Consistently Estimating Internal Climate Variability from Climate Model Simulations

DIRK OLONSCHECK AND DIRK NOTZ

Max Planck Institute for Meteorology, Hamburg, Germany

(Manuscript received 2 June 2016, in final form 29 July 2017)

ABSTRACT

This paper introduces and applies a new method to consistently estimate internal climate variability for all models within a multimodel ensemble. The method regresses each model's estimate of internal variability from the preindustrial control simulation on the variability derived from a model's ensemble simulations, thus providing practical evidence of the quasi-ergodic assumption. The method allows one to test in a multimodel consensus view how the internal variability of a variable changes for different forcing scenarios. Applying the method to the CMIP5 model ensemble shows that the internal variability of global-mean surface air temperature remains largely unchanged for historical simulations and might decrease for future simulations with a large CO_2 forcing. Regionally, the projected changes reveal likely increases in temperature variability in the tropics, subtropics, and polar regions, and extremely likely decreases in midlatitudes. Applying the method to sea ice volume and area shows that their respective internal variability likely or extremely likely decreases proportionally to their mean state, except for Arctic sea ice area, which shows no consistent change across models. For the evaluation of CMIP5 simulations of Arctic and Antarctic sea ice, the method confirms that internal variability can explain most of the models' deviation from observed trends but often not the models' deviation from the observed mean states. The new method benefits from a large number of models and long preindustrial control simulations, but it requires only a small number of ensemble simulations. The method allows for consistent consideration of internal variability in multimodel studies and thus fosters understanding of the role of internal variability in a changing climate.

1. Introduction

Internal variability of the climate system, caused by the system's chaotic nature, limits the predictability of climate (e.g., Deser et al. 2014) and represents a major source of uncertainty for climate projections (e.g., Hawkins and Sutton 2009, 2011; Deser et al. 2012; Swart et al. 2015). Knowledge of internal climate variability is a prerequisite for climate-change attribution (e.g., Swanson et al. 2009; Trenberth 2011; Marotzke and Forster 2015) and climate-model evaluation (e.g., Flato et al. 2013; Stroeve et al. 2014; Notz 2015). However, robustly quantifying internal variability in climate studies remains challenging. Here we examine how internal variability in global climate models estimated from preindustrial climate simulations relates to internal variability estimated from the ensemble spread of historical and future climate simulations. This allows us to develop a new method to estimate internal variability for individual model simulations, which we apply for assessing changes in internal variability over time and for evaluating climate model simulations.

The magnitude of the internal variability of climate model simulations is usually estimated by using one of two different approaches (Collins et al. 2013). The first approach, here called "control-simulation approach," is based on the analysis of preindustrial control simulations with constant external forcing (for diverse applications see, e.g., Schneider and Kinter 1994; Swanson et al. 2009; Huber and Knutti 2014; Palmer and McNeall 2014; Resplandy et al. 2015; Schindler et al. 2015; Roberts et al. 2015). Apart from model drift, any climate variability of the preindustrial control simulations is internal variability. Preindustrial control simulations, typically spanning many centuries, are commonly sufficiently long to also include multidecadal and longer-term internal variability. They are usually available for any climate model and are, for example, part of the entrance criteria for a model to participate in phase 6 of the Coupled Model Intercomparison Project (Meehl et al. 2014). However, the control-simulation approach

Corresponding author: Dirk Olonscheck, dirk.olonscheck@mpimet.mpg.de

DOI: 10.1175/JCLI-D-16-0428.1

is commonly considered unsuitable for representing the internal variability of simulations with a different or changing external forcing (e.g., Kay et al. 2015).

The second approach, here called ''ensemble-spread approach,'' addresses this possible limitation (e.g., Deser et al. 2012, 2014; Wettstein and Deser 2014). The ensemble-spread approach is based on ensemble simulations with slightly different initial conditions, with each realization subject to the same external forcing. The ensemble spread between these different realizations of a single model measures the internal variability for different forcing scenarios. Some modeling groups run large ensembles of a single model to disentangle the internally and externally forced contributions in a simulation, because a sufficiently high number of ensemble simulations is required to estimate the model's total ensemble spread. However, running multiple realizations with any given global climate model consumes substantial computational power. As a consequence, many modeling groups provide only a single realization or a small number of ensemble simulations. This inhibits a robust and consistent estimation of model-specific internal variability for different forcing scenarios for a given multimodel ensemble.

We here address this common problem of multimodel studies by examining the relationship between the estimate of preindustrial internal variability from the control-simulation approach and the estimate of historical or future internal variability from the ensemble-spread approach. We expect similarities between both estimates of internal variability following the quasi-ergodic assumption, which states that the variance of one sequence of events over time equals the ensemble variance at a given time (e.g., von Neumann 1932; Hingray and Said 2014). However, changes in external forcing might alter the internal variability of a climate variable over time (Lu et al. 2014; Sutton et al. 2015). By relating both approaches across a multimodel ensemble, we derive estimates of internal variability for different forcings of each given model.

We apply our method to the model ensemble of phase 5 of the Coupled Model Intercomparison Project (CMIP5) for near-surface air temperature (SAT) and sea ice volume and sea ice area (i) to investigate whether the magnitude of internal variability changes over time and (ii) to robustly evaluate CMIP5 sea ice simulations.

Possible changes of internal variability over time were investigated by Huntingford et al. (2013), Thompson et al. (2015), and Holmes et al. (2016). Huntingford et al. (2013) used model output from 17 CMIP5 models to investigate the time-evolving global temperature variability. They examined 11- and 31-yr detrended ensemble-mean historical simulations and RCP8.5 simulations, and found that so far, the globally averaged temperature variability has been stable but is projected to decrease in future. Holmes et al. (2016) analyzed future changes in winter and summer temperature variability in a 17-member ensemble from a global climate model forced by the SRES A1B emission scenario. They removed 40-yr linear trends in ensemble-mean temperature and found strong regional changes in temperature variability. In contrast to their approaches, ours does not require the removal of any trend and thus allows for a clear separation of internal variability from external forcing based on ensemble simulations of individual models. Thompson et al. (2015) estimated the uncertainty in projections of future climate trends arising from internal variability. Using an analytic model that requires a time-stationary standard deviation for different forcing scenarios, they related the statistics of the preindustrial control simulation to the spread of trends in the 40-member ensemble of the global climate model CCSM3. They concluded that for most regions, the preindustrial control simulation is sufficient to represent the internal variability derived from the CCSM3 large ensemble. Based on our multimodel approach, we derive estimates of internal variability for different forcing scenarios for all CMIP5 models; our estimates are largely independent of the ensemble size of a single model. Further, our approach does not require the standard deviation of a variable to be stationary in time and thus can be used for variables whose standard deviation changes for different forcing scenarios.

For the evaluation of CMIP5 simulations, we limit our analysis to sea ice. Recent studies evaluated CMIP5 sea ice simulations of Arctic sea ice extent (e.g., Massonnet et al. 2012; Stroeve et al. 2012; Flato et al. 2013; Notz 2014), Antarctic sea ice extent (e.g., Mahlstein et al. 2013; Zunz et al. 2013), and Arctic sea ice thickness and volume (Stroeve et al. 2014; Shu et al. 2015). All of these studies stressed the large influence of internal variability. The influence of internal variability on Arctic sea ice trends was specifically estimated by Swart et al. (2015), who investigated both the CMIP5 models and the 30-member ensemble of the global climate model CESM1. They concluded that internal variability must be carefully accounted for when evaluating sea ice simulations, which was also spelled out in a dedicated study by Notz (2015). Nevertheless, model-specific estimates of internal variability for the satellite period (1979–today) across the CMIP5 model ensemble do not exist yet. Our method now provides such estimates and consistently considers internal variability in an evaluation of the CMIP5 sea ice simulations.

Section 2 introduces our method to robustly estimate internal variability for different forcing scenarios across models. Section 3 presents the data used. Sections 4 and 5

demonstrate usage of this method by applying it to specific climate observables, namely, annual near-surface air temperature, sea ice volume, and sea ice area. Section 6 summarizes our findings.

2. Method and applications

a. Method

We regress the standard deviation from the control-simulation approach on that of the ensemble-spread approach. We specifically explain here how this regression allows us to derive a model-specific estimate of internal variability across models for simulations with different forcing scenarios. Our method consists of four steps:

1) We calculate the standard deviation of the preindustrial control simulation for each model.
2) We calculate the ensemble standard deviation for each model that provides ensemble simulations.
3) We regress the estimates from step 1 onto those from step 2 and fit a regression line through these estimates. We use this basic version of our method to assess changes of internal variability over time.
4) If the regression can robustly be determined, we use it in a final step to translate the estimates of the preindustrial standard deviations to estimates corresponding to different forcing scenarios for models with a single simulation. This extended version of our method gives us consistent estimates of internal variability for all models.

We now explain these four steps in detail.

1) CONTROL-SIMULATION APPROACH

We calculate the standard deviation σ_{piC} for a variable x from the preindustrial control simulation of length T of each model across each output interval of interest t (e.g., year):

$$\sigma_{\mathrm{piC}}(T) = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(\overline{x} - x_t)^2}. \qquad (1)$$

However, for two reasons we cannot obtain a reliable estimate of internal variability from directly applying this control-simulation approach to all simulations. First, some preindustrial control simulations still drift substantially, likely because the models are at the beginning of the control simulation not yet in equilibrium with the preindustrial forcing (e.g., Knutson et al. 2013; Frankcombe et al. 2015). We hence remove the least squares linear trend from each model's preindustrial control simulation to minimize model drift. Second, because of multidecadal and longer-term internal variability, the standard deviation also depends on the T of the control simulation, which strongly varies among models (cf. Table 1). When a model provides too short a control simulation, which we assess from its power spectrum as described in appendix A, we assume that the internal variability as given by the control simulation is not the total internal variability of this particular model. We here disregard such models.

2) ENSEMBLE-SPREAD APPROACH

We calculate the ensemble standard deviation σ_{ens} for a variable x as the square root of the ensemble variance across the different ensemble simulations n of a model with N ensemble simulations for each output interval of t (e.g., each year) averaged over the simulation length T:

$$\sigma_{\mathrm{ens}}(N,T) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{N-1}\sum_{n=1}^{N}(\overline{x_t} - x_{n,t})^2\right]}. \qquad (2)$$

The ensemble-spread approach also includes long-term internal variability, because calculating the ensemble standard deviation does not require one to remove any model trend. The approach avoids possible underestimates of a model's total internal variability due to small numbers of ensemble simulations because it benefits from a largely increased sample size caused by considering the estimates of ensemble variance at every output interval. For example, $N = 3$ ensemble simulations with $T = 150$ yr correspond to a simulation with $T = 450$ yr. Hence, the ensemble-spread approach then relies on a sample size of 450, which is similar to the standard length of a control simulation. To nevertheless test the sensitivity of the ensemble standard deviations to very low numbers of available ensemble simulations, we assess the range of ensemble standard deviations from any pairwise combination of a model's ensemble simulations (i.e., $N = 2$; see section 4a). The upper bound of this range is given by the two ensemble members with the most contrarily temporal evolution, while the lower bound results from the two ensemble members that are most similar. We use this range to estimate the uncertainty of the ensemble standard deviation.

The ensemble-spread approach purely samples internal climate variability and circumvents uncertainties due to inconsistent natural external forcings over time in CMIP5. While preindustrial and future RCP8.5 simulations generally do not include natural external forcing from volcanic eruptions, historical simulations do contain volcanic eruptions (e.g., Santer et al. 2014). Since volcanic eruptions considerably contribute to the variability of climate variables, such as global-mean surface temperature (e.g., Bradley and Jones 1992; Briffa et al. 1998) and sea ice area (Rosenblum and Eisenman 2016), this would bias estimates of preindustrial and future natural

TABLE 1. CMIP5 simulations and single-model large ensembles used in this study. (Expansions of acronyms are available online at http://www.ametsoc.org/PubsAcronymList.)

Model name	T of preindustrial control simulation (yr)	No. of SAT simulations		No. of sea ice simulations		
		Historical	RCP8.5	Historical	RCP8.5	Extended to 2014
ACCESS1.0	500	2	1	3	1	1
ACCESS1.3	500	3	1	3	1	3
BCC_CSM1.1	500	3	1	3	1	1
BCC_CSM1.1(m)	500	3	1	3	1	1
BNU-ESM	559	1	1	1	1	1
CanESM2	1096	5	5	5	5	5
CCSM4	501	6	6	6	6	6
CESM1(BGC)	500	1	1	1	1	1
CESM1(CAM5)	319	3	3	3	3	3
CESM1(FASTCHEM)	222	3	—	3	—	—
CESM1(WACCM)	200	1	3	1	3	—
CMCC-CESM	277	1	1	1	1	1
CMCC-CM	330	1	1	1	1	1
CMCC-CMS	500	1	1	1	1	1
CNRM-CM5	850	10	5	10	5	5
CNRM-CM5.2	150	1	—	1	—	—
CSIRO Mk3.6.0	500	10	10	10	10	10
EC-EARTH	452	8	8	10	10	10
FGOALS-g2.0	700	4	1	4	1	1
FIO-ESM	800	3	3	3	3	3
GFDL CM3	500	5[a]	1	5[a]	1	5[a]
GFDL-ESM2G	500	3[a]	1	1[a]	1	1[a]
GFDL-ESM2M	500	1[a]	1	1[a]	1	1[a]
GISS-E2-H	780	6	1	6	1	5
GISS-E2-H-CC	251	1	1	1	1	1
GISS-E2-R	850	6	1	6	1	6
GISS-E2-R-CC	251	1	1	1	1	1
HadCM3	1200	10[a]	—	10[a]	—	10[a]
HadGEM2-CC	240	1[a]	3	1[a]	3	1[a]
HadGEM2-ES	576	5[a]	4	4[a]	4	4[a]
INM-CM4.0	500	1	1	1	1	1
IPSL-CM5A-LR	1000	6	4	6	4	4
IPSL-CM5A-MR	300	3	1	3	1	3
IPSL-CM5B-LR	300	1	1	1	1	1
MIROC5	670	5	3	5	3	5
MIROC-ESM	630	3	1	3	1	3
MIROC-ESM-CHEM	255	1	1	1	1	1
MPI-ESM-LR	1000	3	3	3	3	3
MPI-ESM-MR	1000	3	1	3	1	3
MPI-ESM-P	1156	2	—	2	—	—
MRI-CGCM3	500	3	1	3	1	1
MRI-ESM1	250	1	1	1	1	1
NorESM1-M	501	3	1	3	1	3
NorESM1-ME	252	1	1	1	1	1
CESM1-CAM5-BGC-LE	1800	35[b]	35	35[b]	35	—
MPI-ESM1.1-LE	2000	100	—	100	—	—

[a] Historical simulations that start in year 1860 only.
[b] Historical simulations that start in year 1920 only.

variability low compared to estimates of historical natural variability. However, for our estimates of historical internal variability derived from the ensemble-spread approach, each ensemble member has experienced the same volcanic forcing time series. Apart from the minor effects of synchronized ensemble spread in years of volcanic eruptions, our ensemble standard deviation is thus independent from volcanic forcing.

3) BASIC VERSION OF THE METHOD

Depending on the specific application, our method can be used in two versions, here called "basic version"

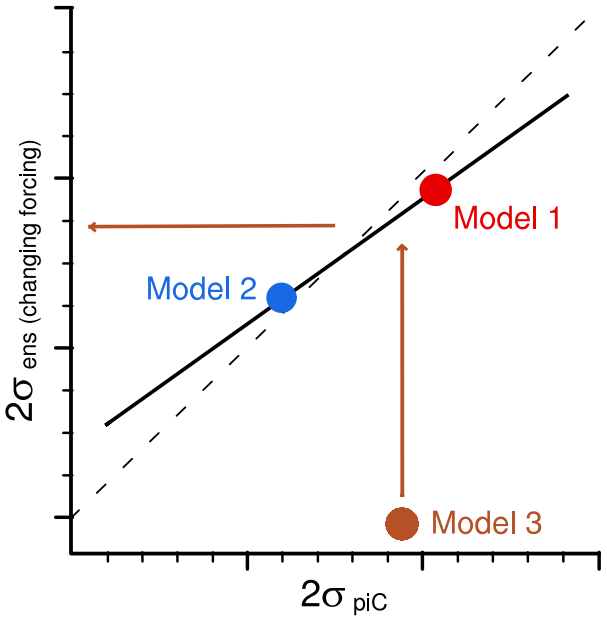

FIG. 1. Schematic view of the method for estimating internal variability for different forcing scenarios. The basic version of the method regresses the estimate of internal variability derived from the preindustrial control simulation of a model (x axis) on the ensemble standard deviation of models with ensemble simulations such as models 1 and 2 (y axis). The unity line as a reference is indicated by the dashed black line. For the extended version, a constructed ensemble standard deviation can be derived for models with a single simulation (model 3) using the regression line through models 1 and 2. The extended version requires a consistent response of the models with ensemble simulations.

and "extended version." To present both versions of the method, we use a fictitious example as sketched in Fig. 1. For the basic version, we simply relate the estimate of internal variability from the control-simulation approach ($2\sigma_{\mathrm{piC}}$, x axis) to that of the ensemble-spread approach ($2\sigma_{\mathrm{ens}}$, y axis). This can be done for a single model that provides ensemble simulations, such as model 1, but also across all models of a given multimodel ensemble that provide ensemble simulations by linear least squares regression, such as model 1 and model 2. This basic version allows one to directly compare the preindustrial internal variability to the internal variability for different forcing scenarios for models with ensemble simulations. We exploit the multimodel relationship of the basic version to test for a change of internal variability over time in application 1.

4) EXTENDED VERSION OF THE METHOD

The extended version builds on the basic version. By using the regression line through the estimates of internal variability of model 1 and model 2, we derive an ensemble standard deviation for models with a single simulation from the model's preindustrial estimate of internal variability (model 3, situated at the zero line of the ensemble standard deviation). This procedure is based on the key assumption that a relationship found for many models is valid for other models as well. We justify the assumption by the underlying theory of quasi ergodicity that we show to hold for the models with ensemble simulations. To be applicable to a variable, the procedure requires a similar response from models with ensemble simulations for this variable. If this prerequisite is fulfilled, then the procedure allows us to circumvent limitations associated with models that have a single realization only. We keep the original estimates for models with ensemble simulations and do not adjust their standard deviations to the regression line. We use these original estimates of internal variability for models with ensemble simulations and the derived estimates from the extended version of our method for models with a single simulation to consistently evaluate sea ice simulations from all models in application 2.

b. *Applications*

To obtain a consensus estimate of the projected direction and magnitude of a possible change in internal variability over time (application 1), we evaluate the regression line obtained in the simple version of our method at the location of the multimodel mean. To examine the likelihood of an identified change, we test whether the confidence interval of the regression line includes the unity line. Following the IPCC terminology, we define a change as likely when the 66% confidence interval of the regression line does not include the unity line and as extremely likely when the 95% confidence interval of the regression line does not include the unity line. In contrast, we define a change as possible when the 66% confidence interval of the regression line does include the unity line.

To evaluate the CMIP5 sea ice simulations (application 2), we consider both the simulated internal variability σ_{mod} that we derive from the extended version of our method and the observational or reanalysis uncertainty δ_{ref} (see appendix B). To combine both sources of uncertainty, we follow an approach of Santer et al. (2008) that was adopted by Stroeve et al. (2012). This approach uses a plausibility variable as a measure of model fidelity,

$$\phi = \frac{\overline{\mathrm{mod}} - \overline{\mathrm{ref}}}{\sqrt{\sigma_{\mathrm{mod}}^2 + \delta_{\mathrm{ref}}^2}}, \tag{3}$$

that weights the distance between any time-averaged CMIP5 model simulation ($\overline{\mathrm{mod}}$) and the time-averaged

reference data ($\overline{\text{ref}}$) by the internal variability of the simulations and the observational or reanalysis uncertainty (see appendix B for details). The plausibility variable ϕ thus quantifies how far the model deviates from the reference data in units of the associated quantity- and model-specific uncertainty.

3. Data

To demonstrate usage of our method, we apply it to the internal variability of annual near-surface air temperature and of Arctic and Antarctic sea ice volume and sea ice area as simulated by the models that took part in CMIP5 (Taylor et al. 2012). These simulations are available from the Earth System Grid data portal of the Earth System Grid Federation (http://esgf-node.llnl.gov/search/cmip5/) and from the data portal from the Centre for Environmental Data Analysis (ftp://ftp.ceda.ac.uk/badc/cmip5/). For both near-surface air temperature and sea ice volume and area, we analyze the preindustrial control simulation, the historical simulations (1850–2005), and future simulations (2006–2100) driven by the representative concentration pathway (RCP) 8.5 scenario (Moss et al. 2010) of each CMIP5 model.

For near-surface air temperature, we analyze gridded monthly mean data of 145 historical simulations from 44 different climate models and 87 RCP8.5 simulations from 40 different climate models (Table 1). To account for regional changes in temperature variability across models, we regrid all CMIP5 simulations by bilinear interpolation on a grid resolution of 1.8947° × 3.75°. We use the regridded data to calculate annual global-mean surface temperature by weighting the near-surface air temperature with the area of the model grid cells and then averaging annually and globally.

For sea ice, we analyze 145 historical simulations from 44 different climate models and 88 RCP8.5 simulations from 40 different climate models that provide gridded monthly mean data of sea ice concentration and sea ice thickness (Table 1). From these, sea ice area is calculated by multiplying the area of the model grid cells with their sea ice concentration, which is then added up over all grid cells for either the Northern or the Southern Hemisphere. The model output of sea ice thickness is the equivalent thickness averaged over the grid cell assuming that the grid cell is entirely ice covered. Sea ice volume is calculated as the product of the area of the model grid cells and their equivalent sea ice thickness, which again is added up over all grid cells for both hemispheres. For reasons explained by Notz (2014), such as differences in grid geometry and misleading results with respect to model quality as a result of synthetic biases in sea ice extent, we focus on the more direct and more physical metric sea ice area instead of sea ice extent.

To evaluate the applicability of small ensemble sizes of individual CMIP5 models, we additionally analyze single-model large ensemble simulations of updated versions of two models that are part of CMIP5: the 35-member ensemble of the NCAR Community Earth System Model (CESM1-CAM5-BGC) covering the period 1920–2100 with 2006–2100 forced by RCP8.5 (Kay et al. 2015) and the 100-member ensemble of the Max Planck Institute Earth System Model (MPI-ESM1.1) in low resolution covering the period 1850–2005.

4. Application 1: Assessing changes of internal climate variability over time

a. *Surface air temperature*

We first apply the basic version of our method to annual surface air temperature (Fig. 2). For the historical simulations, plotting the global-mean estimates of internal variability from the control-simulation approach ($2\sigma_{\text{piC}}$, x axis) against that from the ensemble-spread approach ($2\sigma_{\text{ens}}$, y axis) across the CMIP5 models results in a linear one-to-one relationship (black regression line in Fig. 2a). Model estimates of ensemble standard deviation that deviate from the one-to-one relationship are usually very uncertain as shown by the test on how representative these estimates are for a model's total ensemble standard deviation described in section 2a (see vertical bars in Figs. 2a,b). We have additional support for the one-to-one relationship from the two models that provide large ensemble simulations, CESM1-CAM5-BGC and MPI-ESM1.1 (black triangle and diamond, respectively). We thus detect no robust change in time-averaged internal variability of annual global-mean surface temperature between the preindustrial and the historical period across the CMIP5 models.

For the future simulations forced by RCP8.5, we detect a possible decrease in time-averaged internal variability of annual global-mean surface temperature compared to the preindustrial climate (Fig. 2b).

To examine the direction and relative magnitude of regional changes in temperature variability, we use our method to calculate the ratio $\sigma_{\text{ens}}/\sigma_{\text{piC}}$ averaged for all models for every grid box (colored patterns in Figs. 2c,d). For the historical simulations, we find weak regional changes in temperature variability (Fig. 2c). While tropical regions show a possibly (no stippling) increased temperature variability compared to the preindustrial climate (red), many midlatitude regions show a possibly decreased temperature variability (blue). The pattern of temperature-variability change for the historical period

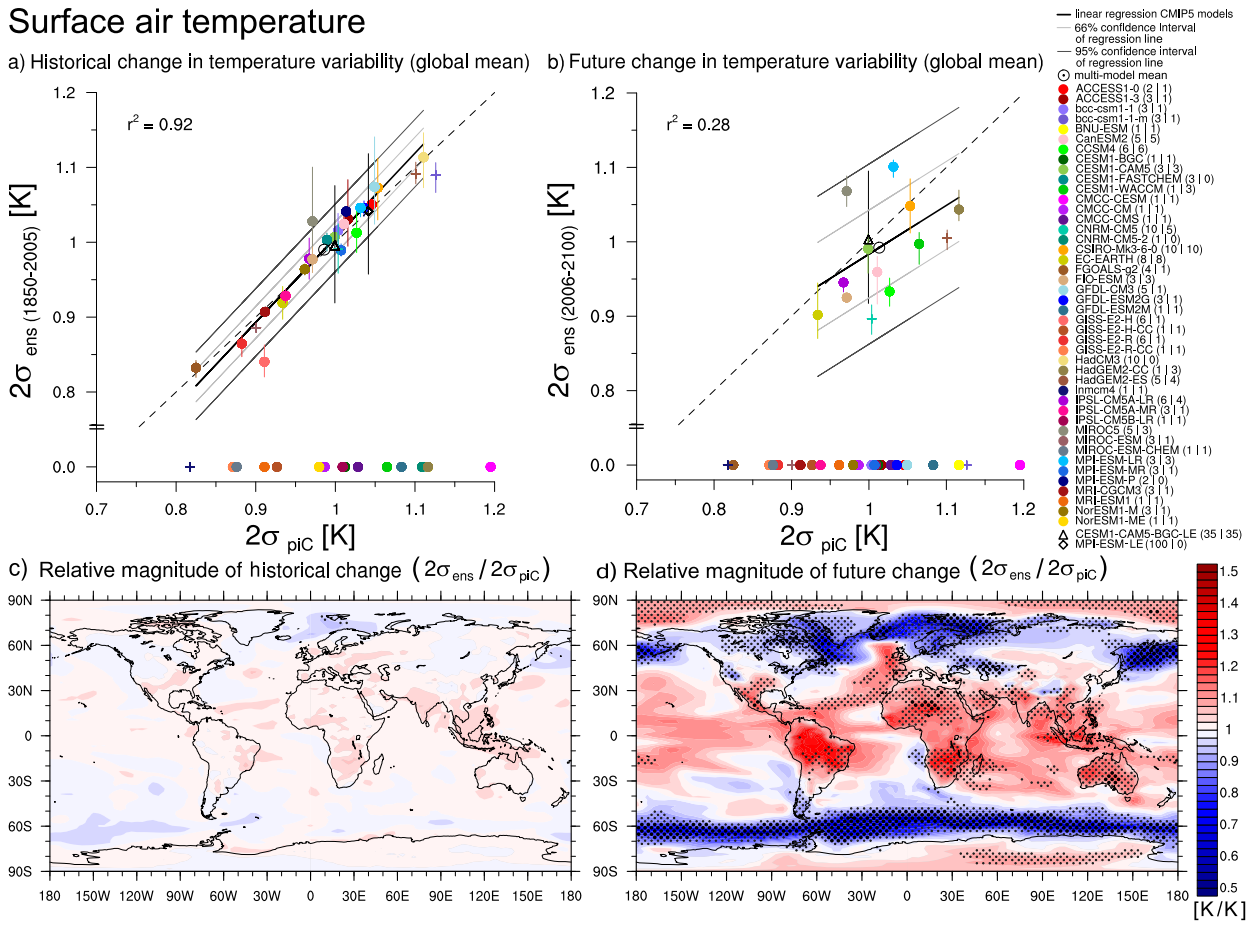

FIG. 2. Relationship between the standard deviation of each CMIP5 model preindustrial control simulation (x axis) and the ensemble standard deviation of the corresponding (a) historical simulations and (b) RCP8.5 scenario runs (y axis) for annual global-mean surface temperature. CMIP5 models that provide ensemble simulations (colored filled dots with nonzero ensemble standard deviation) are considered for calculating the regression line, the r^2, and the 66% and the 95% confidence intervals, and for calculating the multimodel mean (black circled dot). CMIP5 models that have a too short preindustrial control-simulation length to cover their total internal variability (+ symbols). Models with a single simulation are situated at the zero line of the ensemble standard deviation. Large ensembles are denoted by a triangle or a diamond. The number of ensemble simulations used to calculate the ensemble standard deviation is given in parentheses first for the historical simulations and second for the RCP8.5 scenario runs. Uncertainty as a result of the different numbers of ensemble simulations by assessing the ensemble standard deviation from any pairwise combination of a model's ensemble simulations (vertical bars). (c) Magnitude of relative change between the preindustrial temperature variability and the historical temperature variability, and (d) the magnitude of relative change between the preindustrial temperature variability and the temperature variability of a future climate forced by RCP8.5. A possible increase (red shades) and a possible decrease (blue shades) in temperature variability, and likely changes (light stippling) and extremely likely changes (strong stippling) are shown in (c) and (d).

already depicts the pattern of change projected for the future.

For the future simulations forced by RCP8.5, we detect strong regional changes in temperature variability (Fig. 2d). Whereas many tropical, subtropical, and polar regions show a likely (light stippling) or possibly increased temperature variability, many midlatitude regions show a likely decreased temperature variability. Many mid-to-high-latitude oceans show extremely likely changes (strong stippling).

Our findings of a globally averaged stable internal variability of surface air temperature for the time-averaged historical period and a possibly decreased internal variability for the time-averaged future climate forced by RCP8.5 agree with the result by Huntingford et al. (2013) that so far the variability of global-mean surface temperature has not changed but is projected to decrease in the future. They are also in line with the single-model result by Hawkins et al. (2016), who found a decreased global-mean surface-temperature variability of about 10% for

idealized initial condition ensembles of a global climate model forced by a 1% CO_2 increase per year.

Regionally, our findings confirm the result by Huntingford et al. (2013) of increased variability in regions of low variation and decreased variability in regions of high variation. Our results also agree with the pattern of future temperature-variability change found by Holmes et al. (2016, their Figs. 2c,d, 4c,d) and are further consistent with Screen (2014) and Schneider et al. (2015). They all argue that polar amplification decreases the temperature variability in Northern Hemisphere midlatitudes, because northerly winds are warming more rapidly than southerly winds, especially in winter. However, our multimodel result often identifies likely and possible changes because the model estimates differ widely in the direction and/or magnitude of regional changes.

In contrast to these previous studies that show regional changes in future temperature variability, Thompson et al. (2015) found a time-stationary internal variability. They show that for most regions, the preindustrial control simulation is sufficient to represent the future internal variability derived from the CCSM3 large ensemble. Kay et al. (2015) support this independence of regional internal variability from external forcing based on 34-yr trends in winter surface air temperature from CESM1-CAM5-BGC-LE.

These inconsistent results, which are mainly based on ensembles of individual models, demands a multimodel approach. Our multimodel approach discloses often very different model estimates of projected temperature-variability change that hinders a robust projection of changes in many regions of the globe. Our approach therefore cautions one to use only individual models for assessing future changes in temperature variability and now allows one to interpret single-model results in a multimodel context.

b. *Sea ice metrics*

We now apply the basic version of our method to sea ice. In an analogy to surface air temperature, we find remarkable similarity between the preindustrial and the historical internal variability for both Northern and Southern Hemisphere sea ice volume and area (Fig. 3). The single-model large ensembles of CESM1-CAM5-BGC and MPI-ESM1.1 (black triangle and diamond, respectively) confirm the one-to-one-relationship.

When testing for a change in sea ice internal variability with respect to a future RCP8.5-forced climate, we separately analyze winter sea ice (Fig. 4) and summer sea ice (Fig. 5). For winter Arctic and Antarctic sea ice volume, we detect an extremely likely decreased internal variability compared to the preindustrial climate (Figs. 4a,b), while for winter Antarctic sea ice area we find a likely decreased internal variability (Fig. 4d).

For winter Arctic sea ice area (Fig. 4c), the model responses differ substantially and the future variability proves largely independent of a model's preindustrial variability. We suggest two counteracting effects that cause the models to disagree even on the direction of change. On the one hand, the variability of Arctic sea ice area decreases because the mean sea ice area decreases (see section 4c). On the other hand, the variability of Arctic sea ice area increases when the sea ice is detached from continental boundaries (Eisenman et al. 2011), when the sea ice becomes thinner (e.g., Bitz and Roe 2004; Notz 2009), and when the high-latitude temperature variability increases (cf. Fig. 2d). The inconsistent model responses even on the direction of change might reflect the different manifestation and timing of these counteracting processes in each model.

For summer sea ice, we consider only sea ice volume larger than $1 \times 10^3\,\mathrm{km}^3$ and sea ice area larger than $1 \times 10^6\,\mathrm{km}^2$ to prevent artifacts arising from touching the lower bound of zero sea ice. As for winter, we find an extremely likely decreased internal variability of summer Arctic and Antarctic sea ice volume and a likely decreased internal variability of summer Antarctic sea ice area compared to preindustrial conditions (Figs. 5a,b,d). Our result for summer Antarctic sea ice area is consistent with Goosse et al. (2009, their Fig. 1b), who show a decreasing sea ice variability with a decreasing mean state for March Antarctic sea ice extent.

In contrast to these variables, the future internal variability of summer Arctic sea ice area possibly increases and becomes largely independent of the different preindustrial manifestations of internal variability (Fig. 5c). The possible increase agrees with the result for CCSM3 (Holland et al. 2008) and is consistent with the study by Goosse et al. (2009) on September Arctic sea ice extent. Based on 14 global climate models, Goosse et al. (2009, their Fig. 1a) showed that the variability in September Arctic sea ice extent increases until the mean sea ice area has decreased to around $3 \times 10^6\,\mathrm{km}^2$ and then decreases for a lower mean September sea ice extent. Averaged over the period 2005–2100, we thus find increased variability in summer Arctic sea ice area across all models (Fig. 5c). As for winter, the variability in future summer Arctic sea ice area is possibly increased because the sea ice gets detached from continental boundaries, becomes thinner, and is vulnerable to increases in high-latitude temperature variability.

In summary, the magnitude of modeled annual mean internal variability of sea ice volume and area remains largely unchanged for the historical period. In contrast, the magnitude of internal variability in winter and summer sea ice decreases in the RCP8.5 scenario, except for the

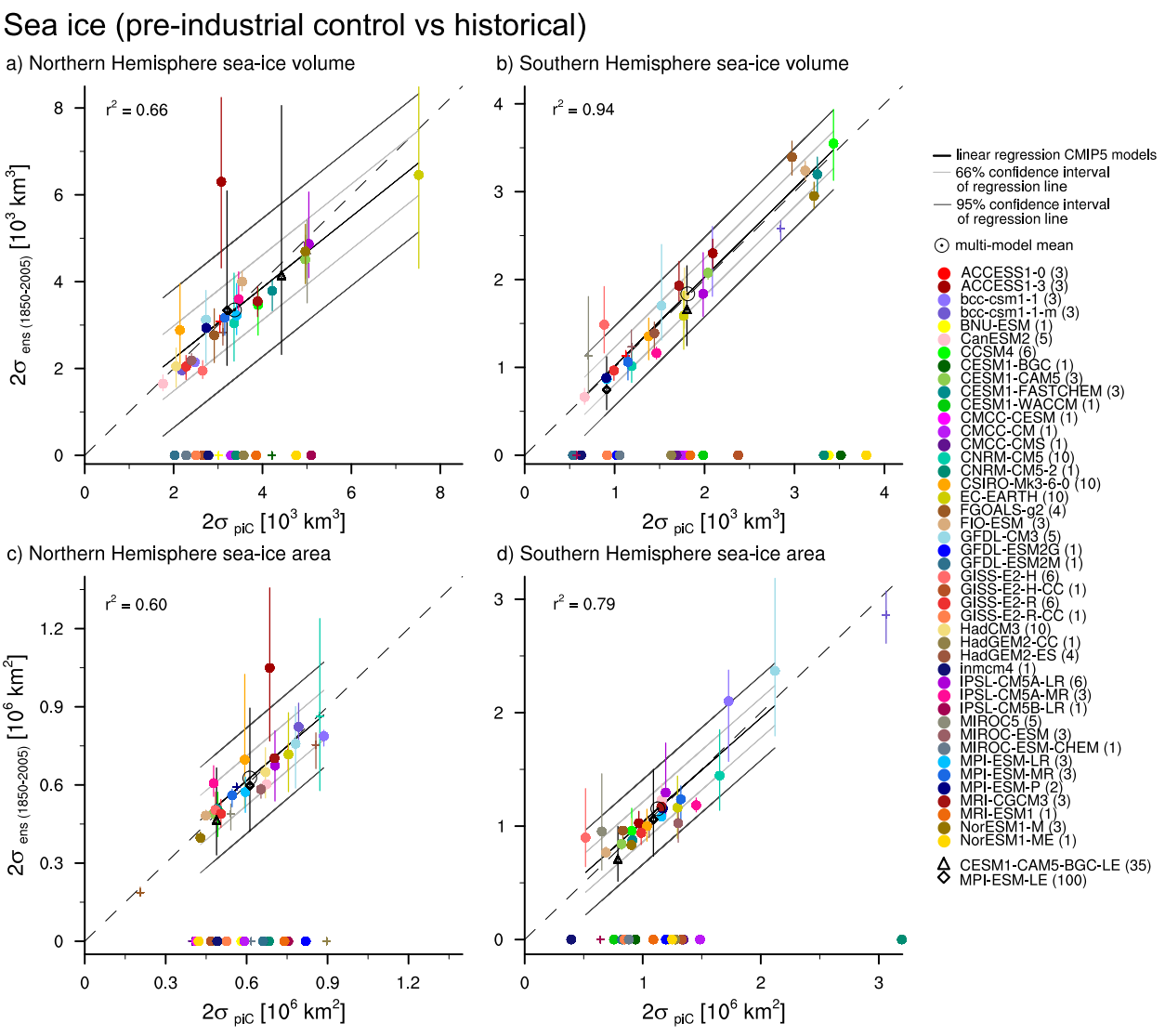

FIG. 3. Relationship between the standard deviation of each CMIP5 model preindustrial control simulation (x axis) and the ensemble standard deviation of the corresponding historical simulations (y axis) for annual (a) Northern and (b) Southern Hemisphere sea ice volume, and (c) Northern and (d) Southern Hemisphere sea ice area. CMIP5 models that provide ensemble simulations (colored filled dots with nonzero ensemble standard deviation) are considered for calculating the regression line, the r^2, and the 66% and the 95% confidence intervals, and for calculating the multimodel mean (black circled dot). CMIP5 models that have a too short preindustrial control-simulation length to cover their total internal variability (+ signs). Large ensembles are indicated by a triangle or a diamond. Uncertainty as a result of the different numbers of ensemble simulations by assessing the ensemble standard deviation from any pairwise combination of a model's ensemble simulations (vertical bars).

variability of Arctic sea ice area, which shows inconsistent model responses on the way to ice-free conditions. The inconsistent model responses for Arctic sea ice area highlight the benefit of our multimodel approach compared to single-model studies, as it allows one to interpret single-model results in a multimodel context.

c. *Linking sea ice variability to the mean sea ice state*

Previous studies showed a linear relationship between mean temperature and temperature variability in the range of high-latitude annual mean temperatures (Esau et al. 2012, their Fig. 3b) and a linear relationship between mean Arctic temperatures and mean Arctic sea ice area (e.g., Gregory et al. 2002; Mahlstein and Knutti 2012) for global climate models. To test whether sea ice variability is also linked to the mean sea ice state in CMIP5 models, we apply the extended version of our method. The time-averaged evolution of internal variability for the sea ice metrics in winter (Figs. 6a–d) shows that models with a high preindustrial mean

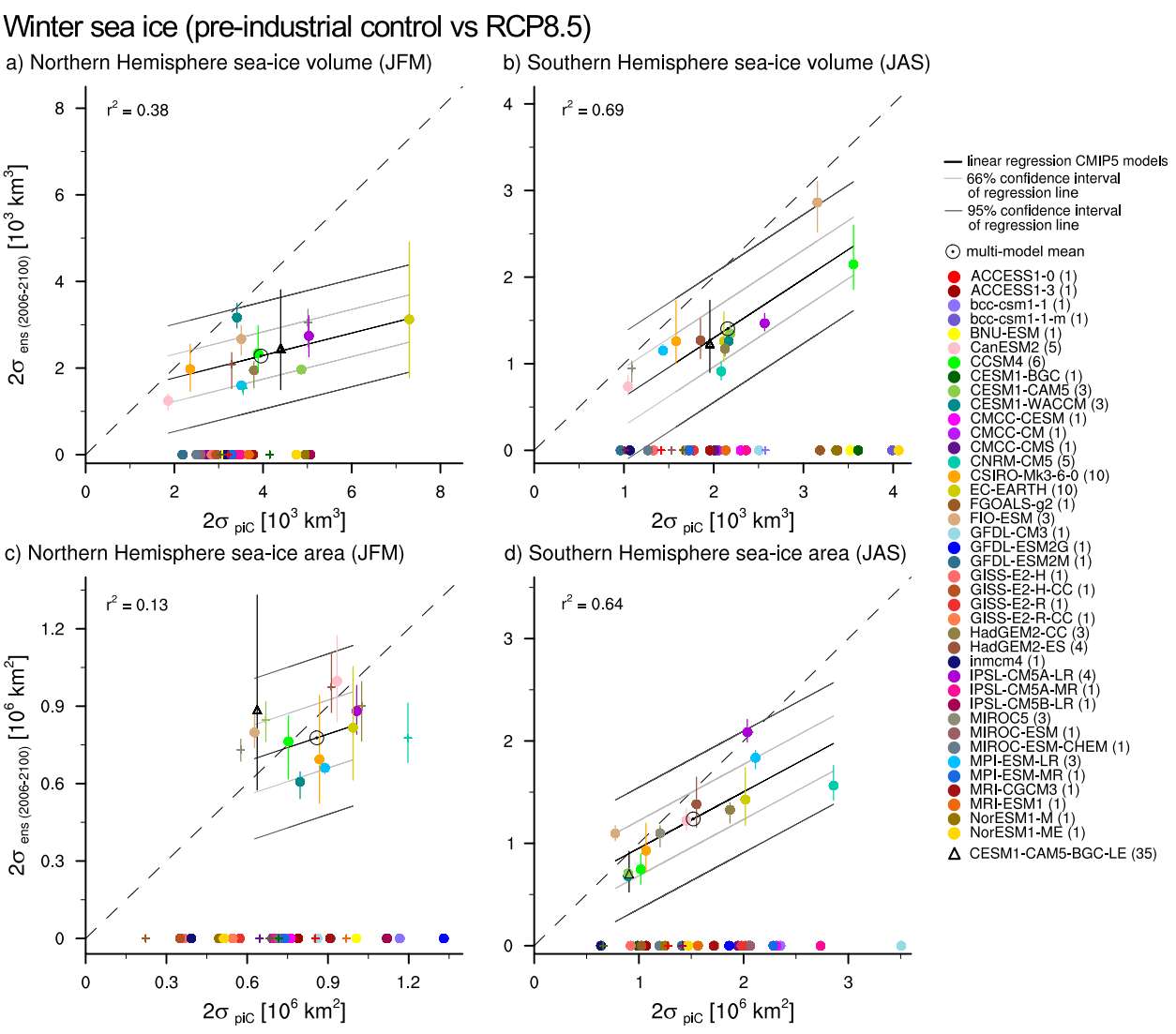

FIG. 4. As in Fig. 3, but for the ensemble standard deviation of the corresponding RCP8.5 scenario runs (y axis) for winter conditions.

state simulate a higher preindustrial estimate of internal variability of the corresponding sea ice variable than models with a low preindustrial mean state (filled dots). The cross-model relationship for the preindustrial state also holds for a single model over time, which one can infer from following individual models over time in Fig. 6. If for the reference data period 1979–2014 (triangles) the modeled time-averaged mean sea ice state is decreased compared to the preindustrial mean state (filled dots), then the internal variability is usually reduced similarly. The decrease in internal variability continues for the RCP8.5-forced decrease in the mean sea ice state (circles) for all metrics except for Arctic sea ice area, whose future variability is largely independent of the future mean Arctic sea ice area.

Except for future Arctic sea ice area, our analysis implies that the stronger the mean sea ice state is altered, the more the internal sea ice variability is reduced. From this relationship between the mean state and the variability of sea ice in CMIP5, we learn that when assuming that the models are realistic, knowledge about the mean state of the observable can be an emergent constraint for the system's internal variability, as also stated by Bathiany et al. (2016) based on two box models and a comprehensive Earth system model. However, because of the spread of modeled estimates, the relationship between the mean state and the variability permits only a rough estimate of the system's total variability from the mean state of a short observational time series. For comparison to observational or reanalysis data shown as dashed lines in Figs. 6a,c,d, we show the model simulations for the reference data period 1979–2014 instead of those from the historical period.

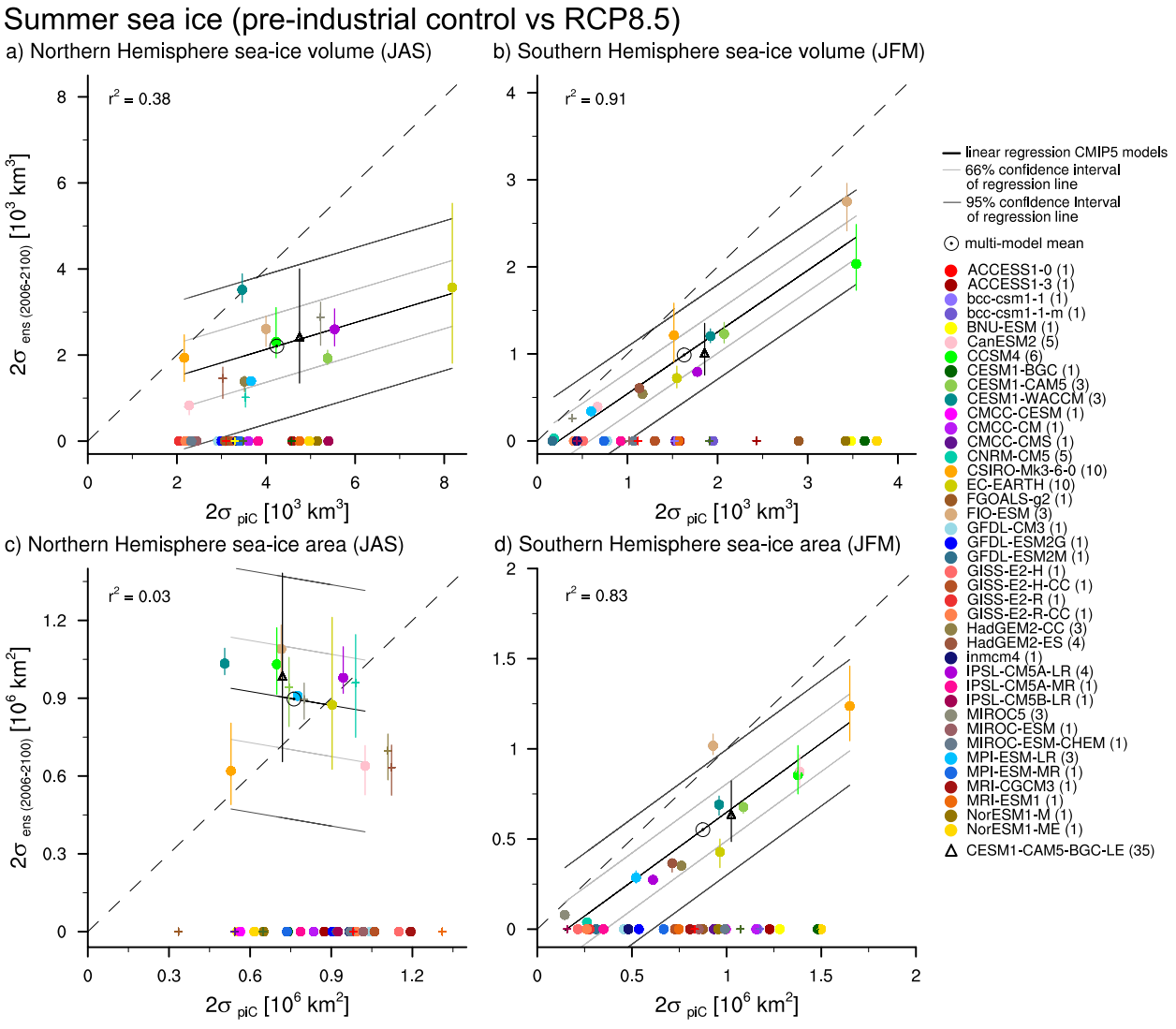

FIG. 5. As in Fig. 4, but for summer conditions. Note that only a sea ice volume $> 1 \times 10^3$ km^3 and a sea ice area $> 1 \times 10^6$ km^2 are considered when calculating the ensemble standard deviation.

5. Application 2: Plausibility of sea ice simulations

To present a second application of our method, we use its extended version for a robust and consistent assessment of the plausibility of sea ice simulations from all CMIP5 models. We specifically stress here that only our robust estimation and consideration of modeled internal variability for evaluating sea ice simulations is needed, since we do not know the system's true internal variability that otherwise could be used. We lack this knowledge because the observational record of sea ice evolution is short and no robust estimate of the internal variability of the preindustrial sea ice state exists. When applying the emergent constraint between internal variability and the mean state as discussed in section 4c, we face too broad a range of modeled estimates to derive a robust relationship that could be used to properly estimate the system's true internal variability based on the observed mean state.

The plausibility of the CMIP5 sea ice simulations is tested for the metrics sea ice volume and sea ice area for the Northern Hemisphere and sea ice area for the Southern Hemisphere both with respect to 30-yr trends and the mean state. No evaluation of CMIP5 simulations of Southern Hemisphere sea ice volume is provided because we lack a consistent long-term reanalysis dataset as a reference. Reference observational or reanalysis data for the considered measures are available from 1979 until today. To maximize overlap of this period with the CMIP5 simulations, we prolong the historical

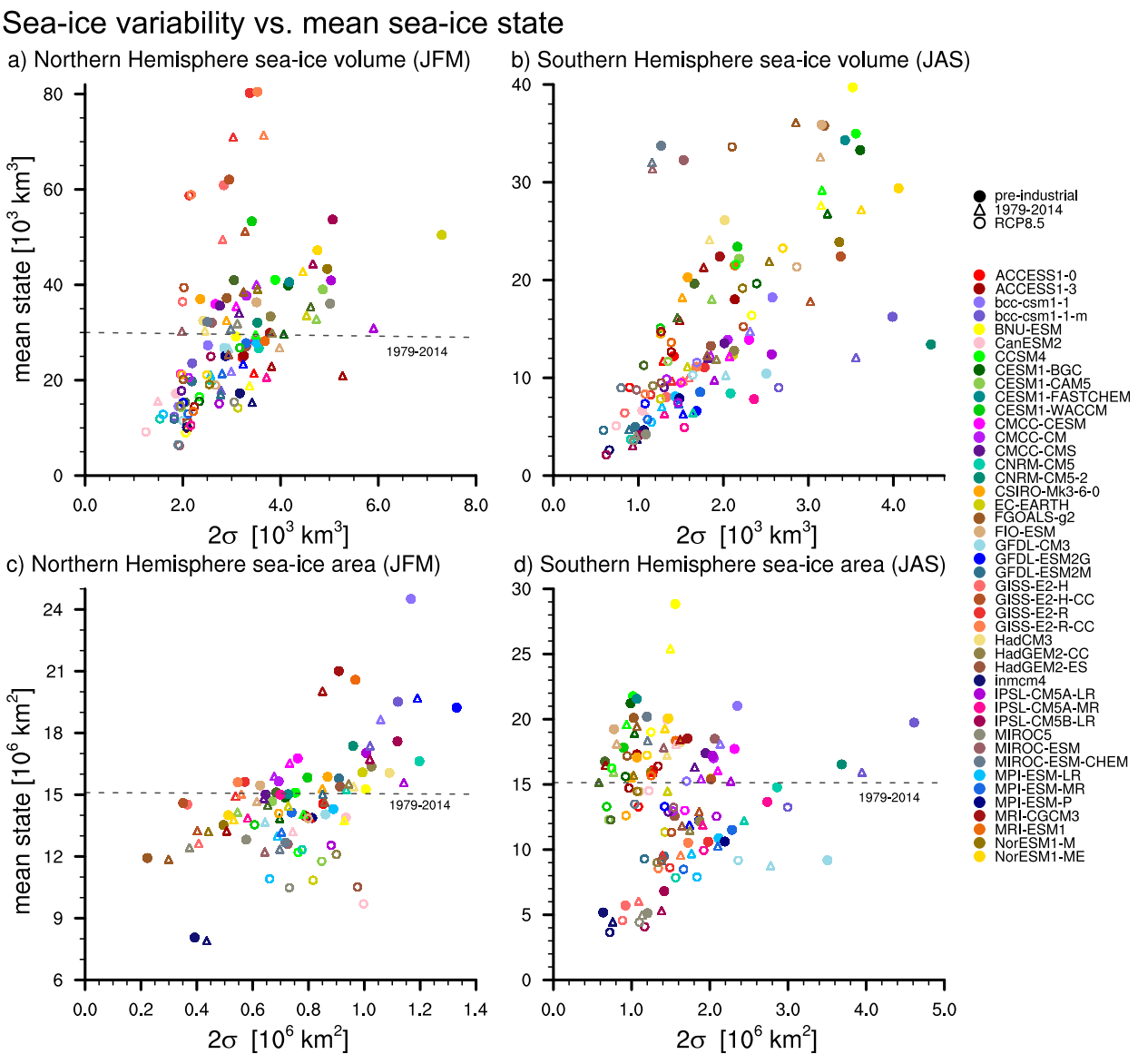

FIG. 6. Relationship between the standard deviation and the mean state of the preindustrial (filled dots), reference data period (triangles), and future RCP8.5-forced climate (circles) of winter (a) Northern and (b) Southern Hemisphere sea ice volume, and (c) Northern and (d) Southern Hemisphere sea ice area for the CMIP5 models. Symbols are shown in the model-specific color. The mean state of reference data from 1979 to 2014 is shown (dashed lines). Note that for future Arctic sea ice area, only the estimates from models with ensemble simulations are shown, because the regression line based on these models is highly uncertain and does not allow one to derive robust estimates for models with a single simulation (see Fig. 4c).

(1850–2005) sea ice simulations until 2014 using simulations based on future RCP emission scenarios (Moss et al. 2010). Depending on availability and to maximize the number of models included in the analysis, we use RCP4.5 (preferred) or RCP8.5 for the extension of the historical sea ice simulations. The choice of using either RCP4.5 or RCP8.5 for the extension of the historical runs until 2014 does not influence the evaluation result because both RCPs differ only slightly during this period. The extension of the historical sea ice simulations reduces the total number of available simulations to 119 (see Table 1). For the evaluation of the mean sea ice state, each model simulation is averaged over the period 1979–2014. The same is done for the reference data. In case of 30-yr trends, both the model output and the reference data are averaged over the six 30-yr linear trends obtained from the available 36-yr-long time series from 1979 to 2014.

CMIP5 model plausibility subdivided for each individual simulation and each month is presented as a

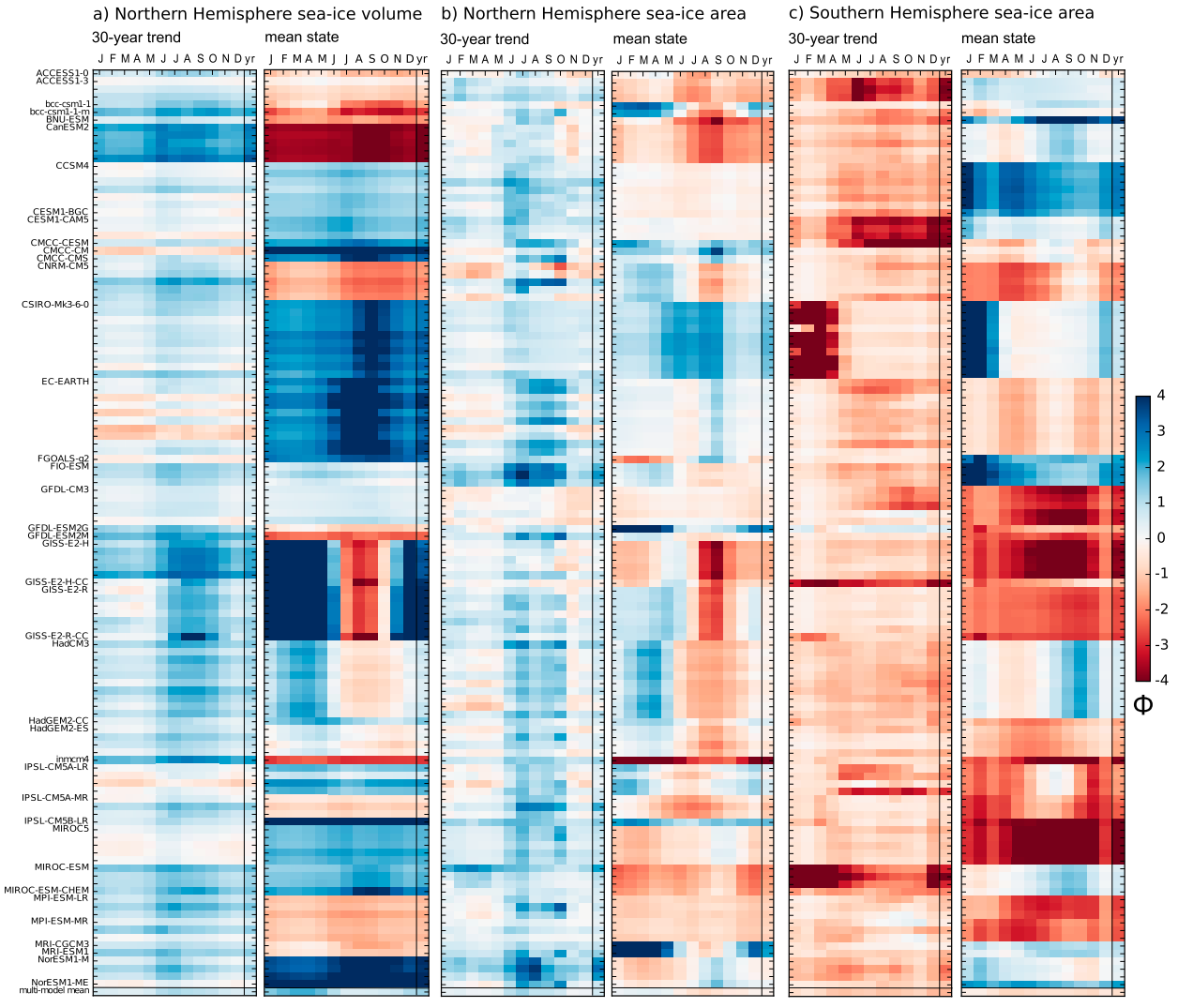

FIG. 7. Portrait plot of the plausibility of CMIP5 sea ice simulations for the 30-yr trend and the mean state of (a) Northern Hemisphere sea ice volume, (b) Northern Hemisphere sea ice area, and (c) Southern Hemisphere sea ice area based on the distance between each extended historical CMIP5 model simulation and reference data (PIOMAS for Northern Hemisphere sea ice volume and the CDR satellite retrieval for sea ice area). Deviations are shown in units of ϕ, which combines δ_{ref} and σ_{mod}; a model's negative (red) and positive (blue) deviation with respect to reference data are indicated. Note that each model name is attached to the first ensemble simulation only.

portrait plot as introduced by Gleckler et al. (2008) (Fig. 7). This is a condensed color-coded way to compare different variables of different model simulations to each other. The color indicates the likelihood of the model simulation to be consistent with the reference data. Red corresponds to a model's negative deviation with respect to the reference data, whereas blue indicates a model's positive deviation. While $\phi = 0$ describes perfect agreement between the model output and the reference data, simulations are plausible at a likelihood of 95% when their deviations from the reference data result in $-2 < \phi < 2$. Deviations larger than $-3 < \phi < 3$ are plausible at a likelihood of 1%.

By now consistently taking model-specific internal variability and reference data uncertainty into account, we find that internal variability can explain much of the differences between the models and the reference data for 30-yr trends in sea ice volume and sea ice area (Figs. 7a–c), whereas for some models it cannot explain the model biases of the mean state (Figs. 7a–c). More specifically, our results reveal that for most models their internal variability is sufficiently high to explain the models' deviations from observed Northern Hemisphere trends in sea ice volume and in sea ice area, from Southern Hemisphere trends in sea ice area, and the mean state of Northern Hemisphere sea ice area. In

contrast, for many models their internal variability cannot explain the model's deviation from the reanalyzed mean state of Northern Hemisphere sea ice volume and observed Southern Hemisphere sea ice area.

Our results confirm previous findings that modeled Northern Hemisphere trends in sea ice area are generally less negative than the observed trends, especially in summer (e.g., Stroeve et al. 2012), and that modeled Southern Hemisphere trends in sea ice area are more negative/less positive than the ones observed (e.g., Mahlstein et al. 2013; Zunz et al. 2013; Haumann et al. 2014). Nevertheless, the internal variability of most of the models can explain the annual mean deviations of the modeled trends in sea ice area from the ones observed at a likelihood of 95%, in line with the discussion by Notz (2015).

Overall, the results show that the plausibility of models differs widely, both within one metric and across metrics. On the one hand, this variety is simply caused by the different model performances in simulating a sea ice metric. On the other hand, this variety is also a result of the different model-specific internal variability taken into account for the model evaluation. For example, the simulations of two different models having the same distance to the reference data will have a different plausibility if one model is characterized by a different internal variability than the other. Evaluating the different model performances in the light of different model-specific internal variability is the strength of this evaluation approach.

6. Summary and conclusions

We present a method that allows us to derive a robust estimate of internal variability in climate model simulations. We combine the control-simulation approach and the ensemble-spread approach that are commonly used for estimating internal variability from model simulations. The method provides practical evidence that the quasi-ergodic assumption holds as long as the internal variability is not changed by external forcing. Across different metrics, we find a linear relationship between the estimates from both approaches, which allows us to translate the modeled internal variability of the preindustrial control climate to that of the historical and future climate. This multimodel approach also allows for robust estimates of historical and future internal variability for models with a single simulation. This new method for estimating internal climate variability for different forcing scenarios is readily transferable to other variables and other applications in future multimodel studies. The applicability of our method is limited when the ensemble spread is averaged over much shorter periods than several decades, or when the ensemble standard deviations estimated for models with ensemble simulations have inconsistent directions of change, such as for future Arctic sea ice area.

We present results from two applications of this method, namely, the assessment of changes of internal variability over time and the evaluation of climate-model simulations. From applying our method to annual global-mean surface air temperature and sea ice volume and area to assessing large-scale changes of internal variability over time, we find

1) a stable internal variability of annual global-mean surface air temperature and sea ice volume and area for the historical climate compared to the preindustrial climate
2) a possibly decreased multimodel mean internal variability of annual global-mean surface air temperature for the RCP8.5 scenario
3) an extremely likely decreased internal variability of winter and summer Arctic sea ice volume and winter and summer Antarctic sea ice volume and a likely decreased internal variability of winter and summer Antarctic sea ice area for a future climate forced by the RCP8.5 scenario, while winter and summer Arctic sea ice area show inconsistent model responses
4) changes in sea ice variability to be largely controlled by changes in the mean sea ice state, except for future Arctic sea ice area, which gets detached from continental boundaries and is vulnerable to increased surface-temperature variability

On a regional scale, the method offers a multimodel consensus view on how and where internal temperature variability is projected to change in future. We find that

1) the pattern of possible temperature-variability change for the historical period already depicts the pattern of change projected for the future, in agreement with temperature extremes that occurred in the last decade [for an overview see Coumou and Rahmstorf (2012)]
2) many midlatitudes are extremely likely to experience or will likely experience strong decreases in surface-temperature variability, while many subtropical, tropical, and polar regions will likely experience strong increases in temperature variability under a future RCP8.5-forced climate. The multimodel consensus pattern of a projected increased temperature variability with possibly associated extreme events, especially on land, has major implications for society

In the context of previous studies (Deser et al. 2000; Screen 2014; Schneider et al. 2015; Holmes et al. 2016), our study suggests a close interplay between the mean

state and the variability of surface air temperature and sea ice volume and area. Under global warming, the mean Arctic sea ice cover decreases linearly with mean Arctic temperature (e.g., Gregory et al. 2002; Mahlstein and Knutti 2012). Our results show that decreases in sea ice variability are directly linked to the decrease in the mean sea ice state as also found by Bathiany et al. (2016). This link between surface-temperature variability and sea ice variability via their mean states is enforced by the high sensitivity of thin sea ice to ocean and atmosphere temperatures (e.g., Bitz and Roe 2004; Notz 2009; Bathiany et al. 2016). For Arctic sea ice area, the link between sea ice variability and the mean sea ice state is less prominent, likely because of the competing effect of increased temperature variability. When sea ice that keeps surface air temperature close to the melting temperature is replaced by open ocean, the temperature variability can increase and thus allows for an increased variability of the remaining sea ice area.

For both annual global-mean surface air temperature and sea ice volume and area, applying our method reveals that the small CMIP5 ensemble size of a model is already representative of the model's total internal variability as shown for CESM1(CAM5) and MPI-ESM-LR based on their large ensembles. This representativeness of only few ensemble simulations of a model and the hugely different manifestation of internal variability in CMIP5 models suggests that a multimodel approach offers more robust estimates for changes in internal variability than results based on single-model large ensembles. The method proves powerful in addressing questions of regional temperature variability such as extreme events, which are commonly investigated using single-model large ensembles. Consequently, we consider the method as a useful tool for studies on internal climate variability complementary to large ensemble simulations of a single model.

For the evaluation of climate model simulations, the method permits a uniform consideration of model-specific internal variability for all models independent of whether they provide several realizations or not. When applied to CMIP5 simulations of sea ice volume and area, we conclude the following:

1) Our multimodel approach discloses a highly variable model-specific internal variability of sea ice volume and area. The different manifestation of internal variability in CMIP5 models hence must be considered in climate model evaluation, as discussed previously by Stroeve et al. (2014), Notz (2014, 2015), and Swart et al. (2015).
2) The consideration of model-specific internal variability in evaluating CMIP5 sea ice simulations is crucial for understanding the discrepancies between model output and reference data. The results allow for a distinction between model deviations that are plausible due to internal variability and reference data uncertainty and those that cannot be explained by these sources of uncertainty and thus point to model biases.

The applications discussed here show the potential of our simple method for estimating internal variability for individual models and across multimodel ensembles. It allows us to gain both a robust assessment of temporal changes in variability and a robust evaluation of model plausibility. In addition, our method allows us to directly quantify the agreement among models, which we often find to be quite low. We hence caution against the overinterpretation of possible changes in internal variability obtained from single-model studies. Finally, while we limited ourselves to an assessment of surface air temperature and sea ice volume and area, our method should be applicable for a wide range of climate variables and thus hopefully contribute to further understanding of internal variability and its role for the climate evolution of our planet.

Acknowledgments. We thank all of the reviewers for their very valuable comments and suggestions, which were essential for shaping the final version of this manuscript. We are grateful to Daniela Matei and David Thompson for their helpful comments and discussions on an earlier version of this manuscript. We further thank the participants of the WWRP/WCRP/Bolin Centre School on Polar Prediction for the valuable discussions, PCMDI for its management of CMIP5, the coupled modeling groups for providing the simulations used here, and NCAR for making the Community Earth System Model large ensemble data publicly available. The research was made possible through a Max Planck Research Fellowship of the Max Planck Society. Scripts used in the analysis and other supporting information that may be useful in reproducing the authors' work are archived by the Max Planck Institute for Meteorology and can be obtained by contacting publications@mpimet.mpg.de.

APPENDIX A

Method

Power spectra of preindustrial control simulations

We analyze the spectral distribution of variability to test whether individual control simulations are too short to cover the model's total internal variability. The logarithmic power spectra (see Fig. A1 for HadCM3 and HadGEM2-ES for annual global-mean surface temperature) reveal that for some models, the maximum

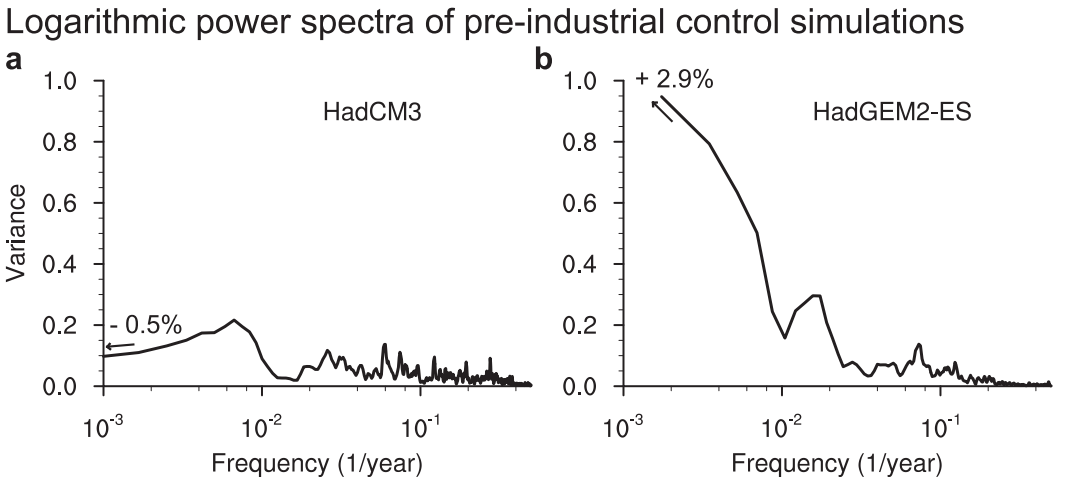

FIG. A1. Logarithmic power spectra for (a) a preindustrial control simulation that is sufficiently long (HadCM3) and (b) a preindustrial control simulation that is too short (HadGEM2-ES) to represent the model's total internal variability of annual global-mean surface temperature. The number indicates the change in variance at the last time-equivalent time step of the spectral distribution of variability. Note that preindustrial control simulations shorter than the one of HadGEM2-ES can be fully adequate because the suitability depends on the manifestation of internal variability in a model. The suitability of a preindustrial control simulation further depends on the variable analyzed.

spectral power of the control simulation does still considerably increase at the time scale provided by the length of the control simulation (e.g., HadGEM2-ES). We here consider an increase in the spectral power at the time scale of the length of the control simulation as considerable when

$$\mathrm{var}(T) - \mathrm{var}(T - 1_{\mathrm{equiv}}) > \left(\sum_{i=1}^{T} \mathrm{var}\right)/100; \quad \text{(A1)}$$

that is, the control simulation's variance increases by more than 1% of the total variance for the last time-equivalent time step $\Delta[T - (T - 1_{\mathrm{equiv}})]$ of the spectral distribution of variability. We choose this criterion because it still allows for slight increases in variance and does not exclude too many models.

APPENDIX B

Reference Data and Uncertainties

a. Sea ice volume

As reference data for the evaluation of Northern Hemisphere sea ice volume, we use reanalysis data from the Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS) (Zhang and Rothrock 2003) that cover the period from 1979 to today. PIOMAS is considered useful for climate-model evaluation (Laxon et al. 2013), as the sea ice–ocean model (i) assimilates sea ice concentrations from satellite retrievals and is forced by NCEP atmospheric reanalysis data and (ii) simulates a sea ice thickness estimate that agrees with past and recent airborne and in situ point measurements and with recent satellite measurements of ICESat (Kwok and Rothrock 2009; Schweiger et al. 2011) and complemented data of CryoSat (Laxon et al. 2013). In a detailed assessment of PIOMAS March sea ice thickness, including additional satellite, submarine, and mooring data, Stroeve et al. (2014) confirmed that PIOMAS is suitable for model evaluation of long-term trends.

Schweiger et al. (2011) discussed the uncertainties in PIOMAS sea ice volume and provided conservative uncertainty estimates for March and October sea ice volume and sea ice volume trends. Based on model sensitivity studies, they stated an uncertainty of the 32-yr trend in sea ice volume of $1.0 \times 10^3\,\mathrm{km}^3\,\mathrm{decade}^{-1}$ and a conservative uncertainty range of $2.25 \times 10^3\,\mathrm{km}^3$ for the mean state of sea ice volume in March and $1.35 \times 10^3\,\mathrm{km}^3$ in October. We interpolate the uncertainty ranges for the other months by weighting them with the monthly mean sea ice volume averaged over the period 1979–2010. To reach a smooth curve of monthly uncertainty estimates that is fixed at the March value of $2.25 \times 10^3\,\mathrm{km}^3$ and inspired by the mean seasonal cycle of Arctic sea ice volume, the July–December values are increased by a factor of 1.12–1.25. Adapting the uncertainty estimate in summer rather than in winter is justified by the uncertainty related to melt ponds in the

sea ice concentration products from satellite retrievals that are assimilated to PIOMAS. These monthly uncertainty estimates define δ_{ref} used when evaluating Northern Hemisphere sea ice volume.

b. Sea ice area

As reference data for the evaluation of modeled sea ice area, we use satellite retrievals of sea ice concentration. The sea ice concentration data product used here is the Climate Data Record of Passive Microwave Sea Ice Concentration (CDR; Meier 2013). The CDR combines different satellite algorithms that, when applied individually, result in different estimates of sea ice concentration dependent on the applied transfer function that translates the passive-microwave signature into sea ice concentration. The reliability of satellite retrievals based on a single algorithm is questioned mainly because of the different treatment of the impact of surface properties (e.g., Lindsay et al. 2014; Titchner and Rayner 2014). The CDR aims to reduce the uncertainty originating from the use of specific algorithms. Therefore, we consider the CDR time series as a best estimate of the ''true'' evolution of sea ice concentration.

To account for the area around the North Pole that is not covered by satellite data, we fill this data hole following the procedure by Olason and Notz (2014). The first satellite observations from 1979 to August 1987 reached only 84.5°N. However for this period, filling the data hole with a sea ice concentration of 1 is reasonable because the latitudes to the south show a constantly dense sea ice concentration as well. This assumption does not hold for the period from August 1987 onward, although the observations now reach 87.2°N. The sea ice concentration starts to become too variable in the central Arctic. We therefore use the mean concentration of the outer rim of the large pre-1987 data hole (i.e., between 84.5° and 87.2°N) to fill the remaining post-1987 data hole of sea ice concentration.

As δ_{ref} estimates for sea ice area, we use the standard deviation estimate provided by the CDR. Note that this standard deviation estimate is available only since August 1987.

REFERENCES

Bathiany, S., B. van der Bolt, M. Williamson, T. M. Lenton, M. Scheffer, E. van Nes, and D. Notz, 2016: Statistical indicators of Arctic sea-ice stability—Prospects and limitations. *Cryosphere*, **10**, 1631–1645, doi:10.5194/tc-10-1631-2016.

Bitz, C., and G. Roe, 2004: A mechanism for the high rate of sea ice thinning in the Arctic Ocean. *J. Climate*, **17**, 3623–3632, doi:10.1175/1520-0442(2004)017<3623:AMFTHR>2.0.CO;2.

Bradley, R. S., and P. D. Jones, 1992: Records of explosive volcanic eruptions over the last 500 years. *Climate Since A.D. 1500*, R. S. Bradley and P. D. Jones, Eds., Routledge, 606–622 .

Briffa, K. R., P. D. Jones, F. H. Schweingruber, and T. J. Osborn, 1998: Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years. *Nature*, **393**, 450–455, doi:10.1038/30943.

Collins, M., and Coauthors, 2013: Long-term climate change: Projections, commitments and irreversibility. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1029–1136.

Coumou, D., and S. Rahmstorf, 2012: A decade of weather extremes. *Nat. Climate Change*, **2**, 491–496, doi:10.1038/nclimate1452.

Deser, C., J. Walsh, and M. Timlin, 2000: Arctic sea ice variability in the context of recent atmospheric circulation trends. *J. Climate*, **13**, 617–633, doi:10.1175/1520-0442(2000)013<0617:ASIVIT>2.0.CO;2.

——, A. S. Phillips, V. Bourdette, and H. Teng, 2012: Uncertainty in climate change projections: The role of internal variability. *Climate Dyn.*, **38**, 527–546, doi:10.1007/s00382-010-0977-x.

——, ——, M. A. Alexander, and B. V. Smoliak, 2014: Projecting North American climate over the next 50 years: Uncertainty due to internal variability. *J. Climate*, **27**, 2271–2296, doi:10.1175/JCLI-D-13-00451.1.

Eisenman, I., T. Schneider, D. S. Battisti, and C. M. Bitz, 2011: Consistent changes in the sea ice seasonal cycle in response to global warming. *J. Climate*, **24**, 5325–5335, doi:10.1175/2011JCLI4051.1.

Esau, I., R. Davy, and S. Outten, 2012: Complementary explanation of temperature response in the lower atmosphere. *Environ. Res. Lett.*, **7**, 044026, doi:10.1088/1748-9326/7/4/044026.

Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866.

Frankcombe, L. M., M. H. England, M. E. Mann, and B. A. Steinman, 2015: Separating internal variability from the externally forced climate response. *J. Climate*, **28**, 8184–8202, doi:10.1175/JCLI-D-15-0069.1.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *J. Geophys. Res.*, **113**, D06104, doi:10.1029/2007JD008972.

Goosse, H., O. Arzel, C. M. Bitz, A. de Montety, and M. Vancoppenolle, 2009: Increased variability of the Arctic summer ice extent in a warmer climate. *Geophys. Res. Lett.*, **36**, L23702, doi:10.1029/2009GL040546.

Gregory, J. M., P. A. Stott, D. J. Cresswell, N. A. Rayner, C. Gordon, and D. M. H. Sexton, 2002: Recent and future changes in Arctic sea ice simulated by the HadCM3 AOGCM. *Geophys. Res. Lett.*, **29**, 2175, doi:10.1029/2001GL014575.

Haumann, F. A., D. Notz, and H. Schmidt, 2014: Anthropogenic influence on recent circulation-driven Antarctic sea ice changes. *Geophys. Res. Lett.*, **41**, 8429–8437, doi:10.1002/2014GL061659.

Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1107, doi:10.1175/2009BAMS2607.1.

——, and ——, 2011: The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dyn.*, **37**, 407–418, doi:10.1007/s00382-010-0810-6.

——, R. Smith, J. Gregory, and D. Stainforth, 2016: Irreducible uncertainty in near-term climate projections. *Climate Dyn.*, **46**, 3807–3819, doi:10.1007/s00382-015-2806-8.

Hingray, B., and M. Said, 2014: Partitioning internal variability and model uncertainty components in a multimember multimodel ensemble of climate projections. *J. Climate*, **27**, 6779–6798, doi:10.1175/JCLI-D-13-00629.1.

Holland, M. M., C. M. Bitz, L. B. Tremblay, and D. A. Bailey, 2008: The role of natural versus forced change in future rapid summer Arctic ice loss. *Arctic Sea Ice Decline: Observations, Projections, Mechanisms, and Implications*, *Geophys. Monogr.*, Vol. 180, Amer. Geophys. Union, 133–150.

Holmes, C. R., T. Woollings, E. Hawkins, and H. de Vries, 2016: Robust future changes in temperature variability under greenhouse gas forcing and the relationship with thermal advection. *J. Climate*, **29**, 2221–2236, doi:10.1175/JCLI-D-14-00735.1.

Huber, M., and R. Knutti, 2014: Natural variability, radiative forcing and climate response in the recent hiatus reconciled. *Nat. Geosci.*, **7**, 651–656, doi:10.1038/ngeo2228.

Huntingford, C., P. D. Jones, V. N. Livina, T. M. Lenton, and P. M. Cox, 2013: No increase in global temperature variability despite changing regional patterns. *Nature*, **500**, 327–330, doi:10.1038/nature12310.

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. Soc.*, **96**, 1333–1349, doi:10.1175/BAMS-D-13-00255.1.

Knutson, T. R., F. Zeng, and A. T. Wittenberg, 2013: Multimodel assessment of regional surface temperature trends: CMIP3 and CMIP5 twentieth-century simulations. *J. Climate*, **26**, 8709–8743, doi:10.1175/JCLI-D-12-00567.1.

Kwok, R., and D. A. Rothrock, 2009: Decline in Arctic sea ice thickness from submarine and ICESat records: 1958–2008. *Geophys. Res. Lett.*, **36**, L15501, doi:10.1029/2009GL039035.

Laxon, S. W., and Coauthors, 2013: CryoSat-2 estimates of Arctic sea ice thickness and volume. *Geophys. Res. Lett.*, **40**, 732–737, doi:10.1002/grl.50193.

Lindsay, R., M. Wensnahan, A. Schweiger, and J. Zhang, 2014: Evaluation of seven different atmospheric reanalysis products in the Arctic. *J. Climate*, **27**, 2588–2606, doi:10.1175/JCLI-D-13-00014.1.

Lu, J., A. Hu, and Z. Zeng, 2014: On the possible interaction between internal climate variability and forced climate change. *Geophys. Res. Lett.*, **41**, 2962–2970, doi:10.1002/2014GL059908.

Mahlstein, I., and R. Knutti, 2012: September Arctic sea ice predicted to disappear near 2°C global warming above present. *J. Geophys. Res.*, **117**, D06104, doi:10.1029/2011JD016709.

——, P. R. Gent, and S. Solomon, 2013: Historical Antarctic mean sea ice area, sea ice trends, and winds in CMIP5 simulations. *J. Geophys. Res. Atmos.*, **118**, 5105–5110, doi:10.1002/jgrd.50443.

Marotzke, J., and P. M. Forster, 2015: Forcing, feedback and internal variability in global temperature trends. *Nature*, **517**, 565–570, doi:10.1038/nature14117.

Massonnet, F., T. Fichefet, H. Goosse, C. M. Bitz, G. Philippon-Berthier, M. M. Holland, and P. Y. Barriat, 2012: Constraining projections of summer Arctic sea ice. *Cryosphere*, **6**, 1383–1394, doi:10.5194/tc-6-1383-2012.

Meehl, G. A., R. H. Moss, K. E. Taylor, V. Eyring, R. J. Stouffer, S. Bony, and B. Stevens, 2014: Climate model intercomparisons: Preparing for the next phase. *Eos, Trans. Amer. Geophys. Union*, **95**, 77–84, doi:10.1002/2014EO090001.

Meier, W., F. Fetterer, M. Savoie, S. Mallory, R. Duerr, and J. Stroeve, 2013: NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, version 2. National Snow and Ice Data Center, Boulder, CO, accessed 9 December 2015, https://doi.org/10.7265/N55M63M1.

Moss, R. H., and Coauthors, 2010: The next generation of scenarios for climate change research and assessment. *Nature*, **463**, 747–756, doi:10.1038/nature08823.

Notz, D., 2009: The future of ice sheets and sea ice: Between reversible retreat and unstoppable loss. *Proc. Natl. Acad. Sci. USA*, **106**, 20 590–20 595, doi:10.1073/pnas.0902356106.

——, 2014: Sea-ice extent and its trend provide limited metrics of model performance. *Cryosphere*, **8**, 229–243, doi:10.5194/tc-8-229-2014.

——, 2015: How well must climate models agree with observations? *Philos. Trans. Roy. Soc. London*, **373A**, 20140164, doi:10.1098/rsta.2014.0164.

Olason, E., and D. Notz, 2014: Drivers of variability in Arctic sea-ice drift speed. *J. Geophys. Res. Oceans*, **119**, 5755–5775, doi:10.1002/2014JC009897.

Palmer, M. D., and D. J. McNeall, 2014: Internal variability of Earth's energy budget simulated by CMIP5 climate models. *Environ. Res. Lett.*, **9**, 034016, doi:10.1088/1748-9326/9/3/034016.

Resplandy, L., R. Séférian, and L. Bopp, 2015: Natural variability of CO_2 and O_2 fluxes: What can we learn from centuries-long climate models simulations? *J. Geophys. Res. Oceans*, **120**, 384–404, doi:10.1002/2014JC010463.

Roberts, C. D., M. D. Palmer, D. McNeall, and M. Collins, 2015: Quantifying the likelihood of a continued hiatus in global warming. *Nat. Climate Change*, **5**, 337–342, doi:10.1038/nclimate2531.

Rosenblum, E., and I. Eisenman, 2016: Faster Arctic sea ice retreat in CMIP5 than in CMIP3 due to volcanoes. *J. Climate*, **29**, 9179–9188, doi:10.1175/JCLI-D-16-0391.1.

Santer, B. D., and Coauthors, 2008: Consistency of modelled and observed temperature trends in the tropical troposphere. *Int. J. Climatol.*, **28**, 1703–1722, doi:10.1002/joc.1756.

——, and Coauthors, 2014: Volcanic contribution to decadal changes in tropospheric temperature. *Nat. Geosci.*, **7**, 185–189, doi:10.1038/ngeo2098.

Schindler, A., A. Toreti, M. Zampieri, E. Scoccimarro, S. Gualdi, S. Fukutome, E. Xoplaki, and J. Luterbacher, 2015: On the internal variability of simulated daily precipitation. *J. Climate*, **28**, 3624–3630, doi:10.1175/JCLI-D-14-00745.1.

Schneider, E., and J. Kinter, 1994: An examination of internally generated variability in long climate simulations. *Climate Dyn.*, **10**, 181–204, doi:10.1007/BF00208987.

Schneider, T., T. Bischoff, and H. Plotka, 2015: Physics of changes in synoptic midlatitude temperature variability. *J. Climate*, **28**, 2312–2331, doi:10.1175/JCLI-D-14-00632.1; Corrigendum, **29**, 3471, doi:10.1175/JCLI-D-16-0096.1.

Schweiger, A., R. Lindsay, J. Zhang, M. Steele, H. Stern, and R. Kwok, 2011: Uncertainty in modeled Arctic sea ice volume. *J. Geophys. Res.*, **116**, C00D06, doi:10.1029/2011JC007084.

Screen, J. A., 2014: Arctic amplification decreases temperature variance in northern mid- to high-latitudes. *Nat. Climate Change*, **4**, 577–582, doi:10.1038/nclimate2268.

Shu, Q., Z. Song, and F. Qiao, 2015: Assessment of sea ice simulations in the CMIP5 models. *Cryosphere*, **9**, 399–409, doi:10.5194/tc-9-399-2015.

Stroeve, J. C., V. Kattsov, A. Barrett, M. Serreze, T. Pavlova, M. Holland, and W. N. Meier, 2012: Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations. *Geophys. Res. Lett.*, **39**, L16502, doi:10.1029/2012GL052676.

——, A. Barrett, M. Serreze, and A. Schweiger, 2014: Using records from submarine, aircraft and satellites to evaluate climate model simulations of Arctic sea ice thickness. *Cryosphere*, **8**, 1839–1854, doi:10.5194/tc-8-1839-2014.

Sutton, R., E. Suckling, and E. Hawkins, 2015: What does global mean temperature tell us about local climate? *Philos.*

Trans. Roy. Soc. London, **373A**, 20140426, doi:10.1098/rsta.2014.0426.

Swanson, K. L., G. Sugihara, and A. A. Tsonis, 2009: Long-term natural variability and 20th century climate change. *Proc. Natl. Acad. Sci. USA*, **106**, 16 120–16 123, doi:10.1073/pnas.0908699106.

Swart, N. C., J. C. Fyfe, E. Hawkins, J. E. Kay, and A. Jahn, 2015: Influence of internal variability on Arctic sea-ice trends. *Nat. Climate Change*, **5**, 86–89, doi:10.1038/nclimate2483.

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.

Thompson, D. W. J., E. A. Barnes, C. Deser, W. E. Foust, and A. S. Phillips, 2015: Quantifying the role of internal climate variability in future climate trends. *J. Climate*, **28**, 6443–6456, doi:10.1175/JCLI-D-14-00830.1.

Titchner, H. A., and N. A. Rayner, 2014: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *J. Geophys. Res. Atmos.*, **119**, 2864–2889, doi:10.1002/2013JD020316.

Trenberth, K. E., 2011: Attribution of climate variations and trends to human influences and natural variability. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 925–930, doi:10.1002/wcc.142.

von Neumann, J., 1932: Proof of the quasi-ergodic hypothesis. *Proc. Natl. Acad. Sci. USA*, **18**, 70–82, doi:10.1073/pnas.18.1.70.

Wettstein, J. J., and C. Deser, 2014: Internal variability in projections of twenty-first-century arctic sea ice loss: Role of the large-scale atmospheric circulation. *J. Climate*, **27**, 527–550, doi:10.1175/JCLI-D-12-00839.1.

Zhang, J., and D. Rothrock, 2003: Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates. *Mon. Wea. Rev.*, **131**, 845–861, doi:10.1175/1520-0493(2003)131<0845:MGSIWA>2.0.CO;2.

Zunz, V., H. Goosse, and F. Massonnet, 2013: How does internal variability influence the ability of CMIP5 models to reproduce the recent trend in Southern Ocean sea ice extent? *Cryosphere*, **7**, 451–468, doi:10.5194/tc-7-451-2013.