doi:10.1038/nature09763

## 1. Observation and model data processing

As observations of annual maximum daily and five-day precipitation amounts (RX1D and RX5D respectively), we use the Hadley Centre global land-based climate extremes dataset (HadEX)[1]. HadEX covers the period 1951-2003 and provides the most comprehensive gridded dataset of indices for temperature and precipitation extremes over the global land currently available. It is based on the station observations from various sources including pre-existing datasets, results from a number of workshops held in data-sparse regions, and contributions from individual institutions world wide[2]. Data were quality controlled and homogeneity was taken into account comprehensively in order to be used for analyzing long-term trends of climate extremes (see below for more discussion). Daily precipitation data are available at about 6,000 stations. The HadEX dataset is presented on a 2.5° latitude × 3.75° longitude grid. We calculated the probability-based index (PI, see below) for 1951-1999 for both RX1D and RX5D on the HadEX grid.

For comparison with the model datasets (see below), and due to better observational coverage, we confine our analysis to the period ending in 1999. There is a sudden drop in observational coverage after 2000. For 2001-2003, fewer than 60% of grid-points have data compared to the 1961-1990 mean. Supplementary Fig. 5 shows the geographical pattern of the observed PI trends for 1951-2003, where we consider only grid-points with at least two years of data out of 2000-2003. The sudden drop in observed data coverage can be clearly seen in the plot, with almost no grid-points over Eurasia except for China and part of India for RX1D. Overall, trend patterns of RX1D and RX5D for 1951-2003 are found to be

similar to those for 1951-1999 (Fig. 1) over the reduced area. This is consistent with regional studies for North America[3,4], China[5,6], and Europe[7] that include the recent years in their analyses.

The spatial coverage of RX5D is better than for RX1D (Fig. 1) because RX5D has a larger decorrelation length scale than RX1D, and therefore was "infilled" more extensively[1]. This might raise concerns that the RX5D dataset could be less reliable than RX1D, especially at grid-boxes where RX5D value is available but RX1D is not. We therefore tested the sensitivity of our RX5D detection results to coverage by restricting our RX5D analyses to RX1D grid-boxes. We found little influence of different spatial coverage of RX5D on our main detection results (Supplementary Fig. 1). Nevertheless, results will depend to some extent on assumptions in gridding (for both RX1D and RX5D) and on the robustness of individual records of extreme precipitation. In this context we note that extremely high precipitation values are generally more robustly measured than very low amounts (see below).

Simulated daily precipitation data were obtained from the Coupled Model Intercomparison Project Phase 3 (CMIP3) archive[8] and from individual modelling centres (see Supplementary Table 1). We use 20$^{th}$ century simulations (20C3M) from the eight models that provided daily precipitation datasets. A total of 19 anthropogenic forcing (ANT, greenhouse gases and sulphate aerosols as major factors) simulations are available from six models. In addition, 16 simulations conducted under natural (solar and volcanic activity) plus anthropogenic forcings (ALL) are also available from five models. Three models provide datasets from both experiments. In order to test sensitivity to model sampling, we have redone our detection analysis using the three models that provide both ANT and ALL runs. Our detection results are found to be robust when this subset of

models is selected (Supplementary Fig. 6). We use a total of 4,300 years of preindustrial control (CTL, no external forcing) simulations from eight models. A 1,000-year present-day control simulation from the ECHO-G model has also been used to increase CTL samples, resulting in total 106 non-overlapping 49-year CTL chunks. The number of simulations used from each model for different forcing experiments is listed in the Supplementary Table 1 (also see Supplementary Fig. 2).

In order to reduce uncertainties arising from different horizontal resolutions in observations and multiple models, we first calculate PI (see below) at different resolutions from individual models and then interpolated onto the 3.75° × 2.5° grid of the observations. We used an interpolation scheme that calculates grid-box values as a weighted average of the values on the original grid where the weighting is determined by the degree of overlap of the original boxes onto the target grid-box. The scheme requires that a minimum of 50% of the area of the target grid-box be covered by the original non-missing grid-boxes in order to obtain an interpolated value. In order to consider the availability of observed data, only grid-points with at least 40 years of observations during 1951-1999 are considered to calculate area-averaged time series of PI for both observations and model simulations (see below for the influence of using different thresholds). For each 49-year model run, area-averages of PI were obtained for each model year over the entire Northern Hemisphere (NH) land area and over other latitudinal or continental scale sub-areas. Five-year mean anomalies were then calculated with respect to its 49-year mean (four-year means are used for 1951-1954).

**2. Probability-based index calculation**

We assume that the annual extreme precipitation amounts analyzed here, RX1D and RX5D, follow the generalized extreme value (GEV) distribution which incorporates Gumbel, Frechet, and Weibull distributions and has a cumulative distribution function (CDF) as described in Eq. (1) in the Methods Summary of the main paper. The choice of GEV distribution is well motivated by statistical extreme value theory[9]. The GEV distribution is fitted in order to be able to convert the 49-year time series of extreme precipitation amounts, into 49-year time series of a standardized index, i.e. probability-based indices PI. In fitting the GEV distribution, we assume that its location ($\mu$), scale ($\sigma$), and shape ($\xi$) parameters are constant in time. Distributions are fitted separately at each grid-point with sufficient data. Fitting is performed by using maximum likelihood estimation (MLE) to estimate the GEV parameters, for which a simplex function minimization procedure is employed with L-moment estimates taken as the initial values for the maximization[10]. The fitted distribution is then used to transform time series of annual precipitation extremes to time series of PI values. This is simply done by evaluating the fitted CDF at the values of the annual extremes. The resulting dimensionless index ranges in value from 0 to 1, as explained in the Methods Summary.

There might be a concern about uncertainty in the GEV parameter estimation. However, 49-year samples are generally adequate when fitting the GEV distribution, e.g., problems with MLE fitting of the GEV generally only turn up in small samples ($\leq 25$)[11], and resulting parameter uncertainty would exert a very small impact on optimal regression. Our MLE calculations converge very quickly for either 49 or 30 samples, and this also supports confidence for our GEV parameter estimation. In addition, we have conducted a simple sensitivity test by using a short period (1961-90) to estimate GEV parameters, i.e. using 30 samples rather than 49 samples, and found negligible influence on large-scale area mean PI series (not shown). Many previous studies have demonstrated that annual

precipitation extremes can be well modelled by the GEV distribution. However, even if there were some misfit, the primary objective of transforming precipitation onto a smooth dimensionless scale taking values between 0 and 1 would have been achieved.

### 3. Estimating and evaluating internal variability

Two sets of 49-year CTL samples are prepared as follows. We divide the available CTL simulations into 100-year chunks performed with different models (53 chunks in total, Supplementary Table 1). Each of the 100-year chunks is again split into two 50-year chunks, and for each 49-year series (excluding 1st year) RX1D and RX5D amounts are transformed into PI values (see above and Methods Summary in the main paper). The 49-year time series of annual PI patterns are interpolated into the common 3.75° × 2.5° grid of the observations and masked by the observational availability (see above). Five-year area mean anomalies are then calculated as processed for the observations (see Supplementary Fig. 7), which we use to estimate internal variability. One set of 53 49-year samples is used to obtain the EOF projection and the noise covariance matrix required for the estimation of the regression coefficient, and the other set is used for the residual consistency test[12].

Supplementary Fig. 8 shows the sensitivity of residual consistency test to different EOF truncations. Two cases are compared for five-year mean PI when using area averages over the whole NH (10 dimensional time vector) and when using two regional averages combined (northern mid-latitudes + northern tropics, 20 dimensional space-time vector). ANT results show that the model variability is not rejected when compared to the observed residual variance and the detection results are robust when up to nine EOF patterns are retained for both cases. In contrast, ALL results often cannot pass the residual consistency test without doubling the model variability and the detection results are still a bit sensitive

to EOF truncations after the variability inflation. This indicates that there might be signal errors in ALL, which contribute to the residual variance (see below).

Along with the residual consistency test, we evaluate model-simulated internal variability by comparing variances estimated from observations and CTL runs (Supplementary Table 2). The observed NH mean increases lie outside the range of internal variability as represented by CTL simulations (Supplementary Fig. 7). We remove linear trends from the observed five-year mean PI time series before calculating variances, thereby leaving out anthropogenic influence in a simple manner. For models, we calculate variances for each five-year mean time series from CTL 49-year samples (106 chunks in total) after detrending them individually, and then estimated the mean and the $10^{th}$ and $90^{th}$ percentiles of variance estimates. Variances obtained from the observations lie well within the 10-90% ranges of CTL variance. This can be clearly seen when looking at actual time series of the detrended observations in comparison with CTL ranges (dashed black lines vs. grey lines in Supplementary Fig. 7); supporting the conclusion that multi-model CTL datasets used here can reasonably reproduce the internal variability of extreme precipitation indices that are averaged over large hemispheric land areas.

**4. Comparison of signal-to-noise ratios**

Using area-averaged time series of PI for the NH, we compared the detectability of the ANT signal in RX1D and RX5D changes during 1951-1999. In order to use the same spatial domain for averages, we here use $RX5D_{m1d}$ (RX5D masked by RX1D data availability, See Fig. 1 in the main paper) although results are found to be almost the same when using the original RX5D. We calculated the slope of linear trends in the PI time series for RX1D and $RX5D_{m1d}$ using ANT and CTL simulations. The mean ANT trend slope is

defined as "signal", the standard deviation of CTL trend slopes is defined as "noise" and the signal-to-noise ratio (SNR) is calculated as their ratio (Supplementary Table 3). We found that the RX1D signal amplitude is about 8% larger than that of RX5D in multi-model mean ANT runs. However, the SNR is very similar (about 1.85) in RX1D and RX5D, which is due to reduced noise amplitude in RX5D. Observational results show a stronger positive trend (about 11%) in RX1D than RX5D.

The same analysis of the ALL simulations produces estimated SNRs of 0.93 and 0.30 for RX1D and $RX5D_{m1d}$ respectively, which are much less than for ANT. This is because the natural forcing (NAT) employed in the ALL runs, produces a decreasing tendency in extreme precipitations on average, reducing the overall increasing trend in ALL as compared to the ANT runs, as discussed in the main text (see Supplementary Fig. 3).

As a consequence, local changes in ALL simulations will be noisier than in ANT simulations, giving more grid-points with negative and positive changes just because more noise variability will prevail against the signal (Fig. 1 in the main paper). This means that large-scale area-mean PI series will also become noisier for ALL, reducing some detection power. The larger error bounds of the scaling factors represent weaker signals simulated by the ALL models (Fig. 3)[12]. Considering that models underestimate observed changes[13], this implies that the ANT fingerprint would better match the observed changes than ALL, becoming more (and earlier) detectable. When we retain longer-term trend components of the area-mean PI time series (e.g., 4-5 EOFs, see Supplementary Fig. 9), ALL shows better agreement with observed patterns and becomes detectable as well as ANT (Supplementary Fig. 8). Using such low-frequency "fingerprints" patterns in the detection analysis gives rise to temporal smoothing, which reduces short-term increases at both ends of the record while retaining long-term trend components particularly for observations. That is, observed

changes become weaker than unfiltered raw series during both end periods, partly explaining better agreement with model simulations.

**5. Influence of the observed weakness of precipitation extremes during the 1950s**

Observations are characterized by lower PI values in the 1950s (Fig. 2 in the main paper). When comparing decadal mean PI patterns between the 1950s and 1990s (Supplementary Fig. 10), anomalous weakening of extreme precipitation events (lower PI) in the 1950s is pronounced over North America while intensification of extreme events (higher PI) in the 1990s is found over a large continental area. This localized feature indicates the possibility that observed lower PI in the 1950s might be driven by variability generated within the climate system, possibly in relation to the North American multi-decadal drought that some have suggested may be attributable to ocean influences[14,15] (a negative phase of the Pacific Decadal Oscillation combined with a positive phase of the Atlantic Meridional Oscillation in 1950s). Regarding long-term changes, there might be a concern that the relatively weak extremes in the first decade of the analysis period may enhance observational trends for 1951-1999 and therefore improve the detectability of anthropogenic signals.

We have therefore repeated our optimal detection analysis for the period of 1955-1999 (Supplementary Fig. 11). The power of detection becomes weaker for a shorter record due to decrease in the signal-to-noise ratio. Nevertheless, we can detect ANT robustly in both RX1D and RX5D, and also marginally for ALL depending upon the choice of EOF truncation. Residual consistency has been improved, indicating better agreement between modelled and observed variability due to the exclusion of the 1951-1954 period.

## 6. Influence of ENSO on observed changes

El Niño and Southern Oscillation (ENSO) significantly influences extreme precipitation[16,17] at local and regional scales in some parts of the world. However, the influence of ENSO on extreme precipitation is positive in some areas and negative in other areas, or the reverse, depending upon the phase of the ENSO cycle. That is, El-Nino events followed by La-Nina have opposing effects on extremes at a given location. As a result, the net influence of ENSO on continental scale means of precipitation extremes on times scales of a full ENSO cycle or longer is anticipated to be rather small. Our analysis as described below confirms this, and indicates that ENSO has not influenced on our detection and attribution results.

We have tested the sensitivity of our detection results by using ENSO-residual observations. In order to keep all calculations within the scope of extreme value theory, we have removed the ENSO influence from the observed PI time series by employing the observed Southern Oscillation Index (SOI) as a covariate in our GEV model[16]. The SOI has not been detrended and thus change in PI associated with the secular trend in ENSO would also have been removed although it is unclear whether the ENSO trend is due to global warming or not. We make the location and scale parameter of the GEV distribution dependent upon the observed SOI time series, and then calculate annual PI series by using the SOI-adjusted GEV parameters. We used the SOI index averaged for the winter months (Nov to Mar) when the ENSO influence on extreme precipitation is expected to be strongest[17]. We found a weaker influence on our detection results when using summer mean (May to Sep) or annual SOI (not shown). Results show that the ENSO influence on area mean PI series is almost negligible on longer time scales (Supplementary Fig. 12).

During 1995-1999, we find that ENSO reduced PI in 1998 but increased PIs for 1997 and 1999, resulting in five-year mean values similar to raw data, consistent with our expectation that the ENSO cycle influences at any one place would tend to average out during a given five-year period. Larger impacts rather occur in the early 1980s and early 1990s, consistent with the SOI time series. Detection results for 1951-1999 are found to be insensitive to whether or not ENSO-residual PI observations are used in the analysis (Supplementary Fig. 12).

**7. Influence of observational data coverage change and spatial interpolation**

There might be a concern about possible influence of changes in observational data coverage on detection results. Here we examine the sensitivity of the observed PI time series to different thresholds on data availability. Supplementary Fig. 13 compares the number of grid-points and corresponding observed PI time series for RX1D averaged over the whole NH, North America, and Eurasia when applying different thresholds (a single criterion requiring more than 40 years out of 1951-1999 versus dual criteria that require at least 40 years in total and at least six years in each decade). Overall we find only a negligible influence from data coverage in the observed PI time series. The main difference occurs in the Eurasian region, which displays a weaker increasing trend in PI compared to North America in the early decade (See above for discussion on weakened extreme events in 1950s). In contrast, note that there is no coverage change with time in the model time series because we have applied a "fixed" observational mask to model simulated PI, excluding all 3.75° × 2.5° grid-points that had less than 40 years of observations out of 49 years. This was to avoid the influence of the coverage artefact and was also partly related to our interpolation method of multi-model datasets with different horizontal resolutions.

When additionally considering the influence of applying a space-time varying observational mask to models, there was little changes in detection results (see Supplementary Fig. 14).

While our main analysis was performed on the HadEX 3.75° × 2.5° grid, we also performed the analysis on a coarser 5° × 5° grid that results in a smoother spatial representation of PI for both observations and models since 5° × 5° grid have approximately 2.7 times the area of 3.75° × 2.5° grid boxes. When interpolating both observed and simulated datasets onto the same 5° × 5° grid prior to optimal detection analysis, overall detection results are not changed (not shown). This suggests that the degree of smoothing that results from interpolation to different grid spacings does not affect our results.

**8. Discussion on reliability of observational record**

Here we provide some simple analyses and discussion that attempts to address questions about the reliability of the observational record.

First, in an attempt to increase our confidence in the HadEX dataset, we have conducted an experimental comparison of spatial patterns of trends in PI for RX1D over North America where we have a relatively high density of precipitation stations, which may provide some hint on uncertainties arising from different station datasets and gridding methods. We have compared three trend patterns for 1951-1999 (Supplementary Fig. 15) - (1) trends in local PI values at stations (GHCN data set[18], requiring at least 40 yrs; left panel), (2) trends in PI values from gridded RX1D from stations (requiring at least 2 stations for gridding and taking a simple average within each grid-box; upper right panel), and (3) trends in PI values from HadEX (middle right panel). Note that the pattern (2) is

from a HadEX-like product and that we use the HadEX grid (3.75° × 2.5°) for this comparison. Trend patterns show generally similar structure, especially with similar amplitudes of area-averaged trends. The difference pattern between (2) and (3) (bottom right panel) is characterized by spatial noise such that larger positive or negative values appear randomly over some grid-boxes and the area average becomes near zero. Note also that the amplitude of differences between the two gridded products is similar across the domain, even though the density of observing stations is very much lower in the Western USA than in the Eastern USA. This experimental result indicates the reliability of HadEX datasets in terms of large-scale trend patterns.

We also have compared, in models, five-year mean time series of regional mean PI from the full region and from the observation-masked region. Supplementary Fig. 16 shows results for NH, NH mid-latitude land areas, and Eurasia. The five-year mean series show very similar temporal variations and long-term trends. Therefore, assuming gridbox quantities are reasonably estimated (see above), low-frequency large-scale area means can be estimated with increased confidence.

In addition, the response of extreme precipitation to an external forcing is likely to be spatially homogeneous on scales much larger than the gridbox scale, even if spatial variation in unforced precipitation variability does not exhibit strong spatial covariance structure. A recently published study, for example, illustrates that precipitation extremes in the far tails respond in a fairly homogenous manner to ENSO sea surface temperature (SST) forcing[16]. Presumably this might also be true for greenhouse-gas forcing.

Change in observing practices would have influenced the measurement of precipitation and therefore extreme rainfall events at individual stations. This includes the use of more efficient rain gauges with smaller wetting losses and changes in station location

that may have increased or decreased the measured amount. A change in the type of gauge that is used to measure precipitation is usually system wide over an entire country and is usually implemented over a few years. Such changes have the potential to cause a gradual shift in precipitation measurements, but the dominant influence is typically on light precipitation events, and is negligible for heavy precipitation events as we used in this study[19,20]. Changes in station locations can result in large shifts in precipitation measurements, affecting both light and heavy rainfall events. As a result, a station location change could cause a shift in the annual extreme precipitation. However, station location changes are localized and more or less randomly distributed in space and time, as opposed to being system-wide. It is generally accepted that, to first order, the impacts from such changes, resulting in upward shifts in precipitation measurement at some locations and downward shifts elsewhere, should cancel out over large regions. As a result, the average of extremes over a large region should not be affected by changes in station location.

**9. Discussion of the physical mechanisms for extreme precipitation changes**

The physical mechanisms for extreme precipitation changes have been extensively investigated by recent studies. Therefore, we provide arguments here based on those studies. There is a general consensus that extreme precipitation changes are expected to be reasonably well explained by the Clausius-Clapeyron relationship over the NH mid-latitudes where the atmospheric circulation is expected to change little[21-24]. There is also a recent study supporting the Clausius-Clapeyron relationship even on smaller scales over the NH mid-latitude lands. Isolating mechanisms for future extreme precipitation changes focusing on Europe, Kendon et al. (2010, ref. 25) show that large scale changes (> 2,000 km) in winter extreme precipitations, which are expected to determine annual maxima over

the region, are dominated by moisture increase with warming. It should, however, be noted that there might be relatively weaker changes over land than over oceans due to possible change of relative humidity with stronger warming[26].

Nevertheless, the studies cited above are largely based on future simulations under strong GHG forcing, and so we cannot confirm that such a mechanism is reliably applicable to the historical period with weaker GHG forcing. In this respect, we have conducted a simple analysis that considers relationships between PI and surface air temperature (SAT). We regressed the five-year mean PI series onto SAT series for 1951-1999, for which all datasets have been interpolated into the 5° × 5° grid of SAT observations (CRUTEM3v)[27] and the observed mask for PI has been considered before analysis. Results show that long-term changes in extreme precipitation generally match temperature changes supporting the Clausius-Clapeyron relationship (Supplementary Fig. 17). We obtain an observed regression coefficient of ~7.2%/K for RX1D while the multi-model mean ANT and ALL results show 3.6-3.8%/K, which is about a half of observed change. Considering that observed SAT trends are well simulated by the models, we must attribute this difference to the models' underestimation of the observed changes in extreme precipitation. When looking at individual model runs, there is large spread due to inter-model uncertainties with a standard deviation 1.7%/K as discussed in O'Gorman and Schneider (2009, ref. 23). When considering SAT averaged over a broader area by comparing PI for RX1D with SAT over the RX5D mask, model results are not affected much and only a small decrease appears in the observations, giving 6.9%/K (not shown). When using 1955-1999, we get 6.5%/K for observations, still much stronger than models (not shown). RX5D gives weaker PI-SAT relationship than RX1D (5.5%/K for 1951-1999, Supplementary Fig. 17), as might be expected.

We now briefly discuss some possible causes of inconsistency with the C-C relation for both observations and models. For models, deviations from the C-C relation can occur depending upon latitude and whether the region being considered is over ocean or land[22,26]. We consider mainly land areas over the Northern mid-latitudes, which may be one of the major factors causing the inconsistency of models with the theory. Second, it is known that under strong GHG forcing, model-simulated increases in precipitation extremes generally do follow the C-C equation[28], although there is great variability between models. It is possible that the relationship may become weaker or noisier under weaker GHG forcing such as for the latter half of the 20th century. For observations, the influence of low-frequency natural noise may be more apparent in the diagnosed C-C relationship compared to models for which ensemble averaging will cancel out the influence of natural mode's that are in different phases in different ensemble members.

On the other hand, the aerosol influence on extreme precipitation is an issue that is still scientifically unclear and speculative[29]. One can think of several mechanisms related to aerosols, such as change in shortwave forcing, control on energy balance in the atmospheric column through heating by absorbing aerosols, and changes in microphysical properties, that may affect precipitation.

## References


1. Alexander, L. V. *et al.* Global observed changes in daily climatic extremes of temperature and precipitation. *J. Geophys. Res.* **111**, D05109, doi:10.1020/2005JD006290 (2006).

2. Caesar, J., Alexander, L. & Vose, R. Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *J. Geophys. Res.* **111**, D05101, doi:10.1029/2005JD006280 (2006).

3. Peterson, T. C., Zhang, X., Brunet-India, M. & Vázquez-Aguirre, J. L. Changes in North American extremes derived from daily weather data. *J. Geophys. Res.* **113**, D07113, doi:10.1029/2007JD009453 (2008).

4. Gleason, K. L., Lawrimore, J. H., Levinson, D. H., Karl, T. R. & Karoly, D. J. A revised U.S. Climate Extremes Index. *J. Clim.* **21**, 2124–2137 (2008).

5. Yao, C., Yang, S., Qian, W., Lin, Z. & Wen, M. Regional summer precipitation events in Asia and their changes in the past decades. *J. Geophys. Res.* **113**, D17107, doi:10.1029/2007JD009603 (2008).

6. Ning, L. & Qian Y. F. Interdecadal change in extreme precipitation over South China and its mechanism. *Adv. Atmos. Sci.* **26**, 109-118 (2009).

7. Zolina, O., Simmer, C., S. K. Gulev, S. K. & Kollet, S. Changing structure of European precipitation: Longer wet periods leading to more abundant rainfalls. *Geophys. Res. Lett.* **37**, L06704, doi:10.1029/2010GL042468 (2010).

8. Meehl, G. A. *et al.* The global coupled climate multi-model dataset: A new era in climate change research. *Bull. Am. Meteorol. Soc.* **88**, 1383-1394 (2007).

9. Coles, S. *An Introduction to Statistical Modeling of Extreme Values.* (Springer, Berlin, 2001)

10. Kharin, V. V. & Zwiers, F. W. Estimating extremes in transient climate change simulations. J. Climate, 18, 1156–1173 (2005).

11. Martins, E. S. & Stedinger, J. R. Generalized maximum likelihood generalized extreme-value quantile estimators for hydrological data. *Water Resour. Res.* **36**, 737–744 (2000).

12. Allen, M. R. & Tett, S. F. B. Checking for model consistency in optimal fingerprinting. *Clim. Dyn.* **15**, 419-434 (1999).

13. Allan, R. P. & Soden, B. J. Atmospheric warming and the amplification of precipitation extremes. *Science* **321**, 1481-1484 (2008).

14. McCabe, G. J., Palecki, M. A. & Betancourt, J. L. Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. *Proc. Natl Acad. Sci. USA* **101**, 4136–4141 (2004).

15. Sutton, R. T. & Hodson, D. L. R. Atlantic Ocean forcing of North American and European summer climate. *Science* **309**, 115-118 (2005).

16. Zhang, X., Jiafeng, W., Zwiers, F. W. & Groisman, P. Ya. The Influence of large-scale climate variability on winter maximum daily precipitation over North America. *J. Clim.* **23**, 2902-2915 (2010).

17. Kenyon, J. & Hegerl, G. C. Influence of modes of climate variability on global precipitation extremes. *J. Clim.* in press (2010).

18. Gleason, B., Groisman, P., Peterson, T. C., Vose, R. & Ezell R. A new global daily temperature and precipitation dataset, presented at 13th Symposium on Global Change and Climate Variations, Am. Meteorol. Soc. Orlando, FL, USA (2002).

19. Habib, E., Krajewski, W. F. & Kruger, A. Sampling errors of tipping-bucket rain gauge measurements. *J. Hydrol. Eng.* **6**, 159-166 (2001).

20. Ciach, G. J. Local random errors in tipping-bucket rain gauge measurements. *J. Atmos. Oceanic Technol.* **20**, 752-759 (2003).

21. Emori, S. & Brown, S. J. Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate. *Geophys. Res. Lett.* **32**, L17706, doi:10.1029/2005GL023272 (2005).

22. Pall, P., Allen, M. R. & Stone, D. A. Testing the Clausius-Clapeyron constraint on changes in extreme precipitation under $CO_2$ warming. *Clim. Dyn.* **28**, 351-363 (2007).

23. O'Gorman, P. A. & Schneider, T. The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proc. Natl Acad. Sci. USA* **106**, 14773-14777 (2009).

24. Sugiyama, M., Shiogama, H. & Emori, S. Precipitation extreme changes exceeding moisture content increases in MIROC and IPCC climate models, *Proc. Natl Acad. Sci. USA* **107**, 571-575 (2010).

25. Kendon, E. J., Rowell, D. P. & Jones, R. G. Mechanisms and reliability of future projected changes in daily precipitation. *Clim. Dyn.* **35**, 489–509 (2010).

26. O'Gorman, P. A. & Muller, C. J. How closely do changes in surface and column water vapor follow Clausius–Clapeyron scaling in climate change simulations? *Environ. Res. Lett.* **5** 025207, doi:10.1088/1748-9326/5/2/025207 (2010).

27. Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B. & Jones, P. D. Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *J. Geophys. Res.* **111**, D12106, doi:10.1029/2005JD006548 (2006).

28. Kharin, V. V., Zwiers, F. W., Zhang. X. & Hegerl, G. C. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Clim.* **20**, 1419-1444 (2007).

29. Khain, A. P. Notes on state-of-the-art inverstigations of aerosol effects on precipitation: a critical review. *Environ. Res. Lett.* **4**, 015004 (2009).

30. Collins, W. D. *et al.* The Community Climate System Model: CCSM3. *J. Clim.* **19**, 2122–2143 (2006).

31. Flato, G. M. The Third Generation Coupled Global Climate Model (CGCM3) (and included links to the description of the AGCM3 atmospheric model). http://www.cccma.bc.ec.gc.ca/models/cgcm3.shtml (2005).

32. Gordon, H. B. *et al.* The CSIRO Mk3 Climate System Model. CSIRO Atmospheric Research Technical Paper No. 60, Commonwealth Scientific and Industrial Research Organisation Atmospheric Research, Aspendale, Victoria, Australia, 130 pp, http://www.cmar.csiro.au/e-print/open/gordon_2002a.pdf (2002).

33. Jungclaus, J. H. *et al.* Ocean circulation and tropical variability in the AOGCM ECHAM5/MPI-OM. *J. Clim.* **19**, 3952–3972 (2006).

34. Min, S.-K., Legutke, S., Hense, A. & Kwon, W.-T. Climatology and internal variability in a 1000-year control simulation with the coupled climate model ECHO-G - I. Near-surface temperature, precipitation and mean sea level pressure. *Tellus* **57A**, 605–621 (2005).

35. Delworth, T. *et al.* GFDL's CM2 global coupled climate models - Part 1: Formulation and simulation characteristics. *J. Clim.* **19**, 643–674 (2006).

36. Washington, W. M. *et al.* Parallel Climate Model (PCM) control and transient simulations. *Clim. Dyn.* **16**, 755–774 (2000).

**Supplementary Table 1.** List of the eight coupled climate models used in this study: ANT (anthropogenic forcing) and ALL (natural plus anthropogenic forcing) runs for 1951-1999 and CTL (preindustrial control) runs. The number indicates the number of ensemble members for ANT and ALL and the number of 49-year non-overlapping chunks for CTL. Note that we have also used the ECHO-G 1,000-year present-day control simulation (20 49-year chunks, marked as asterisk). See SI text for detailed model data processing. Access to the data archives of individual modelling centres was obtained via the URLs listed below the Table.

| Model[#] | ANT 1951-1999 [# of runs] | ALL 1951-1999 [# of runs] | CTL [# of 49-year chunks] |
|---|---|---|---|
| CCSM3[30] | 3 | 4 | 10 |
| CGCM3[31] | 5 | - | 20 |
| CSIRO-Mk3.0[32] | 1 | - | 6 |
| ECHAM5/MPI-OM[33] | 3 | - | 10 |
| ECHO-G[34] | 3 | 3 | 6+20* |
| GFDL CM2.0[35] | - | 3 | 10 |
| GFDL CM2.1[35] | - | 3 | 10 |
| PCM[36] | 4 | 3 | 14 |
| Sum | 6 models, 19 runs | 5 models, 16 runs | 8 models, 106 chunks |

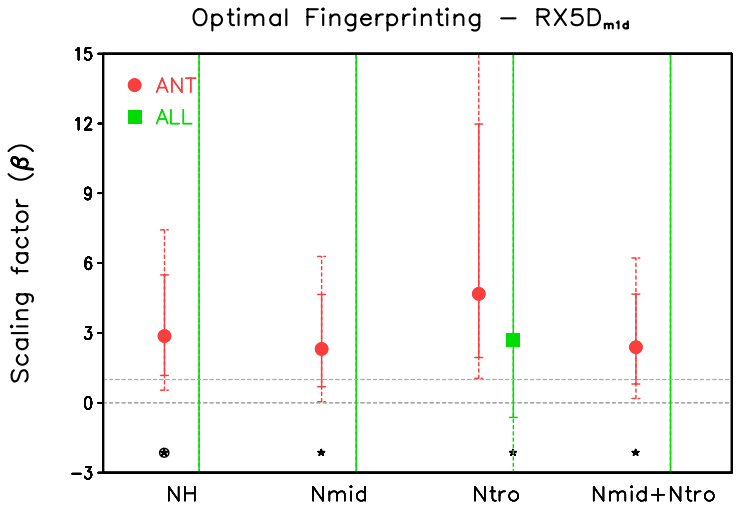
[#]Data archive URLs
CCSM3 and PCM: http://www.earthsystemgrid.org
CGCM3: http://www.cccma.bc.ec.gc.ca/data/
CSIRO-Mk3.0: http://ngportal.sf.utas.edu.au/gridsphere/
ECHAM5/MPI-OM and ECHO-G: http://www.mad.zmaw.de/wdc-for-climate/
GFDL CM2.0 and GFDL CM2.1: http://nomads.gfdl.noaa.gov

**Supplementary Table 2.** Comparison of temporal variances of five-year mean NH area averaged PI for RX1D and RX5D between detrended observations ($OBS_{dtr}$) and CTL runs. Each CTL PI time series (106 chunks in total, see Supplementary Table 1) is detrended individually before analysis. $P_{10}$ and $P_{90}$ are the 10th and 90th percentiles of CTL variances. See Supplementary text for details. Units are $\%^2$.

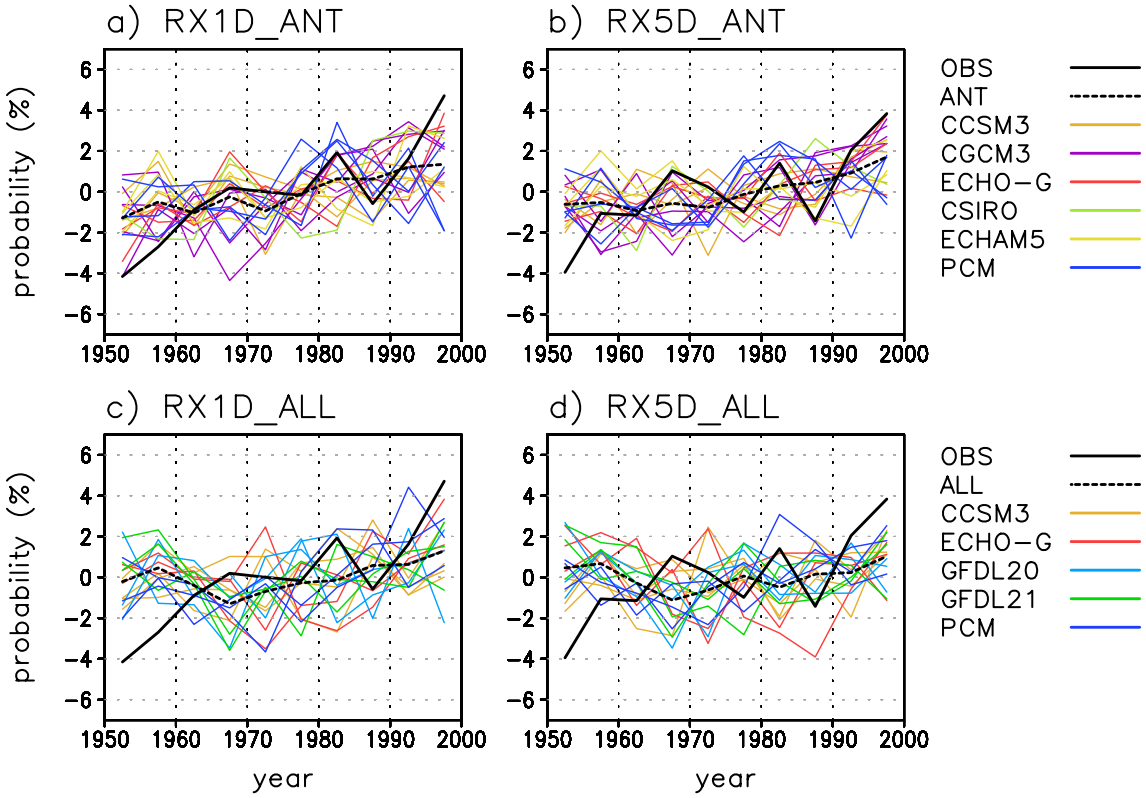
| | $OBS_{dtr}$ | CTL | | |
|---|---|---|---|---|
| | | Mean | $P_{10}$ | $P_{90}$ |
| RX1D | 1.45 | 1.85 | 0.70 | 3.21 |
| RX5D | 2.04 | 1.60 | 0.59 | 2.76 |

**Supplementary Table 3.** Comparison of signal amplitude and signal-to-noise ratios (SNR) between PI for RX1D and $RX5D_{m1d}$ (RX5D masked by RX1D observations) in climate model simulations. SNR is estimated from linear trend slopes of five-year mean PI series averaged over the Northern Hemisphere for 1951-1999 from ANT, ALL and CTL runs. The ANT or ALL mean slope is defined as the signal ($S_{ANT}$ or $S_{ALL}$) and the standard deviation of 106 CTL trends is defined as the noise (N). SNR (values in parentheses) is then defined as the ratio of S to N. Observational trend slopes ($S_{OBS}$) are given for comparison. Units for S and N are % per 49 years.

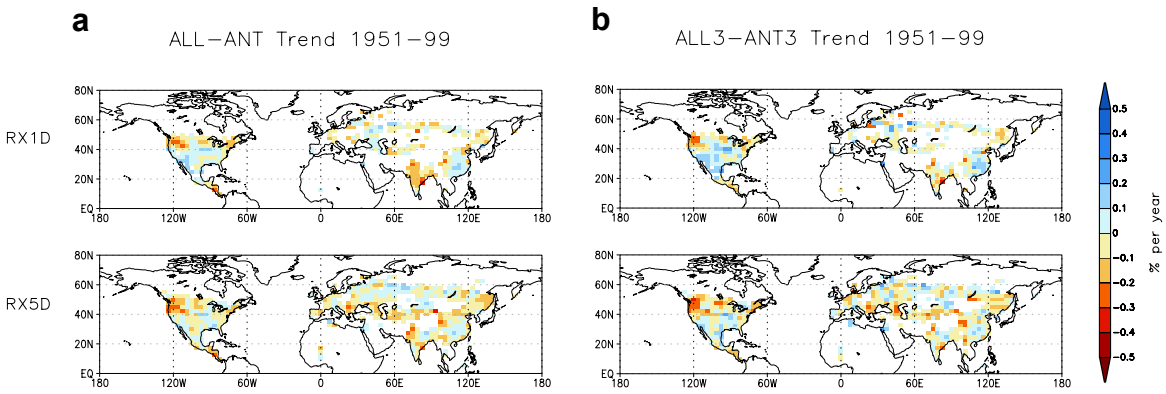
| | $S_{ANT}$ ($SNR_{ANT}$) | $S_{ALL}$ ($SNR_{ALL}$) | N | $S_{OBS}$ |
|---|---|---|---|---|
| RX1D | 2.88 (1.86) | 1.43 (0.93) | 1.54 | 7.07 |
| $RX5D_{m1d}$ | 2.67 (1.85) | 0.42 (0.30) | 1.44 | 6.39 |
| Ratio ($RX1D/RX5D_{m1d}$) | 1.08 (1.01) | 3.37 (3.14) | 1.08 | 1.11 |



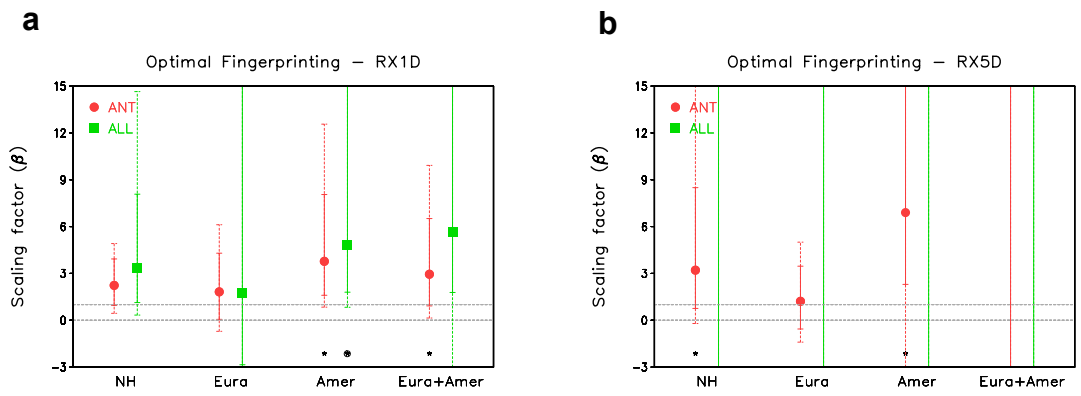

**Supplementary Figure 1 | Results from optimal detection analysis for RX5D when using the observational mask of RX1D**. Same as in Fig. 3b but for when considering the observational availability of RX1D (referred to as RX5D$_{m1d}$). Here the asterisk or circled asterisk indicates failure of the residual consistency test[12] when the original or the doubled internal variability was used respectively due to too small variability of the models.



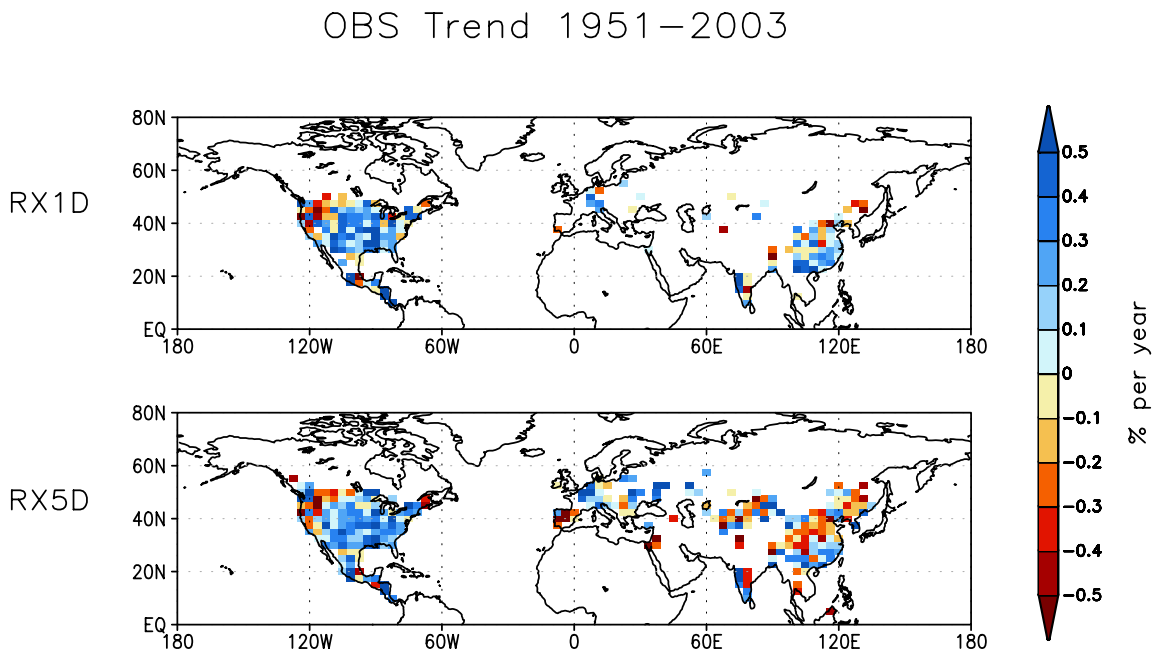

**Supplementary Figure 2 | Time series of five-year mean area-averaged PI anomalies over the NH land.** Same as in Fig. 2 except individual ANT and ALL simulations are shown (thin coloured lines).



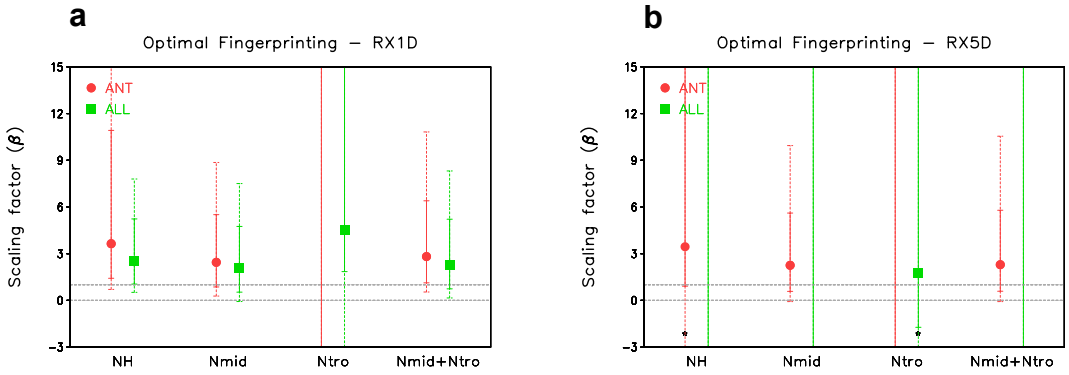

**Supplementary Figure 3 | Geographical patterns of differences between ALL and ANT trends.** Same as in Fig. 1 except for difference in ALL and ANT trends using all available models (left) and using the three models (right) that provide both ALL and ANT datasets (CCSM3, ECHO-G, and PCM, see Supplementary Table 1). Results suggest that natural forcing over that period leads to a reduction in precipitation extremes in many regions, although uncertainty due to climate variability is substantial.



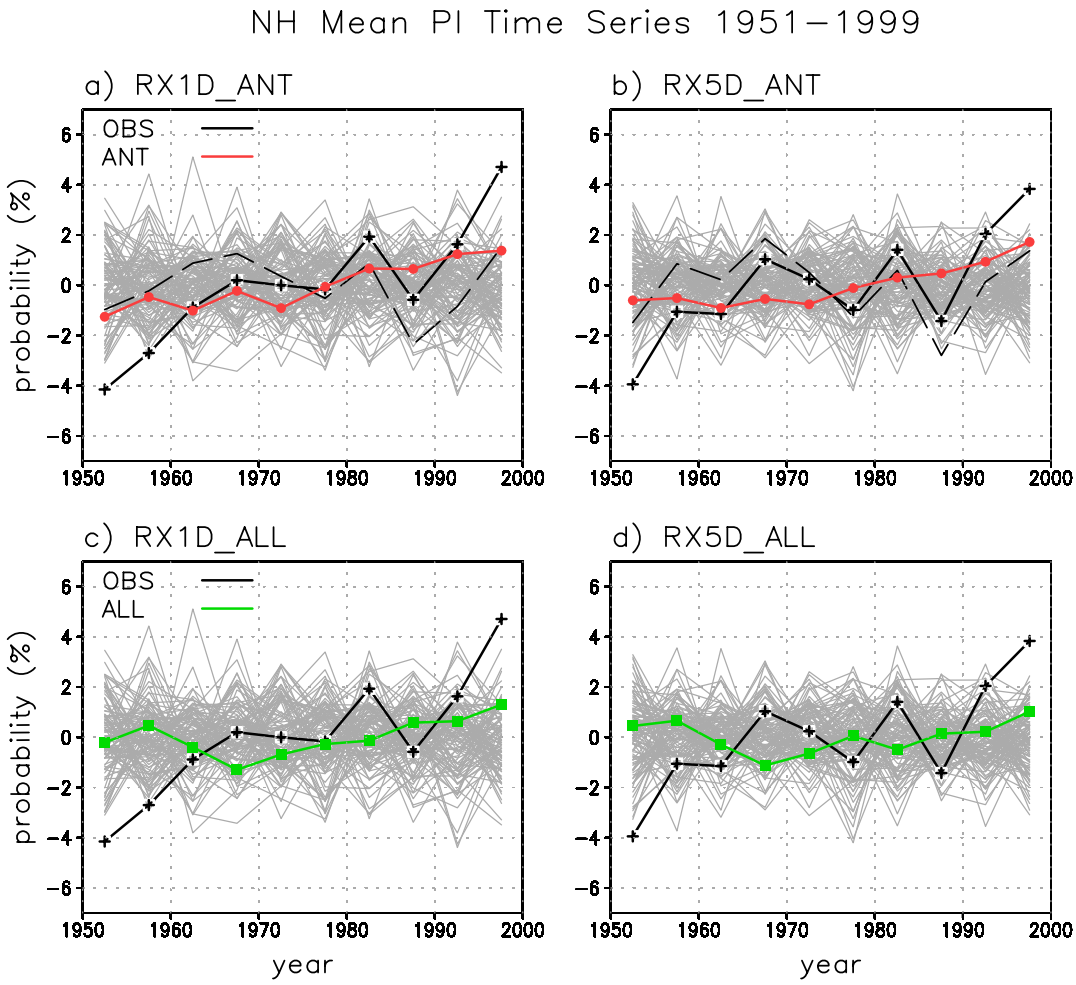

**Supplementary Figure 4 | Results from optimal detection analyses when considering Eurasia and America as spatial domains.** Same as in Fig. 3 but for when using PI averaged over NH (displayed again for easier comparison), Eurasia (Eura), America (Amer), and both combined (Eura + Amer). An asterisk or circled asterisk indicates failure of the residual consistency test when the original or the doubled internal variability was used respectively. Five-year mean PI anomalies are projected onto the four leading EOFs of internal variability, which explain about 43-64% of the total variance in these regions.



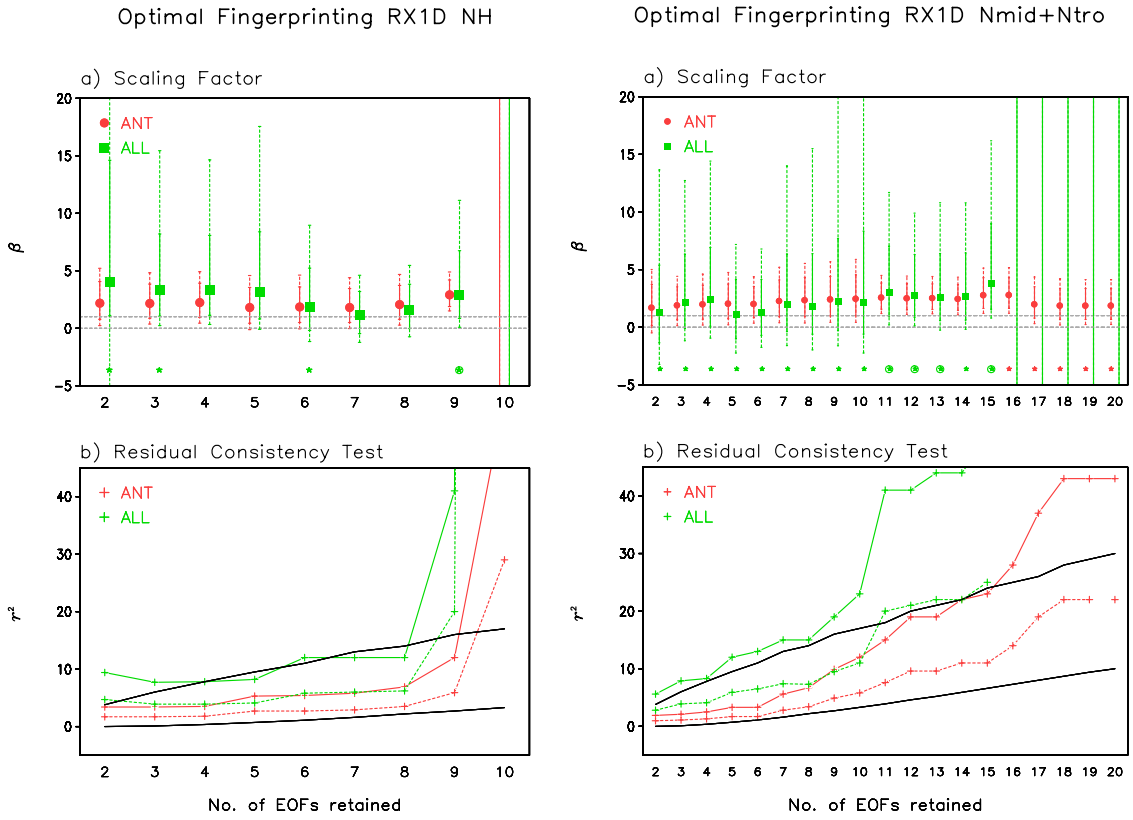

**Supplementary Figure 5 | Geographical patterns of observed trends in extreme precipitation indices during 1951-2003.** Note that here we consider only grid-points that have at least two years of data during 2000-2003, in addition to the requirement for more than 40 years during 1951-1999, so as to also represent a sudden drop of observed data availability after 2000. See SI text for more details.



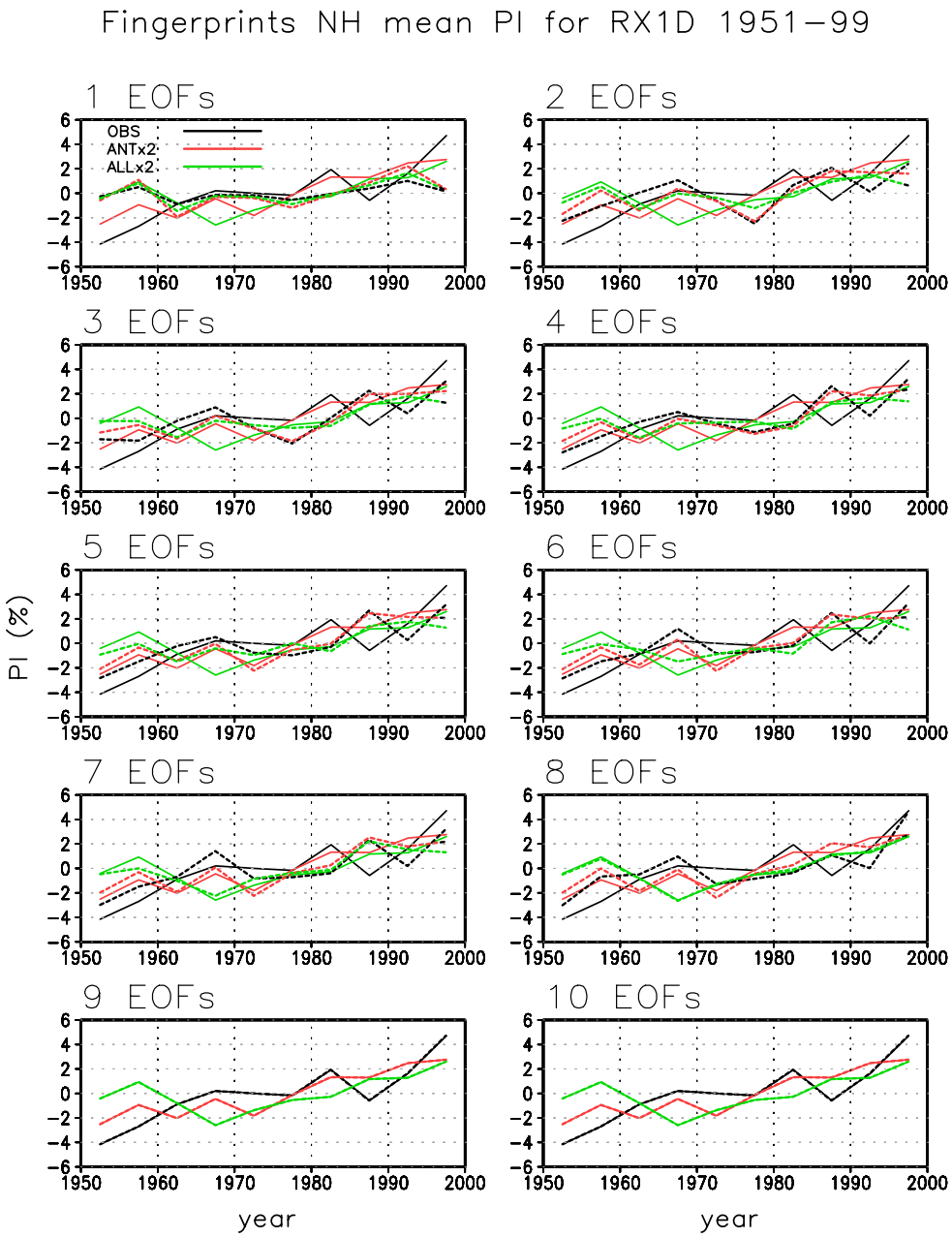

**Supplementary Figure 6 | Sensitivity of optimal detection results to model sampling.** Same as Fig. 3 but when using three models that provide both ANT and ALL runs to estimate fingerprints (CCSM3, ECHO-G, and PCM, see Supplementary Table 1 for details).



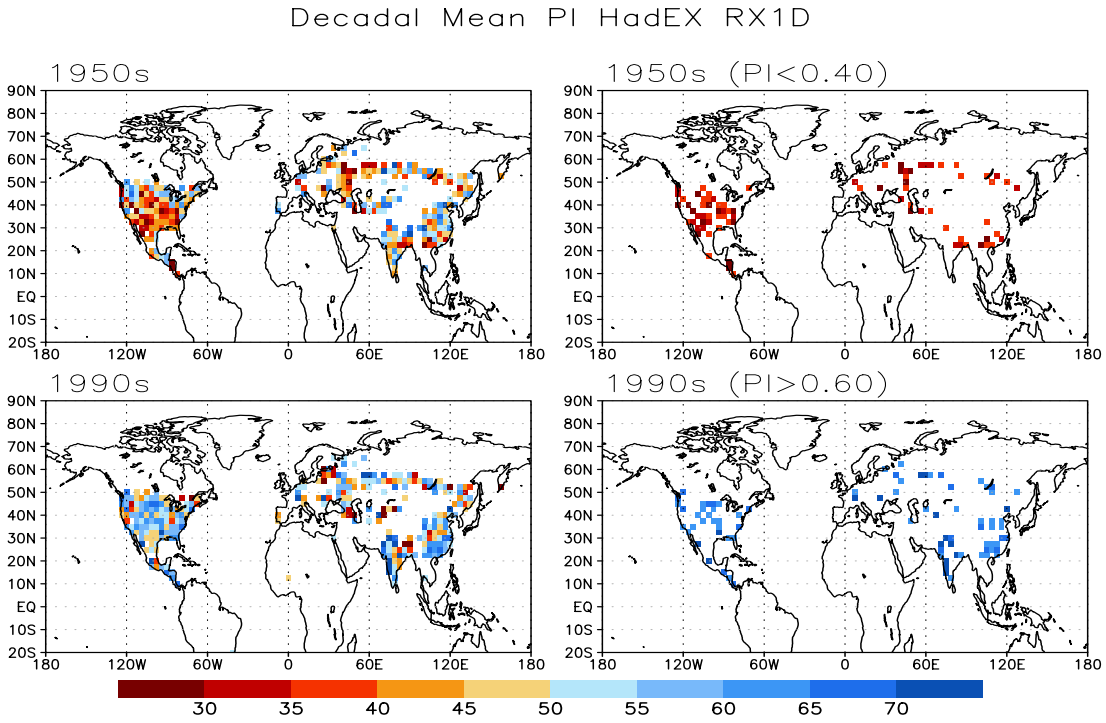

**Supplementary Figure 7 | Time series of area-averaged PI anomalies over the NH land.** Same as in Fig. 2 except for displaying ranges (grey lines) of CTL simulations together with observations (black solid) and multi-model mean ANT (red) and ALL (green) simulations. Black dashed lines represent detrended observations.



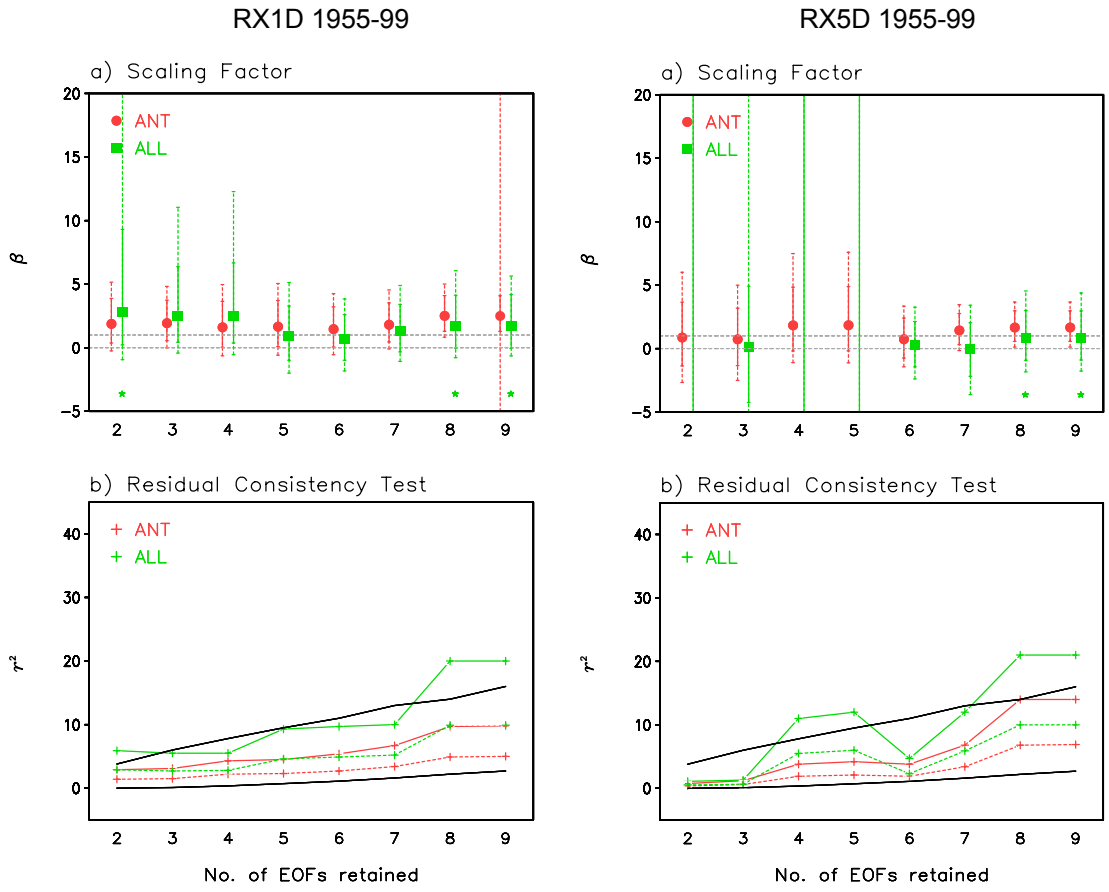

**Supplementary Figure 8 | Sensitivity of optimal detection results to the number of EOF patterns retained.** (left) Five-year mean PI time series for RX1D averaged over the NH (10 dimensions), and (right) a space-time vector with northern mid-latitude and tropics combined (Nmid + Ntro, 20 dimensions). In each panel, **a** regression coefficients and **b** residual consistency tests[12]. In **a** an asterisk or circled asterisk indicates failure of the residual consistency test when the original or the doubled internal variability was used respectively. In **b** coloured solid (dashed) lines represent residual variances ($r^2$) when the original (doubled) internal variability was used. Black lines depict 5-95% ranges of $\chi^2$ distribution. If the coloured line moves outside of the range, the model variability will be unreliable ("too small" if it moves above 95% line or "two large" if below 5% line).



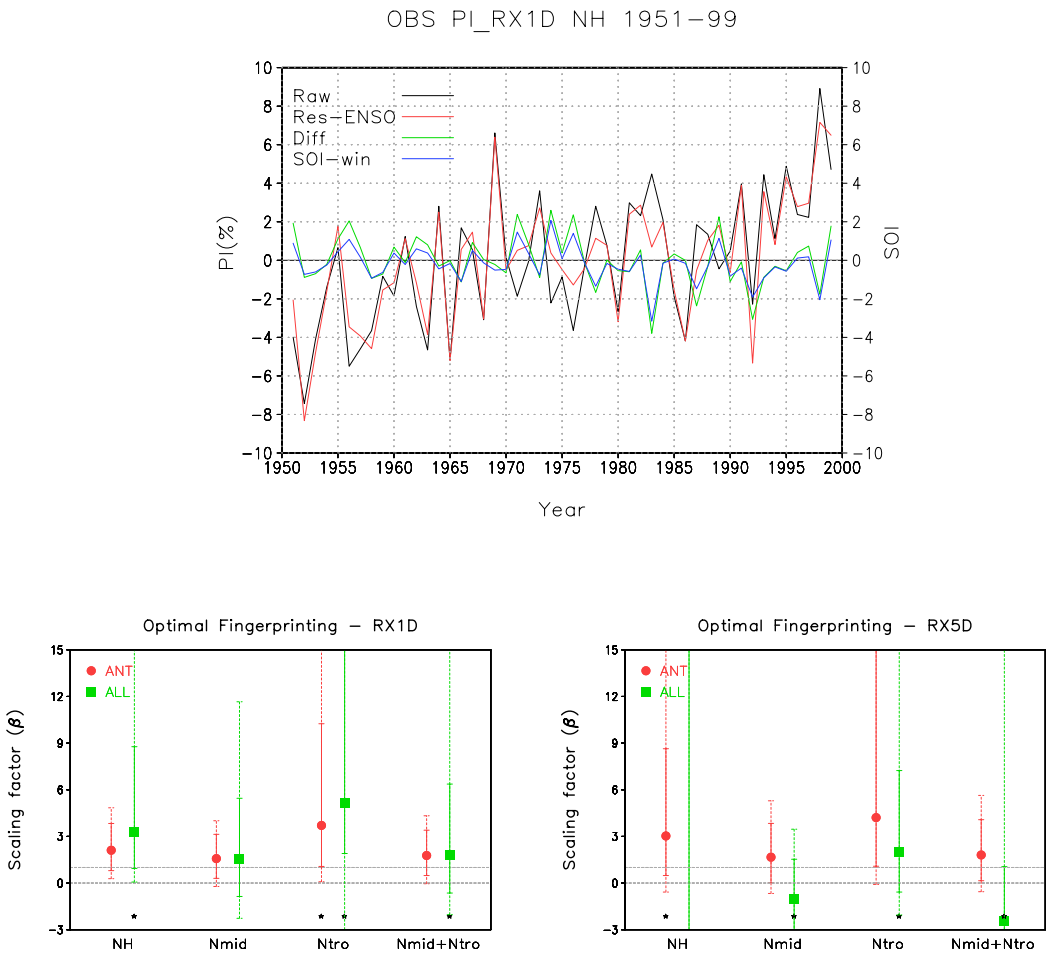

**Supplementary Figure 9 | Fingerprint patterns for RX1D.** Five-year mean NH mean PI anomaly series during 1951-1999 for different number of leading EOFs retained. EOF filtered series (dashed lines) are compared with unfiltered raw series for observations (OBS, black), ANT (green), and ALL simulations (red). Note that model values are doubled for display purpose.



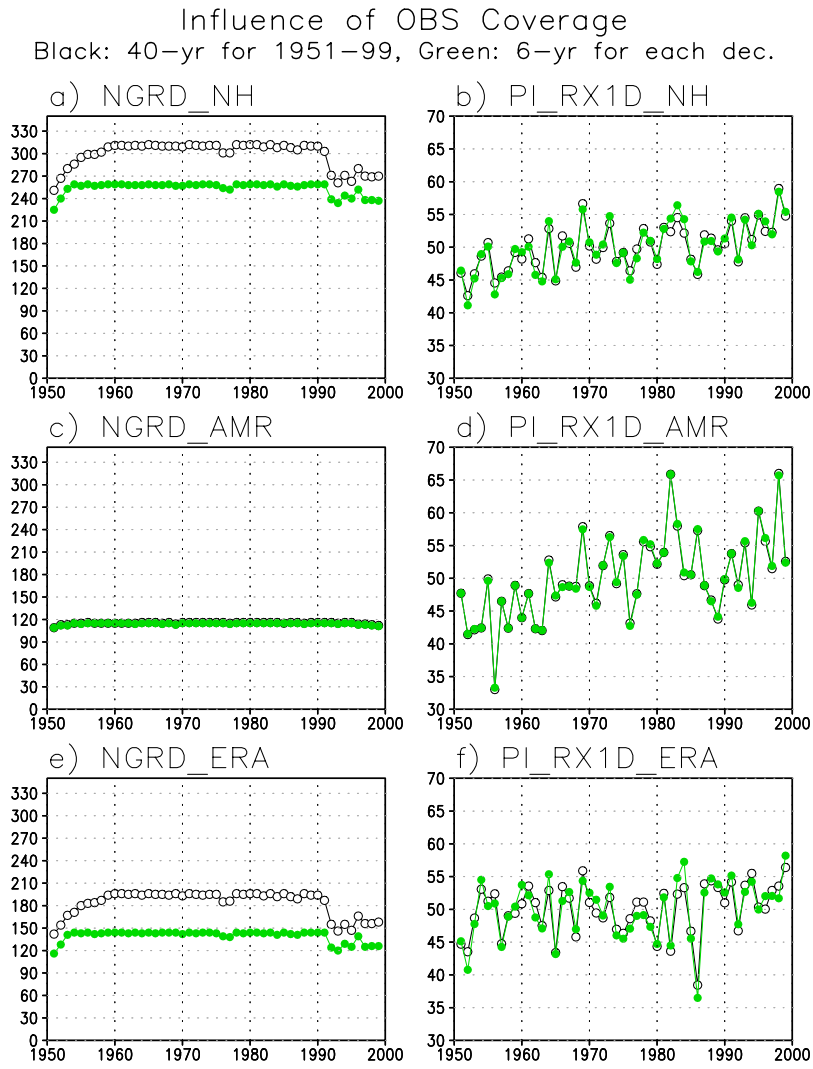

**Supplementary Figure 10 | Comparison of decadal mean PI patterns for RX1D between 1950s (1951-1959) and 1990s (1990-1999).** Left panels show decadal mean values over all available grid-points while right hand panels display grid-points with stronger amplitudes (PI less than the 40% in 1950s and larger than 60% in the 1990s respectively). Units are %.



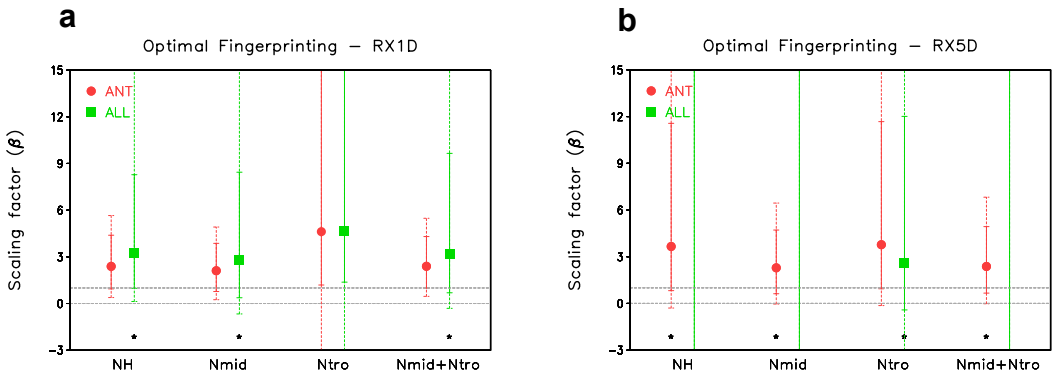

**Supplementary Figure 11 | Sensitivity of optimal detection results to a shortened analysis period.** Same as in Supplementary Fig. 8 but for the shorter period 1955-1999 (9 dimensions) using five-year mean NH averaged PI for RX1D (left) and RX5D (right).



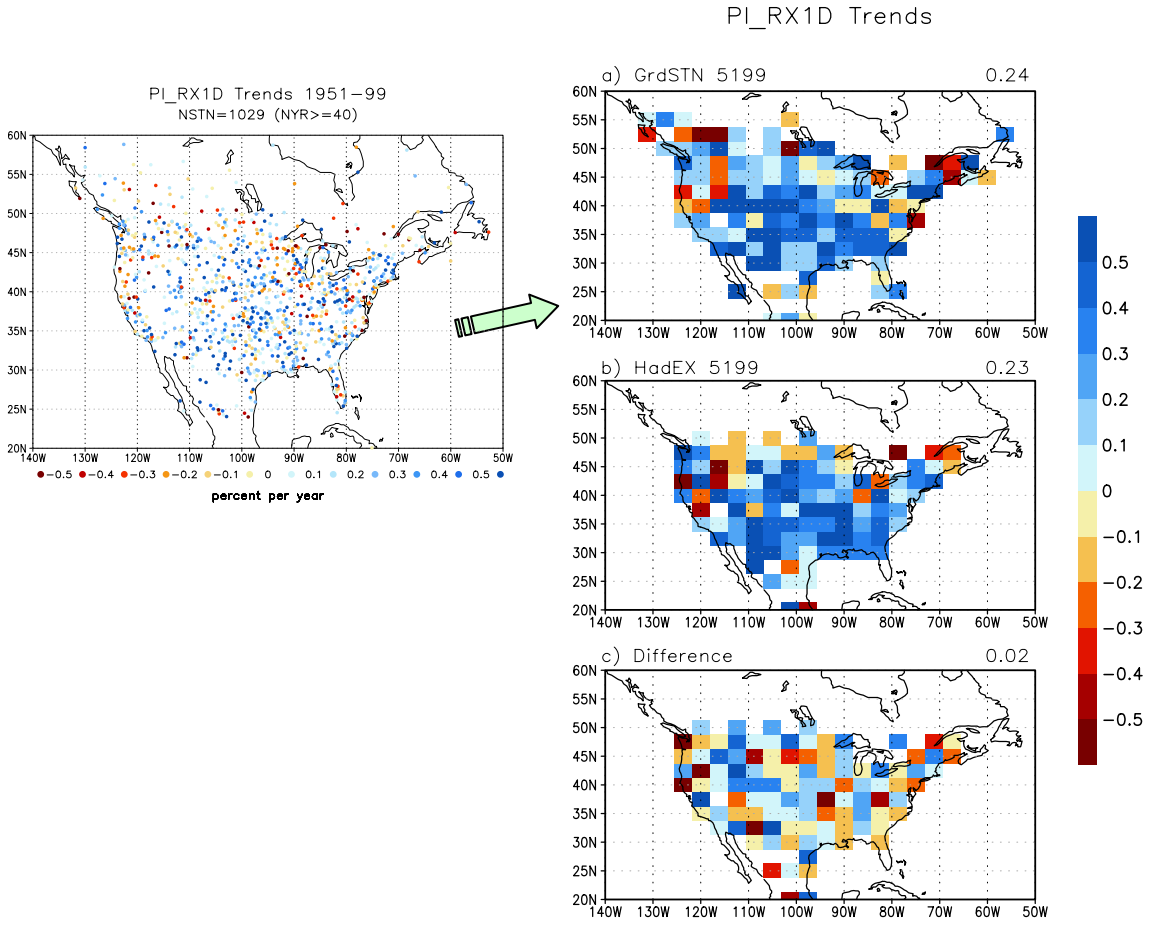

**Supplementary Figure 12 | Influence of ENSO-residual observations.** (upper) Annual time series of RX1D observations, ENSO-residual observations, their difference, and winter SOI index for 1951-1999, (lower) detection results same as Fig. 3 but for when using ENSO-residual observations.



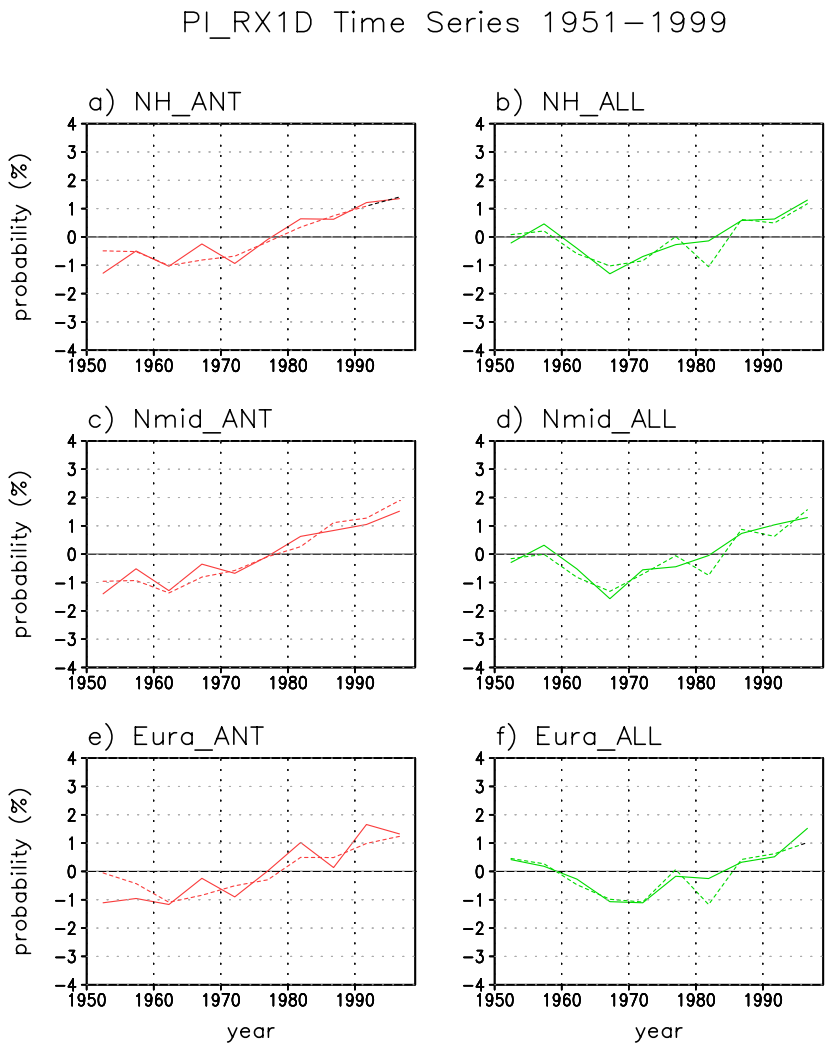

**Supplementary Figure 13 | Influence of observational data coverage change.** Annual time series of number of grid-points with datasets (left) and observed time series of PI for RX1D (right) averaged over the Northern Hemisphere (top), North America (middle), and Eurasia (lower) for different thresholds for observational data coverage. Black lines represent results when more than 40 years in the period 1951-1999 are required (as in this study) while green lines show a more strict case which requires more than six annual data for each decade as well as a minimum of 40 years.



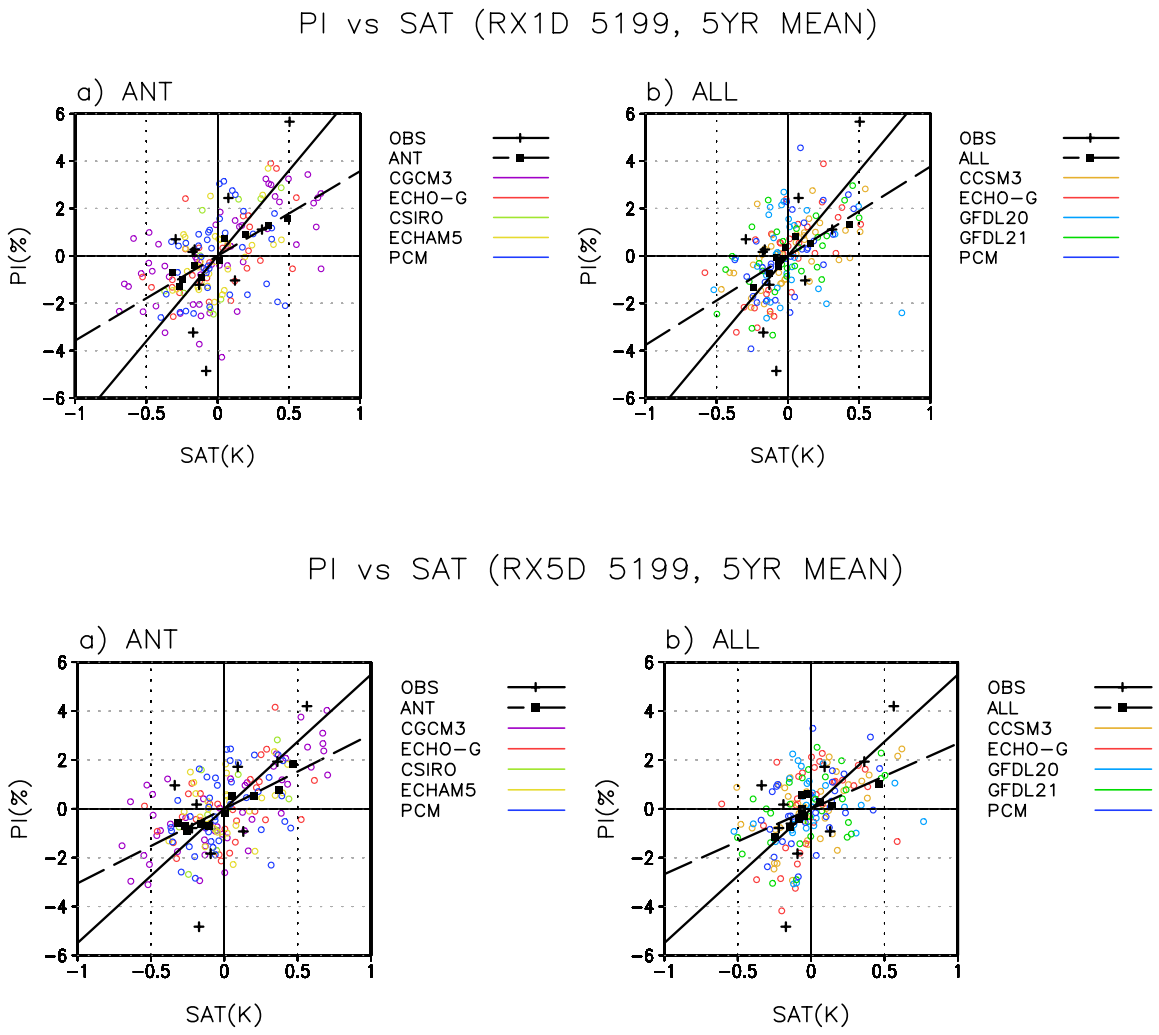

**Supplementary Figure 14 | Sensitivity of optimal detection results to a different observational availability.** Same as Fig. 3 but additionally taking account of space-time varying observational mask. See SI text for details.




**Supplementary Figure 15 | Comparison of 1951-1999 trend patterns for PI for RX1D.** (left) Trends in local PI values at stations with at least 40 years data, (right) **a** trends in PI values calculated using gridded RX1D from station datasets, for which we have required at least 2 stations within each grid-box and taken simple averages of station values to generate gridded values, **b** trends in PI values from HadEX, and **c** difference between **a** and **b**. Area means of trends or difference in trends are depicted on the right top of each panel, for which we have applied the same HadEX data coverage to **a**. See SI text for details. Units are % probability per year.




**Supplementary Figure 16 | Influence of observational mask on large-scale area mean PIs.** Five-year mean anomalies of PI for RX1D averaged over the entire NH land area (top), NH mid-latitude land (middle), and Eurasia (bottom) from ANT (left) and ALL (right) multi-model averages. Solid line represents results when considering the observational availability, and dashed line indicates results from the full region without considering the observational data coverage. See SI text for details.




**Supplementary Figure 17 | Relationship between surface air temperature (SAT) and extreme precipitation indices.** Scatter plots of five-year mean NH land averaged anomalies of PIs for (upper) RX1D and (lower) RX5D and SAT during 1951-1999 from ANT and ALL model runs. Observed and simulated PI and SAT fields are interpolated onto the 5° × 5° grid of the CRU SAT observations before analysis. The solid line represents regression coefficient from least square fitting with units of % per K. CCSM3 ANT runs are omitted here due to unavailable SAT datasets. See SI text for details.