Detection of Human Influence on 20th Century Precipitation Trends

Xuebin Zhang[1], Francis W. Zwiers[1], Gabriele C. Hegerl[2], F. Hugo Lambert[3], Nathan P. Gillett[4], Susan Solomon[5], Peter Stott[6], Toru Nozawa[7]
[1]*Climate Research Division, Environment Canada, Toronto, Ontario, M3H 5T4, Canada*
[2]*Nicholas School for the Environment and Earth Sciences, Box 90227, Duke University, Durham, NC 27708, USA*
[3]*Department of Geography, 507 McCone Hall, University of California, Berkeley, CA 94720, USA*
[4]*Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK*
[5]*NOAA Earth System Research Laboratory, 325 Broadway, Boulder, CO 80305, USA*
[6]*Met Office Hadley Centre (Reading Unit), Meteorology Building, University of Reading, Reading RG6 6BB, UK*
[7]*National Institute for Environmental Studies, Tsukuba, Ibaraki 305-8506, Japan*

Details of Data Processing

The Global Historical Climatology Network (GHCN) data set[1] has been collected, maintained, and quality controlled at the National Climatic Data Centre, NOAA. The most recent version (second version) has been used in this analysis. In addition to updates for both spatial and temporal coverage, data homogeneity has been addressed for the US, and measurement errors for data from Canada and the former USSR have also been corrected[2]. Given the amount of data and the need to update regularly, erroneous data might still exist in the data set. We conducted some additional quality control before analysing the data by screening for possible spurious trends, such as a trend that is too large to be reasonable, in the station data and other non-physical artifacts such as change points that might be indicative of changes in instrumentation, observing practices, station location and so on. This was done, in part, by assessing consistency of changes at adjacent stations. There appears to have been a change in the unit of measurement for Bangladesh stations around 1990 that caused large apparent negative trends in annual precipitation. This error, which has been corrected, coincides with the introduction of recent World Weather Records for that country into the GHCN. Missing values at some tropical stations may have been coded as zero, also causing spurious trends. Data from those stations were corrected to the extent possible by comparing seasonal means with those of surrounding stations, or were excluded from the analysis. The overall number of stations that were judged to have spurious trends is very small.

To ensure that trends computed from observational data are representative for the period, we include in our analysis only those stations that meet the following two criteria: 1) the station has at least 25 years of data during the 1961-1990 base period, and 2) the station has at least 5 years of data in every decade during the period 1950-1999. The latter constraint was not applied to the 1925-1949 period. However, to determine the longest period in the 20th century during which a trend representative for 10-degree longitudinal bands can be reliably estimated, we computed the fraction of land area sampled by the observations during the base period in each 10-degree band (Supplementary Figure 1) for the 20th century. Sixty percent or more of land area sampled during the base period is covered by observations since 1925 for every latitude band used in this analysis. Data coverage deteriorates rapidly for several latitude bands prior to 1925. As a result, the longest time period on which our detection analysis is conducted is 1925-1999 (note that the broad feature of changes remains unaltered if the entire 20th century is analyzed; see Fig 9.18 of Hegerl et al., 2007). The best data coverage starts around 1950. We thus conducted our main detection analysis for the 50-year period 1950-1999.

Precipitation variability estimated from climate models and observations

A challenge in detecting an anthropogenic climate change signal in observed precipitation is the discrepancy between observed and simulated variance. Supplementary Figure 2 illustrates the problem. It shows that the observational estimate of the variance of 5-yr mean precipitation anomalies in 10° latitude bands is greater than that simulated by models, typically by a factor of 2. Power spectra of zonal average annual precipitation anomalies also confirm that modeled zonal mean precipitation typically has smaller variability than the observed counterpart (Supplementary Figure 3). A substantial part of the discrepancy may be the result of the use of data from a few stations and sometimes from only one station to represent grid box mean precipitation. Supplementary Figure 4 shows that the power spectrum of grid box averaged precipitation anomalies is likely lower than that computed from individual stations by illustrating that precipitation averaged from 56 stations shows substantially less variability on all timescales than precipitation averaged from subsets of 1, 3 and, for short timescales, even for 10 stations within the grid box. This means that the use of available station precipitation data overestimates the variability of grid box mean precipitation in a large number of grid boxes. In fact, this problem affects the majority of grid boxes because the median of available stations is 3 per grid box; exceptions occur in a few regions where there are dense networks of observing stations.

Our detection analysis is conducted on a sub-space spanned by the first 5 leading EOFs of zonal precipitation trends, thereby retaining scales where model simulated variability in zonal trends is more often consistent with the observationally based estimate according to a standard test[3]. Cases where internal variability is problematic are labelled in Figure 3 (main paper). Because the observationally based estimate of variability in precipitation trends is likely too high, our detection results should be robust and conservative. Furthermore, most of our results are also robust to inflating the model variability by a factor of 2.

Details of estimation of internal variability

The covariance matrices required for optimization and testing are estimated from independent parts of a 1700 year control simulation performed with HadCM3, for which there is no change in external forcing, and from variability between members in ensembles of forced simulations. The latter consist of a total of 9200 years of climate simulation with ALL, ANT, or NAT forcing in ensembles of 3 to 9 100-yr simulations (Supplementary Table 1 and references therein). The 1700-yr HadCM3 control simulation is divided into 17 non-overlapping 100-yr samples.

For the 75-yr detection analysis, trends are estimated from the last 75-yr of the 100-yr sample of model output that is masked to mimic the availability of gridded observations over the period of record. We then split the control and forced simulations in such a way that every model is represented in the covariance matrix estimates $\hat{\mathbf{C}}_{N1}$ and $\hat{\mathbf{C}}_{N2}$. In the case of ensembles with an odd number of members *k*, *(k+1)/2* members are randomly assigned to either sample N1 or N2, and the remaining *(k-1)/2* members are allocated to the other sample;

the allocation of the larger number of members is alternated between models to ensure that N1 and N2 contain the same number of members in total. In the case of ensembles with an even number of members, half of the members are randomly allocated to each sample. Overall mean trend patterns are removed separately from the control, ALL, ANT, and NAT simulations in each sample N1 and N2, and the residual trends are used to estimate $\mathbf{C}_{N1}$ and $\mathbf{C}_{N2}$. The covariance estimates are appropriately adjusted to reflect the fact that ensemble mean trend patterns have been removed from individual ensembles.

For the 50-yr analyses, we split each of the 100-yr chunks of data into two 50-yr periods, and compute trends in each 50-yr data segment masked to mimic the availability of gridded observations during 1950-99. Those trends are then partitioned similarly as those of 75-yr trends for the estimation of $\hat{\mathbf{C}}_{N1}$ and $\hat{\mathbf{C}}_{N2}$.

The estimates of internal variability necessarily contain a contribution from inter-model and intra-ensemble variability. They therefore reflect structural uncertainty as well as internal variability. In this case the residual trends deviate from the model mean trend due to both internal climate variability and systematic differences between individual models, yielding an inflated estimate of internal climate variability. Using such a covariance estimate to optimize signal-to noise ratios will have the effect of down-weighting features of zonal precipitation change that are strongly affected by model uncertainty[17]. The residual consistency test[3] suggests that model simulated variability is generally not inconsistent with observed variability at the 10% significance level when 5 or fewer EOFs are retained. More than 76% of signal and noise variance is retained at this truncation. The shapes of the precipitation trend patterns are also preserved at this truncation (Supplementary Figure 5)

Uncertainty in the estimate of internal variability

Detection and residual consistency tests use a statistical formalism that is designed to account for sampling variability assuming that trends follow a Gaussian distribution. There are many different ways to partition the available sets of simulations into samples N1 and N2 when estimating the noise covariance. This allows us to take sampling uncertainty in the estimation of internal variability into account more robustly by comparing results from different configurations of model data for noise estimation. To do so, we randomly partition model simulations, as detailed above, using 2000 different configurations, and thereby obtained 2000 samples of $\hat{\mathbf{C}}_{N1}$ and $\hat{\mathbf{C}}_{N2}$. We then repeated our detection analyses 2000 times using these different estimates of internal variability.

Supplementary Table 2 summaries the number of times detection occurs in the one-signal analyses. For the 50-yr trend, ALL is robustly detected. ANT is also detected, but the residual consistency test passes only infrequently in this case (see discussion below). Results from the 4 models that conducted simulations with ALL, ANT, and NAT forcing are similar, but here also, the residual consistency test fails frequently. The greater difficulty with residual variability may indicate that the signals are not as well defined when a smaller number of models is used, or that the residual contains more unexplained parts of the signal and thus shows larger variance than the estimate of internal variability. Results for the 75-yr trends are consistent with those for 50-yr trends. However, the problem with residual variability is more apparent. This is perhaps not unexpected given that poorer spatial coverage of station data towards the beginning of the period should yield poorer estimation of observed trends and greater overestimation of grid box scale variability. Detection results, particularly when using the ALL signal, are robust to doubling the estimated internal variance, in which case the residual consistency test passes much more frequently.

Results for the 2-signal analyses are summarized in Supplementary Table 3. For the 50-yr trends, both ANT and NAT are detected relatively robustly when ALL and ANT signals are used and when model simulated variance is doubled. Results with simulations from the 4 models are consistent, but residual variability appears to be more problematic, which is consistent with the results from the one-signal analyses. ANT is not separated from NAT or internal variability if the NAT4 signal is used in the analyses. This may indicate problems with the NAT simulations from the 4 models. For the 75-yr trend, ANT may be separated, but less robustly, from NAT and natural variability in the ALL and ANT analyses as explained in the methods section.

Effect of area weighting on detection

Both the observed vector **y** and signals **X** in the regression can be weighted to account for surface area differences between latitude bands. However, in that case the EOF analysis focuses largely on the tropics (which represent large areas and large rainfall variability), making it difficult to represent the important features of high latitude rainfall increases in a truncated space. We therefore did not use area weighting in our analysis. However, results largely hold if area weighting is used, although in this case, detection can only be conducted in a sub-space spanned by the first 4 EOFs according to the residual consistency test[3].

Effect of coverage changes on zonal precipitation trends

We used the ALL simulations to assess the effect of changes in data coverage over time on zonal precipitation anomaly trend estimation. To do so, we proceeded as follows. We first produced two masks from the observations: a) a fixed mask in time that includes all grid points where there is at least one annual observation, and b) the usual mask (Mask A) that indicates the timing and location of available observations. Having produced the two masks, two quantities were calculated for each 10 degree band from each of the 50 ALL runs:

a) Zonally averaged precipitation anomalies calculated over the fixed mask consisting of all points that were sampled at least once over the period of record of the observations (this data set is called "complete")
b) Zonally averaged precipitation anomalies calculated over points where masking mimics the availability of real observations over time and space (this data set is called Mask A)

Trends corresponding to 1925-1999 and 1950-1999 are computed for the "complete" and the Mask A precipitation anomaly series. The difference in "complete" trends between different runs represents uncertainty in the trend estimate due to natural variability. The difference between the "complete" and the Mask A trends for the same run indicates uncertainty in the trend estimate due to changes in the spatial coverage in the observational data. This difference across the 50 ALL simulations is shown in the box plots in Supplementary Figure 6a, as a fraction of the standard deviation of the "complete" trends from the 50 ALL simulations. The change in the spatial coverage does not appear to have resulted in a systematic bias in the trend estimate. Uncertainty over the 1950-

1999 period is much smaller than that over the longer 1925-1999 period. Over the Northern Hemisphere, the average absolute difference between the "complete" and the Mask A trends is generally smaller than 0.1 and 0.2 standard deviations for the 1950-1999 and 1925-1999 periods, respectively. However, the differences are larger over the tropical regions. Supplementary Figure 6b displays the number of times the "complete" trend and the Mask A trend have the same sign in 50 ALL simulations; the "complete" trend and the Mask A trend usually have the same sign. In the worst case of the 1925-1999 precipitation trend, the "complete" trend and the Mask A trend match their signs 86% of time. Because the sign of the trend is unlikely to be altered due to the change in the spatial coverage, we conclude that it is unlikely that the moistening and drying pattern seen in the observed zonal mean precipitation anomaly trends is affected by change over time in the spatial coverage of observational data in the GHCN data set.

Overall, spatial coverage change may have resulted in uncertainty in the trend estimate, especially for 1925-1999. However, this uncertainty is far smaller than that due to natural variability (or errors caused by structural differences between models), and is unlikely to have altered the spatial patterns of drying and moistening.

Joint estimation of scaling factors for ANT and NAT

Scaling factors estimated jointly in two-signal detection analysis are generally correlated, and thus their joint uncertainty due to sampling variability is best expressed by means of joint confidence regions. Supplementary Figure 7 shows 90% joint confidence regions for the scaling factors on the model estimated responses to NAT and ANT forcing, along with the corresponding one-dimensional marginal error ranges, when ALL and ANT trends from all available simulations are used in the two-way regression analysis, and when ALL4 and ANT4 trends are used in the analysis.

Precipitation trends simulated by individual models

The scaling factors shown on Figure 3 (main paper) are larger than unity, indicating that multi-model response to external forcings may be underestimated. As shown in Figure 1 (main paper) which displays ensemble mean trends from individual models, some models such as the GFDL 2.0 and MIROC do have response features, particularly in the tropics, that are as large observed. However, these large trends are situated somewhat differently latitudinally, impeding detection and damping the amplitude of the multi-model pattern of response. However, both of these models exhibit a quite weak response in high latitudes. Other models contribute to the multi-model mean increase that is seen in high latitudes, such as PCM or GISS, both of which have little structure in the tropics. The pattern correlation (including the global mean trend in the calculation) between multi-model mean ALL or ANT trends and observed trends is higher than the pattern correlation between the trends simulated by any individual model and the observed trends. Therefore, it appears that in this case, multi-model averaging improves the agreement between models and observations, although it occurs at the cost of dampening the response of the more strongly responding models.

References:

1. Vose, R.S., Schmoyer, R.L., Steurer, P.M., Peterson, T.C., Heim, R., Karl, T.R., and Eischeid, J. The Global Historical Climatology Network: long-term monthly temperature, precipitation, sea level pressure, and station pressure data. ORNL/CDIAC-53, NDP-041. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee (1992).
2. Groisman, P.Ya. & Rankova, E. Ya. Precipitation trends over the Russian permafrost-free zone: removing the artifacts of pre-processing. *Internat. J. Climatol.* **21**, 657-678 (2001).
3. Allen, M.R. & Tett, S.F.B. Checking for model consistency in optimal fingerprinting. *Clim. Dyn.* **15**, 419-434 (1999)
4. Min, S.-K., Legutke, S., Hense, A. & Kwon, W.-T. 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).
5. Delworth, T.L. & co-authors. GFDL's CM2 global coupled climate models -- Part 1: Formulation and simulation characteristics. *J. Clim.*, **19**, 643-674 (2006).
6. Schmidt, G.A., Ruedy, R., Hansen, J.E., Aleinov, I., Bell, N., Bauer, M., Bauer, S., Cairns, B., Canuto, V., Cheng, Y., Del Genio, A., Faluvegi, G., Friend, A. D., Hall, T.M., Hu, Y., Kelley, M., Kiang, N. Y., Koch, D., Lacis, A.A., Lerner, J., Lo, K.K., Miller, R.L., Nazarenko, L., Oinas, V., Perlwitz, Ja., Perlwitz, Ju., Rind, D., Romanou, A., Russell, G.L., Sato, Mki., Shindell, D.T., Stone, P.H., Sun, S., Tausnev, N., Thresher, D., and Yao M.-S. Present day

atmospheric simulations using GISS ModelE: Comparison to in-situ, satellite and reanalysis data. *J. Clim.,* **19**, 153-192 (2006).

7. K-1 model developers. K-1 coupled model (MIROC) description, K-1 technical report, 1, H. Hasumi and S. Emori (eds.), Center for Climate System Research, University of Tokyo, 34pp (2004).
8. Seiji Yukimoto, Akira Noda, Akio Kitoh, Masato Sugi, Yoshiteru Kitamura, Masahiro Hosaka, Kiyotaka Shibata, Shuhei Maeda and Takao Uchiyama, 2001: The New Meteorological Research Institute Coupled GCM (MRI-CGCM2) --- Model Climate and Variability, *Pap. Meteor. and Geophys.*, **51**, 47-88 (2001).
9. Collins, W.D., Bitz, C.M., Blackmon, M.L., Bonan, G.B., Bretherton, C. S., Carton, J. A., Chang, P., Doney, S. C., Hack, J. J., Henderson, T. B., Kiehl, J. T., Large, W. G., McKenna, D. S., Santer, B. D., & Smith, R. D. The Community Climate System Model: CCSM3. *J. Clim.*, **19**, 2122-2143 (2006).
10. Meehl, G.A., Washington, W. M., Ammann, C., Arblaster, J. M., Wigley, T. M. L., & Tebaldi, C. Combinations of natural and anthropogenic forcings and 20th century climate. *J. Clim.*, **17**, 3721—3727 (2004).
11. Johns T.C., Gregory, J. M., Ingram, W. J., Johnson, C. E., Jones, A., Lowe, J. A., Mitchell, J. F. B., Roberts, D. L., Sexton, B. M. H., Stevenson, D. S., Tett, S. F. B., & Woodage, M. J. Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios, *Clim. Dyn*, **20**, 583-612 (2003).
12. Flato, G.M. & Boer, G. J. Warming Asymmetry in Climate Change Simulations. *Geophys. Res. Lett.*, **28**, 195-198 (2001).

13. Gordon, H. B., Rotstayn, L. D., McGregor, J. L., Dix, M. R., Kowalczyk, E. A., O'Farrell, S. P., Waterman, L. J., Hirst, A. C., Wilson, S. G., Collier, M. A., Watterson, I. G., and Elliott, T. I. The CSIRO Mk3 Climate System Model [Electronic publication]. Aspendale: CSIRO Atmospheric Research. (CSIRO Atmospheric Research technical paper; no. 60). 130 pp (2002).
14. Cai, W., Collier, M.A., Durack P.D., Gordon H.B., Hirst A.C., O'Farrell S.P. and Whetton P.H. The response of climate variability and mean state to climate change: preliminary results from the CSIRO Mark 3 coupled model. *CLIVAR Exchanges*, **28**, 8-11 (2003).
15. Yu Yongqiang, Zhang Xuehong, Guo Yufu. Global coupled ocean-atmosphere general circulation models in LASG/IAP. *Adv. Atmos. Sci.*, **21**, 444-455 (2004).
16. Jungclaus, J.H., Botzet, M., Haak, H., Keenlyside, N., Luo, J.-J., Latif, M., Marotzke, J., Mikolajewicz, U. & Roeckner, E. Ocean circulation and tropical variability in the AOGCM ECHAM5/MPI-OM. *J. Clim.*, **19**, 3952-3972 (2006).
17. Sun,Y., Solomon, S., Dai, A. & Portmann, R.W. How often does it rain? *J. Clim.,* **19**, 916-934 (2006).

Supplementary Table 1. Model simulations and ensemble size available for this analysis

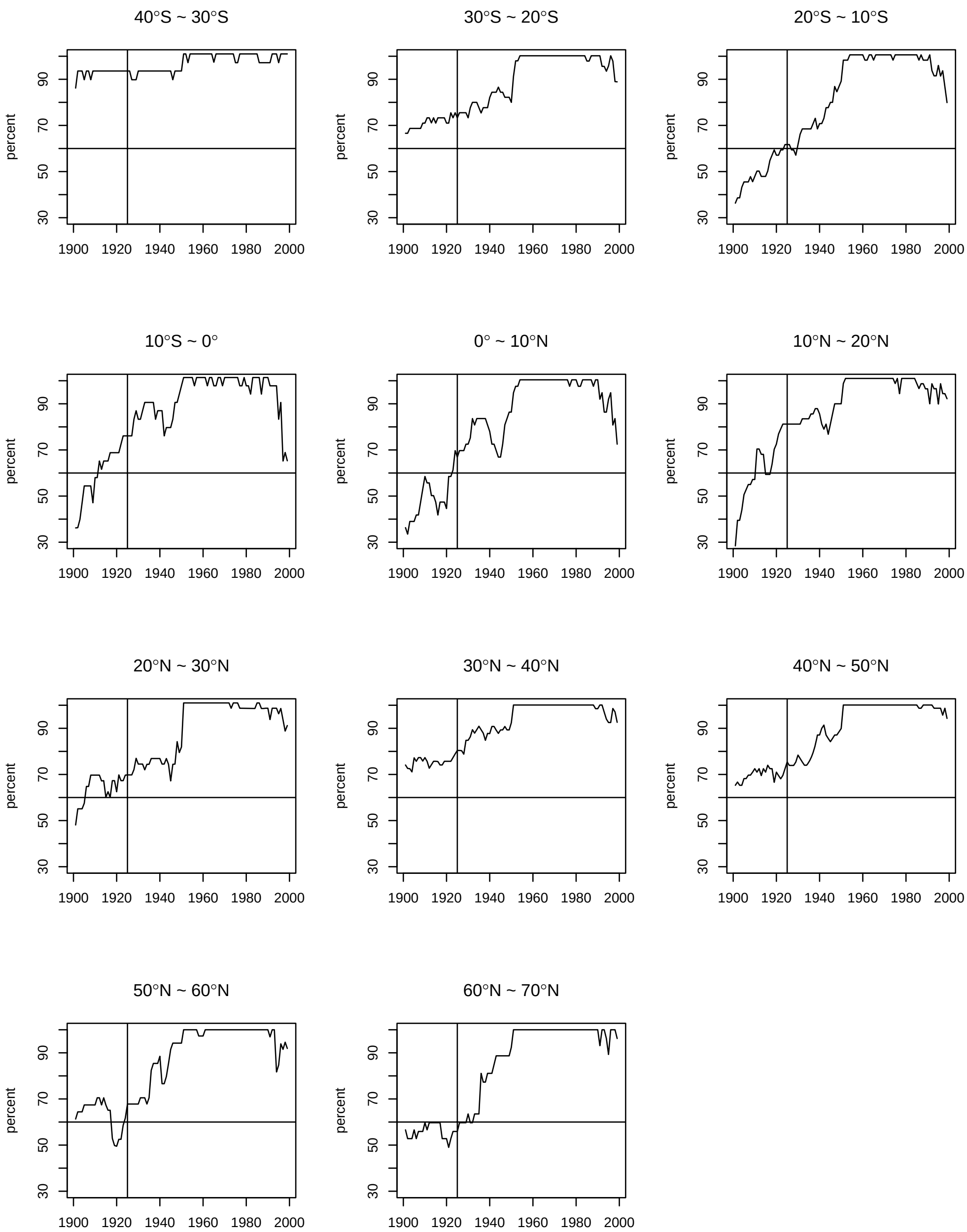
Model	ALL	ANT	NAT
ECHO-G[4]	5	3	3
GFDL CM2 0[5]	3		
GFDL CM2 1[5]	3		
GISS Model E H[6]	5		
GISS Model E R[6]	9		
MIROC 3.2 medres[7]	4	4	4
MRI CGCM2 3 2a[8]	5		
NCAR CCSM 3[9]	8		
NCAR PCM1[10]	4	3	4
UKMO HadCM3[11]	4	4	4
CCCma CGCM3[12]		4	
CSIRO[13,14]		3	
FGOALS[15]		3	
ECHAM5/MPI-OM[16]		3	

Supplementary Table 2. Robustness of attribution results. The table lists the number of times detection occurs in 2000 repetitions of one-signal detection analysis where each of the 2000 analyses uses a different sampling of the available internal variability information. Blue numbers indicate the number of times detection occurred, while black and red numbers correspond to the number of times the residual test was passed or failed in those cases, respectively. Note that a failed residual test can be due to underestimated model internal variability in the space used, or due to the model fingerprint not explaining the observed change well, as is the case if important forcings are omitted. The mean and standard deviation of the scaling factor when detection occurred are also shown.

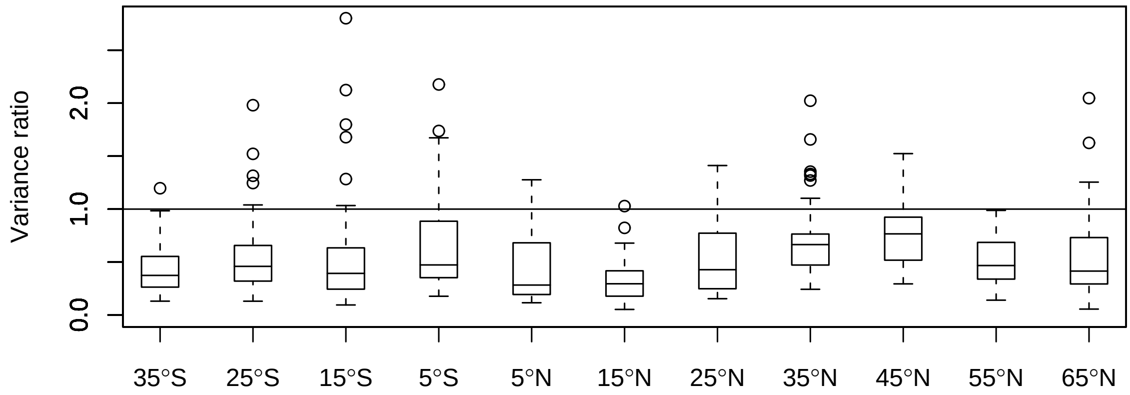
	75-yr trend			50-yr trend		
Simulations	ALL	ANT	NAT	ALL	ANT	NAT
All models	1946 51/1895 $\beta = 5.23 \pm 1.53$	1958 636/1322 $\beta = 5.15 \pm 1.40$	0 0/0	1979 1846/133 $\beta = 6.89 \pm 1.72$	1582 0/1582 $\beta = 3.32 \pm 0.96$	0 0/0
All models, but variance is inflated by a factor of 2	1745 1319/426	1767 1553/214	0 0/0	1916 1915/1	1144 2/1142	2 0/2
4 models	1023 0/1023 $\beta = 6.77 \pm 1.90$	1381 17/1364 $\beta = 4.74 \pm 1.81$	3 0/3	1687 77/1610 $\beta = 7.33 \pm 1.60$	1647 0/1647 $\beta = 4.35 \pm 1.19$	0 0/0
4 models, but variance is inflated by a factor of 2	400 120/280	967 446/521	0 0/0	977 746/231	1367 2/1365	0 0/0

Supplementary Table 3. The number of times detection occurs in 2000 repetitions of two-signal detection analysis where each of the 2000 analyses uses a different sampling of the available internal variability information. Blue numbers indicate the number of times detection occurred, while black and red numbers correspond to the number of times the residual test was passed or failed in those cases, respectively. The table demonstrates that results given in the body of the paper are fairly typical.

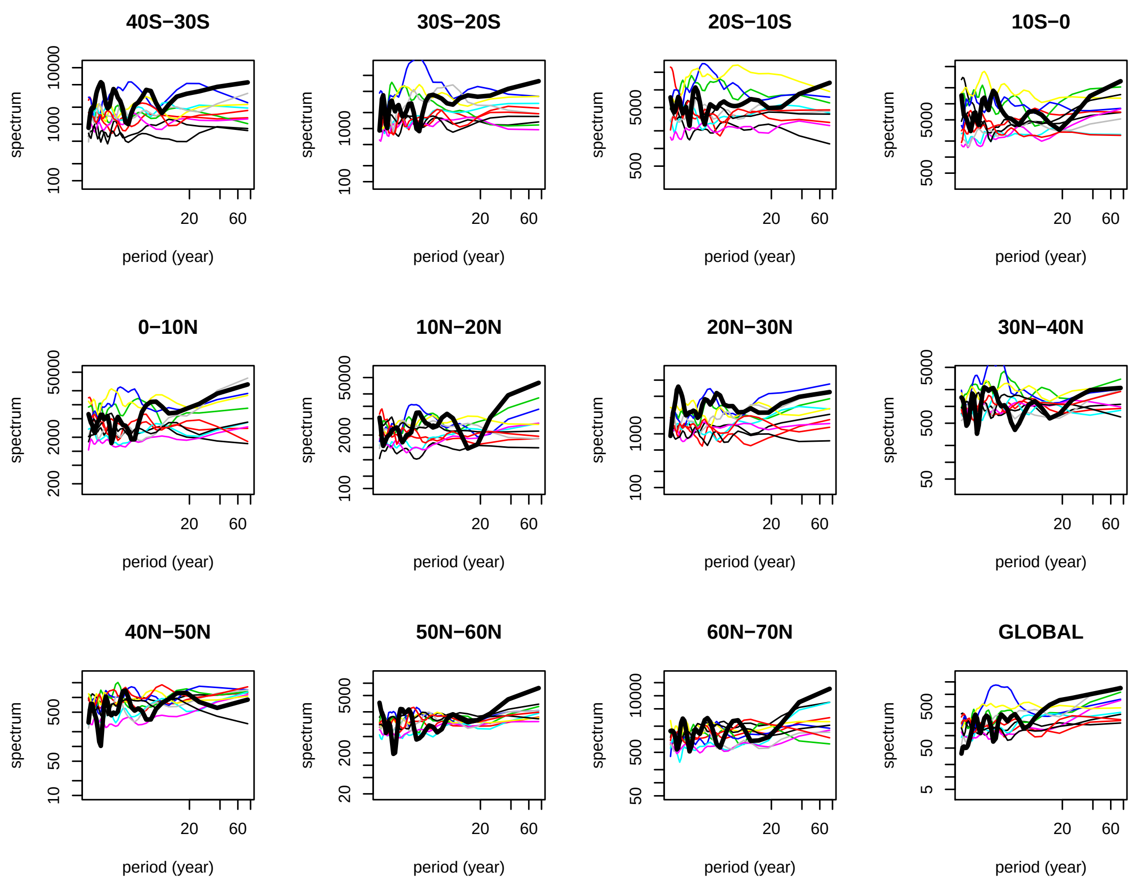
	75-yr trend		50-yr trend	
Signals combinations	ANT	NAT	ANT	NAT
ALL+ANT, all models	1576 667/909	179 106/73	1824 62/1762	1938 141/1797
ALL+ANT, all models, variance doubled	1233 1091/143	34 21/13	1854 1853/2	1956 1951/5
ALL+ANT, 4 models	23 1/22	6 0/6	1682 36/1646	1827 127/1700
ALL+ANT, 4 models, variance doubled	10 3/7	5 3/2	339 118/221	499 264/235
ALL+NAT, all models	939 171/768	3 0/3	7 5/2	6 4/2
ALL+NAT, all models, variance doubled	71 69/2	0 0/0	19 15/4	5 4/1
ALL+NAT, 4 models	300 0/300	3 0/3	48 19/29	36 17/19
ALL+NAT, 4 models, variance doubled	11 3/8	2 0/2	1 1/0	0 0/0
ANT+NAT, all models	490 222/268	76 22/54	0 0/0	1 1/0
ANT+NAT, all models, variance doubled	60 52/8	4 0/4	0 0/0	0 0/0
ANT+NAT, 4 models	200 6/194	0 0/0	0 0/0	0 0/0
ANT+NAT, 4 models, variance doubled	19 9/10	3 0/3	0 0/0	0 0/0



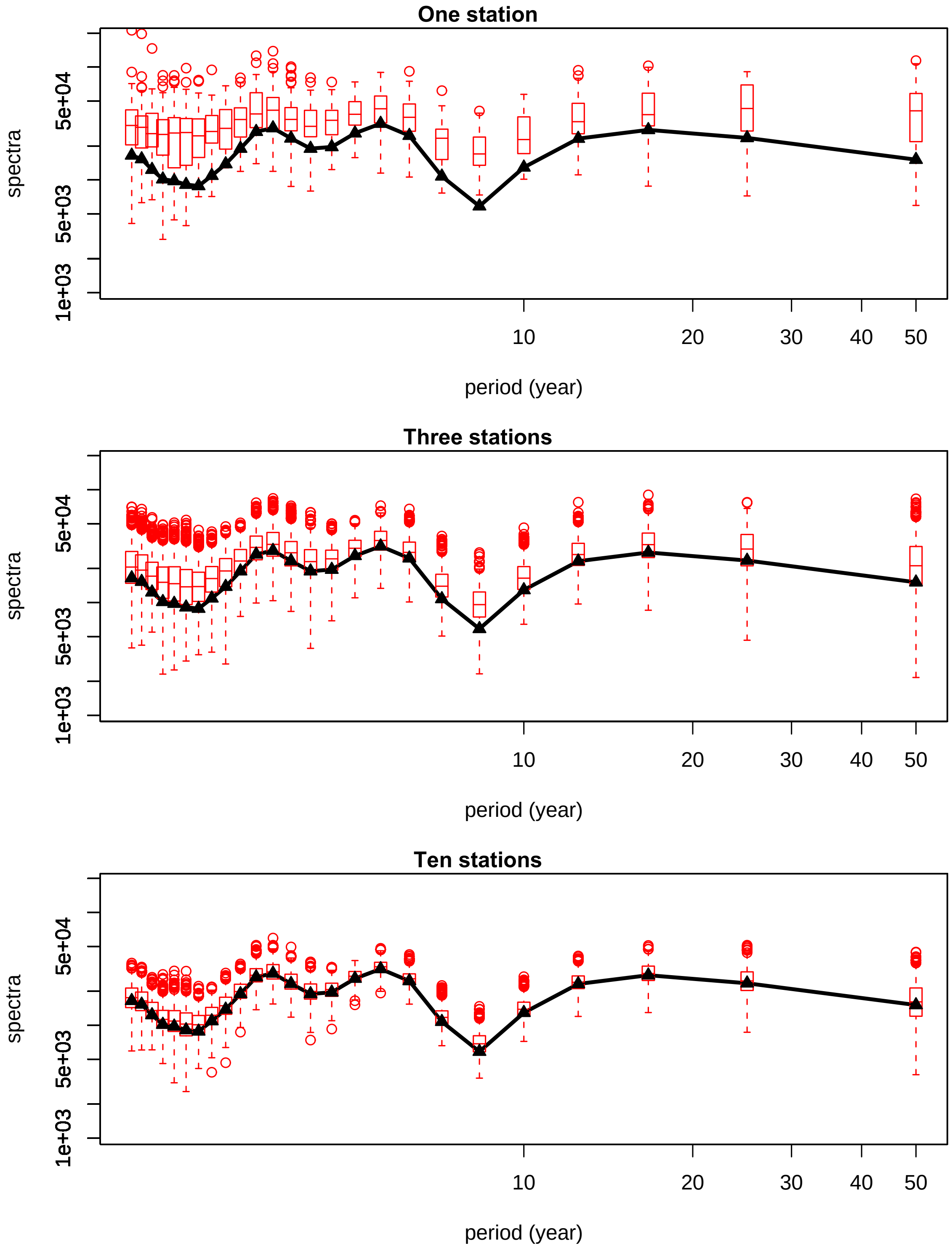
Supplementary Figure 1: Changes in data coverage for 10-degree bands as a fraction of best data coverage during the 1961-1990 base period. Vertical lines mark the year 1925. The fractional coverage is calculated as the ratio of the area covered by 5°×5° latitude-longitude grid boxes for which annual precipitation anomalies could be calculated (see main paper) to the best coverage by 5°×5° grid boxes during the 1961-1990 base period.



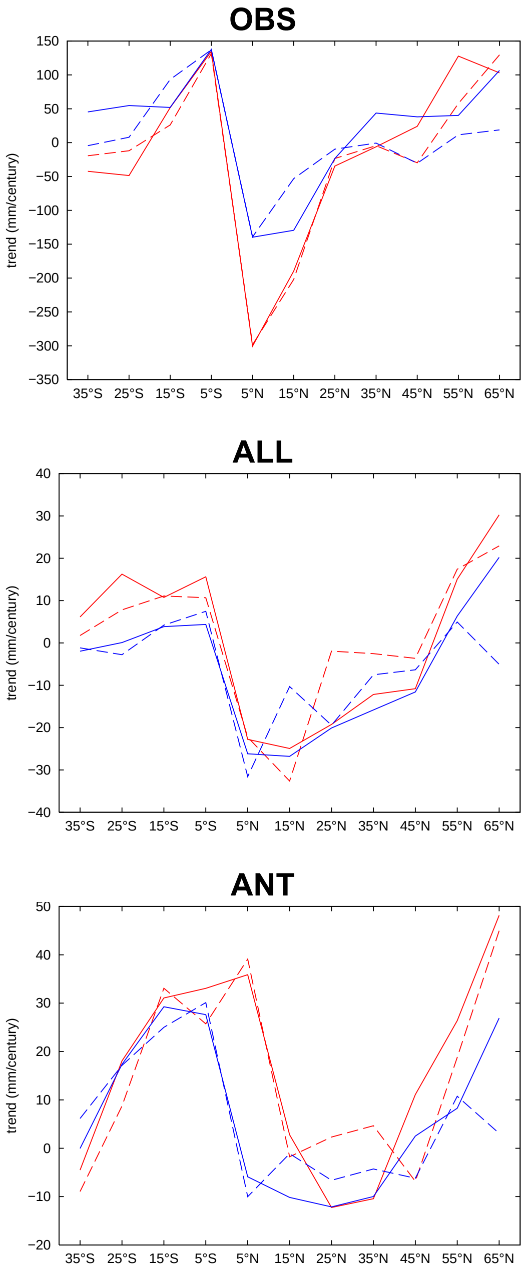
Supplementary Figure 2: Zonal pattern of variance ratio between climate models and observations. The figure shows box and whisker plots of the ratio of 5-yr 10° zonal mean precipitation variances between ALL simulations and that estimated from station observations. The upper and lower ends of each box are drawn at the 75th and 25th quartiles, and the bar through each box is drawn at the median. The two bars indicate the range that would cover approximately 90% of variance ratios if the upper or lower halves of the variance ratio distribution were roughly Gaussian in shape. Individual points beyond the horizontal bars indicate outliers.



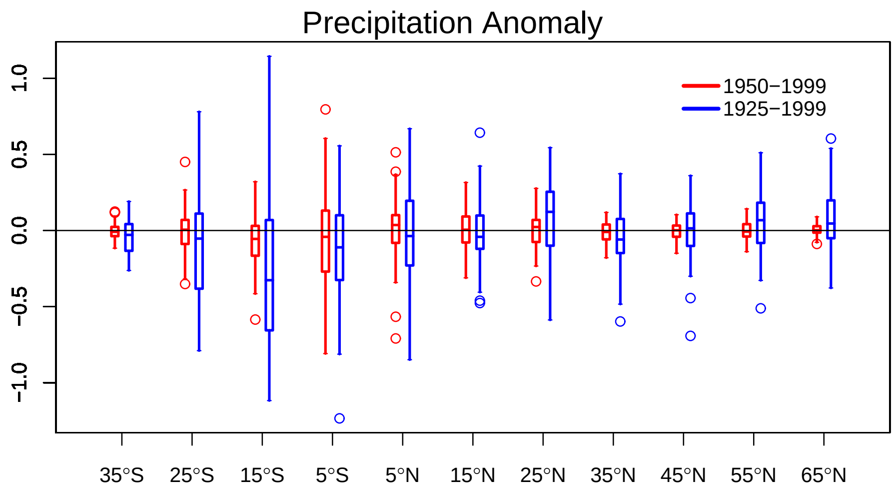
Supplementary Figure 3: Power spectra of zonal mean precipitation anomalies in 10° latitude bands estimated from 1925-1999 data. The thick black curves represent the power spectra of observed precipitation anomalies, while the thin coloured curves represent the ensemble averages of the power spectra for ALL simulations from individual models.



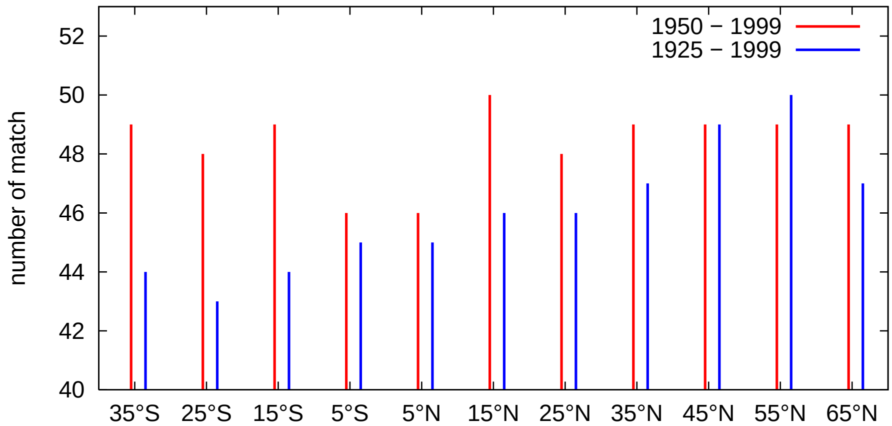
Supplementary Figure 4. Comparison between well-sampled and sub-sampled precipitation variance in a 5°x5° latitude-longitude grid box. Power spectra of averaged annual precipitation anomaly (thick black line and triangles) are shown for the grid box covering 35-40°N, 75-80°W in the central United States, estimated from 1950-1999 data. There are 56 stations that have at least 40 years of data in the grid box. Box and whisker plots (red; see Supplementary Figure 2 for a description) of power spectra computed from individual observing stations (upper panel), mean precipitation anomalies of three randomly selected stations in the grid box (middle panel), and mean precipitation anomalies from ten randomly selected stations from the grid box (lower panel) are also shown. It is clear that the variability of precipitation for the grid box average is much lower than that for individual stations, and that it is also lower than that of an average of three stations. Note that the median number of stations for all available grid boxes is 3. This suggests that observationally based estimates of precipitation variability are too high in at least half of all grid boxes.



Supplementary Figure 5: Observed (OBS) and model simulated (ALL and ANT) precipitation trends during 1950-1999 (red) and 1925-1999 (blue). Solid lines represent the original trend, while trends reconstructed from the first 5 leading EOFs of the estimated covariance are dashed. The EOFs capture more than 76% of variance and key physical aspects of the observed pattern of change.

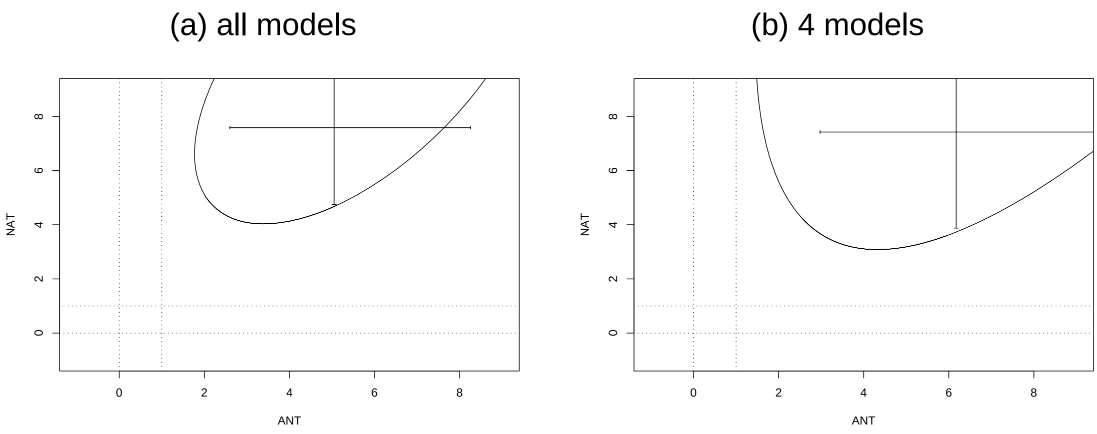


a) Box and whisker plot (1950-1999, red;1925-1999, blue)



b) number of matches in the sign of the trend

Supplementary Figure 6. a) Sensitivity of zonal precipitation trends to sampling, Box and whisker plots are shown for the differences between the "complete" trend and the "Mask A" trend computed from the same simulation expressed as a fraction of the standard deviation of the "complete" trend in ALL simulations. b) The number of times when the "complete" trend and the "Mask A" trend have the same sign in the 50 ALL simulations.



Supplementary Figure 7. Combined uncertainty ranges from 2-signal attribution analysis. 90% uncertainty regions for NAT and ANT scaling factors for the 1950-1999 trends are estimated jointly in a 2-way regression with signals estimated as specified below. The error bars, which cross at the best estimates of the scaling factors, indicate 5%-95% one-dimensional confidence limits. a) 2-way regression with ALL and ANT signals estimated from all available simulations, b) 2-way regression with ALL and ANT estimated from the four GCMs that conducted separate ALL, ANT, and NAT simulations.