

A Bayesian Assessment of Climate Change Using Multimodel Ensembles. Part II: Regional and Seasonal Mean Surface Temperatures

SEUNG-KI MIN AND ANDREAS HENSE

Meteorologisches Institut, Universität Bonn, Bonn, Germany

(Manuscript received 15 December 2005, in final form 18 October 2006)

ABSTRACT

A Bayesian approach is applied to the observed regional and seasonal surface air temperature (SAT) changes using single-model ensembles (SMEs) with the ECHO-G model and multimodel ensembles (MMEs) of the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) simulations. Bayesian decision classifies observations into the most probable scenario out of six available scenarios: control (CTL), natural forcing (N), anthropogenic forcing (ANTHRO), greenhouse gas (G), sulfate aerosols (S), and natural plus anthropogenic forcing (ALL). Space-time vectors of the detection variable are constructed for six continental regions (North America, South America, Asia, Africa, Australia, and Europe) by combining temporal components of SATs (expressed as Legendre coefficients) from two or three subregions of each continental region.

Bayesian decision results show that over most of the regions observed SATs are classified into ALL or ANTHRO scenarios for the whole twentieth century and its second half. Natural forcing and ALL scenarios are decided during the first half of the twentieth century, but only in the low-latitude region (Africa and South America), which might be related to response patterns to solar forcing. Overall seasonal decisions follow annual results, but there are notable seasonal dependences that differ between regions. A comparison of SME and MME results demonstrates that the Bayesian decisions for regional-scale SATs are largely robust to intermodel uncertainties as well as prior probability and temporal scales, as found in the global results.

1. Introduction

Recently there have been increasing studies of climate change detection and attribution focusing on regional-scale surface air temperatures (SATs) and other variables such as precipitation, sea level pressure, and ocean heat contents (International Ad Hoc Detection and Attribution Group 2005, and references therein). Another effort that has been made in the studies is to consider uncertainties originating from intermodel differences that affect the estimation of not only internal variability (noise) but also model responses to given external forcing (signal) and hence detection and attribution results. In this context, Bayesian methods have been suggested as a useful tool (Leroy 1998; Berliner et al. 2000; Min et al. 2004, 2005a; Schnur and Hasselmann 2005; Lee et al. 2005).

The large number of ensemble simulations from 22 coupled climate models, which were integrated for contributing to the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC), enables one to test the sensitivity of climate change assessment to the intermodel uncertainties. Using the dataset of AR4 multimodel ensembles (MMEs) and single-model ensembles (SMEs) of the ECHAM and the global Hamburg Ocean Primitive Equation (HOPE-G) model (ECHO-G), Min and Hense (2006b, hereafter referred to as MH06) classified observed global mean SAT changes over the twentieth century into four scenarios, which are control (CTL), natural (N), greenhouse gas (G), and natural plus anthropogenic forcing (ALL) scenarios, based on a Bayesian method. They found that observed global SAT changes over the whole twentieth century and its second half are classified into ALL scenarios while there is evidence of N and ALL forcing scenarios for the first half of the century. Comparing results from SMEs with MMEs, they demonstrated that the Bayesian assessments for the global mean SATs are not sensitive to intermodel

Corresponding author address: Dr. Seung-Ki Min, Climate Research Division, Environment Canada, 4905 Dufferin Street, Downsview, ON, M3H 5T4, Canada.
E-mail: seung-ki.min@ec.gc.ca

TABLE 1. Climate change scenarios for Bayesian decision analysis.

Abbreviation	Forcing	Relevant model simulations used for "mean" estimation*	
		SMEs	MMEs
CTL	Natural internal variability (control)	ECHO-G_PD	MME_PI
N	Natural forcing (solar and volcanic activities)	ECHO-G_N	—
ANTHRO	Anthropogenic forcing (greenhouse gases and sulfate aerosols)	ECHO-G_ANTH	MME_ANTH
G	Greenhouse gases	ECHO-G_G	—
S	Sulfate aerosols	ECHO-G_S	—
ALL	Natural plus anthropogenic forcing	ECHO-G_ALL	MME_ALL

* PD: present-day control run; PI: preindustrial control run.

uncertainties, supporting previous studies mostly based on conventional approaches such as optimal fingerprinting (e.g., Hegerl et al. 1997; Allen and Tett 1999). Consistently, through evaluating the performance of the 22 AR4 models on reproducing observed global mean SAT variations over the twentieth century as well as its first and second halves, Min and Hense (2006a) showed that models with natural plus anthropogenic forcing together have better skills than those with anthropogenic forcing only, indicating the important role of the natural forcing in explaining observed changes.

Regional-scale SATs are analyzed in several recent studies of detection and attribution (Stott 2003; Karoly et al. 2003; Karoly and Braganza 2005; Zwiers and Zhang 2003; Zhang et al. 2006; Knutson et al. 2006; Min et al. 2005a). Stott (2003) used the Third Hadley Centre Coupled Ocean–Atmosphere GCM (HadCM3) simulations to attribute observed changes of decadal mean SATs over six continental regions for the twentieth century. Applying an optimal detection method, he detected consistent greenhouse warming and sulfate aerosol cooling signals separately over all regions. Using Canadian climate model simulations and an optimal fingerprinting method, Zwiers and Zhang (2003) assessed detectability of anthropogenic signals in SAT patterns from global to continental scales over the twentieth century, and detected greenhouse gas and sulfate aerosol (GS) signals over continental scales including North America and Europe. Zhang et al. (2006) extended the work of Zwiers and Zhang (2003) by using simulations from four models and considering smaller regions. They showed that GS signals are detectable in all domains, confirming previous findings.

Karoly et al. (2003) and Karoly and Braganza (2005) compared trends of simple SAT indices from observations and several coupled climate model simulations over North America and Australia, respectively, and detected GS signals for the second half of the twentieth century. Knutson et al. (2006) assessed skills of the Geophysical Fluid Dynamics Laboratory (GFDL) Cli-

mate Model version 2 (CM2) at simulating regional SAT trends on several regional areas from land and ocean over the twentieth century and its second half and found a better consistency with observations in all-forcing and anthropogenic-only forcing runs than natural-only forcing or no external forcing. Unlike the studies based on conventional statistics (optimal detection or trend analysis), Min et al. (2005a) applied a Bayesian method to the East Asian SAT patterns from ECHAM3/LSG simulations and detected G signals that have been robust to prior changes and spatial scales were retained.

The objective of this study is to extend the multi-model Bayesian study of MH06 to the regional and seasonal mean SATs over six continental regions following Stott (2003). A space–time data vector is constructed by combining Legendre coefficients of regional mean SATs for two or three subregions that constitute continental regions. The block averages will reduce the spatial degree of freedom while the Legendre expansions concern temporal dimensions. Another extension is applied on scenarios. In addition to four scenarios (CTL, N, G, and ALL) used in our previous study, we consider anthropogenic forcing only (ANTHRO) and sulfate aerosol forcing only (S) as other possible explanations of regional SAT changes. As in MH06, SMEs come from the ECHO-G model (Legutke and Voss 1999; Min et al. 2005b,c) and MMEs are composed of the IPCC AR4 22 models. The list of the six scenarios and relevant SME and MME simulations used to define them is given in Table 1.

The methods of Bayesian climate change assessment and temporal refinement (Legendre series expansions) are briefly described in the next section. Observations and the model dataset from ECHO-G and IPCC AR4 models are introduced in section 3. In section 4, structures of detection variables are shown with regional mean SAT time series and their Legendre coefficients for three analysis periods (1900–99, 1900–49, and 1950–99). Bayesian decision results for annual and seasonal

space–time SAT patterns over six continental regions are explained in section 5, where sensitivity to priors, Legendre degrees retained, and intermodel uncertainties are tested. Finally conclusions are given in section 6.

2. Methodology

a. Bayesian decision method

Here we provide a brief description of the Bayesian decision method. A detailed explanation can be found in Min et al. (2004) and MH06. Given the observational vector \mathbf{d} and N defined scenarios m_i ($i = 1, \dots, N$), we want to estimate the probability of each scenario given the observation $P(m_i|\mathbf{d})$, that is, the posterior probability of the scenario. Applying Bayes' rule, the posterior can be evaluated from the prior probability $P(m_i)$ and likelihood function $l(\mathbf{d}|m_i)$:

$$P(m_i|\mathbf{d}) = \frac{l(\mathbf{d}|m_i)P(m_i)}{\sum_{j=1}^N l(\mathbf{d}|m_j)P(m_j)}. \quad (1)$$

The prior represents a subjective belief in the scenario and the likelihood shows the observational probability given the scenario. Assuming multivariate Gaussian distributions for the data vector from the observation and scenarios, the likelihood function can be calculated as

$$l(\mathbf{d}|m_i) = \frac{1}{\sqrt{(2\pi)^q}} \sqrt{\frac{\det \mathbf{A}_i^{-1}}{\det \boldsymbol{\Sigma}_i \det \boldsymbol{\Sigma}_0}} \exp\left(-\frac{1}{2} \Lambda_i\right), \quad (2)$$

where q is the dimension of the data vector, $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_i$ are the covariance matrices of the observation \mathbf{d} and the scenario m_i , respectively, \mathbf{A}_i is a function of the covariance matrices as $\mathbf{A}_i = \boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_0^{-1}$, and Λ_i indicates a generalized distance between the observation and scenario expressed as $\Lambda_i = (\mathbf{d} - \boldsymbol{\mu}_i)^T (\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_0)^{-1} (\mathbf{d} - \boldsymbol{\mu}_i)$, where $\boldsymbol{\mu}_i$ is mean of the scenario m_i (simplified from Min et al. 2004).

The posterior probability calculated from Eq. (1) can be used as a decision function (Duda and Hart 1973; Berger 1985). We select the scenario with maximum posterior (the most probable scenario) into which observations are classified. In case of identical prior assumption between scenarios, the Bayes factor (likelihood ratio) itself becomes a decision function. According to the descriptive scales suggested by Kass and Raftery (1995), the logarithm of the Bayes factors larger than 1, 2.5, and 5 represent “substantial,” “strong,” and “decisive” observational evidences for the scenario (or against a reference scenario, here

CTL), respectively (Lee et al. 2005; Schnur and Haselmann 2005; MH06).

Since a sort of significance is tested many times using different scenarios and different prior values, one might think that this Bayesian method might be related with traditional multiple comparisons. In the multiple comparisons (or testing) one is required to adjust the significance level to a much smaller magnitude in order to consider the possible increase of “false positive” results (type I errors) arising from repeated tests. This is not the case in our Bayesian approach. Given a prior, we always test significance once using two scenarios, one of forced scenarios (N, ANTHRO, G, S, and ALL), and a reference scenario (here CTL), through the Bayes factor.

One can test the sensitivity of Bayesian decision results to covariance matrices of the observation and scenario ($\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_i$), which are normally estimated from model simulations. MH06 found little effect of the uncertainties in the estimated covariance matrices when they carried out several sensitivity tests for global assessment results to different covariance matrices from the multimodel (Table 4 of MH06). In this study we assume that the insensitivity to covariance matrices holds for regional-scale assessment as well.

As discussed in MH06, a major limitation of the Bayesian decision method is that its result definitely depends on the scenarios that a user applies. Hence one should define proper scenarios carefully, which can explain observed change better. In the case that all scenarios (model simulations) are a bad fit to the observations, the Bayesian decision might be made for the best of the worst. It would be useful to make some additional measures of “absolute” skills for the scenario. One way can be to compare the likelihood of the scenario to that of an idealized scenario (rather than to that of CTL). Evaluating coupled climate models, Min and Hense (2006a) suggested constructing a reference model (m_r), which has a mean identical to the observation [$\boldsymbol{\mu}_r = \mathbf{d}$ or $\Lambda_r = 0$ in Eq. (2)], while its internal variability is obtained from CTL. That method, however, tends to be too strict for higher-dimensional analysis like this study using space–time vectors (Min and Hense 2007). Yet another possibility is to monitor the absolute values of the likelihood. If these are small it would mean that the simulations are on the fringes of the data distribution. Min et al. (2004) gave an example of this type of behavior.

A main difference between our Bayesian decision method and the optimal fingerprint methods, for example, by Stott (2003), exists in the way of comparing the model simulations with the observation. The opti-

mal fingerprint uses the regression coefficient (i.e., beta factor) for the rescaling of the model signal amplitude to provide a better fit to the observed change. The Bayesian decision method measures a generalized distance Λ_i between the observation and model simulations (defined as scenario) as a sort of error measure in the likelihood of Eq. (2), and weights the prior probability of the scenario with the likelihood (negative exponential function of Λ_i) to finally produce the posterior probability. Thereby a scenario (or model) with larger Λ_i will have smaller likelihood, and hence posteriors, and vice versa.

b. Legendre series expansions

To avoid singular or near-singular matrices in the likelihood calculation a dimension reduction method is required. We take Legendre series expansions following MH06. Regional SAT time series are decomposed into temporal components representing overall warming (scale), linear trend (trend), and shorter-term decadal and interannual fluctuations. The truncation point of Legendre polynomials (LPs) is based on the models' ability to simulate internal variability. For the evaluation, we analyze power spectra of regional mean SATs from the MME control runs. These results, to be presented below in section 4b, show that models can simulate well the internal variability over most of the subregions on decadal time scales. This implies that one can truncate LPs at up to the 12th degree, which corresponds to a 20-yr (10 yr) period for 100-yr (50 yr) time series (MH06).

Additionally, in our space–time approach, the dimension of detection variable \mathbf{d} also increases by a factor of 2 or 3 according to the number of subregions (see next section for regional domains). Actually we apply LP truncation up to the first five degrees only (from LP0 to LP4) so that the maximum dimension of space–time vector becomes 10 or 15. Then the resolved time scales by the first 5 LPs are longer than 50 yr for the 100-yr period, which would be 25 yr for the 50-yr period, estimated from power spectra of LP4 (not shown). However, it should be noted that time scales are not clearly filtered in the LP structure since multiple time scales are contained (see MH06). Testing the computational limits to avoid singular matrices leads to different maximum dimensions of order 20 to 30 depending on the range of internal variability and signal amplitude. The finally chosen conservative limit of 10 or 15 is well below the critical dimension.

3. Data and experiment

As the SAT observations we use the Climate Research Unit (CRU) data over land (CRUTEM2v) for

1900–99 (Jones and Moberg 2003). Regional domains are applied following Stott (2003). There are six continental-scale regions—North America (NAM), Asia (ASI), South America (SAM), Africa (AFR), Australia (AUS), and Europe (EUR)—with each region comprising two or three subregions (Fig. 1). Combining Legendre coefficients of area-averaged SATs over the subregions constructs the space–time data vector for the six continental regions from observations and model simulations.

There are two sources for the model dataset. At first, data come from SME simulations with ECHO-G. ECHO-G is a coupled climate model of the Model and Data Group at the Max Plank Institute for Meteorology that has the atmospheric component ECHAM4 with a horizontal resolution of T30 and 19 vertical levels, and the ocean model HOPE-G with a T42 equivalent horizontal resolution (a gradual equatorial refinement in meridional direction up to 0.5° near equator) and 20 vertical levels (Legutke and Voss 1999). We use the SAT dataset from a 1000-yr present-day control run (ECHO-G_PD; Min et al. 2005b,c) and historical ensemble simulations over the period 1860–2000 for the five forced scenarios (ECHO-G_N, ECHO-G_ANTH, ECHO-G_G, ECHO-G_S, and ECHO-G_ALL; Table 1). Detailed descriptions about the historical simulations can be found in MH06. ECHO-G_PD provides 91 overlapping samples of 100-yr regional SAT time series for the CTL scenario with a moving window of 100-yr length shifted by 10 yr. From the ECHO-G forced simulations, we obtain three realization samples of 1900–99 (five from ECHO-G_ALL).

Simulations from 22 coupled models are used as the second source of model data, which can be downloaded from the IPCC AR4 archive (http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php). Information about models and experiment details can be found at the same place. We use historical climate change (twentieth-century climate in coupled models; 20C3M) and preindustrial (PI) control runs. The 20C3M simulations are divided into two groups on the basis of the external forcing implemented: MME_ANTH (12 models with 25 members) and MME_ALL (12 models with 48 members). MME_PI and MME_ALL are exactly the same simulations as in MH06. MME_ANTH (anthropogenic forcing only) is newly added in this study and consists of 12 models with total 25 members (Table 2). Only two models, ECHO-G and the Hadley Centre Global Environmental Model version 1 (HadGEM1), provide samples for MME_ANTH as well as MME_ALL. We extract overall 25 and 48 nonoverlapping samples of 100-yr time series of regional mean SATs from MME_ANTH and MME_ALL, respectively. For

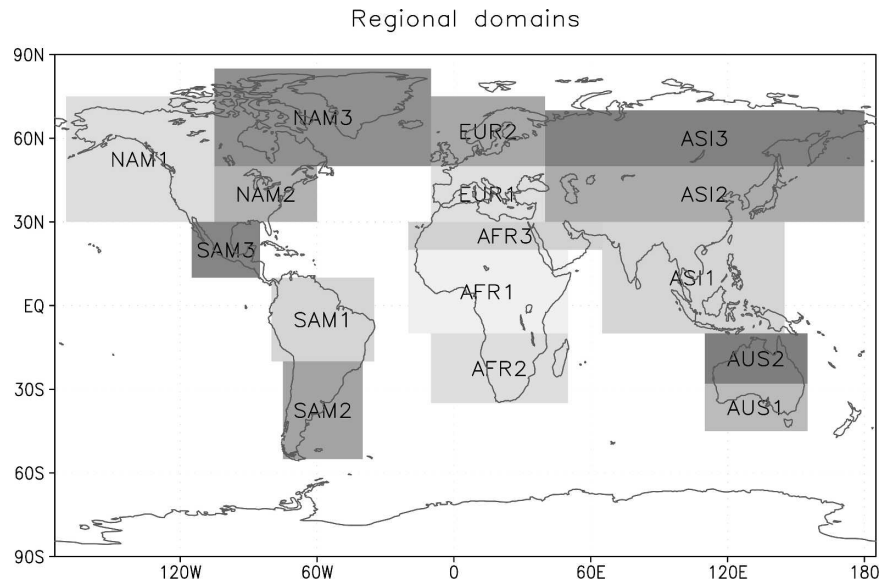


FIG. 1. Regional domains used in this study following Stott (2003). There are six continental regions: NAM, ASI, SAM, AFR, AUS, and EUR, with each composed of two or three subregions. The subregions are expressed as the attached number after the abbreviation of the continental region.

TABLE 2. List of IPCC AR4 coupled climate models used in this study. The twentieth-century climate change simulations of the IPCC AR4 models are divided into MME_ALL (natural plus anthropogenic forcing) and MME_ANTH (anthropogenic forcing only) according to external forcing implemented. Ensemble members of the models are used as nonoverlapping samples for MME_ALL and MME_ANTH. Preindustrial control simulations of the IPCC AR4 models provide overlapping samples for MME_PI that are obtained from a moving window of 100-yr length with a 10-yr shift. More detailed information about MME_PI sampling can be found in MH06.

Model	MME_ALL (ensemble member)	MME_ANTH (ensemble member)	MME_PI (No. of 100-yr overlapping subsections)
BCCR_BCM2.0	—	1	16
CCSM3	8	—	55
CGCM3.1(T47)	—	5	41
CGCM3.1(T63)	—	1	26
CNRM-CM3	—	1	30
CSIRO-Mk3.0	—	3	29
ECHAM5/MPI-OM	—	3	58
ECHO-G	5	3	25
FGOALS-g1.0	—	3	3
GFDL-CM2.0	3	—	41
GFDL-CM2.1	3	—	41
GISS-AOM	—	2	32
GISS-EH	5	—	21
GISS-ER	9	—	41
INM-CM3.0	1	—	24
IPSL-CM4	—	1	14
MIROC3.2 (high res.)	1	—	1
MIROC3.2 (med. res.)	3	—	41
MRI_CGCM2.3.2	5	—	26
PCM	4	—	49
HadCM3	—	1	25
HadGEM1	1	1	5
SUM (models)	48 (12)	25 (12)	644 (22)

MME_PI, overlapping 100-yr-long moving windows with a 10-yr shift produce 644 samples. MH06 found a negligible effect of using a different number of samples from MME_PI (80 nonoverlapping versus 644 overlapping samples) on the Bayesian decisions for global mean SATs.

For each 100-yr-long sample, Legendre coefficients are obtained for the whole period and the first and second 50 yr, which correspond to observational periods of 1900–99, 1900–49, and 1950–99. Then the coefficients are used to estimate means and covariance matrices for the likelihood calculation in Eq. (2). For the mean estimation, we use only SMEs for the G, S, and N scenarios for which MMEs are not available, while SMEs or MMEs are selectively used for CTL, ANTHRO, and ALL. Covariance matrices from the forced scenarios are assumed to be identical to that of CTL, which is estimated from SMEs or MMEs. We refer to these two kinds of settings with SMEs and MMEs as the SINGLE and MULTI experiments, respectively. The same assumption is applied for the covariance matrix of the observation. The SAT anomaly is relative to the first 20 yr in every 100-yr sample. Prior to calculating regional averages, model data are interpolated to and masked with the observational grids.

It should be noted that differences between MME_ANTH and MME_ALL may partly result from model differences as well as forcing differences. For instance, MME_ANTH includes more models from different modeling groups than MME_ALL although one cannot assess independence between the models correctly. Besides, here we have not removed the possible effect of climate drift in the 20C3M runs. That is to say, 20C3M SAT anomalies are calculated directly from the 20C3m run, not as differences from the appropriate PI control run. We have only excluded a few model simulations that have noticeable climate drift defining MME_PI as in Table 3 of MH06. The remaining drift would increase the range of internal-model uncertainty as discussed in MH06.

4. Structures of detection variables

a. Time series of regional temperatures

Figure 2 shows low-pass-filtered time series of area-averaged SAT anomalies over the 16 subregions from the CRU observations and the model simulations with different forcing relevant to the six scenarios. Filtered series for 1900–99 are obtained from Legendre degrees retained at the 12th degree (see below for the truncation). Results from SMEs and MMEs are displayed together for the three scenarios of CTL, ANTHRO, and ALL, while only SME outputs are shown for N, G, and

S. Observational SAT changes exhibit consistent warming trends over most of the subregions in recent decades. Observations are also characterized by an early warming near the 1940s, which is most obvious for the areas NAM1–3 and SAM3. However, temporal behavior and amplitude of the early warming varies significantly between subregions. The subregions ASI1, AUS1, and AUS2 do not carry any early-warming signals.

Unforced model simulations (CTL) reveal different ranges of internal variabilities in the twentieth-century SAT changes in different subregions. As a whole, the variabilities are larger over high latitudes, especially near the North Atlantic (NAM3 and EUR2) while they get smaller over low latitudes such as ASI1, SAM1, AFR1, AFR2, AUS1, and AUS2. The former is consistent with regions of principal natural variability like the Arctic Oscillation (AO) or North Atlantic Oscillation (NAO), as shown in previous analyses of SAT variability patterns (e.g., Stouffer et al. 2000; Collins et al. 2001; Min et al. 2005b). Min et al. (2005b) reported that the ECHO-G model overestimates the SAT variability over the North Atlantic and North Pacific in time scales from years to decades. This explains why over NAM3 the variability range of ECHO-G_PD is larger than that of MME_PI. On the other hand, the variability in MME_PI is stronger in EUR2. One might deduce a conclusion that the recent observed warming is outside of this range of internal variability in some subregions while within the range in other subregions. However, such interpretation highly depends on model simulations that provide samples for internal variability estimation. Therefore, the multimodel approach should be pursued in this respect. We will examine the effect of the multimodel on Bayesian decisions, where the internal variability range is a key factor as a background noise, by comparing results from SINGLE with MULTI below.

Natural forcing simulations (N) with ECHO-G are characterized by a warming near the middle of the twentieth century that is profound over a few subregions (NAM3, ASI3, and EUR1), but with large uncertainty between ensemble members. All together they cannot explain the recent increasing trend in the observations. Greenhouse gas forcing simulations (G) represent a steady warming over the whole century in most of the places. However, G runs seldom reproduce the observed warming around the 1940s, for example, over the NAM subregions. On the other hand, cooling trends after the 1960s are dominant in the sulfate aerosol runs (S), which are stronger ($< -1.0^{\circ}\text{C}$ by 1999) over the northern subregions NAM2, ASI2, ASI3, and EUR2.

The ANTHRO simulations with ECHO-G show less

Regional Mean SATs 1900–1999

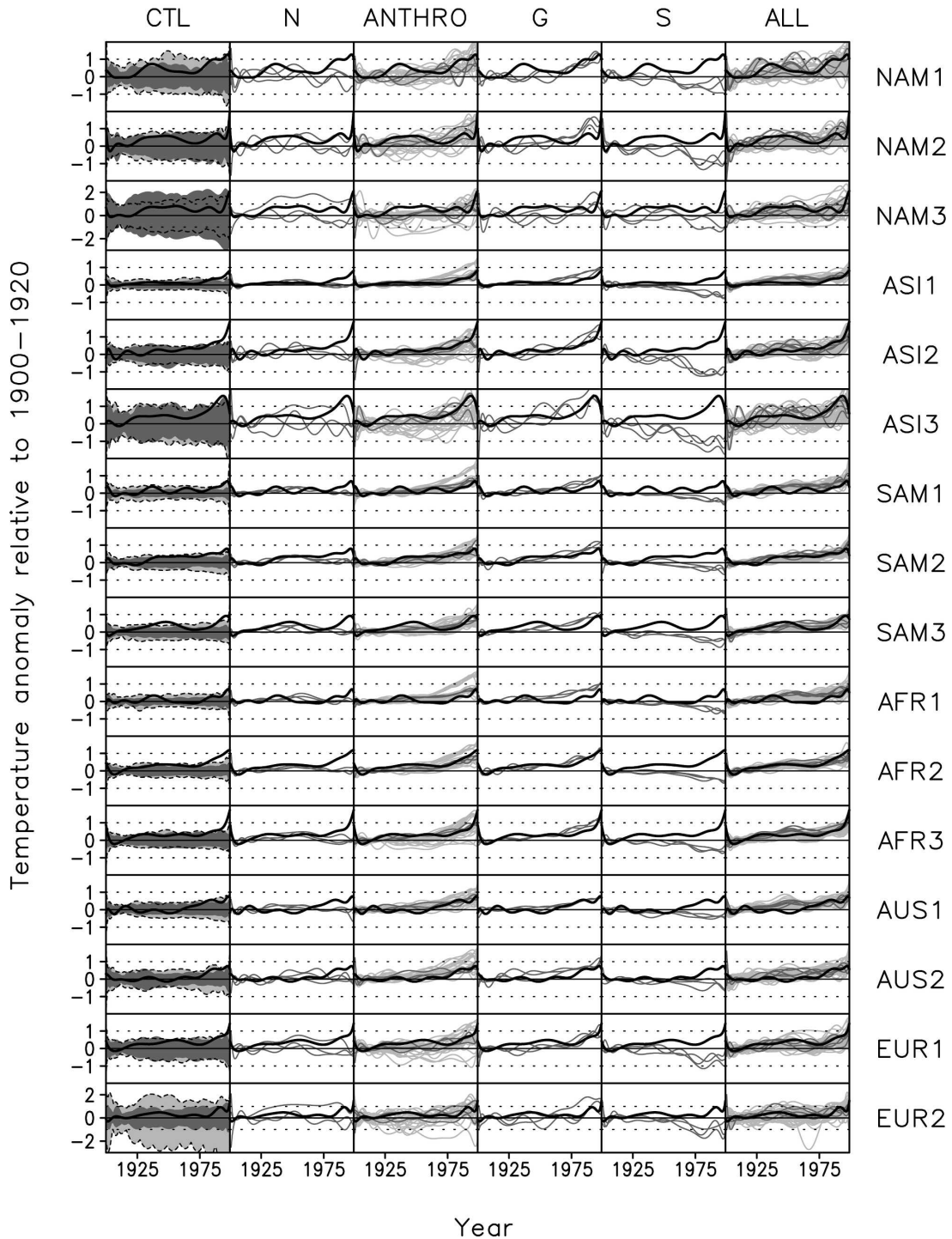


FIG. 2. Reconstructed time series (retained at the 12th degree of Legendre polynomials) of regional annual mean SAT anomalies during 1900–99 for the 16 subregions (Fig. 1) from CRU observations (thick black), SME (light gray) simulations, and MME (dark gray) simulations under different forcing factors. Note that MMEs are available only for CTL, ANTHRO, and ALL and that only max and min ranges are plotted for SME (also dashed) and MME in CTL. Temperature anomalies are relative to 1900–20 means.

recent warming than G through an offset by the S cooling. MME results have a larger range than SMEs in the simulations but tend to overestimate the recent warming in some subregions like AFR1, AFR2, SAM1, and AUS2. This indicates possible effects of intermodel difference in responses to external forcing. Although ALL runs from SMEs or MMEs capture well the observed behavior over most of the subregions, deficiencies occur over ASI1, SAM1, AFR1, AUS1, and AUS2 where models overestimate the early century warming. The intuitive and qualitative comparisons described above are quantified below through Bayesian analyses considering model uncertainties as well as natural variability.

b. Power spectrum analysis

We apply a power spectrum analysis to evaluate the model performance at simulating internal variability of regional-scale SAT. Based on this analysis, we can also decide on the temporal truncation of the LP up to which models represent reasonable variabilities compared to those observed. As model simulations, we use 10 and 80 nonoverlapping samples of ECHO-G_PD and MME_PI (MH06), respectively. Model power spectra are calculated for each sample and the ensemble average of the power spectrum and its range (maximum and minimum) are analyzed in comparison to observations. The 100-yr time series from the observations and simulations are detrended prior to analysis to remove influences of external long-term forcing. It should be kept in mind, however, that such detrending may not be enough to eliminate effects of external forcing that are delivered from decadal or interannual time scales, especially from observations (Jones and Hegerl 1998; Stouffer et al. 2000).

Power spectra of regional-mean SATs obtained from ECHO-G_PD are given in Fig. 3 with those from CRU observations. As in the result for the global mean SATs (MH06; Stone et al. 2007), regional SATs from observations and SME simulations over most of the subregions have larger power in the lower frequency, which corresponds to the red variance spectra of stochastic climate models where slowly varying “climate” variability is explained as the integral response of randomly excited short-term “weather” disturbances (Hasselmann 1976). In terms of decadal variability, ECHO-G simulations show a similar power spectrum compared to observations in most subregions. Exceptions occur in some subregions (SAM1, SAM3, AFR1, AFR2, and AFR3) where the model underestimates the SAT decadal variability and in NAM3 where the variance is overestimated by the model. The larger variability over NAM3 is a specific feature of ECHO-G as discussed in Fig. 2. Min et al. (2005b) found a 2-yr spectral peak in

the global mean SATs from the ECHO-G_PD, which is hardly seen in the observation. Its occurrence is related to too strong and frequent El Niño events simulated by the model (Min et al. 2005c). The 2-yr spectral peak can also be found in the regional SATs mainly over lower latitudes (ASI1, SAM1, AFR1, AFR2, and AUS2), which are the areas of significant responses to the El Niño–Southern Oscillation (ENSO) in the model (see Fig. 10 of Min et al. 2005c).

If we extend the power spectrum analysis to MME_PI (Fig. 4), we get two major changes of model spectra. The first is a broadening of the spectral ranges and the other is an improved consistency between simulated multimodel mean powers and observed ones. These changes happen over most of the subregions. Especially the improved model performance at simulating decadal variability emerges over subregions where overestimation or underestimation by SMEs occurs (see above). This indicates that the advantages of using MMEs over SMEs hold in the regional climate change assessment as well as global studies (MH06). Considering the limit in the spectral comparison (e.g., using linear detrending to remove external influences from observations), overall results suggest that, even over smaller subregions, models can reproduce decadal variability of SATs reasonably and one can refine temporal decompositions at least up to decadal components. According to MH06, this corresponds to retaining Legendre coefficients at the 12th degree for the period of 1900–99 (see below).

c. Legendre coefficients

Figure 5 displays the Legendre coefficients of regional SATs for 16 subregions during 1900–99. The observed patterns show that in general there is a warming over most of the subregions (positive values of LP0 and LP1), but with amplitude varying from region to region relative to each range of the internal variability. Observed Legendre coefficients at the 4th degree (LP4) are also positive over some subregions (e.g., NAM1, NAM2, SAM3, AFR3, and EUR1). These are associated with the early century warming near the 1940s, although LP4 may contain warming signals from recent decades as well (for LP patterns, see Fig. 1 of MH06). Model simulations with different forcing factors show that the coefficients from ALL and ANTHRO runs are closer to observational patterns than those from the other runs (N, G, and S). In some regions, positive coefficients of LP4 are dominant in N and ALL runs, which are also evident in the observations. This represents a possible role of natural forcing since LP4 contains the early warming near the 1940s as explained above. The distribution of internal variabilities of Leg-

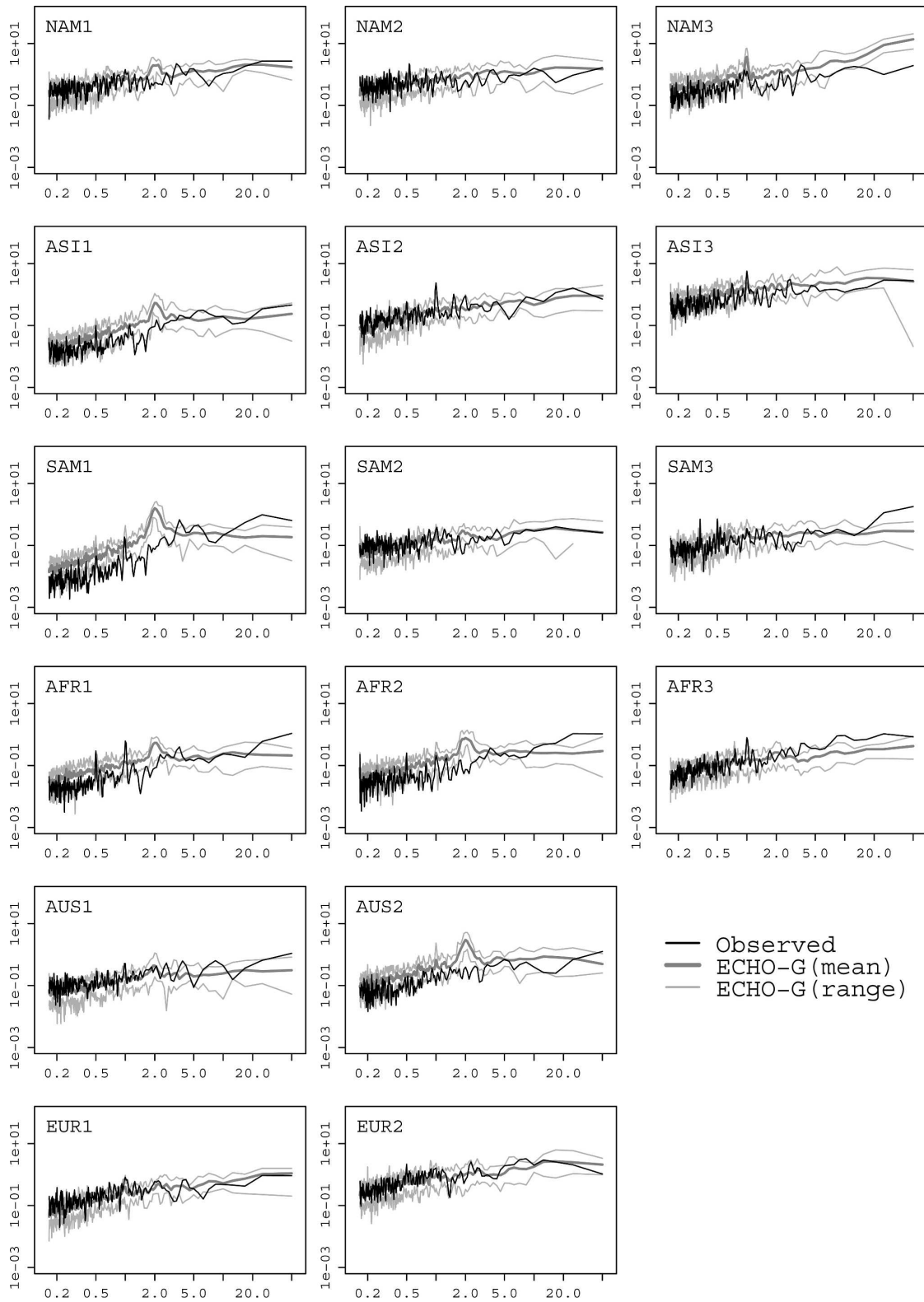


FIG. 3. Power spectra of regional monthly mean SAT anomalies during 1900–99 for the 16 subregions (Fig. 1) obtained from a 1000-yr present-day control run with ECHO-G (ECHO-G_PD). The abscissa is frequency (year) and the ordinate is spectral density (nondimensional).

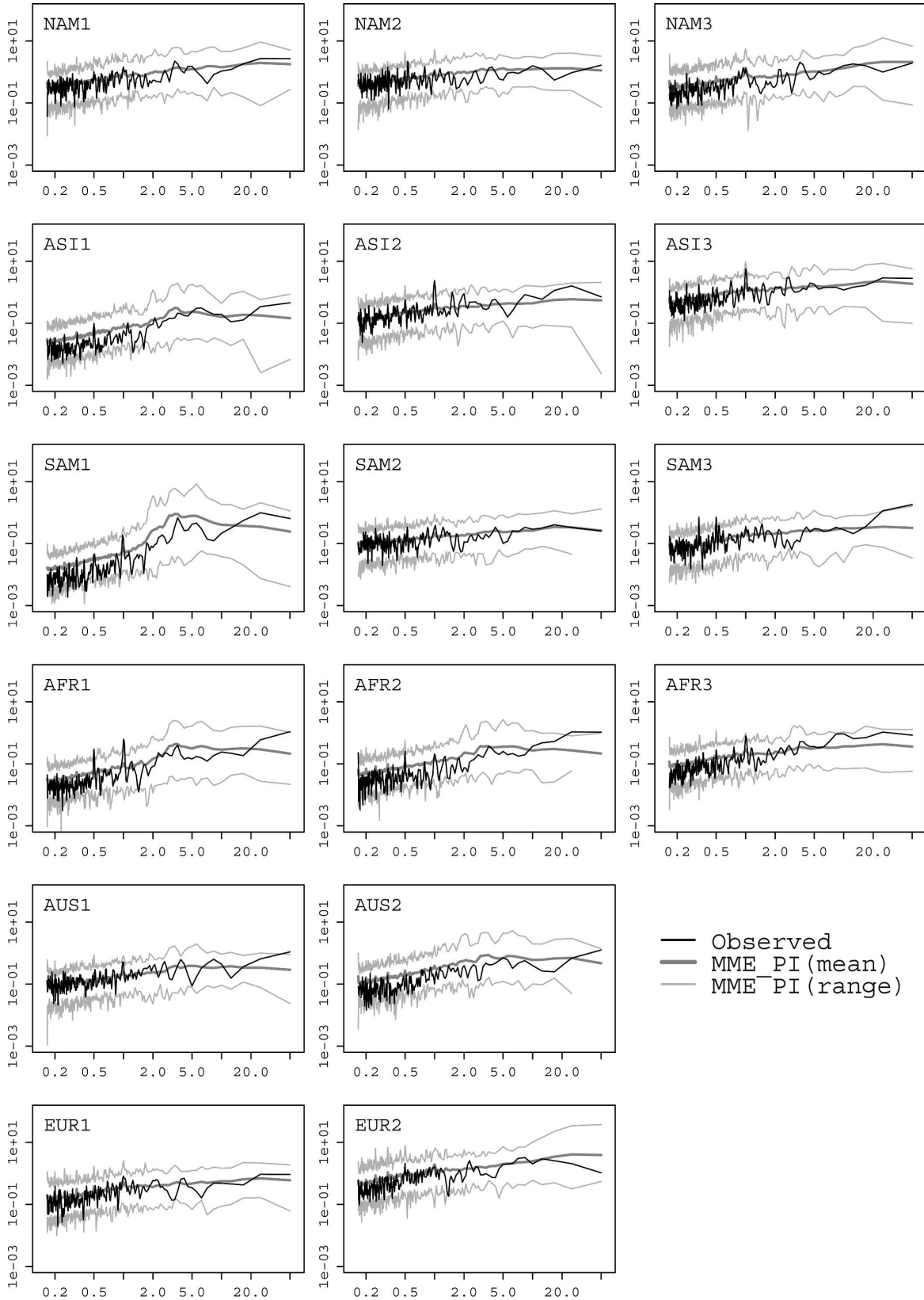


FIG. 4. Same as in Fig. 3, but for preindustrial control runs with IPCC AR4 models (MME_PI).

Legendre Coefficient for SATs 1900–1999

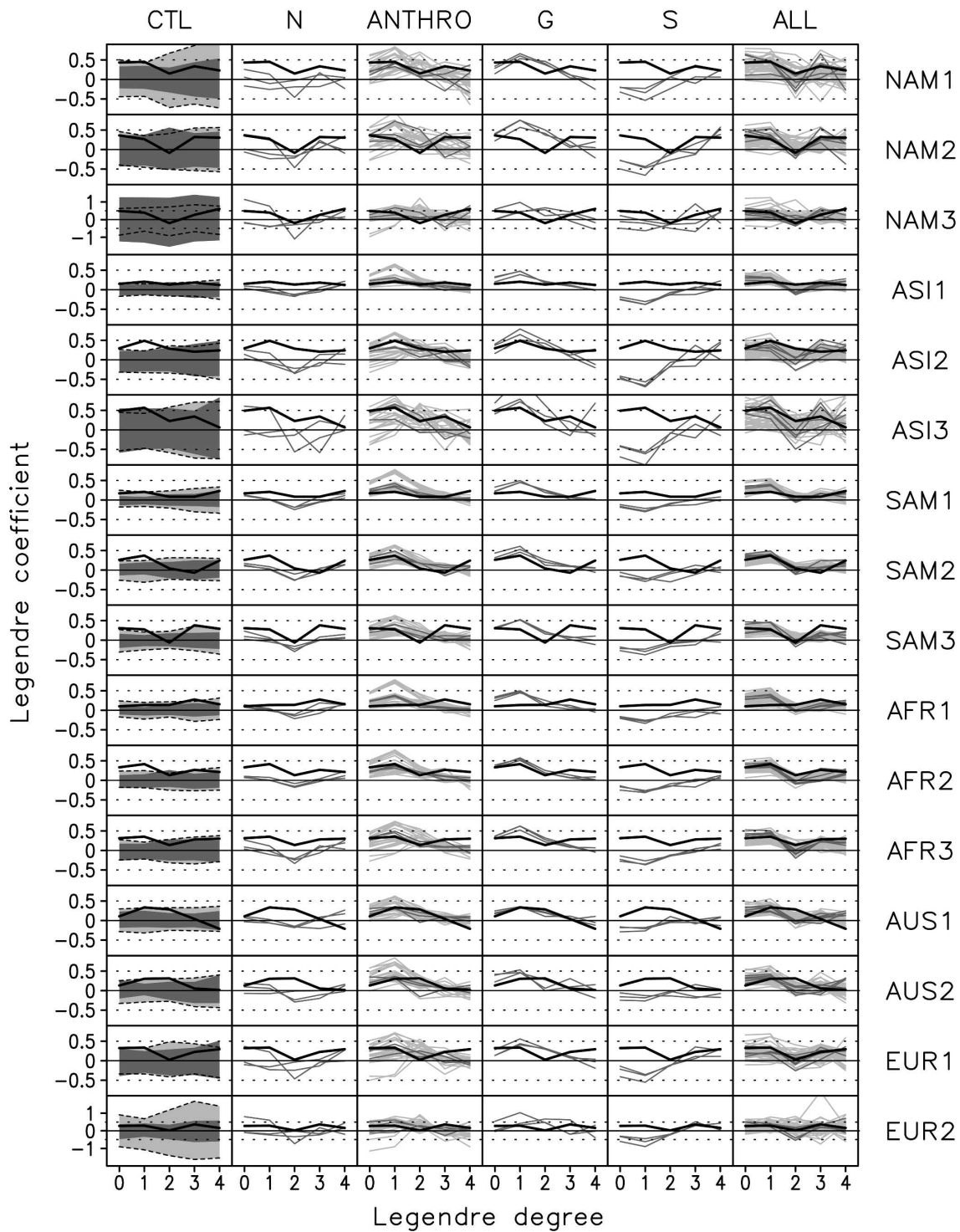


FIG. 5. Legendre expansion coefficients (at zeroth to fourth degree) of regional annual mean SAT anomalies for 1900–99 over the 16 subregions from CRU observations (thick black), SME (light gray) simulations, and MME (dark gray) simulations under different forcing factors. Note that MME results are available only for CTL, ANTHRO, and ALL and that only maximum and minimum ranges are plotted for SME (also dashed) and MME in CTL.

endre coefficients is well matched with SAT variabilities in Fig. 2. The ranges of SMEs are similar to those of MMEs over large parts of subregions, while notable differences are found over a few subregions: SME variability is smaller than MMEs over NAM1 and EUR2, while SME has a larger range over NAM3.

Simulated and observed Legendre coefficients for the first half of the twentieth century are shown in Fig. 6. Except for several subregions such as ASI1, ASI2, SAM1, AUS1, and AUS2, observed LP0 and LP1 are showing positive values, which means overall warming and increasing trend of SATs for this period. Model simulations with different forcing factors show that N and ALL runs are able to capture the observed warming while other runs cannot reproduce positive values of LP0 and LP1.

Figure 7 shows Legendre coefficients for the second half of the twentieth century (1950–99). Overall distribution of LP coefficients is very close to that for 1900–99 (Fig. 5). Compared to the results in 1900–49 when some regions do not exhibit any warming (Fig. 6), observations in 1950–99 are characterized by clearer warming patterns in most of the subregions where the warming signals are stronger than background noise simulated with models without external forcing (i.e., internal variability ranges of CTL). However, the warming is very weak in NAM2, NAM3, and EUR2 compared to the noise arising from larger SAT variability over the North Atlantic region. In the model simulations, ANTHRO and ALL runs exhibit better skill at reproducing the amplitude of observed warming. The G runs tend to overestimate the warming trend (LP1) over some subregions while S runs are characterized by consistent cooling over most of the subregions. The N runs display positive values at LP0 (warming relative to 1900–20) but weak negatives at LP1 (the trend within this period). Through Bayesian decision analysis below, we will quantify the comparison or similarity measure between the observed and simulated Legendre coefficients with consideration of spatial and temporal patterns together.

5. Bayesian decision results

a. Identical-prior case

In case of identical priors, the Bayes factor (or likelihood ratio) can be used as a decision function to classify observations into the most probable scenario. Figure 8 shows the distributions of the Bayes factors for five scenarios (N, ANTHRO, G, S, and ALL) with respect to CTL over six continental regions from the SINGLE experiment. The dimension of the data vector increases hierarchically in steps of 2 or 3 depending on

the number of subregions (Fig. 1) when going stepwise from LP0 to LP4. Three analysis periods of 1900–99, 1900–49, and 1950–99 are applied as before. The descriptive scales of Bayes factors by Kass and Raftery (1995) are represented as shadings in the figure, which provide substantial, strong, and decisive observational evidence for the forced scenario (for the CTL) when the logarithm of the Bayes factor is larger than 1, 2.5, and 5 (less than -1 , -2.5 , and -5), respectively.

For the regional SATs of the entire twentieth century, ALL (for NAM and SAM), ANTHRO (for ASI and AFR), and G (for AUS) are the scenarios of maximum Bayes factors. Signals are amplified with increasing LP retained and decisive pieces of evidence (log of Bayes factor >5) are found over all regions except for EUR. Over EUR, ALL and ANTHRO have similar Bayes factors, but with weaker substantial amplitudes. When applying the Bayesian decision method to search for the most probable scenario, one should take into account the closeness (or discernibleness) of the scenarios. For regional SATs during 1900–99, ANTHRO and ALL signals are very close to each other over all regions, and the three scenarios of ALL, ANTHRO, and G are grouped in AUS. This implies indistinguishable responses of models to different combinations of external forcing factors that prevent one from selecting single scenario. Hence it will not be reasonable to select or decide only one scenario as a cause of observed change. Rather one should consider common signals emerging from multiple scenarios with at least strong scales in this case. For the twentieth-century SAT over continental regions, ANTHRO is the common signal (except for AUS where G is the common signal).

Bayes factors for 1900–49 show that only two regions of SAM and AFR have larger-than-strong signals of N, ANTHRO, G, and ALL. The N signal appears slightly dominant over SAM while disorganized patterns are seen over AFR. It is hard to deduce a single common signal from the four signals since N and G are independent. Instead we conclude that both N and G contribute separately to the total changes with similar amplitudes for this period. Results for the second half (1950–99) resemble those for the whole twentieth century. A difference is that the ANTHRO signal is larger than ALL in SAM. Again ANTHRO and ALL share a similarity over all regions except for AUS where the G signal is strongest. Concerning this G signal in AUS, this might be related with the underestimated response of ECHO-G to greenhouse gas forcing over the region (see below).

Figure 9 shows the distributions of the Bayes factors from the MULTI experiment. The MULTI experiment has some different settings compared to SINGLE. First

Legendre Coefficient for SATs 1900–1949

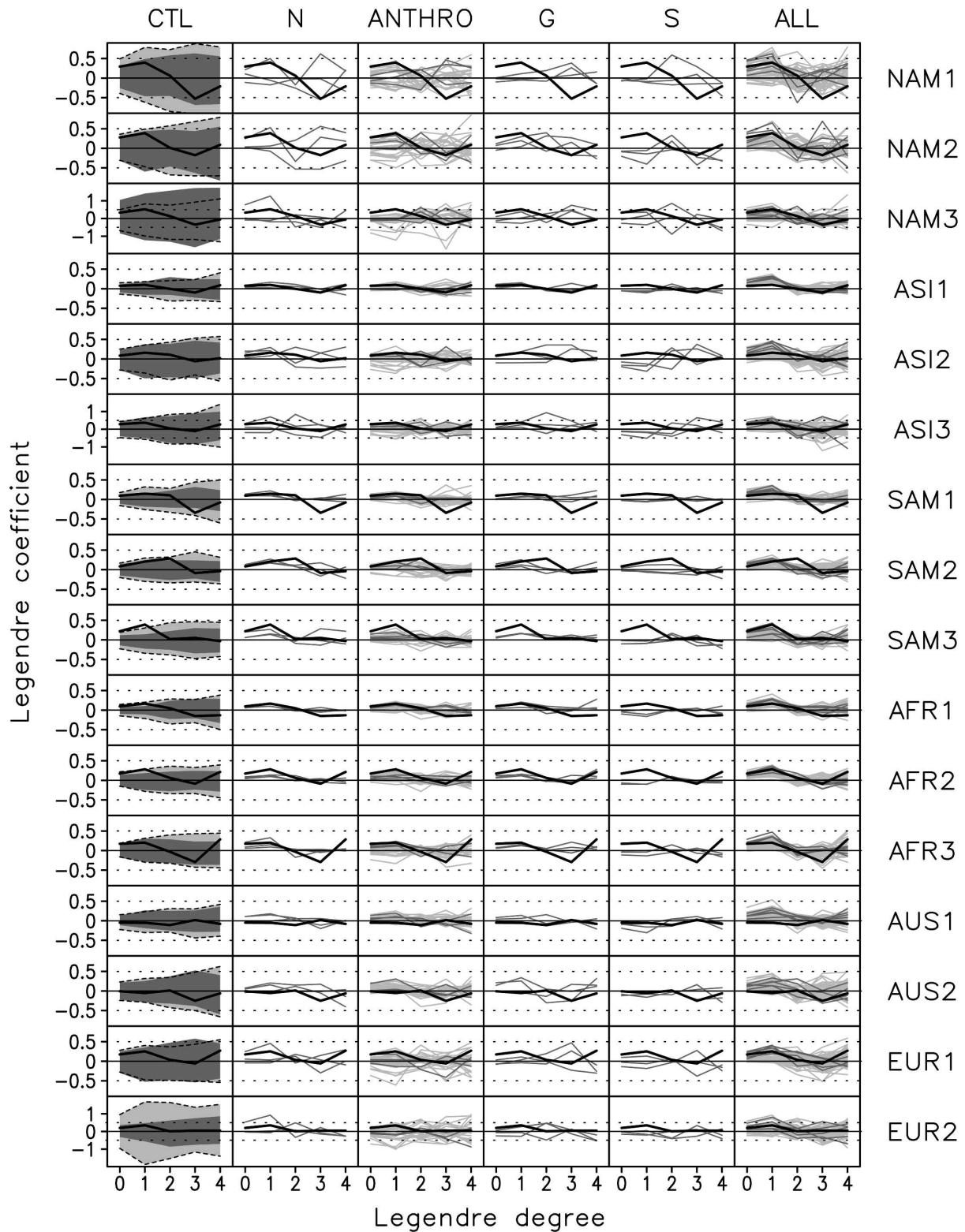


FIG. 6. Same as in Fig. 5, but for 1900–49.

Legendre Coefficient for SATs 1950–1999

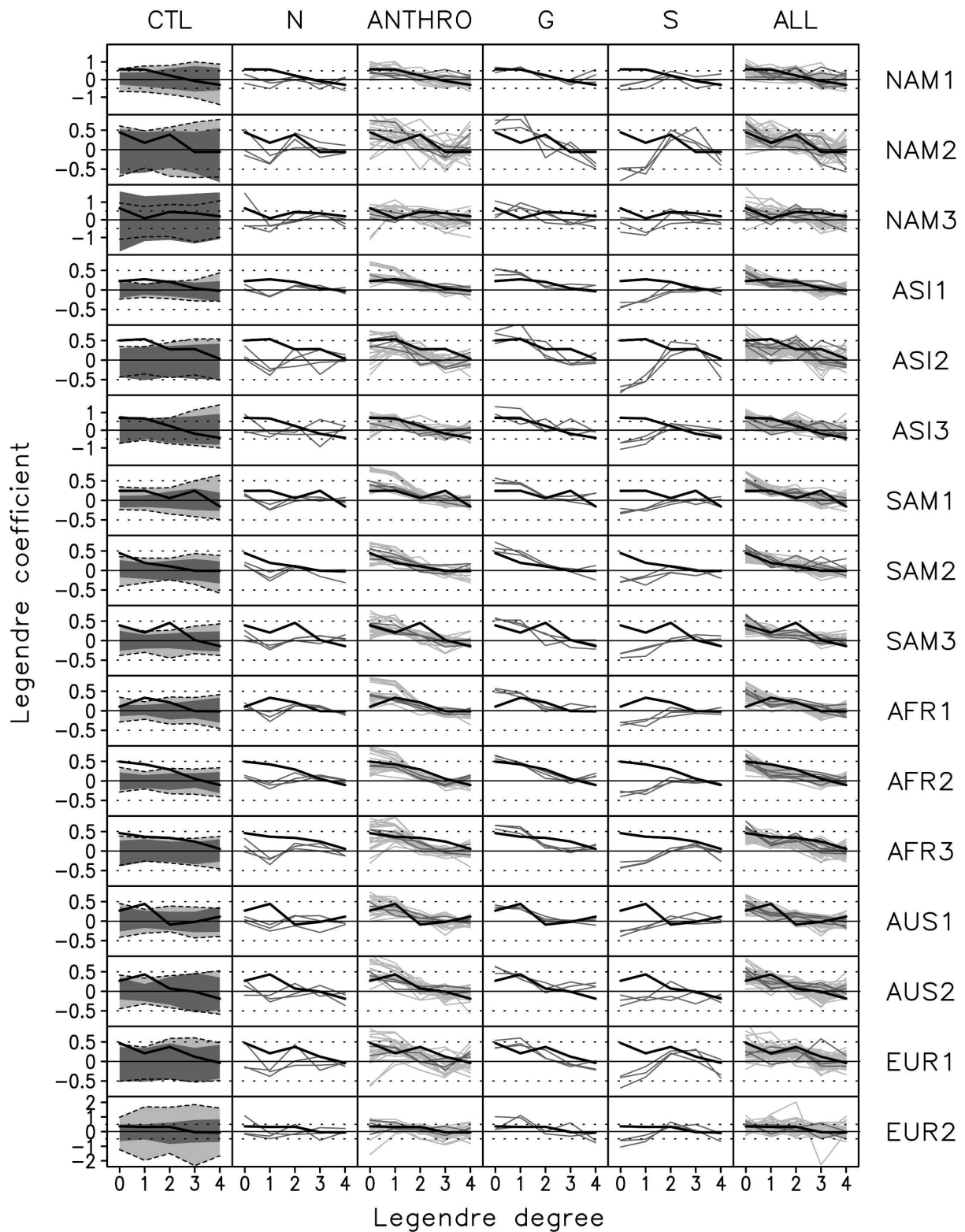


FIG. 7. Same as in Fig. 5, but for 1950–99.

Bayes Factors for Regional SATs from SINGLE

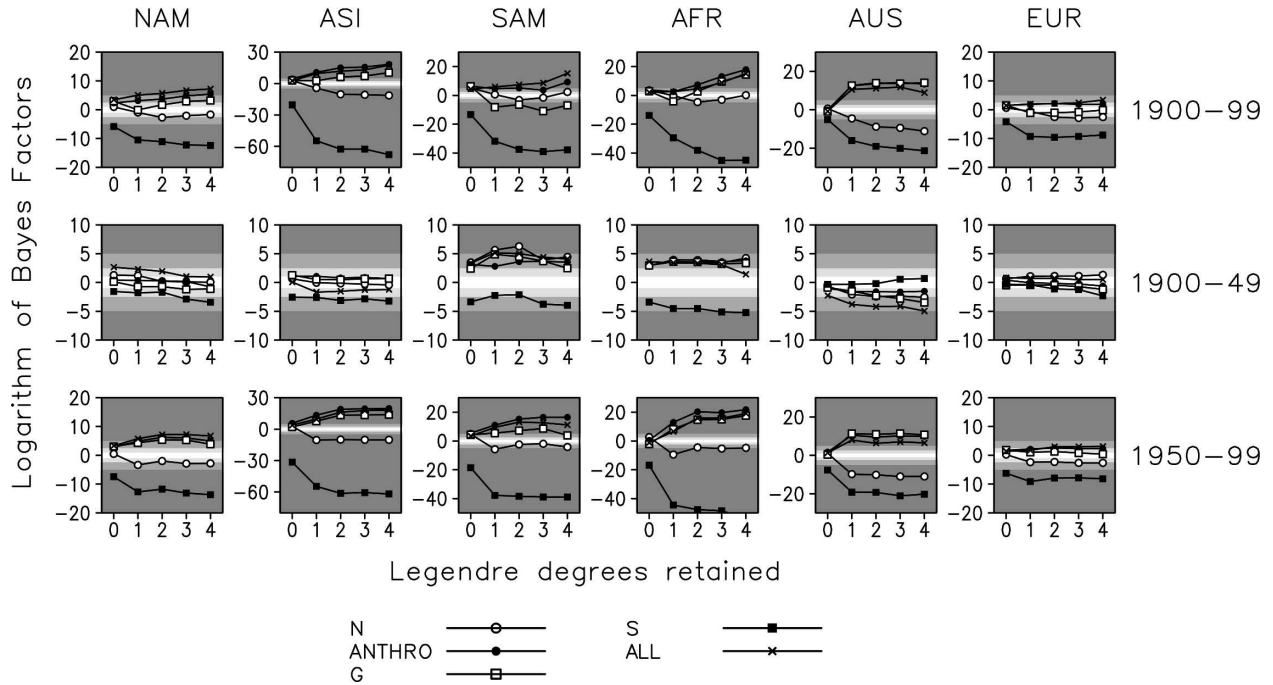


FIG. 8. Bayes factors of the five forced scenarios (N, ANTHRO, G, S, and ALL) with respect to the CTL for annual mean SAT anomalies during 1900-99, 1900-49, and 1950-99 over six continental regions (Fig. 1) from the SINGLE experiment. Gray shadings represent (from light to dark) substantial, strong, and decisive observational evidence, respectively, for the forced scenario.

Bayes Factors for Regional SATs from MULTI

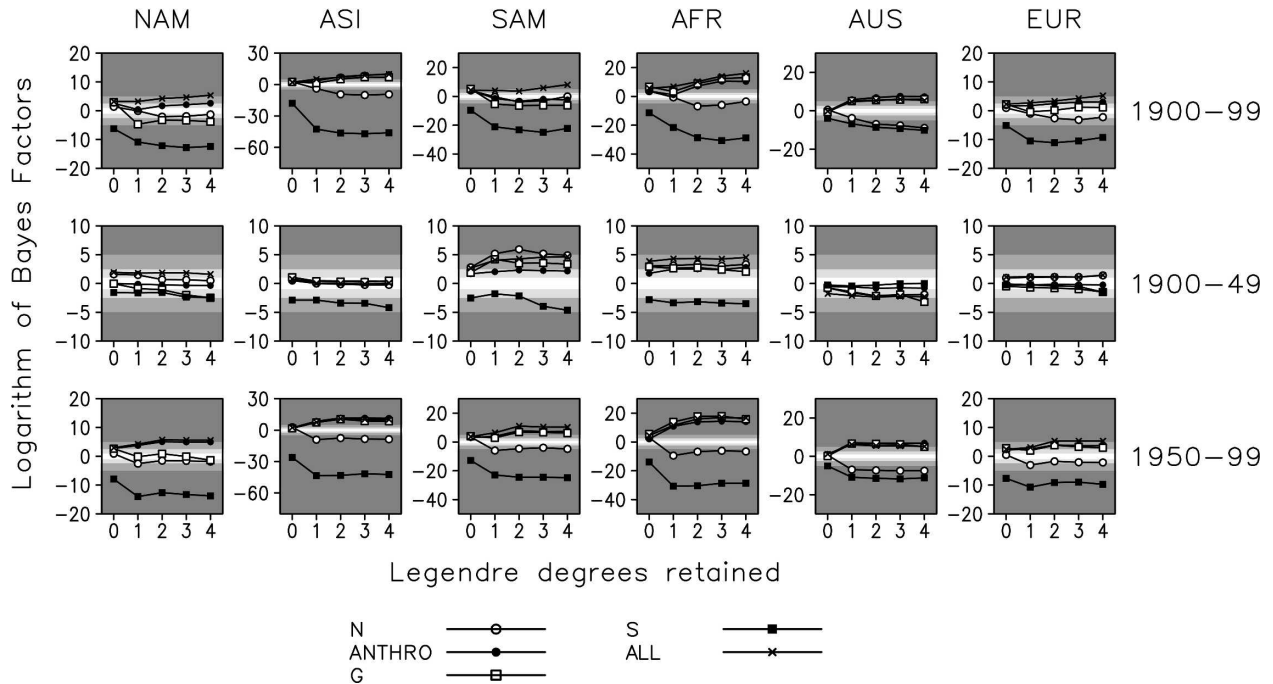


FIG. 9. Same as in Fig. 8, but for the result from the MULTI experiment.

we take MME_PI to estimate the parameters (means and covariance matrices) for the CTL scenario, which affects Bayes factors for the other scenarios. Additionally MME_ANTH and MME_ALL runs are used for defining ANTHRO and ALL scenarios. For the three analysis periods, Bayes factors from MULTI are very similar to those from SINGLE. Some minor changes can be identified. For 1900–99, ALL becomes more dominant over NAM, SAM, AFR, and EUR, while the ANTHRO signal is largest over AUS with the closeness between the signals unaffected. For 1900–49, the ALL signal is enhanced over AFR. The MME effects on Bayes factors for 1950–99 are characterized by a strengthening of ALL signals over SAM and EUR and of ANTHRO signal over AUS as in 1900–99, while G has maximum Bayes factors over AFR. As a whole, patterns of Bayes factors for regional SATs are insensitive to intermodel uncertainties in cases of identical priors, which is consistent with previous studies for global mean SATs (MH06). We test the sensitivity of this conclusion to varying priors below.

b. Varying-prior case

Priors of six scenarios are changed following the method by MH06 using “uninformed” uniform priors. We assume that the total sum of priors of six scenarios is always unity and that priors of the forced scenarios except for CTL are assumed to be identical. Then the prior of CTL is shifted from 0.01 to 0.99, which corresponds to varying priors of the forced scenarios from 0.198 to 0.002. By this we consider a range of prior probabilities from weighting the forced scenarios by 19.8 ($=0.198/0.01$) times of CTL (when CTL prior is 0.01) to weighting CTL by 495 ($=0.99/0.002$) times of the other scenarios (when CTL prior is 0.99). Introducing a similar definition of priors to ours with three scenarios (CTL, G, and GS), Schnur and Hasselmann (2005) applied two settings of varying priors. They assigned 90% and 10% probabilities to CTL (degree of belief in that the observed climate change can be explained by natural variability), which represent two “extreme” views from a “climate change skeptic” and a “climate change advocate,” respectively. We here take a wider range of CTL priors from 1%–99%, which provides more extensive views on climate change although other methods of prior modeling could be envisaged.

In the case of this generalized setting with varying priors, posterior probability is used as a decision function. Given Legendre coefficients and prior values, we can evaluate posterior probabilities of six scenarios using Eq. (1). Then the most probable scenario is decided from a decision rule of selecting the scenario of maximum posteriors. Figure 10 shows the decision results as

a function of Legendre degrees and priors that are obtained from the SINGLE experiment. Each decision box over six continental regions can be interpreted as an extension of the one-dimensional results in Fig. 8 into two-dimensional plots utilizing the prior as a vertical axis. It should be noted that, however, only maximum scenarios are displayed in Fig. 10 without other information such as signal amplitudes of each scenario and their similarities or differences. Therefore distributions of Bayes factors in Fig. 8 should be analyzed together to interpret the Bayesian decision results more clearly, especially when treating various scenarios.

Bayesian decision results for 1900–99 in Fig. 10a are characterized by three scenarios of ALL, ANTHRO, and G, which are largely insensitive to prior changes (vertical axis) and temporal truncations (horizontal axis) as in the global results (MH06). Interestingly, a regional dependence of the decision patterns appears: ALL signals over NAM and SAM, ANTHRO signals over ASI and AFR, and G signals over AUS. Europe has the weakest external forcing signals, losing ALL and ANTHRO signals for CTL prior larger than 0.6, which might be related to the large uncertainty range of ECHO-G over EUR2 (Fig. 2 and Fig. 5).

For 1900–49, forced scenarios are decided only over SAM and AFR and are composed of N, ANTHRO, and ALL (Fig. 10b). As discussed above in Fig. 8, the N signal is dominant over SAM, but a mixture of N, ANTHRO, and ALL appears over AFR. The ALL signal over NAM is seen only for smaller priors of CTL. Observations over EUR, ASI, and AUS are classified into CTL. This means that N signals arising from global mean SATs for this period (MH06) are contributed mainly by the two low-latitude continental regions. Considering that temperature responses to solar forcing are more profound over the Tropics in some coupled model simulations (Cubasch et al. 1997; Meehl et al. 2003; Rind et al. 2004) and observations (van Loon et al. 2004), the decision patterns evident over SAM and AFR in the first half of the twentieth century might be related to model response to solar forcing. It is also possible that this pattern might be related to low-frequency variations in the Atlantic Ocean, affecting the tropical Atlantic but not the Pacific (cf. Delworth and Knutson 2000; Knutson et al. 2006).

The stronger solar influence on the tropical climate originates from spatially heterogeneous solar input, which affects more directly relatively cloud-free areas, compared to spatially uniform forcing of greenhouse gases (Meehl et al. 2003). However, the amplitude of the regional response to solar forcing in Cubasch et al. (1997), Meehl et al. (2003), and ECHO-G results used

Bayesian Decision for SATs from SINGLE

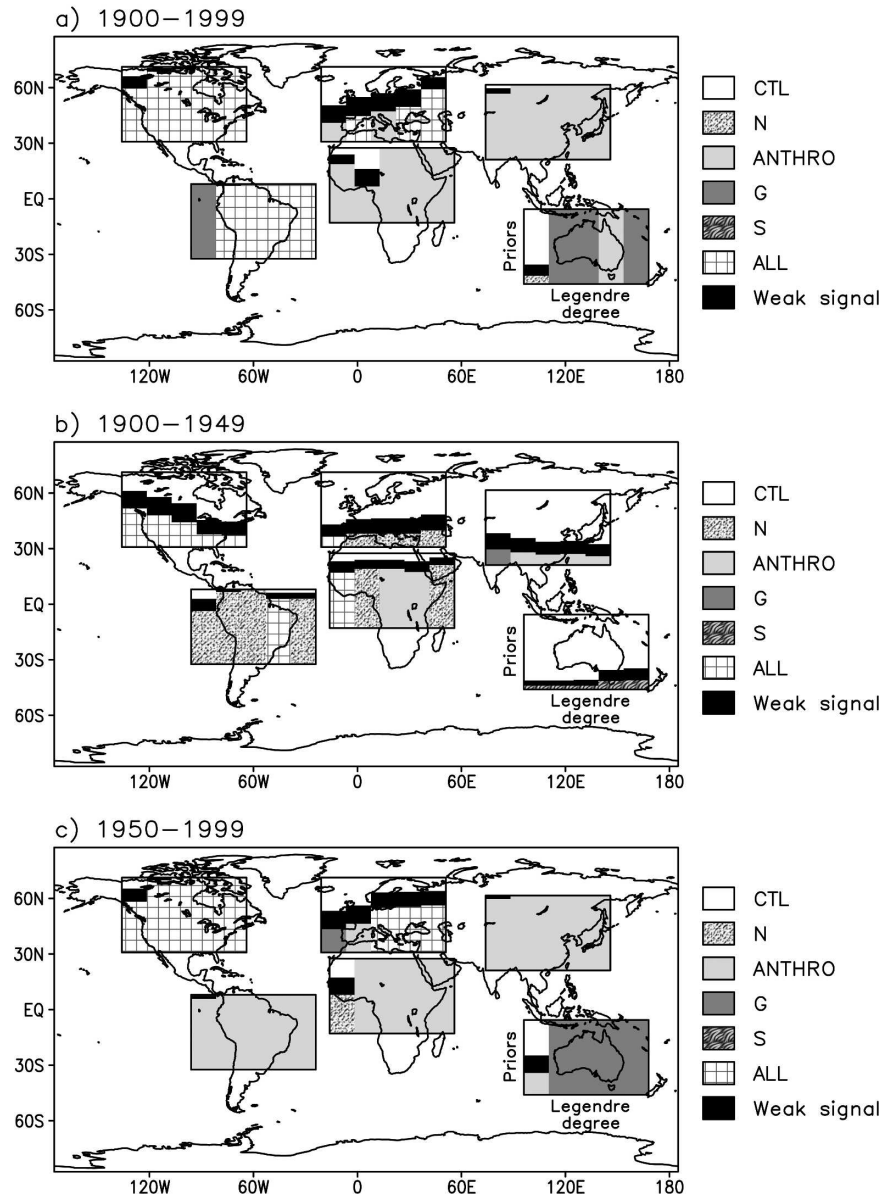


FIG. 10. Bayesian decisions for annual mean SAT anomalies during 1900–99, 1900–49, and 1950–99 over six continental regions (defined as in Fig. 1) from SINGLE. Each two-dimensional box represents the most probable scenario out of the six scenarios (CTL, N, ANTHRO, G, S, and ALL, as indicated by shadings and hatching patterns) when Legendre truncations are shifted from zeroth to fourth degree in the horizontal axis and priors of CTL (and corresponding priors of the five forced scenarios) are varied from 0.01 to 0.99 (0.18 to 0.002) along the vertical axis. Black area in the decision plot indicates area of weak signal where the logarithm of the posterior odds of the decided scenario with respect to CTL are less than 1; i.e., the climate change signal is “not worth more than a bare mention” (Kass and Raftery 1995). Under the situation of no significant signals, decisions will be white or black.

here might be a bit overestimated by using solar irradiance rather than spectral descriptions. Rind et al. (2004) found effects of total and spectral solar forcing to be that the tropical response was somewhat greater

with total irradiance while the stratospheric response was greater with the spectral forcing. Meehl et al. (2003) also suggested a possible nonlinear amplification of the regional response when solar forcing occurs in

Bayesian Decision for SATs from MULTI

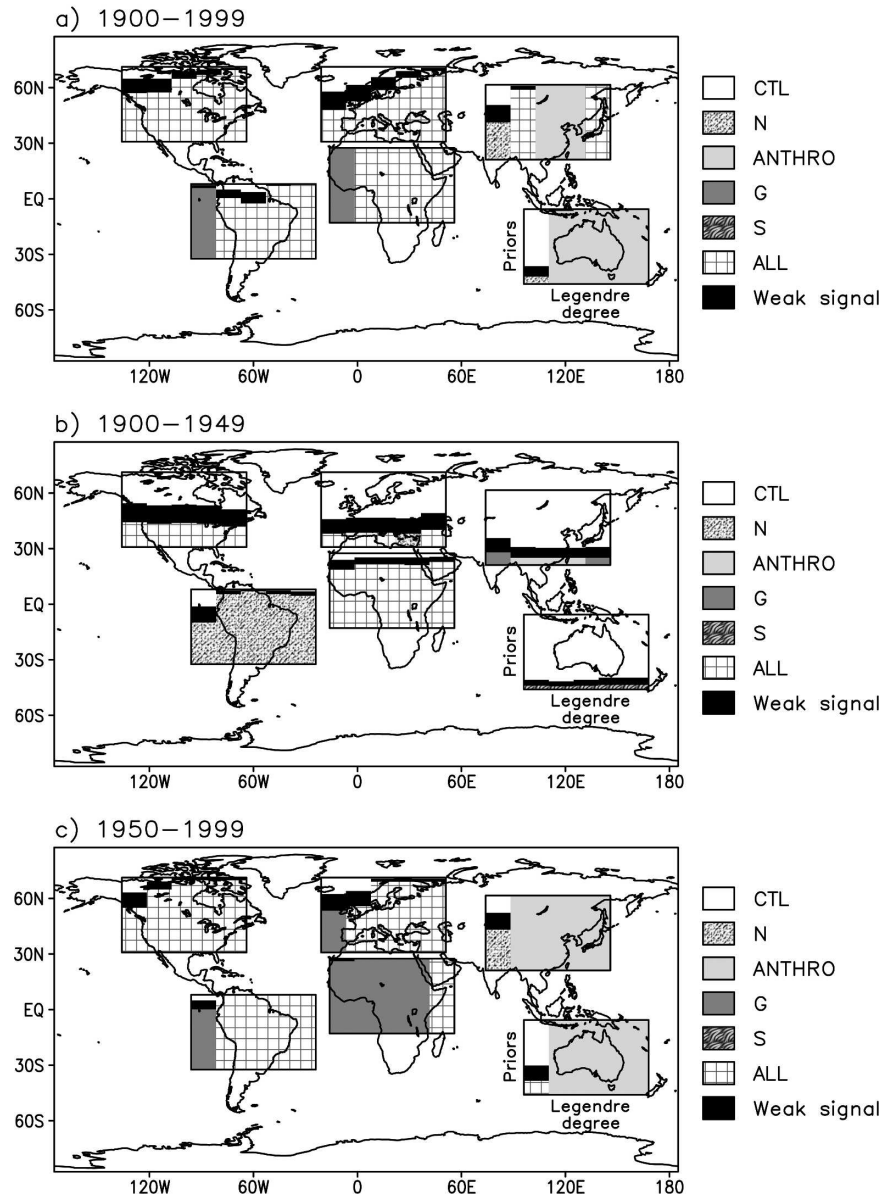


FIG. 11. Same as in Fig. 10, but for the result from MULTI.

combination with anthropogenic forcing, which might be model dependent with the cloud physics parameterization (Rind et al. 2004).

The decision results for the second half of the twentieth century share main features of those for the whole twentieth century (Fig. 10c). ALL, ANTHRO, and G are major scenarios explaining observed SAT changes. One notable difference is that ANTHRO is decided over SAM rather than ALL. However, since Bayes factors of the two scenarios are similar over the region (Fig. 8), this is a minor change dependent on model simulations.

SME simulations might not be sufficient to reproduce observed regional SAT changes due to possible model errors in simulating responses to external forcing. This is explored below by comparing results using SMEs with MMEs. Figure 11 represents the decision results for six continental-scale regions from MULTI. Decision results for 1900–99 (Fig. 11a) show that NAM, SAM, AFR, and EUR have dominant ALL signals while ASI and AUS prefer the ANTHRO scenario. In 1900–49, the N signal over SAM and the ALL signal over AFR are manifest (Fig. 11b). Results for 1950–99

(Fig. 11c) resemble those for 1900–99 except that AFR is dominated by the G signal, which indicates an underestimated warming trend over AFR in the late century simulated by MMEs. In comparison with the results from SINGLE (Fig. 10), some differences can be found. The ALL signal extends over AFR for 1900–99 and 1900–49, SAM in 1950–99, and EUR for 1900–99 and 1950–99. Also ANTHRO is dominant over AUS for 1900–99 and 1950–99. These changes originate from different model responses to external forcing as well as different ranges of internal variability. For instance, the MME_ANTH response matches best with observations over AUS (Figs. 11a,c) while for SME the ECHO-G_G matches best over the region (Figs. 10a,c). Another example of a different model response comes over AFR for 1950–99 where MME_ALL and MME_ANTH have worse skill than ECHO-G_G (Fig. 11c) while ECHO-G_ANTH shows the best consistency with observations (Fig. 10c). The effect of changes in internal variability is responsible for different decision results over EUR between SINGLE and MULTI. The range over EUR2 is much larger in SINGLE than in MULTI, which hinders one from detecting external forcing signals.

Generally regional-scale decisions from MULTI are insensitive to the priors and temporal truncations, which is similar to that in SINGLE. According to the patterns of the Bayes factors (Fig. 9), the strength of ALL and ANTHRO signals over most of the regions is very similar. Hence the prevailing ALL and ANTHRO signals in the Bayesian decisions indicates that the effect of intermodel uncertainties appears to be weak even for regional-scale climate change assessment.

c. Seasonal dependence

To examine seasonal dependence of Bayesian decision results, seasonal mean SAT series are analyzed over six continental regions. Four seasons are defined from calendar months as December–February (DJF), March–May (MAM), June–August (JJA), and September–November (SON). One hundred samples are constructed for each season during 1900–99, and the first and second halves are used for 50-yr subperiod analysis for 1900–49 and 1950–99. For DJF, 99 (49) samples are available for 1900–99 (1950–99). The same space–time approach is applied by combining Legendre series expansions for two or three subregions as above.

Figure 12 shows the Bayesian decision results for four seasons for three analysis periods from the SINGLE experiment. Annual results (ANN) are plotted together for comparison. They are identical to those in Fig. 10. Seasonal dependence or contribution of the attributed signals is very different from region to region. MAM is a season of the strongest signals over

NAM, while SON is strongest over SAM. JJA has the strongest signal over EUR where the ANN signal is not evident. On the other hand JJA is the only season of a weak signal for ASI. Over AFR, different signals emerge for different seasons such as the ALL signal for DJF while the ANTHRO signal emerges for MAM and JJA. This suggests that we might enhance signal detectability in managing global patterns by confining analysis to signal-favorable seasons.

Seasonal results for 1900–49 reveal some complex decision patterns. Over SAM a mixture of ALL, G, and ANTHRO signals is found while the N signal is dominant for ANN. A similar complexity holds for AFR. It is interesting to note that NAM and EUR have evident ALL signals in JJA while their ANN signals are very weak. No external signals show up for all seasons over ASI and AUS. The most consistent picture is that the low-latitude regions of SAM and AFR exhibit a clear signal in the two seasons with strong insolation in the Southern Hemisphere (SON and DJF). Results for 1950–99 are very similar to those for 1900–99. One marked difference is that, as in the ANN results discussed above, SAM has ANTHRO signals rather than ALL in the seasonal decisions.

Figure 13 shows seasonal decisions from the MULTI experiment. Seasonal patterns largely follow ANN decisions for each region. Compared to SINGLE, major signals in the observed SAT changes are ALL for the whole twentieth century as well as its first and second halves, which are partly shared by ANTHRO. It seems that multimodel ensembles have an effect of producing more consistent signals across regions through improved model response as discussed above. Here again it should be noted that the signal amplitudes of ALL, ANTHRO, and in some cases G are very similar to each other according to patterns of Bayes factors (not shown). Therefore we conclude that the main results of climate change assessments are largely insensitive to the intermodel uncertainties as well as the prior probabilities and temporal scales retained, consistent with global results by MH06.

6. Conclusions

A Bayesian approach is applied extensively to the observed regional and seasonal SAT changes using MMEs of the IPCC AR4 simulations and SMEs with the ECHO-G model. A Bayesian decision method is used as a tool for classifying spatially and temporally varying observed SAT patterns over six continental-scale regions into six scenarios (CTL, N, G, S, ANTHRO, and ALL). Observed and simulated spatial mean SATs are decomposed into temporal components of overall

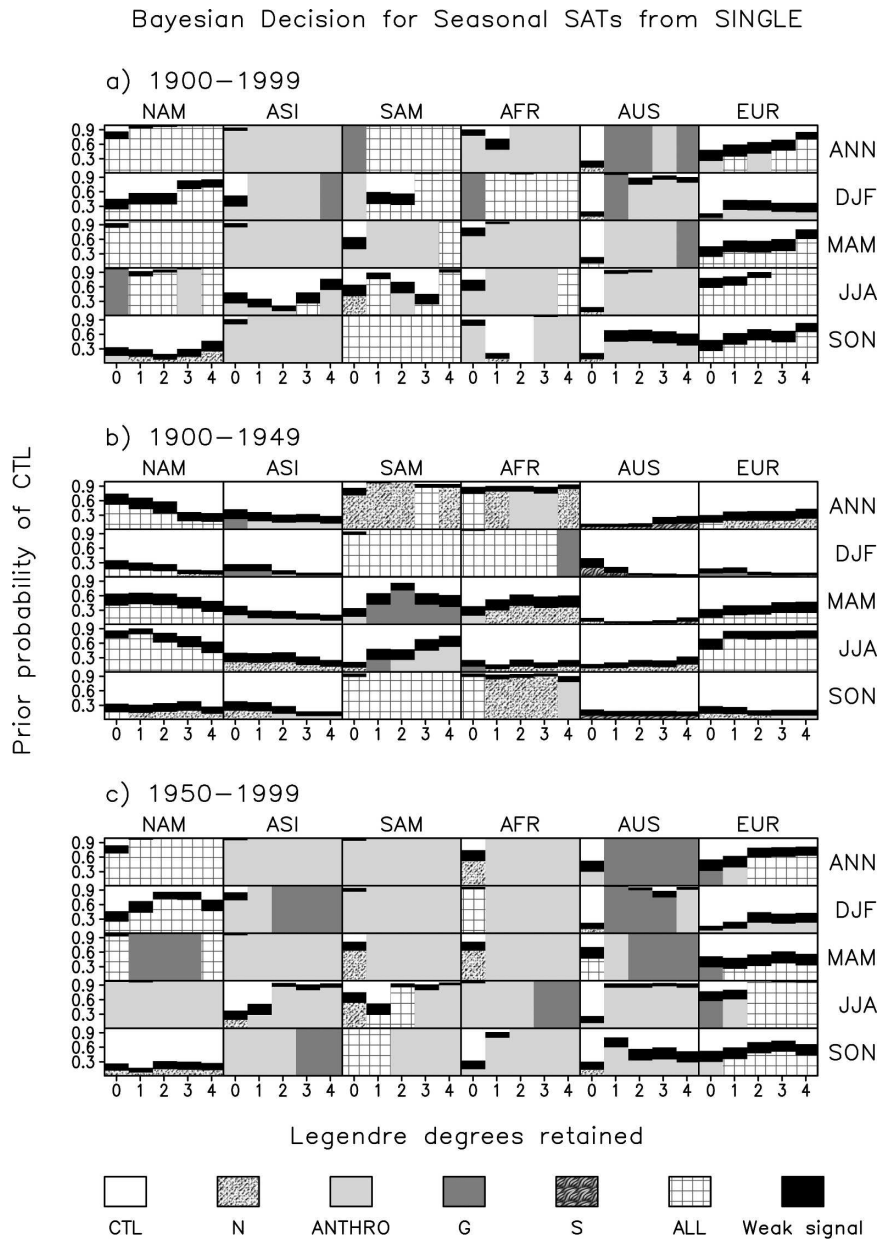


FIG. 12. Bayesian decisions for seasonal mean SAT anomalies during 1900–99, 1900–49, and 1950–99 over six continental regions from SINGLE. Each two-dimensional plot represents decision scenarios as a function of Legendre truncation (zeroth to fourth degree) and priors of CTL (0.01 to 0.99) as explained in Fig. 10.

mean, linear trend, and decadal variabilities through Legendre series expansions. The coefficients are used as detection variables. Parameters (means and covariance matrices for likelihood calculation) for defining each scenario are estimated from SMEs or MMEs, by which sensitivity of Bayesian decision results to inter-model uncertainties is examined.

Application results show that observed SAT changes over continental regions are classified into ALL or

ANTHRO scenarios for the twentieth century and its second half (1950–99), which corroborates previous studies (Stott 2003; Karoly et al. 2003, Karoly and Braganza 2005; Min et al. 2005a; International Ad Hoc Detection and Attribution Group 2005). For the first half of the twentieth century, the N or ALL signals are dominant over Africa and South America only, which might be related to response patterns to solar forcing centered over the Tropics, especially when considering

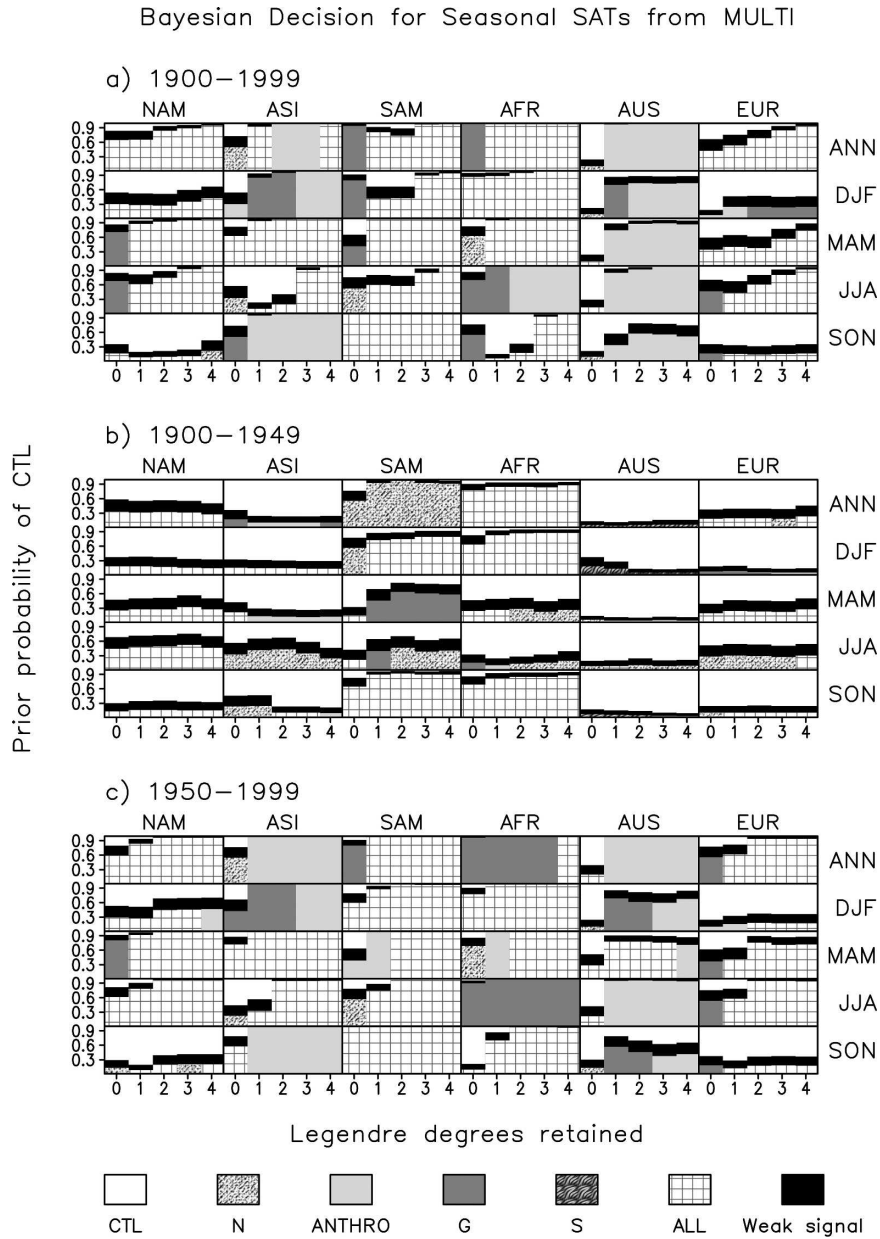


FIG. 13. Same as in Fig. 12, but for MULTI.

the seasonal results for these regions (Cubasch et al. 1997; Meehl et al. 2003). Seasonal patterns of Bayesian decisions in general are similar to annual results. However, we found notable seasonal dependences of detected signals, which vary across regions. The implication is that the signal detectability may be enhanced by combined assessments of seasons with stronger signals from different regions. Overall decision results from MMEs (MULTI) are not changed much from those from SMEs (SINGLE), indicating the robustness of Bayesian assessment to intermodel uncertainties, al-

though MULTI exhibits more prevailing signals of ALL across regions than SINGLE. In most cases, the Bayesian decisions for regional-scale SATs are largely insensitive to prior probability and temporal scales, as in the global results by MH06.

Acknowledgments. This work was supported by the German Research Foundation (DFG) with Grant He1916/8. We thank two anonymous reviewers for their constructive comments, and Won-Tae Kwon and Hyo-Shin Lee for providing their ECHO-G data.

ECHO-G model simulations have been performed at NEC supercomputers in DKRZ, Germany, and in KMA and KISTI, South Korea. We also acknowledge the international modeling groups for providing their data for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the model data, the JSC/CLIVAR Working Group on Coupled Modelling (WGCM) and their Coupled Model Intercomparison Project (CMIP) and Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, U.S. Department of Energy.

REFERENCES

- Allen, M. R., and S. F. B. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Climate Dyn.*, **15**, 419–434.
- Berger, J. O., 1985: *Statistical Decision Theory and Bayesian Analysis*. 2d ed. Springer-Verlag, 617 pp.
- Berliner, L. M., R. A. Levine, and D. J. Shea, 2000: Bayesian climate change assessment. *J. Climate*, **13**, 3805–3820.
- Collins, M., S. F. B. Tett, and C. Cooper, 2001: The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.*, **17**, 61–81.
- Cubasch, U., G. C. Hegerl, R. Voss, J. Waszkewitz, and T. J. Crowley, 1997: Simulation of the influence of solar radiation variations on the global climate with an ocean-atmosphere general circulation model. *Climate Dyn.*, **13**, 757–767.
- Delworth, T. L., and T. R. Knutson, 2000: Simulation of early 20th century global warming. *Science*, **287**, 2246–2250.
- Duda, R. O., and P. E. Hart, 1973: *Pattern Classification and Scene Analysis*. John Wiley, 482 pp.
- Hasselmann, K., 1976: Stochastic climate models. Part I. Theory. *Tellus*, **28**, 463–485.
- 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.
- International Ad Hoc Detection and Attribution Group, 2005: Detecting and attributing external influences on the climate system: A review of recent advances. *J. Climate*, **18**, 1291–1314.
- Jones, P. D., and G. C. Hegerl, 1998: Comparisons of two methods of removing anthropogenically related variability from the near-surface observational temperature field. *J. Geophys. Res.*, **103**, 13 777–13 786.
- , and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Climate*, **16**, 206–223.
- Karoly, D. J., and K. Braganza, 2005: Attribution of recent temperature changes in the Australian region. *J. Climate*, **18**, 457–464.
- , —, P. A. Stott, J. M. Arblaster, G. A. Meehl, A. J. Broccoli, and K. W. Dixon, 2003: Detection of a human influence on North American climate. *Science*, **302**, 1200–1203.
- Kass, R. E., and A. E. Raftery, 1995: Bayes factors. *J. Amer. Stat. Assoc.*, **90**, 773–795.
- Knutson, T. R., and Coauthors, 2006: Assessment of twentieth-century regional surface temperature trends using the GFDL CM2 coupled models. *J. Climate*, **19**, 1624–1651.
- Lee, T. C. K., F. W. Zwiers, G. C. Hegerl, X. Zhang, and M. Tsao, 2005: A Bayesian approach to climate change detection and attribution assessment. *J. Climate*, **18**, 2429–2440.
- Legutke, S., and R. Voss, 1999: The Hamburg atmosphere-ocean coupled circulation model ECHO-G. DKRZ Tech. Rep. 18, German Climate Computer Centre, Hamburg, Germany, 62 pp.
- Leroy, S. S., 1998: Detecting climate signals: Some Bayesian aspects. *J. Climate*, **11**, 640–651.
- Meehl, G. A., W. M. Washington, T. M. L. Wigley, J. M. Arblaster, and A. Dai, 2003: Solar and greenhouse gas forcing and climate response in the 20th century. *J. Climate*, **16**, 426–444.
- Min, S.-K., and A. Hense, 2006a: A Bayesian approach to climate model evaluation and multi-model averaging with an application to global mean surface temperatures from IPCC AR4 coupled climate models. *Geophys. Res. Lett.*, **33**, L08708, doi:10.1029/2006GL025779.
- , and —, 2006b: A Bayesian assessment of climate change using multimodel ensembles. Part I: Global mean surface temperature. *J. Climate*, **19**, 3237–3256.
- , and —, 2007: Hierarchical evaluation of IPCC AR4 coupled climate models with systematic consideration of model uncertainties. *Climate Dyn.*, in press.
- , —, H. Paeth, and W.-T. Kwon, 2004: A Bayesian decision method for climate change signal analysis. *Meteor. Z.*, **13**, 421–436.
- , —, and W.-T. Kwon, 2005a: Regional-scale climate change detection using a Bayesian decision method. *Geophys. Res. Lett.*, **32**, L03706, doi:10.1029/2004GL021028.
- , S. Legutke, A. Hense, and W.-T. Kwon, 2005b: Internal variability in a 1000-year control simulation with the coupled climate model ECHO-G—I. Near-surface temperature, precipitation and sea level pressure. *Tellus*, **57A**, 605–621.
- , —, —, and —, 2005c: Internal variability in a 1000-year control simulation with the coupled climate model ECHO-G—II. El Niño Southern Oscillation and North Atlantic Oscillation. *Tellus*, **57A**, 622–640.
- Rind, D., D. Shindell, J. Perlwitz, J. Lerner, P. Lonergan, J. Lean, and C. McLinden, 2004: The relative importance of solar and anthropogenic forcing of climate change between the Maunder Minimum and the present. *J. Climate*, **17**, 906–929.
- Schnur, R., and K. Hasselmann, 2005: Optimal filtering for Bayesian detection and attribution of climate change. *Climate Dyn.*, **24**, 45–55.
- Stone, D. A., M. R. Allen, and P. A. Stott, 2007: A multimodel update on the detection and attribution of global surface warming. *J. Climate*, **20**, 517–530.
- Stott, P. A., 2003: Attribution of regional-scale temperature changes to anthropogenic and natural causes. *Geophys. Res. Lett.*, **30**, 1728, doi:10.1029/2003GL017324.
- Stouffer, R. J., G. Hegerl, and S. Tett, 2000: A comparison of surface air temperature variability in three 1000-yr coupled ocean-atmosphere model integrations. *J. Climate*, **13**, 513–537.
- van Loon, H., G. A. Meehl, and J. M. Arblaster, 2004: A decadal solar effect in the tropics in July–August. *J. Atmos. Sol.-Terr. Phys.*, **66**, 1767–1778.
- Zhang, X., F. W. Zwiers, and P. A. Stott, 2006: Multimodel multilateral climate change detection at regional scale. *J. Climate*, **19**, 4294–4307.
- Zwiers, F. W., and X. Zhang, 2003: Toward regional-scale climate change detection. *J. Climate*, **16**, 793–797.

Copyright of *Journal of Climate* is the property of *American Meteorological Society* and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.