Moving beyond the Total Sea Ice Extent in Gauging Model Biases

DETELINA P. IVANOVA

*Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory,
Livermore, California, and Nansen Environmental and Remote Sensing Center, and Bjerknes
Centre for Climate Research, Bergen, Norway*

PETER J. GLECKLER, KARL E. TAYLOR, AND PAUL J. DURACK

*Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore
National Laboratory, Livermore, California*

KATE D. MARVEL

NASA Goddard Institute for Space Studies, Columbia University, New York, New York

(Manuscript received 20 January 2016, in final form 5 September 2016)

ABSTRACT

Reproducing characteristics of observed sea ice extent remains an important climate modeling challenge. This study describes several approaches to improve how model biases in total sea ice distribution are quantified, and applies them to historically forced simulations contributed to phase 5 of the Coupled Model Intercomparison Project (CMIP5). The quantity of hemispheric total sea ice area, or some measure of its equatorward extent, is often used to evaluate model performance. A new approach is introduced that investigates additional details about the structure of model errors, with an aim to reduce the potential impact of compensating errors when gauging differences between simulated and observed sea ice. Using multiple observational datasets, several new methods are applied to evaluate the climatological spatial distribution and the annual cycle of sea ice cover in 41 CMIP5 models. It is shown that in some models, error compensation can be substantial, for example resulting from too much sea ice in one region and too little in another. Error compensation tends to be larger in models that agree more closely with the observed total sea ice area, which may result from model tuning. The results herein suggest that consideration of only the total hemispheric sea ice area or extent can be misleading when quantitatively comparing how well models agree with observations. Further work is needed to fully develop robust methods to holistically evaluate the ability of models to capture the finescale structure of sea ice characteristics; however, the “sector scale” metric used here aids in reducing the impact of compensating errors in hemispheric integrals.

1. Introduction

Some of the most dramatic changes evident in Earth's climate system are in the observed reduction in Arctic sea ice cover, raising concern from the scientific community,

policy-makers, and the general public. However, many challenges remain in effectively simulating observed sea ice loss, and improved simulation of sea ice has been described as a “grand challenge” for climate modeling (Kattsov et al. 2010). In this study, we focus on developing improved methods to objectively compare the large-scale climatological characteristics of simulated sea ice with available observations.

In the previous generation of climate models (CMIP3), Parkinson et al. (2006) found that the phase and amplitude of the annual cycle of total sea ice extent agreed reasonably well with observations, performing better in the Northern Hemisphere (NH) than the Southern Hemisphere (SH). Recent studies examining simulated sea ice in the CMIP5 multimodel ensemble have analyzed

Denotes Open Access content.

Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JCLI-D-16-0026.s1>.

Corresponding author address: Detelina P. Ivanova, Nansen Environmental and Remote Sensing Center, Thormøhlens gate 47, N-5006, Bergen, Norway.
E-mail: detelina.ivanova@nersc.no

DOI: 10.1175/JCLI-D-16-0026.1

TABLE 1. Total sea ice area (10^6 km²) statistics [mean (std)].

	Arctic			Antarctic		
	Winter	Summer	Annual	Winter	Summer	Annual
NT	12.79 (0.44)	5.32 (0.37)	9.54 (3.07)	13.67 (0.21)	2.57 (0.29)	8.47 (4.57)
BT	13.67 (0.43)	6.56 (0.41)	10.46 (2.94)	15.24 (0.27)	2.84 (0.30)	9.45 (5.11)
CMIP5 MMEM	14.15 (1.77)	5.86 (1.59)	10.38 (1.52)	13.66 (4.13)	2.31 (2.18)	8.31 (3.45)

the mean distribution and annual cycle of total sea ice extent, as well as trends in the historical and future projections of the Arctic and Antarctic (e.g., Pavlova et al. 2011; Stroeve et al. 2012; Massonnet et al. 2012; Zunz et al. 2013; Wang and Overland 2012; Shu et al. 2015). Collectively, these studies suggest that the newer CMIP5 models have somewhat improved sea ice simulation in comparison with CMIP3. Behrens et al. (2012) and Semenov et al. (2015) analyzed the CMIP3 and CMIP5 simulated sea ice area in different Arctic regions and identified some improvement in CMIP5 regional performance. Agreement with observations varies substantially with location, being for example quite good in the central Arctic but rather poor in the Barents Sea.

In the majority of the earlier studies, the total sea ice extent has been evaluated as a hemispheric quantity. This approach is attractive because it is straightforward to compute and interpret. However, the results may be misleading if there is substantial error compensation due to simulated sea ice being excessive in some areas while at the same time deficient in others. More recently, Notz (2014) has discussed the limitations of the total sea ice extent as a model performance metric, some of which we summarize here. 1) Even though total sea ice extent (TSIE) is preferable due to the smaller observational uncertainty compared to the total sea ice area (TSIA), the estimates of the sea ice cover derived from it are only approximate since by definition it includes an open water fraction as well. Also, 2) TSIE may be misleading as a metric for model performance since models with the same sea ice extent may have different sea ice area distributions. We concur with the conclusion of Notz (2014) that sea ice area, which is the actual physical quantity of the sea ice cover, is more appropriate than sea ice extent for gauging the consistency between model and observations. In an attempt to take this further, a primary motivation for this study is to develop and apply improved measures for gauging the differences between observed and simulated total sea ice by reducing the possible influence of compensating errors and to extend the analysis beyond the hemispheric total.

In section 2 we discuss the observational and model data used in this study. We then summarize several existing metrics and introduce a new metric. In section 3, we discuss our results from CMIP5 models evaluation

using some of the metrics described in the previous section and propose a new approach to examining the contributions to total errors in simulating sea ice distribution and extent. In section 4 we summarize our methods and results and discuss their implications.

2. Data and methods

a. Observations

Most CMIP5 ice evaluation studies (Pavlova et al. 2011; Stroeve et al. 2012; Massonnet et al. 2012) to date have solely relied upon the Sea Ice Index (Fetterer et al. 2016), which is derived from two daily gridded products based on the NASA Team algorithm. To provide a simple characterization of observational uncertainty in our model evaluation, we use two observational datasets for the sea ice concentration (SIC) prepared by the National Snow and Ice Data Center (NSIDC), differing only by the algorithms used to process the SSM/I satellite sensor records—NASA Team (NT; Cavalieri et al. 1996) and Bootstrap (BT; Comiso 2000). The known climatological differences between the two sets are $\sim 1\%$ for the total ice extent and $\sim 7.5\%$ – 10% for the ice area [Comiso et al. 1997; see also Notz (2014) and section 3c herein]. Several improved algorithms have been subsequently developed using higher resolution and higher-frequency channels (Markus and Cavalieri 2000; Meier 2005; Andersen et al. 2007), but the length of their record is shorter and for the purposes of our study the consistent and continuous 30-yr records of the NT and BT datasets are more appropriate for comparison to long-term model climatologies. Our choice is further supported by the recent findings of Schneider et al. (2012), who compared climatologies of various products based on single or multiple algorithms or data sources, and concluded that the single-algorithm datasets such as NT and BT are currently the most reliable for climate studies. The hemispheric TSIA mean and standard deviation (std) of the observations are shown in Table 1 along with the model results, which are described below.

b. Model simulations

Our focus is on the historically forced simulations contributed to CMIP5. We consider results from 41

models, with many models performing small ensembles of simulations (“realizations”), resulting in a total of 154 historical runs (Table 2). The CMIP5 sea ice output is archived on the native grid of each model, which means that sea ice area can be accurately calculated. Our analysis is based on the monthly mean sea ice concentration (see below), but we also make use of ancillary data including the time-invariant grid cell area of each model. When computing multimodel ensemble mean (MMEM) results, we use all 41 CMIP5 models considered in this study. Since many of the model groups have contributed only one realization to ensure an equal contribution from each model, we consider only a single ensemble member of each participating model. Even though the focus of our study is on the evaluation of long-term climatological mean characteristics of the sea ice, one must always consider the potential importance of internal variability, which has been found previously to be a major factor of uncertainty when evaluating model performance (Notz 2014, 2015; Zunz et al. 2013; Santer et al. 2011). We note, however, that for our hemispheric climatological tests we have found single model interrealization differences to be small compared to intermodel differences. Additional information about the individual models and their sea ice components is summarized in Table 2.

c. Regions

We define six regions based on each ocean basin sector of the polar areas: the North Atlantic sector of the Arctic (NA), North Pacific sector of Arctic (NP), Central Arctic (CA), South Atlantic sector of Antarctic (SA), South Pacific sector of Antarctic (SP), and South Indian sector of Antarctic (IO). The boundaries of each region (or hereafter equivalently “sectors”) used in our analysis are shown in Fig. 1 and the domains summarized in Table 3. Rather than referring to the Northern and Southern Hemisphere results, we refer to the Arctic and Antarctic results. We note, however, the important distinction between our use of the terms “Arctic” (hemispheric total) and “Central Arctic” (one of our regions).

d. Sea ice metrics

Before describing our results, we summarize several sea ice metrics commonly used to define the distribution of sea ice. We also introduce a new sea ice diagnostic, defined by our analysis.

Recent studies of simulated sea ice in CMIP5 have evaluated the mean, annual cycle, and trends of the *total sea ice extent* in each hemisphere, which is usually defined as the cumulative sum of the area of all model grid cells (or observed pixels) with at least 15% ice concentration (e.g., Pavlova et al. 2011; Stroeve et al. 2012;

Massonnet et al. 2012; Zunz et al. 2013). Similarly to Notz (2014), we compare this approach of evaluating sea ice “extent” with the actual sea ice covered area which excludes the area of the open water part in the grid cells (pixels), and for this purpose we define the *total sea ice area* as the cumulative sum of the area of the grid cells (pixels) with at least 15% ice concentration multiplied by the ice fraction in the grid cell (pixel). To ensure an accurate calculation of TSIA, we use the model and observational data on their native grids and apply their native land–sea masks, rather than first interpolating data to a common grid.

Another metric commonly used to assess the spatial characteristics of the sea ice cover is the 2D climatological spatial patterns of sea ice concentration (SIC). The statistical description of the Arctic SIC is complicated due to its bimodal type of distribution (Figs. 2a–d), with two major modes—one in the range of >90% SIC and another in the range of <10%. This SIC bimodality results generally due to the fact that the sea ice cover can be considered a mixture of patches of ice and open water (leads). At fine enough grid resolution the distribution would become purely binary, with grid elements either being 100% ice covered or ice free. As non-Gaussian variables, the interpretation of statistics (such as means and standard deviations) commonly used to characterize the SIC distributions is not straightforward. One way to avoid this difficulty is to use more integrated measures (Bernstein et al. 2015).

The September and March Arctic SIC distributions of the CMIP5 models together with NASA Team and Bootstrap observations have been analyzed before by Notz (2014). Here we expand upon his findings by showing the seasonal (winter and summer) distributions in both Arctic and Antarctic for the two observational sets (NT and BT) and the CMIP5 multimodel ensemble mean. Consistent with the results by Notz (2014), we find that all Arctic observed distributions are in a close agreement in winter (Fig. 2a) when they are highly compact (>90%) but disagree in summer (Fig. 2b), particularly in the high concentration ranges (80%–100%). One challenge is interpreting the difference between the two observational datasets; while the BT observational distribution stays highly compact (maximum frequency of 90%–100% range), the NT observations feature a somewhat broader summer distribution (maximum frequency in 80%–90% range). These differences have been explained as arising from the different treatments of the melt ponds in the two satellite algorithms (Meier and Notz 2010). The CMIP5 MMEM is in better agreement with the NT summer distribution (Fig. 2b). In his study, Notz (2014) classified the individual CMIP5 models in two groups, one with compact (agreeing better with

TABLE 2. CMIP5 model descriptions. (Expansions of acronyms available online at <http://www.ametsoc.org/PubsAcronymList>.)

Model	Institution	Country	Sea ice model	Spatial resolution	Sea ice component description	Number of historical ensemble members	Reference	Web site
ACCESS1.0	CSIRO-BOM	Australia	CICE, v4.1	Tripolar, $1^\circ \times 1^\circ$, refinement at the equator	Energy-conserving thermo, ITD, EVP	1	Bi et al. 2013a,b	https://wiki.csiro.au/display/ACCESS/Home
ACCESS1.3	CSIRO-BOM	Australia	CICE, v4.1	Tripolar, $1^\circ \times 1^\circ$, refinement at the equator	Energy-conserving thermo, ITD, EVP	3	Bi et al. 2013a,b	https://wiki.csiro.au/display/ACCESS/Home
BCC-CSM1.1-M	BCC	China	SIS	Tripolar, $\sim 1^\circ \times (1-1/3)^\circ$	Semtner 3-layer, ITD, EVP	3	Xin et al. 2013	http://bcc.cma.gov.cn/bccsm/web/channel-43.htm
BCC-CSM1.1	BCC	China	SIS	Tripolar, $\sim 1^\circ \times (1-1/3)^\circ$	Semtner 3-layer, ITD, EVP	3	Xin et al. 2013	http://bcc.cma.gov.cn/bccsm/web/channel-43.htm
CCSM4	NCAR	United States	CICE, v4	Dipolar, $1.11^\circ \times (0.27-0.54)^\circ$, NP in Greenland	Energy-conserving thermo, ITD, EVP	6	Gent et al. 2011	http://www.cesm.ucar.edu/experiments/cmp5.html
CESM1-BGC	NSF-DOE-NCAR	United States	CICE, v4	Dipolar, $1.11^\circ \times (0.27-0.54)^\circ$, NP in Greenland	Energy-conserving thermo, ITD, EVP	1	Gent et al. 2011	http://www.cesm.ucar.edu/experiments/cmp5.html
CESM1-CAM5-1-FV2	NSF-DOE-NCAR	United States	CICE, v4	Dipolar, $1.11^\circ \times (0.27-0.54)^\circ$, NP in Greenland	Energy-conserving thermo, ITD, EVP	4	Gent et al. 2011	http://www.cesm.ucar.edu/experiments/cmp5.html
CESM1-CAM5	NSF-DOE-NCAR	United States	CICE, v4	Dipolar, $1.11^\circ \times (0.27-0.54)^\circ$, NP in Greenland	Energy-conserving thermo, ITD, EVP	3	Gent et al. 2011	http://www.cesm.ucar.edu/experiments/cmp5.html
CESM1-FASTCHEM	NSF-DOE-NCAR	United States	CICE, v4	Dipolar, $1.11^\circ \times (0.27-0.54)^\circ$, NP in Greenland	Energy-conserving thermo, ITD, EVP	3	Gent et al. 2011	http://www.cesm.ucar.edu/experiments/cmp5.html
CESM1-WACCM	NSF-DOE-NCAR	United States	CICE, v4	Dipolar, $1.11^\circ \times (0.27-0.54)^\circ$, NP in Greenland	Energy-conserving thermo, ITD, EVP	4	Gent et al. 2011	http://www.cesm.ucar.edu/experiments/cmp5.html
CMCC-CM	CMCC	Italy	LIM2	ORCA-2° tripolar	Semtner 3-layer+brine pockets, Virtual ITD, VP	1	N/A	http://www.cmcc.it/
CMCC-CMS	CMCC	Italy	LIM2	ORCA-2° tripolar	Semtner 3-layer+brine pockets, Virtual ITD, VP	1	N/A	http://www.cmcc.it/
CNRM-CM5	CNRM-CERFACS	France	GELATO, v5	ORCA-1° tripolar	ITD, EVP	10	Volodire et al. 2013	http://www.cnrm-meteo.fr/cmp5/spip.php?article7

TABLE 2. (Continued)

Model	Institution	Country	Sea ice model	Spatial resolution	Sea ice component description	Number of historical ensemble members	Reference	Web site
CSIRO-Mk3.6.0	CSIRO-QCCCE	Australia	Component of Mk3.6	T63 Gaussian	Cavitating fluid	10	Jeffrey et al. 2013	https://wiki.csiro.au/display/CSIROMk360/Home
CanCM4	CCCMA	Canada	CanSIM1	T63 Gaussian	Cavitating fluid	10	N/A	http://www.cccma.ec.gc.ca/models
CanESM2	CCCMA	Canada	CanSIM1	T63 Gaussian	Cavitating fluid	5	N/A	http://www.cccma.ec.gc.ca/models
FGOALS-s2	LAGS-IAP	China	CSIM5	$(0.5-1)^{\circ} \times (0.5-1)^{\circ}$	Energy-conserving thermo, ITD, EVP	3	Bao et al. 2013	http://www.lasg.ac.cn/fgoals/index2.asp
GFDL-CM2p1	NOAA GFDL	United States	SISp2	Tripolar, $\sim 1^{\circ} \times 1^{\circ}$	Modified Semtner 3-layer, ITD, EVP	10	N/A	http://www.gfdl.noaa.gov/gfdl-cmip5-ipcc-ar5-database
GFDL-CM3	NOAA GFDL	United States	SISp2	Tripolar, $\sim 1^{\circ} \times 1^{\circ}$	Modified Semtner 3-layer, ITD, EVP	5	Griffies et al. 2011	http://www.gfdl.noaa.gov/coupled-physical-model-cm3
GFDL-ESM2G	NOAA GFDL	United States	SISp2	Tripolar, $\sim 1^{\circ} \times 1^{\circ}$	Modified Semtner 3-layer, ITD, EVP	1	Dunne et al. 2012	http://www.gfdl.noaa.gov/earth-system-model
GFDL-ESM2M	NOAA GFDL	United States	SISp2	Tripolar, $\sim 1^{\circ} \times 1^{\circ}$	Modified Semtner 3-layer, ITD, EVP	1	Dunne et al. 2012	http://www.gfdl.noaa.gov/earth-system-model
GISS-E2-H-CC	NASA GISS	United States	Russell	Tripolar, $\sim 1^{\circ} \times 1^{\circ}$	Modified Semtner 0-layer, VP	1	Schmidt et al. 2014	http://www.giss.nasa.gov/tools/modelE/
GISS-E2-H	NASA GISS	United States	Russell	Tripolar, $\sim 1^{\circ} \times 1^{\circ}$	Modified Semtner 0-layer, VP	5	Schmidt et al. 2014	http://www.giss.nasa.gov/tools/modelE/
GISS-E2-R-CC	NASA GISS	United States	Russell	$1^{\circ} \times 1.25^{\circ}$	Modified Semtner 0-layer, VP	1	Schmidt et al. 2014	http://www.giss.nasa.gov/tools/modelE/
GISS-E2-R	NASA GISS	United States	Russell	$1^{\circ} \times 1.25^{\circ}$	Modified Semtner 0-layer, VP	6	Schmidt et al. 2014	http://www.giss.nasa.gov/tools/modelE/
HadCM3	MOHC	United Kingdom	Component of HadOM3	$1.25^{\circ} \times 1.25^{\circ}$	Semtner 0-layer, VP	10	Gordon et al. 2000	http://www.metoffice.gov.uk/research/modelling-systems/unified-models/climate-models
HadGEM2-AO	NIMR/KMA	South Korea	Inspired from CICE	$1^{\circ} \times 1^{\circ}$	Semtner 0-layer, ITD, EVP	1	Martin et al. 2011	http://www.metoffice.gov.uk/research/modelling-systems/unified-model/climate-models

TABLE 2. (Continued)

Model	Institution	Country	Sea ice model	Spatial resolution	Sea ice component description	Number of historical ensemble members	Reference	Web site
HadGEM2-CC	MOHC	United Kingdom	Inspired from CICE	$(1-0.3)^{\circ} \times 1^{\circ}$	Semtnr 0-layer, ITD, EVP	3	Martin et al. 2011	http://www.metoffice.gov.uk/research/modelling-systems/unified-model/climate-models
HadGEM2-ES	MOHC	United Kingdom	Inspired from CICE	$(1-0.3)^{\circ} \times 1^{\circ}$	Semtnr 0-layer, ITD, EVP	5	Martin et al. 2011	http://www.metoffice.gov.uk/research/modelling-systems/unified-model/climate-models
IPSL-CM5A-LR	IPSL	France	LIM2	ORCA-2°-tripolar	Semtnr 3-layer+brine pockets, virtual ITD, VP	6	Dufresne et al. 2013	http://icmc.ipsl.fr/research/international-projects/cmip5
IPSL-CM5A-MR	IPSL	France	LIM2	ORCA-2°-tripolar	Semtnr 3-layer+brine pockets, virtual ITD, VP	3	Dufresne et al. 2013	http://icmc.ipsl.fr/research/international-projects/cmip5
IPSL-CM5B-LR	IPSL	France	LIM2	ORCA-2°-tripolar	Semtnr 3-layer+brine pockets, virtual ITD, VP	1	Dufresne et al. 2013	http://icmc.ipsl.fr/research/international-projects/cmip5
MIROC-ESM-CHEM	MIROC	Japan	Component of COCO v3.4	$\sim 1.4^{\circ} \times 1^{\circ}$	Semtnr 0-layer, 2 ice categories, EVP	1	Watanabe et al. 2011	http://amaterasu.ees.hokudai.ac.jp/~fswiki/pub/wiki.cgi?page=CMIP5
MIROC-ESM	MIROC	Japan	Component of COCO v3.4	$\sim 1.4^{\circ} \times 1^{\circ}$	Semtnr 0-layer, 2 ice categories, EVP	3	Watanabe et al. 2011	http://amaterasu.ees.hokudai.ac.jp/~fswiki/pub/wiki.cgi?page=CMIP5%2FEExperiments%2FModels+and+Grid
MIROC4h	MIROC	Japan	Component of COCO v3.4	$0.28^{\circ} \times 0.19^{\circ}$, Spherical poles rotated	Semtnr 0-layer, 2 ice categories, EVP	3	Sakamoto et al. 2012; Komuro et al. 2012	http://amaterasu.ees.hokudai.ac.jp/~fswiki/pub/wiki.cgi?page=CMIP5%2FEExperiments%2FModels+and+Grid
MIROC5	MIROC	Japan	Component of COCO v4.5	$\sim 1.4^{\circ} \times (0.5-1.4)^{\circ}$, Generalized spherical, shifted poles	Semtnr 0-layer, ITD, EVP	5	Komuro et al. 2012	http://amaterasu.ees.hokudai.ac.jp/~fswiki/pub/wiki.cgi?page=CMIP5%2FEExperiments%2FModels+and+Grid

TABLE 2. (Continued)

Model	Institution	Country	Sea ice model	Spatial resolution	Sea ice component description	Number of historical ensemble members	Reference	Web site
MPI-ESM-LR	MPI-M	Germany	Component of MPI-OM	$\sim 1.5^\circ \times 1.5^\circ$	Semtner 0-layer, Virtual ITD, VP	3	Notz et al. 2013	http://www.mpimet.mpg.de/en/science/models/mpi-esm.html
MPI-ESM-MR	MPI-M	Germany	Component of MPI-OM	$\sim 0.4^\circ \times 0.4^\circ$	Semtner 0-layer, Virtual ITD, VP	3	Notz et al. 2013	http://www.mpimet.mpg.de/en/science/models/mpi-esm.html
MPI-ESM-P	MPI-M	Germany	Component of MPI-OM	$\sim 1.5^\circ \times 1.5^\circ$	Semtner 0-layer, Virtual ITD, VP	2	Notz et al. 2013	http://www.mpimet.mpg.de/en/science/models/mpi-esm.html
NorESM1-M	NCC	Norway	CICE, v4	Dipolar, $1.11^\circ \times (0.27\text{--}0.54)^\circ$, NP in Greenland	Energy-conserving thermo, ITD, EVP	3	Bentsen et al. 2013, Iversen et al. 2013	http://www.uib.no/People/ngfhd/EarthClim/index.htm
NorESM1-ME	NCC	Norway	CICE, v4	Dipolar, $1.11^\circ \times (0.27\text{--}0.54)^\circ$, NP in Greenland	Energy-conserving thermo, ITD, EVP	1	Bentsen et al. 2013, Iversen et al. 2013	http://www.uib.no/People/ngfhd/EarthClim/index.htm
EC-EARTH	EC-EARTH consortium	EU	LIM2	ORCA-1 $^\circ$ tripolar	Semtner 3-layer + brine pockets, virtual ITD, VP	—	Hazeleger et al. 2012	http://eearth.knmi.nl/
INM-CM4	INM	Russia	Component of INM-CM4 ocean	$1^\circ \times 0.5^\circ$	Semtner 0-layer, EVP	—	Volodin et al. 2010, Yakovlev 2009	http://ksv.inm.ras.ru/GCM_DATA_PLOTTING/GCM_INM_DATA_en.html

a) Arctic Regions Mask

b) Antarctic Regions Mask

FIG. 1. Region bounds: (a) Arctic regions: NA (North Atlantic sector), NP (North Pacific sector), and CA (Central Arctic sector) and (b) Antarctic regions: SA (South Atlantic sector), SP (South Pacific sector) and IO (Indian Ocean sector).

BT observations) and another with loose (in better agreement with NT observations) Arctic summer distribution. In contrast to the Arctic, the observed and simulated Antarctic summer SIC distributions shown here (Figs. 2c,d) are in better agreement than in winter, simply because the high concentrations (the problematic one for the satellite algorithms) are less common and the distributions are dominated by the low ice concentrations. The Antarctic winter SIC distributions have the same problem as the Arctic summer, loose ice pack in

TABLE 3. Region boundaries.

Central Arctic (CA)	80°–90°N, 120°W–90°E; 65°–90°N, 90°E–120°W
North Atlantic (NA)	45°–80°N, 120°W–90°E
North Pacific (NP)	45°–65°N, 90°E–120°W
South Atlantic (SA)	90°–55°S, 60°W–20°E
South Pacific (SP)	90°–55°S, 60°W–90°E
Indian Ocean (IO)	90°–55°S, 20°–90°E

NT and CMIP5 MMEM and compact in BT (Fig. 2c); similarly, the CMIP5 MMEM agrees better with the NT data.

Another measure of sea ice spatial distribution often used in various socioeconomic applications (e.g., navigation, fishery) is the sea ice edge (SIEd), usually defined as the latitude to which the sea ice extends (more specifically the latitude of the 15% SIC contour, the upper bound of the observational measurement uncertainty). The distribution of SIEd, sampled over all longitudes, is mostly unimodal (Figs. 2e–h) and close to normal in the winter season, which facilitates interpretation because, for example, the mean of the distribution is more meaningful than in the bimodal case. Moreover, there is probably less observational uncertainty due to usage of different satellite algorithms in this localized measure since the two observational sets are in better agreement than in the case of SIC. On the other hand, correctly simulating SIEd seems to be a tougher challenge for the CMIP5 models as evidenced by inconsistencies between the observed and MMEM distributions, particularly in the Antarctic summer (Fig. 2h).

We note that, compared with TSIA, the accuracy with which the current generation of climate models (with nominal grid resolution of $\sim 1^\circ$) are capable of reproducing this fine spatial characteristic is more limited. Moreover, if we were to plot SIEd (not shown) as a function of longitude in the Arctic, we would find some models with huge differences along longitudes intersected by Greenland. This is because whether or not there is a small amount of ice just south of Greenland changes SIEd considerably (~ 20 degrees of longitude). This local difference might be considered of minor importance to global climate but would dominate any error statistic based on SIEd. Another limitation of SIEd is that a model could get the correct extent of sea ice but have open ocean regions between the ice edge and the pole that might be entirely unrealistic. A downside of SIEd in terms of standard test for model evaluation is that its frequency distribution characteristics change hemispherically and seasonally. For example, in the Antarctic it shifts from a 68° – 52° S range of latitudes with distribution maxima (or the mode of the frequency

FIG. 2. Histograms of NASA Team, Bootstrap, and CMIP5 multimodel ensemble mean (MMEM) seasonal normalized frequency distributions for (a) Arctic winter SIC, (b) Arctic summer SIC, (c) Antarctic winter SIC, (d) Antarctic summer SIC (but eliminating from consideration grid cells with SIC less than 1×10^{-6} %), (e) Arctic winter SIED, (f) Arctic summer SIED, (g) Antarctic winter SIED, (h) Antarctic summer SIED, (i) Arctic winter MISIA, (j) Arctic summer MISIA, (k) Antarctic winter MISIA, and (l) Antarctic summer MISIA.

distribution) at 64°S in the winter to $76^{\circ}\text{--}64^{\circ}\text{S}$ range with mode at 70°S in summer.

To avoid such problems with SIED, we introduce here an alternative diagnostic that can be reckoned as intermediate between the TSIA and SIED: the meridionally integrated sea ice area (MISIA), which we define

as the cumulative sum of all grid cells/pixels of sea ice area in 1° sectors of longitude extending from a pole to the equator. Sampling all 360 longitude sectors, this quantity has a mostly unimodal frequency distribution (Figs. 2i–l) in contrast to the SIC, and also yields more information about the large-scale sea ice spatial distribution

compared to the SIEd. In comparisons between model-simulated and observed MISIA, the statistic can clearly reveal discrepancies between unrealistic open ocean regions poleward of the ice edge and is less sensitive than SIEd to small errors in sea ice concentration near Greenland. It is also relatively insensitive to observational dataset compared to SIC, perhaps indicating a more reliable reference for evaluation of models.

The MISIA distribution ranges are similar between the two hemispheres, but not unexpectedly the upper bound changes seasonally from $120 \times 10^3 \text{ km}^2$ in the winter to $50 \times 10^3 \text{ km}^2$ in the summer (Note the change of the scales in Figs. 2i–l). Although this measure can be used to gauge how well models capture the longitudinal structure of observed sea ice, it, like other measures, has some limitations. Most notably, at some longitudes the sea ice distribution is constrained naturally by the coastlines. Except for an exceptionally inferior model, temperatures throughout the Arctic in winter are cold enough to ensure sea ice formation all the way to the northern continental coasts at nearly all longitudes. This means that MISIA is only useful for evaluating sea ice in the wintertime Arctic along longitudes that lead to open ocean. On the other hand, these natural constraints apply to any spatial metric of sea ice and appear unavoidable. The point is that this limitation needs to be kept in mind when interpreting measures of sea ice cover.

The MISIA results are derived by first interpolating (bilinear) all observations and model data onto a $1^\circ \times 1^\circ$ grid. To ensure a consistent comparison of models and observations, we mask out, in all models and the two observational datasets, all locations where the SIC is “missing,” which includes land and the area poleward of 87.2°N (which is not seen by the satellites) in the reference NT dataset.

In the analysis that follows, we examine the climatological annual cycle as well as seasonal means, both of which are basic characteristics included in any routine evaluation of climate models. For both the models and observations, we create climatologies of annual mean and annual cycle based on 27 years of data from 1979 to 2005 (2005 is the last year of the CMIP5 historical simulations). Our evaluation, focusing on the overall annual cycle behavior of the spatial distribution of sea ice, does not include examination of more targeted features, such as, for example, the September ice extent for the Arctic, but the metrics developed here (section 3d) could readily be adapted to meet more specific needs.

3. Results

a. Seasonal SIC patterns

We begin our evaluation of the CMIP5 historical simulations by comparing the seasonal mean spatial

distributions of the simulated SIC to the observed SSM/I fields. We choose the NT dataset as our baseline reference because it has been most commonly used in previous studies, but we also include a comparison between the NT and BT data. In Fig. 3 seasonal (January–March/July–September) spatial maps are shown for subset of 14 (one representative for each modeling group of the 41 CMIP5 models used in this study). Results from the full set can be found in online supplemental material (Figs. S1 and S2); also see the appendix of Shu et al. (2015) for February/September CMIP5 monthly climatologies. The simulated patterns are in better agreement with observations in the winter season in both the Arctic and Antarctic (Figs. 3a,c), when the sea ice distribution is dominated by high ice concentrations ($>85\%$). Discrepancies are evident in both compact and marginal ice characteristics. There is better agreement in the Arctic than in the Antarctic, in part because the sea ice pack is more restricted by the northern continents. In contrast, the distribution of observed summer season ice is highly diversified in the Arctic and almost ice free in the Antarctic (Figs. 3b,d). Overall, the seasonal maps of the sea ice concentration reveal large differences between the modeled and observed distributions (especially in the summer and in Antarctic), and also between the models themselves.

b. Zonal structure

To further characterize the observed and simulated differences in the seasonal spatial distributions of sea ice, in Fig. 4 we examine the MISIA defined in section 2. The largest MISIA values in Arctic are found during the winter in the Sea of Okhotsk (at $\sim 150^\circ\text{E}$), in the Bering Sea (at $\sim 190^\circ\text{E}$) in the NP sector, in the Labrador Sea (at $\sim 300^\circ\text{E}$), and in the Greenland Sea (at $\sim 350^\circ\text{E}$) in the NA sector (Fig. 4a). Consistent with the spatial distributions discussed earlier (Fig. 3) the best MISIA agreement between models and observations appears to be for the winter Arctic (Fig. 4a) with largest disagreement in the Sea of Okhotsk and the Labrador and Greenland Seas. In the summer, when the NP and the NA sectors are almost ice free, the MISIA in Arctic can be interpreted as representative of the CA sector (Fig. 4b). We find large model spread between 150° and 250°E where the Beaufort Gyre is located. Note that this is the area where the two observational datasets (Fig. 4b; NT–black stars, BT–black triangles) have the largest disagreement as well. A fairly large group of models fails to represent the ice extent in the Barents Sea. In both seasons in Arctic the MMEM (Figs. 4a,b, black stars) seems to follow well the observations with better agreement with NT in summer.

In Antarctic (Figs. 4c,d) the largest MISIA is found all year long in the Weddell Sea (at $\sim 315^\circ\text{E}$), part of the SA

sector, and in the Ross Sea (150°–210°E) in the SP sector, which are the only areas significantly covered with ice in the summer (Fig. 3d). The sea ice is minimal in the rest of the Antarctic and confined close to the coast line. In general, the CMIP5 model ensemble spread is large in both seasons in Antarctic, but perhaps larger in summer (Figs. 4c,d). Besides the substantial model disagreement, here we also find the largest differences with the observations. In the winter, the majority of the models lack ice cover between 10°W and 30°E (Fig. 4c; see MEM), and in the summer, there is a common tendency in many models to develop excessive ice cover in the west Ross Sea and insufficient cover in the Weddell Sea (Fig. 4d; MEM).

To analyze contributions to the total error in MISIA attributable to a model's general tendency to form too little or too much sea ice and attributable to inaccurate placement of sea ice we partition the total errors into two orthogonal components,¹ namely 1) a zonal mean bias and 2) a “pattern” error accounting for differences in the longitudinal distribution of sea ice (once the zonal mean bias has been removed). The errors can conveniently be expressed as mean squared errors (MSEs), and the orthogonal components can be summed in quadrature to yield the total MSE. For most models, the errors in the Arctic winter ice distributions are primarily pattern based (Fig. 5a). The bias contribution is smaller in the Arctic winter because the ice is constrained by the continents and most models simulate temperatures cold enough to form ice as far south as the northern edge of the continents (even if those temperatures are inaccurate). In the summer, however, accurate simulation of temperature is more critical to getting the ice extent correct, which results in nonnegligible bias in many models. In the Antarctic, Fig. 5 suggests that the mean bias errors dominate in some models and are an important contributor to the total MSE of most models. The largest errors are found in the Antarctic winter fields (Fig. 5c). The MSE of the MISIA is an objective metric for quantifying the model errors shown in Fig. 3, which is statistically easier to understand than metrics based on the SIC maps, which have a bimodal frequency distribution (cf. Figs. 2a–e and 2i–l).

We revisit Fig. 2 to summarize the comparison of the three sea ice spatial measures considered here. The distributions of the MISIA, SIC, and SIED (Fig. 2) are clearly distinct, and at the same time in each case the MEM is reasonably consistent with one or both

observational datasets. The MISIA has a unimodal distribution compared to the bimodal SIC distribution (Figs. 2a,c), which can simplify its statistical characterization and at the same time capture more detail from the large-scale sea ice spatial distribution compared to SIED. Still, one factor that needs to be kept in mind is that statistics based on the 1D MISIA and SIED are probably more strongly constrained by continental boundaries than those derived from the 2D SIC.

To further explore use of these measures in testing models, we constructed seasonal Taylor diagrams (Taylor 2001) for the MISIA, SIC, and SIED (Fig. S3a) comparing the CMIP5 simulations with observations. We found substantial differences between the results; thus, while it may be desirable to identify a single metric that effectively captures how well models simulate the longitudinal structure of the observed sea ice climatology, more work is necessary to fully explain the characteristics of the scalars used here and in other studies. However, our analysis demonstrates several advantageous characteristics of the MISIA, and therefore recommends that it be added to the mix of measures that are currently used for model evaluation.

As a cumulative measure, MISIA is still subject to compensation errors but presumably to a lesser degree than TSIA. In the following section we offer a simple but more holistic error analysis approach that can uncover errors occurring throughout the annual cycle (rather than just individual seasons) and guard against spatially compensating errors.

c. Sector-scale analysis

1) ANNUAL CYCLE OF TOTAL SEA ICE AREA

While in the previous section we evaluated the performance of the CMIP5 models in terms of their spatial distribution during a particular season, here we consider the sea ice area seasonal evolution in each of sectors defined in section 2. Figure 6a depicts the climatological annual cycle of the TSIA, hemispherically in the Arctic, and regionally in the CA, NA, and NP sectors. Results are shown for the CMIP5 multimodel ensemble and the NT and BT observational dataset means (solid lines). The red shading is the intermodel standard deviation and the blue and the green shading shows the year-to-year standard deviation of NT and BT observations, respectively. The observed and simulated annual cycle of Arctic TSIA most closely resembles the NA Arctic sector (Fig. 6a). The observed TSIA annual cycle in the CA region has a nearly constant value of $6.5 \times 10^6 \text{ km}^2$ from December to May, decreasing to $4 \times 10^6 \text{ km}^2$ in August and September. The NP sector is ice free from July to October and attains its winter maximum of up to

¹ The orthogonal decomposition of a vector x in R^n is the sum of a vector in a subspace X of R^n and a vector in the orthogonal complement $X^\perp$. See Golub and Van Loan (1996).

FIG. 4. Seasonal hemispheric meridionally integrated sea ice area (MISIA) found in individual CMIP5 models (color lines), in the multimodel ensemble mean [MMEM (circles)] and observed [NASA Team (stars) and Bootstrap (triangles)]: (a) Arctic winter, (b) Arctic summer, (c) Antarctic winter, and (d) Antarctic summer. The vertical black lines mark region boundaries for the NP, NA, SA, SP, and IO sectors.

FIG. 5. Seasonal hemispheric mean-square-error (MSE) of the MISIA for individual CMIP5 models compared with NASA Team observations, resolved into contributions from the hemispheric mean bias (red) and from spatial deviations from the overall mean (blue): (a) Arctic winter, (b) Arctic summer, (c) Antarctic winter, and (d) Antarctic summer.

$1\text{--}1.5 \times 10^6 \text{ km}^2$ in March. Consistent with the results for the annual cycle of CMIP5 TSIE (Stroeve et al. 2012), we find the largest disagreement with the observations in the total Arctic ice area in the winter and spring (January–May) when the CMIP5 MME mean exceeds both observational estimates. Regionally the largest mean biases are found in the CA sector (Fig. 6a). The best agreement with the observations is in the NA sector, where the MME is closer to the BT estimate in

the winter and to the NT in the summer. In the NP sector the MME is overestimated in May and June. Despite the CMIP5 intermodel spread being substantial in some cases, it is worth noting that the MME tracks the observations quite well in generally capturing the distinct evolution of each sector.

The observed TSIA in the Antarctic (Table 1) has a minimum in February and maximum in September, with an annual amplitude of $\sim 12.5 \times 10^6 \text{ km}^2$ and closely

FIG. 6. Hemispheric and regional total sea ice area (TSIA) annual cycle comparison of the CMIP5 MMEM (red), NASA Team observations (blue), and Bootstrap observations (green): (a) all Arctic, (b) CA, (c) NA, (d) NP, (e) all Antarctic (f) IO, (g) SA, and (h) SP. The red shading is the intermodel ensemble standard deviation and the blue and the green shadings are the year-to-year std of NT and BT observations, respectively.

resembles the SP sector annual cycle (Fig. 6b). Compared to the Arctic results, the CMIP5 MMEM annual cycle of the total Antarctic SIA seems to be in better agreement with the observed (NT), although the model spread is larger. The MMEM TSIA is slightly underestimated from April through June. Regionally, the largest biases in the MMEM occur in the autumn (May–September) in the SA sector and in the winter (August–October) in the IO sector where the MMEM is beyond the year-to-year range of the observations. The greatest model spread is found in the SP sector, although here the MMEM is in better agreement with observations compared to the other Antarctic sectors.

2) QUANTIFYING MODEL ERRORS BY SECTOR

To gauge how well the individual models agree with the observed annual cycle estimates of TSIA for each of our sectors, we examine the regional MSE differences of

the CMIP5 models with the NT observations (Fig. 7). We again partition the MSE into two orthogonal components: 1) a time mean bias (red bars) and 2) departures from the annual mean errors (blue bars).

The largest MSE in the Arctic regions for a majority of the models is in the Central Arctic sector ($>0.01 \times 10^{12} \text{ km}^4$), and in many cases the annual mean bias is the dominant contributor to the total MSE. In Arctic regions (Figs. 7a–c), sizable biases persist in each sector suggesting a systematic problem (e.g., GFDL-ESM2G, BCC-CSM1.1), whereas in others they are confined to a particular sector (e.g., the NA sector for IPSL-CM5B-LR and CSIRO-Mk3.6.0, and the NP sector for CESM1-WACCM and HadGEM2-CC), which might be indicative of biases originating from local processes. Splitting the errors into contributions associated with the annual mean and annual cycle amplitude reveals the origin of the total errors. For example, with some models the mean

FIG. 7. Regional MSE of TSIA annual cycle of the individual CMIP5 models relative to NASA Team observations, resolved into contributions from the annual mean error and errors in the annual cycle departures from the mean: (a) CA, (b) NA, (c) NP, (d) IO, (e) SA, and (f) SP.

biases dominate, whereas in others discrepancies in the amplitude and phase of the annual cycle contribute substantially.

In North Atlantic and North Pacific sectors, there are a few outliers, whereas in the Antarctic there is a more continuous range of model errors, with some being smaller than the difference between the NT and BT estimates. There is a tendency for some of the models (e.g., BCC-CSM1.1, CSIRO-Mk3.6.0, GFDL-ESM2G) that are outliers in the Arctic to perform better in the Southern Hemisphere, whereas for others (e.g., MIROC5) the opposite behavior is apparent. Similar to the Arctic regions, in some cases (e.g., IPSL-CM5B-LR, MIROC5) the biases seen in the regional TSIA persist in all of the Antarctic sectors whereas in others they are localized (SP sector for BCC-CSM1.1-m, SA sector for all GFDL models and HadCM3; IO sector for all GISS models) (Figs. 7d–f).

d. Exposing compensating errors in hemispheric totals

As discussed in the introduction, one motivation for our study is to examine the potential impacts of compensating errors in the quantification of model biases in total ice area. As a first step beyond the traditional measure (of global integrals), we consider the total hemispheric sea ice area error as the sum of errors of each contributing hemispheric sector. We do this by decomposing the total hemispheric error into a bias error in the hemispheric mean and a term that represents departures from the hemispheric mean in each sector. In addition, we resolve both the hemispheric and sector errors into annual mean and annual cycle (departures from the mean) components. In this analysis, we use the SIC rather than the SIA in order to eliminate the discrepancies due to the different grid types and resolutions in the model sector areas. We construct an orthogonal decomposition of the model SIC (f) as sum of four components:

$$f = \bar{f}^{(0)} + f^{(0)*} + \bar{f}^{(1)} + f^{(1)*},$$

where $\bar{f}$ is the annual mean component, f^* a temporal deviation from the annual mean, $f^{(0)}$ the “global” spatial mean, and $f^{(1)}$ a sector deviation from the “global” annual mean/cycle.

Similarly, we can decompose the observational SIC (g):

$$g = \bar{g}^{(0)} + g^{(0)*} + \bar{g}^{(1)} + g^{(1)*}.$$

Furthermore we can resolve the total global SIC error into four components applying the decomposition of f and g above:

$$E^2 = \bar{E}_0^2 + E_0^{*2} + \bar{E}_1^2 + E_1^{*2},$$

where the error components are defined as

$$\bar{E}_0^2 = [\bar{f}^{(0)} - \bar{g}^{(0)}]^2 \text{ (global annual mean error),}$$

$$E_0^{*2} = (1/12) \sum_m [f_m^{(0)*} - g_m^{(0)*}]^2 \text{ (global annual cycle error),}$$

$$\bar{E}_1^2 = \sum_j [\bar{f}_j^{(1)} - \bar{g}_j^{(1)}]^2 \hat{w}_j / \sum_j \hat{w}_j \text{ (sector annual mean error),}$$

and

$$E_1^{*2} = (1/12) \sum_m \sum_j [f_{jm}^{(1)*} - g_{jm}^{(1)*}]^2 \hat{w}_j / \sum_j \hat{w}_j \text{ (sector annual cycle error),}$$

where m is the monthly index ($m = 1:12$), j is the sector index (in our case varies from 1:3 for each of the hemispheres), and $\hat{w}_j$ is the geographical area of sector j .

This error decomposition methodology can be expanded to include components on the grid cell level (see the [appendix](#)).

The results from this TSIA error decomposition for each of the CMIP5 models are shown on [Fig. 8](#) for the Arctic and Antarctic. The sum of the global error amounts (global annual mean error in red and global annual cycle error in blue) is related to the commonly discussed error of the total Arctic/Antarctic sea ice extent (e.g., [Stroeve et al. 2012](#); [Shu et al. 2015](#)), while the sector components (sector deviation from the global mean in yellow and sector deviation from the global annual cycle in green) has not been considered until now. The two sector components summed in quadrature is a measure of compensating errors that are not evident in the hemispheric total result. We can compare the MSE error of hemispheric mean quantities ($E_0^2 + E_0^{*2}$) with the total MSE (E^2) to evaluate the importance of this error compensation. This is illustrated in the scatterplots of [Figs. 9a and 9b](#) for Arctic and Antarctic correspondingly. Since total MSE is always greater than the error in hemispheric means, all results lie above the unity line. If there were no error compensation caused by an unrealistic distribution of sea ice among the sectors, then the models would lie on the diagonal. The models closest to unity have less error compensation, but can still have large errors due to hemispheric mean discrepancies. Overall the error magnitude is smaller in the Arctic where all but one model have a total MSE $< 25 \times 10^{12} \text{ km}^4$ while in the Antarctic about a quarter of the models have a larger error ([Figs. 8 and 9](#)). This might be related to the fact that the Antarctic seasonal amplitude is twice as large than in the Arctic ([Table 1](#)).

To examine the impact of error compensation on a ranking of the models, for both error measures in [Figs. 9a and 9b](#) we sort the results by the error magnitude ([Figs. 9c,d](#)). Contrasting our sector scale measure with the traditional global analysis demonstrates that there

FIG. 8. Hemispheric MSE of TSIA for individual CMIP5 models relative to NASA Team observations, resolved into error contributions from hemispheric mean TSIA, annual cycle departures from the hemispheric mean, sector time-mean deviations from the hemispheric mean, and residual sector annual cycle errors (the last bar is the mean-square difference between the NASA Team and Bootstrap observations): (a) Arctic and (b) Antarctic.

can be considerable shifts in the ranking of CMIP5 models, particularly in the Arctic. For some models such as MIROC-ESM-CHEM and all MPI models, relatively large compensating errors in the sectorial distribution of SIC in the Arctic are hidden in a ranking based on hemispheric means (Fig. 9c). Less error compensation is found in the Antarctic (Figs. 9b,d), where the largest part of the errors are due to the error in the global mean (Fig. 8b). It is also evident that models that perform relatively well in the Arctic generally do not perform well in the Antarctic and vice versa. Only two of the 41 CMIP5 models are ranked among the top 10 both in the Arctic and the Antarctic—MIROC4h and ACCESS1.0. An interesting aspect of our results is that the models with largest magnitude of the MSE errors (Figs. 8a and 9a) have the least error compensation (e.g., GFDL-ESM2G, BCC-CSM1.1, IPSL-CM5B-LR, MIROC5) while others with a relatively small MSE feature large error compensation (MPI-ESM-P, MIROC-ESM-CHEM). One explanation for this is that the relatively small errors in hemispheric quantities may be due at least in part to careful tuning of models to reproduce the observations. As highlighted by Mauritsen et al. (2012), during model tuning, model parameters are adjusted to reduce the model biases in the simulated climate, which can result in error compensation without necessarily improving the

physical representation. For example, the sea ice volume (product of the sea ice area and thickness) is sometimes used by the modeling groups (Mauritsen et al. 2012; Gent et al. 2011) to tune via adjusting parameters such as the sea ice albedo (Gent et al. 2011) or a nondimensional freezing parameter (Mauritsen et al. 2012). Further analysis of the impacts of model tuning is outside the scope of the current paper; however, we point out that in gauging model fidelity, an error measure that penalizes compensating errors in the spatial distribution of sea ice is less likely to be reduced through simple tuning approaches.

4. Summary and conclusions

We have analyzed the annual cycle of sea ice distribution as simulated in the CMIP5 historical simulations. Motivated by a need to move beyond the traditional evaluation of hemispherically integrated Arctic and Antarctic sea ice area or extent, we have explored two approaches for gauging how well modeled sea ice distribution agrees with observations. Our work has involved methodical consideration of how best to capture key aspects of “regional” behavior in large-scale tests of model agreement with observations. We propose a new method (the second approach discussed below) for objectively quantifying the large-scale agreement between

FIG. 9. (top) Mean-square errors and (bottom) model ranking in the (left) Arctic and (right) Antarctic based on the annual cycle of TSIA for the hemisphere (abscissa) and based on the total error (ordinate), which includes a contribution attributable to errors in the distribution of TSIA among the polar sectors.

observed and simulated sea ice distribution that includes some large-scale spatial information.

The two approaches described in section 3, while complementary, serve different purposes. The first yields quantitative information about the zonal structure of sea ice distribution by examining the meridionally integrated sea ice area as a function of longitude. This provides a challenging test of the current generation of models, and, as model resolution and our understanding of sea ice behavior increase, may be useful in documenting improved agreement with observations. As we moved beyond the simple global integrals to include structural information, we found that conclusions regarding relative model skill were sensitive to the measure used to quantify errors. Thus, while it is desirable to identify a metric that effectively captures how well models simulate the longitudinal structure of the observed sea ice climatology, we believe it is premature

to advocate the use of a single statistic for gauging how well models simulate finescale structures of sea ice distribution. More work is necessary to fully explain the characteristics of these scalar quantities derived from well-established measures of sea ice distribution.

With our second method, we attempt to identify an informative balance between hemispheric integrals and error measures that attempt to gauge the detailed variations of sea ice extent. To do so, we add some spatial detail by first calculating the total sea ice area of individual sectors. Using established methods to decompose errors into orthogonal components, our “sector-scale” approach enables an evaluation of the role of compensating errors in the hemispheric total sea ice area. Indeed, we demonstrate that compensating errors in the CMIP5 models are substantial, and if not properly accounted for can lead to a misrepresentation of how models are compared to observations and to one another. We have presented several

variants of our sector-scale analysis, including a space–time metric that gauges the departures from the annual and global mean as well as departures about both quantities. As a less complex case, we have also illustrated metrics for a seasonal mean, which emphasize the sector mean departures about the hemispheric total. This spatial sector-scale error measure could easily be applied for other applications (e.g., to gauge the September minimum sea ice extent). We believe these sector-scale metrics are well suited for quantifying model agreement with observations, and thus we advocate their use in conjunction with traditional hemispheric integral approaches.

As new model “weighting” strategies are explored that might improve upon the traditional “one model one vote” approach to presenting multimodel projections of sea ice change (Wang and Overland 2012; Massonnet et al. 2012), reliable performance metrics will be needed. Already in the most recent IPCC assessment, unequal weighting was used for the first time in plotting the multimean model estimate of projected decrease of Arctic sea ice extent (Collins et al. 2013). Ultimately, metrics that target key processes may be best suited for weighting model projections, and for some purposes the additional detail offered by our sector-scale analysis may not be necessary. However, we contend that our sector-scale metrics can only help to advance a more objective and comprehensive approach to the evaluation of how well modeled sea ice is compared with observations.

Acknowledgments. This work was supported by the U.S. Department of Energy’s (DOE’s) Office of Science (Biological and Environmental Research) through its Regional and Global Climate Modeling Program and was performed at Lawrence Livermore National Laboratory as a contribution to the U.S. Department of Energy, Office of Science, Climate and Environmental Sciences Division, Regional and Global Climate Modeling Program under Contract DE-AC52-07NA27344. The research was partly supported by the Centre for Climate Dynamics at the Bjerknes Centre and the Norwegian Research School on Climate Dynamics. We acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 2 of this paper) for producing and making available their model output. For CMIP, the U.S. DOE’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We also thank the anonymous reviewers and the editor for their constructive comments, which helped to substantially improve the quality of the paper.

APPENDIX

Methods: Error Decomposition

The goal is to determine contributions to total sea ice concentration error from

- 1) “global” mean error (in our case the *hemispheric* mean error for either the Arctic or Antarctic),
- 2) sector error (error in distribution of ice among sectors) and/or
- 3) grid cell error (error in the distribution of ice within each sector).

In addition we want to know contributions from

- 1) errors in the annual mean (denoted with an overbar) and
- 2) errors due to annual cycle deviations from the annual mean (denoted by an asterisk).

We can decompose orthogonally (Golub and Van Loan 1996) the sea ice concentration into six components:

$$f = \bar{f}^{(0)} + f^{(0)*} + \bar{f}^{(1)} + f^{(1)*} + \bar{f}^{(2)} + f^{(2)*},$$

where

$\bar{f}$ is the annual mean,

f^* is the deviation from the annual mean,

$f^{(0)}$ is the “global” mean,

$f^{(1)}$ is the sector deviation from the “global” mean, and

$f^{(2)}$ is the grid cell deviation from sector mean.

Furthermore we can resolve the total error into six components analogous to the decomposition of f :

$$E^2 = \bar{E}_0^2 + E_0^{*2} + \bar{E}_1^2 + E_1^{*2} + \bar{E}_2^2 + E_2^{*2}.$$

Considering one hemispheric field, let f_{ijm} be the simulated field and g_{ijm} the observed, where m , j , and i are indexes for month of the year, sector, and grid cell, respectively. Now we define the following:

w_{ij} is the area of an ocean grid cell,

$\hat{w}_j = \sum_{i=1}^{N(j)} w_{ij}$ is the ocean area of sector j , where $N(j)$ is the number of grid cells in sector j , and

$\hat{f}_{jm} = \sum_{i=1}^{N(j)} (f_{ijm} w_{ij} / \hat{w}_j)$ is the mean sea ice concentration in sector j .

Then we can calculate the components of f as follows:

$\bar{f}^{(0)} = [(1/12) \sum_m \sum_j \hat{f}_{jm} \hat{w}_j] / \sum_j \hat{w}_j$ (global annual mean),

$f_m^{(0)*} = (\sum_j \hat{f}_{jm} \hat{w}_j / \sum_j \hat{w}_j) - \bar{f}^{(0)}$ (global mean annual cycle deviation from the annual mean),

$\bar{f}_j^{(1)} = (1/12) \sum_m \hat{f}_{jm} - \bar{f}^{(0)}$ (sector annual mean deviation from global mean)

$f_{jm}^{(1)*} = \hat{f}_{jm} - \bar{f}^{(0)} - f_m^{(0)*} - \bar{f}_j^{(1)}$ (sector annual cycle deviation from global mean annual cycle deviation),
 $\bar{f}_{ij}^{(2)} = (1/12)\sum_m f_{ijm} - \bar{f}_j^{(1)} - \bar{f}^{(0)}$ (grid cell annual mean deviation from sector mean), and
 $f_{ijm}^{(2)*} = f_{ijm} - \hat{f}_{jm}$ (grid cell annual cycle deviation from the sector mean).

The error components then are defined as follows:

$\bar{E}_0^2 = [\bar{f}^{(0)} - \bar{g}^{(0)}]^2$ (global annual mean error),
 $E_0^{*2} = (1/12)\sum_m [f_m^{(0)*} - g_m^{(0)*}]^2$ (global annual cycle error),
 $\bar{E}_1^2 = \sum_j [\bar{f}_j^{(1)} - \bar{g}_j^{(1)}]^2 \hat{w}_j / \sum_j \hat{w}_j$ (sector annual mean error),
 $E_1^{*2} = (1/12)\sum_m \sum_j [f_{jm}^{(1)*} - g_{jm}^{(1)*}]^2 \hat{w}_j / \sum_j \hat{w}_j$ (sector annual cycle error),
 $\bar{E}_2^2 = \sum_j \sum_i [\bar{f}_{ij}^{(2)} - \bar{g}_{ij}^{(2)}]^2 w_{ij} / \sum_j \hat{w}_j$ (grid cell annual mean error), and
 $E_2^{*2} = (1/12)\sum_m \sum_j \sum_i [f_{ijm}^{(2)*} - g_{ijm}^{(2)*}]^2 w_{ij} / \sum_j \hat{w}_j$ (grid cell annual cycle error),

where the weights are observation based.

REFERENCES

- Andersen, S., R. Tonboe, L. Kaleschke, G. Heygster, and L. T. Pedersen, 2007: Intercomparison of passive microwave sea ice concentration retrievals over the high-concentration Arctic sea ice. *J. Geophys. Res.*, **112**, C08004, doi:10.1029/2006JC003543.
- Bao, Q., and Coauthors, 2013: The Flexible Global Ocean–Atmosphere–Land System model, spectral version 2: FGOALS-s2. *Adv. Atmos. Sci.*, **30**, 561–576, doi:10.1007/s00376-012-2113-9.
- Behrens, L. K., T. Martin, V. A. Semenov, and M. Latif, 2012: The Arctic sea ice in the CMIP3 climate model ensemble—Variability and anthropogenic change. *Cryosphere Discuss.*, **6**, 5317–5344, doi:10.5194/tcd-6-5317-2012.
- Bentsen, M., and Coauthors, 2013: The Norwegian Earth System Model, NorESM1-M—Part 1: Description and basic evaluation. *Geosci. Model Dev.*, **6**, 687–720, doi:10.5194/gmdd-5-2843-2012.
- Bernstein, E. R., C. A. Geiger, T. L. DeLiberty, and M. D. Lemcke-Stampone, 2015: Antarctic sea-ice thickness and volume estimates from ice charts between 1995 and 1998. *Ann. Glaciol.*, **56**, 383–393, doi:10.3189/2015AoG69A763.
- Bi, D., and Coauthors, 2012a: The ACCESS coupled model: Description, control climate and preliminary validation. *Aust. Meteor. Oceanogr. J.*, **63** (1), 41–64.
- , and Coauthors, 2012b: ACCESS-OM: The ocean and sea ice core of the ACCESS coupled model. *Aust. Meteor. Oceanogr. J.*, **63**, 213–232.
- Cavalieri, D. J., C. L. Parkinson, P. Gloersen, and H. J. Zwally, 1996 (updated yearly): Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave data, version 1. [1979–2010]. National Snow and Ice Data Center, doi:10.5067/8GO8LZQVLOVL.
- 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.
- Comiso, J., 2000 (updated 2015): Bootstrap sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS, version 2. [1979–2010]. National Snow and Ice Data Center. [Available online at http://nsidc.org/data/docs/daac/nsidc0079_bootstrap_seaice.gd.html.]
- , D. Cavalieri, C. Parkinson, and P. Gloersen, 1997: Passive microwave algorithms for sea ice concentrations: A comparison of two techniques. *Remote Sens. Environ.*, **60**, 357–384, doi:10.1016/S0034-4257(96)00220-9.
- Dufresne, J.-L., and Coauthors, 2013: Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. *Climate Dyn.*, **40**, 2123–2165, doi:10.1007/s00382-012-1636-1.
- Dunne, J. P., and Coauthors, 2012: GFDL's ESM2 global coupled climate-carbon Earth System Models. Part I: Physical formulation and baseline simulation characteristics. *J. Climate*, **25**, 6646–6665, doi:10.1175/JCLI-D-11-00560.1.
- Fetterer, F., K. Knowles, W. Meier, and M. Savoie, 2016 (updated daily): Sea ice index, version 2. National Snow and Ice Data Center, doi:10.7265/N5736NV7.
- Gent, P. R., and Coauthors, 2011: The Community Climate System Model version 4. *J. Climate*, **24**, 4973–4991, doi:10.1175/2011JCLI4083.1.
- Golub, G. H., and C. F. Van Loan, 1996: *Matrix Computations*. 3rd ed. Johns Hopkins Studies in the Mathematical Sciences, Vol. 50, Johns Hopkins University Press, 694 pp.
- Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.*, **16**, 147–168, doi:10.1007/s003820050010.
- Griffies, S. M., and Coauthors, 2011: The GFDL CM3 coupled climate model: Characteristics of the ocean and sea ice simulations. *J. Climate*, **24**, 3520–3544, doi:10.1175/2011JCLI3964.1.
- Hazeleger, W., and Coauthors, 2012: EC-Earth V2.2: Description and validation of a new seamless Earth system prediction model. *Climate Dyn.*, **39**, 2611–2629, doi:10.1007/s00382-011-1228-5.
- Iversen, T., and Coauthors, 2013: The Norwegian Earth System Model, NorESM1-M—Part 2: Climate response and scenario projections. *Geosci. Model Dev.*, **6**, 389–415, doi:10.5194/gmdd-5-2933-2012.
- Jeffrey, S. J., L. D. Rotsteyn, M. A. Collier, S. M. Dravitzki, C. Hamalainen, C. Moeseneder, K. K. Wong, and J. I. Syktus, 2013: Australia's CMIP5 submission using the CSIRO Mk3.6 model. *Aust. Meteor. Oceanogr. J.*, **63**, 1–13.
- Kattsov, V. M., V. Ryabinin, J. Overland, M. Serreze, M. Visbeck, J. Walsh, W. Meier, and X. Zhang, 2010: Arctic sea-ice change: A grand challenge of climate science. *J. Glaciol.*, **56**, 1115–1121.
- Komuro, Y., and Coauthors, 2012: Sea-ice in twentieth-century simulations by new MIROC coupled models: A comparison between models with high resolution and with ice thickness distribution. *J. Meteor. Soc. Japan*, **90A**, 213–232, doi:10.2151/jmsj.2012-A11.
- Markus, T., and D. J. Cavalieri, 2000: An enhancement of the NASA Team sea ice algorithm. *IEEE Trans. Geosci. Remote Sens.*, **38**, 1387–1398, doi:10.1109/36.843033.
- Martin, G. M., and Coauthors, 2011: The HadGEM2 family of Met Office Unified Model Climate configurations. *Geosci. Model Dev.*, **4**, 723–757, doi:10.5194/gmdd-4-765-2011.
- 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.

- Mauritsen, T., and Coauthors, 2012: Tuning the climate of a global model. *J. Adv. Model. Earth Syst.*, **4**, M00A01, doi:[10.1029/2012MS000154](https://doi.org/10.1029/2012MS000154).
- Meier, W. N., 2005: Comparison of passive microwave ice concentration algorithm retrievals with AVHRR imagery in Arctic peripheral seas. *IEEE Trans. Geosci. Remote Sens.*, **43**, 1324–1337, doi:[10.1109/TGRS.2005.846151](https://doi.org/10.1109/TGRS.2005.846151).
- , and D. Notz, 2010: A note on the accuracy and reliability of satellite-derived passive microwave estimates of sea-ice extent. CliC Arctic Sea Ice Working Group Consensus Doc., World Climate Research Programme, 2 pp.
- Notz, D., 2014: Sea-ice extent and its trend provide limited metric of model performance. *Cryosphere*, **8**, 229–243, doi:[10.5194/tc-8-229-2014](https://doi.org/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](https://doi.org/10.1098/rsta.2014.0164).
- , F. A. Haumann, H. Haak, J. H. Jungclauss, and J. Marotzke, 2013: Arctic sea-ice evolution as modeled by Max Planck Institute for meteorology's Earth system model. *J. Adv. Model. Earth Syst.*, **5**, 173–194, doi:[10.1002/jame.20016](https://doi.org/10.1002/jame.20016).
- Parkinson, C. L., K. Y. Vinnikov, and D. J. Cavalieri, 2006: Evaluation of the simulation of the annual cycle of Arctic and Antarctic sea ice coverages by 11 major global climate models. *J. Geophys. Res.*, **111**, C07012, doi:[10.1029/2005JC003408](https://doi.org/10.1029/2005JC003408).
- Pavlova, T., V. Kattsov, and V. Govorkova, 2011: Sea ice in CMIP5 models: Closer to reality? (in Russian). *Proc. MGO*, **564**, 7–18.
- Sakamoto, T., and Coauthors, 2012: MIROC4h—A new high-resolution atmosphere–ocean coupled general circulation model. *J. Meteor. Soc. Japan*, **90**, 325–359, doi:[10.2151/jmsj.2012-301](https://doi.org/10.2151/jmsj.2012-301).
- Santer, B. D., and Coauthors, 2011: Separating signal and noise in atmospheric temperature changes: The importance of timescale. *J. Geophys. Res.*, **116**, D22105, doi:[10.1029/2011JD016263](https://doi.org/10.1029/2011JD016263).
- Schmidt, G. A., and Coauthors, 2014: Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *J. Adv. Model. Earth Syst.*, **6**, 141–184, doi:[10.1002/2013MS000265](https://doi.org/10.1002/2013MS000265).
- Schneider, D., D. Shea, and C. Deser, 2012: Antarctic sea ice climatologies and trends in multiple data sets. *2012 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract C41B–0555.
- Semenov, V. A., T. Martin, L. K. Behrens, and M. Latif, 2015: Arctic sea ice area in CMIP3 and CMIP5 climate model ensembles—Variability and change. *Cryosphere Discuss.*, **9**, 1077–1131, doi:[10.5194/tcd-9-1077-2015](https://doi.org/10.5194/tcd-9-1077-2015).
- 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](https://doi.org/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](https://doi.org/10.1029/2012GL052676).
- Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192, doi:[10.1029/2000JD900719](https://doi.org/10.1029/2000JD900719).
- Voldoire, A., and Coauthors, 2013: The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dyn.*, **40**, 2091–2121, doi:[10.1007/s00382-011-1259-y](https://doi.org/10.1007/s00382-011-1259-y).
- Volodin, E., N. Dianskii, and A. Gusev, 2010: Simulating present-day climate with the IN-MCM4.0 coupled model of the atmospheric and oceanic general circulations. *Izv. Atmos. Ocean. Phys.*, **46**, 414–431, doi:[10.1134/S000143381004002X](https://doi.org/10.1134/S000143381004002X).
- Wang, M., and J. E. Overland, 2012: A sea ice free summer Arctic within 30 years: An update from CMIP5 models. *Geophys. Res. Lett.*, **39**, L18501, doi:[10.1029/2012GL052868](https://doi.org/10.1029/2012GL052868).
- Watanabe, S., and Coauthors, 2011: MIROC-ESM: Model description and basic results of CMIP5-20c3m experiments. *Geosci. Model Dev.*, **4**, 845–872, doi:[10.5194/gmdd-4-1063-2011](https://doi.org/10.5194/gmdd-4-1063-2011).
- Xin, X., T. Wu, J. Li, Z. Wang, W. Li, and F. Wu, 2013: How well does BCC_CSM1.1 reproduce the 20th century climate change over China? *Atmos. Ocean. Sci. Lett.*, **6**, 21–26.
- Yakovlev, N. G., 2009: Reproduction of the large-scale state of water and sea ice in the Arctic Ocean in 1948–2002: Part I. Numerical model. *Izv. Atmos. Oceanic Phys.*, **45**, 357–371.
- 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](https://doi.org/10.5194/tc-7-451-2013).