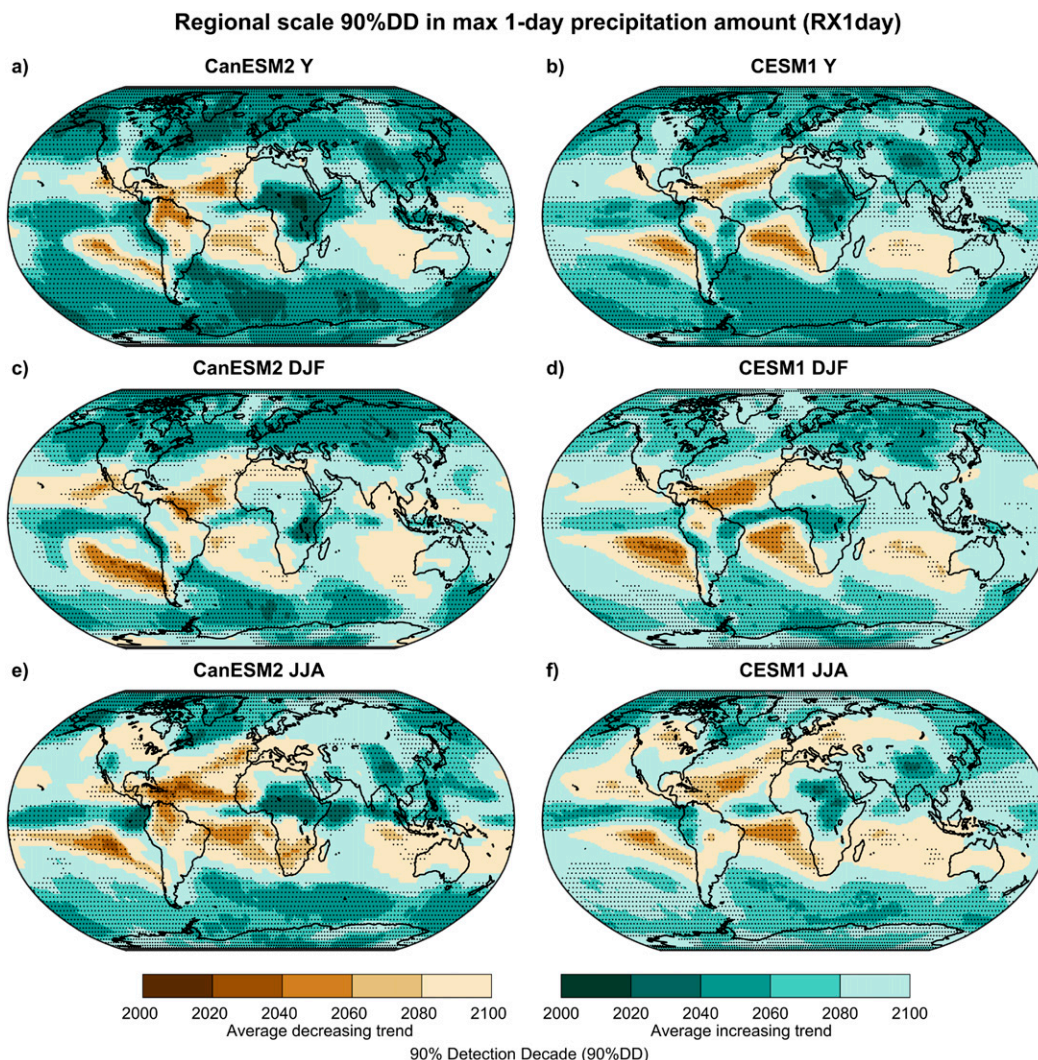**Regional scale 90%DD in max 1-day precipitation amount (RX1day)**

FIG. 13. As in Fig. 9, but for the RX1day index.

sampling errors and uncertainties in HadEX2 and GHCNDEX datasets.

Furthermore, the comparison of trends between models and observations for both indices suggests that it is difficult to directly compare the global spatial distribution of trends obtained by the different members of each ensemble to observed trends. When comparing one realization (the observed recent past) against a probabilistic distribution (ensemble members), the best possible outcome is to frame this realization within the possible predicted range according to the expected statistical frequency. However, the large variability of trends extracted for each ensemble members demonstrates the challenge of detecting the climate change signal at the local scale. Comparison of observed and simulated trends was only achieved at the local scales.

A comparison for each region listed in Table 1 could also have been performed, but a qualitative analysis of all members for each ensemble (not shown due to lack of space) clearly outlined a very large intermember variability at the scale of the regions and would not have changed the above conclusion. Other difficulties when dealing with local and regional comparisons arise from the different sources of uncertainty in observation datasets, such as short observational records, homogeneity problems, and missing data (Hegerl et al. 2015). Furthermore, since gridded observed datasets are typically constructed by interpolated point values (e.g., station), various upscaling/downscaling problems are always present (Avila et al. 2015; Chen and Knutson 2008; Herold et al. 2017; Sillmann et al. 2013a).

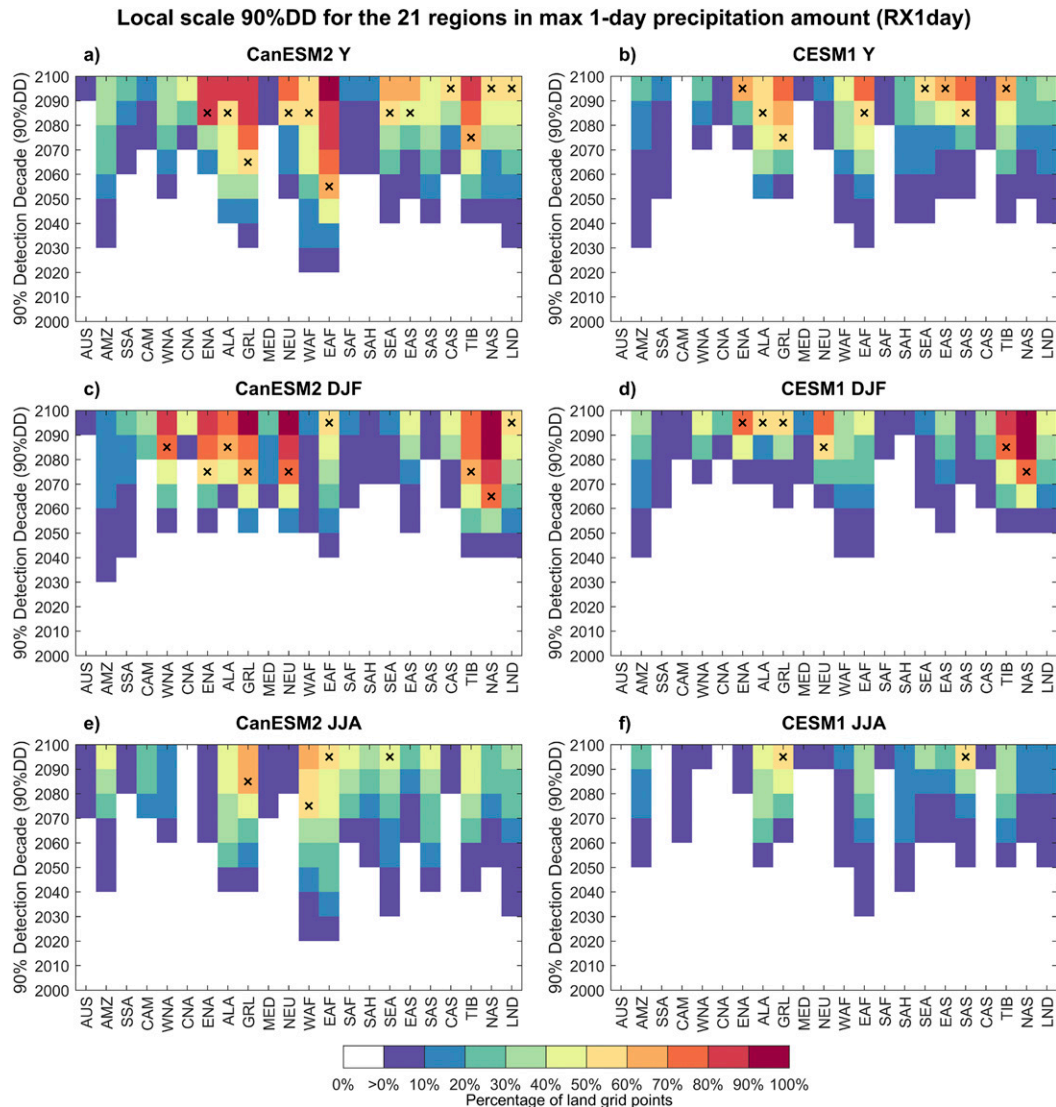The 90%DD is shown to be conservative estimate as it corresponds to the decade where the climate change

FIG. 14. As in Fig. 10, but for the RX1day index.

signal is detected in most of simulations from a large ensemble of simulations, as compared to a "single realization" of the climate system when dealing with the real world. This is well illustrated in the supplemental material (Figs. S3–S6), showing the probability of detecting a significant trend during a given decade at both the local and regional scales. While the probability increases overall as we move further into the twenty-first century, very high probabilities are only reached after the midcentury, and even later for many grid points. For the 1950–2010 period, this probability remains relatively low for most regions.

A limited qualitative comparison of CanESM2 and CESM1 against the CMIP5 multimodel mean signal was made to frame the general behavior of both climate models against other GCMs or ESMs. Globally, spatial patterns of increasing and decreasing trends match the multimodel average changes obtained by Sillmann et al. (2013b) for both annual total and extreme precipitation indices. Furthermore, both the sign of the change and robustness of the climate change signal (characterized in this study by an early 90%DD) match the signal obtained by Fischer et al. (2014) remarkably well (especially for RX1day) in regions where at least 12 out of 15 CMIP5 models agreed on the direction of change. The regions with the most robust climate change signal for precipitation extremes obtained by Scoccimarro et al. (2013) are consistent with the regions with the earliest 90%DD for the majority of land grid points for both ensembles. There are no reasons to assume that the
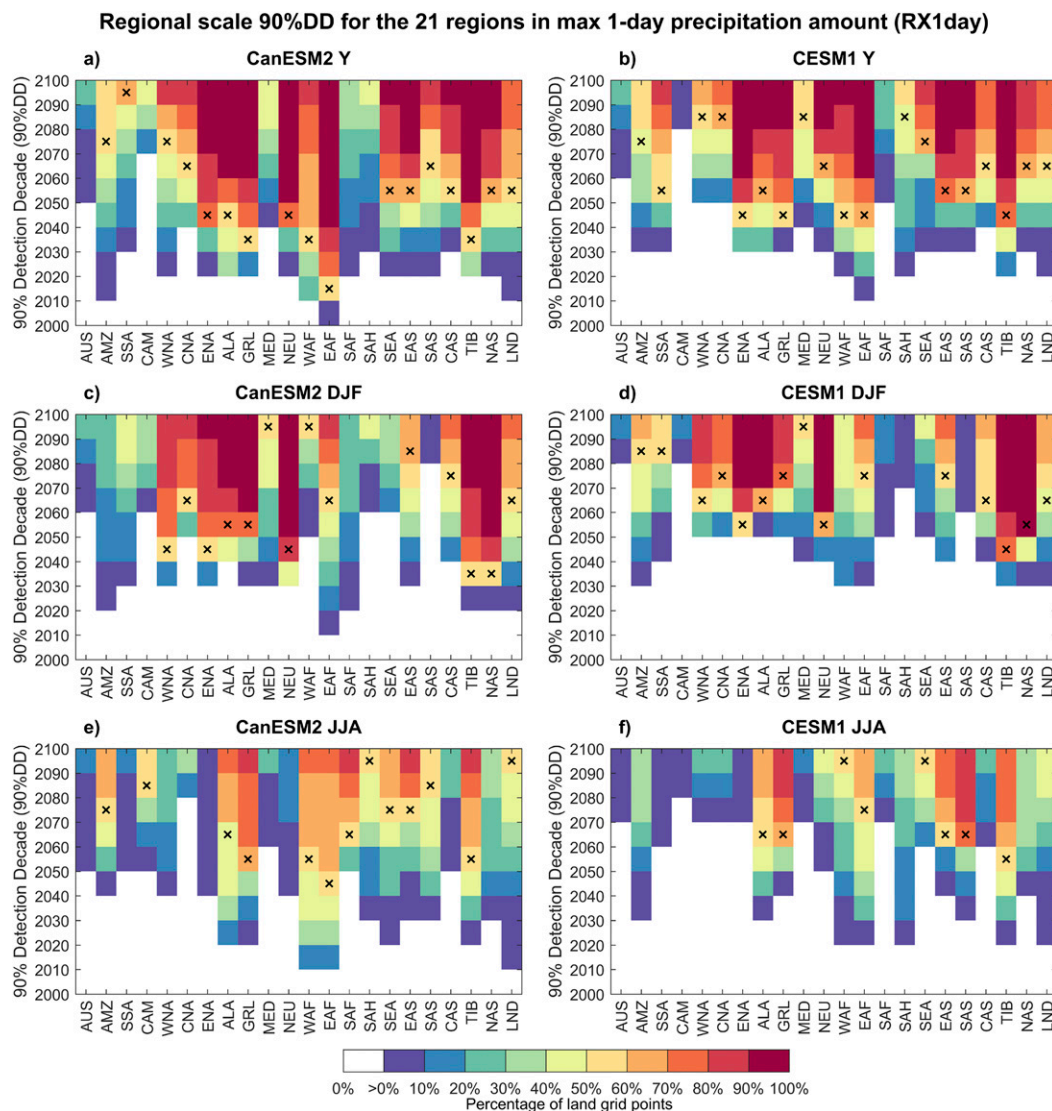
FIG. 15. As in Fig. 11, but for the RX1day index.

b. Impact of natural variability at the local and regional scales

Both local and regional trend-based analyses were performed to determine how the spatial correlation affects results for the PRCPTOT and RX1day indices. In general, the field significance resampling approach showed that a more robust climate change signal can be detected from natural variability at the regional scale as compared to the local scale. Figures S3–S6 show that the increasing probability of detecting a significant trend is initially larger and grows faster at the regional scale.

For the PRCPTOT index, results from both the local and regional scales are quite similar for the annual and seasonal scales. This suggests that mean precipitation trends can likely be detected at the local scale. However, for RX1day, spatial dependence was shown to have a great influence, as the results for the regional scale were markedly different from those at the local scale. Figures 14 and 15 clearly show the difference between the local and regional scales for RX1day.

These results show that when investigating extreme precipitation at the local scale, it is likely that natural variability will strongly impede the detection of a statistically significant climate change signal over a long period. Overall, this is also in agreement with Fischer et al. (2013), who concluded that it is not possible to provide stakeholders with reliable information for changes in extreme precipitation when investigating at the local scale.

conclusions drawn from both ensembles would be markedly different when using another GCM/ESM.

TABLE 2. Comparison of the percentage of grid points (either all grid points or land grid points) with 90%DD before the end of the simulations in 2100. Results for CanESM2 and CESM1 ensembles at the annual (Y) and seasonal (DJF and JJA) scales are shown for both indices at both the local and regional scales.

Model	Scale	All grid points				Land grid points (LGP)			
		Local scale		Regional scale		Local scale		Regional scale	
		PRCPTOT	RX1day	PRCPTOT	RX1day	PRCPTOT	RX1day	PRCPTOT	RX1day
CanESM2	Y	60.5%	59.6%	64.1%	74.1%	66.0%	53.2%	71.2%	78.2%
	DJF	46.4%	43.8%	52.6%	59.7%	56.0%	55.0%	64.1%	69.3%
	JJA	44.0%	41.5%	51.7%	59.1%	30.0%	26.8%	38.4%	42.7%
CESM1	Y	54.6%	52.2%	60.3%	78.8%	57.3%	38.9%	67.1%	81.5%
	DJF	42.1%	37.0%	49.2%	61.1%	50.7%	47.1%	60.2%	68.8%
	JJA	38.7%	34.7%	47.3%	60.9%	23.8%	21.1%	33.6%	43.2%

Westra et al. (2013) investigated trends on the HadEX2 dataset for the RX1day index using a field significance resampling approach. The areas that showed the most significant trends were the United States, Europe, South Africa, and some parts of India and Southeast Asia. With the exception of South Africa, the results obtained here for these areas (Figs. 12 and 13) also showed a relatively early 90%DD, corresponding to areas with a robust climate change signal.

Further comparisons were made using the CESM1 ensemble to investigate the effect of using an increasing region size in the field significance resampling approach. Regions made of 1 (1 × 1), 9 (3 × 3), 25 (5 × 5), 49 (7 × 7), and 81 (9 × 9) grid points were used for this analysis. Results can be seen in Fig. S7 (PRCPTOT) and Fig. S8 (RX1day). Overall, for RX1day, the results indicate a convergence around the 5 × 5 domain, with minor changes seen as we move to a larger domain. As for the PRCPTOT index, there was no significant difference at any of the sizes tested, which is consistent with the previously discussed results. It is expected that using a larger region would eventually lead to an overlap of wetter and dryer regions, which could impair the ability to detect trends at the regional scale.

c. Impact of natural variability on the PRCPTOT and RX1day indices

The discussion from the previous section clearly outlines one of the main differences between both indices, which is the strong influence of natural variability at the local scale for RX1day, and its much smaller influence for the PRCPTOT index.

Fischer and Knutti (2014); and Fischer et al. (2014) show that there is a greater expectation of extreme precipitation to emerge from natural variability than mean precipitation. They argue that natural variability is indeed greater in the case of extreme precipitation. However, this difference is likely due because precipitation extremes respond more strongly to global warming than does mean precipitation (Fischer and Knutti 2014; Fischer et al. 2014). Results in Table 2

show a comparison between both indices of the percentage of grid points that have reached their 90% DD before the end of the century. When looking at the local scale, we see that the RX1day index has fewer grid points reaching their 90%DD as compared to PRCPTOT (e.g., the CESM1 annual-scale LGP percentage for PRCPTOT is 57.3% vs 38.9% for RX1day). However, at the regional scale, the RX1day ends up with a larger number of grid points reaching their 90% DD than the PRCPTOT index (e.g., the CESM1 annual-scale LGP percentage for PRCPTOT is 67.1% vs 81.5% for RX1day). These results indicate that the climate change signal for RX1day is indeed more robust than for PRCPTOT at the global scale, which is in agreement with previous studies.

For PRCPTOT, many regions will experience an increase in precipitation (especially at high latitudes), while a considerable number of regions will also see a decrease in precipitation (see Figs. 8 and 9). However, for RX1day (see Figs. 12 and 13), nearly all land grid points show an increasing trend due to climate change. Globally, the RX1day index shows more increasing trends than the PRCPTOT index, both for the annual and seasonal scales (with the smallest percentage at the JJA scale). Thus, there will be regions that will see a decrease in annual total precipitation, but an increase in RX1day. While the RX1day index increases globally at the annual scale, many regions will see a decrease at the JJA [e.g., AUS, central North America (CNA), MED, northern Europe (NEU), and SAF]. The Amazon basin and central North America seem to be the only regions where decreases are observed year round. Overall, these spatial patterns of average increasing or decreasing trends agree with the general behavior of the expected climate change signal described by the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) and other published studies (Fischer et al. 2014; Hegerl et al. 2015; IPCC 2013; King et al. 2015; Maraun 2013).

The results indicate that for both indices the climate change signal will be affected by natural variability until

past the midcentury for most land grid points at the local and regional scales. At the global scale, Fischer and Knutti (2014) showed that a significant fraction of grid points will experience increases. It is also likely that this influence will be stronger during summer (JJA) than during winter (DJF) or at the annual scales. When looking at the 21 geographical regions, high-latitude (e.g., GRL, ALA, and NAS) and tropical (e.g., AMZ, WAF, and EAF) climate change signals will be detected much earlier than in other regions for both ensembles; other regions will see their 90%DD reached later in the century.

Overall, natural variability represents a considerable source of uncertainty and it can mask or amplify the climate change signal at both the local and regional scales. This conclusion agrees with those from previous studies (Deser et al. 2012a,b; Fischer and Knutti 2014; Kay et al. 2015; Thompson et al. 2015; Sanderson et al. 2018).

5. Discussion of limitations

The following issues need to be discussed as their outcome may impact the conclusions of this study:

a. Coarse resolution of the ESM

There are indications that even with their coarse spatial resolutions, both GCMs and ESMs do a reasonably good job capturing the large-scale events usually associated with synoptic weather patterns (IPCC 2013; Sillmann et al. 2013a). However, smaller-scale weather events in GCMs or ESMs are not directly simulated but considered through convection parameterization schemes (Chan et al. 2014; Jones and Randall 2011; Kendon et al. 2012, 2016; Prein et al. 2015, 2017). A spatial resolution of the order of the kilometer would be required to adequately simulate the deep convection that plays a significant role in the generation of extreme rainfall in some regions at the daily scale (Prein et al. 2015). Thus, the impact of spatial resolution and deep convection parameterization needs to be investigated using a large ensemble of simulations at very high resolutions (approximately a few kilometers). The only available simulations are still limited to small regions (Prein et al. 2015, 2017).

b. Representative concentration pathway

There is evidence to suggest that the rate of increase in extreme precipitation does not depend specifically on the emission scenario (as it does for mean precipitation) but rather on the total amount of warming (Pendergrass et al. 2015). The RCP8.5 used in this study represents the scenario with the largest increase in greenhouse gas concentrations typically used in climate change studies (IPCC 2013). It is reasonable to think that under less significant anthropic forcing, natural variability could be expected to hide the anthropogenic climate change signal over longer time periods since forcing is weaker. This hypothesis could only be validated by comparing two large ensembles of simulations from the same model with different forcing scenarios.

A study by Sanderson et al. (2018) used two large ensembles from the Community Earth System Model with identical settings (30 members using RCP8.5 and 15 members using RCP4.5) to explore the role played by greenhouse gas concentration trajectories. Their results suggest a considerable overlap in possible outcomes for both ensembles even in the 2080 decade. Some significant changes between both scenarios started appearing, albeit with considerable overlap after 2040 at the regional scale in northern Europe, while no difference was observed at the local scale.

By extending these conclusions to this work, under the weaker RCP2.6 or RCP4.5 scenarios, lower probabilities of detecting the climate change signal could be expected resulting in later 90%DD than those obtained for the RCP8.5 at the regional scale, but with little difference at the local scale.

c. Simulation period

Trend analyses were performed on subperiods of the 1950–2100 simulations. Extending this to the pre-1950s period, and ultimately to the nineteenth century, when anthropogenic forcing began, could possibly have an impact on trends detection in the climate change signal. This is because trends detection will very likely be impacted when using longer time series, which could in turn have an impact on the estimated trend detection probability during forthcoming periods. Such work could only be performed if both large ensembles had simulations using extended periods prior to 1950.

This was tested to see the impact of using the longer simulation period available for CESM1 (from 1920 to 2100) and shown in Figs. S9 and S10 for PRCPTOT and Figs. S11 and S12 for RX1day. Using an extended period early in the twentieth century did not provide different conclusions from those obtained using the simulations starting in 1950. Thus, it is reasonable to assume that this limitation should not have a significant impact on the results and conclusions obtained in the present paper.

6. Concluding remarks

For precipitation extremes, natural variability is likely to dominate the climate change signal at the local scale until the next century in many parts of the world. To

properly estimate trends in extreme precipitation it is essential to take into account spatial dependence. This is less critical for annual and seasonal total precipitation, which is comparably less affected by natural variability at the local scale. When accounting for spatial dependence, trend detection for precipitation extremes is expected to occur for a larger number of grid points than for annual and seasonal total precipitation.

In some instances, natural variability may undermine our ability to detect the climate change signal at the local and regional scales. This should not prevent us from implementing adaptation measures, especially when dealing with precipitation extremes. In other words, the uncertainty linked to natural variability should not detract decision makers from underlying anthropogenic changes. Nonetheless, results from this study clearly show that natural variability can impede the detection of the anthropogenic signal for a few to several decades over many parts of the world, and this should be considered when implementing adaptation strategies.

REFERENCES

Arora, V. K., and Coauthors, 2011: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophys. Res. Lett.*, **38**, L05805, https://doi.org/10.1029/2010GL046270.

Avila, F. B., S. Dong, K. P. Menang, J. Rajczak, M. Renom, M. G. Donat, and L. V. Alexander, 2015: Systematic investigation of gridding-related scaling effects on annual statistics of daily temperature and precipitation maxima: A case study for south-east Australia. *Wea. Climate Extremes*, **9**, 6–16, https://doi.org/10.1016/j.wace.2015.06.003.

Chan, S. C., E. J. Kendon, H. J. Fowler, S. Blenkinsop, N. M. Roberts, and C. A. T. Ferro, 2014: The value of high-resolution Met Office regional climate models in the simulation of multihourly precipitation extremes. *J. Climate*, **27**, 6155–6174, https://doi.org/10.1175/JCLI-D-13-00723.1.

Chen, C.-T., and T. Knutson, 2008: On the verification and comparison of extreme rainfall indices from climate models. *J. Climate*, **21**, 1605–1621, https://doi.org/10.1175/2007JCLI1494.1.

Deser, C., R. Knutti, S. Solomon, and A. S. Phillips, 2012a: Communication of the role of natural variability in future North American climate. *Nat. Climate Change*, **2**, 775–779, https://doi.org/10.1038/nclimate1562.

——, A. Phillips, V. Bourdette, and H. Teng, 2012b: Uncertainty in climate change projections: The role of internal variability. *Climate Dyn.*, **38**, 527–546, https://doi.org/10.1007/s00382-010-0977-x.

——, ——, M. A. Alexander, and B. V. Smoliak, 2014: Projecting North American climate over the next 50 years: Uncertainty due to internal variability. *J. Climate*, **27**, 2271–2296, https://doi.org/10.1175/JCLI-D-13-00451.1.

Dittus, A. J., D. J. Karoly, S. C. Lewis, and L. V. Alexander, 2015: A multiregion assessment of observed changes in the areal extent of temperature and precipitation extremes. *J. Climate*, **28**, 9206–9220, https://doi.org/10.1175/JCLI-D-14-00753.1.

Donat, M. G., and Coauthors, 2013a: 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.

——, L. V. Alexander, H. Yang, I. Durre, R. Vose, and J. Caesar, 2013b: Global land-based datasets for monitoring climatic extremes. *Bull. Amer. Meteor. Soc.*, **94**, 997–1006, https://doi.org/10.1175/BAMS-D-12-00109.1.

Douglas, E. M., R. M. Vogel, and C. N. Kroll, 2000: Trends in floods and low flows in the United States: Impact of spatial correlation. *J. Hydrol.*, **240**, 90–105, https://doi.org/10.1016/S0022-1694(00)00336-X.

Fischer, E. M., and R. Knutti, 2014: Detection of spatially aggregated changes in temperature and precipitation extremes. *Geophys. Res. Lett.*, **41**, 547–554, https://doi.org/10.1002/2013GL058499.

——, U. Beyerle, and R. Knutti, 2013: Robust spatially aggregated projections of climate extremes. *Nat. Climate Change*, **3**, 1033–1038, https://doi.org/10.1038/nclimate2051.

——, J. Sedláček, E. Hawkins, and R. Knutti, 2014: Models agree on forced response pattern of precipitation and temperature extremes. *Geophys. Res. Lett.*, **41**, 8554–8562, https://doi.org/10.1002/2014GL062018.

Giorgi, F., and R. Francisco, 2000: Uncertainties in regional climate change prediction: A regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. *Climate Dyn.*, **16**, 169–182, https://doi.org/10.1007/PL00013733.

——, and X. Bi, 2009: Time of emergence (TOE) of GHG-forced precipitation change hot-spots. *Geophys. Res. Lett.*, **36**, L06709, https://doi.org/10.1029/2009GL037593.

Hawkins, E., and R. Sutton, 2011: The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dyn.*, **37**, 407–418, https://doi.org/10.1007/s00382-010-0810-6.

——, and ——, 2012: Time of emergence of climate signals. *Geophys. Res. Lett.*, **39**, L01702, https://doi.org/10.1029/2011GL050087.

——, T. Edwards, and D. McNeall, 2014: Pause for thought. *Nat. Climate Change*, **4**, 154–156, https://doi.org/10.1038/nclimate2150.

Hegerl, G. C., and Coauthors, 2015: Challenges in quantifying changes in the global water cycle. *Bull. Amer. Meteor. Soc.*, **96**, 1097–1115, https://doi.org/10.1175/BAMS-D-13-00212.1.

Herold, N., A. Behrangi, and L. V. Alexander, 2017: Large uncertainties in observed daily precipitation extremes over land. *J. Geophys. Res. Atmos.*, **122**, 668–681, https://doi.org/10.1002/2016JD025842.

IPCC, 2013: *Climate Change 2013: The Physical Science Basis*. Cambridge University Press, 1535 pp.

Jones, T. R., and D. A. Randall, 2011: Quantifying the limits of convective parameterizations. *J. Geophys. Res.*, **116**, D08210, https://doi.org/10.1029/2010JD014913.

Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) Large Ensemble Project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1.

Kendall, M. G., 1975: *Rank Correlation Methods*. 4th ed. Charles Griffin, 202 pp.

Kendon, E. J., N. M. Roberts, C. A. Senior, and M. J. Roberts, 2012: Realism of rainfall in a very high-resolution regional climate model. *J. Climate*, **25**, 5791–5806, https://doi.org/10.1175/JCLI-D-11-00562.1.

——, and Coauthors, 2016: Do convection-permitting regional climate models improve projections of future precipitation change? *Bull. Amer. Meteor. Soc.*, **98**, 79–93, https://doi.org/10.1175/BAMS-D-15-0004.1.

Kiktev, D., D. M. H. Sexton, L. Alexander, and C. K. Folland, 2003: Comparison of modeled and observed trends in indices of daily climate extremes. *J. Climate*, **16**, 3560–3571, https://doi.org/10.1175/1520-0442(2003)016<3560:COMAOT>2.0.CO;2.

King, A. D., and Coauthors, 2015: The timing of anthropogenic emergence in simulated climate extremes. *Environ. Res. Lett.*, **10**, 094015, https://doi.org/10.1088/1748-9326/10/9/094015.

Klein Tank, A. M. G., F. W. Zwiers, and X. Zhang, 2009: Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation. WMO, 52 pp., www.wmo.int/pages/prog/wcp/wcdmp/wcdmp.../WCDMP_72_TD_1500_en__1.pdf.

Lins, H. F., and J. R. Slack, 1999: Streamflow trends in the United States. *Geophys. Res. Lett.*, **26**, 227–230, https://doi.org/10.1029/1998GL900291.

Maraun, D., 2013: When will trends in European mean and heavy daily precipitation emerge? *Environ. Res. Lett.*, **8**, 014004, https://doi.org/10.1088/1748-9326/8/1/014004.

Mora, C., and Coauthors, 2013: The projected timing of climate departure from recent variability. *Nature*, **502**, 183–187, https://doi.org/10.1038/nature12540.

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.

Prein, A. F., and Coauthors, 2015: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Rev. Geophys.*, **53**, 323–361, https://doi.org/10.1002/2014RG000475.

——, R. M. Rasmussen, K. Ikeda, C. Liu, M. P. Clark, and G. J. Holland, 2017: The future intensification of hourly precipitation extremes. *Nat. Climate Change*, **7**, 48–52, https://doi.org/10.1038/nclimate3168.

Sanderson, B. M., K. W. Oleson, W. G. Strand, F. Lehner, and B. C. O'Neill, 2018: A new ensemble of GCM simulations to assess avoided impacts in a climate mitigation scenario. *Climatic Change*, **146**, 303–318, https://doi.org/10.1007/s10584-015-1567-z.

Scoccimarro, E., S. Gualdi, A. Bellucci, M. Zampieri, and A. Navarra, 2013: Heavy precipitation events in a warmer climate: Results from CMIP5 models. *J. Climate*, **26**, 7902–7911, https://doi.org/10.1175/JCLI-D-12-00850.1.

Sen, P. K., 1968: Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Stat. Assoc.*, **63**, 1379–1389, https://doi.org/10.1080/01621459.1968.10480934.

Sigmond, M., and J. C. Fyfe, 2016: Tropical Pacific impacts on cooling North American winters. *Nat. Climate Change*, **6**, 970–974, https://doi.org/10.1038/nclimate3069.

Sillmann, J., V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh, 2013a: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res. Atmos.*, **118**, 1716–1733, https://doi.org/10.1002/jgrd.50203.

——, ——, F. W. Zwiers, X. Zhang, and D. Bronaugh, 2013b: Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *J. Geophys. Res. Atmos.*, **118**, 2473–2493, https://doi.org/10.1002/jgrd.50188.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

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

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.

Wilks, D. S., 1997: Resampling hypothesis tests for autocorrelated fields. *J. Climate*, **10**, 65–82, https://doi.org/10.1175/1520-0442(1997)010<0065:RHTFAF>2.0.CO;2.

——, 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 676 pp.

Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. Klein Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers, 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.