



Detection and Attribution of Observed Changes in Northern Hemisphere Spring Snow Cover

DAVID E. RUPP AND PHILIP W. MOTE

Oregon Climate Change Research Institute, College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon

NATHANIEL L. BINDOFF

Institute of Marine and Antarctic Studies, and CSIRO Marine and Atmospheric Research, University of Tasmania, Hobart, Tasmania, Australia

PETER A. STOTT

Met Office Hadley Centre, Exeter, United Kingdom

DAVID A. ROBINSON

Department of Geography, Rutgers, The State University of New Jersey, New Brunswick, New Jersey

(Manuscript received 31 July 2012, in final form 19 February 2013)

ABSTRACT

Significant declines in spring Northern Hemisphere (NH) snow cover extent (SCE) have been observed over the last five decades. As one step toward understanding the causes of this decline, an optimal fingerprinting technique is used to look for consistency in the temporal pattern of spring NH SCE between observations and simulations from 15 global climate models (GCMs) that form part of phase 5 of the Coupled Model Intercomparison Project. The authors examined simulations from 15 GCMs that included both natural and anthropogenic forcing and simulations from 7 GCMs that included only natural forcing. The decline in observed NH SCE could be largely explained by the combined natural and anthropogenic forcing but not by natural forcing alone. However, the 15 GCMs, taken as a whole, underpredicted the combined forcing response by a factor of 2. How much of this underprediction was due to underrepresentation of the sensitivity to external forcing of the GCMs or to their underrepresentation of internal variability has yet to be determined.

1. Introduction

The accumulation of greenhouse gases in the atmosphere has the potential to substantially alter the earth's climate in profound ways. Understanding the unfolding details of those changes poses an important scientific and societal challenge, especially for those details of climate that directly affect human well-being. A research approach called "detection and attribution" compares observed and modeled changes in one or more climate

variables over some past period in order to determine rigorously whether the changes are (i) outside of natural variability and (ii) consistent with human modification of the atmosphere through the addition of greenhouse gases (and often sulfate aerosol).

Detection and attribution studies (beginning with Hasselmann 1993) initially focused on global temperature means and patterns (Hegerl et al. 1997; Stott et al. 2000), in both longitude–latitude and latitude–altitude (Stott et al. 2001; Santer et al. 2003). Subsequently, detection and attribution has been applied to zonal mean precipitation patterns (e.g., Lambert et al. 2005), surface temperature extremes (e.g., Tebaldi et al. 2006; Stott et al. 2011), ocean heat content (Barnett et al. 2005), Arctic sea ice (Min et al. 2008), western U.S. hydroclimate (Barnett

Corresponding author address: David E. Rupp, Oregon Climate Change Research Institute, Oregon State University, 326 Strand Hall, Corvallis, OR 97330.
E-mail: david.rupp@oregonstate.edu

et al. 2008; Pierce et al. 2008), northern and southern annular modes (Gillett et al. 2005), and more. A related approach, fractional attributable risk, has been applied to specific extreme events like the 2003 European heat wave (e.g., Stott et al. 2004). Other important aspects of global climate, for example, the El Niño–Southern Oscillation and tropical cyclones, have not exhibited a detectable change (Hegerl et al. 2007).

Northern Hemisphere (NH) snow cover extent (SCE) is among the most important indicators of global climate variability and change. An increase in global temperature should cause a decline in total snow cover extent given the 0°C-threshold response of snow formation and melt. However, increases in winter precipitation could be sufficient to offset this response regionally, particularly at high latitudes and elevations (Brown and Mote 2009). Of course, a decrease in winter precipitation would also bring less snow, temperature being equal.

To date, no detection and attribution studies of spring NH SCE have been published. Brown and Mote (2009) did compare satellite-derived snow cover patterns with simulated snow cover from global climate models (GCMs) for the years 1970–99 but did not compare the temporal trends. Very recently, Derksen and Brown (2012) showed that late spring–early summer (May–June) NH snow cover, which is predominantly over the Arctic, decreased significantly over the last four decades and that the decrease in June SCE was greater than that simulated by an ensemble of eight GCMs in phase 5 of the Coupled Model Intercomparison Project (CMIP5). In contrast, our study compares 90 years of observed snow cover in March–April (when the snow cover extent is much greater) to simulations from 15 CMIP5 GCMs and performs a formal detection and attribution exercise known as optimal fingerprinting.

2. Data

Our source for observed SCE was March and April snow covered area for the Northern Hemisphere over the period from 1922 to 2010 from Brown (2000) and Brown and Robinson (2011). Simulated monthly NH SCE was obtained from the CMIP5 multimodel ensemble (Taylor et al. 2012). We acquired SCE for two CMIP5 experiments for the period 1850–2005. The first experiment, “historicalNat,” used natural external forcings only, which include solar irradiance and volcanic gases. The second experiment, “historical,” used both natural and anthropogenic forcing; the latter includes long-lived greenhouse gases, aerosols and chemically active gases, though not all models include the identical suite of anthropogenic forcing agents. Simulated monthly SCE that excluded any time-varying forcing came from long-duration

runs under the CMIP5 preindustrial control experiment “piControl.”

We used the following criteria for selecting a subset of the available (as of 1 December 2012) CMIP5 datasets for this analysis: 1) preindustrial control had to have lengths of at least 500 continuous years and 2) a GCM had to have at least three ensemble members for a historicalNat or historical scenario (ensemble members for a given GCM vary by their initial conditions only). A total of 30 historicalNat runs from 7 GCMs and 67 historical runs from 15 GCMs were used for testing whether the observed decline in SCE could be explained by either forcing. A total of 15 preindustrial simulations (1 per model) were used representing 7500 years. These preindustrial simulations were used to estimate internal variability (see section 3). Table 1 provides a summary of the selected historicalNat and historical CMIP5 datasets, and the time series of anomalies are shown in Figs. 1 and 2, respectively.

An initial examination of the CMIP5 historical datasets revealed climatological biases in spring NH SCE ranging from –34% (MPI-ESM-MR) to +22% (CNRM-CM5), with BCC-CSM1.1 showing the least bias (+0.2%). Given that a GCM’s bias could influence its sensitivity in SCE to external forcing, we performed the optimal fingerprint analysis on both absolute and relative anomalies.

Prior to the optimal fingerprint analysis, all observed and simulated time series were subjected to a Gaussian smoother to reduce the higher frequency variability, which is noise in our case, in favor of the multidecadal scale response to forcings that is of interest here. We applied the smoother with two values (1.5 and 2.5 yr) of the standard deviation of the Gaussian kernel to examine the sensitivity of the detection and attribution methodology to a particular amount of smoothing.

3. Methodology

The method of detection and attribution known as optimal fingerprinting is discussed in detail elsewhere (e.g., Allen and Tett 1999; Allen and Stott 2003), so we merely give a brief overview here. The method can be presented in the form of “total” least squares (TLS) regression,

$$\mathbf{y} = (\mathbf{x} - \nu_1)\beta + \nu_0, \quad (1)$$

where \mathbf{y} is the filtered version of the observed temporal, or spatiotemporal, pattern; \mathbf{x} contains the filtered simulated response pattern to the external forcings; β is the unknown scaling factor; and ν_0 represents the internal climate variability or noise. The additional noise term ν_1

TABLE 1. Summary of CMIP5 historical datasets used in the analysis.

GCM*	Model	Institution	No. of ensemble members by expt	
			historical	historicalNat
BCC-CSM1.1	Beijing Climate Center, Climate System Model, version 1.1	Beijing Climate Center, China Meteorological Administration	3	0
CanESM2	Second Generation Canadian Earth System Model	Canadian Centre for Climate Modelling and Analysis	5	5
CCSM4.0	Community Climate System Model, version 4.0	National Center of Atmospheric Research, United States	6	4
CNRM-CM5	Centre National de Recherches Météorologiques Coupled Global Climate Model, version 5	National Centre of Meteorological Research, France	6	6
CSIRO Mk3.6.0	Commonwealth Scientific and Industrial Research Organisation Mark, version 3.6.0	Commonwealth Scientific and Industrial Research Organization/Queensland Climate Change Centre of Excellence, Australia	10	5
FGOALS-g2	Flexible Global Ocean–Atmosphere–Land System Model gridpoint, version 2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences	4	3
FIO-ESM	First Institute of Oceanography (FIO) Earth System Model (ESM)	The First Institute of Oceanography, State Oceanic Administration (SOA), China	3	0
GISS-E2-H	Goddard Institute for Space Studies Model E, version 2, coupled with the HYCOM ocean model	NASA Goddard Institute for Space Studies, United States	5	0
GISS-E2-R	Goddard Institute for Space Studies Model E, version 2, coupled with Russell ocean model	NASA Goddard Institute for Space Studies, United States	5	4
MIROC-ESM	Model for Interdisciplinary Research on Climate, Earth System Model	Japan Agency for Marine–Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	3	3
MIROC5	Model for Interdisciplinary Research on Climate, version 5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine–Earth Science and Technology	5	0
MPI-ESM-LR	Max Planck Institute Earth System Model, low resolution	Max Planck Institute for Meteorology, Germany	3	0
MPI-ESM-MR	Max Planck Institute Earth System Model, medium resolution	Max Planck Institute for Meteorology, Germany	3	0
MRI-CGCM3	Meteorological Research Institute Coupled Atmosphere–Ocean General Circulation Model, version 3	Meteorological Research Institute, Japan	3	0
NorESM1-M	Norwegian Earth System Model, version 1 (intermediate resolution)	Norwegian Climate Centre, Norway	3	0

* For all models, the first version of the perturbed physics model (denoted in CMIP5 as “p1”) is used except for GISS-E2-H and GISS-E2-R, where versions “p2” and “p3” are used, respectively.

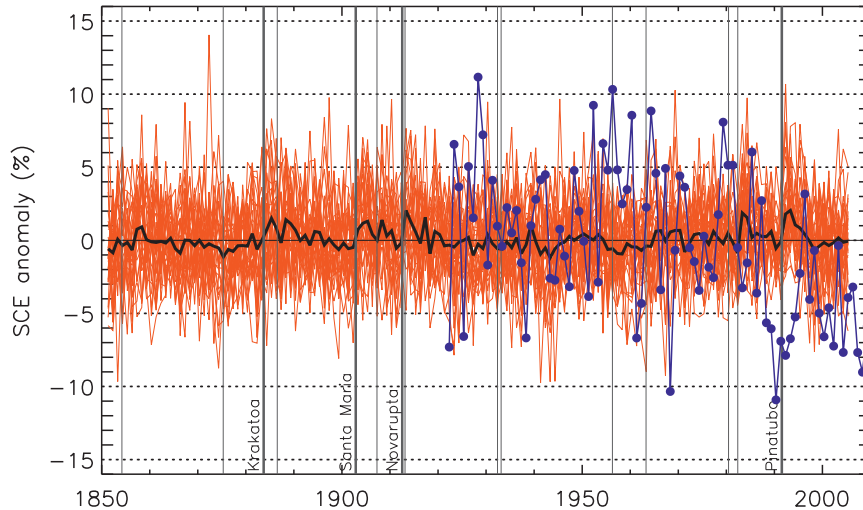


FIG. 1. Observed and simulated spring (average of March and April) Northern Hemisphere snow cover extent (SCE) anomalies from the 1922–2005 baseline period. Observed anomalies are shown by the blue symbols. The red lines show 30 simulations from seven models under natural forcing only (“historicalNat” experiment), and the thick black line is the full ensemble average. Vertical gray lines denote major volcanic events (thin and thick lines signify volcanic explosivity index of 5 and 6, respectively).

accounts for the fact that we are using a finite ensemble of model-simulated response \mathbf{x} ; thus, \mathbf{x} differs by a quantity v_1 from the underlying response from a hypothetical infinite ensemble. Given that our ensemble sizes were mostly small (as few as three members per model) it would be improper to assume v_1 is negligible (Allen and Stott 2003; Stott et al. 2003).

Representing the problem as a linear regression with the assumption that the noise terms are normally distributed allows for the calculation of a best estimate $\tilde{\beta}$ of

the true scaling factor β , as well as confidence intervals on $\tilde{\beta}$, based on established techniques (Allen and Stott 2003). While the noise terms are Gaussian, they include strong correlations in time and consequently are often highly correlated with the signal itself. This correlation between internal variability and forced response is the challenge in detecting and attributing the observed changes to external forcing. From these estimates of the scaling factors $\tilde{\beta}$, we performed a test of amplitude consistency (Tett et al. 2002). Amplitude consistency

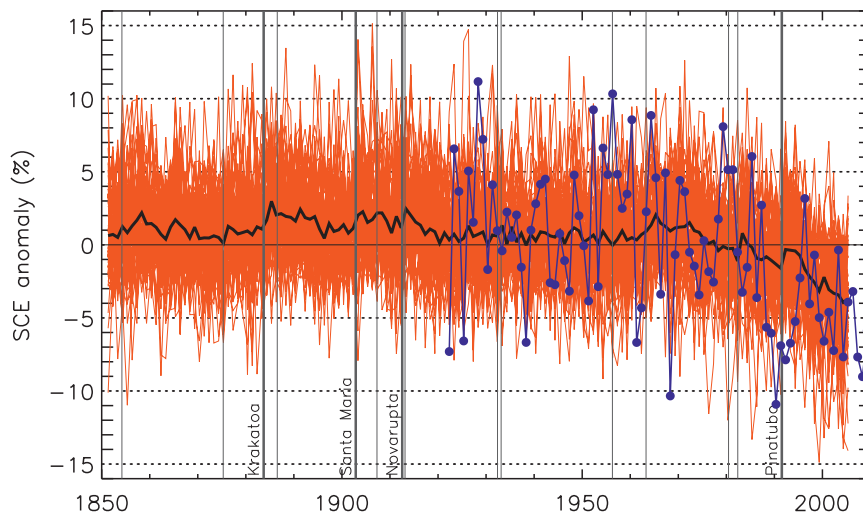


FIG. 2. As in Fig. 1, but for the 67 simulations from 15 models using natural and anthropogenic forcing (historical experiment).

tests the null hypothesis that the observed response to forcing is consistent with the amplitude of the simulated response: that is, that the two-tailed uncertainty range about $\tilde{\beta}$ includes unity. Rejection of the null hypothesis implies that the simulated signal amplitude is inconsistent with the observations.

The noise v_0 was extracted from samples of the control simulations (Allen and Tett 1999) where the samples were the same length as that of the observations. Samples of the control simulations were offset by 10 yr to increase the number of samples available for the analysis (e.g., Stott et al. 2001), resulting in 41 control samples of 84 yr each per GCM. Overlapping samples was justified because the time series, including the Gaussian-smoothed series, show essentially no autocorrelation at lags of 10 yr and greater. Even so, the use of overlapping samples requires an adjustment to the degrees of freedom for significance testing. As in Stott et al. (2001) and Tett et al. (2002), the degrees of freedom were given as 1.5 times the number of nonoverlapping samples. The control samples were divided into two sets: one set of 21 samples was used for the “prewhitening” operator, which was used to produce the optimized fingerprints from the SCE anomaly components, and the second set of 20 samples was used for the model consistency test (Allen and Tett 1999).

As stated above, both the observations and simulations were filtered prior to fitting Eq. (1): note that this filtering is distinct from the smoothing of the time series. Filtering was done in order to provide a robust estimate of the inverse noise covariance, which is needed in order to normalize model-simulated responses and observations by internal variability so as to improve the signal to noise ratio (Allen and Tett, 1999). Filtering was accomplished through an empirical orthogonal function (EOF) analysis. The EOFs were generated from the preindustrial control samples, which were treated as “pseudo” observations of internal variability. Our EOF analysis contained N variables, each being spring NH SCE z_i at points in time $i = 1, 2, \dots, N$, for $N = 84$ (1922–2005). For each z_i , there were $M = 21$ samples of pseudo observations. (Note that this differs from the more familiar way of conducting an EOF analysis where z_i has position i in space and the samples are taken through time.) All data (observed and simulated) were then projected onto the leading κ eigenvectors, for $\kappa < M$. We looked for truncations (κ) that provided consistency with the linear statistical model in Eq. (1). The linear model was accepted when an F test of the ratio of the variance of the residuals of the regression and the variance of the control simulation returned p values that exceeded 0.05 and did not exceed 0.95 (Allen and Tett 1999). The upper limit of 0.95 was used as a check

against overfitting. In effect, we chose κ that returned the highest p value under 0.95.

In addition to applying Eq. (1) to each GCM separately, we also applied it to the multimodel ensemble (e.g., Gillett et al. 2002). For simplicity, we treated all ensemble members of each model equally: thus, individual models were, in effect, weighted by the number of ensemble members each model contained. The number of control samples was the number of models multiplied by M .

4. Results and discussion

The records of observed spring NH snow cover extent (SCE) show a decline over the last 90 years, with most of this decline occurring during the last 40 years. From 1970 to 2010, the rate of decline was $\sim 0.8 \times 10^6 \text{ km}^2$ per decade (Brown and Robinson 2011). Putting these numbers in perspective that is equivalent to a loss of one-third of the area of Canada, or a 9% reduction, from average pre-1970 values.

Under natural forcing only, the GCMs as a group did not simulate any such multidecadal decline in spring NH SCE (Fig. 1). The most evident naturally forced response was from volcanic activity; in the multimodel average, SCE increased by a little over 2% following each of the four eruptions since 1850 that had the largest volcanic explosivity index (6). However, the volcanic influence appeared to become negligible by the third or fourth year following the eruption. Unfortunately (i.e., from a statistical standpoint), only one of these eruptions (Mount Pinatubo, June 1991) has occurred since 1922, and the observed SCE record actually shows a small decrease in the first spring following the Pinatubo eruption. The large year-to-year variability in observed spring NH SCE relative to the apparent simulated response to volcanic forcing necessitates a much larger sample size to permit a satisfactory statistical analysis of the observed SCE response to volcanic emissions.

Another important natural forcing agent, total solar irradiance (TSI), has a dominant mode of variability with an approximately 11-yr cycle. In addition to this solar cycle, TSI reconstructions also show a gradual increase from the late nineteenth century to about 1960 (e.g., Lean and Rind 2008). Though both of these features of the TSI time series are included in the CMIP5 historicalNat experiment, neither produced a signal that is visually evident in the multimodel ensemble (Fig. 1).

In contrast to the CMIP5 historicalNat experiment, the simulations that included anthropogenic forcing show a clear decline in snow cover since 1970 (Fig. 2). The two largest short-term reversals of this longer trend in the simulations followed large volcanic eruptions

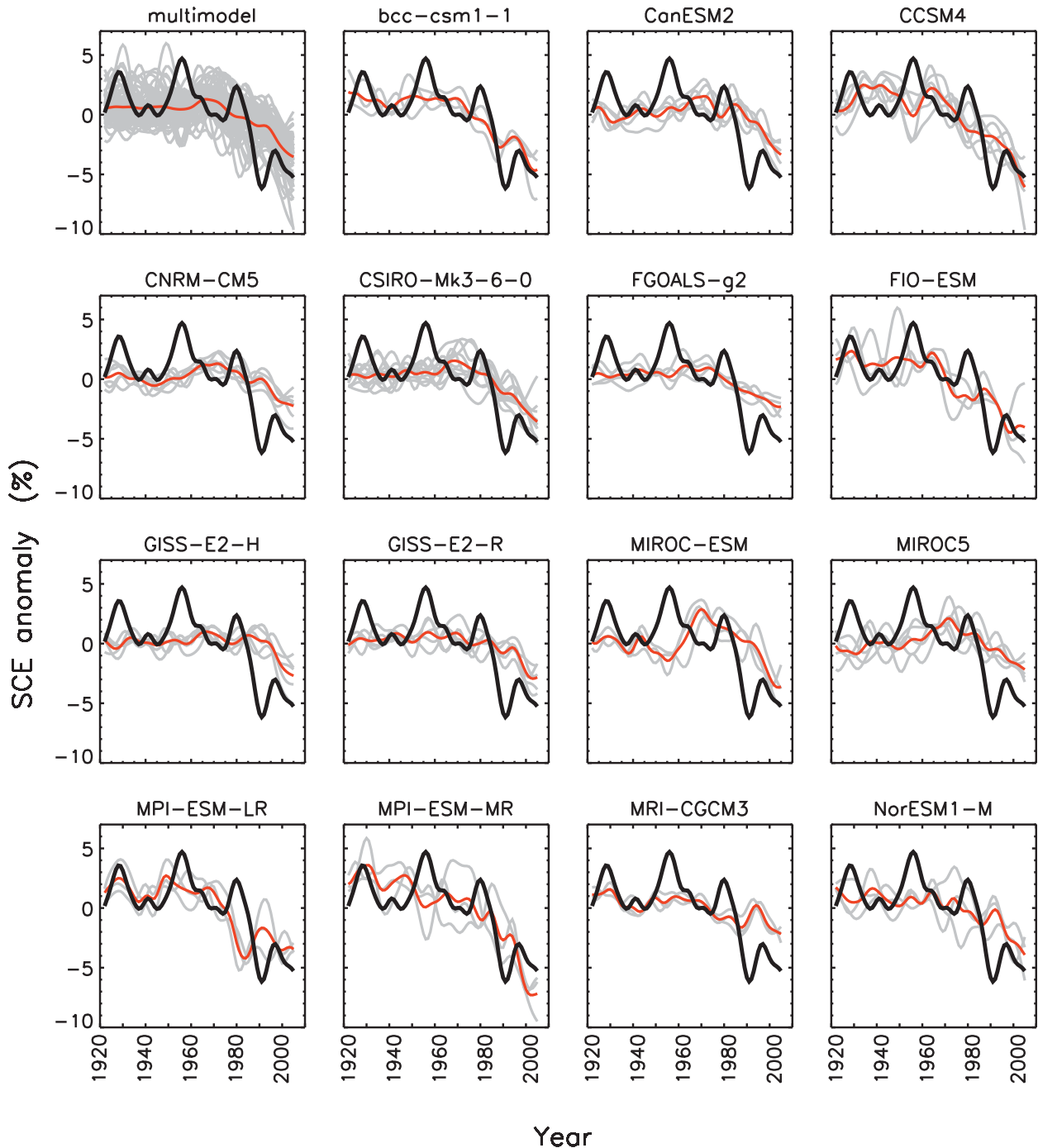


FIG. 3. Time series of spring Northern Hemisphere SCE anomalies. Observations (black) and simulations (gray) with combined natural and anthropogenic forcing were smoothed with a Gaussian kernel standard deviation of 2.5 yr. The red trace shows the mean of the simulations.

(El Chichón, March 1982; Mount Pinatubo). Inspection of the smoothed time series highlights the magnitude of the post-1970 decline against the (multi)decadal variability in both observations and simulations (Fig. 3). While the smoothed observations show high variability

at time scales of 20–30 yr, the magnitude of the marked decline in SCE over the last four decades appears to exceed the prior historical variability back to 1922. The individual GCMs also produced a post-1970 decline in SCE to varying degrees. However, while several

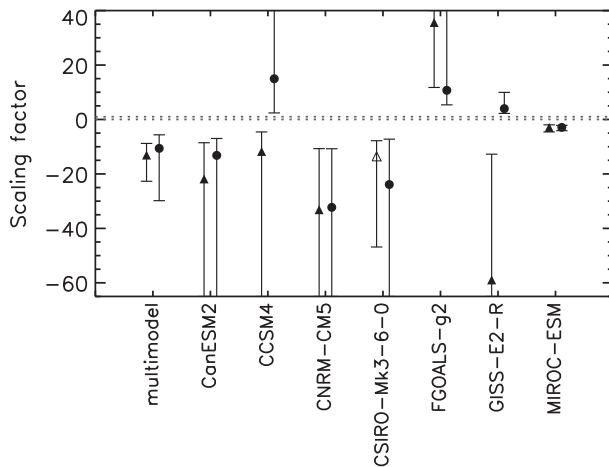


FIG. 4. Best estimates of scaling factors $\hat{\beta}$ and their 5%–95% uncertainty estimates for spring NH SCE with natural forcing only. Filled (open) symbols indicate where the linear model in Eq. (1) was accepted (rejected). Triangles (circles) indicate where the relative SCE anomalies were smoothed with a Gaussian kernel with a standard deviation of 1.5 (2.5) yr. Note that in some instances the 95% uncertainty limit is outside the plotting area.

GCMs generated multidecadal variability with magnitudes similar to what has been observed (e.g., CCSM4.0, FIO-ESM, and MPI-ESM-LR), many others produced too little variability (e.g., CNRM-CM, FGOALS-g2, and MRI-CGCM3).

Results of the optimal fingerprinting further demonstrated a lack of consistency between observations and simulations with only natural forcing. The majority of GCMs in the historicalNat experiment, including the multimodel ensemble, had negative values of the scaling factor (Fig. 4, Table 2), which imply an observed response to forcing that was of opposite sign of the simulated response. The minority of GCMs with positive scaling factors had values that were much larger than unity, thus inconsistent with observations. Though the degree of smoothing of the time series did influence the

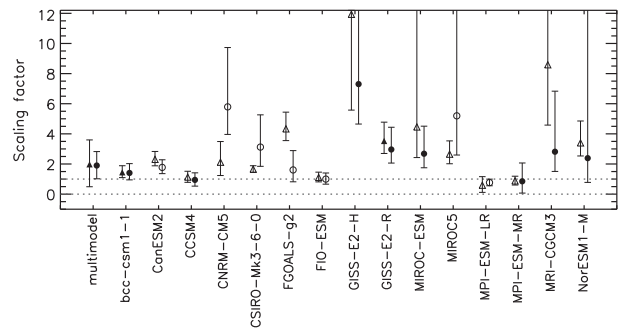


FIG. 5. As in Fig. 4, but for simulations with combined natural and anthropogenic forcing.

scaling factor estimate (more smoothing generally broadened the confidence intervals, and in two cases caused a sign reversal), it did not affect in any meaningful way the overall results.

With combined natural and anthropogenic forcings, the optimal fingerprinting revealed much greater consistency between observations and simulations. In the historical experiment, all GCMs had a scaling factor significantly greater than 0 (Fig. 5, Table 3). Of the eight GCMs that were adequately represented by the linear model in Eq. (1), four had 5%–95% confidence intervals on $\hat{\beta}$ that included unity, implying that those four GCMs were also consistent with the observations in the amplitude of their response.

However, the scaling factors were generally greater than unity across GCMs and ranged from 0.9 to 7.3 where the linear model [Eq. (1)] was accepted. The multimodel ensemble had a scaling factor of just under 2, implying that the GCMs as a group underrepresented the SCE response to external forcings by a factor of roughly 2. It is worth noting that Pierce et al. (2008), when examining 1 April western U.S. snow water equivalent normalized by precipitation, detected the same degree of underrepresentation ($\times 2$) for two GCMs.

TABLE 2. Scaling factor β by GCM for the CMIP5 historicalNat experiment using relative and absolute SCE anomalies, with 5%–95% confidence intervals (CI) and p values. Anomalies were smoothed with Gaussian kernel with a standard deviation of 2.5 yr.

GCM	Relative anomalies			Absolute anomalies		
	Estimated β	CI	p value	Estimated β	CI	p value
Multimodel	–11	–30 to –6	0.106	–10	–31 to –5	0.160
CanESM2	–13	–86 to –7	0.904	–12	–69 to –6	0.903
CCSM4.0	15	2 to –4*	0.813	14	2 to –4*	0.799
CNRM-CM5	–33	33 to –10*	0.277	–26	27 to –9*	0.289
CSIRO Mk3.6.0	–23	20 to –7*	0.070	–33	26 to –10*	0.401
FGOALS-g2	11	5–188	0.689	8.8	4–93	0.682
GISS-E2-R	4.0	2–10	0.263	3.7	2–10	0.258
MIROC-ESM	–2.9	–4 to –2	0.294	–3.0	–4 to –2	0.274

* CI passes through \pm infinity.

TABLE 3. Scaling factor β by GCM for the CMIP5 historical experiment using relative and absolute snow cover extent anomalies, with 5%–95% CI and p value. Anomalies were smoothed with Gaussian kernel with a standard deviation of 2.5 yr.

GCM	Relative anomalies			Absolute anomalies		
	Estimated β	CI	p value	Estimated β	CI	p value
Multimodel	1.9	1.0–2.8	0.897	1.9	1.1–2.7	0.560
BCC-CSM1.1	1.4	0.9–2.0	0.409	1.4	0.9–2.0	0.476
CanESM2	1.8	1.4–2.3	0.007*	1.5	1.1–2.0	0.022*
CCSM4.0	0.9	0.5–1.4	0.793	0.9	0.5–1.4	0.827
CNRM-CM5	5.8	4.0–9.7	0.013*	3.2	2.0–5.6	0.033*
CSIRO Mk3.6.0	3.1	1.8–5.3	0.000*	4.6	3.1–7.3	0.000*
FGOALS-g2	1.0	0.8–2.9	0.002*	1.7	1.1–2.6	0.002*
FIO-ESM	1.0	0.7–1.4	0.000*	1.3	0.9–1.8	0.000*
GISS-E2-H	7.3	4.7–15.2	0.889	5.9	3.7–12.0	0.906
GISS-E2-R	3.0	2.1–4.4	0.198	2.7	1.9–4.1	0.275
MIROC-ESM	2.7	1.8–4.5	0.781	2.8	1.8–4.6	0.759
MIROC5	5.2	2.6–22	0.000*	4.2	1.9–16.7	0.000*
MPI-ESM-LR	0.8	0.6–2.1	0.000*	2.2	1.7–1.8	0.000*
MPI-ESM-MR	0.9	0.1–2.1	0.612	1.3	0.3–3.7	0.050
MRI-CGCM3	2.8	1.5–6.8	0.938	2.6	1.5–5.2	0.720
NorESM1-M	2.4	0.8–16.2	0.905	1.7	1.2–2.5	0.005*

* Linear model [Eq. (1)] rejected.

The best estimates of the scaling factors for a given GCM were generally similar whether a standard deviation of 1.5 or 2.5 was used with the Gaussian smoother (Fig. 5), though in some cases the effect was dramatic (e.g., CNRM-CM5). However, the results of the significance testing varied with degree of smoothing. With the greater smoothing, six more GCMs were found to show consistency with Eq. (1). However, increasing the smoothing also tended to broaden the confidence intervals about β .

While there appears to be general consistency between the scaling factors in Table 3 and the magnitude of the modeled SCE declines in the time series in Fig. 3, the relationship is not direct because the data are first projected onto a user-selected number κ of leading eigenvectors that are derived from an independent set of control runs. To demonstrate the effect of this EOF filtering, we provide more detailed results from the historical experiment for four example cases: the multimodel ensemble mean, two GCMs with a scaling factor close to unity (CCSM4.0 and MPI-ESM-MR), and one GCM with a high scaling factor (GISS-E2-H). In each case, the selection of κ followed an examination of scaling factors calculated for all possible κ (Fig. 6, left column). Stability in the scaling factor across several values of κ increased confidence in the results (see multimodel ensemble, CCSM4.0, and MPI-ESM-MR examples in Fig. 6). In contrast, our confidence in the estimated scaling factor for GISS-E2-H was lower because the value was less stable across values of κ .

The observed and simulated time series, both projected onto the selected κ eigenvectors, are shown in Fig. 6 (center column). These two projected time series

form \mathbf{y} and \mathbf{x} , respectively, in Eq. (1). Recall that the noise in \mathbf{y} , or v_0 , derives from internal variability in the control runs. The TLS procedure also assumes there is noise v_1 in the simulated pattern (owing to finite ensemble size), which must also be estimated. Plotting \mathbf{y} against the estimated noise-free model response ($\mathbf{x} - v_1$) illustrates the relationship between the noise and the estimated scaling factor $\hat{\beta}$ (Fig. 6, right column); large simulation noise v_1 relative to observation noise v_0 is generally associated with larger values of $\hat{\beta}$ (e.g., multimodel ensemble and GISS-E2-H).

A conspicuous result of optimal fingerprinting using the historicalNat experiment is the very wide range in the values of the scaling factor among GCMs (Fig. 4). This result by itself is evidence that the naturally forced simulations do not have variations that project easily onto the observations. In other words, the natural experiments are more or less approximately orthogonal to the observed changes.

Though a much narrower range of scaling factors was estimated from the historical experiment, some of these scaling factors did vary among GCMs and in some cases significantly (Fig. 5). These differences in the historical experiment scaling factors may be caused by differences across GCMs in, among other things, unforced variability, bias, and climate sensitivity due to a different physics representation. Though an in-depth analysis of the causes of the variability in scaling factors is beyond the scope of this paper, we took an exploratory look at the possible influences of unforced variability and bias (focusing on the time series smoothed with a 2.5 standard deviation kernel). These two factors may be related, as GCMs that

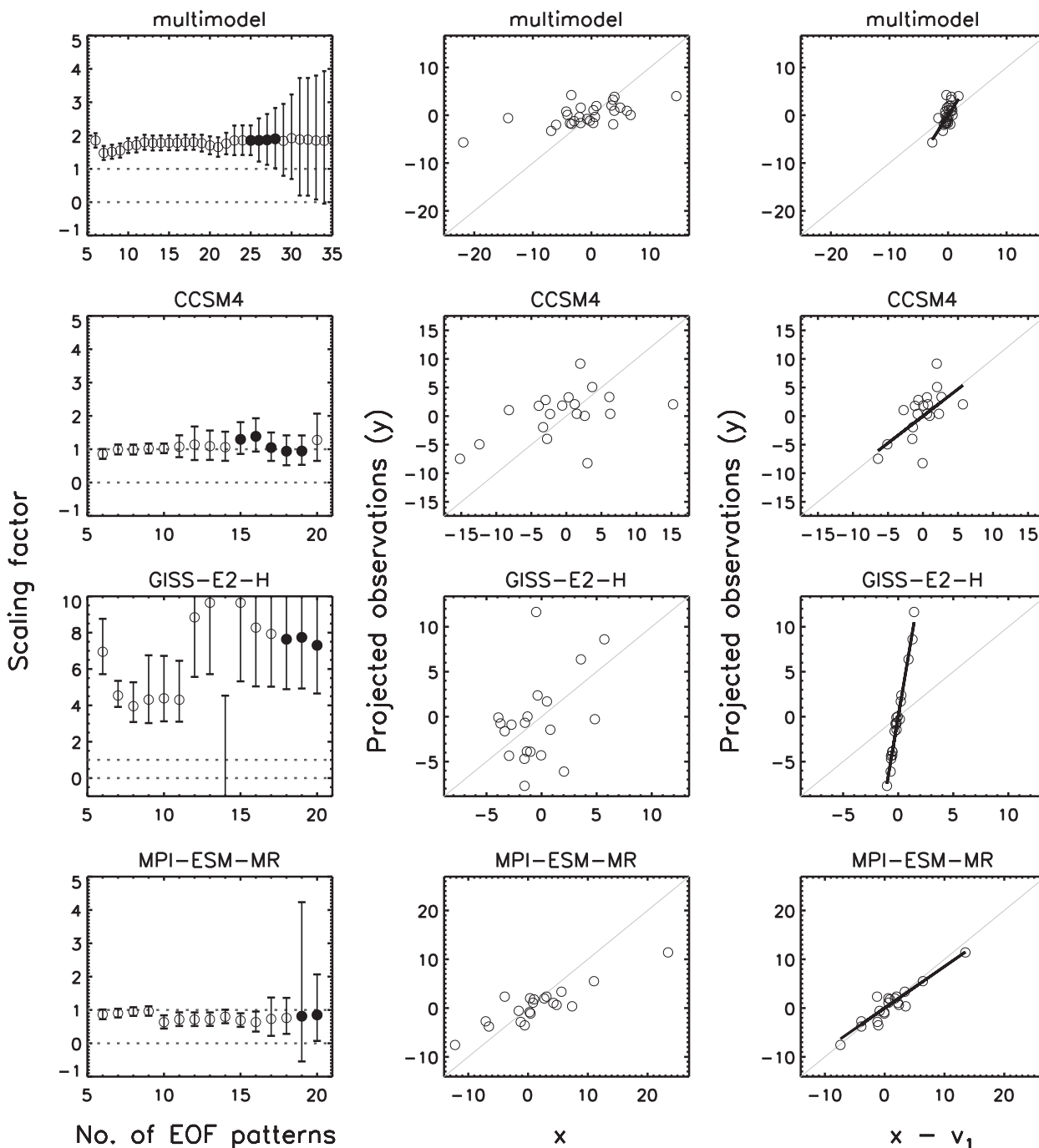


FIG. 6. (left) Scaling factor estimate $\bar{\beta}$ and its 5%–95% uncertainty against number of EOF patterns retained κ for the multimodel ensemble and three example GCMs. Filled (open) symbols indicate where the linear model was accepted (rejected). (center) Observations vs simulations after projecting each onto selected number of eigenvectors, or y vs x [see Eq. (1)]. (right) Projected observed pattern (y) against estimates of noise-free projected simulated pattern ($x - v_1$). The thick black line in the right column shows the estimated noise-free projected observed pattern ($y - v_0$) vs the estimated noise-free projected simulated pattern ($x - v_1$).

tend to generate more snow cover may tend to have higher variance in their absolute anomalies. In fact, the standard deviations of the preindustrial control runs were positively, albeit weakly, correlated with model bias ($r = 0.34$).

It is possible that underrepresentation in internal variability can inflate the scaling factor. However, we found that the standard deviations of the preindustrial absolute anomalies explained none of the scaling factor

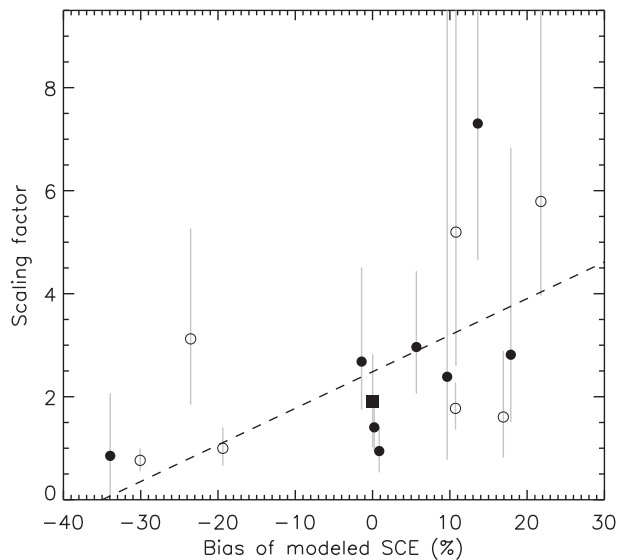


FIG. 7. Best estimate of scaling factor against climatological bias in modeled spring NH SCE (1922–2005 baseline). Circles represent individual models, and the square shows the multimodel ensemble. Filled (open) symbols indicate where the linear model was accepted (rejected). Gray lines show the 5% and 95% uncertainty estimates. The dashed line is the regressed linear fit ($R^2 = 0.30$) through the accepted scaling factors (solid symbols). Scaling factors were estimated from the Gaussian-smoothed relative anomalies (std dev = 2.5 yr).

variance ($R^2 = 0.00$). Interestingly, we conducted the optimal fingerprinting with both absolute and fractional anomalies because of the possibility that bias-driven model differences in internal variability could be a factor. However, using relative anomalies in place of absolute anomalies did not reduce the variability in scaling factor values and, for most GCMs, had only a modest effect (Table 3).

For combined natural–anthropogenic forcing, we found that climatological SCE bias explained a noteworthy proportion of the scaling factor variance ($R^2 = 0.30$). Curiously, the trend of scaling factor with bias was positive (Fig. 7), whereas we expected smaller scaling factors resulting from greater positive bias if greater positive bias implied greater internal variability.

This leads us toward the conclusion that physics-representation-driven differences in climate sensitivity to external forcings among GCMs are the principal cause of the scaling factor variability. However, a thorough investigation of the causes requires much more analysis than we have presented. One step would be to compare our results with a similar optimal fingerprinting analysis of both temperature and precipitation response among the same set of CMIP5 models. Given, for example, that Brown and Robinson (2011) found that near-surface air temperature (40° – 60° N) explained 49% of the variability

in spring NH SCE, we would expect that scaling factors estimated from NH surface air temperature in particular would explain much of the variability seen among models in this study. Furthermore, the sensitivities should vary in space (e.g., Brown and Mote 2009); therefore, an analysis of the spatiotemporal consistency between observations and models is a logical follow up to the study presented here.

Acknowledgments. We acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their model output. For CMIP, the U.S. Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. Dáithí Stone provided the core code for the optimal fingerprinting analysis and Rachel Calmer assisted with the CMIP5 data acquisition. PAS was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). NLB acknowledges the support of the ARC Centre of Excellence for Climate System Science.

REFERENCES

- 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, doi:10.1007/s00382-003-0313-9.
- Barnett, T. P., D. W. Pierce, K. M. AchutaRao, P. J. Gleckler, B. D. Santer, J. M. Gregory, and W. M. Washington, 2005: Penetration of human-induced warming into world’s oceans. *Science*, **309**, 284–287.
- , and Coauthors, 2008: Human-induced changes in the hydrology of the western United States. *Science*, **319**, 1080–1083, doi:10.1126/science.1152538.
- Brown, R. D., 2000: Northern Hemisphere snow cover variability and change, 1915–97. *J. Climate*, **13**, 2339–2355.
- , and P. W. Mote, 2009: The response of Northern Hemisphere snow cover to a changing climate. *J. Climate*, **22**, 2124–2145.
- , and D. A. Robinson, 2011: Northern Hemisphere spring snow cover variability and change over 1922–2010 including an assessment of uncertainty. *Cryosphere*, **5**, 219–229, doi:10.5194/5-219-2011.
- Derksen, C., and R. Brown, 2012: Spring snow cover extent reductions in 2008–2012 period exceeding climate model projections. *Geophys. Res. Lett.*, **39**, L19504, doi:10.1029/2012GL053387.
- 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.

- , R. J. Allan, and T. J. Ansell, 2005: Detection of external influence on sea level pressure with a multi-model ensemble. *Geophys. Res. Lett.*, **32**, L19714, doi:10.1029/2005GL023640.
- Hasselmann, K., 1993: Optimal fingerprints for the detection of time-dependent climate change. *J. Climate*, **6**, 1957–1971.
- Hegerl, G. C., K. Hasselmann, U. Cubasch, J. F. B. Mitchell, E. Roeckner, R. Voss, and J. Waszkewitz, 1997: Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. *Climate Dyn.*, **13**, 613–634, doi:10.1007/s003820050186.
- , 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.
- Lambert, F. H., N. P. Gillett, D. A. Stone, and C. Huntingford, 2005: Attribution studies of observed land precipitation changes with nine coupled models. *Geophys. Res. Lett.*, **32**, L18704, doi:10.1029/2005GL023654.
- Lean, J. L., and D. H. Rind, 2008: How natural and anthropogenic influences alter global and regional surface temperatures: 1889 to 2006. *Geophys. Res. Lett.*, **35**, L18701, doi:10.1029/2008GL034864.
- Min, S.-K., X. Zhang, F. W. Zwiers, and T. Agnew, 2008: Human influence on Arctic sea ice detectable from early 1990s onwards. *Geophys. Res. Lett.*, **35**, L21701, doi:10.1029/2008GL035725.
- Pierce, D. W., and Coauthors, 2008: Attribution of declining western U.S. snowpack to human effects. *J. Climate*, **21**, 6425–6444.
- Santer, B. D., and Coauthors, 2003: Behavior of tropopause height and atmospheric temperature in models, reanalyses, and observations: Decadal changes. *J. Geophys. Res.*, **108**, 4002, doi:10.1029/2002JD002258.
- Stott, P. A., S. F. B. Tett, G. S. Jones, M. R. Allen, J. F. B. Mitchell, and G. J. Jenkins, 2000: External control of 20th century temperature by natural and anthropogenic forcings. *Science*, **290**, 2133–2137.
- , —, —, —, W. J. Ingram, and J. F. B. Mitchell, 2001: Attribution of twentieth century temperature change to natural and anthropogenic causes. *Climate Dyn.*, **17**, 1–12.
- , M. R. Allen, and G. S. Jones, 2003: Estimating signal amplitudes in optimal fingerprinting. Part II: Application to general circulation models. *Climate Dyn.*, **21**, 493–500, doi:10.1007/s00382-003-0314-8.
- , D. A. Stone, and M. R. Allen, 2004: Human contribution to the European heatwave of 2003. *Nature*, **432**, 610–614.
- , G. S. Jones, N. Christidis, F. W. Zwiers, G. Hegerl, and H. Shiogama, 2011: Single-step attribution of increasing frequencies of very warm regional temperatures to human influence. *Atmos. Sci. Lett.*, **12**, 220–227, doi:10.1002/asl.315.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 1689–1696.
- Tebaldi, C., K. Hayhoe, J. M. Arblaster, and G. A. Meehl, 2006: Going to extremes: An intercomparison of model-simulated historical and future changes in extreme events. *Climatic Change*, **79**, 185–211.
- Tett, S. F. B., and Coauthors, 2002: Estimation of natural and anthropogenic contributions to twentieth century temperature change. *J. Geophys. Res.*, **107**, 4306, doi:10.1029/2000JD000028.