

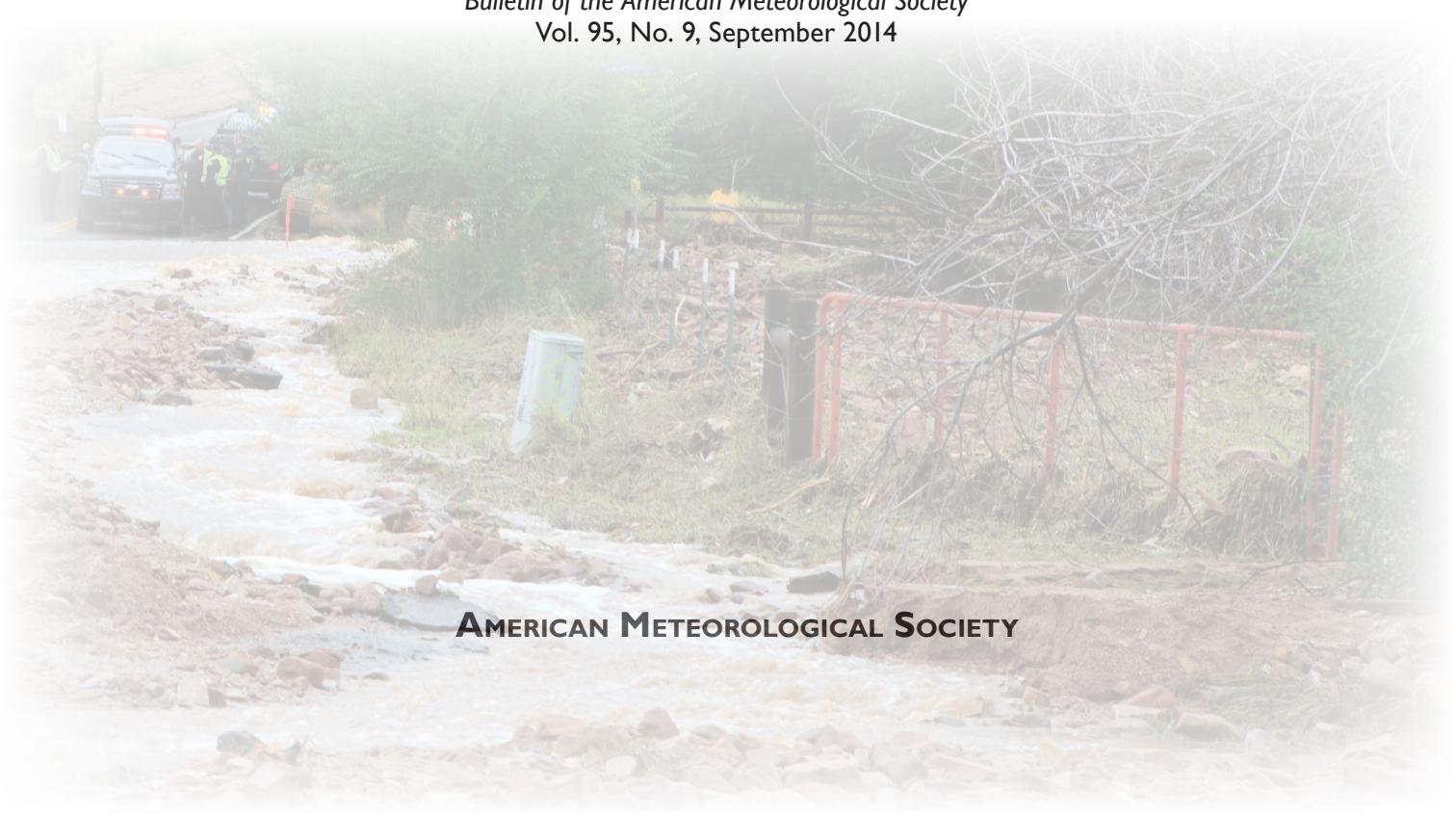
 SUPPLEMENT

EXPLAINING EXTREME EVENTS OF 2013 FROM A CLIMATE PERSPECTIVE

Editors

Stephanie C. Herring, Martin P. Hoerling, Thomas C. Peterson, Peter A. Stott,

Special Supplement to the
Bulletin of the American Meteorological Society
Vol. 95, No. 9, September 2014



AMERICAN METEOROLOGICAL SOCIETY

S2. THE EXTRAORDINARY CALIFORNIA DROUGHT OF 2013/14: CHARACTER, CONTEXT, AND THE ROLE OF CLIMATE CHANGE

DANIEL L. SWAIN, MICHAEL TSIANG, MATZ HAUGEN, DEEPTI SINGH, ALLISON CHARLAND, BALA RAJARATNAM, AND NOAH S. DIFFENBAUGH

Empirical analysis. We calculate the probability of a 2013-magnitude event by analyzing the National Centers for Environmental Prediction reanalysis (NCEP1; Kalnay et al. 1996) and the Coupled Model Intercomparison Project Phase 5 (CMIP5) global climate model ensemble (Taylor et al. 2012). We define a “2013-magnitude event” as the mean January–December 2013 500-mb geopotential height (GPH) averaged over 35°–60°N and 210°–240°E, which emphasizes the area in which 2013 GPHs were unprecedented in the 66-year reanalysis (Fig. 2.1e in the main report; Fig.

S2.1). We justify the use of GPH to characterize the event given both the rarity of the GPH anomalies and the observed strength of the relationship between northeastern Pacific GPH and California precipitation (Fig. 2.1f,i in the main report; Mitchell and Blier 1997; Chen and Cayan 1994).

We restrict our probability analysis to the period from 1979 onwards, in order to focus on the satellite era, for which there is higher confidence in the accuracy of three-dimensional atmospheric fields in the reanalysis. Our CMIP5 model ensemble consists of 13

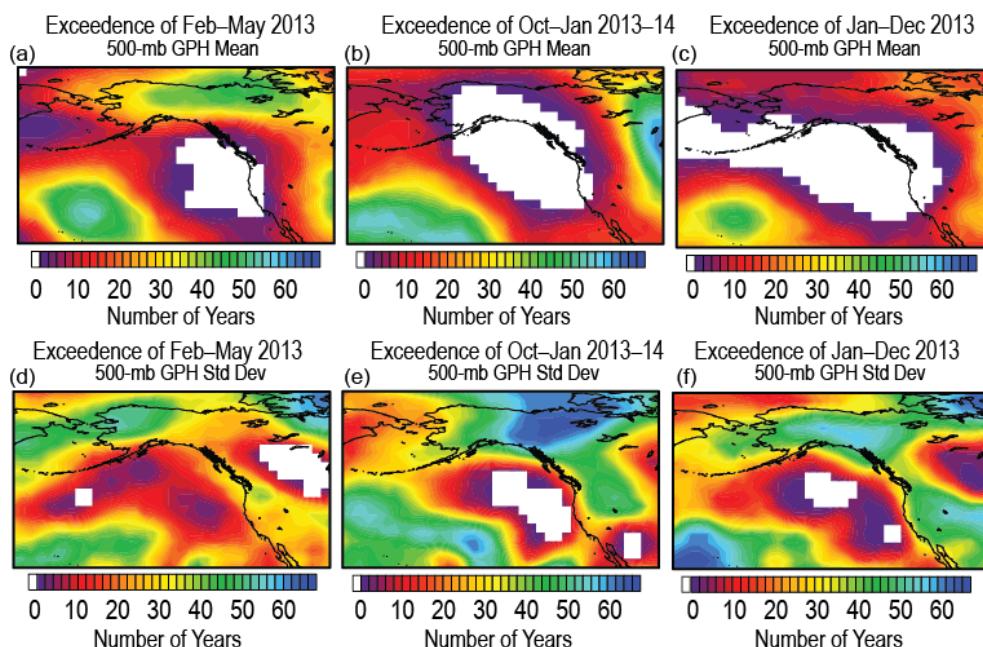


FIG. S2.1. Temporal and spatial structure of observationally-unprecedented 500-mb GPH in 2013–14. Shown are the number of Feb–May (a), Oct–Jan (b), and calendar year Jan–Dec (c) periods during 1948–2012 in which 500mb GPH exceeded the respective 2013 values. (d), (e), and (f) are the same as (a), (b), and (c), respectively, but for standard deviation of daily 500-mb GPH.

models for which GPH data are available for the long preindustrial control (pre-1850) simulations (Table S2.1). We compare the preindustrial (P.I.) simulation of each model with the 1979–2005 period of the 20th century (20C) simulations. (The CMIP5 20C simulations end in 2005.) We aggregate the 1979–2005 period from each model’s 20C realizations into a single 20C dataset for each model.

We first bias-correct the North Pacific 500-mb GPH data from each model by adjusting all simulated values by the difference between the NCEP1 reanalysis mean and the single-model mean:

$$GPH_{bias,corrected} = GPH_{model} + (\overline{GPH}_{reanalysis} - \overline{GPH}_{model}) \quad (1)$$

We then perform a Kolmogorov–Smirnov (K-S) test on each model to assess goodness of fit between the bias-corrected simulated distribution of annual 500-mb GPH and the reanalysis distribution. We find that the GPH distributions in one model (CCSM4; Fig. S2.2) are significantly different from the NCEP1 reanalysis at the $p = 0.2$ level (i.e., agreement between the bias-corrected simulated distributions and the reanalysis distribution is poor). We, therefore, exclude CCSM4 from the remainder of our analysis.

To assess changes in the frequency of occurrence of extreme GPH values between the P.I. and 20C pe-

riods in each model, we first estimate the GPH value associated with each of the 90–99th P.I. percentiles (i.e., the P.I. 0.90–0.99 quantiles). We then count the number of times that a value equaling or exceeding this threshold occurs in the 20C simulations and calculate the relative change:

$$\Delta Freq = (Freq_{20C} - Freq_{PI}) / Freq_{PI} \quad (2)$$

Parametric analysis of CMIP5 data. We calculate the return period of the 2013-magnitude event by first finding a parametric distribution that is appropriate for both the reanalysis and CMIP5 data. In 2013, GPH fell far in the upper tail of the reanalysis record (and is clearly the most extreme event in the reanalysis time series; Fig. 2.2a in the main report). However, given the relatively short period of record for satellite-era observations, there is considerable uncertainty regarding the rarity of the event in a broader climate context. In order to provide a “lower bound” estimate on both the present-era probability of a 2013-magnitude event and the relative change in probability between a preindustrial control climate (absent the effect of anthropogenic forcing) and the historical period, we fit Pareto III-type distributions to the 1979–2012 reanalysis distribution and the 1979–2005 CMIP5 distribution (Fig. 2.2 in the main report). We use a Pareto distribution because of its characteristic heavy upper tail. The use of a heavy-tailed distribution means that our estimates for return periods associated with a 2013-like extreme upper-tail event are smaller than they would be if we had used a distribution with a more quickly attenuating tail, which is justified given the inherent uncertainty in both the true underlying distribution and in the selection of a parametric distribution. However, given the large variability of GPH and small sample size in the post-satellite era (Fig. 2.2a in the main report), we conduct extensive uncertainty quantification for both the reanalysis return period (Fig. 2.2b in the main report) and the CMIP5 20C and P.I. return periods (Fig. 2.2c in the main report; Fig. S2.3).

To find the three parameters in each Pareto III distribution, we graphically obtain the global maximum by examining the 3D empirical probability space given the observations. We then define a cube that surrounds the global minimum and sample from a uniform distribution with the graphically obtained bounds.

For fitting the Pareto to the CMIP5 simulations, we choose a restrictive “best 3” (B3) subset of CMIP5 models by selecting those models with

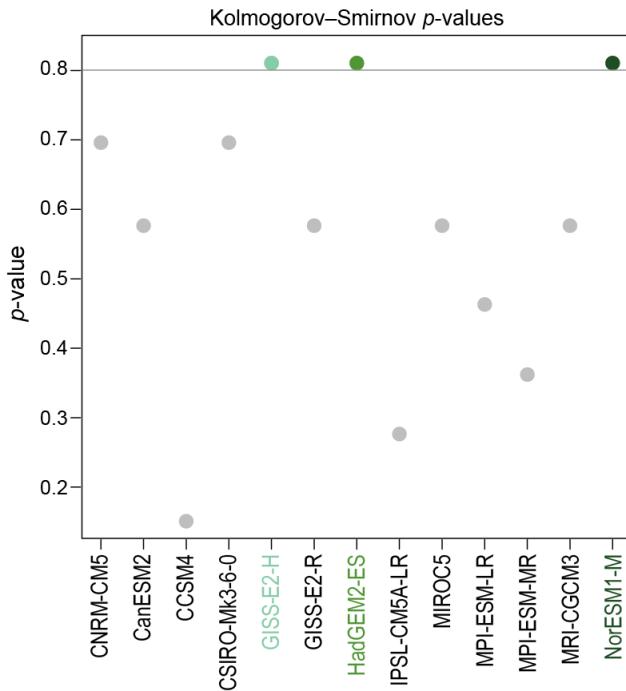


FIG. S2.2. Kolmogorov–Smirnov test results for each of the 13 CMIP5 models for which geopotential heights are archived for the P.I. simulations. Higher p -values suggest higher confidence that the mean-corrected model distributions match the NCEP reanalysis distribution. The B3 models are highlighted in green.

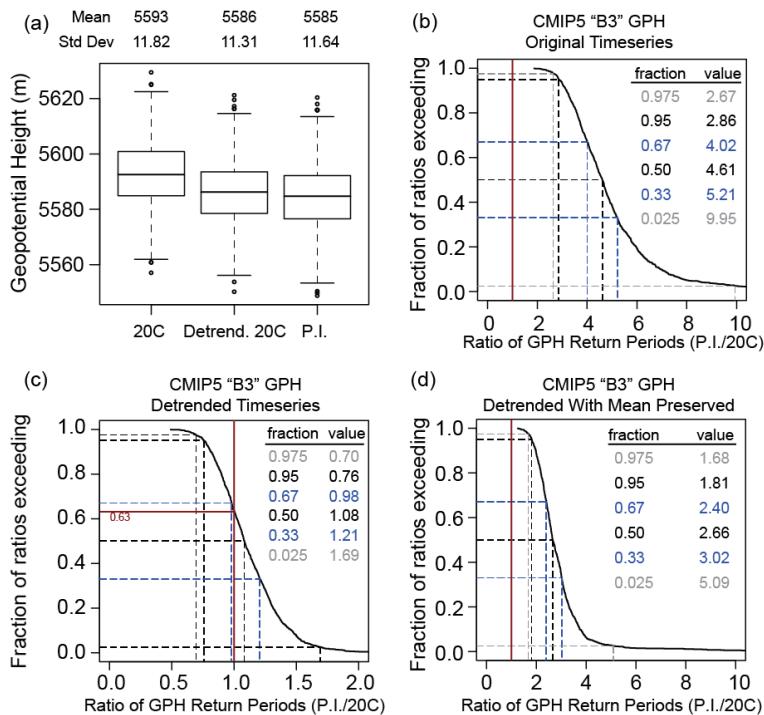


FIG. S2.3. Impact of trend in GPH on distribution of return period ratios. (a) Distribution of Jan–Dec GPH values in the pre-industrial control (P.I.) simulations of the B3 CMIP5 models, the historical 20th century (20C) simulations of the B3 CMIP5 models, and the detrended 20C simulations of the B3 CMIP5 models. (b) Cumulative distribution of bootstrapped return period ratios for the P.I. and 20C simulations in the B3 models, calculated as (P.I./20C). (d) As in Fig. S2.2b, but for detrended (c) and mean-preserved detrended (d) 20C simulations of the B3 CMIP5 models.

K-S test p values ≥ 0.8 (i.e., those models whose mean-corrected North Pacific GPH distributions are closest to the full NCEP reanalysis distribution, including the tails). The three models meeting this criterion are GISS-E2-H, HadGEM2-ES, and NorESM1-M (Fig. S2.2). We then aggregate the individual-model bias-corrected GPH values over all of the 1979–2005 model years from all 20C realizations of each of the B3 models. We then fit Pareto III distributions to this aggregate data in order to estimate the overall change in probability/shift in return period for a 2013-magnitude event.

To calculate the return periods in the CMIP5 B3 models, we first estimate the quantile of the 2013 event using the fitted NCEP distribution. We then find the magnitude of an event of the same quantile in the B3 20C simulations. We then estimate the probability of occurrence of an event of this magnitude in the B3 P.I. and 20C simulations. Finally, we calculate the associated return periods in the B3 P.I. and 20C simulations along with the ratio (P.I. divided by 20C).

Uncertainty quantification for return period estimates. To formally account for uncertainty in the parametric fitting and in the representativeness of the empirical distributions, we perform bootstrap resampling upon the original reanalysis distribution ($B = 1000$) prior to fitting the Pareto III parametric distribution. The range of distributions and parametric fits resulting from the bootstrap allows us to quantify the uncertainty in the reanalysis return period for a 2013-magnitude event, including the “likely” (>66% confidence; Mastrandrea et al. 2011) and “very likely” (>95% confidence) range of values (Fig. S2.3b).

A bootstrap resampling methodology (with replacement) is also employed in fitting the P.I. control and historical 20C Pareto III distributions. In this analysis, all three distributions (reanalysis, P.I., and 20C) are simultaneously resampled and used to calculate return periods (and ratios of return periods). We note that the same parameter bounds are used in each iteration of the bootstrap. We report the resulting distribution of return period ratios (i.e., 1000 different estimates of the relative change in probability of a 2013-magnitude event between the P.I. and 20C forcing regimes) and again assign a “likely” and “very likely” range on the basis of the fitted distributions created using the resampled empirical distributions.

Assessing the role of natural versus anthropogenic forcing in the late 20th century period. We perform additional analysis using the CMIP5 20th century natural forcing (NAT) experiments and compare these results to those from the 20C historical experiments and from the P.I. control experiments (Taylor et al. 2012). The 20C simulations include both natural forcings (such as variability in solar irradiance and the inclusion of volcanic aerosols) and anthropogenic forcings (such as increasing concentrations of greenhouse gases), while the NAT experiments are identical to the 20C experiments except for the exclusion of anthropogenic forcings. The P.I. simulations assume constant solar irradiance, and include neither anthropogenic nor volcanic forcings.

We repeat our analysis of the empirical distribution of GPH in each model (described above) in order to determine the relative change in extreme (90–99th percentile) events between the different experiments. Although we find large increases in the frequency of occurrence of extreme GPH events in the 20C simulations relative to the P.I. simulations (Fig. 2.2e in the main report), we find little or no change in the probability of occurrence between the NAT and P.I. experiments (Fig. 2.2f in the main report). That the large simulated increase in the probability of extreme GPH events only occurs in climate model simulations that include natural and anthropogenic forcings and not in simulations that include only natural forcings suggests that the heightened probability cannot be explained without the anthropogenic contribution.

Addressing the role of non-stationarity. Our return period analysis—which requires fitting parametric models to reanalysis and climate model data—is predicated on a statistical assumption of stationarity in each time series. Because we find a statistically significant linear trend in the reanalysis time series, we take two separate approaches in detrending these data and performing additional analysis to quantify the role of the long-term trend in driving the simulated increase in probability of extreme GPH events in the CMIP5 models.

In the first approach, we detrend the reanalysis time series (excluding the extreme 2013 value) by fitting a linear model to the data and subtracting the accumulated trend at each point from the reanalysis

value. We then subtract the mean of the linearly detrended values and add the mean of the original time series:

$$GPH_{DTP} = (GPH_{DT} - \overline{GPH_{DT}}) + \overline{GPH_{original}} \quad (3)$$

The “detrended” 2013 value is defined as the difference between the actual 2013 value and its expected value in the linear model. We then perform an analogous operation on each 27-year 20C (1979–2005) realization from each of the B3 models to obtain mean-preserved, detrended time series. This approach yields a stationary series with the same mean as the original data, which we verify by analyzing the autocorrelation and partial autocorrelation structure of the new time series (using the Ljung–Box statistic). Our initial detrending approach controls for the existence of a linear trend in GPH over the reanalysis/20C period, but it does not remove the difference in the means between the 20C and P.I. periods. In this analysis, we find that the median P.I./20C ratio of extreme GPH probability is 2.66 (and is greater than 2.4 with 66% confidence and 1.81 with 95% confidence, respectively). While slightly lower than the non-detrended P.I./20C ratio, these results suggest that the simulated increase in the probability of extreme GPH is a statistically robust finding, even accounting for non-stationarity during the reanalysis/20C period.

In the second approach, we fit a linear model to the reanalysis time series (and each model realization) as before, but in this instance, we detrend the time

TABLE S2.1. CMIP5 models used in Swain et al. 2014

Model	piControl (r1i1p1) length (years)	Historical Number of Realizations	Natural Number of Realizations
CanESM2	996	5	5
CCSM4	501	8	—
CNRM-CM5	850	10	6
CSIRO-Mk3-6-0	500	10	5
GISS-E2-H	590	17	10
GISS-E2-R	850	23	—
HadGEM2-ES	280	5	4
IPSL-CM5A-LR	1000	6	3
MIROC5	670	5	—
MPI-ESM-LR	620	3	—
MPI-ESM-MR	990	3	—
MRI-CGCM3	500	5	1
NorESM1-M	501	3	—

series without preserving the original mean in each time series. This approach substantially reduces the difference in the mean between the CMIP5 20C and P.I. values (Fig. S2.3a). In this instance, we find that the P.I./20C ratio is much smaller than using either the original data or the mean-preserved detrended data (median value = 1.08), with error bars that make the trend statistically indistinguishable from zero (with 66% confidence that the ratio exceeds 0.98 and 95% confidence that the ratio exceeds 0.75; Fig. S2.3c). This result suggests that the linear trend in GPH over the 20C period has a substantial influence on the in-

crease in probability of a 2013-magnitude event that is simulated by the CMIP5 models. Given the modest simulated increase in GPH standard deviation (Fig. 2.2c in the main report) and the fact that the median of the second detrended distribution of P.I./20C ratios is positive [and the overall distribution is positively skewed (Fig. S2.3c)], there may be an additional contribution to the probability by nonlinear interactions (i.e., a trend in the variability). Further exploration of this possibility is beyond the scope of the present analysis, but will be the focus of future work.

S4. EXAMINING THE CONTRIBUTION OF THE OBSERVED GLOBAL WARMING TREND TO THE CALIFORNIA DROUGHTS OF 2012/13 AND 2013/14

CHRIS FUNK, ANDREW HOELL, AND DÁITHÍ STONE

Supplemental Methods. ENSO and ENSO-residual Coupled Model Intercomparison Project phase 5 decompositions. Using regression, we estimated the ENSO and ENSO-residual components of each set of 1900–2014 Coupled Model Intercomparison Project phase 5 (CMIP5) SST and precipitation. Table S4.1 summarizes these simulations. The ability of these models to evaluate Pacific–North American climate variability has recently been evaluated by Polade et al. (2013). In this study, the skill of the models, based on single value decompositions of SST and precipitation, ranged from 0 to $\sqrt{2}$. Table S4.1 shows the skills of the models used in this study¹. To represent ENSO, we used the first principal component of tropical Pacific SST (125°E–115°W, 15°S–15°N) over the 1900–2014 time period. For each simulation, a new principal component was calculated and regressed against the global SST and precipitation fields. Both the results and the residuals from these regressions were retained, decomposing the CMIP5 SST and precipitation into ENSO and ENSO-residual components.

Empirical estimates of “trend mode” SST warming patterns. Several recent studies have examined trends and EOFs of global SST with the influence of ENSO removed (Compo and Sardeshmukh 2010; Solomon and Newman 2012). Removing ENSO-related variations reduces the influence of interannual fluctuations and increases the level of agreement among trend fields calculated from different SST datasets. Here, we have represented ENSO as the first principal component of NOAA Extended Reconstructed (Smith et al. 2008) tropical Pacific SST (125°E–115°W, 15°S–15°N) over

the 1900–2014 time period. Using regression, we isolated and removed 1900–2014 ENSO-related SST variations, then defined the trend mode of variability as the first covariance-matrix-based EOF of global (70°S–70°N) seasonal SST anomalies. The 1900–2014 seasonal SST grids were converted to anomalies, weighted by cosine-based weights to account for latitudinal variations in area, and used to compute the covariance matrix. To estimate the 2012/13 and 2013/14 trend SST fields, we multiplied the associated 2012/13 and 2013/14 first principal component scores and EOF trend eigenvector.

Community Atmospheric Model version 5 simulations. The impact of trend-related SST changes on precipitation were explored using Community Atmospheric Model version 5 (CAM5) runs. CAM5 is the atmospheric component of the Community Earth System Model version 1 (CESM1; Hurrell et al. 2013). The CAM5 model was run using a finite volume scheme on a 0.9×1.25 degree grid. Shallow convection was simulated using the scheme of Park and Bretherton (2009); deep convection was simulated using a modified parameterization scheme of Zhang and McFarlane (1995). The CAM5 simulations were produced using a data ocean model where ocean SST and sea ice were specified in space and time. Two sets of 21 runs each were based on either the 2012/13 or 2013/14 observed ocean and ice conditions (NOAA Extended Reconstruction SST; Smith et al. 2008). Another two sets of 21 runs were based on the 2012/13 and 2013/14 observed SST with the 2012/13 and 2013/14 SST trend estimates removed (subtracted). For each simulation, CAM5 was initialized on 1 January 2012 or 2013 and run through the end of February of the

¹ These values were provided by Suraj Polade.

following year. Ensembles were initialized with random modifications to the fourth significant digit of the initial conditions, resulting in different weather patterns for each simulation.

TABLE S4.1. CMIP5 Historical climate change simulations.

Modeling Group & Model Name	Model Acronym	Historical Simulations	Skill (from Polade et al. 2013)
Canadian Centre for Climate Modelling and Analysis <i>Canadian Earth System Model version 2</i>	CanESM2	5	0.88
National Center for Atmospheric Research <i>Community Climate System Model version 4</i>	CCSM4	4	1.0
Community Earth System Model Contributors <i>Community Earth System Model version with Community Atmospheric Model version 5</i>	CESM1-CAM5	2	n.a.
Centre National de Recherches Météorologiques / Centre Européen de Recherche et Formation Avancée en Calcul Scientifique <i>Centre National de Recherches Meteorologiques (CNRM) Coupled Global Climate Model, version 5</i>	CNRM-CM5	4	0.89
NOAA Geophysical Fluid Dynamics Laboratory <i>Geophysical Fluid Dynamics Laboratory Earth System Model version 2 with with modular ocean model version 4.1</i>	GFDL-ESM2M	1	1.0
Institut Pierre-Simon Laplace <i>IPSL Community Model version 5</i>	IPSL-CM5A-LR	3	0.57
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology <i>Model for Interdisciplinary Research on Climate, version 5</i>	MIROC5	3	0.92
Max-Planck-Institut für Meteorologie (Max Planck Institute for Meteorology) <i>MPI Earth System Model Low Resolution</i>	MPI-ESM-LR	3	0.77
Total Ensemble		25	

S6. SEASONAL AND ANNUAL MEAN PRECIPITATION EXTREMES OCCURRING DURING 2013: A U.S. FOCUSED ANALYSIS

THOMAS R. KNUTSON, FANRONG ZENG, AND ANDREW T. WITTENBERG

In this supplemental material, we provide additional background, discussion, and analysis, including: region definitions, global precipitation anomaly maps, and locations with extremes in 2013; areal coverage of record or near-record anomalies by year; spatial resolution issues for observed data; additional sliding trend analysis and sensitivity tests; a description of methodology for estimating the 2013 multimodel ensemble All-Forcing anomaly and the fraction of attributable risk (FAR); and an assessment of model-simulated precipitation and internal variability. Table S6.1 provides a list of the Coupled Model Intercomparison Project phase 5 (CMIP5) models used in the study.

Region definitions, global anomaly maps, and locations with seasonal/annual extremes in 2013. Figure S6.1 shows global maps of (a–e, left column) annual and seasonal mean precipitation anomalies for 2013 and (f–j, middle column) the grid locations with record or near-record wet or dry conditions (seasonal or annual mean precipitation ranked first, second, or third highest or lowest in the available record of length at least 100 years). The blue regions in the right column outline the focus regions selected for the study based on their 2013 anomalies. The six focus regions, including the designated name and season or annual mean, are the northern tier region of states along the northern U.S./Canadian border region with extreme positive annual-mean anomalies (“northern tier—ANN”); in March–May (MAM) a similar region of the northern United States extending slightly further south, also with extreme positive anomalies (“upper Midwest—MAM”); during MAM, a region of record or near-record low precipitation occurring over the

southern U.S. Plains (“Southern Plains—MAM”); in Northern Hemisphere (NH) summer (June–August, JJA), extreme positive anomalies occurring over regions of the eastern United States (“eastern U.S.—JJA”); and in NH fall (September–November, SON), extreme positive anomalies occurring in a region of the north-central United States, but slightly to the west of our upper Midwest region (“Northern Plains—SON”). Although the California region, as resolved in the Global Historical Climatology Network (GHCN) gridded data, was not identified in our analysis as having extreme seasonal or annual precipitation in 2013 (i.e., ranked within the lowest three on record), because of notable drought conditions occurring there during 2013, we examined annual precipitation anomalies in this region (“California—ANN”) as well as seasonal anomalies for December 2012–February 2013 (DJF) and MAM 2013.

Percent coverage of extreme annual mean wet and dry anomalies by year (1900–2013). The extremes maps, Fig. S.6.1 (f–j) show that the high and low mean precipitation extremes were well dispersed around the globe during 2013. Figure S6.2 shows that about 2% of the global “available data” area experienced annual mean dry extremes (first, second, or third lowest rainfall) in 2013, while about 5% experienced wet annual mean extremes (first, second, or third highest rainfall). Since extremes are expected to occur in some places around the globe in any given year, an interesting question is whether 2013 is unusual in terms of the percent area with extreme annual mean values. As a preliminary analysis of this issue, we show in Fig. S6.2 the time series of the fraction of area with wet and dry annual mean extremes over

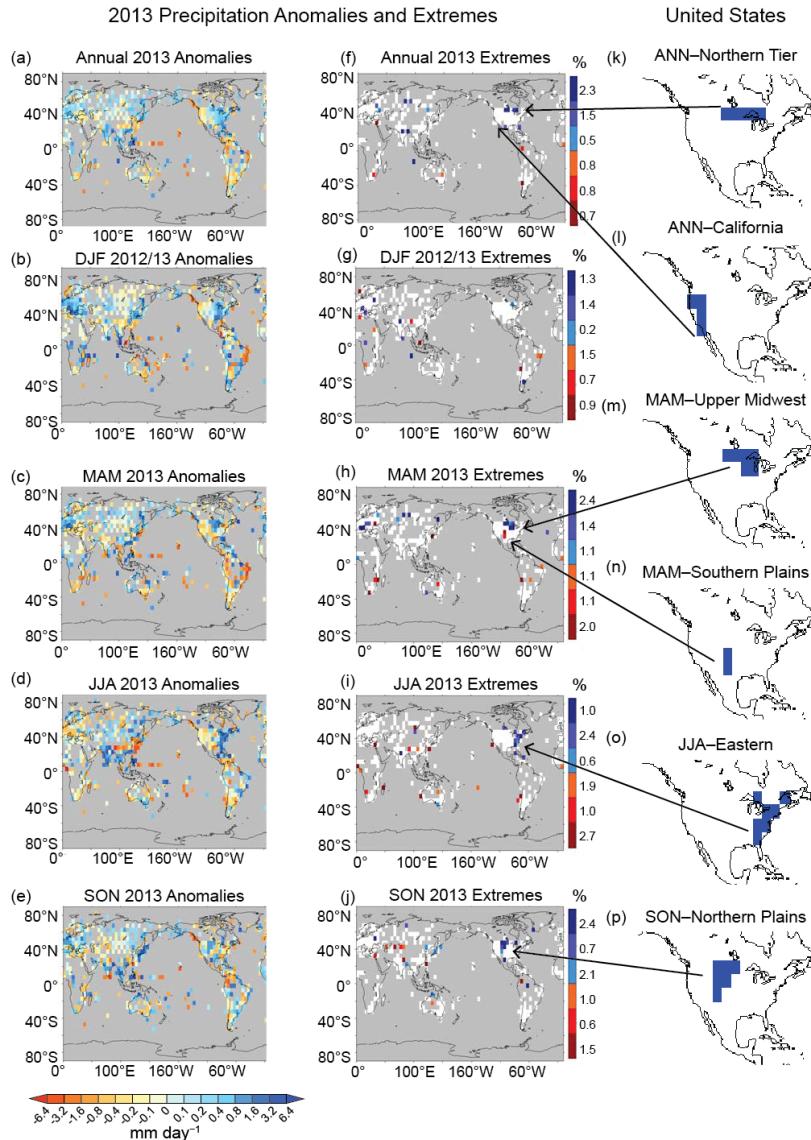


FIG. S6.1. Left column: precipitation anomalies for 2013 (annual or seasonal) in mm day⁻¹. The middle column panels indicate where the anomalies for 2013 are ranked 1st, 2nd, or 3rd wettest or driest in the available record of at least 100 years in length (see Fig. 6.1 legend in the main report). Dark, medium, and light blue depict grid boxes where the 2013 seasonal or annual means rank 1st, 2nd, or 3rd wettest on record. Dark red, red, and orange are 1st, 2nd, and 3rd driest on record. Percent values alongside the color bar in the middle-column panels indicate the percent of global available area with the indicated category of 2013 extreme – where the “available area” has adequate data coverage for at least 100 years (Fig. 6.1 caption in the main report). The blue regions in the right column depict the domains of the six U.S. focus regions selected for our study based on their 2013 extreme anomalies.

the entire record (1900–2013). We include two different measures: (a) the fraction of area with top-three or bottom-three ranked values for each year using the data up to that year and (b) the fraction that is ranked top-three or bottom-three using all years that are eventually available in the series (1900–2013). These metrics show that there has been a tendency

for a larger areal coverage of wet annual mean extremes versus dry extremes in recent decades, and particularly since about 1990. The time series suggest a possible emerging trend in prevalence of wet annual mean precipitation extremes over dry annual extremes. However, the assessment of whether there is a significant trend in these measures is a nontrivial

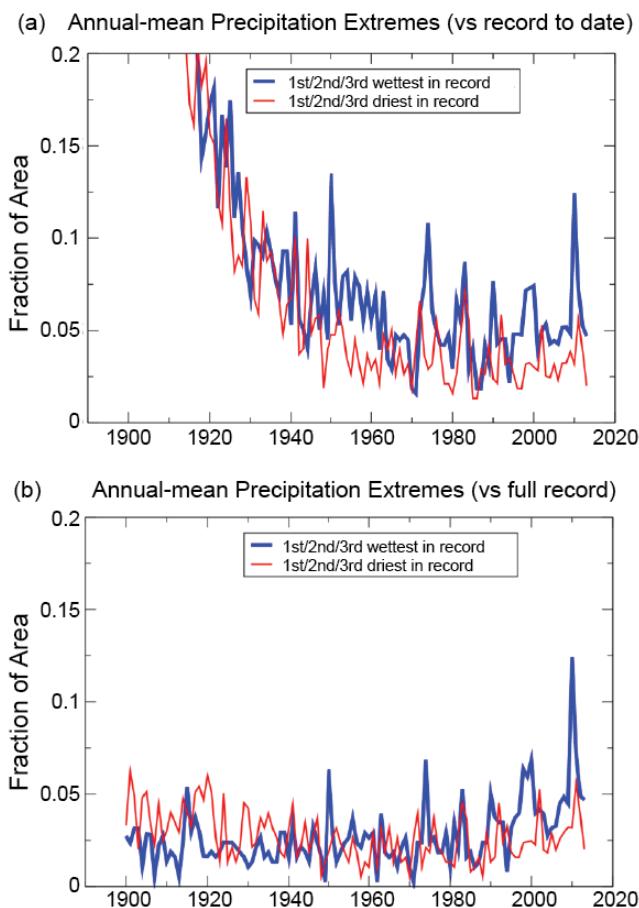


FIG. S6.2. Time series show the fraction of global available area each year with annual mean precipitation anomalies that are ranked 1st through 3rd wettest (blue line) or driest (red line) for each year from 1900 to 2013. The rankings in (a) are based on the available data to date for each particular year (which means that occurrences of new records and near-records are very common in the earlier years). Rankings for (b) are based on the full available record, including years that come after the year in question. This removes the “early year” bias of the method shown in (a).

task. For example, Livezey and Chen (1983) describe issues associated with global significance of areal coverage of locally significant results. Christiansen (2013) addresses the problem of the significance of numbers of record occurrences of warm temperatures and finds that trends in warm daily records for the Northern Hemisphere extratropics since the 1940s are very statistically significant, while trends in monthly warm records are not significant. An assessment of whether the trend suggested in Fig. S6.2 represents a significant change or whether 2013 is a “special” year in any sense in terms of global coverage is beyond the scope of the present study but will be the subject of future investigation.

Resolution dependence (observed data). Owing to the coarse grid on which the data are available, we examine only a very spatially smoothed precipitation anomaly—though the grids are similar to the grid scales of the climate models. Preliminary analysis of an alternative much higher resolution global gridded precipitation dataset (Becker et al. 2013) not only confirms the general occurrence of extreme precipitation anomalies during 2013 in the regions that we focus on, but it also indicates that the record or near-record seasonal anomalies are typically concentrated within relatively small subregions compared to the grid boxes depicted in Fig. S6.1. A separate analysis of about 60 individual U.S. stations in the northern tier region reveals that eight of these stations had unprecedented annual anomalies in 2013 (Cody Hewitt, Rutgers University, 2014, personal communication). Daily timeseries at these eight stations indicate that the unusual annual totals were typically comprised of several large precipitation events, typically occurring in the spring, summer, or autumn.

Sliding trend analysis for all focus regions and sensitivity to excluding 2013 data. Figure S6.3 shows the sliding trend analysis (All-Forcing runs versus Control runs) for each of the six focus regions. Note that the reason we use control runs in our analysis to create the distributions of trends and of variability for the All-Forcing cases (e.g., Fig. 6.2 in the main report and Fig. S6.3) is that the CMIP5 archive does not contain enough individual All-Forcing ensemble members to sample the internal variability of the individual models

adequately for various start dates. Therefore, we have chosen to use samples of variability from the control runs to estimate the range of possible trends around the mean estimates provided by the CMIP5 individual model All-Forcing runs. The All-Forcing 5th to 95th percentile range in Fig. 6.2 in the main report and Fig. S6.3 is based on the aggregate distribution of All Forcing trends and includes a spread due to both differences in ensemble mean response of the various individual models as well as intrinsic (control run) variability, with all control runs sampled equally frequently (see Knutson et al. 2013 for further methods details).

Figure S6.3 (middle column) shows the sliding trend analysis for trends extending to 2013, while the

2013 Precipitation Anomalies and Extremes
for Regional United States

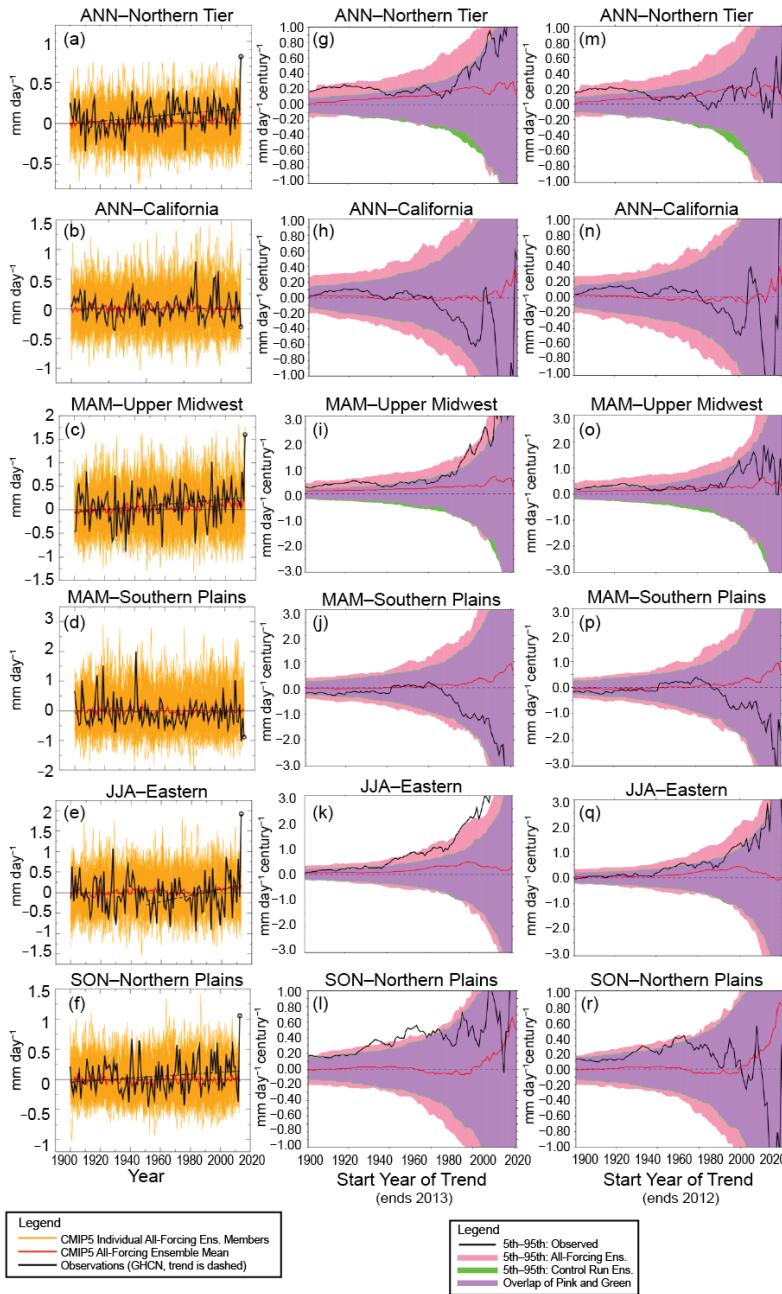


Fig. S6.3. Left (a–f) column shows time series for each of six regions and is analogous to the left column of Fig. 6.2 in the main report. The center (g–l) column is as in Fig. 6.2 in the main report, but shows the sliding trend analysis for each of the six focus regions for trends to 2013. The right column (m–r) analyzes trends for the same regions but excludes 2013 as a sensitivity test. See Fig. 6.2 in the main report caption and text for further details.

right-hand column shows trends to 2012, to test the effect of leaving out the highly anomalous end year (2013). While excluding 2013 has only a minor impact on the detection results for the northern tier—ANN region Fig. S6.3 (g,m) for trends starting prior to 1930,

it has a major impact on the late 20th century trend detection results for the eastern U.S.—JJA region Fig.6.3 (k,q). The latter detection result is not robust to excluding 2013 and thus depends quite critically on the one highly unusual year. Results for the upper Midwest—MAM are intermediate between these results; leaving out 2013 substantially reduces the robustness of the trend detection, but the trends from the early 20th century to 2012 are still generally detectable according to the models.

We now discuss the three remaining U.S. regions from Fig. 6.1 in the main report, which were not discussed in detail in the main text (California—ANN, Southern Plains—MAM, and Northern Plains—SON). Two of these regions did not exhibit significant linear trends according to statistical tests on linear trends over 1900–2013 (California—ANN and Southern Plains—MAM). These regions also do not have detectable long-term trends according to the model-based trend detection tests shown in Fig. S6.3 (h,j). California region trends were also not detectable for the DJF or MAM seasons (not shown). The U.S. Northern Plains—SON region analysis (Fig. S6.3l) indicates some detectable trends to 2013 (All-Forcing runs versus Control runs). Trends to 2013 are detectable beginning in the 1930s, 40s, and 50s, as shown by the black line extending above the purple shaded region. However, for this region the All-Forcing ensemble mean response (red line) is relatively small or even negative, except for trends beginning quite late in the 20th century, at which point the observed trends are not detectable. Note that the pink shaded region (5th to 95th percentile of All-Forcing trends) is slightly broader than the control run trend distribution but is also centered around the purple (control run overlap) shading, in contrast to the positive skewing of the pink shaded region compared to the purple shading for the other regions in Fig.S6.3. This lack of positive

skewing of the All-Forcing shaded region compared to the control is also indicative of the very weak All-Forcing response in the models for this region. The broadening of the pink region relative to the green/purple region is due to the former runs having a diversity of model responses to external forcings, while the latter runs had unchanging preindustrial forcings. The very small CMIP5 century-scale ensemble mean All-Forcing responses (red line) in the Northern Plains—SON region suggest that internal variability is the dominant contributor to the observed long-term trends in this region. This finding assumes that the All-Forcing response is adequately modeled by the CMIP5 models. Finally, our sensitivity tests excluding 2013 (Fig. S6.3r) indicate that the Northern Plains—SON region trend detection results are not very robust to the exclusion of 2013.

The Southern Plains—MAM region time series exhibits some additional interesting behavior. The time series (Fig. S6.3d) shows several decades with

very pronounced variations prior to about 1945, followed by several decades with much smaller variability. Although we find no detectable trends in this region, we suggest that the observed dataset here may require further assessment for possible temporal inhomogeneities, perhaps associated with secular changes in the observing network.

Although not shown here, we also performed some sliding trend analysis comparing the observed trends to CMIP5 Natural Forcing-only distributions. However, we had difficulties with this analysis owing to the relatively few models with available Natural Forcing runs extending to 2012, the relatively small number of ensemble members (in some cases, only one) for the available models, and the relatively few distinct modeling centers that have provided such runs so far. For these reasons, we are not presenting results from the All-Forcing versus Natural Forcing sliding trend analysis in this study.

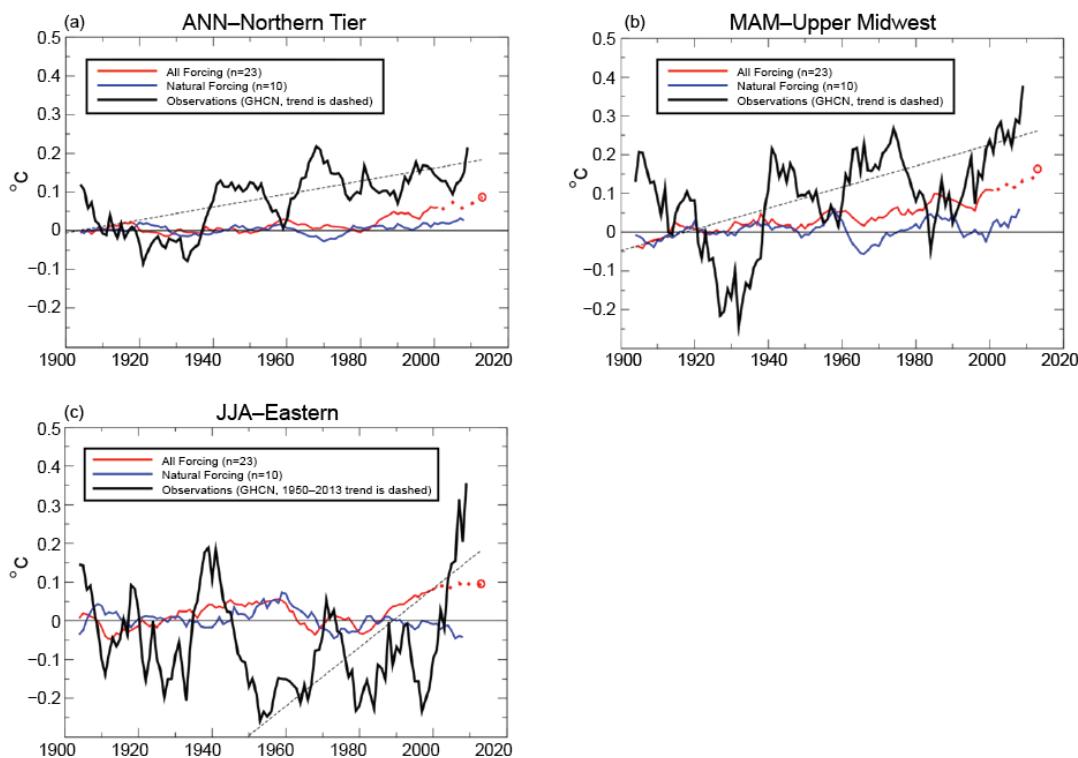


FIG. S6.4. CMIP5 multimodel ensemble mean All-Forcing (red) or Natural Forcing-only (blue) 9-yr running-mean anomalies relative to a 1900–1940 baseline. Time series shown are based on annual (a), March–May (b), or June–August (c) averaged data. Observed 9-yr running means are shown by the black thick lines in each diagram, with the linear trends of annual means shown by the black dashed lines. Results are shown for a) northern tier–ANN, b) upper Midwest—MAM, and c) eastern U.S.—JJA regions/seasons. The red circles at year 2013 show the 9-yr running mean All-Forcing anomaly centered on 2013. All-Forcing ensemble time series values that include any years beyond 2005 (and thus include some RCP4.5 projection values for at least some models) are denoted by the red-dashed segments. See text for further details.

Methodology for estimating the 2013 multimodel ensemble All-Forcing anomaly and the FAR. The time series (Fig. 6.2 a,d,g in the main report) depict the CMIP5 multimodel All-Forcing ensemble mean, relative to the 1900–40 baseline, as a dark red line; the upward-sloping black-dashed lines depict the observed linear trends from 1900 to 2013 (or 1950–2013 for Fig. 6.2g in the main report). The All-Forcing response is small and difficult to see in Fig. 6.2 in the main report, so it is shown in an expanded view (with nine-year running mean smoothing) for each region in Fig. S6.4 (thin red lines). Note that the All-Forcing ensemble mean responses are much smaller than the observed nine-year running mean changes and smaller than the observed linear trends (dashed black lines). The ensemble mean has considerable year-to-year variation (Fig. 6.2 a,d,g in the main report), so estimating the All-Forcing model ensemble’s mean for the year 2013 is difficult. Our approach is to estimate the 2013 value by using a temporally smoothed (nine-year running mean) version of the ensemble mean

time series. To obtain a nine-year running mean value centered on 2013, we extended the All-Forcing response curves to 2017 using the RCP4.5 scenarios in the model archives. The smoothed 2013 All-Forcing values (red circles in Fig. S6.4) are then used to shift the control run distributions in Fig. 6.2c,f,i in the main report to create the All Forcing distributions for 2013 shown in Fig. 6.2c,f,i in the main report and used for our FAR analysis. The red curves in Fig. S6.4 are dashed for values from 2001 on to indicate that these values are at least partly influenced by the RCP4.5 extensions beyond 2005 in the model data.

The blue curves in Fig. S6.4 show the Natural Forcing-only ensemble (10-model) results analogous to those for the All-Forcing ensemble just discussed. In principle, these could be used to create “Natural Forcing-only” versions of the distributions in Fig. 6.2c,f,i in the main report from which a fraction of attributable risk to anthropogenic forcing could be estimated. We have chosen not to do this, however, because of the lack of a long-term detectable trend in

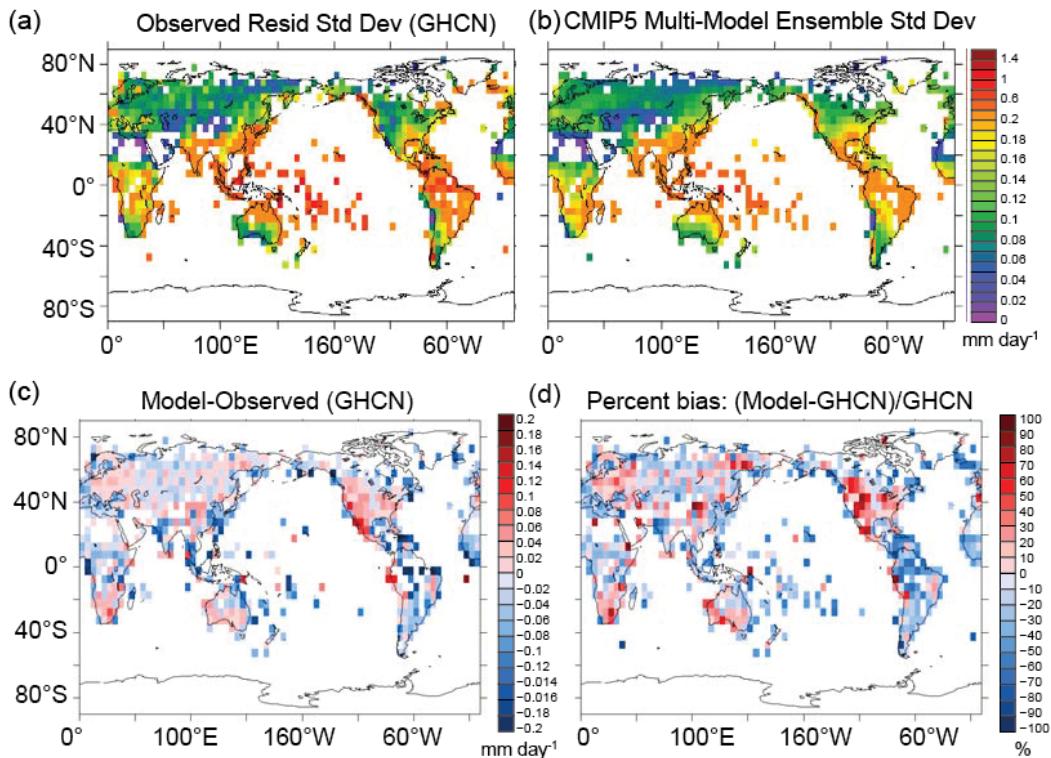


FIG. S6.5. Maps of standard deviation of low-pass (>10 yr filtered) precipitation anomalies (mm day⁻¹) based on annual data for: (a) GHCN observed residuals (1900–2013); and (b) the ensemble-mean standard deviation from the 23 CMIP5 control runs used in this study. The observed residuals were formed by subtracting the CMIP5 multi-model ensemble All-Forcing/RCP4.5 response from the observations to create annual mean residual anomalies series for comparison with the control run (intrinsic variability) simulations. (c) Bias map computed as the modelled standard deviation in (b) minus the observed residuals standard deviation in (a). (d) Same as in (c) but expressed as percent bias: $[(\text{model} - \text{observation}) / \text{observation}] \times 100\%$.

these focus regions compared to the Natural Forcing-only distributions, as discussed previously.

FAR is the fraction of attributable risk for anomalies as large as certain threshold values (here we use the second-ranked year in the observed record), and it is based on the All-Forcing anomaly distributions for 2013 compared to the unforced anomaly distributions. In this case, the fraction of risk is attributable to anthropogenic and natural forcing combined. The occurrence ratio (All-Forcing : Control) is the occurrence rate of anomalies as large as those for the second-ranked year under the All-Forcing scenario, divided by the corresponding rate in the control run distribution. FAR and the occurrence ratio are computed as follows: $FAR = 1 - pc/pf$, and the occurrence ratio is pf/pc , where pf and pc are the occurrence rates within the All-Forcing and Control run distributions, respectively, of anomalies exceeding the defined thresholds shown in the plots (thick gray vertical lines) in Fig. 6.2c,f,i in the main report.

Assessment of model-simulated precipitation and internal variability. How adequate are the CMIP5 models' simulations of precipitation in the focus regions? Of particular interest is the climate variability in their control runs, which we have used to estimate the real world's intrinsic (unforced) climate variability. For

the CMIP5 models, an assessment of the ensemble-mean precipitation geographical distribution and seasonal cycle was done by Flato et al. (2013; see Figs. 9.4, 9.38, 9.39 in the main report). Despite model biases, their figures suggest that the CMIP5 simulations of large-scale precipitation characteristics in our focus regions are sufficiently realistic for our purposes.

Examining variability, Fig. S6.5 shows the geographical distribution of low-frequency (>10 years) (standard deviations of annual precipitation for the observations (GHCN) versus the CMIP5 multimodel ensemble standard deviations from the model control runs. To make the observed standard deviations more comparable to control runs (which represent intrinsic climate variability alone) we subtracted the multimodel mean All-Forcing response from the observed time series to create a residual intrinsic variability estimate, which was then smoothed to include primarily variability with time scales longer than 10 years. The comparison shows (Fig. S6.5) that the models' standard deviation typically exceeds the observed low-frequency standard deviation in the U.S. regions where we are focusing. So, for our trend detection, the models may overestimate the intrinsic variability (i.e., the width of the green/pink/purple bands in Fig. 6.2 b,e,h in the main report and the widths of the distributions in Fig. 6.2 c,f,i in the

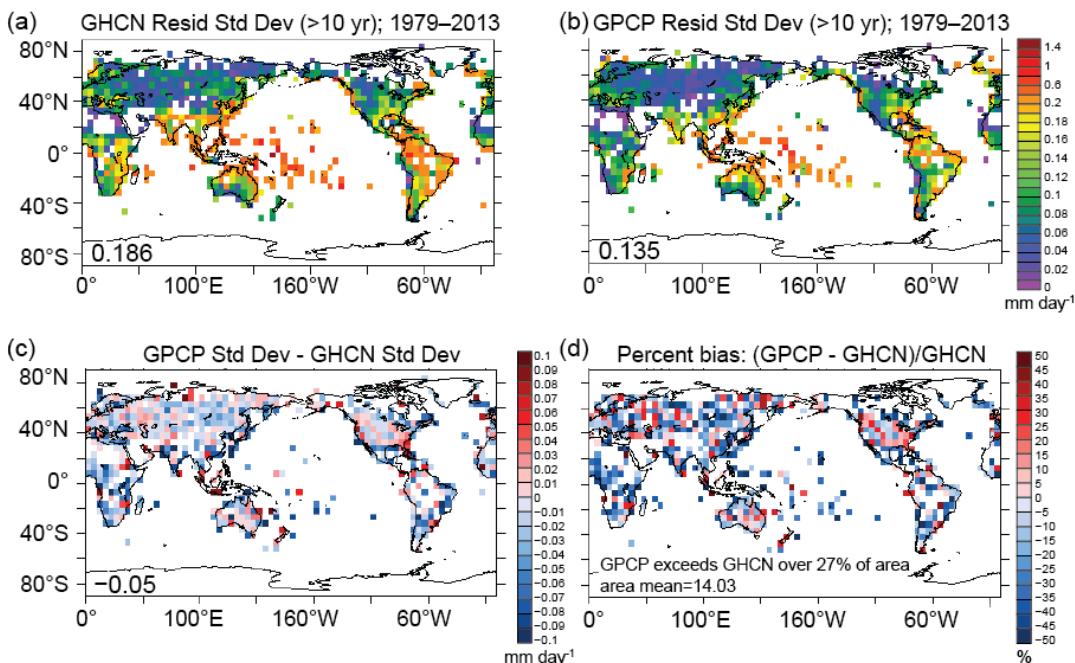


FIG. S6.6. As in Fig. S6.5, but comparing the >10 yr standard deviations of two different observational datasets for the shorter period (1979–2013). The two observational data sets are: (a) GHCN as in Fig. S5a, and (b) GPCP v2.2 (see text for details). (c) and (d) are difference maps and percent difference maps, respectively, for GHCN vs. GPCP. Note the different scales used for Figs. S6.5 vs. S6.6 for panels (c) and (d).

main report). Figure S6.6 also compares the observed GHCN low-frequency (> 10 years) standard deviation for the period 1979–2013 to that of an alternative (combined satellite/rain gauge) dataset over the area of common coverage (Global Precipitation Climatology Project v. 2.2; Adler et al. 2003; <http://www.esrl.noaa.gov/psd/data/gridded/data.gpcp.html>). This shows that there are even substantial uncertainties in estimating the precipitation standard deviations from observations, which is an important further caveat to our analysis and assessment. Also, comparison of the GHCN standard deviation for the short (1979–2013) versus full (1900–2013) data period (Fig. S6.5 versus Fig. S6.6) shows that the standard deviation is considerably larger for the full period, especially over the Northern Hemisphere continents, which illustrates the impact of the epoch considered for the decadal variability in the GHCN data.

The models' high-frequency (unfiltered annual means) intrinsic variability is also used in our study for the FAR analysis. Therefore, we also need to assess the models' high-frequency

variability. To assess this issue, we performed the following auxiliary calculations. For each of the three key regions where we found a detectable trend (northern tier—ANN, upper Midwest—MAM, and eastern U.S.—JJA) we compute an observed residual variability series by subtracting the multimodel ensemble mean All-Forcing response from the observed series. We remove the mean of these residuals and compare their histogram to that from the multimodel control run ensemble, which was obtained by combining 1000-member random samples from each of the 23 CMIP5 control runs into a 23 000 member aggregate control run distribution. The comparison of modeled and observed distributions for each region (Fig. S6.7) indicates that the multimodel ensemble provides a fairly realistic distribution of intrinsic variability, compared to the observed residual distribution. The standard deviations of the control run distributions for the northern tier—ANN and upper Midwest—MAM series are close to but slightly larger (8% and 4%, respectively) than the standard deviation of the observed residual series. This suggests that the mod-

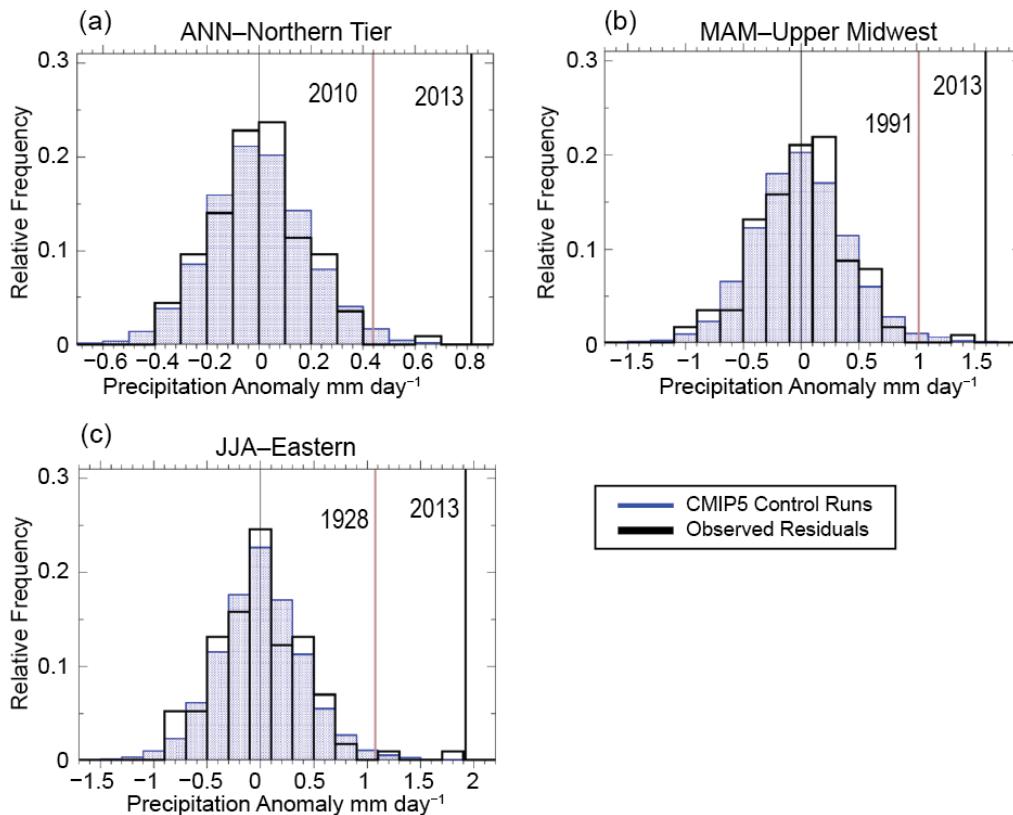


FIG. S6.7. Normalized histograms of annual- or seasonal-mean anomalies from the 23 CMIP5 model control runs (blue bars) vs. observed (GHCN) residuals (black bars). The observed residuals are computed by subtracting the CMIP5 All-Forcing ensemble mean response from the observed time series, and also subtracting the mean of these residuals (resulting in residuals having zero mean). Observed values for 2013 and the alternative threshold value are depicted by the thick black and gray vertical lines (see year labels).

eled estimates are likely adequate for our climate change detection purposes, although they will tend to make it slightly harder to detect forced trends and easier for the All-Forcing estimates to encompass the

observations. For the eastern U.S.—JJA region, the observed residual standard deviation is slightly larger (8%) than for the model control runs. Therefore, as a sensitivity test, we amplified the control run anoma-

TABLE S6.1. Lists of CMIP5 models used in the study for All Forcing (top) and Natural-Only Forcing (bottom) experiments. The lists include the short names of the models, the number of ensemble members included in our analysis in []'s, and a longer name for the modeling center.
All-Forcing Experiments:
BCC-CSM1.1 [3] Beijing Climate Center
CanESM2 [5] Canadian Centre for Climate Modelling and Analysis
CCSM4.0 [6] National Center for Atmospheric Research (U.S.)
CMCC-CM [1] Centro Euro-Mediterraneo per i Cambiamenti Climatici (Italy)
CNRM-CM5 [1] Centre National de Recherches Meteorologiques (France)
CSIRO Mk3.6.0 [1] Commonwealth Scientific and Industrial Research Organisation (Australia)
FGOALS-g2 [5] State Key Lab. Numerical Modeling for Atmos. Sci. and Geophys. Fluid Dyn. (China)
GFDL CM3 [5] Geophysical Fluid Dynamics Laboratory (U.S.)
GFDL-ESM2M [1] Geophysical Fluid Dynamics Laboratory
GFDL-ESM2G [1] Geophysical Fluid Dynamics Laboratory
HadGEM2-ES [4] Hadley Centre (United Kingdom)
INM-CM4 [1] Institute of Numerical Mathematics (Russia)
IPSL-CM5B-LR [1] L'Institut Pierre-Simon Laplace (France)
IPSL-CM5A-MR [3] L'Institut Pierre-Simon Laplace
IPSL-CM5A-LR [6] L'Institut Pierre-Simon Laplace
MIROC5 [5] Model for Interdisciplinary Research on Climate (Japan)
MIROC-ESM [3] Model for Interdisciplinary Research on Climate, Earth System Model
MIROC-ESM-CHEM [1] Model for Interdiscipl. Res. on Climate, Earth Sys. Mod, Chemistry Coupled
MPI-ESM-MR [3] Max Planck Institute (Germany)
MPI-ESM-LR [3] Max Planck Institute
MRI-CGCM3 [3] Meteorological Research Institute (Japan)
NorESM1-M [3] Norwegian Earth System Model
NorESM1-ME [1] Norwegian Earth System Model
Natural-Only Forcing Experiments:
BCC-CSM1.1 [1] Beijing Climate Center
CanESM2 [5] Canadian Centre for Climate Modelling and Analysis
CNRM-CM5 [1] Centre National de Recherches Meteorologiques (France)
CSIRO Mk3.6.0 [5] Commonwealth Scientific and Industrial Research Organisation (Australia)
GISS-E2-H [5] Goddard Institute for Space Studies (U.S.)
GISS-E2-R [5] Goddard Institute for Space Studies (U.S.)
HadGEM2-ES [1] Hadley Centre (United Kingdom)
IPSL-CM5A-MR [3] L'Institut Pierre-Simon Laplace (France)
IPSL-CM5A-LR [3] L'Institut Pierre-Simon Laplace
NorESM1-M [1] Norwegian Earth System Model

lies for this region by a factor of 1.08 and found that this had only a modest impact on our FAR or occurrence ratio estimates in Fig. 6.2i in the main report since this adjustment affects both the control and All-Forcing distribution similarly. The FAR estimate was 0.36 for the unadjusted anomalies versus 0.26 for the amplified anomalies. For the FAR estimates, we used 4000-member random samples from each model for a total sample size of 92 000. Also, the detection results shown for the eastern U.S.—JJA in Fig. 6.2f in the main report are robust to an order 8% increase in

the 95th percentile of the control run trends, but as mentioned previously, the trend detection results for this region are sensitive to the inclusion/exclusion of 2013 observed values.

In summary, our variability assessments suggest that the CMIP5 models can provide a useful assessment of precipitation low-frequency variability (trends) and annual or seasonal anomalies due to intrinsic climate variability. This provides support for using these models for our trend assessments and FAR calculations.

S7. OCTOBER 2013 BLIZZARD IN WESTERN SOUTH DAKOTA

LAURA M. EDWARDS, MATTHEW J. BUNKERS, JOHN T. ABATZOGLOU,
DENNIS P. TODEY, AND LAUREN E. PARKER

TABLE S 7.1. Names of CMIP5 models used in the analysis. The ensemble number used as well as the number of years of preindustrial control runs for daily temperature and precipitation. The number of model years for precipitable water is shown in parenthesis.			
Model Name	Modeling Center	Ensemble Number	PI Years
CanESM2	Canadian Centre for Climate Modelling and Analysis	r1ilpl	1096 (40)
CCSM4	National Center for Atmospheric Research	r2ilpl	156 (20)
CNRM-CM5	Centre National de Recherches Meteorologiques / Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique	r1ilpl	850 (25)
CSIRO-MK3-6-0	Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence	r1ilpl	500
GFDL-ESM2G	NOAA Geophysical Fluid Dynamics Laboratory	r1ilpl	500
MIROC5	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	r1ilpl	670
NorESM1M	Norwegian Climate Centre	r1ilpl	501 (29)
INMCM4	Institute for Numerical Mathematics	r1ilpl	500

S8. MULTIMODEL ASSESSMENT OF EXTREME ANNUAL-MEAN WARM ANOMALIES DURING 2013 OVER REGIONS OF AUSTRALIA AND THE WESTERN TROPICAL PACIFIC

THOMAS R. KNUTSON, FANRONG ZENG, AND ANDREW WITTENBERG

We present here several auxiliary analyses and figures relevant to our study, which were not possible to include in the main report due to space limits. In Fig. S8.1, we show for reference the seasonal mean anomaly maps and seasonal-mean extreme occurrence maps for temperature, which are analogous to Fig. 8.1 in the main text but for the individual seasons. We also present “sliding trend” analyses like those in Fig. 8.2 in the main report (c,d) but comparing 10-model Natural Forcing ensembles with 10-model All-Forcing ensemble subsets of the CMIP5 models. We describe some background on our method and rationale for estimating a Natural-Forcing-only ensemble mean model response for 2013, and the sensitivity of our results to this estimate. We assess the adequacy of simulated internal climate variability in the model for the focus regions in our study. Finally, we assess certain observational issues.

‘Sliding trend’ analysis of Natural Forcing vs. All-Forcing Ensemble. In Fig. S8.2 we present ‘sliding trend’ analysis of trends of varying lengths, all ending in 2012, for the Australian and far western tropical Pacific regions. These analyses are similar to those in Fig. 8.2 in the main report, but compare the All-Forcing trend distributions from a 10-model subset of the CMIP5 models to the Natural-Forcing trend distributions from the same 10 models. The trend analysis is done through 2012 instead of 2013 (as in the main text) because the Natural Forcing runs generally ended in 2012 and we could also test the sensitivity of our trend analysis to leaving out the highly anomalous 2013 values for the observations.

The results show that for all start dates up until about the late 1970s, the trends (to 2012) in the two regions are detectable compared to the multi-model Natural Forcing trend distributions (i.e., outside of the blue envelope). The trends in the Australia region are consistent with the All-Forcing 10-member ensemble (i.e., within the pink envelope) for virtually all start dates examined up to 2000. The trends for the far western Pacific region are consistent with the All-Forcing 10-member ensemble for start dates up to about the late 1970s.

Thus for most start dates beginning in the late 1800s and extending until at least as late as the late 1970s, the CMIP5 model simulations indicate that there is a detectable anthropogenic influence on temperature trends to 2012 in these two regions, according to our testing methodology.

Estimating the Natural-Forcing-only response for 2013. Since the CMIP5 models typically ended their Natural-Forcing runs between 2005 and 2012, we did not have a readily available 23-model estimate of the Natural Forcing ensemble mean response for 2013. However, 10 models had Natural Forcing runs available through 2012. Inspection of the Natural Forcing ensemble time series from those 10 models, in Fig. 8.2 a,b in the main report, suggests that an approximate Natural Forcing ensemble mean component for 2013 would be to reuse the value simulated for 2012 (“Mid Natural”). As sensitivity tests, we also performed our relative risk and fraction of attributable risk calculations assuming a “Low Natural” case of zero Natural Forcing contribution and a “High Natural” case using the maximum of the ensemble mean Natural Forcing response occurring at any point in the time series

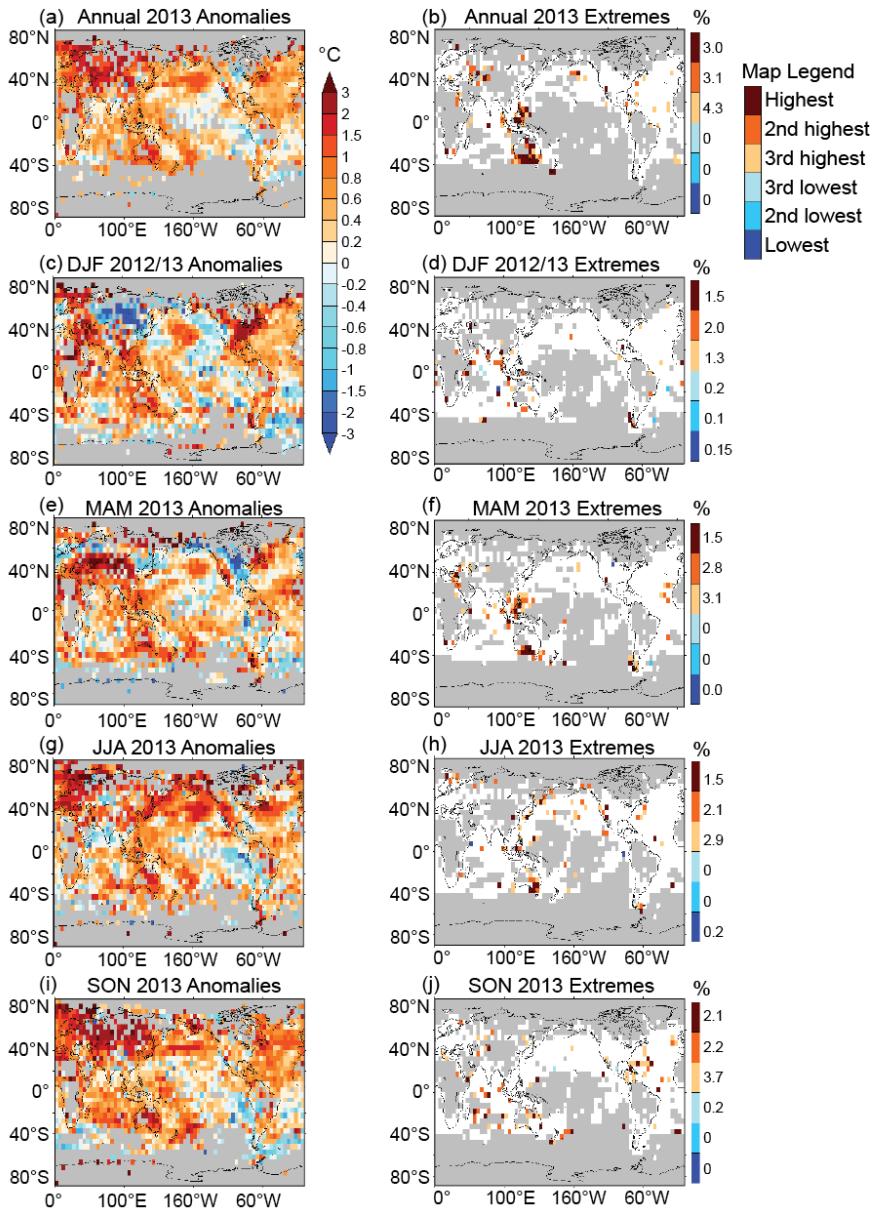


FIG. S8.1. Left column: Annual (a) or seasonal (c,e,g,i) mean surface air temperature anomalies ($^{\circ}\text{C}$) for 2013 (1961–90 base period) from the HadCRUT4 data set. The seasons are DJF (December 2012–February 2013); MAM (March–May); JJA (June–August); and SON (September–November). Right column: Colors identify grid boxes with annual (b) or seasonal (d,f,h,j) mean warm anomalies that rank 1st (dark red), 2nd (orange-red), or 3rd (yellow-orange) in the available observed record. Gray areas did not have sufficiently long records, defined here as containing at least 100 available annual or seasonal means, with a seasonal mean requiring at least one of three months to be available, and an annual mean requiring at least three of four seasons to be available. The percent values (right side of figures in right column) denote the percent of analyzed area for each category.

from around 1880 to 2012 as the estimate for 2013. A “Low Natural” (and not conservative) estimate is equivalent in this case to comparing the observed 2013 anomaly (relative to 1881–1920 baseline) against

ries from the observations, to produce an estimate of the unforced observed residual. The standard deviation, σ , of this “observed” residual for the Australia region is 0.272°C for observations com-

model control run variability alone. The various estimates used are shown as blue circles on Fig. 8.2a,b in the main report and listed in the first column of Table S8.1.

The results in Table S8.1 suggest that for all cases examined and for both regions examined, essentially all of the risk of the 2013 events is attributable to anthropogenic forcing, since anomalies as large as those observed in 2013 are either completely outside of the modeled distribution for the Natural Forcing only scenario or are an extremely rare event within that distribution. The analysis is repeated for a threshold temperature anomaly based on an alternative year (1998 for the western tropical Pacific region, which was slightly warmer than 2013, and 2009 for the Australia region, which was essentially tied for second-ranked year but substantially below 2013’s anomaly in magnitude). The results (Table S8.1) are robust in suggesting that essentially all of the risk of warm anomalies the size of those during these years is attributable to anthropogenic forcing.

Variance consistency test and robustness of findings to standard deviation adjustment. We evaluated the models’ control run interannual variability for the two focus regions for consistency with the internal variability estimated from the observations. The latter was estimated by subtracting the intermodel mean ensemble mean All-Forcing time series

TABLE S8.1. Estimates of observed and modeled temperature anomaly characteristics for 2013 and an alternative similar year (e.g., 1st or 2nd highest) for the two focus regions. See text for description of cases and the column entries. Anomalies for 2013 are relative to a baseline of 1881–1920. “Inf.” indicates cases where the observed anomaly is completely outside of the simulated distribution, so that the relative risk ratio is undefined.

Region/Case (Natural Forcing Estimate in °C)	Observed Anomaly °C for 2013 or Alt. yr.	Observed Percentile in Natural Dist. [%] (2013; Alt. yr)	Observed Percentile in All-Forcing Dist. [%] (2013; Alt. yr.)	Fraction of Attributable Risk (2013; Alt. yr)	Relative Risk (2013; Alt yr)
Australia Region (unadjusted)					
High Natural (0.304)	1.72; 1.17	Inf.; 99.9	99.3; 68.4	1.00; 1.00	Inf.; 376
Medium Natural (0.232)		Inf.; 100.0.	99.3; 68.4	1.00; 1.00	Inf.; 1330.
Low Natural (0.000)		Inf.; Inf.	99.3; 68.4	1.00; 1.00	Inf.; Inf.
Western Tropical Pacific Region (unadjusted)					
High Natural (0.212)	0.97; 1.02	100.0; Inf	75.8; 84.1	1.00; 1.00	5130; Inf.
Medium Natural (0.115)		Inf.; Inf.	75.8; 84.1	1.00; 1.00	Inf.; Inf.
Low Natural (0.000)		Inf.; Inf.	75.8; 84.1	1.00; 1.00	Inf.; Inf.
Western Tropical Pacific Region (adjusted std dev)					
High Natural (0.212)	0.97; 1.02	99.9; 100.0	72.1; 79.9	1.00; 1.00	227; 582
Medium Natural (0.118)		100.0; 100.0	72.1; 79.9	1.00; 1.00	1370; 2790
Low Natural (0.000)		Inf.; Inf.	72.1; 79.9	1.00; 1.00	Inf; Inf.

pared with 0.266°C for the multimodel sample of control runs, indicating good agreement. The standard deviation of the full observed time series is 0.421°C. Thus, while the observed 2013 anomaly of 1.72°C is about a 4 σ event in the observed record ($\sigma = 0.421^\circ\text{C}$), it represents an estimated 6 σ event compared to modeled internal variability.

For the far western tropical Pacific region, the estimated interannual standard deviation from the observed residuals is 0.172°C or 16% higher than the interannual standard deviation of the control runs. The observed 2013 anomaly of 0.97°C is almost a 4 σ event in the total observed distribution of annual temperatures but a 6 σ event compared to the estimated internal variability. As a sensitivity test, we scaled the western Pacific region modeled (control run) anomalies up by a factor of 1.22, which slightly exceeds the amount necessary to adjust for the estimated low variability bias. The results shown in Table S8.1 (adjusted) do not change the basic conclu-

sion that according to the models, the 2013 annual warm anomaly in this region is essentially entirely attributable to anthropogenic forcing in terms of its risk of occurrence.

Assessment of observational uncertainties. Here we consider some observational uncertainty issues. Our sliding trend analyses (e.g., Fig. 8.2c,d in the main report; Fig. S8.2) show via the black shading the 5th–95th percentile range of trends obtained using the 100-member HadCRUT4 observed ensemble (Morice et al. 2012), giving one indication of the observational uncertainty in these trend results. These indicate that our basic findings are robust to this estimate of observational uncertainty. A related issue is whether our results could depend on the use of the HadCRUT4 data, as opposed to an alternative dataset from the Australian Bureau of Meteorology (BOM) that is available for the relatively well-sampled period 1910–2013. We downloaded an all-Australia

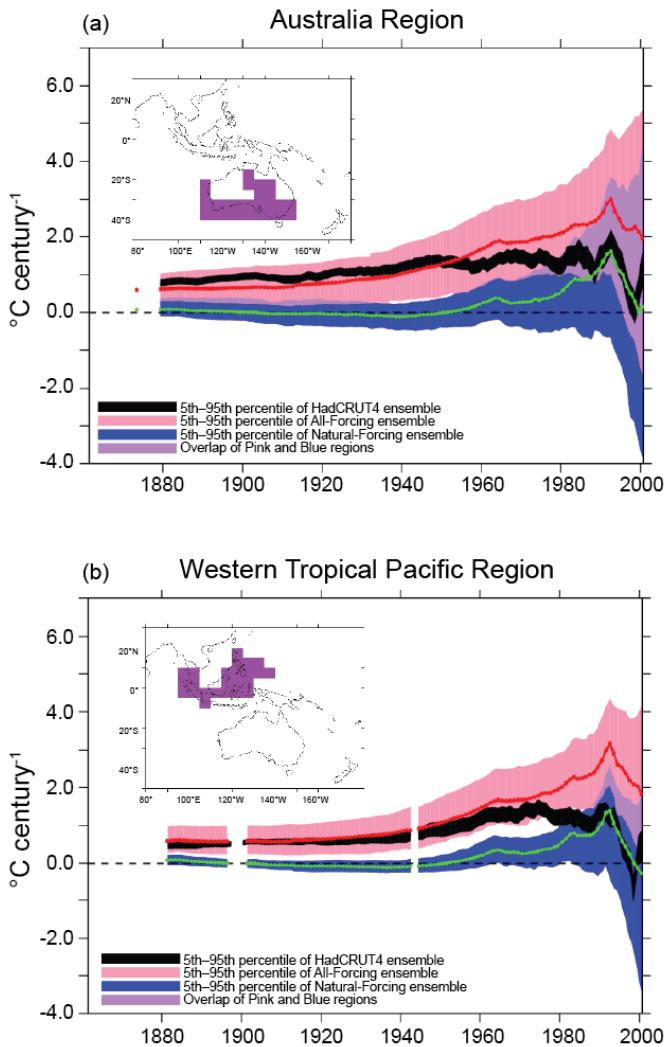


FIG. S8.2. Trends [$^{\circ}\text{C} (100 \text{ yr}^{-1})$] in the area-averaged annual-mean surface temperature series in Fig. S8.2 (a,b) as a function of starting year, with all trends ending in 2012. The black curves show trends from observations (HadCRUT4), indicating the 5th–95th percentile range for the HadCRUT4 observed ensemble (Morice et al. 2012). The red curves show the inter-model mean ensemble mean trends from the 10-member subset of the CMIP5 All-Forcing ensemble that provided natural forcing runs. The pink region represents the ‘All-Forcing’ hypothesis—i.e. the 5th–95th percentile range of trends from the All-Forcing runs. The blue-shaded region shows the 5th–95th percentile range of the alternative ‘Natural-Forcing-Only’ hypothesis using the same 10 models. Purple shading indicates where the pink- and blue-shaded regions overlap. The white spaces in the curves denote years where the initial “start year” was missing due to inadequate spatial or temporal coverage. Temporal coverage was assessed as in Fig. S8.1, and the spatial coverage was assessed for each year by requiring at least 33% non-missing annual means for the region.

index of temperature anomalies from the BOM data at: <http://www.bom.gov.au/climate/change/index.shtml#tabs=Tracker&tracker=timeseries>.

First, we compare the BOM time series cited above (1910–2013) to the HadCRUT4 data averaged over roughly the same Australia region (not the identical region because the HadCRUT4 is available on a $5^{\circ} \times 5^{\circ}$ grid). Figure S8.3 shows a comparison of the seven-year running mean time series derived from the BOM and HadCRUT4 data (reference period 1961–90). This shows that the anomalies in these two datasets are very similar when averaged over the Australian region as a whole. Our main analysis focuses on a sub-region of Australia based on those areas with extreme annual means as identified in Fig. 8.1 in the main report, and for this, we use the HadCRUT4 data, which seems appropriate based on the above comparison.

Another observational issue is the use of different reference periods for estimating the magnitude of the 2013 anomaly relative to preindustrial levels. In general, we would prefer to use as early a reference period as is practical, since earlier periods are closer to preindustrial conditions and we are trying to estimate the anthropogenically forced departure from such conditions. We find, using the HadCRUT4 data averaged over the Australia sub-region in our study, that the anomaly for the available years in 1881–1920 is about 0.2°C lower than that for 1910–49. This difference is much smaller than the 2013 anomaly of 1.72°C . Even adjusting the 2013 anomaly down by 0.2°C (i.e., using the years 1910–49 as the base period), the resulting anomaly for 2013 (1.52°C) remains outside of the range of anomalies in the Natural Forcing distribution shown in Fig. 8.2e in the main report. In addition, the Natural Forcing response (for 2013, if assumed to be equivalent to that simulated for 2012) is about 0.1°C smaller using the 1910–49 base period than using the 1881–1920 period (since the 1881–1920 period featured cooler temperature in the Natural Forcing runs). Taking this adjustment into account implies that the required adjustments for the observations versus the Natural Forcing distribution is a net reduction in their separation by only about 0.1°C . Again, we conclude that the observed anomaly is not simulated in the large multimodel sample of annual means for 2013 Natural Forcing conditions. In short, our finding that the 2013 observed anomaly is outside of the range of model simulated natural variability

HadCRUT4 vs BOM Australia Surface Temperature Anomalies
1961–1990 Base Period: 7-yr Running Means

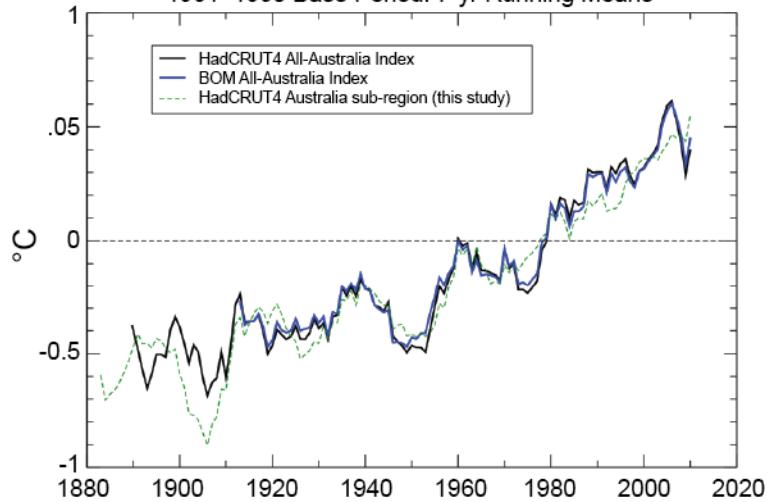


FIG. S8.3. Comparison of all-Australia timeseries of temperature anomalies (relative to 1961–90 base period) for the HadCRUT4 vs. Australian Bureau of Meteorology data set. See text for details. A seven-year running mean was applied to all data sets. The green dashed curve shows the HadCRUT4 data for the sub-region of Australia with near-record high annual-mean temperature anomalies during 2013 (see Figs. 8.1, 8.2 in the main report for region description).

(including Natural Forcing) remains robust to this reference period issue as well.

Considering the far western tropical Pacific region, the use of a later period (1910–49) versus an earlier period (1881–1920) results in a lower observed anomaly magnitude in 2013 by 0.12°C but also, coincidentally, a lower estimated magnitude of the Natural Forcing response in 2012 by almost the same magnitude

(0.12°C). Thus, the estimated occurrence rate of the 2013 anomaly in the Natural Forcing distribution would be essentially the same for the 1910–49 base period as for the 1881–1920 base period, due to these offsetting effects, and our conclusions about exceptional nature of the 2013 anomaly compared to Natural Forcing simulations remain robust.

S9. THE ROLE OF ANTHROPOGENIC FORCING IN THE RECORD 2013 AUSTRALIA-WIDE ANNUAL AND SPRING TEMPERATURES

SOPHIE C. LEWIS AND DAVID J. KAROLY

Table S9.1: List of CMIP5 climate models and ensembles used in this study. Further details of individual models can be found from the Program in Climate Model Diagnosis and Intercomparison (PCMDI; <http://cmip-pcmdi.llnl.gov>). Note that historicalNat contributions with HadGEM2-ES begin only in 1860 and hence provide 145 model years.

Experiment	Major forcings	Years Analyzed	Baseline
Historical	Anthropogenic (greenhouse gases, aerosols, ozone) and natural (solar, volcanics)	1911–2005	1911–40
RCP8.5	Anthropogenic (greenhouse gases, aerosols, ozone scenarios) and natural (solar)	2006–20	1911–40
HistoricalNat	Solar, volcanics	1850–2005	1911–40
piControl	Non-evolving pre-industrial forcings	All	Long-term mean

Table S9.2: CMIP5 model experiments analyzed, major forcings imposed, model years analyzed and climatology used to calculate temperature anomalies [modified from Lewis and Karoly (2013)].

Model	Realizations Historical	RCP8.5	HistoricalNat	piControl
bcc-csm1-l	rlilpl, r2ilpl, r3ilpl	rlilpl	rlilpl	rlilpl
CCSM4	rlilpl, r2ilpl, r3ilpl, r4ilpl, r5ilpl, r6ilpl	rlilpl, r2ilpl, r3ilpl, r4ilpl, r5ilpl, r6ilpl	rlilpl, r2ilpl, r4ilpl, r6ilpl	rlilpl, r2ilpl, r3ilpl
CNRM-CM5	rlilpl, r2ilpl, r3ilpl, r4ilpl, r5ilpl, r6ilpl, r7ilpl, r8ilpl, r9ilpl	rlilpl, r2ilpl, r4ilpl, r6ilpl, r10ilpl	rlilpl, r2ilpl, r4ilpl, r5ilpl, r8ilpl	rlilpl
CSIRO-Mk3-6-0	rlilpl, r2ilpl, r3ilpl, r4ilpl, r5ilpl, r6ilpl, r7ilpl, r8ilpl, r9ilpl, r10ilpl	rlilpl, r2ilpl, r3ilpl, r4ilpl, r5ilpl, r6ilpl, r7ilpl, r8ilpl, r9ilpl, r10ilpl	rlilpl, r2ilpl, r3ilpl, r4ilpl, r5ilpl	rlilpl
FGOALS-g2	rlilpl, r2ilpl, r3ilpl, r4ilpl, r5ilpl	rlilpl	rlilpl, r2ilpl, r3ilpl	rlilpl
GISS-E2-R	rlilpl, r2ilpl, r3ilpl, r4ilpl, r5ilpl, r6ilpl, rlilp2, r2ilp2, r3ilp2, r4ilp2, r5ilp2, r6ilp2, rlilp3, r2ilp3, r3ilp3, r4ilp3, r5ilp3, r6ilp3	rlilpl, rlilp2, rlilp3	rlilpl, r2ilpl, r3ilpl, r4ilpl, r5ilpl, rlilp3, r2ilp3, r3ilp3, r4ilp3, r5ilp3	rlilpl, rlilp2, rlilp3
HadGEM2-ES	rlilpl, r2ilpl, r3ilpl, r4ilpl, r5ilpl	rlilpl, r2ilpl, r3ilpl, r4ilpl	rlilpl, r2ilpl, r3ilpl, r4ilpl	rlilpl
IPSL-CM5A-LR	rlilpl, r2ilpl, r3ilpl, r4ilpl, r5ilpl, r6ilpl	rlilpl, r2ilpl, r3ilpl, r4ilpl	rlilpl, r2ilpl, r3ilpl	rlilpl
NorESM1-M	rlilpl, r2ilpl, r3ilpl	rlilpl	rlilpl	rlilpl

SII. UNDERSTANDING AUSTRALIA'S HOTTEST SEPTEMBER ON RECORD

JULIE M. ARBLASTER, EUN-PA LIM, HARRY H. HENDON, BLAIR C. TREWIN,
MATTHEW C. WHEELER, GUO LIU, AND KARL BRAGANZA

Data and predictors. Monthly maximum temperatures from the Australian Water Availability Project (AWAP) gridded dataset (Jones et al. 2009) were analyzed on a 0.25° grid over Australian land points. Relative to the 1982–2011 base period, the Australian-average maximum temperature anomaly from the AWAP dataset is 2.75°C for September 2013, which is slightly warmer than the 2.73°C anomaly from the homogenized annual temperature dataset of ACORN-SAT (<http://www.bom.gov.au/climate/change/acorn-sat/>) used in Bureau of Meteorology 2013. Note Bureau of Meteorology 2013 also uses 1961–90 as the base period compared to the 1982–2011 used here. Relative to the 1961–90 base period, the Australian average maximum temperature anomaly is 3.32°C (AWAP) and 3.41°C (ACORN). Observed sea surface temperatures from Reynolds et al. (2002) were used for the ENSO index (based on Niño-3.4 SSTs: 5°N–5°S, 170°–120°W) and the Indian Ocean Dipole mode index [western pole (10°S–10°N, 50°–70°E); eastern pole (10°S–0°, 90°–110°E); Saji et al. 1999]. The SAM was calculated as the first EOF of mean sea level pressure (MSLP) anomalies over 20°–75°S (e.g., Lim et al. 2011) from the ERA-Interim reanalysis (Dee et al. 2011). Soil moisture estimates are from Raupach et al. (2009) for the upper-layer (< 0.2m). The anomaly and regression patterns in Figs. S11.2, S11.3, and S11.4 use Reynolds OI v2 for SST (Reynolds et al. 2002) and ERA Interim reanalysis for MSLP (Dee et al. 2011). For all observational analysis, a climatological period of 1982–2011 is used, these being the years in common across all the datasets.

The regression model was built separately for each grid point (Fig. 11.1b–f in the main report) and for the Australian average mean maximum temperatures

(Fig. 11.1l in the main report). Similar Australian average values were found from averaging the spatial plots in Fig. 11.1 (in the main report) compared to those displayed in the bar plot.

For Fig. 11.1 (in the main report) and Fig. S11.1 the 95% prediction interval was computed as $\pm 1.96Se$, where Se is the standard error:

$$Se = \frac{1}{n-2} \sum_{i=1}^n e^2, \quad (1)$$

where e is the residual [$y - \hat{y}$, where \hat{y} is the predicted y by x]. To avoid overfitting, we required at least 10 degrees of freedom ($t - M - 1 \geq 10$, where t is the sample size and M is the number of predictors), with at least five data points per predictor ($M \leq t/5$), following Mo and Straus's (2002) method.

POAMA experiments. The POAMA seasonal forecast system is a fully coupled atmosphere–land–ocean model used operationally at the Bureau of Meteorology. Note it does not include changes in anthropogenic or natural forcings, such as increasing greenhouse gases, though much of the warming signal will be contained in the initial conditions. Retrospective forecast skill from POAMA (1981–2010) for prediction of Australian average maximum temperatures at zero lead time (i.e., initialized on 1 September for the month of September) is 0.75 in September (e.g., White et al. 2014). However, skill for month 1 at 10-day lead time (i.e., initialized on 21 August) drops to 0.42.

In regard to the maximum temperature of September 2013, forecasts initialized on 1 September 2013 produced even stronger September warm anomalies (not shown), but we focus on the earlier start in

order to account for any atmospheric adjustment in the scrambled runs. This allows a separation of the roles of atmospheric/land initial conditions and SST boundary forcing for generating the warm anomaly.

The scrambled experiments are performed by replacing the observed initial conditions for 21 August 2013 by sampling the initial states for 21 August from 30 previous years (Table 11.1 in the main report). In so doing, the spread of the initial states in the

scrambled experiments is necessarily larger than the spread provided by the coupled-ensemble generation strategy used to create the original ensemble forecasts (Hudson et al. 2013). We account for this additional spread in significance testing for the difference of two means by computing the standard deviation of the individual ensemble members that contribute to the ensemble mean forecast.

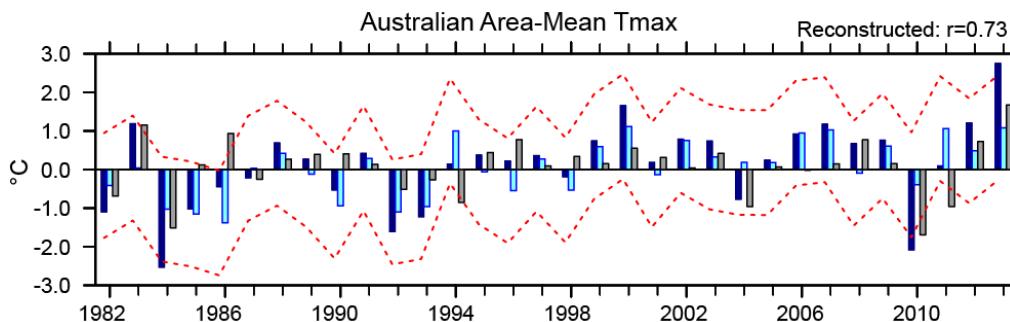


FIG. S11.1. Observed (navy), reconstructed (light blue) and residual (gray) Australian average maximum temperature anomalies over 1982–2013, with anomalies formed from the 1982–2011 base period. The red dotted lines represent the 95% prediction interval. The temperature anomalies in 1982–2011 were reconstructed in calibration mode (i.e, using the training data set that was used for constructing the regression model).

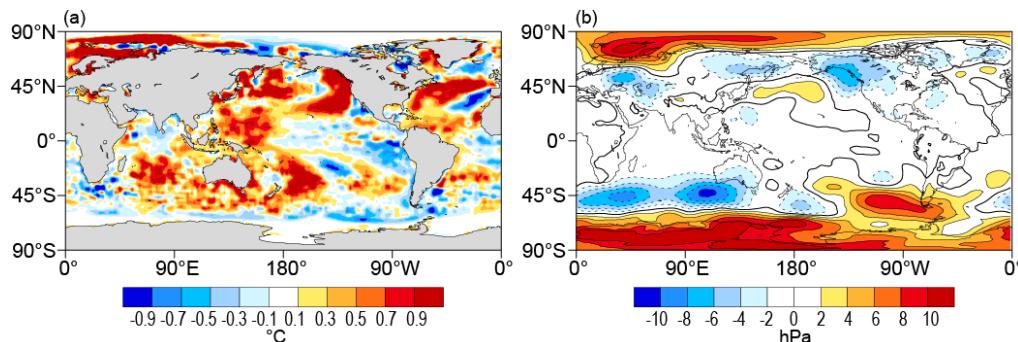


FIG. S11.2. Observed anomalies of (a) SST (°C) and (b) MSLP (hPa) for September 2013 relative to the 1982–2011 base period.

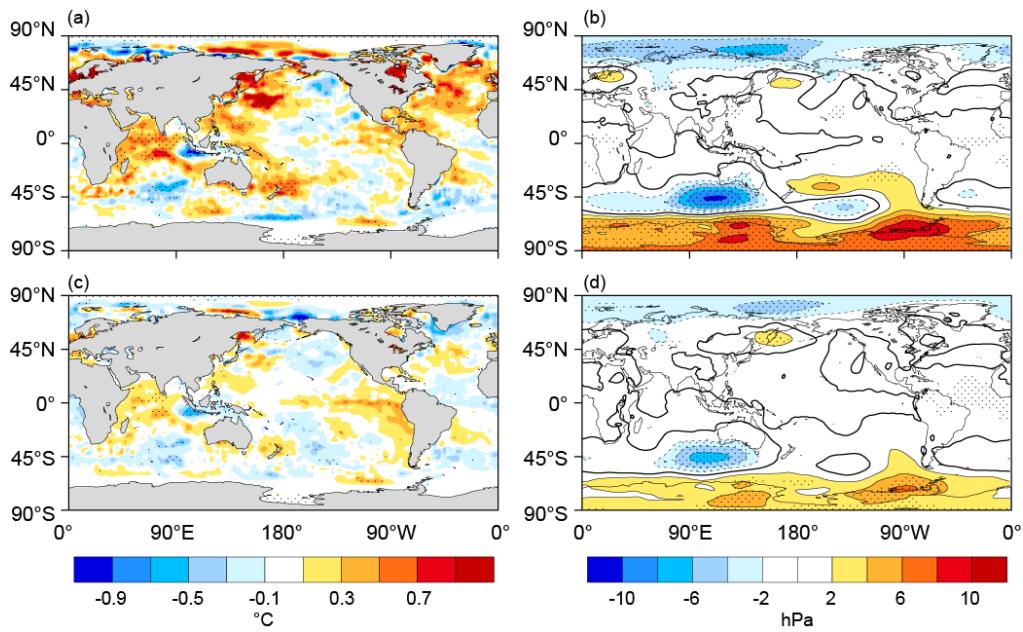


FIG. S11.3. Regressions of (a) SST ($^{\circ}\text{C}$) and (b) MSLP (hPa) onto Australian-average maximum temperatures for September over the 1982–2011 period. (c) and (d) same as (a) and (b) but using detrended timeseries. The regression coefficients are scaled by the magnitudes of 2013 September area averaged Australian maximum temperature with and without trends, respectively. Statistically significant anomalies at the 95% confidence level (based by a two-tailed Student t-test) are stippled.

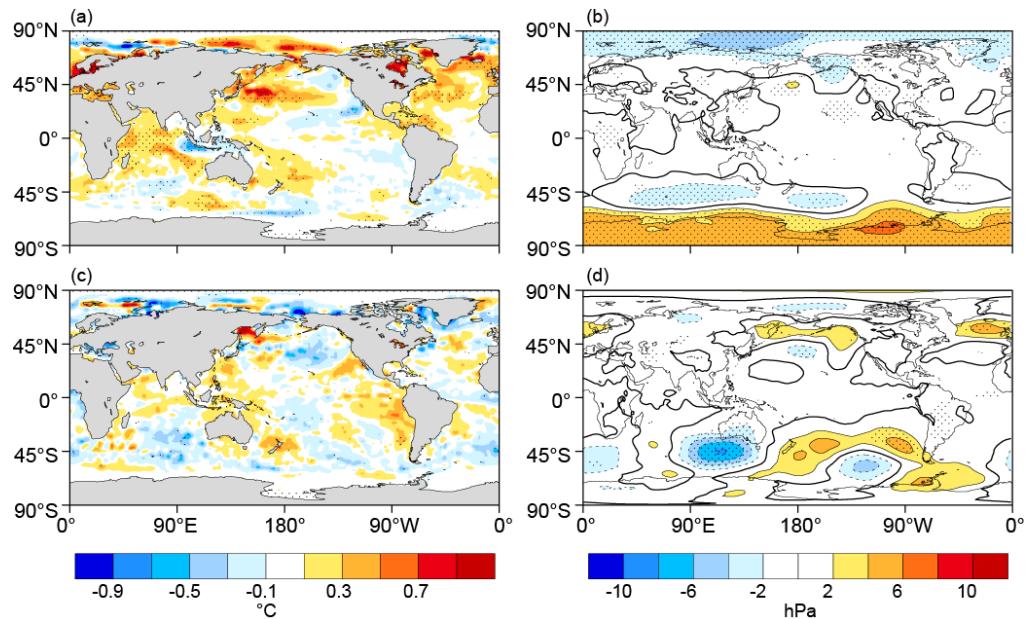


FIG. S11.4. Regression of (a) SST ($^{\circ}\text{C}$) and (b) MSLP (hPa) onto the Australian-average reconstructed maximum temperatures for September from the multiple linear regression analysis over the 1982–2011 period. (c) and (d), same as (a) and (b) but onto the residual time series.

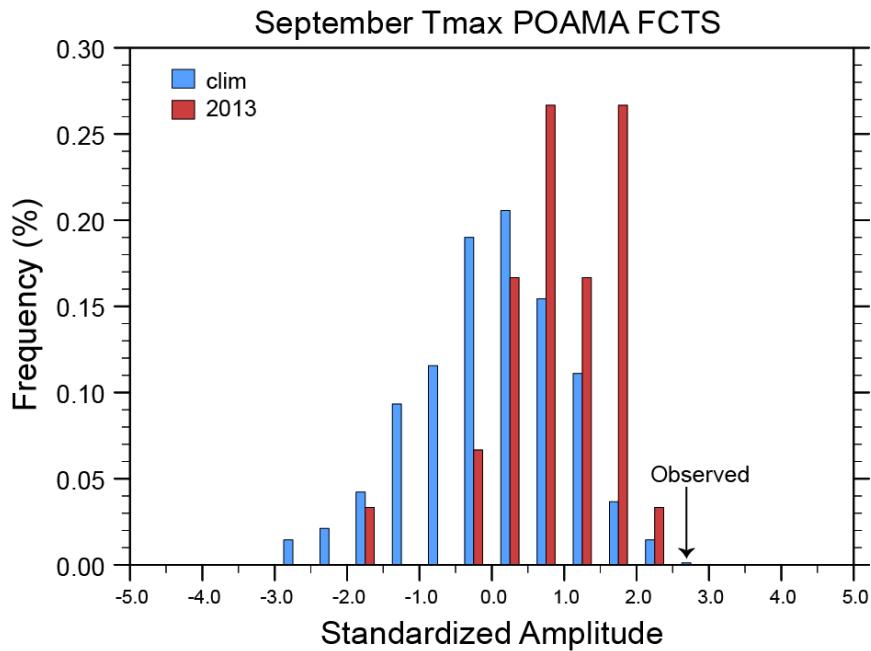


FIG. S11.4. Histograms of POAMA forecasts of Australian-average September maximum temperatures (in units of standard deviation) from the hindcast climatology (blue bars) and for 2013 (red bars). All forecasts consist of 30-members starting from the initial conditions of 21 August. The hindcast climatology is 1981–2010. The arrow indicates the observed standardized anomaly for September 2013 (≈ 2.68). All POAMA forecasts were standardized by the 1981–2010 hindcast standard deviation and the observed value was standardised using observations from 1981–2010.

S12. CLIMATE CHANGE TURNS AUSTRALIA'S 2013 BIG DRY INTO A YEAR OF RECORD-BREAKING HEAT

ANDREW D. KING, DAVID J. KAROLY, MARKUS G. DONAT, AND LISA V. ALEXANDER

TABLE S12.1. List of the 35 CMIP5 models (historical and RCP4.5 runs) used in this analysis.

Model Name	
ACCESSI-0	GISS-E2-H
ACCESSI-3	GISS-E2-H-CC
bcc-csm1-l	GISS-E2-R-CC
bcc-csm1-l-m	HadCM3
BNU-ESM	HadGEM2-AO
CanESM2	HadGEM2-CC
CCSM4	HadGEM2-ES
CESMI-BGC	INM-CM4
CESMI-CAM5	IPSL-CM5A-LR
CMCC-CM	IPSL-CM5A-MR
CMCC-CMS	IPSL-CM5B-LR
CNRM-CM5	MIROC5
CSIRO Mk3-6-0	MIROC-ESM
FIO-ESM	MIROC-ESM-CHEM
GFDL-CM3	MPI-ESM-LR
GFDL-ESM2G	MPI-ESM-MR
GFDL-ESM2M	MRI-CGCM3
	NorESM1-M

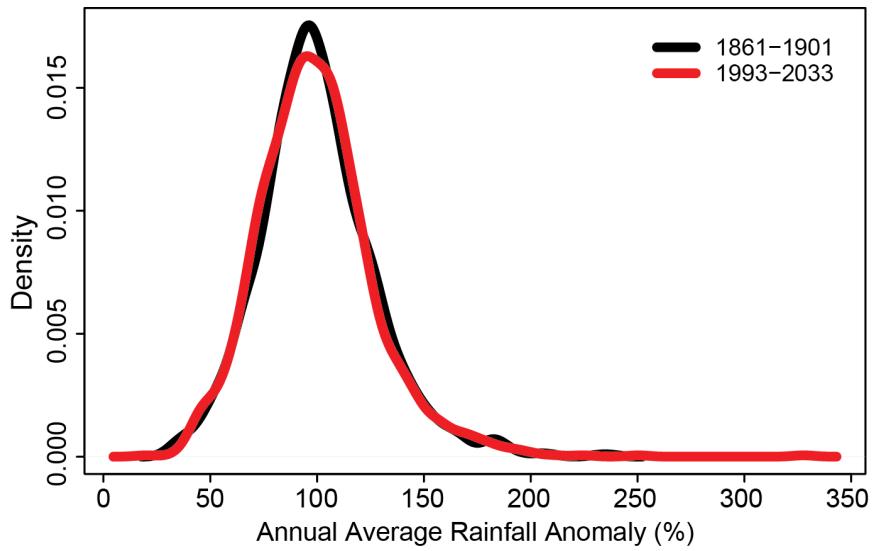


FIG. S12.1. PDFs of annual rainfall anomalies in model years representing 1861–1901 (black curve) and 1993–2033 (red curve).

S13. THE ROLE OF ANTHROPOGENIC CLIMATE CHANGE IN THE 2013 DROUGHT OVER NORTH ISLAND, NEW ZEALAND

LUKE HARRINGTON, SUZANNE ROSIER, SAM M. DEAN, STEPHEN STUART, AND ALICE SCAHILL

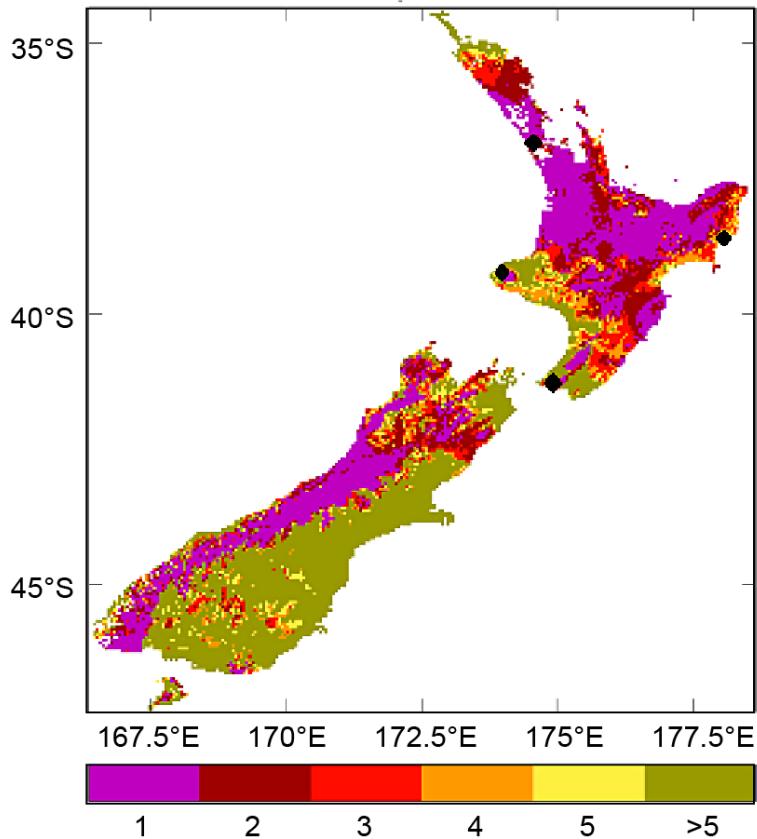


FIG. S13.1. Ranking of July 2012–May 2013 potential evapotranspiration deficit (PED) relative to the other 40 years in the NIWA Virtual Climate Station Network (VCSN) data set (Porteous and Mullan 2013). Rank 1 (purple) means the highest PED in 41 years. Olive colours signify the PED was not in the top five. Black diamonds indicate locations used for the 4-station NIPI calculations.

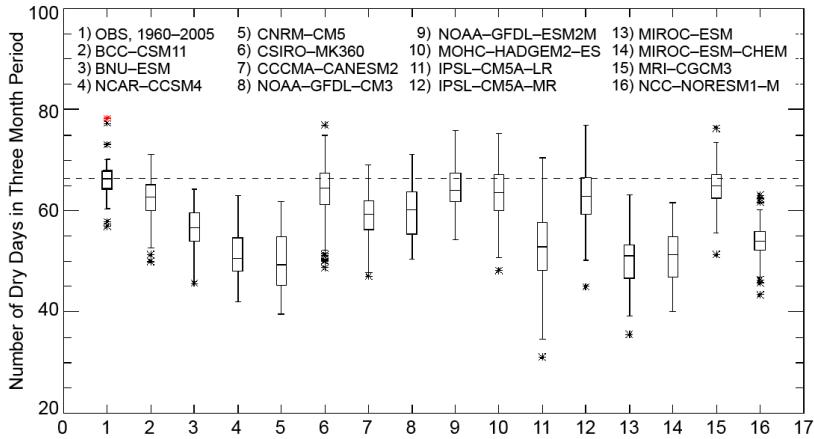


FIG. S13.2. Box plot comparing the observed distribution of extended summer three-month dry day (3MDD) maxima over the North Island NIPI distribution (**bold**) to 15 CMIP5 models, between 1960 and 2005. Each box indicates the median and first and third quartiles, while the whiskers extend to the last values that are 1.5 times the interquartile range above or below the quartiles. The red asterisk marks the 2013 drought event.

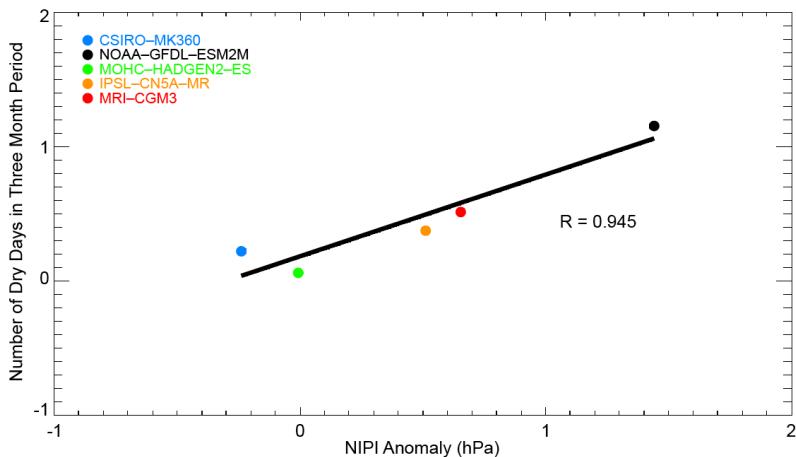


FIG. S13.3. Scatter plot showing the relationship between the shift (**ALL simulations minus NAT**) in mean number of dry days per month and the mean shift in NIPI for five CMIP5 models, between 1960 and 2005. The chosen models accurately simulated the observed distribution of both the NIPI and 3MDD, according to the validation criteria outlined in the main text. A strong positive correlation ($R = 0.95$) exists, with every 1 hPa increase in NIPI corresponding to a mean increase of 0.6 dry days per month.

SI4. ASSESSING HUMAN CONTRIBUTION TO THE SUMMER 2013 KOREAN HEAT WAVE

SEUNG-KI MIN, YEON-HEE KIM, MAENG-KI KIM, AND CHANGYONG PARK

CMIP5 data processing. The “historical” simulations are divided into two periods: 1860–1919 (referred to as ALL_P0) and 1954–2013 (referred to as ALL_P1). ALL_P0 is assumed to represent cold conditions close to the preindustrial period with a weaker human contribution. ALL_P1 represents current conditions with natural and anthropogenic forcings, and it is constructed by extending the historical experiment (1954–2005) up to 2013 using the Representative Concentration Pathways (RCP) 4.5 experiment (2006–13). Because RCP scenarios do not diverge appreciably until the near-term future (Moss et al. 2010), we chose RCP4.5 data, which provides the largest number of model samples, for this extension. We obtained 102 and 105 ensemble members for ALL_P0 and ALL_P1, respectively (Table S14.1). We also use 27 “historicalGHG” (GHG_P1) runs and 38 “historicalNat” (NAT_P1) runs for 1953–2012 (Note that we use a slightly different 60-year period for these experiments since they end in 2012. Results are unchanged when applying the same period of 1953–2012 to ALL_P1). All modeled SST fields, which are internally calculated from atmosphere–ocean coupling, are interpolated onto $2^\circ \times 2^\circ$ grids to match the ERSST observations. SST anomalies from observations are with respect to the 1971–2000 mean while SST anomalies from all model runs are obtained relative to each 1971–2000 mean of ALL_P1 to account for different climatology responses to different forcing factors.

Further details on SST projection method. The SST projection obtained based on linear regression described in the main text represents an area-weighted sum of the regression coefficient multiplied by the SST anomaly in each JJA. This approach is equivalent to carrying out a singular value decomposition (SVD) analysis using Korean Tmin and East Asian SST patterns (see Fig. S14.2). The first SVD mode shows a very similar spatial pattern to the SST regression pattern in Fig.14.2a in the main report. The resulting

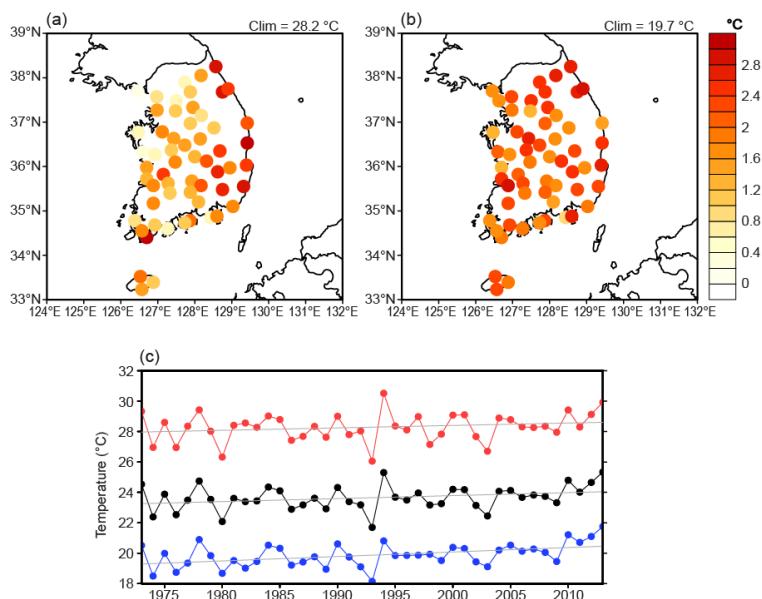


FIG. S14.1. (a) Distribution of JJA mean daily maximum temperature anomalies observed at 59 stations of South Korea in 2013. Anomalies are with respect to 1971–2000 mean. (b) Same as (a) but for daily minimum temperature. (c) Time series of JJA mean daily maximum (red), mean (black), minimum (blue) temperatures averaged over 12 stations. Gray straight lines represent linear trends.

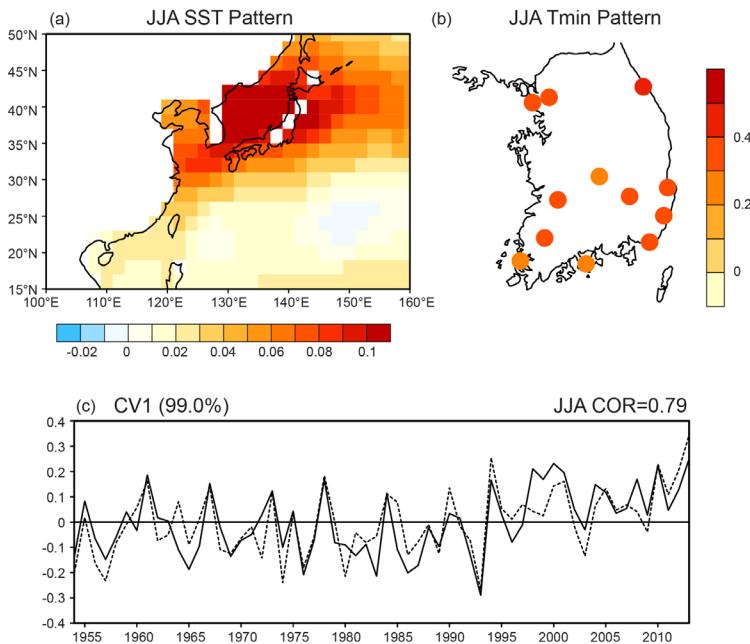


FIG. S14.2. The first SVD mode of (a) observed SST and (b) Tmin in Korea in JJA during 1954–2013. This mode explains 99% of the total covariance. (c) Associated temporal coefficients of the observed SST (solid) and Tmin in Korea (dashed line). Units in all plots are non-dimensional and meaningless.

SVD temporal coefficients for Korean Tmin and East Asian SST display a close relationship with the correlation coefficient of 0.79. We also find our results insensitive to the use of de-trended Korean temperatures when estimating the SST regression pattern (Fig. S14.3), which suggests that this SST pattern is mostly driven by natural internal variability.

Discussion on contribution of individual forcings. In contrast to a linear increase in SST projections from GHG_P1, ALL_P1 is characterized by a weak increasing trend in SST projection until the 1990s and a stronger increasing trend after that. Similar behavior can also be seen from observations (Fig. 14.2b in the main report). Here we discuss possible causes of the long-term cooling during the latter half of the 20th century in ALL_P1. Internal variability is unlikely to cause this cooling pattern because the SST in the coupled models is freely evolving, and the multimodel means will cancel out different SST anomalies across the models. There are two potential external forcings that are likely to exert long-term cooling effects for this period (Bindoff et al. 2013): natural forcings (NAT; due to changes in solar and volcanic activities) and anthropogenic forcings other than greenhouse gases (OA; mainly due to an increase in aerosols). Figure S14.4 compares multimodel mean SST projections obtained from ALL_P1, GHG_P1, and NAT_P1 as well as ANT (anthropogenic forcing response, estimated from ALL-NAT) and OA responses (estimated from ALL-NAT-GHG). Here 27 common model runs that provide data for ALL_P1, GHG_P1, and NAT_P1 (Table S14.1) are used in order to remove possible influence of different model groups across experiments. Results indicate that both NAT and OA can be partly responsible for the long-term cooling until the 1990s, with the ANT response

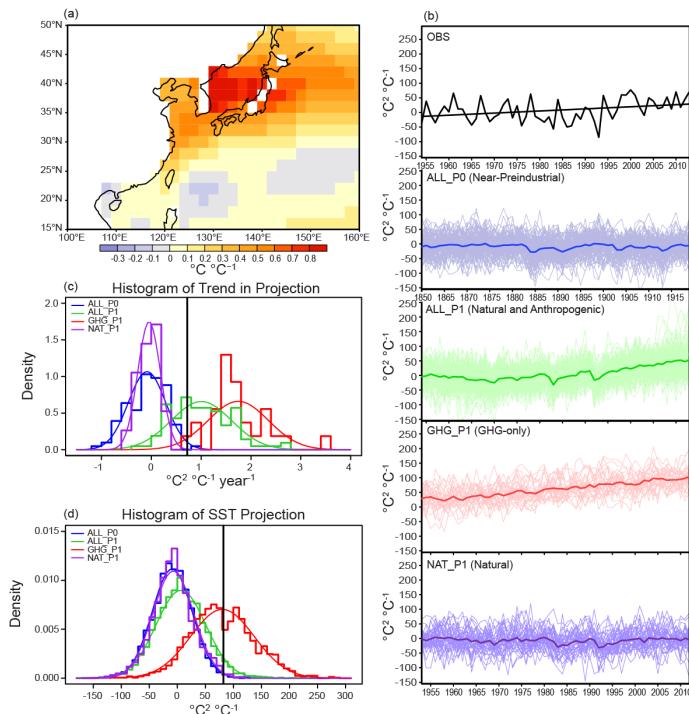


FIG. S14.3. Same as Fig. 14.2 in the main report but for using de-trended series of Korean Tmin. (a) SST pattern regressed onto de-trended series of Korean JJA mean Tmin. (b) Time series of SST projection from observation, ALL_P0, ALL_P1, GHG_P1, and NAT_P1. The SST projection is an area-weighted sum of the regression coefficient ($^{\circ}\text{C } ^{\circ}\text{C}^{-1}$) multiplied by the SST anomaly ($^{\circ}\text{C}$), which results in a unit of $^{\circ}\text{C}^2 \text{ } ^{\circ}\text{C}^{-1}$. The black straight line represents the observed linear trend. Thick colored lines indicate the ensemble mean of each experiment. (c) Normalized histogram for trend in SST projection from ALL_P0, ALL_P1, GHG_P1, and NAT_P1 in comparison with the observed trend (vertical black line). (d) Same as (c) but for SST projection values from models and the observed projection value in 2013.

displaying a steadier increase from the 1970s. However, there are large uncertainties in these indirect estimates, mainly due to substantial uncertainties related to aerosols influence on climate (Bindoff et al.

2013). We also need to assume the linear additivity of different forcings, which may not hold on smaller spatial scales like East Asia.

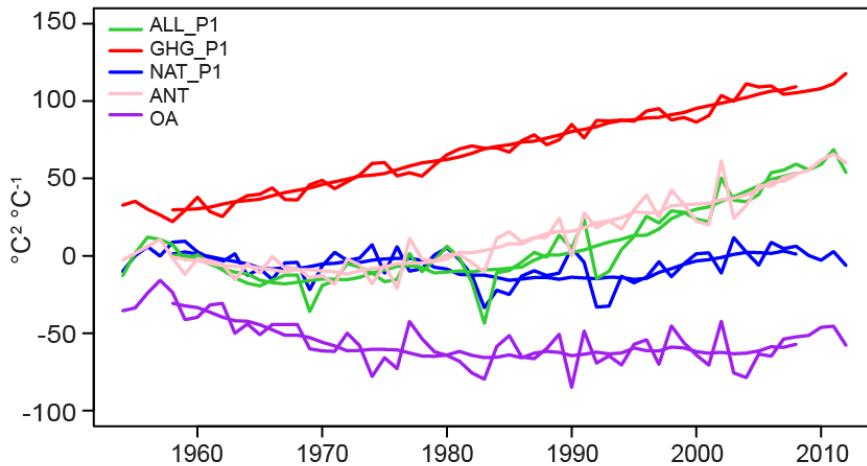


FIG. S14.4. Multimodel mean SST projection time series for 1953–2012 from the ALL_PI, GHG_PI, and NAT_PI experiments. In addition, ANT (anthropogenic forcing response) is estimated from ALL-NAT, and OA (other anthropogenic forcings) is estimated from ALL-NAT-GHG. Light lines represent nine-year moving averages.

TABLE S14.1: List of CMIP5 models used in this study. Numbers represent the number of ensemble members from each model for different experiments. See text for details.

Model	ALL_P0 (1860–1919)	ALL_PI (1954–2013)	GHG_PI (1953–2012)	NAT_PI (1953–2012)
ACCESSI-0	1	1		
ACCESSI-3	1	1		
CanESM2	5	5	5	5
CCSM4	6	6		
CESMI-BGC	1	1		
CESMI-CAM5-1-FV2	1	1		
CMCC-CM	1	1		
CMCC-CMS	1	1		
CSIRO-Mk3-6-0	10	10	5	5
CSIRO-Mk3L-1-2	3	3		
EC-EARTH	1	1		
FGOALS-g2	1	1		
FIO-ESM	3	3		
GISS-E2-H	15	15	5	10
GISS-E2-H-CC	1	1		
GISS-E2-R	16	16	4	9
GISS-E2-R-CC	1	1		
HadCM3	10	10		
HadGEM2-AO	1	1		
HadGEM2-CC	1	1		
HadGEM2-ES	4	4	4	4
IPSL-CM5A-LR	4	4	3	3
IPSL-CM5A-MR	1	1		1
IPSL-CM5B-LR	1	1		
MIROC4h		3		
MIROC5	3	3		
MPI-ESM-LR	3	3		
MPI-ESM-MR	3	3		
MRI-CGCM3	1	1		
NorESM1-M	1	1	1	1
NorESM1-ME	1	1		
Sum Total	102	105	27	38

SI5. THE CONTRIBUTION OF ANTHROPOGENIC FORCING TO THE JAPANESE HEAT WAVES OF 2013

YUKIKO IMADA, HIDEO SHIOGAMA, MASAHIRO WATANABE, MASATO MORI, MASAYOSHI ISHII, AND MASAHADE KIMOTO

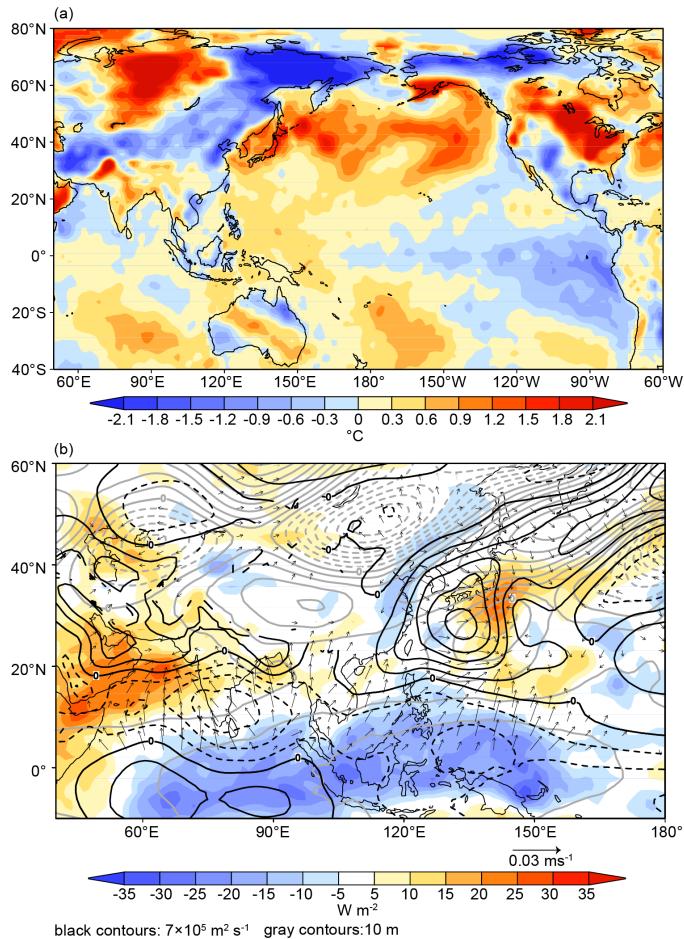


FIG. SI5.1. One example of the simulated extreme cases in the ALL run. (a) SAT (land) and SST (ocean) anomaly from July to August in 2013 ($^{\circ}C$). (b) Same as (a) but for outgoing longwave radiation ($W m^{-2}$, shading), 850-hPa stream function (black contour, $7 \times 10^5 m^2 s^{-1}$ interval), 200-hPa geopotential height (gray contours, 10-m interval), and 200-hPa divergence flow ($m s^{-1}$, arrows).

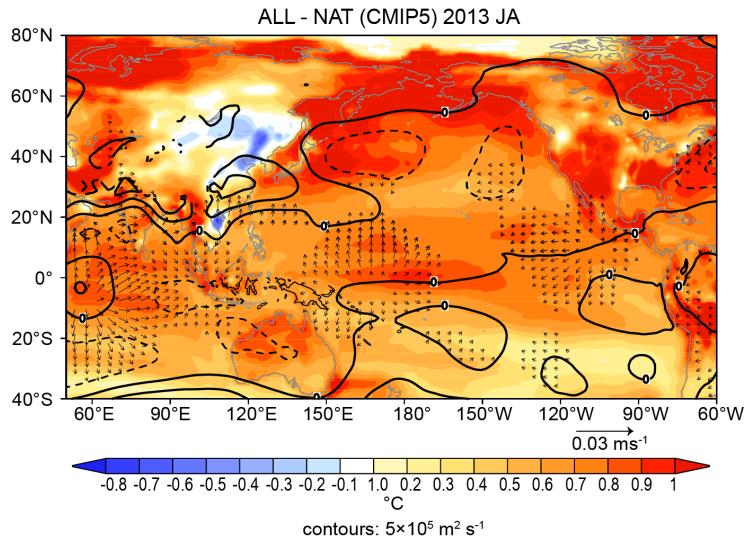


FIG. S15.2. Difference between the ALL and NAT2 runs (ALL minus NAT2): SAT (land) and SST (ocean, $^{\circ}\text{C}$, shading) 850-hPa stream function (black contours, $5 \times 10^5 \text{ m}^2 \text{ s}^{-1}$ interval), and 200-hPa divergence flow (m s^{-1} , arrows).

SI6. UNDERSTANDING A HOT SUMMER IN CENTRAL EASTERN CHINA: SUMMER 2013 IN CONTEXT OF MULTIMODEL TREND ANALYSIS

TIANJUN ZHOU, SHUANGMEI MA, AND LIWEI ZOU

TABLE S16.1: List of 31 CMIP5 models used in the analysis

	Model Name	Historical	Rcp45	Picontrol
1	ACCESS1-0	✓	✓	✓
2	ACCESS1-3	✓	✓	✓
3	BNU-ESM	✓	✓	✓
4	CanESM2	✓	✓	✓
5	CCSM4	✓	✓	✓
6	CESM1-CAM5	✓	✓	✓
7	CMCC-CM	✓	✓	✓
8	CMCC-CMS	✓	✓	✓
9	CNRM-CM5	✓	✓	✓
10	CSIRO-Mk3-6-0	✓	✓	✓
11	EC-EARTH	✓	✓	✓
12	FGOALS-g2	✓	✓	✓
13	GFDL-CM3	✓	✓	✓
14	GFDL-ESM2G	✓	✓	✓
15	GFDL-ESM2M	✓	✓	✓
16	GISS-E2-H	✓	✓	✓
17	GISS-E2-R	✓	✓	✓

CONT. Table S16.1: List of 31 CMIP5 models used in the analysis

	Model Name	Historical	Rcp45	Picontrol
18	HadGEM2-CC	✓	✓	✓
19	HadGEM2-ES	✓	✓	✓
20	IPSL-CM5A-LR	✓	✓	✓
21	IPSL-CM5A-MR	✓	✓	✓
22	MIROC5	✓	✓	✓
23	MIROC-ESM	✓	✓	✓
24	MIROC-ESM-CHEM	✓	✓	✓
25	MPI-ESM-LR	✓	✓	✓
26	MPI-ESM-MR	✓	✓	✓
27	MRI-CGCM3	✓	✓	✓
28	NorESM1-M	✓	✓	✓
29	NorESM1-ME	✓	✓	✓
30	bcc-csm1-1	✓	✓	✓
31	inmcm4	✓	✓	✓

S17. SEVERE PRECIPITATION IN NORTHERN INDIA IN JUNE 2013: CAUSES, HISTORICAL CONTEXT, AND CHANGES IN PROBABILITY

DEEPTI SINGH, DANIEL E. HORTON, MICHAEL TSIANG, MATZ HAUGEN, MOETASIM ASHFAQ, RUI MEI, DEEKSHA RASTOGI, NATHANIEL C. JOHNSON, ALLISON CHARLAND, BALA RAJARATNAM, AND NOAH S. DIFFENBAUGH

Datasets. For the precipitation analysis, we use the $1^\circ \times 1^\circ$ gridded daily dataset from the Indian Meteorological Department (IMD; Rajeevan et al. 2006, 2010). This dataset is developed from approximately 2140 rain gauge stations over India, and it has been extensively used in literature to study characteristics of the Indian Monsoon (Dash et al. 2009; Ghosh et al. 2012; Singh 2013). As shown in Fig. 1a of Rajeevan et al. (2006), all stations included in the development of this dataset have at least 90% data availability within the observational period.

Daily atmospheric variables—including geopotential heights, atmospheric winds, and specific humidity—are obtained from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis (R1) datasets at $2.5^\circ \times 2.5^\circ$ spatial resolution (Kalnay et al. 1996). In order to match the period of record of the IMD precipitation dataset, we analyze the 1951–2013 period in the reanalysis.

Self-organizing maps methodology. The self-organizing map (SOM) is a neural network-based cluster analysis, similar to k-means clustering, that partitions a high-dimensional dataset into a smaller number of representative clusters (Kohonen 2001). In contrast with conven-

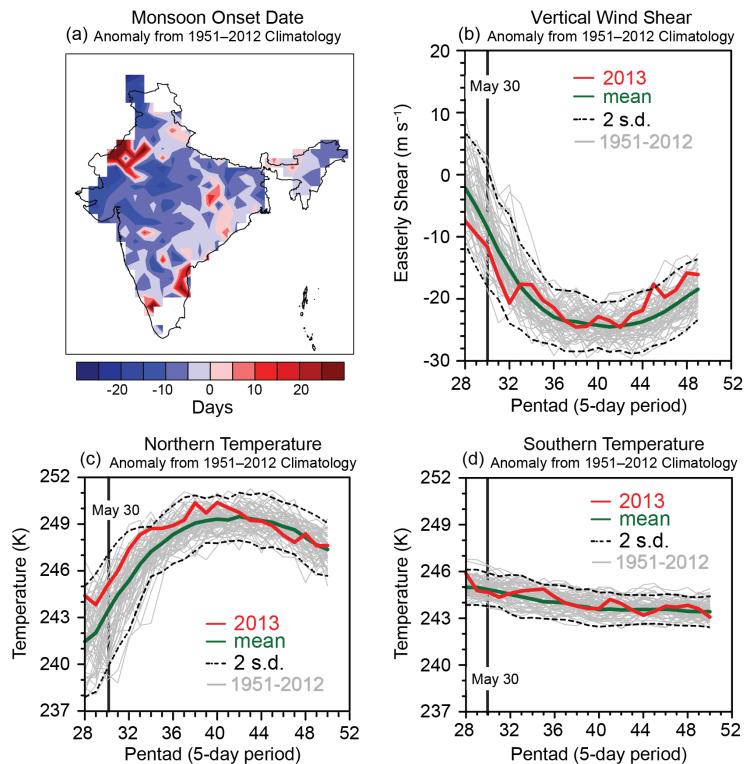


FIG. S17.1. June 2013 monsoon dynamics. (a) June 2013 monsoon onset anomalies from 1951–2012 climatology at each grid-cell. The onset date at each grid-cell is defined as the 5-day mean precipitation exceeding the January mean precipitation of that year and 5 mm day^{-1} . Negative anomalies over most of central and northwestern India suggest an early monsoon onset in June 2013. **(b)** June 2013 vertical easterly zonal wind shear. The easterly zonal wind shear is defined as the difference in winds between the upper (200-mb) and lower (850-mb) atmosphere averaged over the domain ($0^\circ\text{--}30^\circ\text{N}$, $50^\circ\text{--}90^\circ\text{E}$). The vertical wind shear was $\sim 2\sigma$ above the 1951–2012 climatological mean prior to the June 2013 event. **(c,d)** Zonally averaged ($52^\circ\text{--}85^\circ\text{E}$) tropospheric temperature (200–500-mb) above the land at 30°N (c) and the ocean at the equator (d). The temperature over land (“northern temperature”) also exceeded the 1951–2012 climatological mean by 2σ .

tional cluster analysis, these SOM clusters, each of which is associated with a component called a node or neuron, become topologically ordered on a lower-dimensional (typically two-dimensional) lattice so that similar clusters are located close together in the lattice and dissimilar clusters are located farther apart. This topological ordering occurs through the use of a neighborhood function, which acts like a kernel density smoother among a neighborhood of neurons within this low-dimensional lattice. As a result, neighboring neurons within this lattice influence each other to produce smoothly varying clusters that

represent the multidimensional distribution function of the data used to construct the SOM.

In the present study, we perform two separate SOM analysis to categorize daily June 1951–2013 upper-level (200 mb) and lower-level (850 mb) geopotential height anomalies in the domain centered over northern India (0°–60°N, 40°–120°E). In each SOM analysis, the daily geopotential height anomaly field is treated as an M -dimensional vector, where M is the number of grid points. The user specifies the number of clusters, K , and the final clustering is determined through an iterative procedure that approximately

minimizes the Euclidean distance between the daily geopotential height anomaly fields in M -dimensional space and their best-matching SOM patterns. During this iterative procedure, the SOM patterns also “self-organize” into a topologically ordered two-dimensional lattice or grid, as described above. Each daily height field is assigned to a best-matching SOM pattern on the basis of minimum Euclidean distance. Additional details of the SOM methodology can be found in the appendix of Johnson et al. (2008). See also Hewitson and Crane (2002) and Liu et al. (2006) for additional discussion of SOM applications in climate science, and see Chattopadhyay et al. (2008) and Borah et al. (2013) for SOM applications to Indian monsoon intraseasonal variability. Reusch et al. (2005) and Liu et al. (2006) also compare SOM analysis with empirical orthogonal function analysis for the purpose of pattern extraction, demonstrating several advantages of SOM analysis over the more conventional approach.

As mentioned above, the user must specify the choice of K prior to the iterative procedure that determines the SOM clusters. Although there are quite a few approaches for determining an appropriate value of K , there is no universally accepted method for determining the optimal K . In this study, we make the subjective choice of $K = 35$ (i.e., a SOM organized in a two-dimensional lattice with five rows and seven columns), which is a high enough value to resolve regional pattern detail

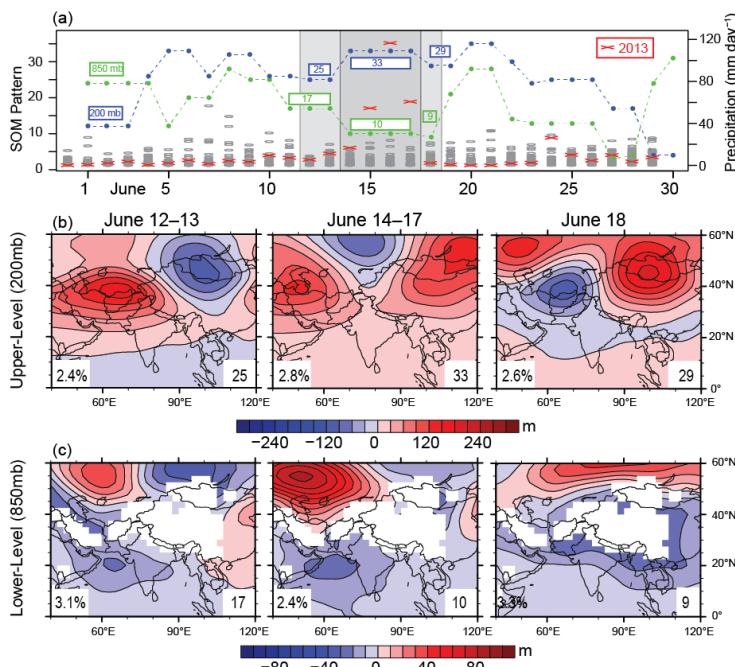


FIG. S17.2. Temporal progression of upper- and lower-level synoptic patterns. We use self-organizing map (SOM) cluster analysis to identify upper- and lower-level atmospheric patterns in June 2013, track their temporal progression, and assess their historical occurrence and co-occurrence (see text). (a) Time-series of the upper-level (blue) and lower-level (green) self-organizing map (SOM) patterns for each day of June 2013. Patterns relevant to the flooding event are highlighted in gray windows in the time-series panel, and their spatial patterns are displayed in panels (b) and (c). SOM pattern numbers display in the lower right of the maps. Daily precipitation accumulation over the flood region (Fig. 1a in the main report, white box) is indicated in the secondary y-axis of (a), with June 2013 values shown in red X's and each year in 1951–2012 shown in gray ovals. The relative frequency of occurrence of each SOM pattern to all historic June days is indicated in the lower left of each SOM map (b,c). From 10 June to 17 June, blocking patterns persisted in the upper atmosphere (200-mb, patterns 25 & 33), while low-pressure troughing strengthened in the lower atmosphere over the flood region (850-mb, patterns 10 & 17). On 18 June, the system lost upper-level support and weakened. 14–17 June 2013 was the first time upper-level pattern 33 co-occurred with lower-level pattern 10 during the month of June within the 1951–2013 historical record.

but small enough to maintain interpretability. We test that our results are robust with different choices for K and that all 35 SOM patterns are statistically distinguishable. For the test of statistical distinguishability, we follow the approach of Johnson (2013). This test is administered by evaluating whether or not all $K(K-1)/2$ SOM pattern pairs are statistically distinguishable on the basis of a field significance test, which in this case is based on the determination of the “false discovery rate” (Benjamini and Hochberg 1995; Wilks 2006). Additional details of the test are given in Johnson (2013).

In the present application, we add an additional step to account for the strong autocorrelation evident in daily data. Because a SOM pattern typically persists for several consecutive days, the number of daily occurrences of each SOM pattern would substantially overestimate the number of temporal degrees of freedom in the local t -tests that determine whether the SOM pattern composite anomalies are significantly different from each other. To correct for this potential source of bias, we perform the local t tests for subsets of the daily geopotential height fields assigned to each SOM pattern. Because the time scale of atmospheric teleconnection patterns is on the order of one to two weeks (e.g., Feldstein 2000), we set the condition that each geopotential height field assigned to the SOM pattern within a subset must be separated by all other geopotential height fields within the subset by at least 15 days. If this separation criterion is not met, then we only keep the daily field with the highest pattern amplitude, where the amplitude is defined as the projection of the daily geopotential height field onto the assigned SOM pattern.

The results of this distinguishability test reveal that all 35 SOM patterns in the 200-mb and 850-mb geopotential height SOMs are statistically distinguishable from each other at the 5% level. Furthermore, all SOM pattern pairs but one (SOM patterns 2 and 8 in the 850-mb geopotential height SOM) are statistically distinguishable from each other at the 1% level. These results suggest that the choice of $K = 35$ is reasonable in that this value of K is not too high such that the SOM patterns become statistically indistinguishable from each other. Furthermore, we also have performed the same analysis for $K = 20$ and found that the main features of the 20-pattern SOMs are quite similar to the main features of the 35-pattern SOMs. Therefore, these evaluations support the robustness of the SOM analysis in this study.

From these 35 distinct nodes, we identify the upper- and lower-level atmospheric patterns in June 2013 to track their temporal progression, and we

assess their historical occurrence and co-occurrence (Fig. 17.1g,h in the main report; Fig. S17.2). Additionally, we determine typical precipitation patterns associated with each geopotential height SOM pattern (e.g., Fig. 17.1i,j in the main report and Fig. S17.2) by creating composites of precipitation from each June day in 1951–2013 on which the atmospheric patterns best matched the corresponding node. This was performed on the basis of assignments of each day to a particular SOM pattern.

Determining a parametric fit for the observed distribution. The event being analyzed lies substantially outside the range of the observed distribution. Therefore, to quantify the probability of occurrence of such an event, we test different parametric distributions (gamma, weibull, lognormal, and pareto) to find a suitable fit for the observations (Fig. S17.3a). We use the maximum log-likelihood estimator to precisely determine the parameters of these potential distributions. Through evaluating the RMS errors (see Table S17.1) between the observed and fitted distributions, we conclude that the Pareto-III distribution most closely represents the observed distribution function.

Probability quantification. We use spatially averaged cumulative June precipitation from 1951 to 2013 over the flood-affected region (Fig. 17.1a in the main report) to generate an observed distribution to examine the probability of the June 2013 event. After fitting the Pareto-III distribution to the observed cumulative June precipitation, we determine the percentile or probability (p) of a specific event magnitude (Pr), and thereby quantify the return period ($T = 1/p$) of the event in the context of the observed climate. We find that the total June 2013 precipitation magnitude exceeds the 99.1th percentile (Fig. 17.1a in the main report) and has a return period of 111 years (Fig. 17.2a in the main report).

TABLE S17.1. Parametric fits of the observed June precipitation distribution.	
Parametric Model	RMS Error
Gamma	0.029
Weibull	0.049
Lognormal	0.021
Pareto	0.018
Root mean square error between fitted and empirical cumulative distribution functions of the observed June precipitation.	

We use the Coupled Model Intercomparison Project phase 5 (CMIP5) historical (20C) and preindustrial (PI) simulations (Table S17.2, S17.3) to quantify the influence of observed anthropogenic forcing on the likelihood of the June 2013 total cumulative precipitation (Fig. 17.2b–d in the main report). The method is illustrated in Fig. S17.3c. After shifting the 20C and PI distributions by the difference between the observed and 20C means, we use the Kolmogorov–Smirnov (K–S) goodness of fit test to measure

TABLE S17.2. Number of years in the historical (20C) and pre-industrial (PI) simulations of select models, and root mean square error between fitted and empirical cumulative distribution functions of model June precipitation.

CMIP5 Model	“20C” years	“PI” years	RMS Error
CNRM-CM5	550	350	0.017
ACCESSI-3	165	500	0.015
ACCESSI-0	110	500	0.025
HadGEM2-CC	147	240	0.019
HadGEM2-ES	275	577	0.016
IPSL-CM5B-LR	55	300	0.02

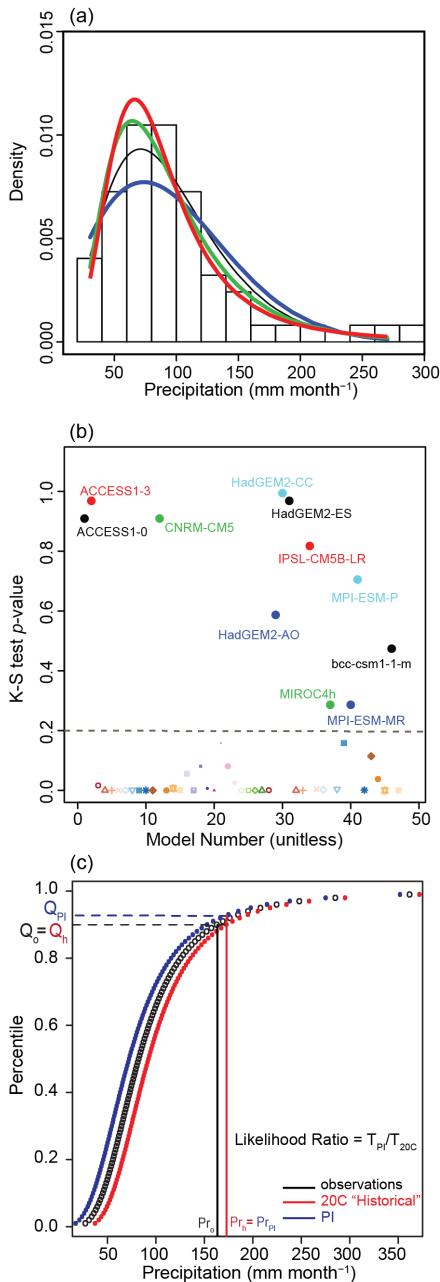


FIG. S17.3. Statistical framework for probability quantification. (a) Results from fitting the observed cumulative June precipitation (1951–2012) with 4 parametric models. The Pareto-III model shows the lowest root mean square error between the empirical and model cumulative distribution functions (shown in Table S17.1). (b) p-values of the Kolmogorov–Smirnov (“K–S”) goodness of fit test measuring the closeness of CMIP5-simulated historical (20C) distributions to the observed distribution. (See Table S17.2 for model names). Higher p-values indicate models that more closely simulate the observed distribution, including the tails. (c) Theoretical methodology for quantifying the probability of a specific event in the 20C and preindustrial PI climates. This involves determining the magnitude of the event (Pr_h) in the 20C distribution corresponding to the percentile of the observed event (Q_o = Q_h). The ratio (PI/20C) of the return periods is calculated by estimating the percentile (or probability) of this simulated event (Pr_i = Pr_h) in the pre-industrial (T_{PI}) and historical (T_{20C}) distributions.

the closeness of CMIP5 simulated historical (20C) distributions to the observed distribution. [Refer to Sperber et al. (2013) for an evaluation of the skill of these models to simulate other characteristics of Indian monsoon rainfall, circulation indices, and their teleconnections.] Because the simulated change in likelihood of extremes can be heavily influenced by biases in the simulated distribution, we restrict our analysis to 11 models whose K–S value exceeds 0.2 (A1; see Fig. S17.3b), ensuring a comparatively good fit of the overall distribution, including in the tails. We then fit the select A1 CMIP5 models with Pareto-III distributions.

We determine the magnitude of the 99.1th percentile event in the 20C distribution as representative of the simulated event magnitude (follow the vertical red line in Fig. S17.3c). By definition, the event has a return period of 111 years in the 20C distribution. Next, we find the corresponding percentile—and hence return period—of this 20C magnitude event in the preindustrial simulations (follow the horizontal blue line in Fig. S17.3c).

By comparing the likelihood of the event in the preindustrial climate to the historical climate, the ratio of the return periods (T_{PI}/T_{20C}) provides an estimate of the influence of 20C forcings on the probability of the June 2013 heavy precipitation.

However, the limited sample of precipitation creates uncertainty in the true fit of the Pareto-III distribution to the observed and CMIP5 simulated precipitation data, thereby introducing uncertainty in the PI/20C return period ratio. We therefore use a bootstrapping approach to quantify this uncertainty and provide confidence estimates for this ratio. We bootstrap the observed, 20C, and PI precipitation data

to generate many distributions to capture the uncertainty around the true parameters of the Pareto-III fits and thus the uncertainty in the return periods. By applying the probability quantification methodology described above to all these bootstrapped distributions, we generate a distribution of the PI/20C return period ratios, which are represented in box plots in Fig. 17.2d in the main report.

TABLE S17.3. Names of the 11 CMIP5 models that meet the K-S test criterion.

Model	Full Model name
CNRM-CM5 C	Centre National de Recherches Meteorologiques - Coupled Global Climate Model version 5
HadGEM2-ES	Hadley Global Environment Model 2 - Earth System
ACCESS1-0	Australian Community Climate and Earth-System Simulator (ACCESS) version 1.0
ACCESS1-3	Australian Community Climate and Earth-System Simulator (ACCESS) version 1.3
BCC-CSM1-I B	Beijing Climate Center - Climate System Model version 1.1
MIROC4h	Model for Interdisciplinary Research on Climate, version 4 High Resolution
MPI-ESM-MR	Max Planck Institute - Earth System Model - Medium Resolution
HadGEM2-CC	Hadley Global Environment Model 2 - Carbon Cycle
HadGEM2-AO	Hadley Global Environment Model 2 - Atmosphere Only
MPI-ESM-P	Max Planck Institute - Earth System Model - Low Resolution and Paleo mode
IPSL-CM5B-LR	Institut Pierre-Simon Laplace - Climate Model version 5B - Low Resolution

SI8. THE 2013 HOT, DRY, SUMMER IN WESTERN EUROPE

BUWEN DONG, ROWAN SUTTON, AND LEN SHAFFREY

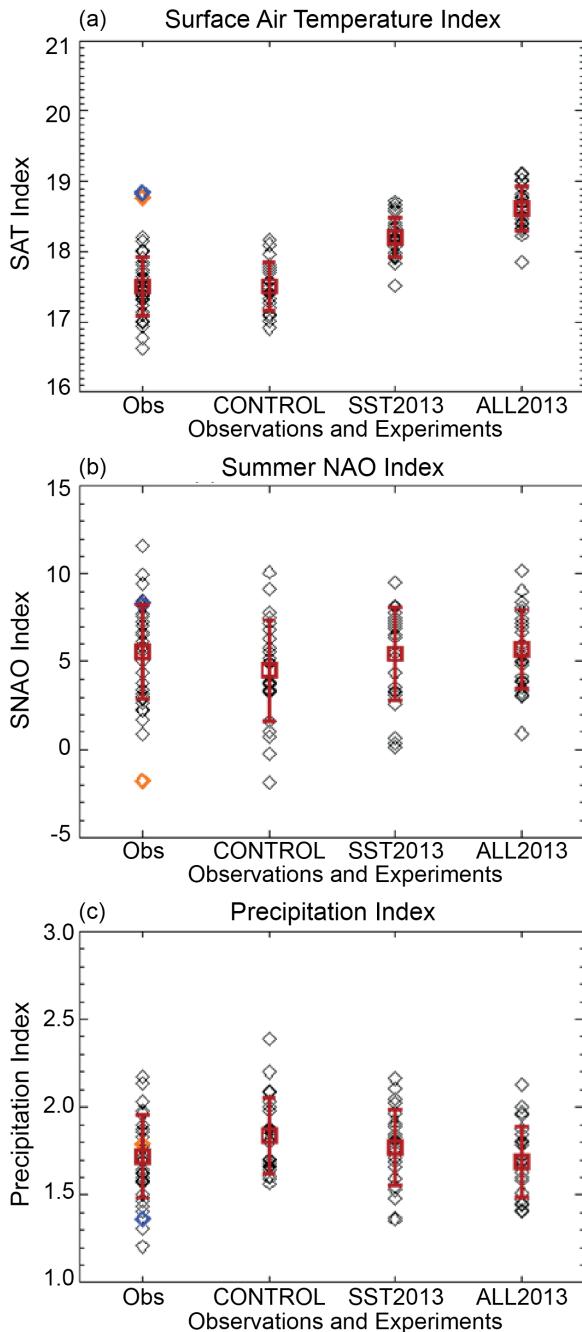


FIG. SI8.1. (a) SAT ($^{\circ}\text{C}$), (b) summer NAO (hPa), and (c) precipitation (mm day^{-1}) indices for observations and model experiments. SAT index is area averaged SAT over region ($35^{\circ}\text{--}75^{\circ}\text{N}$, $10^{\circ}\text{W}\text{--}40^{\circ}\text{E}$, land only; black box in Fig.18.1c in the main report). The SNAO index is defined as the difference of the area mean SLP between two regions around the British Isles ($45^{\circ}\text{--}60^{\circ}\text{N}$, $30^{\circ}\text{W}\text{--}10^{\circ}\text{E}$) and over Greenland ($65^{\circ}\text{--}80^{\circ}\text{N}$, $60^{\circ}\text{--}20^{\circ}\text{W}$; red and blue boxes in Fig.18.1a). Precipitation index is area averaged precipitation over region ($35^{\circ}\text{--}60^{\circ}\text{N}$, $10^{\circ}\text{W}\text{--}20^{\circ}\text{E}$, land only; black box in Fig.18.1b). All black diamonds in observations are for years from 1964 to 1993 with red diamonds for 2012 and blue diamonds for 2013. Red squares and lines are the mean and mean \pm sigma ranges where sigma is the corresponding standard deviation.

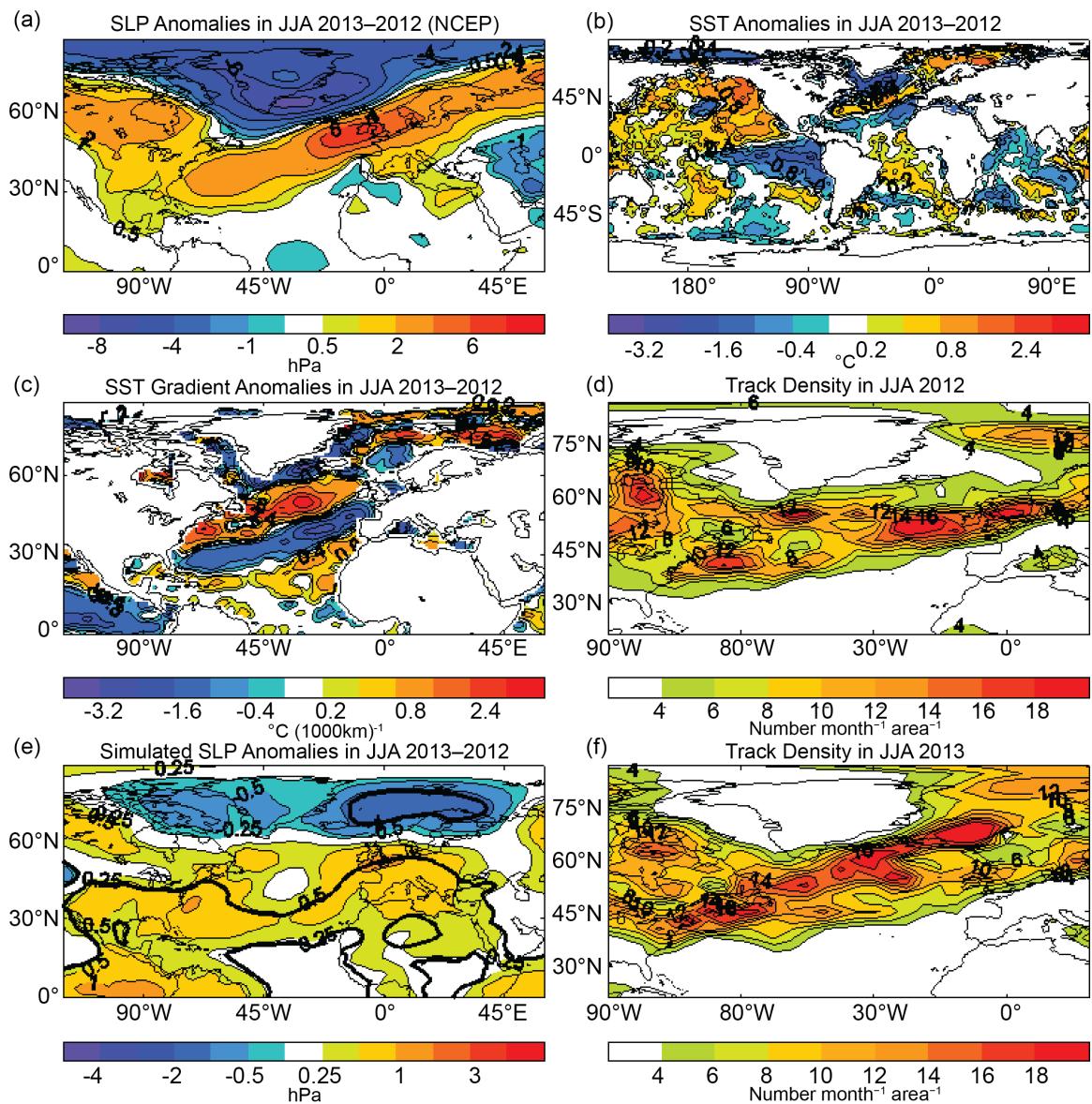


FIG. S18.2. Anomalies in JJA between 2013 and 2012. (a) SLP (hPa), (b) SST ($^{\circ}\text{C}$), and (c) SST gradient [$^{\circ}\text{C} (1000\text{ km})^{-1}$] in observations. (e) Simulated SLP difference between 2013 and 2012 from the changes in SST and SIE. The experiment of 2012 was documented in Dong et al. (2013). (d) and (f) are 2012 and 2013 cyclone track density. Track density is in unit of numbers per month per unit area, where the unit area is equivalent to a 5° spherical cap (about 10^6 km^2). Note that this climatological period is dominated by cold AMO conditions and is the period used for the climatological model simulations. Thick lines in (e) highlight regions where the differences are statistically significant at the 90% confidence level using a two-tailed Student *t*-test.

SI9. CONTRIBUTION OF ATMOSPHERIC CIRCULATION TO WET SOUTHERN EUROPEAN WINTER OF 2013

PASCAL YIOU AND JULIEN CATTIAUX

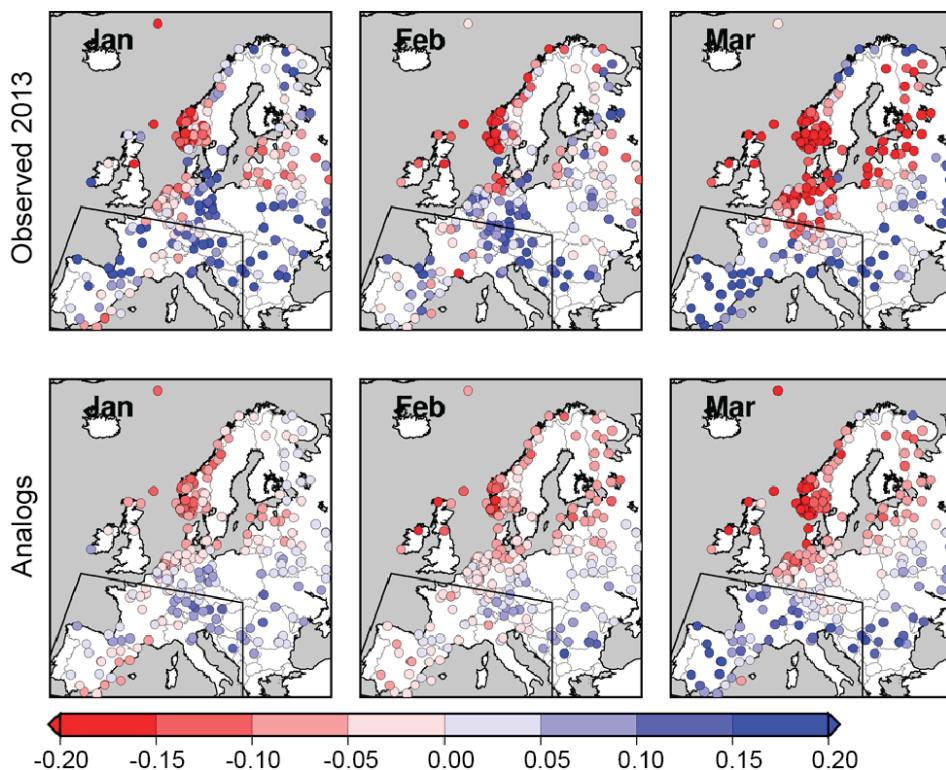


FIG. S19.1. Anomalies of precipitation frequencies over Europe (units are fractions of the number of rainy days over the number of days in a season) for three winter months in 2013 (January to March). The colored points represent the 510 ECA&D stations we retained. The polygon outlines the region over which the averages are computed (10°W – 20°E ; 35° – 49°N). The upper panels represent observed precipitation frequency anomalies; the lower panels represent the median precipitation frequency anomalies obtained from 20 analogues of circulation.

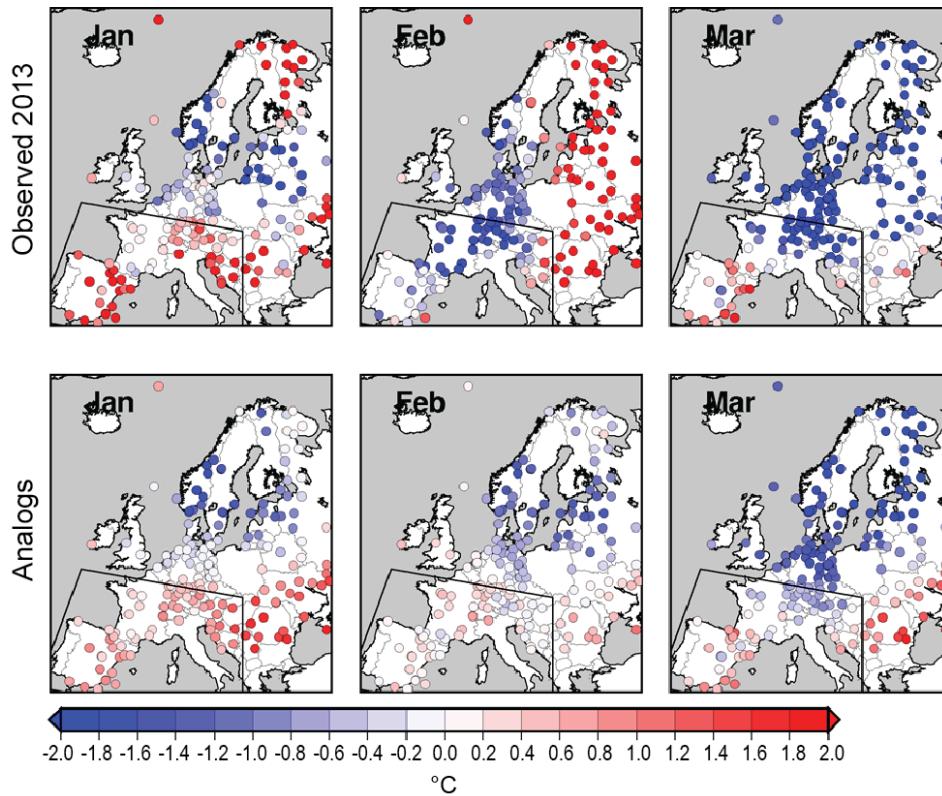


FIG. S19.2. Anomalies of temperature over Europe ($^{\circ}\text{C}$) for three winter months in 2013 (January–March). The colored points represent the 329 ECA&D stations we retained for temperature. The polygon outlines the region over which the averages are computed (10°W – 20°E ; 35° – 49°N). The upper panels represent observed temperature anomalies; the lower panels represent the median temperature anomalies obtained from 20 analogues of circulation.

S20. THE HEAVY PRECIPITATION EVENT OF MAY–JUNE 2013 IN THE UPPER DANUBE AND ELBE BASINS

NATHALIE SCHALLER, FRIEDERIKE E. L. OTTO, GEERT JAN VAN OLDENBORGH, NEIL R. MASSEY,
SARAH SPARROW, AND MYLES R. ALLEN

Modeling experiments setup. For the All Forcing simulations, the atmosphere-only global model is driven by observed SSTs and sea ice fraction (SIF) from the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) dataset (Stark et al. 2007). SSTs and SIF for the Natural runs, however, need to be estimated. Sea ice extent has changed due to anthropogenic climate change; therefore, a composite file of Arctic and Antarctic sea ice was produced. The years with the largest sea ice extents during the winter months for the Arctic and Antarctic, respectively, have been identified. The SIF file used for the Natural ensembles, therefore, contains 1986 SIF values in the Northern Hemisphere and 2008 SIF values in the Southern Hemisphere.

Two different methods are used to estimate the anthropogenic effect on SSTs. The basic idea is to calculate the difference in SSTs for each grid point for each month (referred to as “delta SSTs”) between a decade that experiences anthropogenic forcing and a decade that did not. These delta SSTs are then removed from the observed May–June 2013 SSTs used in the All Forcings runs. If available, the historical scenario (simulations run with all observed atmospheric forcings from 1850 to 2005) and the historicalNat scenario (simulations run with only the natural observed atmospheric forcings from 1850 to 2005) have been used (Taylor et al. 2012).

TABLE S20.1. List of CMIP5 models used to calculate the first set of Delta SSTs.

CMIP5 model	Delta SSTs calculated as difference between
bcc-csm1-l	1997–2007 historical – 1997–2007 historicalNat
CNRM-CM5	1990–2000 historical – 1855–65 historicalNat
inmcm4	1990–2000 historical – 1855–65 historicalNat
IPSL-CM5A-LR	1990–2000 historical – 1855–65 historicalNat
HadCM3	1990–2000 historical – 1864–74 historicalNat
HadGEM2-ES	2002–12 historical combined with RCP8.5 – 2002–12 historicalNat
MPI-ESM-LR	2002–12 historical combined with RCP8.5 – 1855–65 historicalNat
MPI-ESM-P	1990–2000 historical – 1855–65 historicalNat
GISS-E2-H	1990–2000 historical – 1990–2000 historicalNat
GISS-E2-R	1997–2007 historical combined with RCP8.5 – 1997–2007 historicalNat
NorESM1-M	1997–2007 historical combined with RCP8.5 – 1997–2007 historicalNat
HadGEM2-AO	2002–12 historical combined with RCP8.5 – 1865–75 historicalNat
GFDL-ESM2M	1990–2000 historical – 1990–2000 historicalNat
CESM1-CAM5-I-FV2	1990–2000 historical – 1990–2000 historicalNat

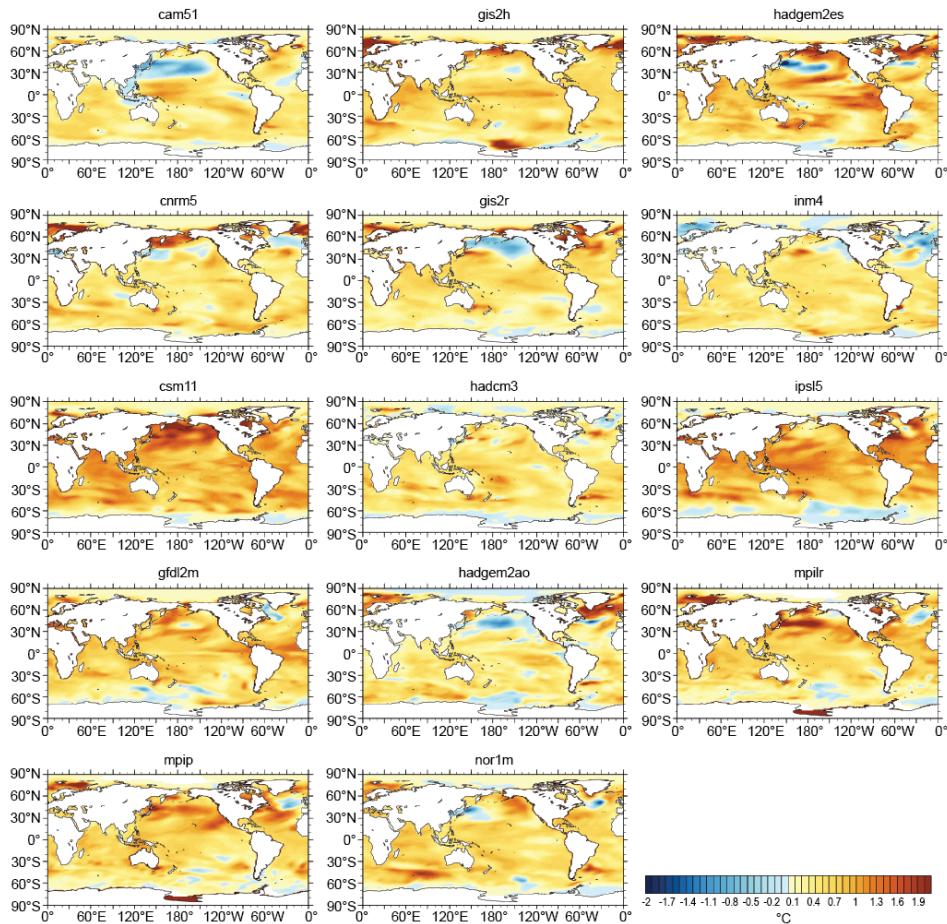


FIG. S20.1. Estimated SST (°C) response pattern to anthropogenic forcing averaged for May/June climatologies for the first set of natural runs (see text for explanations).

The first set of Natural ensembles removes anthropogenic SST patterns estimated from the Coupled Model Intercomparison Project phase 5 (CMIP5) archive before it was completed, which is why not all delta SSTs have been calculated in the same way. This first set of 14 delta SSTs has, in addition, been calculated for one ensemble member only, and it is described in Table S20.1. An example of these SST patterns is shown in Fig. S20.1.

A second set of 11 Delta SSTs was calculated after the CMIP5 archive completed. For this set, only models having run the historical and historicalNat experiments and having three or more ensemble members for both experiments were chosen. The obtained delta SSTs were then averaged over the available ensemble members to get rid of some of the internal variability. These are described in Table S20.2. An example of these SST patterns is shown in Fig. S20.2.

TABLE S20.2. List of CMIP5 models used to calculate the second set of Delta SSTs.

CMIP5 model	Number of ensemble members	Delta SSTs calculated as difference between
CanESM2	5 historical, 5 historicalNat	1996–2005 historical – 1996–2005 historicalNat
CCSM4	6 historical, 4 historicalNat	1996–2005 historical – 1996–2005 historicalNat
CNRM-CM5	10 historical, 6 historicalNat	1996–2005 historical – 1996–2005 historicalNat
CSIRO-Mk3-6-0	10 historical, 5 historicalNat	1996–2005 historical – 1996–2005 historicalNat
GFDL-CM3	5 historical, 3 historicalNat	1996–2005 historical – 1996–2005 historicalNat
GISS-E2-H	5 historical, 5 historicalNat	1996–2005 historical – 1996–2005 historicalNat
GISS-E2-R	6 historical, 5 historicalNat	1996–2005 historical – 1996–2005 historicalNat
HadGEM2-ES	4 historical, 4 historicalNat	1996–2005 historical – 1996–2005 historicalNat
IPSL-CM5A-LR	3 historical, 6 historicalNat	1996–2005 historical – 1996–2005 historicalNat
IPSL-CM5A-MR	3 historical, 3 historicalNat	1996–2005 historical – 1996–2005 historicalNat
MIROC-ESM	3 historical, 3 historicalNat	1996–2005 historical – 1996–2005 historicalNat

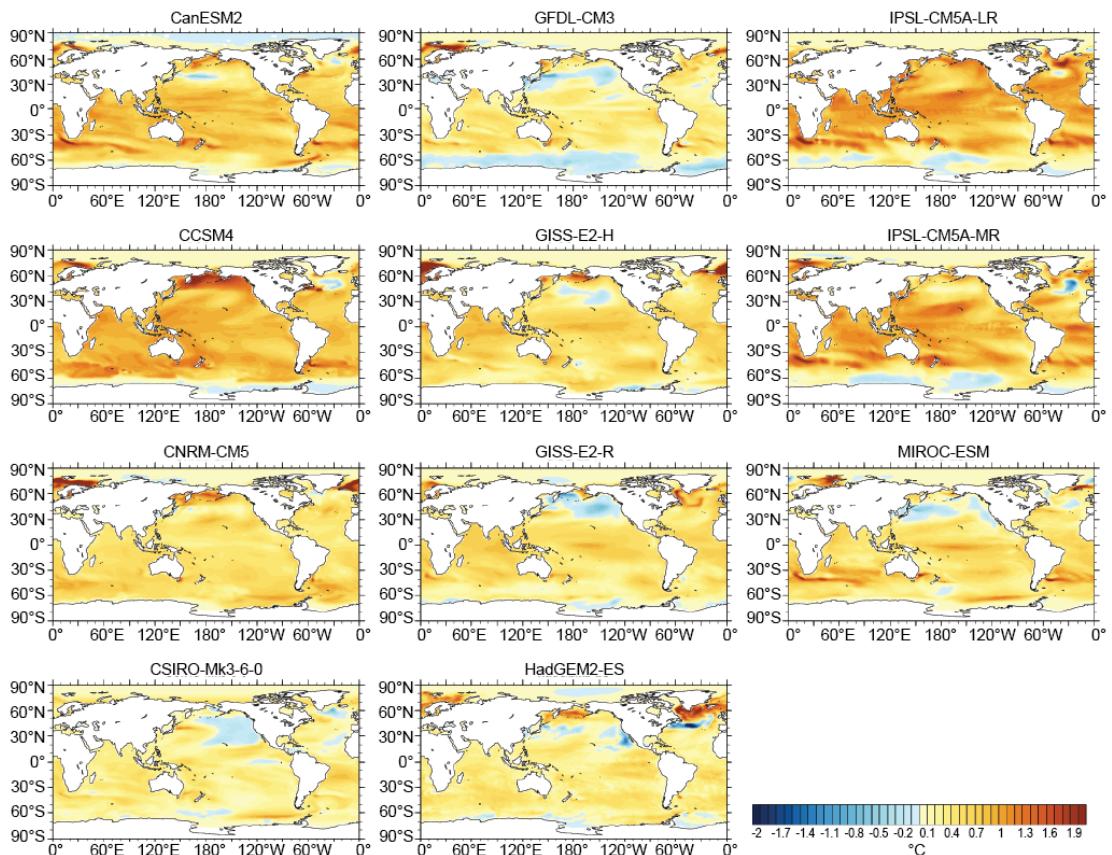


FIG. S20.2. Estimated SST (°C) response pattern to anthropogenic forcing averaged for May/June climatologies for the second set of natural runs (see text for explanations).

S23. THE EFFECT OF ANTHROPOGENIC CLIMATE CHANGE ON THE COLD SPRING OF 2013 IN THE UNITED KINGDOM

NIKOLAOS CHRISTIDIS, PETER A. STOTT, AND ANDREW CIAVARELLA

MJO and stratospheric warming influence on the 2012/13 North Atlantic Oscillation. Weather patterns in the tropical Pacific are known to have an influence on the phase of the North Atlantic Oscillation (NAO). During February and March 2013, the active phase of the MJO was particularly strong, which very likely created favourable conditions for a negative phase of the NAO (Cassou 2008). Moreover, a strong sudden stratospheric warming event in the upper atmosphere over the North Pole in January 2013 led to easterly anomalies over Europe, which also favour negative NAO conditions (Slingo 2013).

The Hadley Centre ACE system. Our Attribution of extreme weather and Climate Events (ACE) system employs version 3A of the Hadley Centre Global Environmental Model (HadGEM3-A), the atmosphere-only component of HadGEM3 (Hewitt et al. 2011). The model is run at N96 horizontal resolution and includes 38 vertical levels. The ensembles of the ALL and NAT experiments are generated using the “perturbed physics” approach, which introduces random perturbations to represent the uncertainty in physical parameters (Murphy et al. 2004). In addition to this approach, wind increments are also added at each time step to account for energy loss due to numerical smoothing and unrepresented sources of kinetic energy (Tennant et al. 2011). The radiative forcings during the analysis period in our simulations come from estimates of historical changes in the forcings that had been previously prescribed in the coupled model experiments described in Stott et al. (2006). The initial conditions for our experiments come from the European Centre for Medium-Range Weather

Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim; Dee et al. 2011).

Boundary conditions. The SSTs in NAT simulations come from the difference between observations and monthly estimates of the anthropogenic change in SST (denoted by Δ SST). In simulations of the first NAT ensemble, we compute from HadISST following the approach introduced in Christidis and Stott (2014). A linear fit is applied to the monthly mean SST time series during 1870–2012 on each sea grid point of the dataset, and is then estimated as the product of the slope and the length of the time series. This estimate is subsequently subtracted from SST observations spanning the analysis period to produce the boundary conditions of the NAT simulations. Although the estimate includes the effects of natural forcings and multidecadal internal variability, these effects are assumed to be minimal given that the length of the time series is over a century long. Variability on centennial to multi-centennial time scales may also introduce uncertainty in the estimated (Galloway et al. 2013; Zanchettin et al. 2010). The approach was introduced by Christidis and Stott (2014) as an alternative to the from individual models, which were shown to often suffer from large discrepancies. Although the observational is only a simple approximation of the anthropogenic change in the oceanic temperature and may be influenced by nonlinearities related to natural factors, it still provides a useful description of the patterns of change, as can be shown using perfect model experiments (not investigated here, as this is beyond the scope of the present study). While the separation of the

natural and anthropogenic components of the SST change remains a challenge to the climate research community, ACE studies resort to approximations, as the ones presented here, which are clearly identified as a source of uncertainty in the results. A more detailed discussion on the observational can be found in Christidis and Stott (2014).

In the second version of NAT, we make the boundary conditions using seven atmosphere-ocean coupled models (HadGEM2-ES, CNRM-CM5, CSIRO-Mk3-6-0, CanESM2, GISS-E2-H, GISS-E2-R, IPSL-CM5A-LR) that participated in the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change. The models selected here are the ones that provide ensembles of simulations with all historical and natural forcings extended to 2012, which are required to compute the change in oceanic temperature. The is estimated from the difference between temperature fields from simulations with all historical forcings and historical natural forcings only. Each model provides ensembles of at least three simulations for each experiment. After constructing the ensemble mean for each experiment and each model, we compute the mean of all the individual model ensemble means for the two experiments and take their difference. The estimate corresponds to the resulting monthly temperature patterns averaged over the most recent decade simulated in the GCM experiments (2003–12) and is again subtracted from observations of the SST during the analysis period (September 2012–August 2013) to produce the boundary conditions. The use of the multimodel ensemble mean instead of a single model in the estimation of has the advantage that the effect of model biases is in principle reduced, unless of course all models suffer from common biases. Again, given the limitations arising from model biases in the simulated SST (e.g., Shin and Sardeshmukh 2011), the estimated change in the oceanic temperature is identified as a source of uncertainty in our results.

Both the observed and modelled SST estimates show an overall anthropogenic warming of the oceans, but regional details may differ (for details and illustrations see Christidis and Stott 2014). We also find that the model mean generally produces more SST warming than the observations. The annual mean anthropogenic warming in the UK region is estimated to be 0.75°C with HadISST and 0.85°C with the AR5 models. As in previous work (e.g., Pall et al. 2011; Christidis et al. 2013), the sea ice in NAT simulations is estimated using empirical linear relationships to adjust the HadISST data, which describe the dependence of sea ice on SST in each hemisphere.

More specifically, a linear fit was applied to gridpoint data from HadISST that cover a climatological period, and a relationship between sea ice and SST was derived for each hemisphere. This relationship was subsequently applied to the estimated to compute the change in sea ice, and the HadISST sea ice data for the analysis period were adjusted accordingly, making sure the sea ice fraction is limited to vary between 0 and 1.

Model evaluation. The evaluation of the attribution system is an essential part of the analysis, as it establishes that the model is a good tool for event attribution. Unlike forecasting systems, an attribution system ought to be able to reproduce the probabilities of the event, rather than the event itself. Common tests to examine how well the model reproduces the climatological frequency of extreme events are based on comparisons between simulations of the climate during a climatological period and observations and/or reanalysis data. We employ such comparisons to demonstrate whether the model gives a realistic distribution of the spring UK temperature and is, therefore, expected to yield realistic probability estimates in the climatological period. A second tool commonly employed for model evaluation is the reliability diagram, which assesses the predictive skill of the model. If the skill is shown to be good, it means we have confidence that the attribution analysis provides the changing odds of the specific event under consideration.

Figure S23.1 shows model evaluation assessments from comparisons between five HadGEM3-A simulations and observations over the period 1960–2010. Figure S23.1a is a reliability diagram that examines whether the model-derived probability of a cold spring (red line) or of cold spring months (black line) in the United Kingdom is consistent with the observed frequency of cold events in the CRUTEM4 dataset. The proximity of the lines to the diagonal indicates the model has good skill in reproducing cold events, which suggests the presence of predictive factors, such as the oceanic state. Details on the construction and interpretation of reliability diagrams can be found in Christidis et al. (2013).

The model time series of spring temperature in the UK region are found to be consistent with the observed time series from CRUTEM4 (Fig. S23.1b), and the model variability is realistic as inferred by power spectrum analysis (Fig. S23.1d). Finally, a Kolmogorov–Smirnov test indicates that the observed and model distributions of the mean spring temperatures (Fig. S23.1c) are not significantly different (when

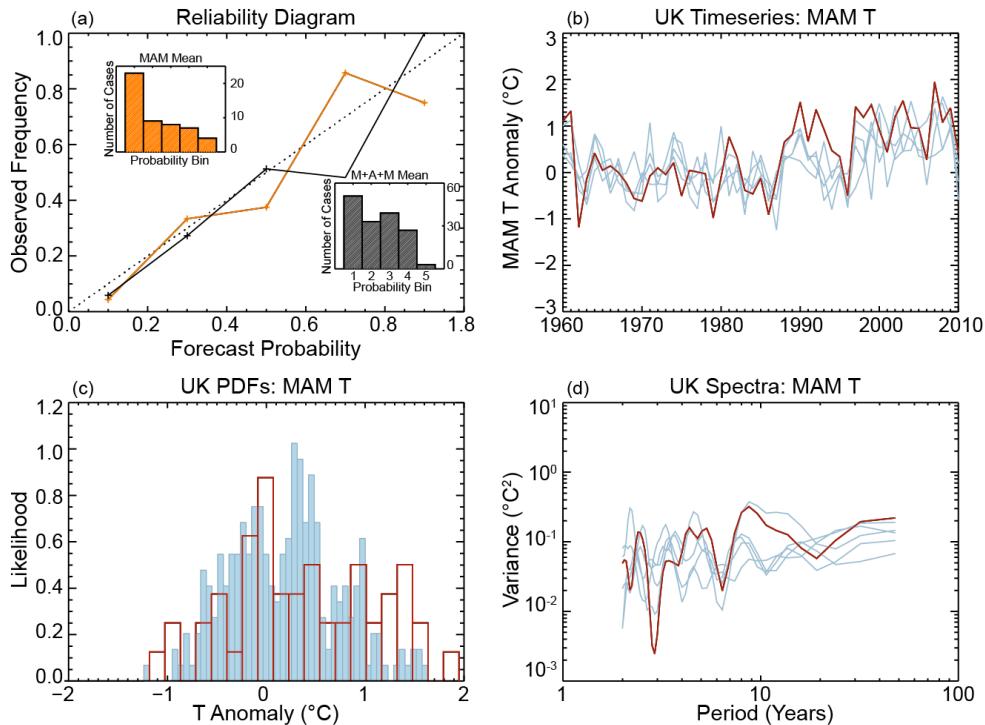


FIG. S23.1. Model evaluation against CRUTEM4 observations. (a) Reliability diagrams for cold springs (in red) and cold spring months (in black) in the UK region (10°E – 5°W , 48° – 60°N). The diagrams are constructed from 51 hindcasts (one for each year during 1960–2010), classified into five probability bins that define the forecast probability. Cold events are characterised by temperatures below the lower tercile of the 1960–2010 climatology. The inset histograms show the number of hindcasts in each probability bin. (b) Spring temperature timeseries in the UK region. (c) Spring temperature probability density functions (PDFs) in the UK region. (d) Power spectra of the spring temperature timeseries. In panels (c)–(d) results plotted in blue correspond to model simulations and results plotted in red to CRUTEM4.

testing at the 5% significance level). On the basis of these assessments we, therefore, expect to obtain realistic estimates of the event probability from our attribution system.

Estimation of probabilities. The probabilities of March–May (MAM) 2013 being cooler than a specific threshold are computed for each ensemble (or for the members of each ensemble that have a specific NAO phase) using the generalized Pareto distribution if the threshold lies at the tails. By randomly resampling the modeled temperatures of each sample 10 000 times, we compute 10 000 estimates of the probability. These 10 000 values provide the best estimate of the probability (50th percentile) and the 5%–95% uncertainty range. Moreover, by applying the same Monte Carlo bootstrap procedure to two different samples, we can obtain estimates of the change in the probability. For example, for the change in probability in cases with negative NAO relative to cases with positive NAO (Fig. 23.2c in the main report), we estimate the two

probabilities from a set of ensemble members with a positive phase and a set of members with a negative phase. We then compute the probability ratio, resample the modelled temperatures of the two sets, and repeat the calculation 10 000 times. Using the resulting estimates of the probability ratio, we can construct the probability density function (PDF) shown in Fig. 23.2c (in the main report).

An analysis of the most negative NAO years. We examine the impact of including only years with the most negative NAO phase (rather than all the years with a negative NAO) in the analysis of the anthropogenic effect. We keep only ensemble members with an NAO index in the lower tercile of all negative index values. Results are shown in Fig. S23.2. We do not find a systematic shift towards higher probabilities of a colder spring, but we do find an increase in the uncertainty in the estimated values, presumably due to the use of smaller samples. The increase in the probability of a cold spring after removing the anthropogenic effect

is in the range of 30–50 (best estimates of the two NAT versions shown in Fig. S23.2c), i.e., the same as in our initial analysis, though the uncertainty in the estimate is higher.

Testing the sensitivity to the definition of the NAO index. Using a spring NAO index based on the leading EOF of MAM sea level pressure anomalies could arguably be an equally, if not more, useful index for our analysis. We test the sensitivity of our results to the NAO index definition by repeating the analysis with the

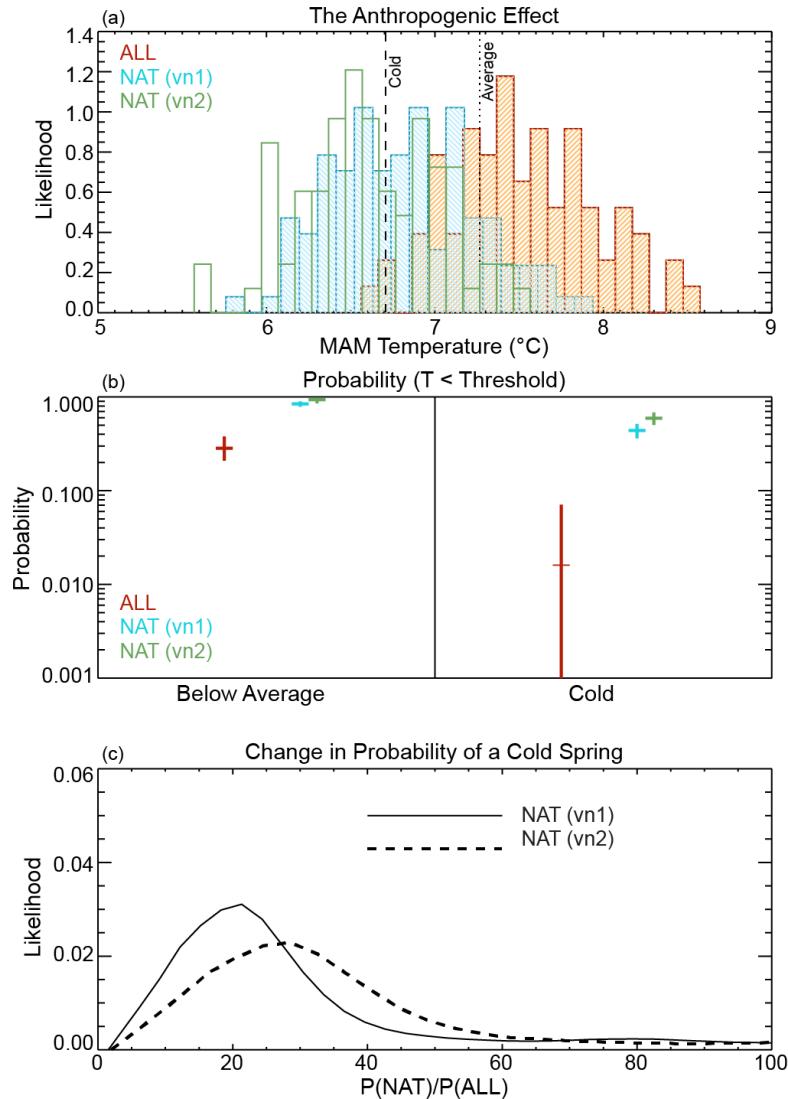


FIG. S23.2. ACE analysis of the anthropogenic influence on the cold UK spring in 2013. The difference between the results plotted here and in Fig. 23.2, in the main report (panels d, e, and f) is that now only ensemble members with the most negative NAO indices have been retained (members with an NAO index in the lower tercile of all the negative index values for each ensemble). (a) Normalized distributions of the regional MAM temperature from the ALL ensemble (red) and the two NAT ensembles (blue and green). The thresholds for cool seasons (temperature below the climatological mean of period 1960–2010) and cold seasons (temperature at least a standard deviation below the climatological mean) are marked by the vertical dotted and dashed lines respectively. (b) Probabilities of a cool and cold spring estimated from the ALL ensemble (red) and the two NAT ensembles (blue and green). The vertical bars around the best estimate (50th percentile of the distribution) correspond to the 5%–95% uncertainty range. (c) Normalized distributions of the change in the odds of a cold spring in 2013 due to the anthropogenic effect computed using the two versions of the NAT ensemble (solid and dashed lines).

alternative MAM NAO index. Figure S23.3a shows a significant relationship between the NAO and spring temperature, as in the case with the December-March DJFM index (Fig. 23.1b in the main report). The relationship, however, between the NAO and the sea ice extent is no longer significant with the new index definition (Fig. S23.3b).

The MAM temperature distributions in 2013 produced with members of the ALL ensemble with positive and negative MAM NAO indices are shown in Fig. S23.3c. The distributions based on the members of the ALL and NAT ensembles with a negative MAM NAO index are shown in Fig. S23.3d. All the PDFs are similar to those constructed based on the DJFM NAO index, albeit, their spread is a little smaller. This does not have much effect on the probabilities of a cold spring with and without human influences shown

in Fig. 23.2e in the main report, and the change in the odds of the event (Fig. 23.2f in the main report) is almost identical. We find that only the probability of a cold spring in positive NAO years (Fig. 23.2b in the main report) is significantly reduced as the corresponding temperature PDF becomes narrower. This means that negative NAO leads to a greater increase in the odds of the event than the conservative estimate of a 10-fold increase obtained with the DJFM index definition.

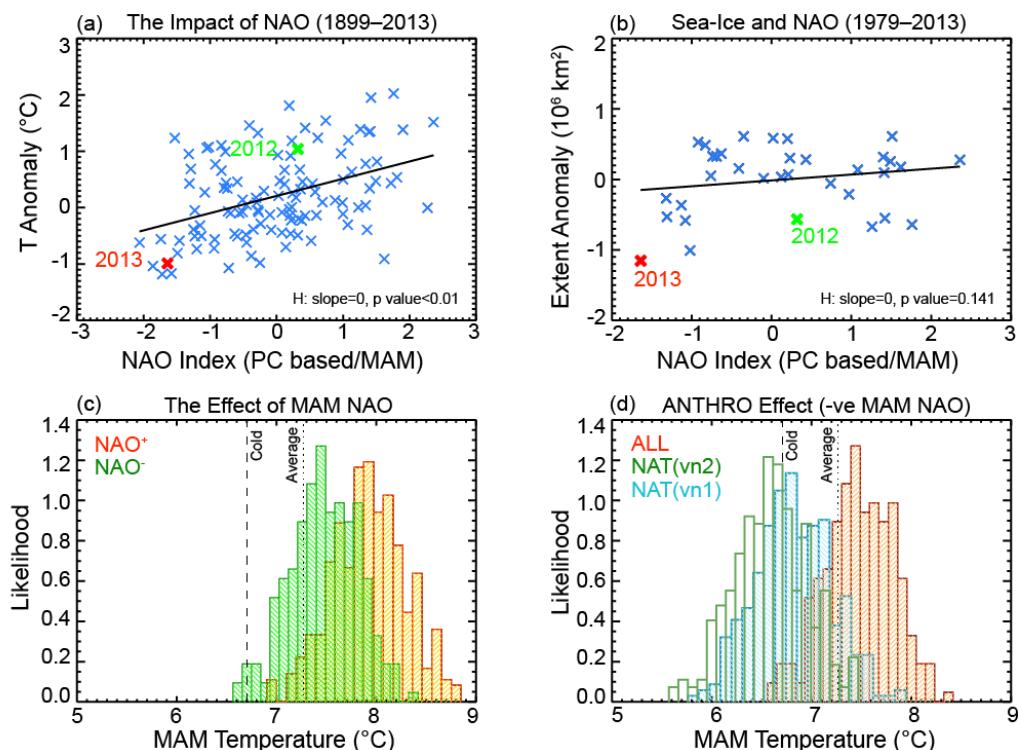


FIG. S23.3. Reproduction of the bottom panels of Fig. 23.1 of the main paper with the MAM NAO index, showing (a) the relationship between the NAO and the spring temperature anomaly in the UK region and (b) the relationship between the NAO and the sea-ice extent anomaly in the Northern Hemisphere. Reproduction of the temperature PDFs shown in Fig. 23.2a and 23.2d of the main paper with the MAM NAO index, showing (c) the temperature distributions from members of the ALL ensemble with positive and negative MAM NAO and (d) the temperature distributions from members of the ALL and NAT ensembles with negative MAM NAO index.

REFERENCES FOR SUPPLEMENTAL MATERIAL

- Adler, R. F., and Coauthors, 2003: The Version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-Present). *J. Hydrometeorol.*, **4**, 1147–1167.
- Becker, A., P. Finger, A. Meyer-Christoffer, B. Rudolf, K. Schamm, U. Schneider, and M. Ziese, 2013: A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901 to present. *Earth Syst. Sci. Data*, **5**, 71–99.
- Benjamini, Y., and Y. Hochberg, 1995: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc.*, **B57**, 289–300.
- Bindoff, N. L., and Coauthors, 2013: Detection and attribution of climate change: From global to regional. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 867–952.
- Borah, N., A. K. Sahai, R. Chattopadhyay, S. Joseph, S. Abhilash, and B. N. Goswami, 2013: A self-organizing map-based ensemble forecast system for extended range prediction of active/break cycles of Indian summer monsoon. *J. Geophys. Res. Atmos.*, **118**, 9022–9034, doi:10.1002/jgrd.50688.
- Bureau of Meteorology, 2013: Australia's warmest September on record. Bureau of Meteorology Special Climate Statement 46, 26 pp. [Available online at <http://www.bom.gov.au/climate/current/statements/scs46.pdf>.]
- Cassou, C., 2008: Intraseasonal interaction between the Madden-Julian Oscillation and the North Atlantic Oscillation. *Nature*, **455**, 523–527.
- Chattopadhyay, R., A. K. Sahai, and B. N. Goswami, 2008: Objective identification of nonlinear convectively coupled phases of monsoon intraseasonal oscillation: Implications for prediction. *J. Atmos. Sci.*, **65**, 1549–1569.
- Chen, S., and D. Cayan, 1994: Low-frequency aspects of the large-scale circulation and West Coast United States temperature/precipitation fluctuations in a simplified general circulation model. *J. Climate*, **7**, 1668–1683.
- Christiansen, B., 2013: Changes in temperature records and extremes: are they statistically significant? *J. Climate*, **26**, 7863–7875.
- Christidis, N., and P. A. Stott, 2014: Change in the odds of warm years and seasons due to anthropogenic influence on the climate. *J. Climate*, **27**, 2607–2621.
- , —, A. Scaife, A. Arribas, G. S. Jones, D. Copsey, J. R. Knight, and W. J. Tennant, 2013: A new HadGEM3-A based system for attribution of weather and climate-related extreme events. *J. Climate*, **26**, 2756–2783.
- Compo, G. P., and P. D. Sardeshmukh, 2010: Removing ENSO-related variations from the climate record. *J. Climate*, **23**, 1597–1978.
- Dash, S. K., M. A. Kulkarni, U. C. Mohanty, and K. Prasad, 2009: Changes in the characteristics of rain events in India. *J. Geophys. Res.*, **114**, D10109, doi:10.1029/2008jd010572.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.
- Dong, B.-W., R. T. Sutton, and T. Woollings, 2013: The extreme European summer 2012 [in “Explaining Extreme Events of 2012 from a Climate Perspective”]. *Bull. Amer. Meteor. Soc.*, **94** (9), S28–S32.
- Feldstein, S. B., 2000: The timescale, power spectra, and climate noise properties of teleconnection patterns. *J. Climate*, **13**, 4430–4440.
- Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866.
- Galloway, J. M., A. Wigston, R.T. Patterson, G.T. Swindles, E. Reinhardt, H.M. Roe, 2013: Climate change and decadal to centennial-scale periodicities recorded in a late Holocene NE Pacific marine record: Examining the role of solar forcing. *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, **386**, 669–689.
- Ghosh, S., D. Das, S.-C. Kao, and A. R. Ganguly, 2012: Lack of uniform trends but increasing spatial variability in observed Indian rainfall extremes. *Nature Climate Change*, **2**, 86–91, doi:10.1038/nclimate1327.
- Hewitson, B. C., and R. G. Crane, 2002: Self-organizing maps: Applications to synoptic climatology. *Climate Res.*, **22**, 13–26.
- Hewitt, H. T., D. Copsey, I. D. Culverwell, C. M. Harris, R. S. R. Hill, A. B. Keen, A. J. McLaren, and E. C. Hunke, 2011: Design and implementation of the infrastructure of HadGEM3: the next-generation Met Office climate modelling system. *Geosci. Model Dev.*, **4**, 223–253.
- Hudson, D., A. G. Marshall, Y. Yin, O. Alves and H. H. Hendon, 2013: Improving intraseasonal prediction with a new ensemble generation strategy. *Mon. Wea. Rev.*, **141**, 4429–4449.
- Hurrell, J., and Coauthors, 2013: The Community Earth System Model: A framework for collaborative research. *Bull. Amer. Meteor. Soc.*, **94**, 1339–1360.

- Johnson, N. C., 2013: How many ENSO flavors can we distinguish? *J. Climate*, **26**, 4816–4827.
- , S. B. Feldstein, and B. Tremblay, 2008: The continuum of Northern Hemisphere teleconnection patterns and a description of the NAO Shift with the use of self-organizing maps. *J. Climate*, **21**, 6354–6371.
- Jones, D. A., W. Wang, and R. Fawcett, 2009: High-quality spatial climate data-sets for Australia. *Aust. Meteor. Ocean. J.*, **58**, 233–248.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Knutson, T. R., F. Zeng, and A. T. Wittenberg, 2013: Multimodel assessment of regional surface temperature trends: CMIP3 and CMIP5 Twentieth Century simulations. *J. Climate*, **26**, 8709–8743.
- Kohonen, T., 2001: *Self-Organizing Maps*. 3rd ed. Springer Series in Information Sciences, Vol. 30, Springer, 501 pp.
- Lewis, S. C., and D. J. Karoly, 2013: Anthropogenic contributions to Australia's record summer temperatures of 2013. *Geophys. Res. Lett.*, **40**, 3705–3709, doi:10.1002/grl.50673.
- Lim, E.-P., H. H. Hendon, D. L. T. Anderson, and A. Charles, 2011: Dynamical, statistical-dynamical, and multimodel ensemble forecasts of Australian spring season rainfall. *Mon. Wea. Rev.*, **139**, 958–975.
- Liu, Y., R. H. Weisberg, and C. N. K. Mooers, 2006: Performance evaluation of the self-organizing map for feature extraction. *J. Geophys. Res.*, **111**, C05018, doi:10.1029/2011GL047658.
- Livezey, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination with Monte Carlo techniques. *Mon. Wea. Rev.*, **111**, 46–59.
- Mastrandrea, M. D., K. J. Mach, G.-K. Plattner, O. Edenhofer, T. F. Stocker, C. B. Field, K. L. Ebi, and P. R. Matschoss, 2011: The IPCC AR5 guidance note on consistent treatment of uncertainties: A common approach across the working groups. *Climatic Change*, **108**, 675–691.
- Mitchell, T., and W. Blier, 1997: The variability of wintertime precipitation in the region of California. *J. Climate*, **10**, 2261–2276.
- Mo, R., and D. M. Straus, 2002: Statistical–dynamical seasonal prediction based on principal component regression of GCM ensemble integrations. *Mon. Wea. Rev.*, **130**, 2167–2187.
- Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset. *J. Geophys. Res.*, **117**, D08101, doi:10.1029/2011JD017187.
- Moss, R. H., and Coauthors, 2010: The next generation of scenarios for climate change research and assessment. *Nature*, **463**, 747–756.
- Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768–772.
- Pall, P., T. Aina, D. A. Stone, P. A. Stott, T. Nozawa, A. G. J. Hilberts, D. Lohmann, and M. R. Allen, 2011: Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature*, **470**, 382–385.
- Park, S., and C. S. Bretherton, 2009: The University of Washington shallow convection and moist turbulence schemes and their impact on climate simulations with the Community Atmosphere Model. *J. Climate*, **22**, 3449–3469.
- Polade, S. D., A. Gershunov, D. R. Cayan, M. D. Dettinger and D. W. Pierce, 2013: Natural climate variability and teleconnections to precipitation over the Pacific-North American region in CMIP3 and CMIP5 models. *Geophys. Res. Lett.*, **40**, 2296–2301, doi:10.1002/grl.50491.
- Porteous, A., and B. Mullan, 2013: The 2012-13 drought: An assessment and historical perspective. MPI Tech. Paper No. 2012/18, Ministry for Primary Industries/National Institute of Water & Atmospheric Research, 57 pp. [Available online at <http://www.niwa.co.nz/sites/niwa.co.nz/files/2013-18-The%202012-13%20drought%20an%20assessment%20and%20historical%20perspective.pdf>].
- Rajeevan, M., J. Bhate, J. Kale, and B. Lal, 2006: A high resolution daily gridded rainfall for the Indian region: Analysis of break and active monsoon spells. *Current Sci.*, **91**, 296–306.
- , S. Gadgil, and J. Bhate, 2010: Active and break spells of the Indian summer monsoon. *J. Earth Sys. Sci.*, **119**, 229–247, doi:10.1007/s12040-010-0019-4.
- Raupach, M. R., P. R. Briggs, V. Haverd, E. A. King, M. Paget, and C. M. Trudinger, 2009: Australian Water Availability Project (AWAP): CSIRO Marine and Atmospheric Research Component: Final Report for Phase 3. CAWCR Technical Report No.013, 67 pp.
- Reusch, D. B., R. B. Alley, and B. C. Hewitson, 2005: Relative performance of self-organizing maps and principal component analysis in pattern extraction from synthetic climatological data. *Polar Geogr.*, **29**, 188–212.
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625.

- Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean. *Nature*, **401**, 360–363.
- Shin, S., and P. D. Sardeshmukh, 2011: Critical influence of the pattern of tropical ocean warming on remote climate trends. *Climate Dyn.*, **36**, 1577–1591.
- Singh, C., 2013: Characteristics of monsoon breaks and intraseasonal oscillations over central India during the last half century. *Atmos. Res.*, **128**, 120–128, doi:10.1016/j.atmosres.2013.03.003.
- Slingo, J., 2013: Why was the start to spring 2013 so cold? Met Office briefing note, April 2013. [Available online at <http://www.metoffice.gov.uk/research/news/cold-spring-2013>.]
- Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's Historical Merged Land-Ocean Surface Temperature analysis (1880–2006). *J. Climate*, **21**, 2283–2296.
- Solomon, A., and M. Newman, 2012: Reconciling disparate twentieth-century Indo-Pacific ocean temperature trends in the instrumental record. *Nature Climate Change*, **2**, 691–699.
- Sperber, K. R., H. Annamalai, I.-S. Kang, A. Kitoh, A. Moise, A. Turner, B. Wang, and T. Zhou, 2013: The Asian summer monsoon: An intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Climate Dyn.*, **41**, 2711–2744.
- Stark, J. D., C. J. Donlon, M. J. Martin, and M. E. McCulloch, 2007: OSTIA: An operational, high resolution, real time, global sea surface temperature analysis system. *Oceans 2007 - Europe*, Aberdeen, Scotland, IEEE, Vols. 1–3, 331–334, doi:10.1109/OCEANSE.2007.4302251.
- Stott, P. A., G. S. Jones, J. A. Lowe, P. Thorne, C. F. Durman, T. C. Johns, and J.-C. Thelen, 2006: Transient climate simulations with the HadGEM1 climate model: Cases of past warming and future climate change. *J. Climate*, **19**, 2763–2782.
- 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.
- Tennant, W. J., G. J. Shutts, A. Arribas, and S. A. Thompson, 2011: Using a stochastic kinetic energy backscatter scheme to improve MOGREPS probabilistic forecast skill. *Mon. Wea. Rev.*, **139**, 1190–1206.
- White, C. J., D. Hudson, and O. Alves, 2014: ENSO, the IOD and the intraseasonal prediction of heat extremes across Australia using POAMA-2. *Climate Dyn.*, doi:10.1007/s00382-013-2007-2.
- Wilks, D. S., 2006: On “field significance” and the false discovery rate. *J. Appl. Meteor. Climatol.*, **45**, 1181–1189.
- Zanchettin, D., A. Rubino, and J. H. Jungclaus, 2010: Intermittent multidecadal-to-centennial fluctuations dominate global temperature evolution over the last millennium. *Geophys. Res. Lett.*, **37**, L14702, doi:10.1029/2010GL043717.
- Zhang, G. J., and N. A. McFarlane, 1995: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre general circulation model. *Atmos.—Ocean*, **33**, 407–446.