Estimating the Anthropogenic Sea Surface Temperature Response Using Pattern Scaling

ADELINE BICHET, PAUL J. KUSHNER, AND LAWRENCE MUDRYK
Department of Physics, University of Toronto, Toronto, Ontario, Canada

LAURENT TERRAY
CERFACS/CNRS, URA1875, Toulouse, France

JOHN C. FYFE
Canadian Centre for Climate Modelling and Analysis, Environment Canada, Victoria, British Columbia, Canada

(Manuscript received 28 August 2014, in final form 3 February 2015)

ABSTRACT

This study seeks to derive the sea surface temperature (SST) response to anthropogenic forcing from observations over the last century, using simple methods inspired from pattern scaling. As in pattern scaling, the spatial response is assumed to scale with global-mean and annual-mean surface temperature. The long-term aim of this work is to generate anthropogenically forced SST and sea ice patterns for the recent past and near-term future, and use them to force atmosphere–land climate models for attribution and prediction purposes. The present work compares estimation methodologies and, within a Monte Carlo framework based on large initial condition ensembles of climate model simulations, examines the robustness of the patterns obtained.

The different methods explored here yield a similar SST spatial response, mostly reflecting the observed SST linear trend map. The different methods nevertheless provide distinctive temporal evolution of the global-mean and annual-mean SST response, which in turn affects the temporal evolution of the global-mean and annual-mean air surface temperature simulated in corresponding prescribed SST simulations. The estimated SST spatial response consists mostly of a warming of the midlatitude coasts near the western boundary currents, the tropical Indian Ocean, and the Arctic Ocean. This pattern generally agrees with previously published observational and modeling studies. Based on Monte Carlo analysis of the large ensembles, it is found that between 36% and 56% of its spatial variance results from anthropogenic forcing.

Overall, the work herein provides constraints on the uncertainty associated with the spatial variability of an anthropogenically forced component of climate change derived from observations, which can potentially be used for climate attribution and prediction.

1. Introduction

It is known that imposing twentieth-century observed sea surface temperatures (SSTs) and sea ice concentrations in a land–atmosphere global climate model (AGCM) determines much of that model's circulation and land surface temperature trend, even in the absence of additional radiative forcing (e.g., Gates 1992; Lau 1997; Hoerling et al. 2008; Deser and Phillips 2009; Compo and Sardeshmukh 2009). This is because oceanic moistening and warming of the air over the ocean leads, by atmospheric advection, to moistening and warming of the air over land, thereby increasing the downward longwave radiation at the land surface, thereby determining the characteristics of the broad climate response. In addition, spatial patterns of SST also affect land climate via their major roles in driving atmospheric teleconnections [e.g., El Niño–Southern Oscillation (ENSO) and the Atlantic multidecadal oscillation (AMO)]. Hence, if one could cleanly separate the anthropogenic component of SST and sea ice change from natural variability, the anthropogenic component of the SST and sea ice change could then be used to drive idealized global or regional land–atmosphere climate models to investigate

Corresponding author address: Adeline Bichet, Department of Physics, University of Toronto, 60 St. George St., Toronto ON M5S 1A7, Canada.
E-mail: abichet@atmosp.physics.utoronto.ca

DOI: 10.1175/JCLI-D-14-00604.1

the associated land climate response. Such simulations can then be used for climate change attribution, climate variability analysis, and climate prediction.

The spatial response of SST and sea ice concentration to anthropogenic forcing can be derived from coupled atmospheric–ocean GCMs (AOGCMs) or directly from observations. Deriving it from AOGCMs is challenging, in particular because current AOGCMs used for climate prediction have difficulty capturing observed SST trends, even after accounting for natural climate variability (Shin and Sardeshmukh 2011). This error can affect applications such as pattern scaling (Santer et al. 1990; Tebaldi and Arblaster 2014) and climate change detection and attribution, and would certainly affect this proposed application as well.

To avoid such issues and make progress, this study evaluates a relatively new method that relies heavily on observations to derive the spatial response of SST and sea ice concentration to anthropogenic forcing (Hoerling et al. 2011). As with Santer et al.'s (1990) model-based pattern scaling, our method consists of scaling with global-mean and annual-mean surface temperature a time-invariant spatial pattern $h(x)$ that corresponds to the spatial response of a given variable (in our case SST and sea ice concentration) to anthropogenic forcing. Note that whereas we derive the spatial pattern $h(x)$ entirely from observations, the global-mean and annual-mean surface temperature used for the scaling may be alternatively derived from observations (in case of past climate attribution) or AOGCM outputs (in case of climate prediction). To our knowledge, this approach has only been used once in the past, by Hoerling et al. (2011). Practically speaking, it is a simple and straightforward approach, which allows a simple and straightforward interpretation of results. In their work, Hoerling et al. (2011) quantify the decadal mean changes (2011–20) in North American surface air temperature and precipitation by forcing three different AGCMs with decadal mean (2011–20) estimates of anthropogenically forced SST computed with the method explored here, and sea ice concentration computed from the persistence of their monthly pattern taken from 2007–09. They predict that the anthropogenically forced component of North American warming in 2011–20 relative to 1971–2000 climatological conditions will be a surface warming over the entire North American continent, a precipitation decrease over the contiguous United States, and a precipitation increase over Canada. The study also provides quantitative estimates of the fraction of change attributable to the predicted forced component. The method shows considerable promise, and in our work we aim to employ it to create a monthly-mean estimate of anthropogenically forced SST and sea ice concentration (reconstructed from a time-invariant spatial pattern and its associated time series) covering historical and future time periods. We then use this estimate to force an AGCM in a time-evolving mode.

Because the method used in Hoerling et al. (2011) remains to be validated and interpreted, the current study explores its characteristics and robustness. We first evaluate three observation-based estimation methods for anthropogenically forced SST and sea ice concentration (described in section 2). We then discuss how the three estimates obtained for anthropogenically forced SST differ in their spatial structure $h(x)$ and associated time evolution. We also use them to force an AGCM, and discuss the corresponding three sensitivity experiments (section 3a). Then (section 3b) we use a Monte Carlo large initial condition ensemble approach (Kay et al. 2015) to quantify the method's ability to correctly isolate the anthropogenically forced signal from internal variability in the SST spatial patterns $h(x)$, as well as in some other aspects of surface temperature change. We also apply our method to several models from phase 5 of the Coupled Model Intercomparison Project (CMIP5), and compare, for SST, these model-based $h(x)$ with the observation-based $h(x)$ (section 3c). Finally, we discuss the uncertainties associated with this method and conclude the study (section 4). Note that even though we apply the method to observed SSTs and sea ice concentrations, this study will focus on analysis performed on SSTs.

2. Method and datasets

a. Overview of the method

We denote by $S(x, t)$ the departure of the annual mean SST from its climatological mean for location x and calendar year t. We wish to estimate the anthropogenically forced component of $S(x, t)$. To do so we first decompose S into the sum of a forced component S_F and a random component S_I, representing the noise associated with internal climate variability that is assumed to be independent of S_F (Ribes et al. 2010). As for pattern scaling, we assumes that $S_F(x, t)$ can be separated as $S_F(x, t) \approx g(t)h(x)$. In classical pattern scaling S_F is obtained from anthropogenically forced climate projections dominated by greenhouse forcing; $h(x)$ is estimated as the epochal difference between two multidecadal periods of S from the simulation separated by a suitably long period, and g is the global mean temperature difference between the two epochs (Santer et al. 1990; Tebaldi and Arblaster 2014). Instead, we here obtain S_F from observations for the reasons outlined in the introduction, and compute $h(x)$, as in Hoerling et al. (2011), by linearly regressing observed

$S(x, t)$ onto $g(t)$. Both variables are high-pass filtered prior to regression to account for serial autocorrelations. We use three different methods to derive $g(t)$ from global-mean and annual-mean SSTs, as described in section 2b. The time-varying S_F is then obtained by multiplying $h(x)$ by $g(t)$, to which the observed climatology is added to yield a monthly-mean time series. Therefore, as for patterns scaling, our method assumes that the SST and sea ice spatial response to anthropogenic forcing (greenhouse gases and aerosols) does not vary in time. This assumption is less justified for Arctic sea ice concentration than for SSTs due to its less linear response to anthropogenic forcing, particularly in late summer/early autumn. This approach nevertheless allows the prescribed sea ice response to be consistent with the prescribed SST response.

b. Methods to estimate $g(t)$

We consider three estimates of $g(t)$, denoted g_L, g_C, and g_{Th}; for all three estimates, g is normalized to have unit standard deviation, meaning that the spatial pattern h represents an estimate of the forced response at each spatial point per unit standard deviation of the global-mean and annual-mean SST response. The ''linear'' estimate $g_L(t)$ is simply a standardized linear function $g_L(t) = (2\sqrt{3}/\Delta t)(t - \bar{t})$, where Δt is the time interval for the analysis and $\bar{t}$ is the time midpoint; note that by definition the standard deviation of g_L is unity over the time interval. In this case, h is a spatial map of the SST linear trend, an approach that is sometimes used in pattern scaling (Tebaldi and Arblaster 2014). The ''cubic'' estimate $g_C(t)$ is obtained by taking the global-mean and annual-mean SST and fitting it to a third-order smoothed spline using the function smooth.spline from the R package with df = 3. A similar approach is used in Hoerling et al. (2011), and is intended to capture the increase in the rate of SST warming through the twentieth century. Finally, the ''Thompson'' estimate g_{Th} is obtained following a three-step procedure. In the first step the signals of ENSO, explosive volcanoes and the cold ocean–warm land (COWL) pattern (Wallace et al. 1995) are extracted from the observed time series of global-mean and monthly-mean SST. The ENSO and volcanic signals are estimated using a simple thermodynamic model while the COWL signal is estimated using the covariance between Northern Hemisphere land–sea temperature difference and sea level pressure. In the second step these signals form the basis of a multivariate linear regression between the signals and the observed time series of global-mean and monthly-mean SST. The terms in this regression associated with these signals of natural variability are then removed from the observed time series of global-mean and monthly-mean SST to produce a ''residual'' time series. This step in the procedure provides a simple, robust, and physically based methodology for removing the major known signals of natural global climate variability (Thompson et al. 2009). Finally, in the third step this residual time series is regressed against our best estimate of observed anthropogenic radiative forcing to yield g_{Th}.

For interpretation of results, our working assumption is that the forced signal from solar variability is relatively weak. Our methods produce $g(t)$ that either temporally filter the short-term impacts of volcanic aerosols, or in the case of the Thompson (Th) method attempts to explicitly removes them. All the maps presented here show the scaled regression coefficient field $\tilde{h}$ representing $h(x)$ divided by its global mean.

c. Datasets and models

We derive $g(t)$ from observations using version 4 of the Hadley Centre surface temperature product (hereinafter HadCRUT4), which includes missing values. We use HadCRUT4 as the basis for $g(t)$ for dataset consistency with the inputs into the g_{Th} (Thompson et al. 2009) index. The possible artificial fluctuations owing to the time-dependent data coverage in the HadCRUT4 global-mean and annual-mean time series are not expected to affect $g(t)$, as $g(t)$ represents a very smoothed version of the time series. Once $g(t)$ is obtained, we are free to obtain $h(x)$ from other datasets. In particular, we derive $h(x)$ from observations using the National Center for Atmospheric Research (NCAR) global SST and sea ice product (Hurrell et al. 2008; the product is hereinafter called ''Hurrell''). We use Hurrell instead of HadCRUT4 to derive $h(x)$ because Hurrell is the standard driving dataset used in the NCAR AGCM simulations that we shall present shortly. In addition, the Hurrell SST twentieth-century trend map agrees with the twentieth-century trend map that is obtained with all the observational SST datasets tested in Solomon and Newman (2012), after statistical correction accounting for sampling of ENSO variability.

For statistical analysis purpose, we also derive $g(t)$ and $h(x)$ from several CMIP5 models (see Table 1) and two large initial condition ensembles. These three ensembles are forced with time-dependent historical greenhouse gases, ozone, aerosols, volcanic emissions, and solar variability taken from the standard datasets of the historical radiative forcing CMIP5 protocol (Taylor et al. 2012). From CMIP5 we use 1900–2005 SSTs from 26 models, using one historical realization per model (Taylor et al. 2012).

The two initial condition large ensembles (e.g., Mudryk et al. 2014; Fischer and Knutti 2014; Kay et al. 2015) are used to test the method in a perfect model

TABLE 1. The 26 CMIP5 models used in this study, distinguishing the models including some representation of the indirect aerosol effect. (Expansions of acronyms are available online at http://www.ametsoc.org/PubsAcronymList.)

Modeling center	Model	Indirect aerosol effect included: Using aerosol emissions instead of aerosol optical depth
CSIRO-BOM	ACCESS1.3	Yes
	ACCESS1.0	Yes
CCCma	CanESM2	Yes
CNRM-CERFACS	CNRM-CM5	Yes
CSIRO-QCCCE	CSIRO-Mk3.6.0	Yes
NASA GISS	GISS-E2-H	Yes
	GISS-E2-R	Yes
MOHC (additional realizations by Instituto Nacional de Pesquisas Espaciais)	HadGEM2-ES	Yes
	HadGEM2-CC	Yes
IPSL	IPSL-CM5A-LR	Yes
	IPSL-CM5A-MR	Yes
	IPSL-CM5B-LR	Yes
MRI	MRI-CGCM3	Yes
Norwegian Climate Centre	NorESM1-M	Yes
	NorESM1-ME	Yes
INM	INM CM4	Yes
MIROC	MIROC5	Yes
NSF–DOE–NCAR	CESM1 (CAM5)	Yes
	CESM1-BGC	No
BCC	BCC-CSM1.1	No
	BCC-CSM1.1m	No
NCAR	CCSM4	No
MPI-M	MPI-ESM-LR	No
	MPI-ESM-MR	No
CMCC	CMCC-CMS	No
FIO	FIO-ESM	No

framework. Realizations in these ensembles can be treated as idealized observations that differ from each other only by their internal variability and so can be used to separate S_I from S_F. First, we use an ensemble of 39 simulations performed with the NCAR Community Climate System version 4 (CCSM4) coupled GCM. It covers the period 1960–2005 at 2° atmosphere and 1° ocean resolution (Mudryk et al. 2014). Second, we use an initial condition ensemble of 30 simulations performed with the NCAR CESM1 (CAM5) coupled GCM (hereinafter simply CESM1; Kay et al. 2015) covering the period 1920–2005 at 1° atmosphere and 1° ocean resolution.

Finally, we briefly analyze a set of four prescribed SST and sea ice concentration experiments over 1920–2005, or "AMIP-type" experiments, with the NCAR atmosphere–land model CAM5 on a horizontal resolution in the atmosphere of approximately 2° (the F_AMIP_CAM5_CN compset). The four experiments are forced with CMIP5 historical radiative forcing (Taylor et al. 2012). The "AMIP" simulation is forced with Hurrell SST and sea ice concentration whereas the linear (L), cubic (C), and Th simulations that will be presented are forced with different estimates of anthropogenically forced SST (S_F) and sea ice concentration, obtained by regressing, for the time period 1900–2005, Hurrell SSTs and Hurrell sea ice concentrations against observation-based g_L, g_C, and g_{Th} respectively. We perform one experiment with each estimate of S_F. For these experiments we consider that areas where the regression coefficients $h(x)$ are not statistically significant at the 95% confidence level represent areas where the signal (anthropogenically forced component) cannot be distinguished from the noise (internal variability). In this case, we set $h(x)$ to zero as an attempt to minimize the effects of regions where internal variability dominates, and apply a spatial smoothing that weights the value of $h(x)$ according its p value to minimize spatial discontinuity.

3. Results

a. Evaluation of the estimation methods

Figure 1 shows the standardized g_L (blue), g_C (red), and g_{Th} (green) derived from observations (solid curves) for 1900–2005. As stated above, g_L increases linearly from $-\sqrt{3}$ to $+\sqrt{3}$ from the beginning to the end of the period (about 0.5 unit of standard deviation per decade). The somewhat nonlinear g_C, similar to what was obtained in Hoerling et al. (2011) shows accelerated warming over time. The more strongly nonlinear g_{Th} has a weak warming trend prior to the early 1970s that transitions to a stronger warming trend thereafter. Factors contributing to the increased warming after the 1970s include enhanced greenhouse forcing and reduced aerosol forcing (e.g., Wild et al. 2007).

For comparison, we include g_C derived from the CMIP5 multimodel mean (dashed curves). By construction, g_L is uniquely defined, and currently g_{Th} has only been calculated for observations. The accelerated warming after the 1970s is stronger in the CMIP5 multimodel mean than in the observations because SSTs increase more quickly in the CMIP5 multimodel mean than in the observations after this date, and more slowly before that.

To illustrate how the spatial pattern of S_F [$h(x)$] can be affected by the choice of $g(t)$, Fig. 2 shows $\tilde{h}(x)$ derived from Hurrell SST over the period 1900–2005 using g_L (Fig. 2a) and g_C (Fig. 2b), and the difference between $\tilde{h}(x)$ using g_{Th} and $\tilde{h}(x)$ using g_L (Fig. 2c). The patterns of $\tilde{h}(x)$ for g_L and g_C are nearly identical. They feature strong warming in the tropical Indian Ocean (stronger warming in the northern as compared to the southern Indian Ocean), the southern Atlantic Ocean, the

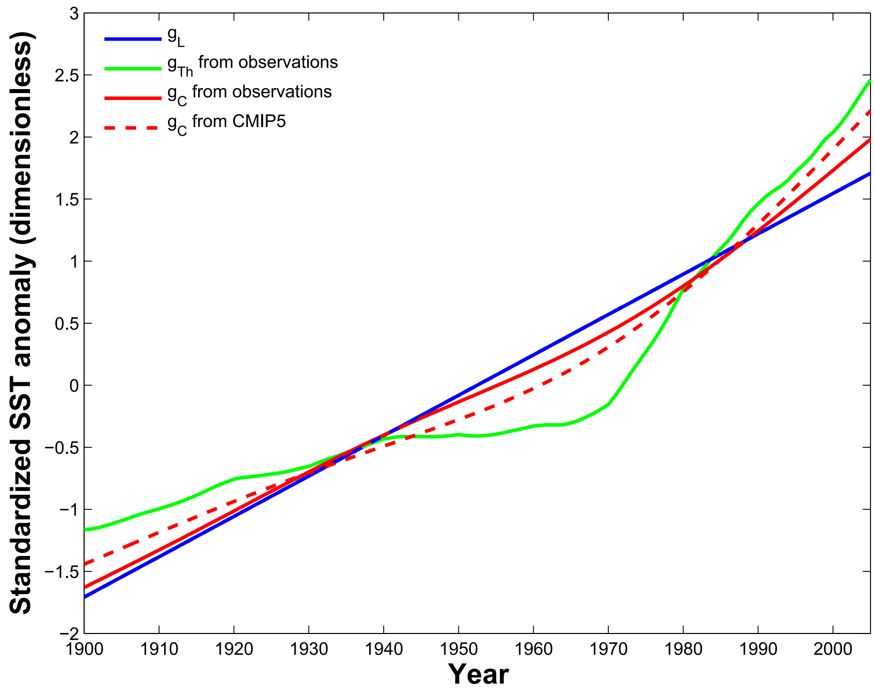

FIG. 1. Standardized $g(t)$ computed with the L (blue), C (red), and Th (green) methods for the period 1900–2005, derived from HadCRUT4 (solid curves). Also plotted is the $g(t)$ derived from the CMIP5 multimodel ensemble mean (dashed red curve) based on the ensemble mean of 26 experiments. All curves have zero mean and unit is standard deviation. Note that L is by construction identical for simulations and observations.

southernmost part of the Atlantic and Indian Oceans (60°–45°S, 60°W–120°E), the northern North Pacific Ocean, the Arctic and subarctic oceans, and the northern and southern midlatitudinal coasts (including the western boundary currents). Cooling or insignificant warming is found in the Greenland Sea, the eastern tropical Pacific, and off the West Antarctic coast. These SST warming patterns are in general agreement with previously published work based on observations (Fig. 7a from Ting et al. 2009; Figs. 2a and 2e from Mohino et al. 2011; Wu et al. 2012) and to some extent on AOGCMs (Fig. 1 from Lu et al. 2008; Figs. 3 and 4a from Ting et al. 2009; Figs. 2 and 8 from Xie et al. 2010). Figures 1 and 2c suggest that relative to g_L, the warming coherent with g_{Th} is enhanced in the tropics and high latitudes and reduced in the midlatitudes before the 1930s and after the 1980s. The overall relative difference between the two patterns is 10%–20%. Not surprisingly $\tilde{h}(x)$ using g_L and g_{Th} is dominated by the linear trend and so is relatively insensitive to the choice of $g(t)$ estimate.

To quantify the overall dependence of the time evolution of $S_F \approx g(t)h(x)$ on the choice of $g(t)$, Fig. 3a shows, for the time period 1920–2005, the global-mean and annual-mean SST anomalies as taken from the observations (Hurrell, magenta curve) and from the different estimates of S_F obtained using observation-based g_L, g_C, and g_{Th} (i.e., the corresponding g multiplied by the spatial mean of the corresponding h). All estimates of S_F show much weaker interannual variability than the observed time series, and the disagreement of the empirical estimates is relatively small compared to interannual-to-decadal variability.

We recall that we wish to use S_F to analyze land climate response to anthropogenically forced SST and sea ice concentration. Based on Fig. 3a, we expect different choices of S_F leading to different responses of the land climate. For example, Fig. 3b shows, for the time period 1920–2005, the 11-yr running mean of global-mean and annual-mean surface air temperature (SAT) anomalies over land, as taken from the observations (HadCRUT4, magenta curve) and from the AMIP, L, C, and Th experiments. Because the simulated global-mean and annual-mean SATs show strong interannual variability, we use an 11-yr running mean to highlight the major differences in the time series. We see that 1) the time evolution of the global-mean and annual-mean land SATs from the AMIP experiment tracks the observed evolution well, with warming trends up to the 1940s, weak cooling from the 1940s to the 1970s, and stronger warming after the 1970s; 2) there is little

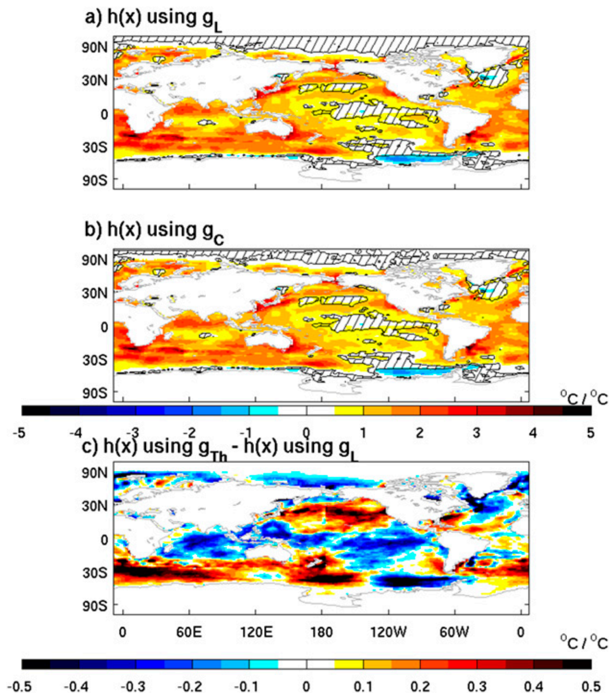

FIG. 2. Shown is $\tilde{h}(x)$ [i.e., regression pattern $h(x)$ divided by its global mean] computed by regressing Hurrell SST against (a) g_L and (b) g_C, both derived from HadCRUT4 over the period 1900–2005, as described in Fig. 1. Also shown is (c) the difference between $\tilde{h}(x)$ computed with g_L and $\tilde{h}(x)$ computed with g_{Th}. All panels have a unit of °C °C^{-1} of global SST warming, and areas in (a) and (b) not significant at 95% according to the Student's t test are hatched.

difference between the time evolution of the global-mean and annual-mean land SATs taken from the L and C experiments; and 3) the global-mean and annual-mean land SAT in Th is nearly constant before the 1970s and increases at about 0.2°C decade^{-1} after the 1970s.

We conclude that the three methods tested produce a similar spatial response of SSTs to anthropogenic forcing [$h(x)$] but a different time evolution of this pattern. In turn, these differences can potentially influence the simulated land climate response. The differences between the observed warming and the L, C, and Th warmings provide distinctive perspectives on the influence of natural variability in land surface temperature in the recent historical era; this idea will be explored further in subsequent work.

b. Evaluation of the methods in a perfect model setting

The sensitivity of the land climate response to different estimates of S_F evident in Fig. 3b motivates a more in-depth investigation of how to estimate uncertainty in the anthropogenically forced signal. In this section, we use large initial condition ensembles (Kay et al. 2015) to quantify our ability to correctly isolate the anthropogenically forced SST spatial response pattern $h(x)$ from internal variability, when $h(x)$ is derived from observations. In the framework of large initial condition ensembles, we can consider each realization to be an identically forced pseudo-observation drawn from the same statistical distribution. For a given method, we can generate one $g(t)$ and one $h(x)$ for each realization, and examine the statistics within the ensemble. Thereby, we obtain one $h(x)$ for each realization of the large ensemble. We then compute the spatial correlation among these $h(x)$ to quantify the amount of internal variability they include. We select g_C for this analysis, although similar results are obtained for the L method.

Figures 4a and 4b show centered spatial correlation coefficients between $\tilde{h}(x)$ derived from SSTs taken from CCSM4 over 1960–2005 and from CESM1 over 1960–2005 and 1920–2005. Figure 4a quantifies the pairwise spatial correlation coefficients, thereby estimating how much of the observation-based spatial pattern $\tilde{h}(x)$ is expected to be reproduced in a second set of observation. According to Fig. 4a, the distribution of pairwise spatial correlation coefficients varies from negative values to about 0.8. For CESM1 over 1920–2005, the median value of the correlation coefficients is about 0.6, indicating that in the CESM1 large ensemble over 1920–2005, on average 36% of the spatial variance of an estimated pattern $\tilde{h}(x)$ is explained by the spatial variance of another $\tilde{h}(x)$, and hence results from anthropogenic forcing. For the period 1960–2005, correlation coefficients tend to be greater for CCSM4 than for CESM1. This could be because the global SST warming trend over the twentieth century is more pronounced in CCSM4 than in CESM1 (Hurrell et al. 2013) or because the warming pattern is more uniform in CESM1 than in CCSM4. Figure 4a also shows that lengthening the averaging period boosts correlation values, suggesting how estimates of S_F derived from observations could be improved quantitatively by using a longer record.

Figure 4b considers a second test that makes additional use of the statistics of the ensemble. In this test, we take the ensemble mean $\tilde{h}(x)$ excluding each time the ensemble member being tested, and find its spatial correlation with the $\tilde{h}(x)$ from each individual ensemble member. The ensemble mean provides a more robust estimate of the forced response by averaging most of the internal variability. To be consistent with Fig. 5, we compute the ensemble mean $\tilde{h}(x)$ by averaging $\tilde{h}(x)$ from individual members instead of computing $\tilde{h}(x)$ from the ensemble mean; the difference between the two patterns obtained is relatively small (on average

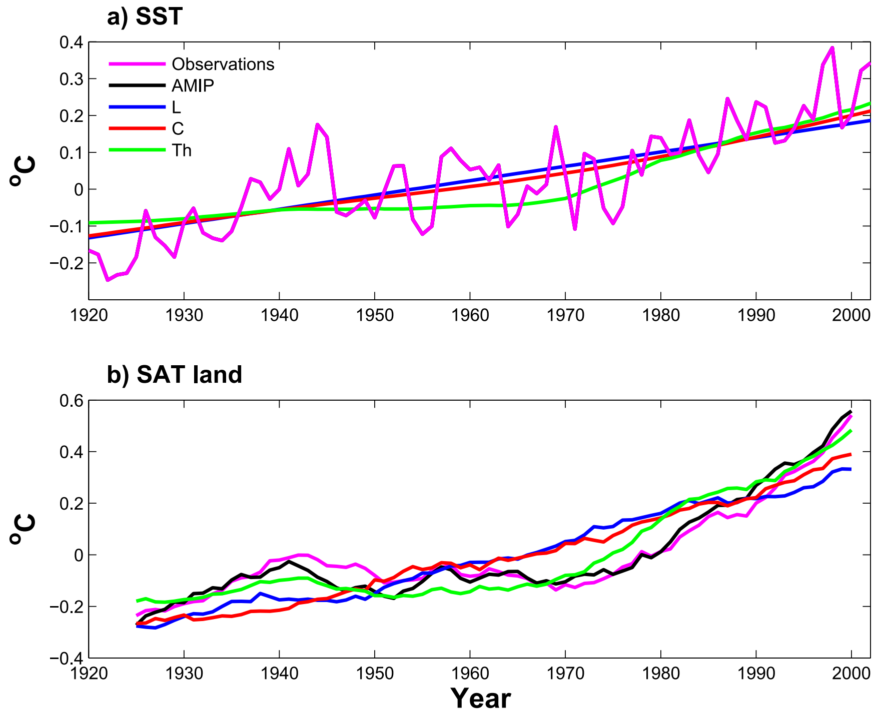

FIG. 3. (a) Global-mean and annual-mean anomalies of the different SSTs used to force the AGCM experiments CAM5 over the period 1920–2005, shown as observed (Hurrell, magenta), and computed by regressing Hurrell against g_L (blue), g_C (red), and g_{Th} (green), all derived from HadCRUT4 over the period 1900–2005, as described in Fig. 1. (b) Simulated SATs averaged over land areas for the AMIP (black), L (blue), C (red), and Th (green) forced simulations, along with the observed HadCRUT4 temperatures (magenta). All time series have been smoothed with an 11-yr running mean.

10%). According to Fig. 4b, the distribution of correlation coefficients is boosted by from 0.1 to 0.2 in all cases and the range of correlation coefficients narrowed. Here we note that the median value of the correlation coefficient in CESM1 over 1920–2005 is about 0.75, indicating that in the CESM1 large ensemble over 1920–2005, on average 56% of the spatial variance of an estimated pattern $\tilde{h}(x)$ is explained by the spatial variance of the ensemble mean $h(x)$ and hence results from anthropogenic forcing.

Because our observation-based estimate of $h(x)$ is computed over the time period 1900–2005 (Fig. 2b), we use the statistics obtained with the longest record available (CESM1; 1920–2005) as an analogy to estimate the amount of internal variability included in our observation-based estimate of $h(x)$. We conclude that between 36% and 56% of the spatial variance of our estimated observation-based pattern $h(x)$ results from anthropogenic forcing. This estimate is model and length dependent.

To quantify the extent to which this test is variable dependent, Figs. 4c and 4d correspond to Figs. 4a and 4b, but for patterns of land SAT excluding the SAT located above sea ice [i.e., the spatial correlation of patterns of land SAT obtained via regression on $g(t)$]. For CESM1, the median values of the correlation coefficients are somewhat lower for land SATs than for SSTs, which illustrates additional uncertainty coming from atmospheric internal variability. However, to round out the picture, when the pattern of global SAT (over ocean, land, and sea ice) response is considered (Figs. 4e,f), the correlations and hence the mutual consistency is increased above that of the SSTs in all cases. Here, the global SAT response including the effects of polar amplification and weaker tropical warming is more robust than the corresponding pattern of SST response, suggesting that some details of the SST response might not be important in determining the global temperature response pattern.

To illustrate how the pattern of $\tilde{h}(x)$ varies within ensembles and across models, Fig. 5 shows the ensemble mean of $\tilde{h}$ (first row), the intraensemble standard deviation of $\tilde{h}$ (second row), and the signal-to-noise ratio of $\tilde{h}$ inferred as the quotient of the first row to the second row (third row), for the CCSM4 (left column) and CESM1 (right column) large ensembles, over the period

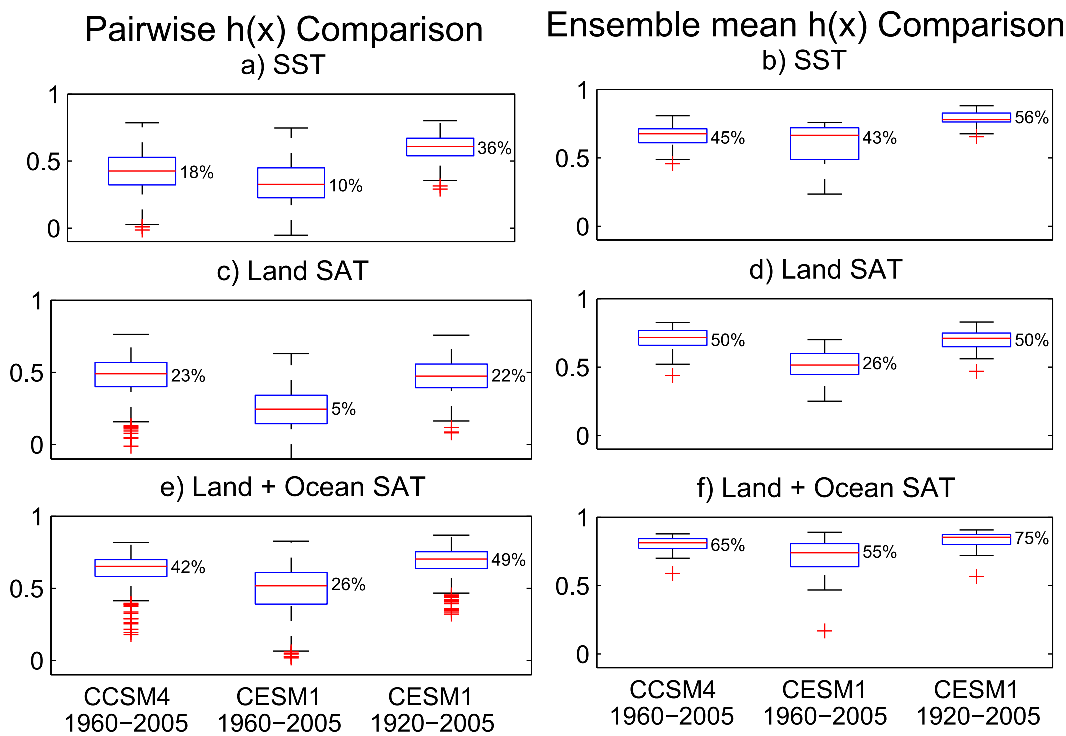

FIG. 4. Perfect-model tests of the method used to estimate $\tilde{h}(x)$. (a) Pairwise centered spatial correlation between the SST $\tilde{h}(x)$ computed from (left) the 39-member CCSM4 ensemble over the time period 1960–2005 (741 pairs), and (middle),(right) the 30-member CESM1 ensemble for the time periods 1960–2005 (465 pairs) and 1920–2005 (465 pairs), respectively. Here, each $\tilde{h}(x)$ is extracted from the specified ensemble member SSTs using g_C derived from the SST of the same specified ensemble member. For each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually in red. (b) As in (a), but here each $\tilde{h}(x)$ is compared to the ensemble mean $\tilde{h}(x)$ excluding the ensemble member being tested. (c),(d) As in (a),(b), but for land SATs excluding SAT located above sea ice; (e),(f) as in (a),(b), but for global (ocean, sea ice, and land) SATs.

1960–2005, which is the common period for the two ensembles. In agreement with the observation-based pattern (Fig. 2b), the two ensemble mean patterns presented in Fig. 5 (top row) show a warming in the southern western boundary currents and almost no warming in the tropical Pacific. On the other hand, they produce more warming in the mid-Atlantic and southern Pacific and less warming in the western North Pacific (in particular the Kuroshio) than in the observation-based pattern (see Fig. 2b). In addition, CCSM4 shows a more uniform warming of the Indian Ocean than the observed pattern, and CESM1 a much stronger Barents/Kara Sea warming. But the question remains whether the differences from observed pattern are due to internal variability or differences in anthropogenically forced response.

To partially answer this question, the second row of Fig. 5 shows that the source of variability in $\tilde{h}(x)$ suggested in Figs. 4a and 4b involves, in both models, internal variability in the northern North Pacific Ocean and the Barents/Kara Sea. Additional hotspots of internal variability include the Greenland Sea in CCSM4, and the western Pacific Ocean in CESM1. The lack of warming in the western North Pacific and the warming in mid-Atlantic in both models is relatively robust across realizations, suggesting that in these regions, the difference with the observations is a systematic bias. According to the third row from Fig. 5, CCSM4 shows a relatively larger signal-to-noise ratio compared to CESM1, in many regions including the Pacific, Indian, and Atlantic Oceans. Both models show a relatively small signal-to-noise ratio in the North Atlantic and the Barents/Kara Sea.

To concretely illustrate how two individual realizations of the same ensemble can differ from each other, Fig. 6 shows $\tilde{h}$ from two individual realizations selected from the CCSM4 (left column) and CESM1 (right column) large ensemble, for the period 1960–2005. According to Fig. 6, the differences in spatial response patterns found in two selected realizations, which

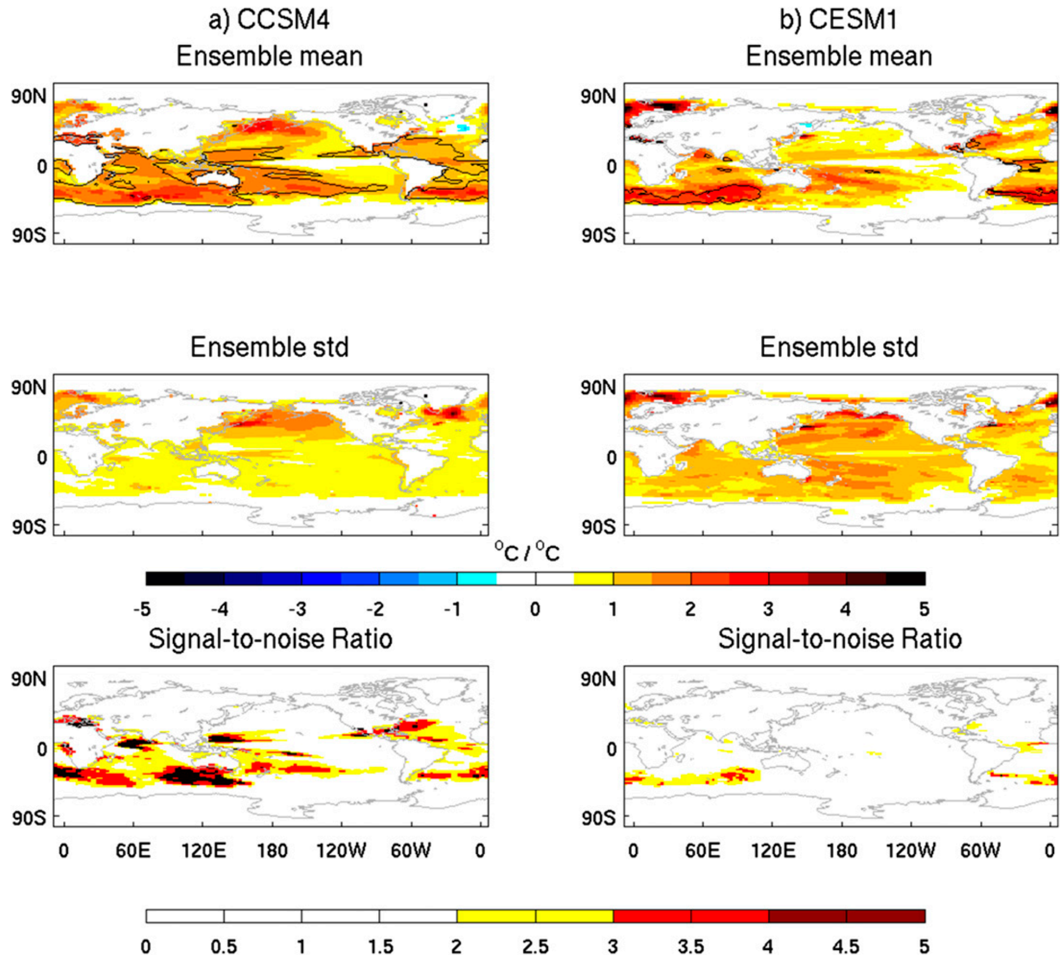

FIG. 5. (a) The value of $\tilde{h}(x)$ computed from the 39-member CCSM4 ensemble over the time period 1960–2005 as described in Fig. 4, and shown as (top) the ensemble mean, (middle) the standard deviation, and (bottom) the signal-to-noise ratio, inferred as the quotient of the first row to the second row. (b) As in (a), but for the 30-member CESM1 ensemble. The top two panels have a unit of °C °C^{-1} of global SST warming. The contours in the top panels show areas where the signal-to-noise ratio is higher than 2.

again could be considered as pseudo observations, can be striking: possible scenarios include a moderate cooling (CCSM4 run X) or pronounced warming (CCSM4 run Y) in the high northern latitudes for CCSM4, and a pronounced (CESM1 run X) or moderate warming (CESM1 run Y) in the western Pacific Ocean for CESM1.

c. *Comparison with the CMIP5 models*

To broaden the comparison, Fig. 7 shows the CMIP5 multimodel mean $\tilde{h}$ (Fig. 7a), the multimodel standard deviation of $\tilde{h}$ (Fig. 7b), and the signal-to-noise ratio of $\tilde{h}$ inferred as the quotient of Figs. 7a and 7b (Fig. 7c), for the period 1900–2005. Here, the multimodel standard deviation reflects differences in external forcing (e.g., different volcanic forcing), model physics (e.g., different implementations of aerosol physics and chemistry), SST response to external forcing, and SST internal variability. Globally, the pattern response derived from CMIP5 (Fig. 7a) shows less spatial structure than the pattern response derived from CCSM4 (Fig. 5a, top), CESM1 (Fig. 5b, top), and the observations (Fig. 2). Whereas the pattern response derived from CMIP5 reproduces some features seen in the pattern response derived from CCSM4, CESM1, and the observations (e.g., warming of the Southern Hemisphere western boundary currents), it fails to reproduce some others (e.g., tropical Pacific cooling). In addition, the CMIP5 intermodel standard deviation (Fig. 7b) is in many regions comparable to the intraensemble standard deviation (Fig. 5, middle row) for the CCSM4 and CESM1 large ensemble, which is particularly pronounced (between 2° and 3°C °C^{-1} of global SST warming) in the Greenland and the Barents/Kara Sea. The largest

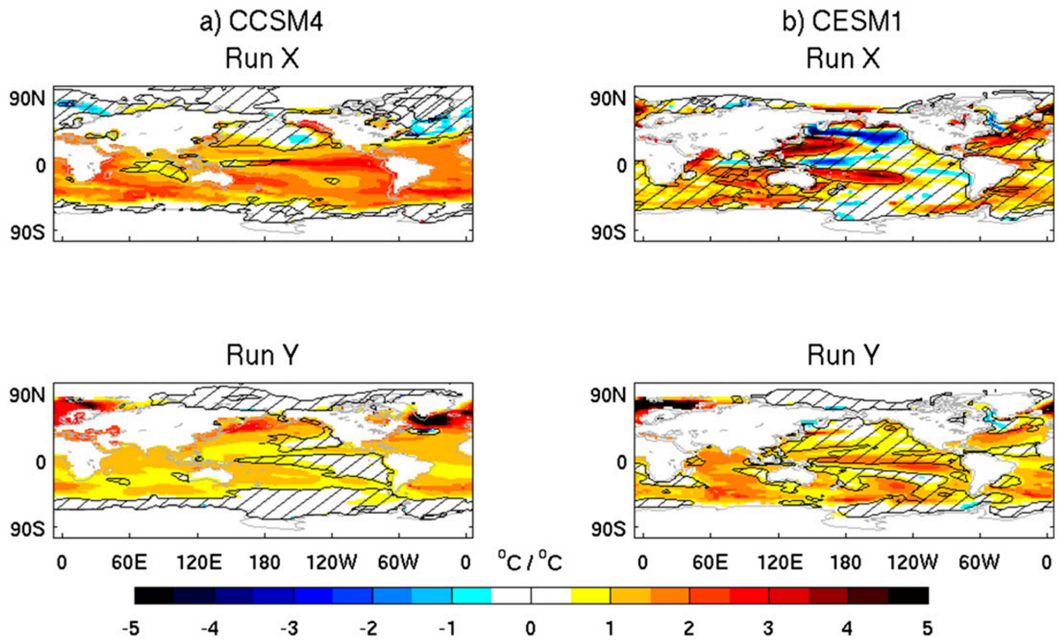

FIG. 6. As in Fig. 5, but for two selected individual members from (a) CCSM4 and (b) CESM1. Areas not significant at 95% according to the Student's t test are hatched.

signal-to-noise ratio is found in the tropics (Fig. 7c). Internal variability must therefore be carefully accounted for when comparing model-based with observation-based patterns. Comparing Fig. 2b with Fig. 7a, we also show that over the period 1900–2005, the median value of the spatial correlation between the observed (Fig. 2b) and the CMIP5 multimodel mean (Fig. 7a) $\tilde{h}(x)$ is about 0.4, with an interquartile range from about 0.3 to 0.5 (not shown). These can be compared to the perfect model potential values of about 0.6 for the median, with an interquartile range from about 0.5 to 0.7 (Fig. 4b). Thus, the correlations obtained are just under half the expected value if the models provided perfect representation of a perfect observational dataset. This low correlation can be explained by imperfections in the method and/or in the CMIP5 models.

We note that there is some sensitivity of these results to whether or not the indirect aerosol effect is included in the CMIP5 models (Table 1; Lohmann and Feichter 2005). Figure 8 is identical to Fig. 7 but separates the CMIP5 models including the indirect aerosol effect (Fig. 8a) from the CMIP5 models excluding them (Fig. 8b). We see that compared to the all-model CMIP5 ensemble mean (Fig. 7a), selecting only the group of CMIP5 models accounting for the indirect aerosol effect (Fig. 8a) 1) does not improve the comparison between the observed and the CMIP5 multimodel mean $\tilde{h}$ discussed, 2) reduces the amount of warming captured in $\tilde{h}$ in particular at high northern latitudes, 3) increases the intermodel spread globally except in the Greenland Sea, and 4) consequently reduces the signal-to-noise ratio.

4. Discussion and conclusions

This work evaluates a novel statistical approach (Hoerling et al. 2011) inspired by classical pattern scaling to estimate, from observations, the spatiotemporal characteristics of the SST response to anthropogenic forcing. We have created various estimates of the SST response based on observations and models, and have performed a series of tests to evaluate the robustness of these responses. We find that the SST spatial response derived from the Hurrell et al. (2008) observational dataset includes warming in the tropical Indian Ocean (in the form of a north–south dipole), the western boundary currents in midlatitudes, the Arctic and subarctic, the southern Atlantic, and the southernmost region of the Atlantic and Indian Ocean (60°–45°S, 60°W–120°E). Principal features of these patterns are mostly reproduced by the CMIP5 multimodel mean and agree with previously published results (including Hoerling et al. 2011). We find that they mostly reflect the linear trend of observed SSTs, within 10%–20% depending on the method chosen to derive the anthropogenic radiative forcing time evolution. We show that despite having a small influence on the spatial pattern of the SST response, the method chosen to derive the anthropogenic radiative forcing time evolution nevertheless

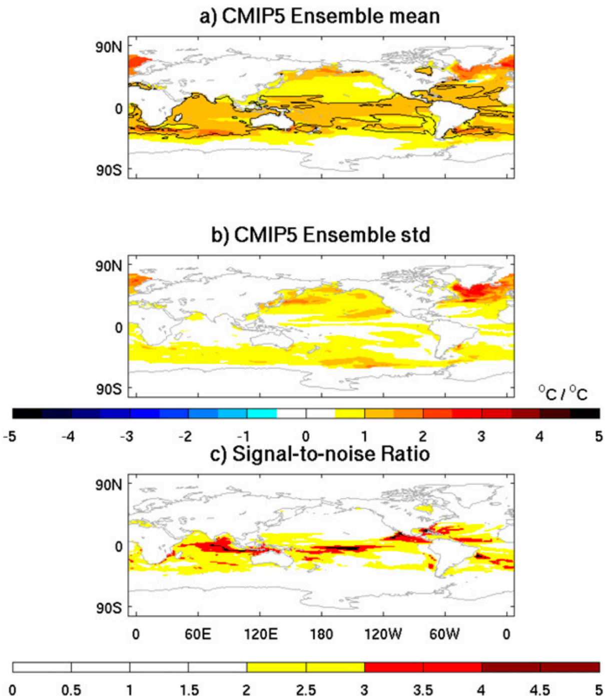

FIG. 7. The value of $\tilde{h}(x)$ computed from the 26 CMIP5 multimodel ensemble member over the time period 1900–2005, and shown as (a) the ensemble mean, (b) the ensemble standard deviation, and (c) the signal-to-noise ratio, inferred as the quotient of (a) to (b). The top two panels have a unit of °C °C^{-1} of global SST warming. The contours in the top panels show areas where the signal-to-noise ratio is higher than 2.

significantly affects the time evolution of the SST response, especially in the mid-twentieth century: In particular, whereas the global-mean and annual-mean SSTs increase more or less linearly through time when produced with the *L* and *C* methods, they show almost no increase until about the 1970s when produced with the Th method. This time evolution difference in global-mean and annual-mean SSTs affects the time evolution of the global warming simulated over land in the corresponding AMIP-type experiments. In turn, this is expected to affect the timing of the associated atmospheric circulation and hydroclimatology changes; these effects will be further considered in subsequent work.

By considering each member from two large initial condition ensembles of coupled climate models as a pseudo-observation, the Monte Carlo (perfect model) analysis highlights the uncertainty associated with deriving the SST spatial response pattern from observations. The test indicates that between 36% and 56% of the spatial variance of the observation-based response pattern described above results from anthropogenic forcing. Our study suggests that similar ratios would apply to other estimates of response patterns derived from observed SSTs over the time period 1920–2005, using either the classical pattern scaling method or a regression method of [e.g., Fig. 7a from Ting et al. (2009), Figs. 2a and 2e from Mohino et al. (2011), and Fig. 1 (top and middle panels) from Hoerling et al. (2011)]. However, we expect higher ratios for the response patterns derived from an ensemble mean of simulations (as opposed to a single ensemble member or to observations), as used in the classical pattern scaling method, since by construction an ensemble mean only includes a limited amount of internal variability (by canceling out the internal variability from individual realization via ensemble averaging). We conclude that deriving the spatial response pattern from an observational dataset instead of an ensemble mean of simulations reduces the uncertainty related to model imperfections, but increases the uncertainty related to internal variability.

The ultimate application of this method is to use our observation-based estimate of anthropogenically forced SSTs and sea ice concentrations to evaluate the land climate response to anthropogenically forced SSTs and sea ice concentrations; an initial test of this application has been presented here (Fig. 3c). We now discuss limitations and uncertainties of this method, and lessons learned to apply going forward.

First, our study shows that the spatial pattern estimate of anthropogenically forced SSTs highly depends on the choice of the SST observational dataset, which can disagree in key areas such as the tropical Pacific (Deser et al. 2010). Then, the time evolution of this estimate depends on the method chosen to derive the anthropogenic radiative forcing time evolution, an issue that might be further explored with model-based approaches (e.g., Ting et al. 2009). Furthermore, our method assumes a linear behavior of the climate system (e.g., SSTs and sea ice concentrations response to anthropogenic forcing is assumed to scale linearly with global-mean and annual-mean SSTs; Santer et al. 1990; Ribes et al. 2010; Tebaldi and Arblaster 2014) that is probably not as simple in reality. In addition, our method does not account for the anthropogenic impact on oceanic internal variability or for the local impact of anthropogenic aerosols on SST and sea ice. The latter leads to additional uncertainties around the mid-twentieth century when aerosol emission rates begin to slow down, and in the future time periods that will be associated with local increase and decrease in aerosol emissions.

Additional uncertainty also comes from the amount of internal variability that is included in the spatial pattern estimate of anthropogenically forced SSTs. This amount can be reduced by increasing the time period chosen for regression (e.g., our study shows that increasing the time

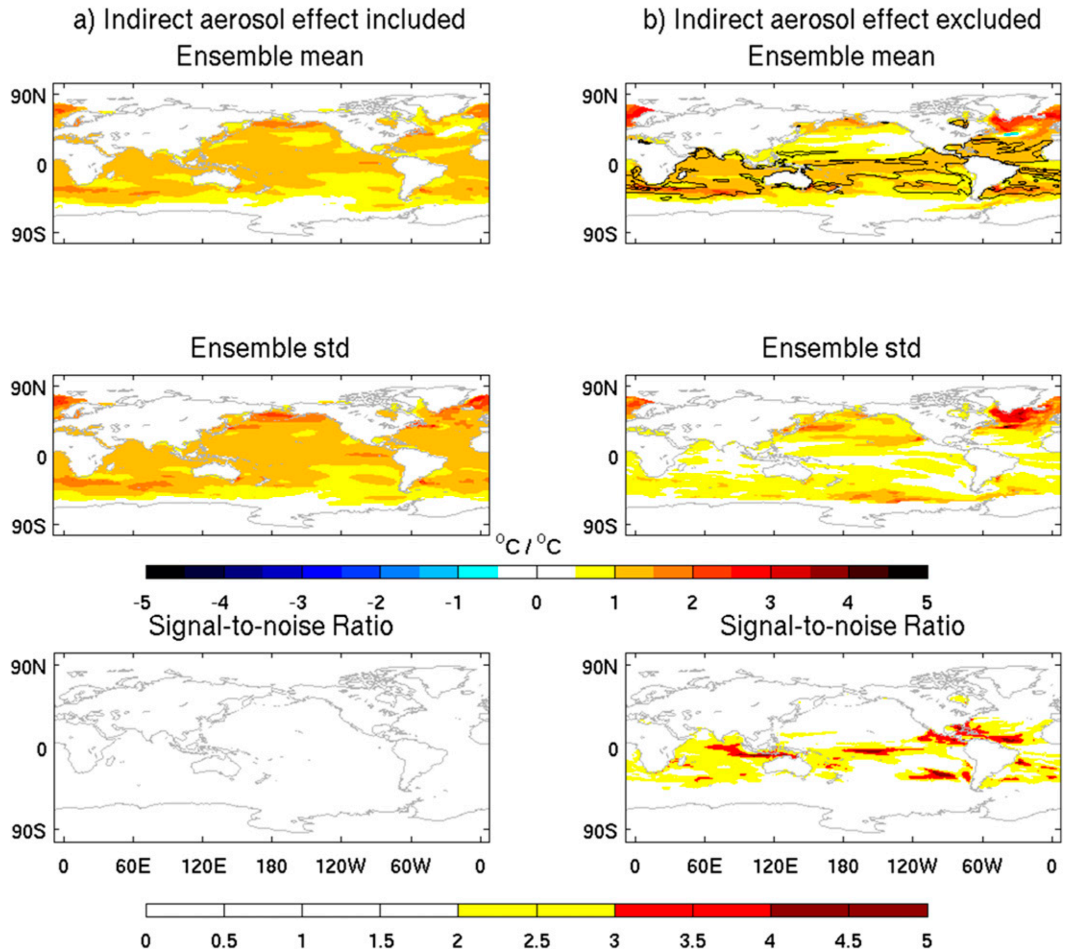

FIG. 8. As in Fig. 7, but separating (a) the 8 CMIP5 models that include the indirect aerosol effect from (b) the 18 models that do not.

period from 1960–2005 to 1920–2005 decreases the amount of internal variability included in the SST spatial response by about 20%) and/or by using an optimal regression (e.g., Hasselmann 1993).

Finally, because AMIP-type experiments, even with prescribed observations of SST, sea ice concentration, and radiative forcing, do not always faithfully reproduce the observed circulation and hydroclimatic response (e.g., Covey et al. 2004), being subject to atmospheric internal variability and inaccuracies in the atmospheric and land models, our proposed approach to simulate the climate response to anthropogenically forced SSTs and sea ice concentrations is still expected to be model dependent, and subject to atmospheric internal variability.

Acknowledgments. The authors thank the National Center for Atmospheric Research (NCAR) for providing access to the CAM5 code and access to the CCSM4 and CESM1 large ensemble data. Original simulations were carried out on the Compute Canada SciNet platform. The authors would also like to acknowledge useful comments by Aurélien Ribes on earlier versions of this manuscript. This research has been supported by the Prévisibilité Climatique Decennale (PRECLIDE) Project, which was funded by the BNP Paribas Foundation and by the Canadian Sea Ice and Snow Evolution Network (CanSISE), which was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) under the Climate Change and Atmospheric Research (CCAR) program.

REFERENCES

Compo, G. P., and P. D. Sardeshmukh, 2009: Oceanic influences on recent continental warming. *Climate Dyn.,* **32,** 333–342, doi:10.1007/s00382-008-0448-9.

Covey, C., K. M. Achutarao, P. J. Gleckler, T. J. Phillips, K. E. Taylor, and M. F. Wehner, 2004: Coupled ocean–atmosphere climate simulations compared with simulations using prescribed sea surface temperature: Effect of a ''perfect ocean.'' *Global Planet. Change,* **41,** 1–14, doi:10.1016/j.gloplacha.2003.09.003.

Deser, C., and A. S. Phillips, 2009: Atmospheric circulation trends, 1950–2000: The relative roles of sea surface temperature forcing and direct atmospheric radiative forcing. *J. Climate,* **22,** 396–413, doi:10.1175/2008JCLI2453.1.

——, ——, and M. A. Alexander, 2010: Twentieth century tropical sea surface temperature trends revisited. *Geophys. Res. Lett.,* **37,** L10701, doi:10.1029/2010GL043321.

Fischer, E. M., and R. Knutti, 2014: Detection of spatially aggregated changes in temperature and precipitation extremes. *Geophys. Res. Lett.,* **41,** 547–554, doi:10.1002/2013GL058499.

Gates, W. L., 1992: AMIP: The Atmospheric Model Intercomparison Project. *Bull. Amer. Meteor. Soc.,* **73,** 1962–1970, doi:10.1175/1520-0477(1992)073<1962:ATAMIP>2.0.CO;2.

Hasselmann, K., 1993: Optimal fingerprint for the detection of time-dependent climate change. *J. Climate,* **6,** 1957–1971, doi:10.1175/1520-0442(1993)006<1957:OFFTDO>2.0.CO;2.

Hoerling, M., A. Kumar, J. Eischeid, and B. Jha, 2008: What is causing the variability in global mean land temperature? *Geophys. Res. Lett.,* **35,** L23712, doi:10.1029/2008GL035984.

——, and Coauthors, 2011: On North American decadal climate for 2011–20. *J. Climate,* **24,** 4519–4528, doi:10.1175/2011JCLI4137.1.

Hurrell, J. W., J. J. Hack, D. Shea, J. M. Caron, and J. Rosinski, 2008: A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. *J. Climate,* **21,** 5145–5153, doi:10.1175/2008JCLI2292.1.

——, and Coauthors, 2013: The Community Earth System Model: A framework for collaborative research. *Bull. Amer. Meteor. Soc.,* **94,** 1339–1360, doi:10.1175/BAMS-D-12-00121.1.

Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) Large Ensemble Project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor.,* doi:10.1175/BAMS-D-13-00255.1, in press.

Lau, N.-C., 1997: Interactions between global SST anomalies and the midlatitude atmospheric circulation. *Bull. Amer. Meteor. Soc.,* **78,** 21–33, doi:10.1175/1520-0477(1997)078<0021:IBGSAA>2.0.CO;2.

Lohmann, U., and J. Feichter, 2005: Global indirect aerosol effects: A review. *Atmos. Chem. Phys.,* **5,** 715–737, doi:10.5194/acp-5-715-2005.

Lu, J., G. Chen, and D. M. W. Frierson, 2008: Response of the zonal mean atmospheric circulation to El Niño versus global warming. *J. Climate,* **21,** 5835–5851, doi:10.1175/2008JCLI2200.1.

Mohino, E., S. Janicot, and J. Bader, 2011: Sahel rainfall and decadal to multi-decadal sea surface temperature variability. *Climate Dyn.,* **37,** 419–440, doi:10.1007/s00382-010-0867-2.

Mudryk, L. R., P. J. Kushner, and C. Derksen, 2014: Interpreting observed Northern Hemisphere snow trends with large ensemble of climate simulations. *Climate Dyn.,* **43,** 345–359, doi:10.1007/s00382-013-1954-y.

Ribes, A., J.-M. Azais, and S. Planton, 2010: A method for regional climate change detection using smooth temporal patterns. *Climate Dyn.,* **35,** 391–406, doi:10.1007/s00382-009-0670-0.

Santer, B. D., T. M. L. Wigley, M. E. Schlesinger, and J. F. B. Mitchell, 1990: Developing climate scenarios from equilibrium GCM results. MPI Rep. 47, 29 pp.

Shin, S.-I., and P. D. Sardeshmukh, 2011: Critical influence of the pattern of tropical ocean warming on remote climate trends. *Climate Dyn.,* **36,** 1577–1591, doi:10.1007/s00382-009-0732-3.

Solomon, A., and M. Newman, 2012: Reconciling disparate twentieth-century Indo-Pacific ocean temperature trends in the instrumental record. *Nat. Climate Change,* **2,** 691–699, doi:10.1038/nclimate1591.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.,* **93,** 485–498, doi:10.1175/BAMS-D-11-00094.1.

Tebaldi, C., and J. M. Arblaster, 2014: Pattern scaling: Its strengths and limitations, and an update on the latest model simulations. *Climatic Change,* **122,** 459–471, doi:10.1007/s10584-013-1032-9.

Thompson, D. W. J., J. M. Wallace, P. D. Jones, and J. J. Kennedy, 2009: Identifying signatures of natural climate variability in time series of global mean surface temperature: Methodology and insights. *J. Climate,* **22,** 6120–6141, doi:10.1175/2009JCLI3089.1.

Ting, M., Y. Kushnir, S. Richard, and C. Li, 2009: Forced and internal twentieth-century SST trends in the North Atlantic. *J. Climate,* **22,** 1469–1481, doi:10.1175/2008JCLI2561.1.

Wallace, J. M., Y. Zhang, and J. A. Renwick, 1995: Dynamic contribution to hemispheric mean temperature trends. *Science,* **270,** 780–783, doi:10.1126/science.270.5237.780.

Wild, M., A. Ohmura, and K. Makowski, 2007: Impact of global dimming and brightening on global warming. *Geophys. Res. Lett.,* **34,** L04702, doi:10.1029/2006GL028031.

Wu, L., and Coauthors, 2012: Enhanced warming over the global subtropical western boundary currents. *Nat. Climate Change,* **2,** 161–166, doi:10.1038/nclimate1353.

Xie, S.-P., C. Deser, G. A. Vecchi, J. Ma, H. Teng, and A. T. Wittenberg, 2010: Global warming pattern formation: Sea surface temperature and rainfall. *J. Climate,* **23,** 966–986, doi:10.1175/2009JCLI3329.1.