

# Multimodel detection and attribution of extreme temperature changes

Seung-Ki Min, Xuebin Zhang, Francis Zwiers, Hideo Shiogama, Yu-Shiang Tung & Michael Wehner

2013

Pacific Climate Impacts Consortium (PCIC)

PCIC Publications

©2013 American Meteorological Society. In compliance with funder open access policies, AMS makes all articles freely and publicly available one year from the date of final publication. <https://www.ametsoc.org/ams/publications/ethical-guidelines-and-ams-policies/ams-licenses-for-journal-article-reuse/>.

Original citation:

Min, S.-K., Zhang, X., Zwiers, F., Shiogama, H., Tung, Y.-S., & Wehner, M. (2013). Multimodel detection and attribution of extreme temperature changes. *Journal of Climate*, 26(19), 7430–7451. <https://doi.org/10.1175/JCLI-D-12-00551.1>

---

Downloaded from UVicSpace Research & Learning Repository

dspace.library.uvic.ca



University  
of Victoria

Libraries

## Multimodel Detection and Attribution of Extreme Temperature Changes

SEUNG-KI MIN,\* XUEBIN ZHANG,<sup>+</sup> FRANCIS ZWIERS,<sup>#</sup> HIDEO SHIOGAMA,<sup>@</sup> YU-SHIANG TUNG,<sup>&</sup>  
AND MICHAEL WEHNER\*\*

\* *CSIRO Marine and Atmospheric Research, Aspendale, Victoria, Australia, and Climate Research Division, Environment Canada, Toronto, Ontario, Canada*

<sup>+</sup> *Climate Research Division, Environment Canada, Toronto, Ontario, Canada*

<sup>#</sup> *Pacific Climate Impacts Consortium, University of Victoria, Victoria, British Columbia, Canada*

<sup>@</sup> *National Institute for Environmental Studies, Tsukuba, Ibaraki, Japan*

<sup>&</sup> *Department of Earth Science, National Taiwan Normal University, Taipei, Taiwan*

\*\* *Lawrence Berkeley National Laboratory, Berkeley, California*

(Manuscript received 31 July 2012, in final form 10 January 2013)

### ABSTRACT

Recent studies have detected anthropogenic influences due to increases in greenhouse gases on extreme temperature changes during the latter half of the twentieth century at global and regional scales. Most of the studies, however, were based on a limited number of climate models and also separation of anthropogenic influence from natural factors due to changes in solar and volcanic activities remains challenging at regional scales. Here, the authors conduct optimal fingerprinting analyses using 12 climate models integrated under anthropogenic-only forcing or natural plus anthropogenic forcing. The authors compare observed and simulated changes in annual extreme temperature indices of coldest night and day (TNn and TXn) and warmest night and day (TNx and TXx) from 1951 to 2000. Spatial domains from global mean to continental and subcontinental regions are considered and standardization of indices is employed for better intercomparisons between regions and indices. The anthropogenic signal is detected in global and northern continental means of all four indices, albeit less robustly for TXx, which is consistent with previous findings. The detected anthropogenic signals are also found to be separable from natural forcing influence at the global scale and to a lesser extent at continental and subcontinental scales. Detection occurs more frequently in TNx and TNn than in other indices, particularly at smaller scales, supporting previous studies based on different methods. A combined detection analysis of daytime and nighttime temperature extremes suggests potential applicability to a multivariable assessment.

### 1. Introduction

Many studies have consistently identified human influence due to increases in greenhouse gases on global and regional surface warming during the past 50 yr by comparing observed changes with model-simulated responses to anthropogenic forcing (fingerprints) using formal statistical techniques (Hegerl et al. 2007; Stott et al. 2010). Anthropogenic influence has also been detected in observed changes in sea level pressure, humidity and moisture content, precipitation, ocean heat content, and Arctic sea ice cover (Stott et al. 2010).

Consistent with mean temperature increases, several observational studies have also reported the warming of extreme temperatures over larger land areas (Frich et al. 2002; Alexander et al. 2006; Caesar et al. 2006; Donat et al. 2013). Model studies have suggested detectability of changes in extreme temperatures comparable to that of mean temperature changes (Hegerl et al. 2004; Kharin et al. 2007). Building on these observational and modeling efforts, there have been an increasing number of detection and attribution studies on temperature extremes that have detected anthropogenic influences at global and regional scales (Christidis et al. 2005; Shioyama et al. 2006; Christidis et al. 2010, 2011; Zwiers et al. 2011; Morak et al. 2011, 2012). Christidis et al. (2005) was the first formal detection and attribution study of observed changes in extreme temperatures. They compared data from Hadley Centre

---

*Corresponding author address:* Seung-Ki Min, School of Environmental Science and Engineering, Pohang University of Science and Technology, Pohang, Gyungbuk 790-784, South Korea.  
E-mail: skmin@postech.ac.kr

Coupled Model, version 3 (HadCM3) climate model ensemble simulations integrated with anthropogenic, natural, and natural plus anthropogenic forcing factors to the Hadley Global Historical Climatology Network-Daily (HadGHCND) gridded observations (Caesar et al. 2006) using an optimal fingerprinting technique (Allen and Stott 2003). Comparing global trend patterns, they detected anthropogenic signals in observed changes in the temperatures of the coldest day, coldest night, and warmest night annually but not in observations of the warmest day annually. Shiogama et al. (2006) obtained the same conclusions using climate simulations with the MIROC3.2(medres) but with noticeable difference in the level of model–observation agreement. Christidis et al. (2010) combined those two models when analyzing changes in the temperature of the warmest night annually and found that the detected anthropogenic signal is also separable from natural forcing at global scale.

More recently two studies have employed non-stationary extreme value theory, which provides a more appropriate statistical treatment of extremes and potentially helps to increase the signal-to-noise ratio (Christidis et al. 2011; Zwiers et al. 2011). They considered time-evolving location parameters when fitting the generalized extreme value distribution to annual extremes. Zwiers et al. (2011) conducted the first multimodel regional detection analysis by comparing changes in annual extremes of daily maximum and minimum temperatures from eight climate models with those from Hadley Centre global land-based climate extreme datasets (HadEX) observations (Alexander et al. 2006) for the period of 1961–2000. They detected the anthropogenic signal in changes in the temperature of the coldest night and day and warmest night over many regions with more frequent detection in warmest night temperatures. They also detected anthropogenic influence on changes in the temperature of the warmest day at the global scale. Christidis et al. (2011) detected the contribution from both anthropogenic and natural forcing signals to changes in the temperature of the warmest day on the global scale for 1950–99 using HadCM3 model runs. On the other hand, Morak et al. (2011) have investigated changes in number of warm nights annually using observations and five climate models. They detected anthropogenic influence in observed increases in the frequency of warm nights in many regions over northern continents, consistent with Zwiers et al. (2011). Extending Morak et al. (2011), Morak et al. (2012) have detected external (natural plus anthropogenic) influence on the observed changes in the frequency of cold and warm extremes during cold and warm seasons using Hadley Centre Global Environment Model, version 1 (HadGEM1) runs.

Although several detection studies on extreme temperatures are available, all with consistent results, further study is still needed. First, previous studies are largely based on a limited number of model simulations and multimodel studies and thus a larger number of models are required to reduce intermodel signal uncertainties, which may improve detection on regional scales. Second and more importantly, reducing signal uncertainty may allow signal separation on subglobal scales. In this study, we apply an optimal detection method to observations of temperature extremes for the period of 1951–2000, using signals that are estimated from an ensemble of 12 climate models. We evaluate whether anthropogenic and naturally forced signals are present in global- and regional-mean indicators of changes in extremes. In addition, we conduct a combined attribution analysis by considering daily minimum and maximum extremes simultaneously as a first step toward a multivariable attribution approach for extremes (Jones et al. 2003; Schnur and Hasselmann 2005; Barnett et al. 2008). To facilitate comparison among different indices and combinations of indices, we standardize extremes indices based on extreme value theory following Min et al. (2009, 2011).

This paper is structured as follows: Section 2 describes the datasets from observations and model simulations that we use as well as the way in which we process the data. Techniques for the standardization of extreme indices and optimal fingerprinting are also explained there. Detection results for global, continental, and subcontinental spatial domains are provided in section 3 with discussions. Conclusions are given in section 4.

## 2. Data and methods

### *a. Observations and models*

The annual minimum and maximum of daily minimum (TN) and maximum (TX) temperatures are analyzed; these indices are referred to as coldest night (TNn), coldest day (TXn), warmest night (TNx), and warmest day (TXx), respectively. We use HadEX as observations, which are gridded data with  $3.75^\circ$  longitude  $\times$   $2.5^\circ$  latitude grid boxes and cover the period from 1951–2003 (Alexander et al. 2006). We use 1951–2000 for our analysis during which daily data from multiple models are available. This selection is also partly due to a sudden drop of data coverage after 2000 in TNx and TXx to lower than 70% of 1961–1990 average (data coverage is also relatively reduced during 1951–54, but coverage is nevertheless higher than 70% of the 1961–90 average).

Model data are obtained from the twentieth century Climate in Coupled Model (20C3M) experiments of the World Climate Research Programme (WCRP) phase 3 of the Coupled Model Intercomparison Project (CMIP3)

TABLE 1. List of CMIP3 coupled climate model simulations analyzed in this study. The number of ensemble members for 1951–2000 is given for ANT and ALL runs except for CCSM3 and PCM, for which the 1950–99 period has been used instead. For CTL (preindustrial control) runs, the number of nonoverlapping 50-yr chunks is provided.

Model	Model expansion	ANT	ALL	CTL (No. of 50-yr chunks)
CCSM3	Community Climate System Model, version 3	3	1	—
CGCM3(t47)	Canadian Centre for Climate Modelling and Analysis (CCCma) Coupled Global Climate Model, version 3.1 (t47 resolution)	5	—	20
CGCM3(t63)	Canadian Centre for Climate Modelling and Analysis (CCCma) Coupled Global Climate Model, version 3.1 (t63 resolution)	1	—	10
CSIRO Mk3.0	Commonwealth Scientific and Industrial Research Organisation Mark, version 3.0	1	—	6
CSIRO Mk3.5	Commonwealth Scientific and Industrial Research Organisation Mark, version 3.5	1	—	12
ECHAM5/MPI-OM	ECHAM5/Max Planck Institute Ocean Model	3	3	10
ECHO-G	ECHAM and the global Hamburg Ocean Primitive Equation	3	3	6
GFDL CM2.0	Geophysical Fluid Dynamics Laboratory Climate Model, version 2.0	—	3	10
GFDL CM2.1	Geophysical Fluid Dynamics Laboratory Climate Model, version 2.1	—	3	10
MIROC3.2(medres)	Model for Interdisciplinary Research on Climate, version 3.2 (medium resolution)	10	10	72
MIROC3.2(hires)	Model for Interdisciplinary Research on Climate, version 3.2 (high resolution)	—	1	2
PCM	Parallel Climate Model	—	2	—
Total		8 models, 27 runs	8 models, 26 runs	10 models, 158 chunks

multimodel dataset (Meehl et al. 2007a) and in part directly from individual modeling centers (Table 1). We use 12 models in total that provide daily data for the period of 1951–2000 (or 1950–99 for the CCSM3 and PCM models). 20C3M models are divided into two groups according to external forcing factors implemented. The first group consists of models that conducted anthropogenic forcing-only experiments (ANT) where historical increases in greenhouse gases and sulfate aerosols are included as major external forcings. The second group consists of models that conducted historical simulations with both natural and anthropogenic forcing (ALL) where natural forcing (NAT) from historical changes in solar and volcanic activities are additionally implemented. We have 27 runs from 8 models available for ANT and 26 runs from 8 models for ALL. Note that the ALL and ANT groups are composed of different models; four models belong to both groups. We use multimodel mean difference between ALL and ANT as a fingerprint of NAT. To estimate the range of internal climate variability (see below), we also use extreme temperature data from preindustrial control simulations (CTL) from 10 models (Table 1).

### b. Probability-based index

The models have different horizontal resolutions, which can affect the comparison of changes in extremes between

observations and models. We therefore standardize annual temperature extremes prior to analysis following the method based on extreme value theory used by Min et al. (2009, 2011). Time series of annual temperature extremes  $T$  at individual grid boxes are converted into probability-based indices (PI) ranging from 0 to 1 as follows: First we fit the generalized extreme value (GEV) distribution to individual samples of 50 annual extremes using the maximum likelihood method (Kharin and Zwiers 2005). Each annual  $T$  is then converted to PI by evaluating the corresponding fitted cumulative density function (CDF) at the value of that annual maximum. This is done separately for each grid box of each dataset.

The PI transformation (probability integral transform in a statistical term) can be based on distributions (e.g., the empirical distribution function) other than the GEV distribution. However, the choice of the GEV is well supported by statistical extreme value theory (e.g., Coles 2001), and in previous applications the GEV distribution has been shown to fit observational and model data well (Kharin et al. 2005). In any case, the specific choice of distribution is not particularly important provided it is able to transform skewed distributions with the long tails associated with extremes to one that is roughly symmetric and has the same width everywhere globally. This will ensure both that all regions will receive

roughly equal weight in the analysis and that the distribution of the quantities used for detection, large area averages, is close to Gaussian (as a consequence of the central limit theorem). Convergence to the normal distribution will be harder to argue for physical temperature extremes because we would be averaging values with different distributions and because local values come from skewed distributions. The convergence will also be slower if extremes at different locations are correlated. There are, nevertheless, some potential benefits of using the GEV distribution. We can use GEV parameters estimated from the historical period when interpreting future changes in extremes (e.g., Kharin et al. 2007). In further work, using the GEV would also enable us to take account of possible influence of climate variability by incorporating climate variability modes into GEV models using a covariate (e.g., Zhang et al. 2010).

For each model dataset, we first calculate PI on the original grid points and then interpolate the PIs onto the HadEX  $3.75^\circ \times 2.5^\circ$  grid. The interpolated PIs from climate models are then masked with observational data availability by considering only grid points with more than 40 yr of observations. To consider consistent data availability among the four extreme temperature indices (TNn, TXn, TNx, and TXx), we use the observational mask of TNx that exhibits the least coverage among indices. This results in the exclusion of India and part of northern Africa from our analysis (see below).

To facilitate the physical interpretation of PI, we examine the relationship between extreme temperature  $T$  and PI using HadEX observations. Figure 1 illustrates scatterplots of annual anomalies of  $T$  and PI from 1951 to 2000 for four extreme temperature indices averaged over the global land area with sufficient data. It shows a strong linear relationship between area-averaged  $T$  and PI with correlation coefficients of 0.99. Regression slopes vary among extreme indices and are steeper for cold extremes [TNn and TXn: around  $0.8\text{--}0.9\text{ K } (10\%)^{-1}$  change in PI] than for warm extremes [TNx and TXx: around  $0.4\text{--}0.5\text{ K } (10\%)^{-1}$  change in PI], representing larger variations in cold extremes during the 50 yr. The linear relationship holds well in five continental domains with high correlation coefficients (0.98–0.99), but regression slopes vary greatly across continents in particular for cold extremes [TNn and TXn:  $0.3\text{--}1.0\text{ K } (10\%)^{-1}$  PI; see Table 2]. Relatively larger regression slopes for cold extremes in Europe represent stronger interannual variability, implying that less weight is given to the region by the PI transformation than other regions when calculating global averages and vice versa. Area-mean PI anomalies can be interpreted as relative changes with respect to typical 2-yr return values because the time-mean

PI is near 0.5 at each grid point and large area averaging reduces temporal variations.

### c. Data processing

We conduct attribution analyses on each of the four standard annual extreme temperature indices TNn, TNx, TXn, and TXx. We obtain anomalies of extreme indices (PI) for each index relative to 1951–2000 mean at each grid point and then calculate area-averaged means. To take account of long-term variations (or reduce interannual noise) and also to reduce the analysis dimension as needed for an optimal detection analysis (see below), we conduct analyses for both time series of decadal means, which gives 5-dimensional vectors, and 5-yr means, which gives 10-dimensional vectors. We also conduct combined attribution analyses so as to consider changes in day and night temperatures simultaneously when comparing observations to models. For this, decadal-mean PI anomalies for cold extremes (TNn + TXn) or warm extremes (TNx + TXx) are combined to give 10-dimensional vectors.

We consider time series of different spatial averages: global (GLB), continental, and subcontinental regions. Spatial coverage is sparse in many regions, and thus to ensure reasonable representation of diverse regions, we require more than 30% of fraction of grid boxes with sufficient data for each domain. This gives five continents as North America (NAM), South America (SAM), Europe (EUR), Asia (ASI), and Australia (AUS) and 16 subcontinents based on the domains defined by Giorgi and Francisco (2000) with two slight modifications: changing the acronym for the Mediterranean Basin (MED) into Southern Europe (SEU) and dividing Australia into northern Australia (NAU) and southern Australia (SAU). Note that all regions have more than 50% spatial data coverage except for southern Africa and northern Australia, which have around 35%. Recall also that grid boxes are required to have at least 40 yr of observations. Refer to Table 3 for more details.

To obtain insight into the possible contribution of additional information from spatial patterns, we combine decadal-mean PI anomalies from five continental regions (area weighted), which gives 25-dimensional space–time vectors of global scale (GLB5), and compare detection results to the case without using spatial information (GLB).

### d. Optimal detection analysis

We compare observed PI changes to model simulations with a standard optimal fingerprinting technique (Allen and Tett 1999; Allen and Stott 2003). This method assumes that observations  $\mathbf{Y}$  are expressed as the sum  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  of scaled fingerprints  $\mathbf{X}$  plus internal variability  $\boldsymbol{\varepsilon}$ .

## PI–T relationship GLB

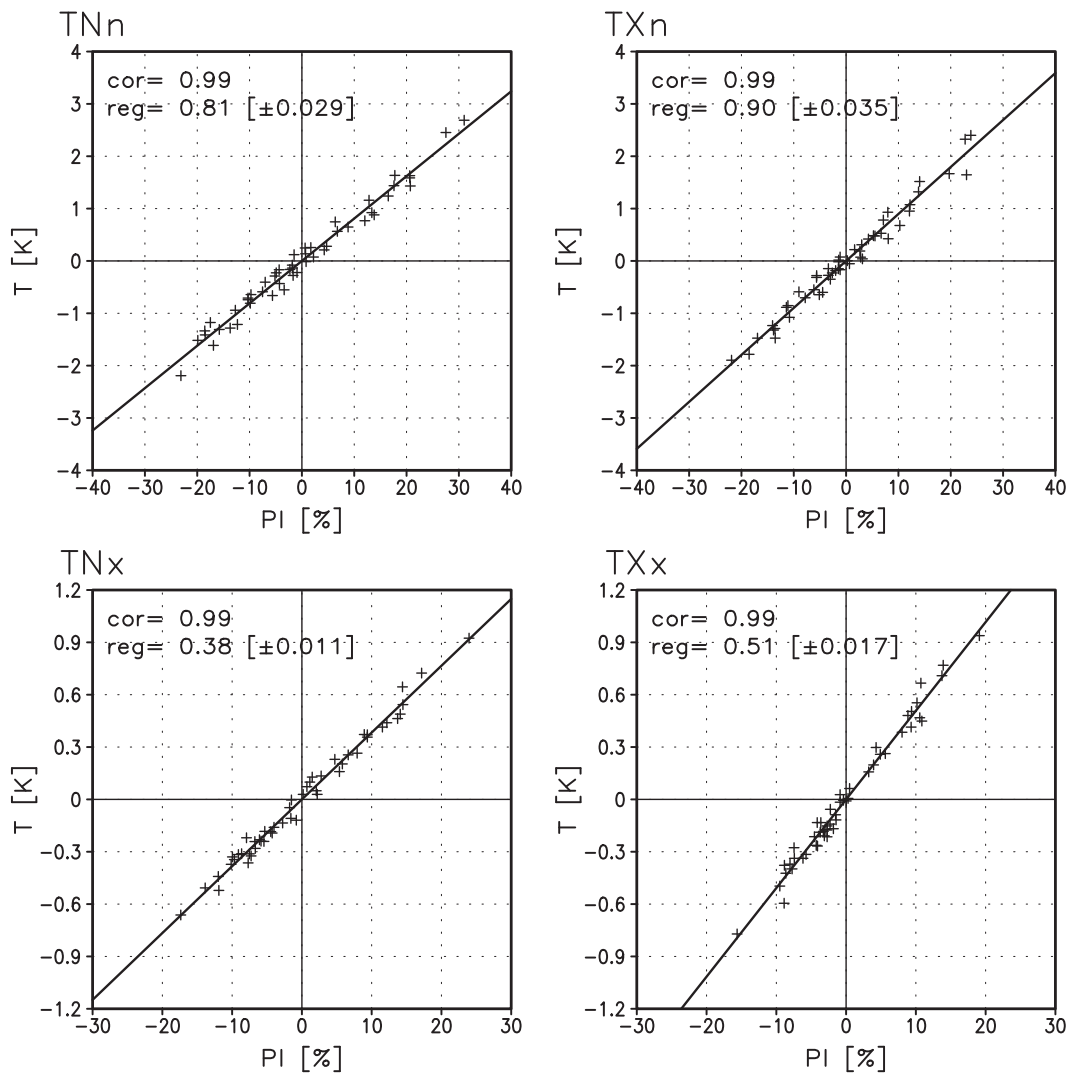


FIG. 1. Scatterplots between global-mean PI (x axis; % probability) and extreme temperature  $T$  (y axis; K) anomalies using annual extreme values from HadEX observations during 1951–2000 for the coldest night (TNn), coldest day (TXn), warmest night (TNx), and warmest day (TXx). Anomalies are relative to the 1951–2000 means. Straight lines represent least squares regression slopes. Correlation coefficients and regression slopes [unit:  $K (10\%)^{-1}$  change in PI] are given in each panel. Numbers in bracket depict 95% uncertainty ranges of the regression slopes estimated with the normality assumption of the regression errors. Note different y-axis ranges between cold extremes (TNn and TXn) and warm extremes (TNx and TXx).

This is equivalent to regressing observations  $\mathbf{Y}$  onto fingerprints  $\mathbf{X}$ . Regression coefficients  $\boldsymbol{\beta}$  (also called scaling factors) are estimated by the total least squares method (Allen and Stott 2003). Fingerprints are estimated from the multimodel mean (ANT and ALL) and internal variability is estimated from CTL runs (see below for details). There are two settings for regression analyses. In the single-signal analysis, observations are regressed onto ANT or ALL separately. In the two-signal

analysis, observations are regressed onto ANT and ALL simultaneously, from which one can examine whether the ANT influence is separable from NAT whose fingerprint is estimated from ALL – ANT. Detection occurs if the estimated scaling factor is positive and the 90% uncertainty range of scaling factor excludes zero, which implies that there is a significant relationship between observed change and fingerprint patterns. The detected modeled fingerprint is assessed to agree with

TABLE 2. Regression slopes (Kelvins per 10% change in PI) between the PI and four extreme temperature indices (TNn, TXn, TNx, and TXx) averaged over the five continents. Anomalies are considered to calculate regression slopes as in Fig. 1. Global-mean values are provided for comparison. Correlation coefficients between PI and  $T$  are very high (0.98–0.99), and the regression slope uncertainty is very small (the half-width of the 95% confidence interval for the slope is less than 0.054 Kelvins per 10% PI) in all cases.

Domain	TNn	TXn	TNx	TXx
Globe	0.81	0.90	0.38	0.51
North America	0.80	0.88	0.39	0.52
South America	0.50	0.52	0.31	0.41
Europe	1.01	0.99	0.37	0.56
Asia	0.86	0.96	0.38	0.49
Australia	0.32	0.35	0.46	0.48

observed changes (from which we would infer attribution in the absence of other plausible explanations for the observed changes) if the uncertainty range of scaling factor additionally includes unity.

We obtain nonoverlapping CTL chunks of 100-yr length from all available individual models and split each 100-yr chunk into two 50-yr chunks (Table 1). The first set of CTL chunks is used to obtain the empirical orthogonal functions (EOF) projection (see below) and also to obtain best estimates of regression coefficients, and the second set is used to estimate the 5%–95% uncertainty ranges of the regression coefficients and to conduct a standard residual consistency test (Allen and Tett 1999). We apply the residual consistency test in order to consider model skill at simulating internal variability. If model-simulated variability is too small, signal-to-noise ratio will be inflated because of reduced

noise (i.e., uncertainty range of the scaling factor will become narrower), leading to spurious detection. To avoid this case, we compare model-simulated variance with observational residual variance which is estimated by removing from observations the portion explained by external forcing signals ( $\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$ ) following Allen and Tett (1999). Here different temporal scales (spatiotemporal scales in case of GLB5), represented by leading EOFs, are considered and we find that overall modeled variability is consistent with observed variability when looking at long-term temporal and larger spatial scales.

All results shown below are obtained using four leading EOFs, which corresponds to the full space analysis for single variable of five decadal means (because the overall mean is removed before calculating EOFs). For combined variable analyses (TNn + TXn and TNx + TXx), four EOFs explain a large portion of internal variability with 74%–98% of explained variance. For the single variable five-continent (GLB5) analyses, 10 leading EOFs are used, which explain 89%–96% of modeled variance of internal variability. Detection results are largely insensitive to the use of different EOF truncations (not shown). The use of decadal means leads to low dimensionality, particularly in the case of single-variable single-domain analyses. We have tested the sensitivity of our single- and two-signal results in these cases to the use of 5-yr means, which approximately doubles the analysis dimension, and obtained similar detection results although the residual consistency test fails more frequently due to increased noise levels arising from shorter-term variations (see below). In addition, results from combined variable analyses (TNn + TXn and TNx + TXx), which use 10-dimensional analysis vectors,

TABLE 3. List of 16 subcontinental regions used in this study. The fraction of grid points with sufficient data is estimated from TNx during 1951–2000.

Continent	Region	Acronym	Latitude	Longitude	Data fraction (%)
North America	Alaska	ALA	60°–72°N	170°–103°W	74
	Greenland	CGI	50°–85°N	103°–10°W	63
	Western North America	WNA	30°–60°N	130°–103°W	97
	Central North America	CNA	30°–50°N	103°–85°W	95
	Eastern North America	ENA	25°–50°N	85°–50°W	88
South America	Central America	CAM	10°–30°N	116°–83°W	52
	Southern South America	SSA	56°–20°S	76°–40°W	77
Europe	Northern Europe	NEU	48°–75°N	10°W–40°E	94
	Southern Europe	SEU	30°–48°N	10°W–40°E	57
Africa	Southern Africa	SAF	35°–12°S	10°W–52°E	35
Asia	North Asia	NAS	50°–70°N	40°E–180°	97
	Central Asia	CAS	30°–50°N	40°–75°E	73
	Tibet	TIB	30°–50°N	75°–100°E	79
	East Asia	EAS	20°–50°N	100°–145°E	95
Australia	Northern Australia	NAU	30°–11°S	110°–155°E	35
	Southern Australia	SAU	45°–30°S	110°–155°E	87

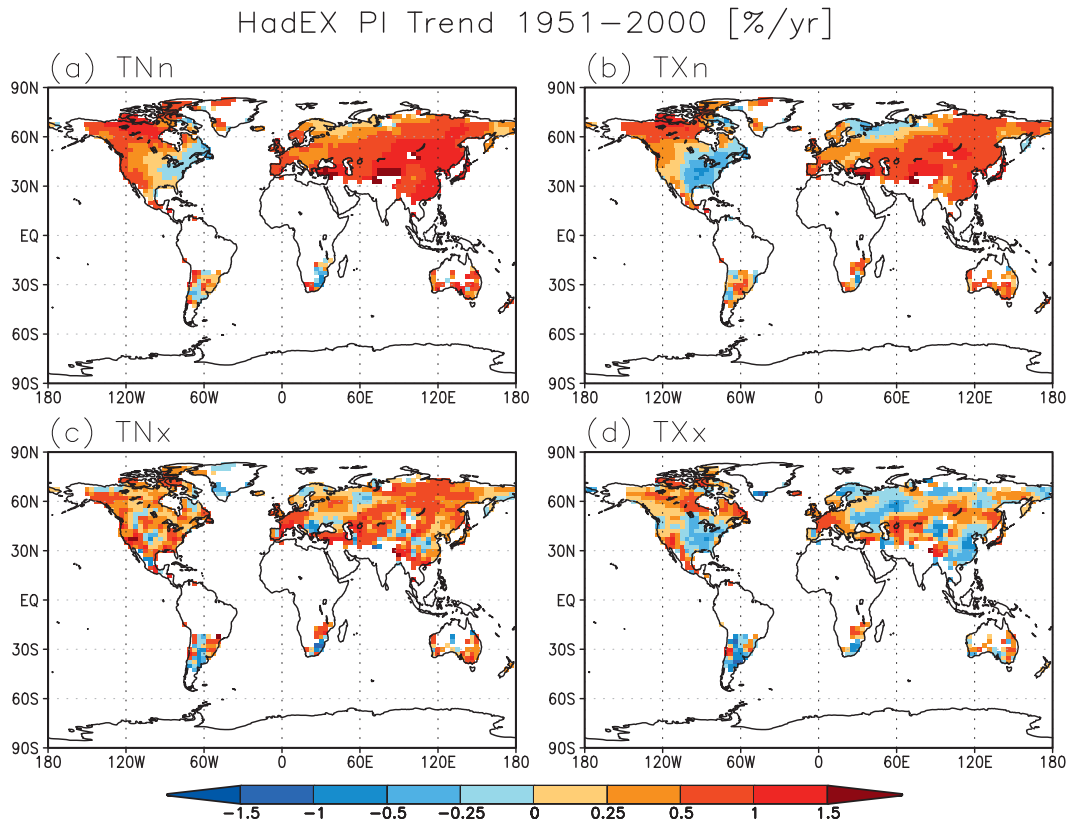


FIG. 2. Geographical patterns of observed trends in PI of temperature extremes (% probability per year) during 1951–2000: (a) TNn, (b) TXn, (c) TNx, and (d) TXx. Positive trends mean weakening of cold extremes in (a),(b) and intensification of warm extremes in (c),(d). Analysis is confined to grid points with sufficient data of longer than 40-yr length for each variable. The observed TNx mask has been applied to all variables for consistency.

increase the robustness of our detection results on both global and regional scales (see below).

We further note the dimension reduction issue. Optimal detection analysis needs to be conducted in a reduced space given that typically, insufficient model output is available to estimate the variance–covariance matrices of the internal variability noise (Allen and Tett 1999). There are different ways to perform the dimension reduction, but almost all detection analyses use a combination of dimension reduction prior to the analysis coupled with further dimension reduction within the analysis. The former, which takes the form of some kind of spatial and temporal filtering, usually still does not reduce dimensionality enough to allow estimation of full rank variance–covariance matrices from the available samples of control simulations, and thus a further reduction is generally performed with an EOF truncation. The balance between how much filtering is done before and within the analysis is a subjective choice based on expert judgment. In our case, computing time/space averages and then conducting EOF analysis on these averages seems to be a better way since one can reduce some

of the subjectiveness related to the use of time/space averaging, and at the same time one can still use the EOF to optimize the dimension reduction (so the largest portion of variance is retained). We are not aware of any previous studies that examined influence of initial smoothing on detection results in a systematic manner, for which further investigation is warranted.

### 3. Results

#### a. Observed and modeled trends

Figure 2 illustrates observed patterns of PI trends for the four extreme temperature indices. Positive trends mean weaker cold extremes and stronger warm extremes. Overall positive trends dominate in TNn and TXn, although there are large areas of decreasing trends in TXn. TNx is also dominated by warming trends while a mixture of positive and negative trends appear in TXx. This is consistent with trend patterns based on percentile-based temperature indices (Frich et al. 2002; Alexander et al. 2006; Caesar et al. 2006; Brown et al. 2008) or

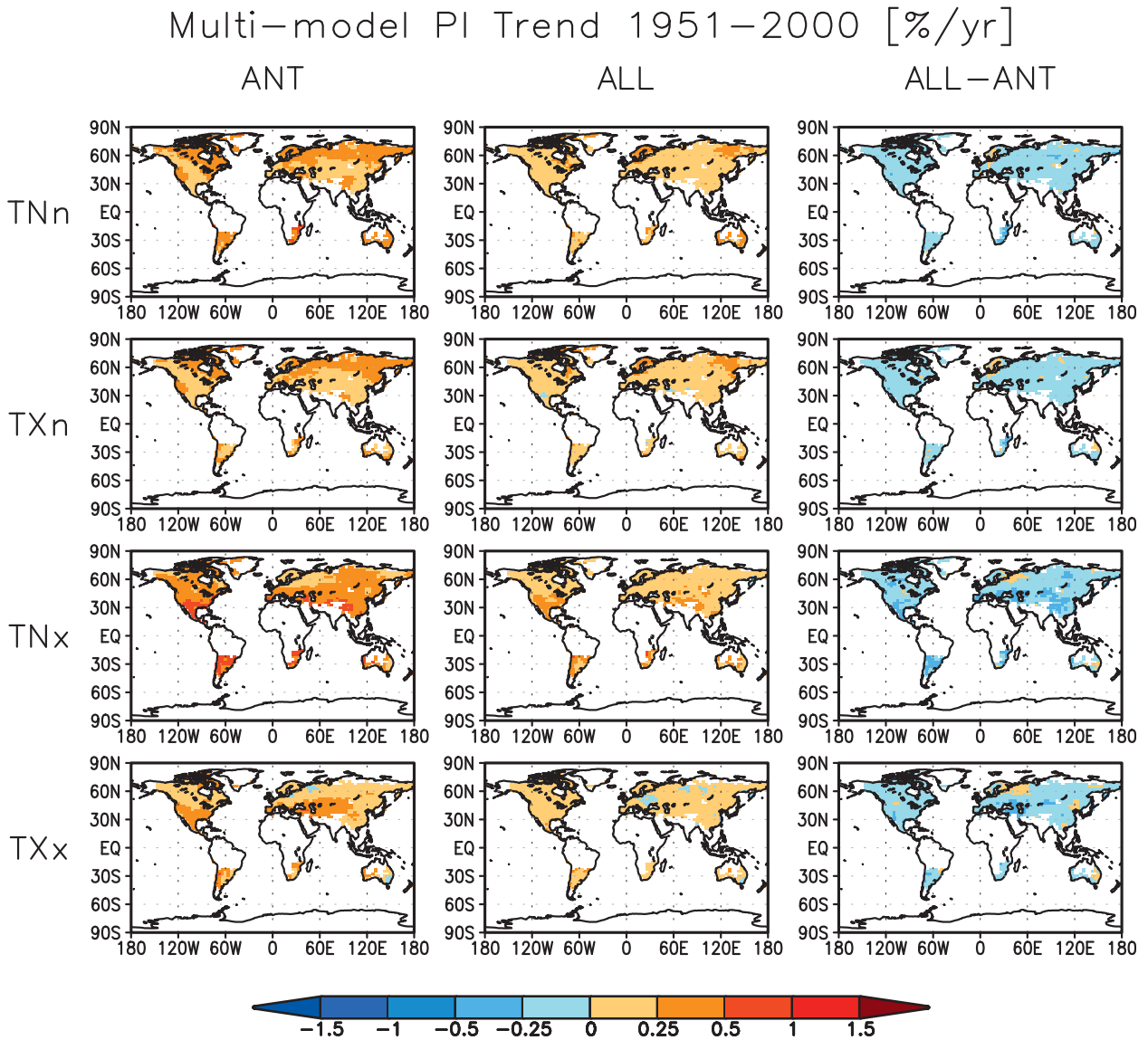


FIG. 3. As in Fig. 2, but for multimodel mean PI trends (% probability per year) from ANT and ALL simulations. Trend differences (ALL minus ANT) are also displayed. Note that same color scales are used as in Fig. 2. Refer to Table 1 for the list of model runs.

absolute temperature indices (Christidis et al. 2005; Alexander et al. 2006; Christidis et al. 2011). In terms of amplitudes, cold extremes (TNn and TXn) exhibit much stronger warming compared to warm extremes (TNx and TXx).

Trend patterns in PI simulated by the ALL and ANT multimodels are displayed in Fig. 3. Trends are first calculated for individual runs and then averaged to estimate multimodel mean trends. ANT runs show that warming occurs almost everywhere for all extreme indices, meaning weakening of cold extremes and intensification of warm extremes in line with global warming. Differences in trend patterns are seen between cold and

warm extremes. Cold extremes are more pronounced over the northern high latitudes while warm extremes are stronger over the subtropics in both hemispheres, consistent with future projection patterns of seasonal-mean surface warming over land: that is, high-latitude winter warming and subtropical summer warming (Giorgi et al. 2001; Meehl et al. 2007b). ALL runs exhibit similar warming patterns to ANT but with weaker magnitudes, which is likely to be due to long-term cooling by natural external forcing for the period (e.g., Huntingford et al. 2006). Difference patterns between ALL and ANT trends (ALL – ANT) clearly indicate the net cooling effect of natural forcing during the latter half of the

## Global Mean Time Series 1951–2000

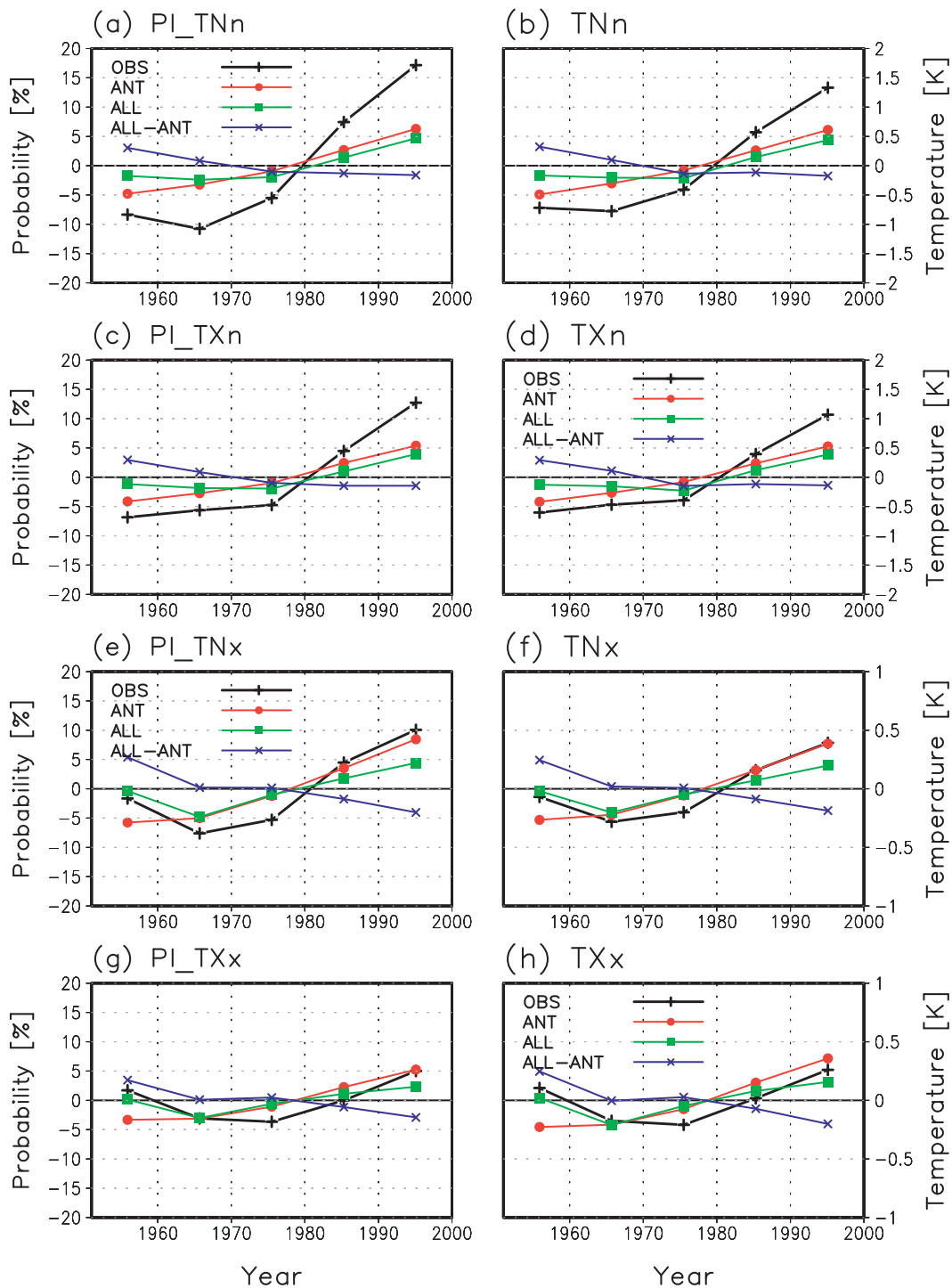


FIG. 4. Time series of global-mean decadal-mean (left) PI (% probability) and (right)  $T$  (K) anomalies from HadEX observations (black) and multimodel means of ANT (red) and ALL (green) simulations. ALL minus ANT is displayed in blue to look at the possible influence of natural (solar and volcanic) forcing. Anomalies are relative to each time mean. Note different y-axis ranges between cold extremes (TNn and TXn) and warm extremes (TNx and TXx) in  $T$  time series.

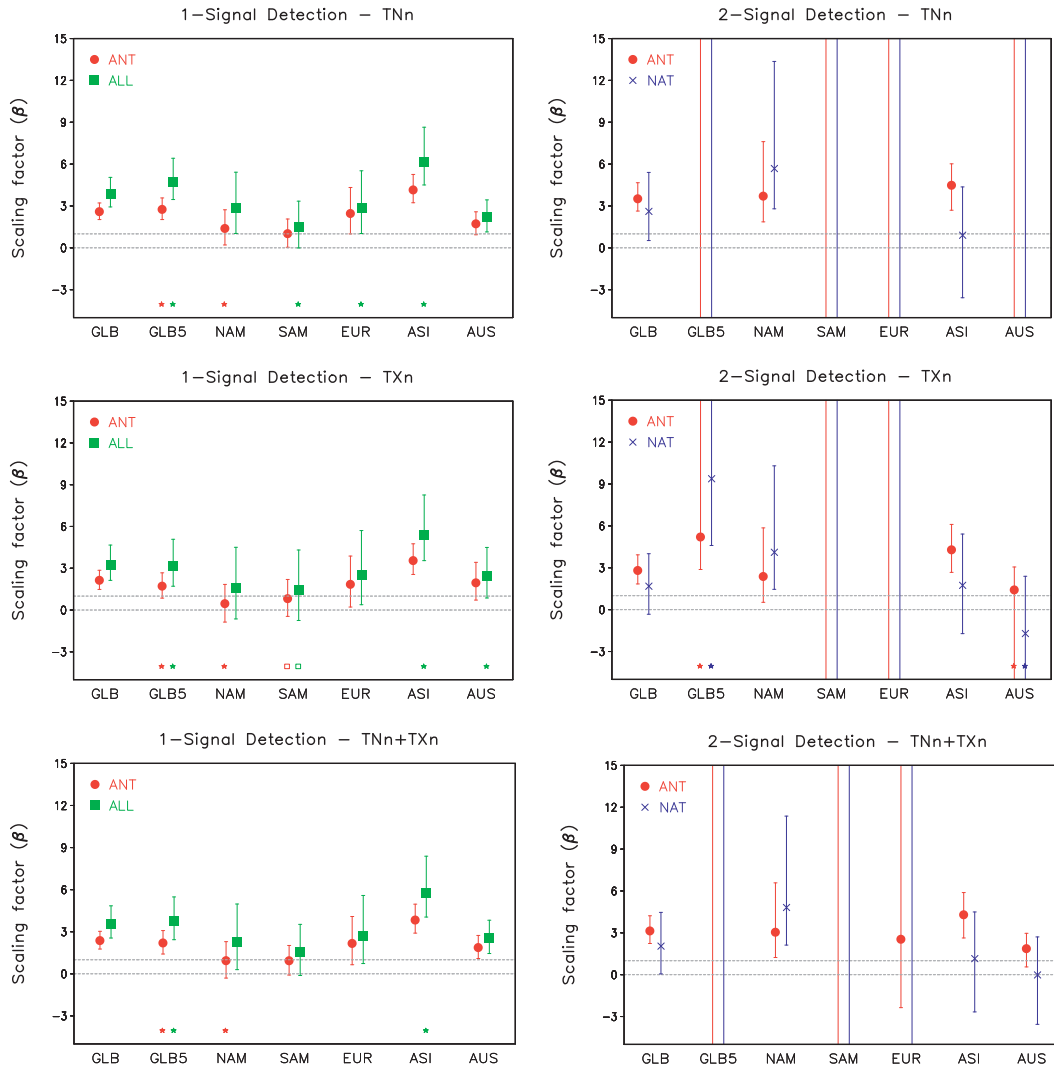


FIG. 5. Results from (left) single-signal and (right) two-signal detection analyses for cold extreme indices of (top) TNn, (middle) TXn, and (bottom) both indices combined (TNn + TXn) for global and continental scales. Results for individual indices are based on time vectors of area-mean decadal-mean PI, while those for GLB5 represent results from space–time analysis where decadal means from five continental regions are combined. Best guess (marks) and 5%–95% ranges (error bars) of scaling factors are shown for each extreme index for global and five continental domains. An error bar without a data point represents unbounded uncertainty in estimating the 90% ranges. The two dashed horizontal lines represent zero and unity. An asterisk (open square) indicates the failure of the residual consistency test because of too small (too large) model variability. See text for more details.

twentieth century. Volcanic cooling seems to be dominant because solar forcing does not change much during 1951–2000 (Hegerl et al. 2007). As a whole, ANT and ALL capture observed warming trends in cold and warm extremes, albeit with somewhat weaker amplitudes. Simulated patterns of change have less spatial variability than observed, presumably because ensemble averaging has substantially reduced the amplitude of the effects of internal variability on 50-yr trends.

Global-mean PI time series of observed and simulated changes are compared in Fig. 4. Decadal-mean anomalies

with respect to each time mean are presented for four extreme indices of TNn, TXn, TNx, and TXx. To aid in the interpretation of PI changes, global-mean  $T$  time series are also shown. The strongest observed warming is found in TNn with about 7% probability increase per decade when estimated based on the linear slope, which corresponds to a warming of about  $+0.5 \text{ K decade}^{-1}$  (Fig. 4; also refer to Fig. 1, which shows PI– $T$  relationships). Observed TNn warming is characterized by little change in early decades and a strong increase since the 1970s. Both ANT and ALL runs reproduce the observed

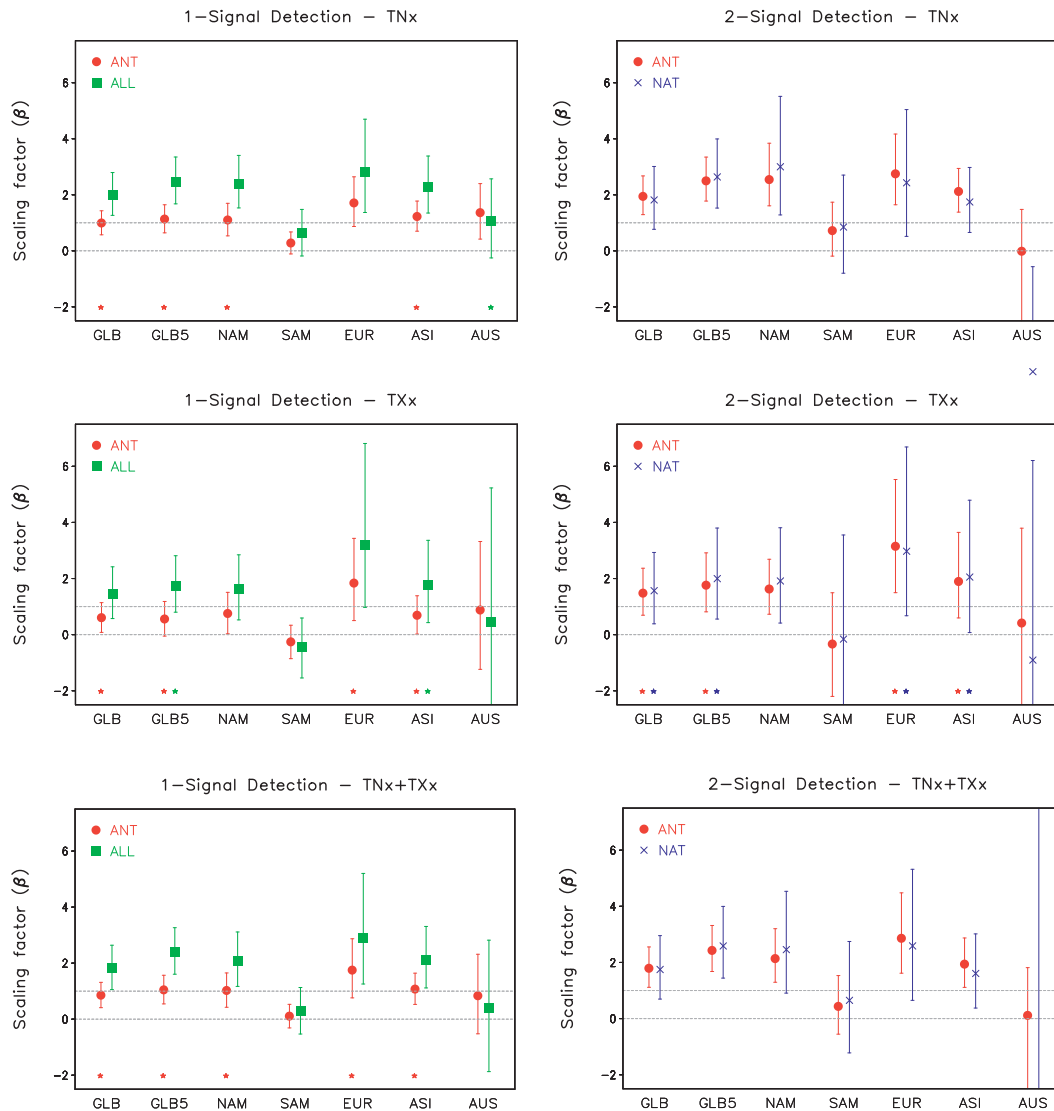


FIG. 6. As in Fig. 5, but for warm extreme indices of (top) TNx, (middle) TXx, and (bottom) TNx + TXx.

increasing trends, but the amplitudes are weaker (around +3% and +2% probability per decade, respectively). The difference between ANT and ALL is clearly identified. While ANT shows a steady increase, ALL displays temporal behavior similar to the observed: that is, stronger warming after the 1970s. This indicates the influence of natural forcing on ALL results and the ALL minus ANT pattern indeed displays a slight cooling impact in the early decades in TNn (Fig. 4). Observed and modeled changes in TXn much resemble those in TNn but with weaker amplitudes especially in observations (+5% probability per decade or +0.4 Kelvins per decade). This is consistent with faster warming of minimum temperature extremes than maximum temperature extremes, reflecting changes in the distribution other than just changes in the mean (Alexander et al. 2006; Caesar

et al. 2006; Donat and Alexander 2012). ANT and ALL runs reproduce the faster warming of TNn than TXn (by about 1% probability per decade) but underestimate the observed difference. This agrees with previous model studies analyzing annual-mean daily maximum and minimum temperatures, which pointed out model deficiencies in simulating changes in cloud cover, precipitation, and soil moisture as possible causes of model under-simulation (Stone and Weaver 2003; Zhou et al. 2010).

Observed changes in warm extremes (TNx and TXx) show different temporal fluctuations compared to cold extremes (Fig. 4). Decreasing trends representing weakening of warm extremes appear more strongly in early decades particularly for TXx, making positive linear trends much weaker during the whole 50 yr. While it is possible that this modulation in observed trends could

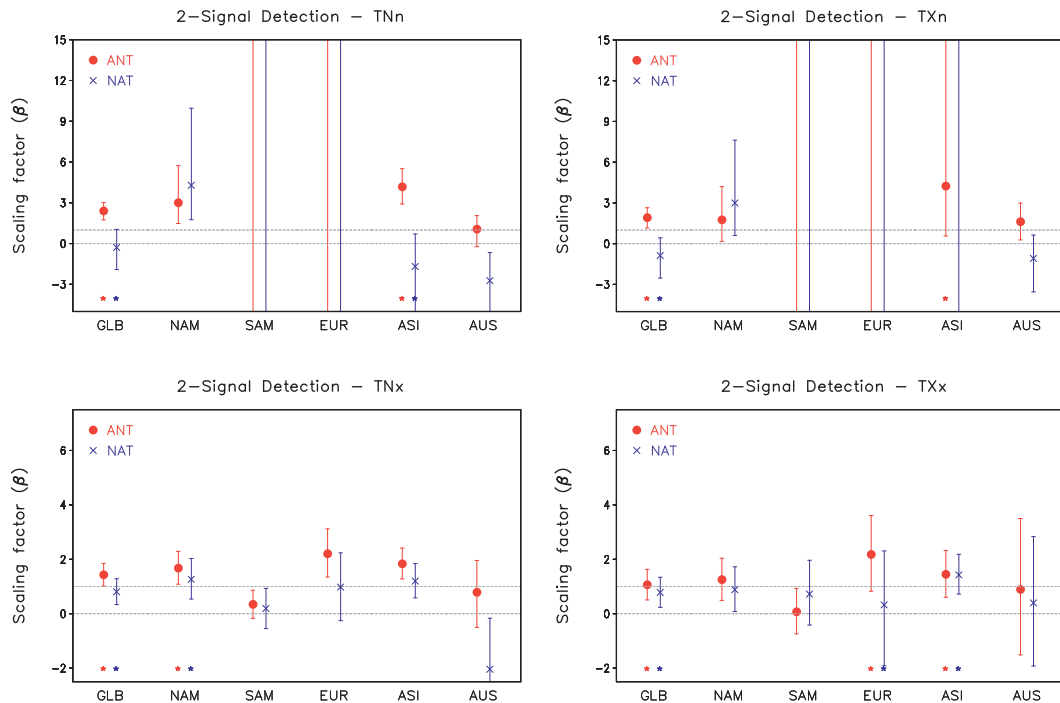


FIG. 7. As in Fig. 5, but for two-signal detection analyses using 5-yr-mean PI anomalies for both (top) cold extremes (TNn and TXn) and (bottom) warm extremes (TNx, and TXx) averaged over the globe and five continental regions. Note that this is based on full-dimensional analyses using 10-dimensional time vectors of 5-yr-mean anomalies (thus having nine independent dimensions).

be due to low-frequency internal variability or missing forcings (such as indirect aerosol effects, which are not included in many models considered), only ALL runs including natural forcing can reproduce these patterns in both TNx and TXx, whereas ANT runs show monotonic increases although providing comparable trend slopes to observations. ALL – ANT patterns are characterized by decreases at the earlier and latter decades, which reflects volcanic-induced long-term global cooling that is expected to particularly affect the lower latitudes in summer (Briffa et al. 1998; Robock 2000). In contrast, volcanic eruptions bring slight warming over northern continents during postvolcanic winters through an advective effect involving stratospheric dynamics, which climate models generally underestimate (Robock 2000; Stenchikov et al. 2006; Driscoll et al. 2012).

Simple comparisons of horizontal trend patterns and global-mean time series indicate the contribution of both ANT and NAT components to changes in extreme temperatures. Below we quantify these comparisons using the optimal fingerprinting technique.

#### *b. Detection results for global and continental domains*

Figure 5 shows results from optimal fingerprinting analyses for cold extreme indices. Best estimates and

90% ranges of regression coefficients  $\beta$  (or scaling factors) are illustrated for TNn, TXn, and coldest night and day combined (TNn + TXn) averaged over GLB and five continental areas of NAM, SAM, EUR, ASI, and AUS and also for the global scale with five continents combined (GLB5). An externally forced signal (ALL, ANT, or NAT) is assessed to be detected when the 90% range of  $\beta$  lies above zero and evidence to support attribution is obtained if the 90% range of  $\beta$  on the detected signals includes unity. Single-signal and two-signal results are displayed together to examine the relative contributions of ANT and NAT forcing. GLB results clearly show that both ALL and ANT are detected for TNn, TXn, and TNn + TXn in single-signal analysis and that ANT is separable from NAT in the two-signal analysis. NAT is detected in TNn and TNn + TXn as well. Best estimates of scaling factors lie above unity, implying that models underestimate observed warming (weakening) of cold extremes by a factor of 2–4 at the global scale. The GLB5 analysis based on five continents shows limited detection results with larger uncertainty range of the scaling factors for cold extremes compared to GLB: simultaneously considering continental-mean changes appears not to strengthen detection results because intercontinental differences are relatively weak in the modeled response (Fig. 3). When using fewer numbers of leading

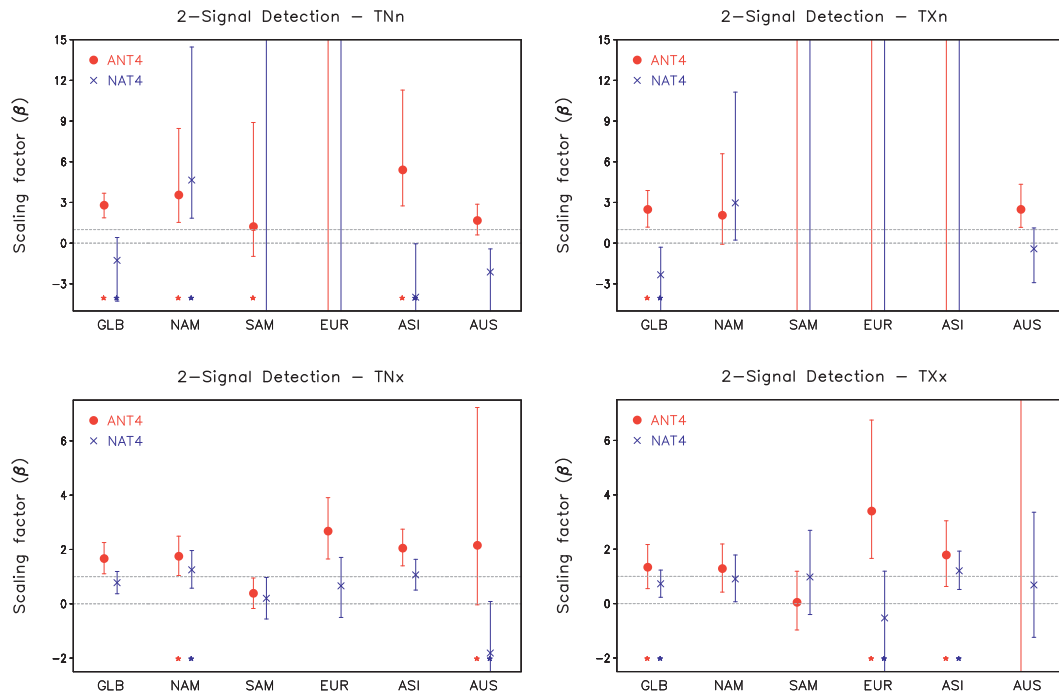


FIG. 8. As in Fig. 7, but for using the same four models, which provide both ANT and ALL simulations when estimating fingerprints [CCSM3, ECHAM5/MPI-OM, ECHO-G, and MIROC3.2(medres); refer to Table 1].

EOFs, detection results become similar to those for the larger continents (ASI and NAM), reflecting the influence of area weighting (not shown).

Single-signal results for continental domains indicate that ALL is detected for TNn over all five continents and detected for TXn over EUR and AUS and for TNn + TXn over all continents but SAM. ANT shows similar results suggesting weak influence from NAT. Detection occurs less frequently for the two-signal analyses reflecting increased noise levels in smaller scales than the global mean. Nevertheless, ANT is detected and separable from NAT over NAM and ASI, indicating some improvement from the addition of NAT. NAT is detected only in NAM. Estimated scaling factors for the detected ALL and ANT signals range from 2 to 5 on the continental scale, implying that model underestimation occurs at continental scales as in global mean for cold extremes.

Detection results for warm extremes are presented in Fig. 6. The single-signal analyses show that ALL is detected for TNx, TXx, and when both are combined (TNx + TXx) for GLB and northern continents including NAM, EUR, and ASI. ANT is detected over the same global and continental domains with better agreement with observations (i.e., scaling factors become closer to unity) but less robustly than ALL (i.e., the residual consistency test fails because of too small model variability). In the single-signal analyses, ANT overestimates

observed changes in TXx for GLB, NAM, and ASI, which is broadly consistent with recent studies based on single-signal analysis with multiple models (Zwiers et al. 2011) and with a single model (Christidis et al. 2011). Two-signal analyses show that both ANT and NAT are jointly detected for both TNx and TXx and separable from each other over GLB, NAM, EUR, and ASI, although TXx results are less robust because of too small model variability. The combined analysis of two warm extremes (TNx + TXx) provides similar results to TNx. Some improvements from the combined analysis can be noticed: scaling factors become slightly closer to unity while the residual consistency test still passes, suggesting improved agreement with observations in term of both long-term changes (signal) and internal variability (noise). Space-time analyses using GLB5 show very similar results to those from time-only analyses using GLB, which reveals negligible influence of spatial patterns of continental-mean changes in warm extremes on detection results as in the case of cold extremes. Compared to cold extremes, warm extremes, and in particular TNx provide clearer signal detection on continental scales. It is interesting to find NAM as the only continental domain where both ANT and NAT are detected for all four temperature extremes with clear signal separation as in the GLB results.

To examine the robustness of our results based on decadal means, we have conducted a sensitivity test using

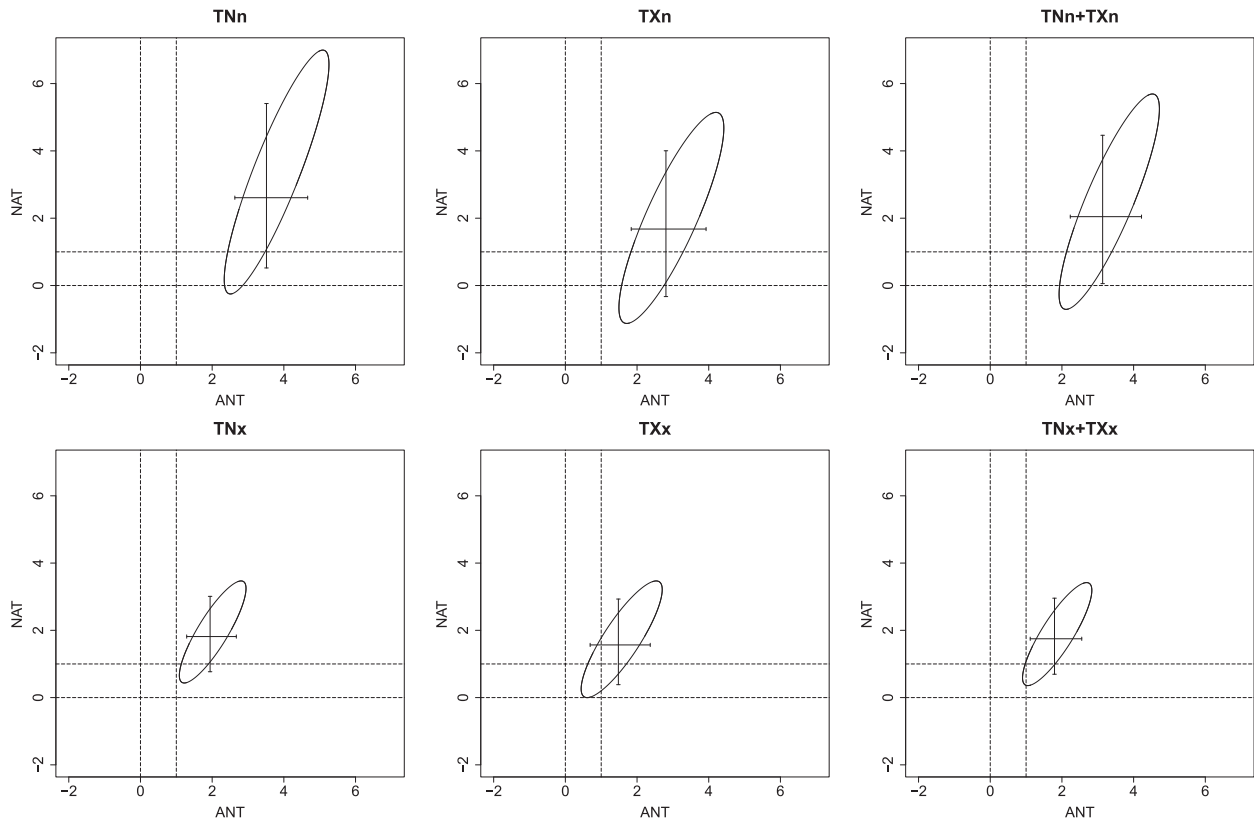


FIG. 9. The joint 90% uncertainty range for the ANT and NAT scaling factors when temperature extreme indices are regressed onto ANT ( $x$  axis) and NAT signals ( $y$  axis) simultaneously: (top) global-mean cold extremes (TNn, TXn, and TNn + TXn) and (bottom) warm extremes (TNx, TXx, and TNx + TXx). The error bars indicate one-dimensional 5%–95% ranges of the scaling factors for each forcing. The dashed horizontal/vertical lines represent zero and unity.

5-yr means. Figure 7 shows two-signal analysis results for four extreme temperature indices when using 5-yr-mean PI anomalies averaged over the globe and continental regions. Overall detection results for ANT are found to resemble those based on decadal means, indicating the robustness of our results to the dimension increase. However, there are some notable differences. NAT detection occurs less frequently and signal separation between ANT and NAT becomes more limited. Also, the residual consistency test fails more frequently than in the low-dimensional case, reflecting larger discrepancies between observed and simulated variability at shorter time scales.

The results shown in Fig. 7 are based on ANT and ALL signals estimated from all available models and thus the estimated NAT signal may be confounded with the influence of model difference (see Table 1). We therefore also test the robustness of our detection results to this model difference by redoing our analysis using the four models that provided both ANT and ALL runs [CCSM3, ECHAM5/MPI-OM, ECHO-G, and MIROC3.2(medres); Table 1]. Figure 8 shows two-signal

detection results for global- and continental-mean extreme temperature PIs obtained when using the same four models to estimate the ANT and NAT signals. Compared with the full model case (Fig. 7), the main results, including ANT signal detection and separation from NAT, are not affected much by the different model samples, suggesting insensitivity of our findings to the model difference.

Signal separation is further described by examining joint 90% uncertainty ranges for the ANT and NAT scaling factors for the GLB domain (Fig. 9). It is shown that the 90% uncertainty contours exclude the origin (0, 0) for all temperature extremes, meaning that ANT and NAT are jointly detected through two-way regression. However, when looking at one-dimensional 90% ranges of the scaling factors, for cold extremes, only ANT is detected and also model underestimation is larger by a factor of 3–4. In warm extremes, both ANT and NAT are detected and model underestimation is not as large, implying better agreement with observations in warm seasons, which may be partly related to the seasonality of volcanic cooling impact as discussed

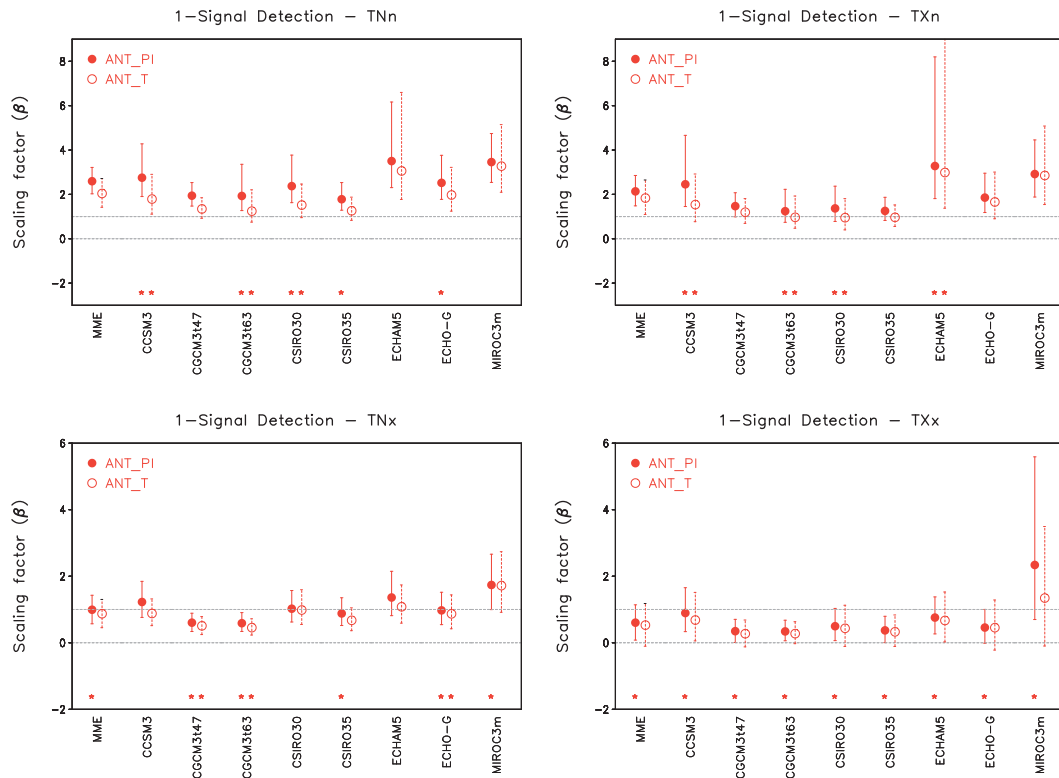


FIG. 10. As in Fig. 5, but for single-signal detection analyses for ANT using global-mean temperature extremes (TNn, TXn, TNx, and TXx) from ensemble mean of individual models as fingerprints. Results are also compared when using PI and when using absolute temperature  $T$  anomalies. MME results are as in Figs. 5 and 6 and displayed again for comparison.

above. Figure 9 also shows that the combined analysis using day and night temperatures together gives very similar results to the better case of single variable results. That is, the joint 90% uncertainty range of TNn + TXn is close to the TXn range, and the TNx + TXx result is close to the TNx case. This suggests the potential advantage of combining multiple variables, keeping noise level of internal variability low and potentially increasing signal-to-noise ratio as discussed above.

The two-signal analysis results for TXx show good agreement with the results of Christidis et al. (2011) that are based on a single model analysis: both ANT and NAT are detected and agree with observed changes (90% ranges of scaling factors include unity in both cases). However, this agreement occurs only when anomalies with respect to the mean annual cycle are used in Christidis et al. (2011). On the contrary, they found model overestimation to ANT when using absolute temperatures, which is consistent with the multimodel analysis of absolute temperatures of Zwiers et al. (2011). Different models used might cause part of this discrepancy and also our standardization procedure by the use of PI may have an impact.

To gain insight into which factor affects detection results more importantly, we repeat our detection analysis using fingerprints from individual model ensembles and also using absolute temperatures  $T$  (i.e., without standardization). In contrast to Zwiers et al. (2011), the current study considers large area means of extreme values, which should therefore be roughly Gaussian even though annual extremes at individual grid boxes will have skewed distributions that are well approximated by the GEV distribution. The choice between PI and  $T$  affects the way in which signal-to-noise ratio is optimized in optimal fingerprinting by altering relative differences in variability between different continents, and also potentially altering signal-to-noise ratios. Figures 10 and 11 illustrate best estimates and 90% confidence intervals of the scaling factors of individual models for ANT and ALL respectively, which are obtained from single-signal detection analyses for the global mean. Results from the multimodel ensemble mean (MME) are plotted again for better comparison. It is interesting to see that the use of standardization tends to slightly increase scaling factors without affecting overall detection results. It is also encouraging to see that the use of

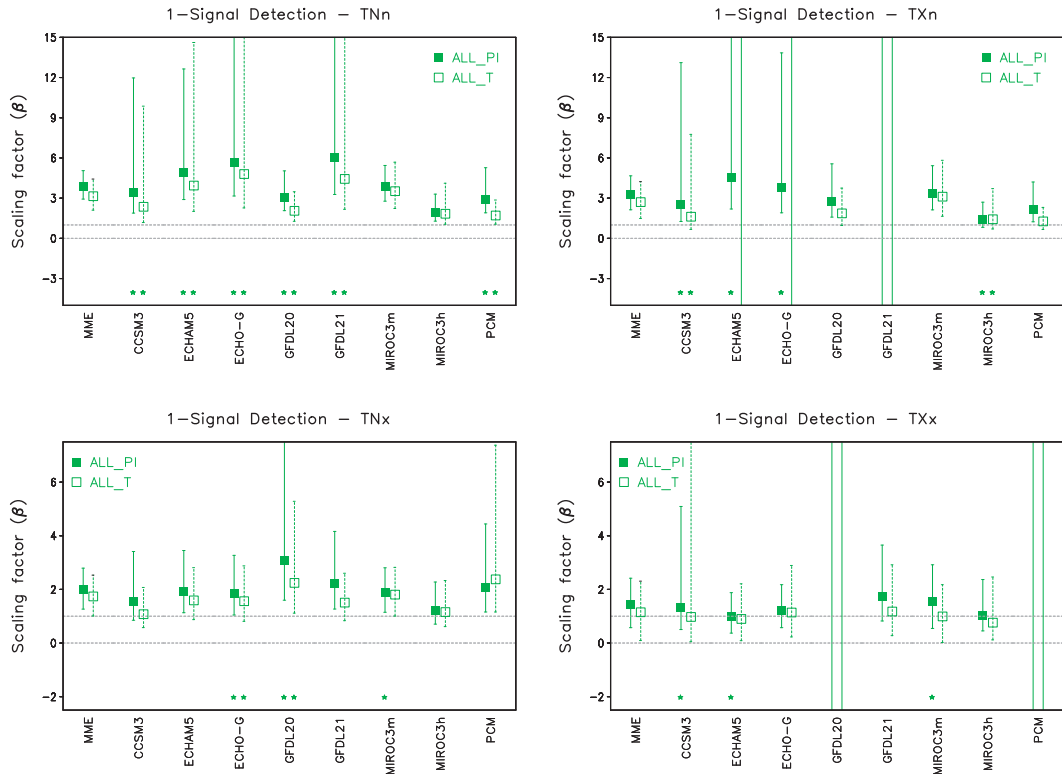


FIG. 11. As in Fig. 10, but for the ALL signal. Note that half of models are different from the ANT case (refer to Table 1).

fingerprints from individual models show largely consistent results compared to MME. However, there exist much larger differences in scaling factors among different models than those between PI and *T*. This suggests that model selection can affect detection results more strongly than standardization and that multimodel approaches are fundamentally needed in order to take account of possible influence of intermodel differences. Further, it is worth noting that best estimates of scaling factors for MIROC3.2(medres) exhibit good agreement with Shiogama et al. (2006), which applied the same fingerprinting method to the same model but considered large-scale space–time patterns rather than taking simple area averages. This additionally confirms that there would be limited influence from additional inclusion of spatial information on detection results.

Another difference from Christidis et al. (2011) is in observational data. They used the HadGHCND dataset (Caesar et al. 2006) in which daily maximum and minimum temperatures from stations were first gridded and extremes were subsequently obtained from the gridded daily data. In contrast, the HadEX observations (Alexander et al. 2006) analyzed here and in Zwiers et al. (2011) are constructed by gridding extremes obtained from stations. Good agreement between these studies indicates the robustness of anthropogenic influence on

temperature extremes to use of different observational datasets.

*c. Detection results for subcontinental regions*

We have conducted the same detection analysis for 16 subcontinental regions (defined in Table 3) for all four extremes indices and day–night combined indices. Here we present TNx results with more details, which give detection in more regions than other indices, consistent with Zwiers et al. (2011). Figure 12 shows single-signal and two-signal detection results for TNx. ALL and ANT are detected over 10 subregions mainly from North America, Europe, and Asia, and southern Australia. These regions with ANT detection are largely in accord with those found in Zwiers et al. (2011), who used different models and a different analysis period (1961–2000). This is also in good agreement with Morak et al. (2011, 2012), who detected the external signal over similar regions using the percentile-based index of the number of nights in a year with a minimum temperature exceeding the 90th percentile of the climatological range (TN90). In addition, our study finds that the detected ANT signal is separable from NAT over nine regions and that NAT is detected over five regions mainly over northern mid-latitudes, in accord with the continental results shown above. These results are also found to be insensitive to

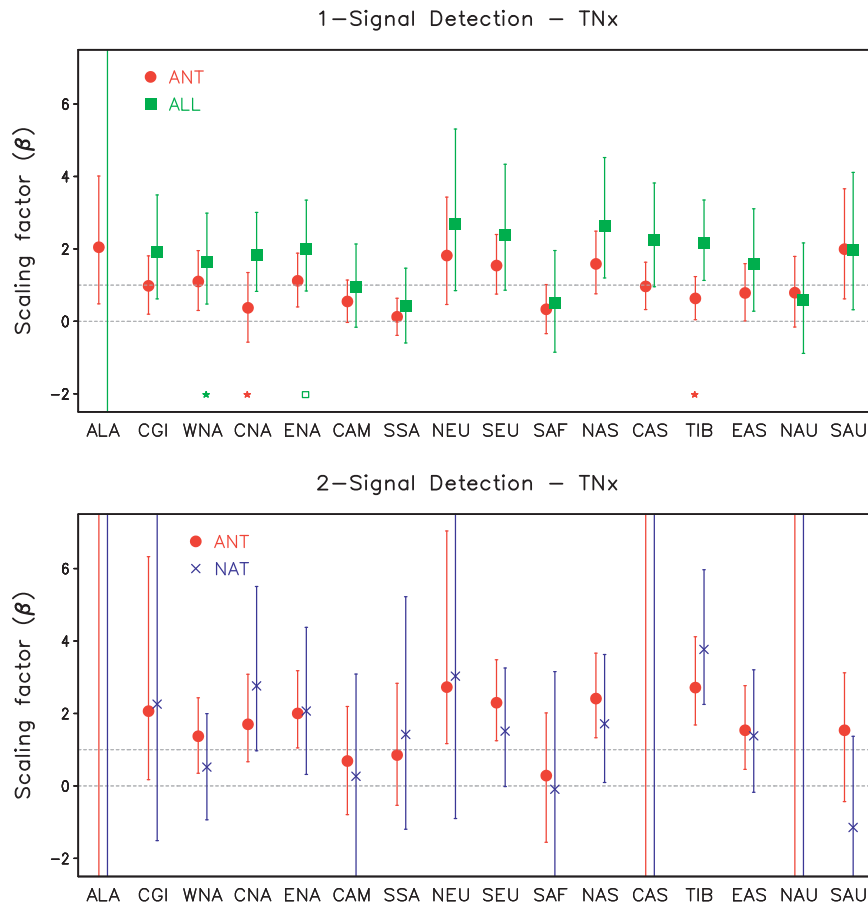


FIG. 12. As in Fig. 5, but for detection results for 16 subcontinental regions for TNx: (top) single-signal analysis and (bottom) two-signal analysis. Refer to Table 3 for the domain of subcontinental regions.

the use of 5-yr-mean PI time series and to the use of the same models to estimate ANT and ALL fingerprints as discussed above for global and continental scales (Figs. 7 and 8, respectively). Figure 13 illustrates detection results for TNx based on 5-yr means averaged over subcontinental regions. Even after excluding regions where the residual consistency test fails, single-signal analyses result in the detection of ANT in eight regions and ALL in six regions. Two-signal analyses also show that ANT is detected and separable from NAT in eight regions. Figure 14 shows the same results when using the four models that provide both ANT and ALL simulations. Results are found to be very similar to those from the full model analysis (Fig. 13), suggesting the robustness of our subcontinental-scale detection results to different model samples between ANT and ALL experiments. It is not surprising to see some less robust results with broader confidence intervals particularly in the two-signal analysis because the use of fewer simulations will generally increase fingerprint uncertainty (Gillett et al.

2002; Hegerl and Zwiers 2011). To our knowledge, this is the first separation of ANT signals from NAT in temperature extremes at continental and subcontinental scales.

Temperature extremes other than TNx provide weaker detection results, which is also consistent with global and continental analyses. Tables 4 and 5 summarize detection results for 16 subcontinental regions, where detection (D) indicates that the 90% confidence ranges of scaling factors lie above zero and attribution (A) is inferred when the confidence intervals of the detected signal include unity, implying agreement of modeled fingerprints with observations. Results from single-signal analyses (Table 4) for TNn show that ANT and ALL are detected in seven and nine subregions, respectively, over northern high-latitude areas such as Alaska (ALA), northern Europe (NEU), north Asia (NAS), and Tibet (TIB) and also over northern and southern Australia (NAU and SAU). For TNx, ANT is detected over fewer regions mainly over Asia and Australia.

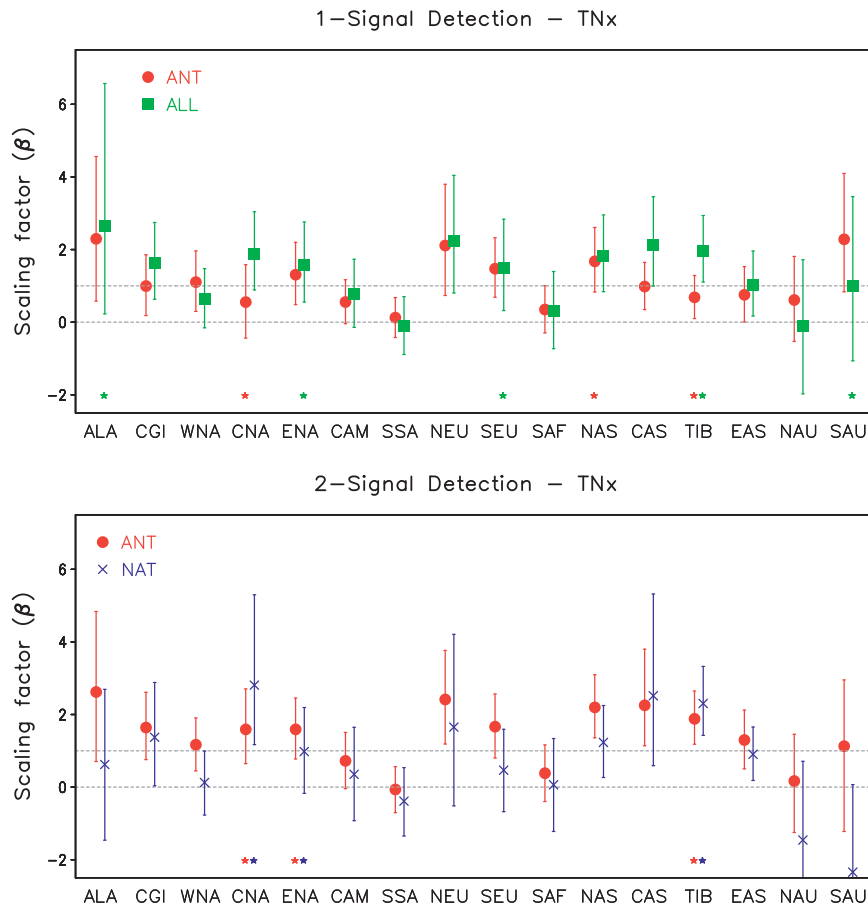


FIG. 13. As in Fig. 12, but using 5-yr-mean PI anomalies. Results are based on full-dimensional analyses (10-dimensional vector of 5-yr-mean anomalies).

Two-signal analyses (Table 5) show that ANT is detected to a more limited extent over ALA, Greenland (CGI), NEU, NAS, and East Asia (EAS) for TNn and over ALA and NAS for TXn. The external signal is rarely detected in TXx at regional scales with some exceptions over North America.

On the whole, results from subcontinental scales support those from global and continental scales. Extremes of daily minimum temperatures (TNn and TNx) give more frequent detection than extremes of daily maximum temperatures (TXn and TXx), in agreement with a stronger warming of nighttime temperature extremes than daytime temperature extremes (Donat and Alexander 2012). Another systematic difference noticeable among indices is the seasonality of model underestimation, which is stronger in cold extremes (TNn and TXn) than in warm extremes (TNx and TXx), as seen in global- and continental-scale results. This is indicated by a greater number of D's than A's in cold extremes, whereas warm extremes have more A's than D's in general (Table 4). This may be partially due to stronger

volcanic cooling in summer, which is relatively well captured by models. It should be noted that climate variability such as the North Arctic Oscillation (Brown et al. 2008; Kenyon and Hegerl 2008) may explain part of the observed stronger warming of cold extremes than multimodel means. It should also be noted that in some regions the observations are more reliable than in other regions, depending on availability of station observations.

#### 4. Conclusions

This study compares HadEX observations with CMIP3 multimodel simulations in terms of changes in temperature extremes during the latter half of twentieth century. Four annual extreme indices of coldest night (TNn) and day (TXn) and warmest night (TNx) and day (TXx) are analyzed using an optimal fingerprinting technique. For better comparisons of extremes across models and regions, we employ a standardization based on GEV distribution, which converts extreme temperatures into 0–1 scales at individual grid points before taking area

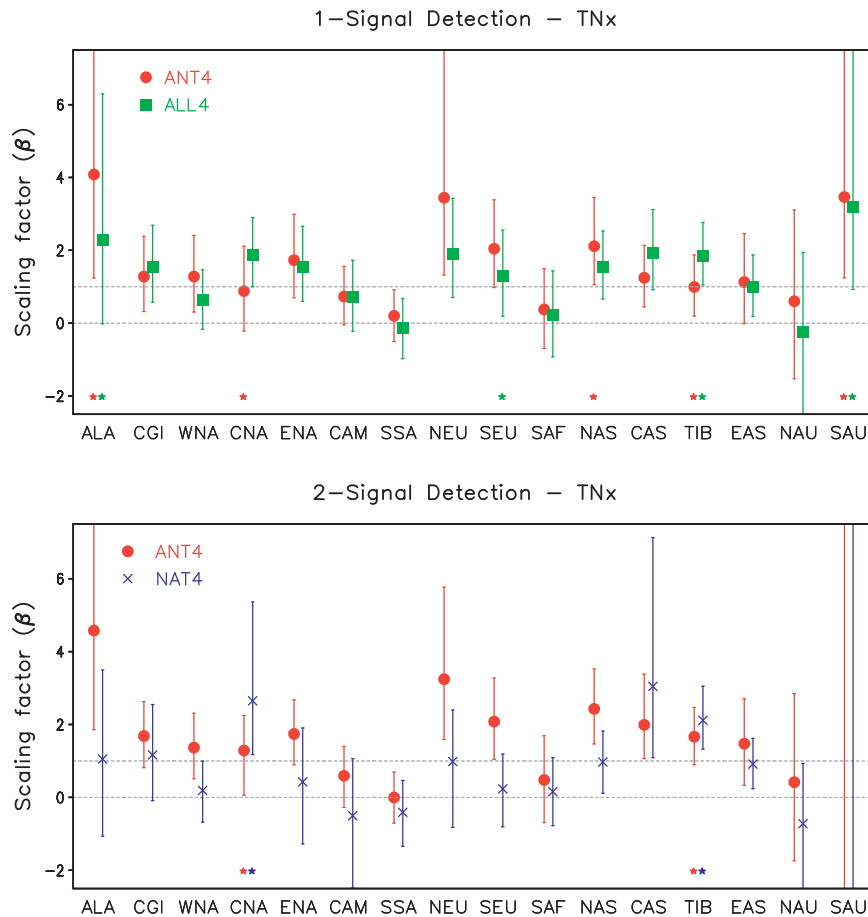


FIG. 14. As in Fig. 13, but using the same four models that provide both ANT and ALL simulations to estimate fingerprints as in Fig. 8.

averages and conducting detection analysis. We also consider different spatial scales from the global mean to 5 continental and 16 subcontinental area means, using regional domains that were selected according to observational data availability.

Results from observed global-mean (GLB) changes show that the anthropogenic (ANT) and natural plus anthropogenic (ALL) signals are detected for all four temperature extremes but less robustly for TXx, consistent with previous findings. GLB results also show that ANT is separable from the natural forcing (NAT) signal for all indices and that NAT is detected as well for warm extremes with better agreement with observed changes. Continental analyses detect ANT and ALL signals in many continents with more frequent detection over northern continents (i.e., North America, Europe, and Asia) as well as Australia. It is found that NAT forcing has contributed to changes in warm extremes over the northern continents through long-term cooling in northern summer, possibly due to volcanic

activity, from which the ANT signal is clearly separated in particular for TNx. Subcontinental-scale results show less frequent detection of external signals because of increased noise levels arising from smaller scales. However, ANT and ALL signals have been detected in TNx and TNn in 7–10 regions, and ANT is separated from NAT over several subregions, particularly for TNx. It is worth noting that the warming of cold extremes (TNn and TXn) is underestimated by the models on global to subcontinental scales in a systematic manner. Further investigation is needed to understand this observation–model discrepancy.

Our results support previous findings based on different observations and models, different periods, and different methods (Christidis et al. 2005, 2011; Zwiers et al. 2011; Morak et al. 2011, 2012). Our study is the first multimodel optimal detection analysis that shows ANT signal separation from NAT at continental and subcontinental scales. We also conducted a combined attribution analysis by considering daytime and nighttime

TABLE 4. Single-signal analysis results for 16 subcontinental areas. Here, “D” represents detection and “A” means attribution (see text for definition). Number of regions with signal detection is given for each extreme temperature indices or combined indices. Bold and nonbold fonts indicate ANT and ALL signals, respectively.

Continent	Subregion	TNn		TXn		TNn + TXn		TNx		TXx		TNx + TXx	
North America	ALA	<b>D</b>	D	<b>D</b>		<b>D</b>		<b>A</b>		<b>A</b>		<b>A</b>	
	CGI		A					<b>A</b>	A	<b>A</b>	A	<b>A</b>	A
	WNA		D			<b>A</b>	D	<b>A</b>		<b>A</b>		<b>A</b>	A
	CNA								A				A
	ENA							<b>A</b>	A				A
South America	CAM		D		D		D				A	<b>A</b>	A
	SSA												
Europe	NEU	<b>A</b>	A					<b>A</b>	A		A	<b>A</b>	A
	SEU			<b>D</b>				<b>A</b>	A	<b>A</b>		<b>A</b>	A
Africa	SAF												
Asia	NAS	<b>D</b>	D	<b>D</b>	D	<b>D</b>	D	<b>A</b>	D			<b>A</b>	A
	CAS				D			<b>A</b>	A			<b>A</b>	A
	TIB	<b>D</b>	D	<b>D</b>		<b>D</b>			D				D
	EAS	<b>D</b>		<b>D</b>		<b>D</b>		<b>A</b>	A				
	NAU	<b>A</b>	A	<b>A</b>		<b>D</b>	D						
Australia	SAU	<b>A</b>	D	<b>A</b>		<b>A</b>	D	<b>A</b>	A				
Counts		<b>7</b>	9	<b>7</b>	3	<b>7</b>	5	<b>10</b>	10	<b>3</b>	3	<b>8</b>	10

temperature extreme indices simultaneously. In general, combined attribution shows similar results to single variable analyses and suggests improvements in terms of internal variability as well as long-term mean changes. It is also found that additional consideration of spatial information does not affect detection results probably due to lack of intercontinental difference in changes in extreme temperatures.

It should be noted that there remain large uncertainties particularly in the regional detection results. There can be influences from missing processes in the global models such as land-cover changes (Portmann

et al. 2009; Avila et al. 2012), which requires further investigation with relevant model experiments. Climate variability is known to exert significant influence on interannual and interdecadal variations of temperature extremes depending on regions and seasons (Brown et al. 2008; Kenyon and Hegerl 2008; Alexander et al. 2009). The possible influence of climate variability on detection and attribution therefore needs to be further examined through sensitivity tests. These issues will be examined in a separate study with the use of updated observations and newly available multimodel datasets for the extended analysis period.

TABLE 5. As in Table 4, but for results from two-signal analysis. Bold and italic fonts represent ANT and NAT signals, respectively.

Continent	Subregion	TNn		TXn		TNn + TXn		TNx		TXx		TNx + TXx	
North America	ALA	<b>D</b>		<b>A</b>		<b>D</b>							
	CGI	<b>A</b>	<i>A</i>			<b>A</b>	<i>A</i>	<b>A</b>					
	WNA							<b>A</b>				<b>A</b>	
	CNA				<i>A</i>			<b>A</b>	<i>A</i>		<i>A</i>	<b>A</b>	<i>A</i>
	ENA							<b>A</b>	<i>A</i>		<i>A</i>	<b>A</b>	<i>A</i>
South America	CAM									<b>A</b>			
	SSA												
Europe	NEU	<b>A</b>				<b>A</b>		<b>D</b>				<b>D</b>	
	SEU							<b>D</b>	<i>A</i>			<b>A</b>	
Africa	SAF												
Asia	NAS	<b>D</b>		<b>D</b>				<b>D</b>	<i>A</i>			<b>A</b>	
	CAS												
	TIB							<b>D</b>	<i>D</i>			<b>D</b>	<i>D</i>
	EAS	<b>A</b>						<b>A</b>					
	NAU					<b>A</b>							
Australia	SAU												
Counts		<b>5</b>	<i>1</i>	<b>2</b>	<i>1</i>	<b>4</b>	<i>1</i>	<b>9</b>	5	<b>1</b>	<i>1</i>	<b>6</b>	<i>3</i>

**Acknowledgments.** We are grateful to two anonymous reviewers for their constructive comments. We also thank Jonas Bhend and Penny Whetton for useful comments. We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multimodel dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy. This study is in part supported by the Goyder Institute for Water Research.

## REFERENCES

- Alexander, L. V., and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.*, **111**, D05109, doi:10.1029/2005JD006290.
- , P. Uotila, and N. Nicholls, 2009: Influence of sea surface temperature variability on global temperature and precipitation extremes. *J. Geophys. Res.*, **114**, D18116, doi:10.1029/2009JD012301.
- Allen, M. R., and S. F. B. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Climate Dyn.*, **15**, 419–434.
- , and P. A. Stott, 2003: Estimating signal amplitudes in optimal fingerprinting, part I: Theory. *Climate Dyn.*, **21**, 477–491.
- Avila, F. B., A. J. Pitman, M. G. Donat, L. V. Alexander, and G. Abramowitz, 2012: Climate model simulated changes in temperature extremes due to land cover change. *J. Geophys. Res.*, **117**, D04108, doi:10.1029/2011JD016382.
- Barnett, T. P., and Coauthors, 2008: Human-induced changes in the hydrology of the western United States. *Science*, **319**, 1080–1083.
- Briffa, K. R., P. D. Jones, F. H. Schweingruber, and T. J. Osborn, 1998: Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years. *Nature*, **393**, 450–455.
- Brown, S. J., J. Caesar, and C. A. T. Ferro, 2008: Global changes in extreme daily temperature since 1950. *J. Geophys. Res.*, **113**, D05115, doi:10.1029/2006JD008091.
- Caesar, J., L. Alexander, and R. Vose, 2006: Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *J. Geophys. Res.*, **111**, D05101, doi:10.1029/2005JD006280.
- Christidis, N., P. A. Stott, S. Brown, G. C. Hegerl, and J. Caesar, 2005: Detection of changes in temperature extremes during the second half of the 20th century. *Geophys. Res. Lett.*, **32**, L20716, doi:10.1029/2005GL023885.
- , —, F. W. Zwiers, H. Shioyama, and T. Nozawa, 2010: Probabilistic estimates of recent changes in temperature: A multi-scale attribution analysis. *Climate Dyn.*, **34**, 1139–1156.
- , —, and S. J. Brown, 2011: The role of human activity in the recent warming of extremely warm daytime temperatures. *J. Climate*, **24**, 1922–1930.
- Coles, S. G., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, 225 pp.
- Donat, M. G., and L. V. Alexander, 2012: The shifting probability distribution of global daytime and nighttime temperatures. *Geophys. Res. Lett.*, **39**, L14707, doi:10.1029/2012GL052459.
- , and Coauthors, 2013: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *J. Geophys. Res. Atmos.*, **118**, 2098–2118, doi:10.1002/jgrd.50150.
- Driscoll, S., A. Bozzo, L. J. Gray, A. Robock, and G. Stenchikov, 2012: Coupled Model Intercomparison Project 5 (CMIP5) simulations of climate following volcanic eruptions. *J. Geophys. Res.*, **117**, D17105, doi:10.1029/2012JD017607.
- Frich, P., L. V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. Tank, and T. Peterson, 2002: Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Res.*, **19**, 193–212.
- Gillett, N. P., F. W. Zwiers, A. J. Weaver, G. C. Hegerl, M. R. Allen, and P. A. Stott, 2002: Detecting anthropogenic influence with a multi-model ensemble. *Geophys. Res. Lett.*, **29**, 1970, doi:10.1029/2002GL015836.
- Giorgi, F., and R. Francisco, 2000: Uncertainties in regional climate change prediction: a regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. *Climate Dyn.*, **16**, 169–182.
- , and Coauthors, 2001: Emerging patterns of simulated regional climatic changes for the 21st century due to anthropogenic forcings. *Geophys. Res. Lett.*, **28**, 3317–3320.
- Hegerl, G. C., and F. W. Zwiers, 2011: Use of models in detection and attribution of climate change. *Wiley Interdiscip. Rev. Climatic Change*, **2**, 570–591.
- , —, P. A. Stott, and V. V. Kharin, 2004: Detectability of anthropogenic changes in annual temperature and precipitation extremes. *J. Climate*, **17**, 3683–3700.
- , and Coauthors, 2007: Understanding and attributing climate change. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 663–745.
- Huntingford, C., P. A. Stott, M. R. Allen, and F. H. Lambert, 2006: Incorporating model uncertainty into attribution of observed temperature change. *Geophys. Res. Lett.*, **33**, L05710, doi:10.1029/2005GL024831.
- Jones, G. S., S. F. B. Tett, and P. A. Stott, 2003: Causes of atmospheric temperature change 1960–2000: A combined attribution analysis. *Geophys. Res. Lett.*, **30**, 1228, doi:10.1029/2002GL016377.
- Kenyon, J., and G. C. Hegerl, 2008: Influence of modes of climate variability on global temperature extremes. *J. Climate*, **21**, 3872–3889.
- Kharin, V. V., and F. W. Zwiers, 2005: Estimating extremes in transient climate change simulations. *J. Climate*, **18**, 1156–1173.
- , —, and X. B. Zhang, 2005: Intercomparison of near-surface temperature and precipitation extremes in AMIP-2 simulations, reanalyses, and observations. *J. Climate*, **18**, 5201–5223.
- , —, —, and G. C. Hegerl, 2007: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Climate*, **20**, 1419–1444.
- Meehl, G. A., C. Covey, K. E. Taylor, T. Delworth, R. J. Stouffer, M. Latif, B. McAvaney, and J. F. B. Mitchell, 2007a: The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, **88**, 1383–1394.
- , and Coauthors, 2007b: Global climate projections. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 747–845.
- Min, S. K., X. B. Zhang, F. W. Zwiers, P. Friederichs, and A. Hense, 2009: Signal detectability in extreme precipitation changes assessed from twentieth century climate simulations. *Climate Dyn.*, **32**, 95–111.
- , —, —, and G. C. Hegerl, 2011: Human contribution to more-intense precipitation extremes. *Nature*, **470**, 378–381.
- Morak, S., G. C. Hegerl, and J. Kenyon, 2011: Detectable regional changes in the number of warm nights. *Geophys. Res. Lett.*, **38**, L17703, doi:10.1029/2011GL048531.

- , —, and N. Christidis, 2012: Detectable changes in the frequency of temperature extremes. *J. Climate*, **26**, 1561–1574.
- Portmann, R. W., S. Solomon, and G. C. Hegerl, 2009: Spatial and seasonal patterns in climate change, temperatures, and precipitation across the United States. *Proc. Natl. Acad. Sci. USA*, **106**, 7324–7329.
- Robock, A., 2000: Volcanic eruptions and climate. *Rev. Geophys.*, **38**, 191–219.
- Schnur, R., and K. I. Hasselmann, 2005: Optimal filtering for Bayesian detection and attribution of climate change. *Climate Dyn.*, **24**, 45–55.
- Shiogama, H., N. Christidis, J. Caesar, T. Yokohata, T. Nozawa, and S. Emori, 2006: Detection of greenhouse gas and aerosol influences on changes in temperature extremes. *SOLA*, **2**, 152–155.
- Stenchikov, G., K. Hamilton, R. J. Stouffer, A. Robock, V. Ramaswamy, B. Santer, and H.-F. Graf, 2006: Arctic Oscillation response to volcanic eruptions in the IPCC AR4 climate models. *J. Geophys. Res.*, **111**, D07107, doi:10.1029/2005JD006286.
- Stone, D. A., and A. J. Weaver, 2003: Factors contributing to diurnal temperature range trends in twentieth and twenty-first century simulations of the CCCma coupled model. *Climate Dyn.*, **20**, 435–445.
- Stott, P. A., N. P. Gillett, G. C. Hegerl, D. J. Karoly, D. A. Stone, X. Zhang, and F. Zwiers, 2010: Detection and attribution of climate change: A regional perspective. *Wiley Interdiscip. Rev. Climate Change*, **1**, 192–211.
- Zhang, X. B., J. F. Wang, F. W. Zwiers, and P. Ya. Groisman, 2010: The influence of large-scale climate variability on winter maximum daily precipitation over North America. *J. Climate*, **23**, 2902–2915.
- Zhou, L. M., R. E. Dickinson, A. G. Dai, and P. Dirmeyer, 2010: Detection and attribution of anthropogenic forcing to diurnal temperature range changes from 1950 to 1999: Comparing multi-model simulations with observations. *Climate Dyn.*, **35**, 1289–1307.
- Zwiers, F. W., X. B. Zhang, and Y. Feng, 2011: Anthropogenic influence on long return period daily temperature extremes at regional scales. *J. Climate*, **24**, 881–892.