

Assessment of Arctic sea ice simulations in CMIP5 models using a synthetic skill scoring method

Liping Wu¹, Xiao-Yi Yang^{1*}, Jianyu Hu^{1,2}

¹ State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, China

² Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519000, China

Received 6 May 2019; accepted 11 June 2019

© Chinese Society for Oceanography and Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The Arctic sea ice cover has declined at an unprecedented pace since the late 20th century. As a result, the feedback of sea ice anomalies for atmospheric circulation has been increasingly evidenced. While climatic models almost consistently reproduced a decreasing trend of sea ice cover, the reported results show a large distribution. To evaluate the performance of models for simulating Arctic sea ice cover and its potential role in climate change, this study constructed a reasonable metric by synthesizing both linear trends and anomalies of sea ice. This study particularly focused on the Barents Sea and the Kara Sea, where sea ice anomalies have the highest potential to affect the atmosphere. The investigated models can be grouped into three categories according to their normalized skill scores. The strong contrast among the multi-model ensemble means of different groups demonstrates the robustness and rationality of this method. Potential factors that account for the different performances of climate models are further explored. The results show that model performance depends more on the ozone datasets that are prescribed by the model rather than on the chemical representation of ozone.

Key words: Arctic sea ice, climate model, Barents and Kara Seas, multi-model ensemble mean

Citation: Wu Liping, Yang Xiao-Yi, Hu Jianyu. 2019. Assessment of Arctic sea ice simulations in CMIP5 models using a synthetic skill scoring method. *Acta Oceanologica Sinica*, 38(9): 48–58, doi: 10.1007/s13131-019-1474-0

1 Introduction

In recent decades, the high latitudes of the Northern Hemisphere showed the most visible signal of climate change and surface warming, which is at least twice as severe as the global average (the Fifth Assessment Report of Intergovernmental Panel on Climate Change, i.e., IPCC AR5). This Arctic amplification effect has been attributed to several factors, including atmospheric and oceanic heat transportation and the solar radiation effect (Holland and Bitz, 2003; Alexeev et al., 2005; Graversen et al., 2008; Serreze et al., 2009; Screen and Simmonds, 2010a; Walsh, 2014). Among these, the retreating sea ice cover plays a central role (Serreze et al., 2009; Screen and Simmonds, 2010a; Chapman and Walsh, 2007).

Climate change in the Arctic is asymmetric and differs in different regions (Overland et al., 1997; Venegas and Mysak, 2000; Semenov and Bengtsson, 2003; Rogers et al., 2013). Several places such as the Barents Sea and the Kara Sea (BK) are important for Arctic climate change, which is likely due to the active exchange of heat and momentum at the ocean-atmosphere interface of these regions. Yang and Yuan (2014) pointed out that the weakening and collapse of the wintertime sea-ice double dipole mode can be largely attributed to the thermal effect of significant BK sea ice reduction during late fall and early winter. In addition to the local effect, several regional signals in the atmosphere-sea ice-ocean system can be amplified, which extend to affect the Arctic as a whole or even further to affect lower latitudes (Semenov and Bengtsson, 2003; Bengtsson et al., 2004; Semenov, 2008;

Semenov and Halford, 2009; Smedsrud et al., 2013). For example, anomalous low BK sea ice may impact the European climate through an atmospheric bridge, leading to unusually cold winters in Europe (Petoukhov and Semenov, 2010). In fact, the BK sea ice has been widely recognized as an important forcing factor for the mid-to-high latitude wintertime climate such as the position and strength of storm tracks, blocking systems, extreme events, and Arctic oscillation (Petoukhov and Semenov, 2010; Francis and Vavrus, 2012; Yang et al., 2016; Ruggieri et al., 2016). The distinction of BK sea ice in the Arctic climate system was further reported by Kim et al. (2016). Unlike other Arctic marginal seas, BK sea ice anomalies can extend throughout an entire year due to the so-called “insulation feedback” mechanism. Increased reception of solar radiation in summer and the heat stored in the BK seas can be released during the fall-winter season, resulting in “delayed warming” (Francis et al., 2009; Deser et al., 2010; Screen and Simmonds, 2010b). This may account for the high correlation between the winter BK sea ice extent (SIE) and the summer Arctic SIE both in model simulations and observations (Li et al., 2017). Therefore, the presented assessment of the model performance plausibly indicates the BK sea ice as casting the Arctic climate change.

With regard to the complex interactions between atmospheric circulation and Arctic sea ice, climate models that couple oceans, atmosphere, and ice were used to diagnose the factors that affect changes of the Arctic climate and predict future climate changes. The Coupled Model Intercomparison Project

Foundation item: The National Natural Science Foundation of China under contract Nos 41576178 and 41630963; the National Basic Research Program (973 program) of China under contract No. 2015CB954004.

*Corresponding author, E-mail: xyyang@xmu.edu.cn

(CMIP) of the World Climate Research Programme (WCRP) consists of a series of contemporary climate models. The IPCC fifth assessment report shows that the models that participate in CMIP5 (<http://pcmdi3.llnl.gov/esgset>) generally achieve higher performance than those in CMIP3 when reproducing long-term trends of sea ice. Rosenblum and Eisenman (2016) reported that the inclusion of volcanic activity, rather than the improvement of sea ice physics or model resolution, accounts for the priority of the CMIP5 over the CMIP3 models for simulating Arctic sea ice trends. Although the observed seasonal cycle and long-term trend of Arctic SIE is well presented by CMIP5 models, the dispersion of the projected SIE throughout the 21st century by CMIP5 models has been similar to that by CMIP3 models (Stroeve et al., 2012; Huang et al., 2017). Stroeve and Notz (2015) further reported that the uncertainty range of future sea-ice evolution simulation remains large for climate models. Massonnet et al. (2012) provided several important metrics that constrain the projections of summer Arctic sea ice. By reducing the inter-model spread of CMIP5 projections, Liu et al. (2013) reproduced consistent Arctic ice-free periods.

To address the problems of inter-model dispersion and uncertainty with regard to Arctic sea ice simulation, this study provides a reasonable assessment of sea ice simulation in CMIP5. In this study, 30 CMIP5 models were objectively and comprehensively evaluated for their capability to simulate Arctic sea ice variability. Given the strong feedback of BK seas to Arctic climate and its interconnection with mid-latitude climate, the BK region was differentiated from the other Arctic regions (exBK) and a larger weight was assigned to the BK region. Both the long-term trends and anomalies of sea ice were considered. In addition to the BK-priority weighting method, a comprehensive and objective assessment framework was constructed to quantify the models' ability to simulate sea ice. Based on this framework, several better models can be identified to constrain model biases and to provide a better basis for the study of Arctic ocean-ice-atmospheric interaction as well as future Arctic climate change predictions. Moreover, the model parameters were further scrutinized and a possible way to improve model performance for Arctic sea ice simulation is suggested.

2 Data and method

The CMIP5 model simulation dataset (Taylor et al., 2012) can be directly downloaded via the website <http://pcmdi3.llnl.gov/esgset/home.htm>. Among others, 30 models were selected due to their intactness and availability of the sea ice dataset, as shown in Table 1. HadISST1 sea ice concentration (SIC) data are applied in this study as observation data to evaluate the models (Rayner et al., 2003). The HadISST1 is a global monthly sea surface temperature (SST) and sea ice dataset with a $1^\circ \times 1^\circ$ grid that ranges from 1871 to the present, and is obtained from various sources including digitized sea ice charts and passive microwave retrieval. The dataset is made more homogeneous by compensating satellite microwave-based SICs for the impact of surface melt effects on retrievals in the Arctic and for algorithm deficiencies in the Antarctic, as well as by matching historical *in situ* concentrations with satellite data. Since *in situ* data prior to the satellite era are sparse and highly inhomogeneous, both the model and observational data were truncated to start from 1979.

The results of historical simulations were analyzed from 1979 to 2005, prolonged with Representative Concentration Pathways 8.5 (RCP8.5) simulations from 2006 to 2014. The exception is HadGEM2-CC and HadGEM2-ES, for which historical simulations from 1979 to 2004 were used, prolonged with RCP8.5 simu-

lations from 2005 to 2014. Given the inconsistency of various grids and projections in CMIP5 models, model outputs were pre-processed by interpolating them onto the same grid as the HadISST1 data.

Due to the significance of BK sea ice variability for recent climate change, a weighted scoring method was applied. The detailed processes of quantification are as follows: firstly, the fraction of the grid cell covered by SIC was multiplied by the area of grid cell to calculate sea ice area (SIA) for BK ($70.5^\circ\text{--}81.5^\circ\text{N}$, $19.5^\circ\text{--}100.5^\circ\text{E}$) and exBK regions, respectively (as shown in Eq. (1)):

$$\begin{aligned} \text{SIA}_{\text{BK}} &= \int_{70.5^\circ\text{N}}^{81.5^\circ\text{N}} \int_{19.5^\circ\text{E}}^{100.5^\circ\text{E}} 2\pi r^2 \left(\sin\left(\frac{\varphi+1}{180}\pi\right) - \sin\left(\frac{\varphi}{180}\pi\right) \right) \frac{\text{SIC}(\lambda, \varphi)}{360} d\lambda d\varphi, \\ \text{SIA}_{\text{exBK}} &= \int_{40.5^\circ\text{N}}^{89.5^\circ\text{N}} \int_{179.5^\circ\text{W}}^{179.5^\circ\text{E}} 2\pi r^2 \left(\sin\left(\frac{\varphi+1}{180}\pi\right) - \sin\left(\frac{\varphi}{180}\pi\right) \right) \frac{\text{SIC}(\lambda, \varphi)}{360} d\lambda d\varphi - \text{SIA}_{\text{BK}}, \end{aligned} \quad (1)$$

where the earth radius $r = 6\,731$ km.

Then, their linear trends were estimated using the least square method. Comparing the SIA trends of model outputs with observations, relative errors of the trends (Eq. (2)) were calculated. A lower absolute value of the relative error indicates a better performance of the models.

$$\text{Relative error} = \left| \frac{X_{\text{mod}} - X_{\text{obs}}}{X_{\text{obs}}} \right|, \quad (2)$$

where X_{mod} and X_{obs} represent the modeled and observational SIA trends, respectively.

Secondly, detrended SIC anomaly time series were obtained for each grid, with both climatology and linear trend being subtracted from the original data. A quantitative comparison between the model results and observations was conducted using the method developed by Willmott (1981) (Eq. (3)):

$$\text{Skill} = 1 - \frac{\sum_{i=1}^N |X_{\text{mod}} - X_{\text{obs}}|^2}{\sum_{i=1}^N (|X_{\text{mod}} - \bar{X}_{\text{obs}}| + |X_{\text{obs}} - \bar{X}_{\text{obs}}|)^2}, \quad (3)$$

where X represents the variable, \bar{X} represents its time mean, and the subscripts mod and obs represent model results and observations, respectively. Skill (Sk) values were calculated in each grid, and the final skill value was obtained after averaging. A higher skill value indicates better model performance. A detailed description of this method can be found in Warner et al. (2005).

To bridge the gap between relative error and skill score, the residual relative error (RRE) (one minus the absolute value of relative error) was used instead of the relative error itself. For exBK and BK regions, four values (RRE_{exBK} , Sk_{exBK} , RRE_{BK} and Sk_{BK}) were obtained. To synthetically evaluate the CMIP5 models with regard to sea ice simulation, these four components were normalized to obtain N-scores (i.e., $\text{N_RRE}_{\text{exBK}}$, $\text{N_Sk}_{\text{exBK}}$, N_RRE_{BK} and N_Sk_{BK}). Finally, a weighted method was used to quantitatively combine trend error (Eq. (2)) and anomaly skill (Eq. (3)). Since the BK sea ice may exert a far-reaching effect on the Arctic

Table 1. Basic characteristics of 30 CMIP5 models

Model Number	Model name	Institution	Country	Sea ice model	Spatial resolution	Sea ice component description	Ensemble members
1	ACCESS1-0	CSIRO-BOM	Australia	CICE, v4.1	tripolar, 1°×1°, refinement at the equator	energy-conserving thermo, ITD, EVP	1
2	ACCESS1-3	CSIRO-BOM	Australia	CICE, v4.1	tripolar, 1°×1°, refinement at the equator	energy-conserving thermo, ITD, EVP	1
3	BCC-CSM1-1-M	BCC	China	SIS	tripolar, ~1°×(1–1/3)°	Semtner 3-layer, ITD, EVP	1
4	BCC-CSM1-1	BCC	China	SIS	tripolar, ~1°×(1–1/3)°	Semtner 3-layer, ITD, EVP	1
5	CanESM2	CCCMA	Canada	CanSIM1	T63 Gaussian	cavitating fluid	5
6	CCSM4	NCAR	USA	CICE, v4	dipolar, 1.11°×(0.27–0.54)°, NP in Greenland	energy-conserving thermo, ITD, EVP	6
7	CESM1-CAM5	NSF-DOE-NCAR	USA	CICE, v4	dipolar, 1.11°×(0.27–0.54)°, NP in Greenland	energy-conserving thermo, ITD, EVP	3
8	CMCC-CM	CMCC	Italy	LIM2	ORCA-2° tripolar	Semtner 3-layer+brine pockets, Virtual ITD, VP	1
9	CNRM-CM5	CNRM-CERFACS	France	GELATO, v5	ORCA-1° tripolar	ITD, EVP	1
10	CSIRO-Mk3-6-0	CSIRO-QCCCE	Australia	Component of Mk3.6	T63 Gaussian	cavitating fluid	10
11	FGOALS-g2	IAP-THU	China	CICE, v4	~1°×1°	energy-conserving thermo, ITD, EVP	1
12	FGOALS-s2	LAGS-IAP	China	CSIM5	(0.5–1)°×(0.5–1)°	energy-conserving thermo, ITD, EVP	3
13	GFDL-CM3	NOAA GFDL	USA	SISp2	tripolar, ~1°×1°	modified Semtner 3-layer, ITD, EVP	3
14	GFDL-ESM2G	NOAA GFDL	USA	SISp2	tripolar, ~1°×1°	modified Semtner 3-layer, ITD, EVP	1
15	GFDL-ESM2M	NOAA GFDL	USA	SISp2	tripolar, ~1°×1°	modified Semtner 3-layer, ITD, EVP	1
16	GISS-E2-H p1	NASA GISS	USA	Russell	tripolar, ~1°×1°	modified Semtner 0-layer, VP	1
17	GISS-E2-R p1	NASA GISS	USA	Russell	1°×1.25°	modified Semtner 0-layer, VP	1
18	IPSL-CM5A-LR	IPSL	France	LIM2	ORCA-2° -tripolar	Semtner 3-layer+brine pockets, Virtual ITD, VP	4
19	IPSL-CM5A-MR	IPSL	France	LIM2	ORCA-2° -tripolar	Semtner 3-layer+brine pockets, Virtual ITD, VP	1
20	IPSL-CM5B-LR	IPSL	France	LIM2	ORCA-2° -tripolar	Semtner 3-layer+brine pockets, Virtual ITD, VP	1
21	MIROC-ESM-CHEM	MIROC	Japan	Component of COCO v3.4	~1.4°×1°	Semtner 0-layer, 2 ice categories, EVP	1
22	MIROC-ESM	MIROC	Japan	Component of COCO v3.4	~1.4°×1°	Semtner 0-layer, 2 ice categories, EVP	1
23	MIROC5	MIROC	Japan	Component of COCO v4.5	~1.4°×(0.5–1.4)°, Generalized spherical, shifted poles	Semtner 0-layer, ITD, EVP	3
24	MPI-ESM-LR	MPI-M	Germany	Component of MPI-OM	~1.5°×1.5°	Semtner 0-layer, Virtual ITD, VP	3
25	MPI-ESM-MR	MPI-M	Germany	Component of MPI-OM	~0.4°×0.4°	Semtner 0-layer, Virtual ITD, VP	1
26	MRI-CGCM3	MRI	Japan	MRI.COM3	1°×~0.49°	N/A	1
27	NorESM1-M	NCC	Norway	CICE, v4	dipolar, 1.11°×(0.27–0.54)°, NP in Greenland	energy-conserving thermo, ITD, EVP	1
28	CESM1-WACCM	NSF-DOE-NCAR	USA	CICE, v4	dipolar, 1.11°×(0.27–0.54)°, NP in Greenland	energy-conserving thermo, ITD, EVP	1
29	HadGEM2-CC	MOHC	UK	Inspired from CICE	(1–0.3)°×1°	Semtner 0-layer, ITD, EVP	2
30	HadGEM2-ES	MOHC	UK	Inspired from CICE	(1–0.3)°×1°	Semtner 0-layer, ITD, EVP	4

climate and a remote effect on the Northern Hemisphere extratropical atmospheric circulation, it was weighted more. Hence, the weight coefficients of four factors including sea ice trends of the exBK, sea ice trends of the BK, sea ice anomalies of the exBK, and sea ice anomalies of the BK were 0.1, 0.3, 0.2 and 0.4, respectively. Then, the final score can be calculated by adding these four weighted N-scores (F-score) and normalizing it

to get the NF-score) (Eq. (4)).

$$NF - score = \text{Normalized}(0.1 \times N_RRE_{exBK} + 0.2 \times N_Sk_{exBK} + 0.3 \times N_RRE_{BK} + 0.4 \times N_Sk_{BK}), \quad (4)$$

The final scores for 30 models are listed in Table 2. The weight coefficients are subjective, which qualitatively reflect the relative

Table 2. Model ranking and grouping according to the final weighted scores of models

Model No.	Model name	BK trend		exBK trend		BK anomaly		exBK anomaly		F-score	NF-score	Group
		RRE	N_RRE	RRE	N_RRE	Sk	N_Sk	Sk	N_Sk			
21	MIROC-ESM-CHEM	0.78	0.85	0.47	0.25	0.36	1.05	0.35	2.11	1.12	1.87	I
24	MPI-ESM-LR	0.59	0.11	0.66	0.63	0.40	1.68	0.29	0.77	0.92	1.53	
25	MPI-ESM-MR	0.84	1.04	0.59	0.49	0.37	1.20	0.27	0.29	0.90	1.49	
13	GFDL-CM3	0.49	-0.29	-0.58	-1.84	0.42	1.90	0.32	1.36	0.76	1.26	
27	NorESM1-M	0.63	0.24	0.77	0.86	0.34	0.84	0.29	0.65	0.62	1.04	
29	HadGEM2-CC	0.81	0.95	0.62	0.55	0.32	0.59	0.27	0.18	0.61	1.01	
9	CNRM-CM5	0.54	-0.10	0.59	0.49	0.36	1.05	0.28	0.37	0.51	0.85	
6	CCSM4	0.62	0.21	0.81	0.92	0.30	0.20	0.32	1.27	0.49	0.81	
19	IPSL-CM5A-MR	0.75	0.71	0.27	-0.14	0.32	0.47	0.26	-0.09	0.37	0.62	
7	CESM1-CAM5	0.34	-0.87	0.63	0.56	0.31	0.43	0.33	1.49	0.27	0.44	
12	FGOALS-s2	0.92	1.35	-0.01	-0.71	0.24	-0.55	0.29	0.61	0.24	0.40	II
30	HadGEM2-ES	0.66	0.36	0.04	-0.61	0.31	0.45	0.26	-0.09	0.21	0.35	
17	GISS-E2-Rp1	0.78	0.81	0.40	0.12	0.29	0.12	0.23	-0.72	0.16	0.26	
1	ACCESS1-0	0.32	-0.97	0.43	0.18	0.35	0.90	0.25	-0.37	0.01	0.02	
5	CanESM2	0.86	1.14	-0.08	-0.84	0.27	-0.17	0.22	-0.98	-0.01	-0.01	
14	GFDL-ESM2G	0.76	0.77	0.25	-0.19	0.26	-0.32	0.22	-0.90	-0.10	-0.16	
2	ACCESS1-3	0.31	-1.00	0.22	-0.25	0.31	0.37	0.26	-0.04	-0.19	-0.31	
23	MIROC5	0.79	0.86	0.71	0.73	0.18	-1.45	0.28	0.31	-0.19	-0.31	
3	BCC-CSM1-1-M	0.68	0.43	-0.13	-0.95	0.28	0.00	0.21	-1.31	-0.23	-0.38	
4	BCC-CSM1-1	0.77	0.79	-1.36	-3.40	0.28	-0.04	0.22	-0.91	-0.30	-0.50	
22	MIROC-ESM	-0.05	-2.41	0.72	0.75	0.27	-0.18	0.35	2.00	-0.32	-0.53	III
16	GISS-E2-Hp1	0.77	0.81	0.27	-0.14	0.23	-0.64	0.19	-1.72	-0.37	-0.62	
15	GFDL-ESM2M	0.07	-1.92	0.36	0.04	0.31	0.41	0.26	-0.01	-0.41	-0.68	
28	CESM1-WACCM	0.45	-0.45	0.80	0.91	0.21	-0.97	0.26	0.03	-0.42	-0.71	
18	IPSL-CM5A-LR	0.38	-0.72	0.09	-0.51	0.23	-0.72	0.28	0.38	-0.48	-0.79	
8	CMCC-CM	0.90	1.29	-0.45	-1.59	0.18	-1.44	0.23	-0.71	-0.49	-0.82	
26	MRI-CGCM3	0.30	-1.02	0.92	1.16	0.24	-0.50	0.23	-0.76	-0.54	-0.90	
20	IPSL-CM5B-LR	0.32	-0.97	0.75	0.82	0.23	-0.68	0.24	-0.58	-0.60	-0.99	
11	FGOALS-g2	0.43	-0.53	0.80	0.90	0.18	-1.37	0.19	-1.57	-0.93	-1.55	
10	CSIRO-Mk3-6-0	0.19	-1.45	0.76	0.83	0.09	-2.64	0.22	-1.07	-1.62	-2.69	

Note: RRE represents the residual relative error for BK and exBK trends, and Sk represents the skill for BK and exBK anomalies. The normalized RRE and Sk are labeled as N_RRE and N_Sk for each component. The F-score was obtained from the weighted integral of four components (N_RREs and N_Sks). This F-score value and its normalized version NF-score are listed as grouping criterion in the last two columns in descending order.

importance of various items rather than quantitatively. Therefore, the metric was chosen based on our understandings of the asymmetric sea ice variability within Arctic regions and the significance of BK sea ice variability for the Arctic climate change. The BK sea ice variability is most remarkable within the Arctic region both for interannual and multi-decadal time scales. Statistical analysis also showed that the BK sea ice variability is closely linked to other Arctic regions (data not shown).

3 Results

As mentioned above, a quantitative score of each model was obtained according to the synthetic skill scoring method. To group the models in a scientific and objective way, they have been sorted in descending order by their NF-scores (as shown in Table 2) and the root mean square (RMS) errors of the multi-model ensemble means (MMEMs) were computed with different numbers of model members (shown in Fig. 1). Three aspects of RMS were investigated: the SIA seasonal cycle (Fig. 1a), the standard deviation (Fig. 1b), and running trends (Fig. 1c). It is clear that almost all three lines of RMS values exhibited a consistent decreasing tendency when the ensemble members were less than 10. For ensemble members exceeding 10, the RMS lines

either level off or ascend slightly with the inclusion of further model members. In the light of this RMS analysis, the first 10 models were grouped into the first group (Group I) and the other 20 models were evenly divided into two groups based on their weighted scores as reference groups (Groups II and III). Then, several metrics were applied to assess the rationality and robustness of this scoring and grouping system.

3.1 Multi-model ensemble mean state

Figure 2 shows the climatological seasonal cycle of SIA for each group. The RMS errors of the MMEM were 0.46, 0.59 and 1.79×10^6 km² for Groups I, II and III, respectively. Apparently, the MMEM of Group I conformed more to observation than that of either Group II or Group III. Observed SIA reached a maximum in March and a minimum in September. Models in Group I consistently reproduced this feature. However, for Groups II and III, the dispersions among model members clearly increased. Although almost all models qualitatively simulated the seasonal waxing and waning of sea ice, the values of monthly climatology of SIA varied as large as 2×10^6 – 5×10^6 km² for Groups II and III. Even the MMEMs of these two groups clearly deviated from the observational data. The SIA was overestimated in wintertime and

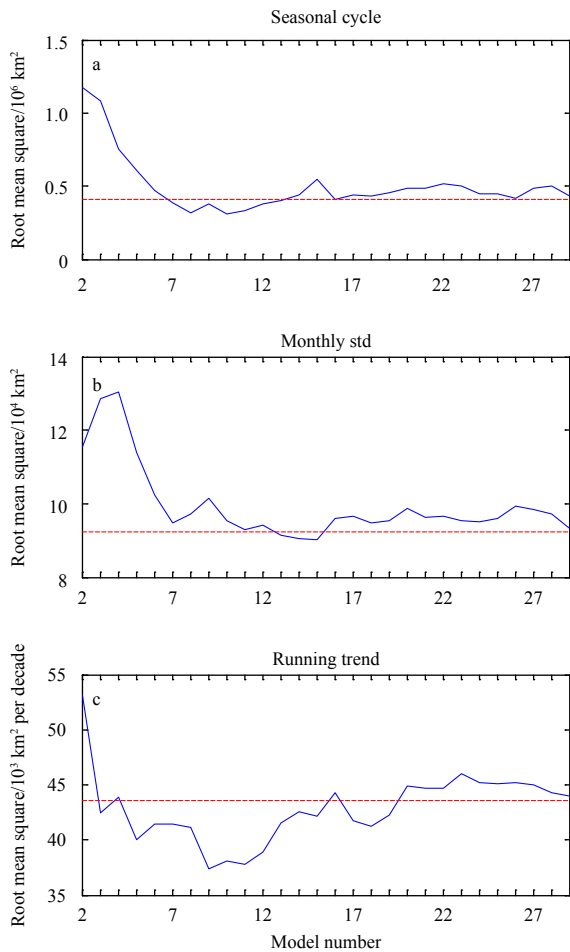


Fig. 1. The root mean square errors of various statistical components for the first n ($n=2-29$) model ensemble means. The model number is arranged in descending order according to their NF-scores. a. The SIA seasonal cycle, b. the SIA standard deviation, and c. the running decadal linear trend.

underestimated in summertime for Group II but was consistently overestimated throughout the year for Group III.

To further probe into the spatial distribution of SIA climatology, the SIA was calculated in each longitude and the climatolo-

gical meridional-integral SIA were presented in March and September for all three groups (Fig. 3). In March (Fig. 3a), MMEM SIA of all three groups were highly conform with observations in longitudes from 60°E to 140°E and 180°W to 122°W, including Kara, Laptev, Bering, and Beaufort Seas. For other longitudes, Group I behaved much better than the other two groups. Group II MMEM tends to overestimate the SIA in Greenland and Barents Seas, while Group III MMEM tends to underestimate (overestimate) the SIA in Hudson (Greenland and Barents Seas) and adjacent regions. The gaps among the three groups were wider in September (Fig. 3b) than in March. Group III models overestimate the SIA in most parts of the Arctic Ocean, with the poorest performance in the Barents and Kara Seas. The overall skills between the MMEM of each group and the observation also showed a large difference, with scores of 0.97, 0.93 and 0.70 for the groups in descending order. The superiority of Group I over both Group II or Group III was most striking for the BK region in March and September, which was predictable due to the larger weights given to the BK in the developed evaluation system.

The SIA trend was also estimated for each longitude (as shown in Fig. 4). All MMEMs of the SIA trend decreased in all investigated regions, which qualitatively agreed with the observational results except for the region near Greenland where the observed SIA increased. The MMEMs of the SIA trend of three groups were well distinguished in the BK. The MMEMs of Group I and II were much closer to the observational data than that of Group III. In other regions, differences between three MMEMs and observation were less significant. The skills between the MMEM of each group and observation were 0.79 (Group I), 0.67 (Group II), and 0.33 (Group III).

The skill score distributions of the MMEM detrended SIC anomalies for each group in both the exBK region and the BK region are displayed in Figs 5 and 6, respectively. In the BK, the skill scores of Group I were much higher than that of either Group II or Group III (Fig. 6). With regard to the exBK region, the differences of the three groups were asymmetric and less significant (Fig. 5). The most rigid hierarchy was found in the Beaufort Sea and the Laptev Sea among the three groups. The mean values of spatial MMEM of skill scores for Groups I, II and III were 0.30, 0.24 and 0.24 in the exBK region and 0.35, 0.28 and 0.22 in the BK region, showing a clear descending order in the sequence of group number both in the BK and exBK. It seems that the models that achieved a better simulation of sea ice in the BK region may generally acquire higher scores in other Arctic regions, too.

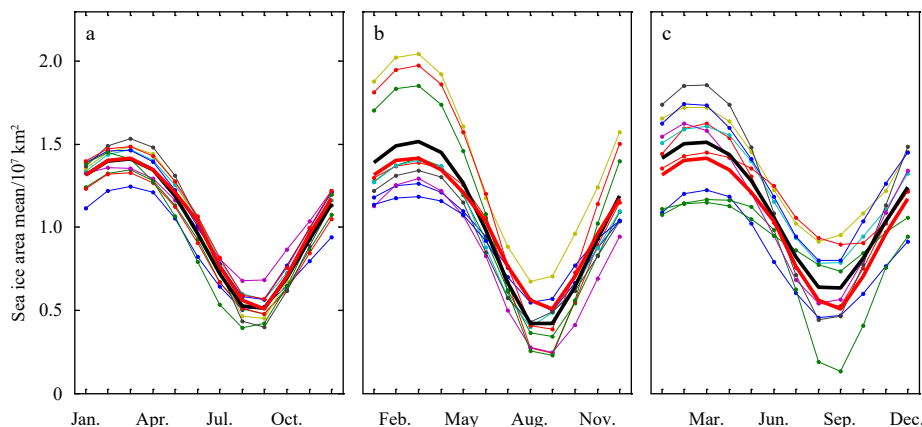


Fig. 2. Climatological seasonal cycles of the Arctic sea ice area during 1979–2014 period for models in Group I (a); Group II (b); and Group III (c). The black bold lines indicate the multi-model ensemble mean. The red bold lines denote observations.

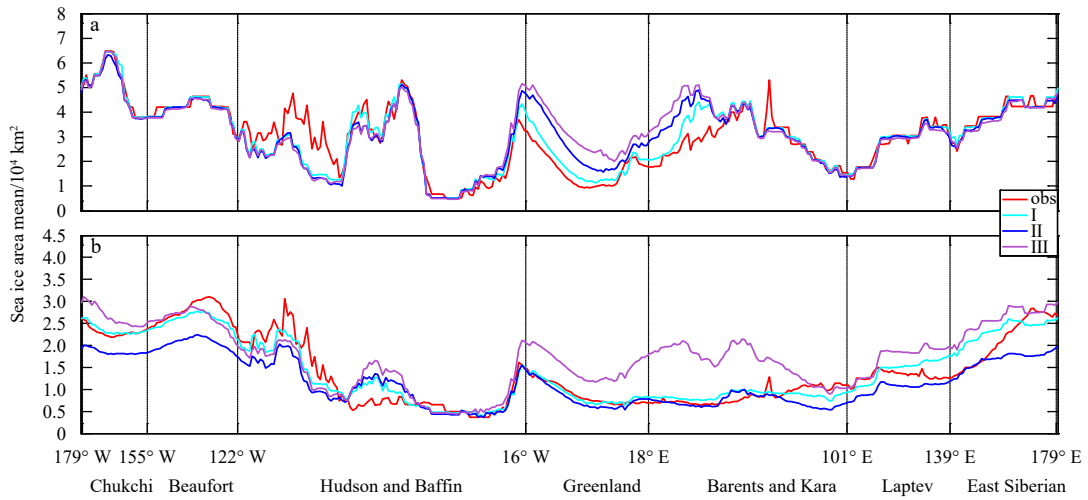


Fig. 3. Meridional-integral sea ice area distribution of three group in March (a) and September (b). The red line represents the observational data. The cyan, blue and mauve lines represent the MMEMs of Group I, Group II and Group III, respectively.

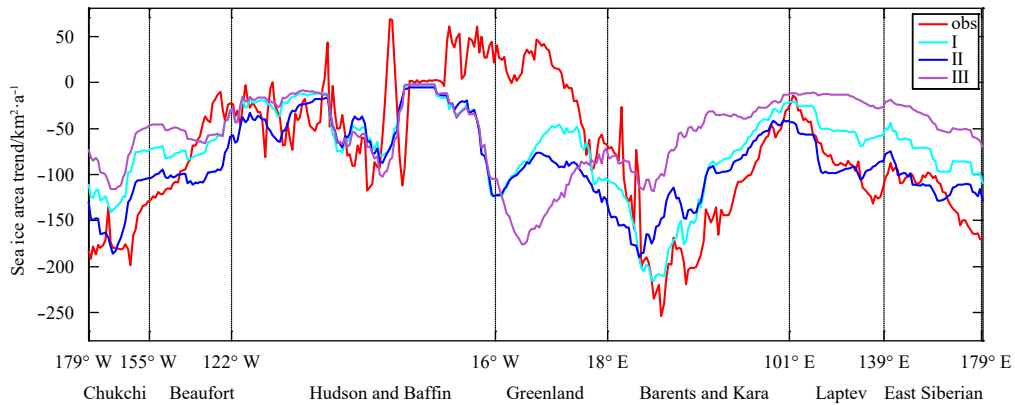


Fig. 4. Linear trends of the meridional sea ice area distribution. The red line represents to observational data. The cyan, blue and mauve lines represent the MMEMs of Group I, Group II and Group III, respectively.

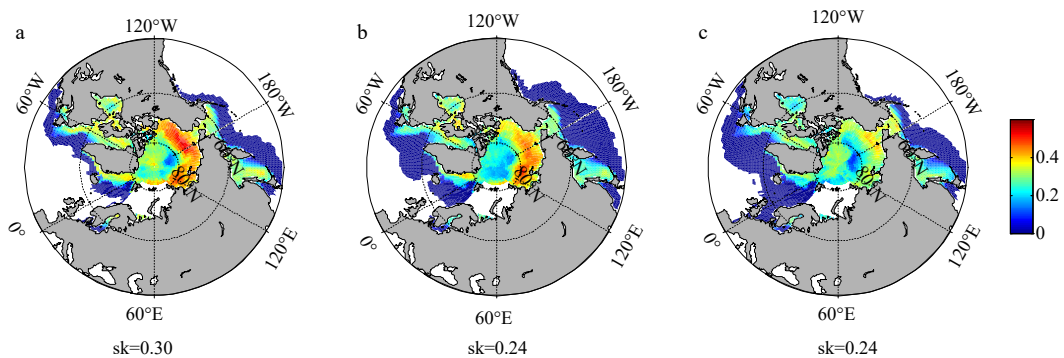


Fig. 5. Multi-model ensemble mean skill distribution for three groups in the Arctic exBK region (with the Barents and Kara Seas removed). a. Group I, b. Group II, and c. Grup III. "sk" refers to the mean skill value for the whole exBK region.

3.2 Individual model contrast

The remarkable contrast among the MMEMs in different groups indicates the applicability of the weighted method to distinguish the models' capability for sea ice simulation in general. However, the ensemble means blur the individual differences between group members. It is therefore necessary to further explore the detailed skill score model by model. Thus, a heatmap

was used to interpret the model differences in various terms of skill scores (Fig. 7). A heatmap is a graphical representation of data where individual values contained in a matrix are represented by colors. In a two-dimensional heatmap, the models were listed in descending order of their NF-scores and the model number is shown on the x-axis, and their scores of various sub-items on the y-axis. The colored squares show the normalized

score values for each model and each sub-item. It is apparent that the first 10 models of Group I generally achieved much higher scores for each sub-item than the models of Group III. The models of Group II exhibit a rather chaotic order in the sub-item scores, which, to some extent, reflects their large dispersion. The contrast between these groups is more pronounced for the BK anomaly than for the exBK trend, since different weights were assigned to these sub-items. It is worth noting that the score of the exBK trend for the 4th model (model No. 13 GFDL-CM3, Group I) was extraordinarily low (≤ -1.0), which stands in contrast to the extraordinary high score (≥ 1.0) of the 27th model (model No. 26 MIROC-ESMMRI-CGCM3, Group III).

To further analyze the result, a detailed comparison of sea ice simulation of both models with observation was conducted (Fig. 8). Figure 8a shows the annual mean SIA time series. Models 26 and 13 both exhibit similar downward trends, which are consistent with the observations. However, Model 13 overestimated the ice decline trend to some extent, particularly for the recent two decades. Model 26, although reproducing the observational sea ice decrease rate, dramatically underestimated the climatological SIA, and the linear correlation coefficient of Model 26 with the observation data was 0.56, which is much smaller than that of Model 13 (0.75). For the seasonal cycle (Fig. 8b), Model 13 was much closer to observation than Model 26. To compare both model simulation results in detail, the difference distributions of trends was presented between Models 13, 26, and observational data (Figs 8c and d). Model 13 underestimated the SIA trends almost everywhere. In contrast, the trends of Model 26 were overestimated in the Beaufort Sea and the East Siberian Sea and underestimated in the Baffin Bay and the Bering Sea. The mean linear trends averaged over the total Arctic region were $-29\,980\text{ km}^2/\text{a}$ for Model 13 and $343\text{ km}^2/\text{a}$ for Model 26. Therefore, the superiority of Model 26 over Model 13 for simulating the exBK trend may be attributed to the asymmetric trend skill distribution in different sub-regions and the cancelation between them. If the

absolute value of the skill score of the exBK trend in each grid is used, the difference of the mean skill distribution between the two models can be neglected.

3.3 Potential factors accounting for the performance of climate models

It seems that the applied weighting scores can well measure the capability of a model to simulate the sea ice. However, the reason for the dispersion of model simulations, particularly in Group II, remains unknown. The model parameters, the grid resolution, and the way that models represent stratospheric ozone have been proposed as potential factors that affect model performance in sea ice simulation (Turner et al., 2009; Sigmond and Fyfe, 2010; Zunz et al., 2013). To investigate the effect of ozone on the model performance, the ozone representation was listed and ozone datasets were prescribed for 30 CMIP5 models as reorganized by Eyring et al. (2013) in Table 3. According to the ozone representation, these 30 models can be roughly grouped into two categories. One contains 10 models with interactive or semi-off-line chemistry ozone representation (CHEM, bold model names in Table 3), and the other contains 19 models with prescribed ozone representation (NOCHEM, normal model names in Table 3). Most of the NOCHEM models apply the prescribed ozone both in the stratosphere and the troposphere. The exception is HadGEM2-ES, which uses prescribed ozone in the stratosphere but an interactive ozone chemistry in the troposphere. Thus, HadGEM2-ES was removed from the following statistics to avoid ambiguity. Unlike the models with prescribed ozone, semi-offline was denoted if the prescribed ozone dataset has been calculated with the underlying CMIP5 chemistry-climate model using prescribed SSTs and SICs following historical emissions as reported by Lamarque et al. (2010) and future emissions under the RCP scenarios as described by Lamarque et al. (2011). These differ from the class of prescribed ozone CMIP5 models (NOCHEM), because their stratospheric ozone evolution responds to changes

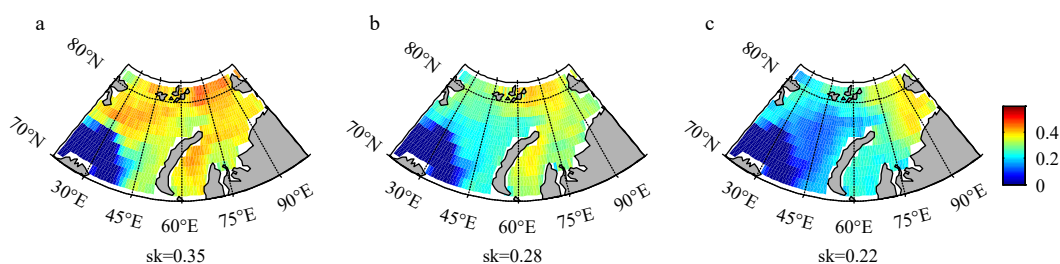


Fig. 6. Multi-model ensemble mean skill distribution for three groups in the Barents and the Kara Seas. a. Group I, b. Group II, and c. Group III. "sk" refers to the mean skill value for the whole exBK region.

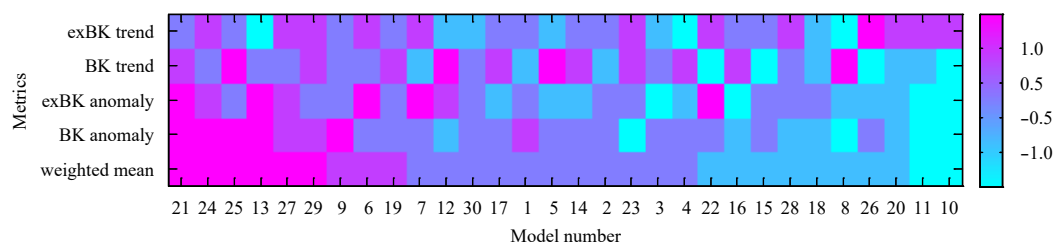


Fig. 7. Score heat map of four metrics and the weighted means of 30 models. One metric is the linear trend of the yearly sea ice area of the Arctic except for the Barents and Kara Seas (exBK), and another one is the anomaly of monthly detrended sea ice concentration of exBK. The other two metrics are the same as above but in the Barents and Kara seas. The magenta solid square represents a score > 1 . Purple represents a score between 0.5 and 1. Blue-purple represents a score between -0.5 and 0.5 . Steelblue represents a score between -1 and -0.5 . Cyan represents a score < -1 .

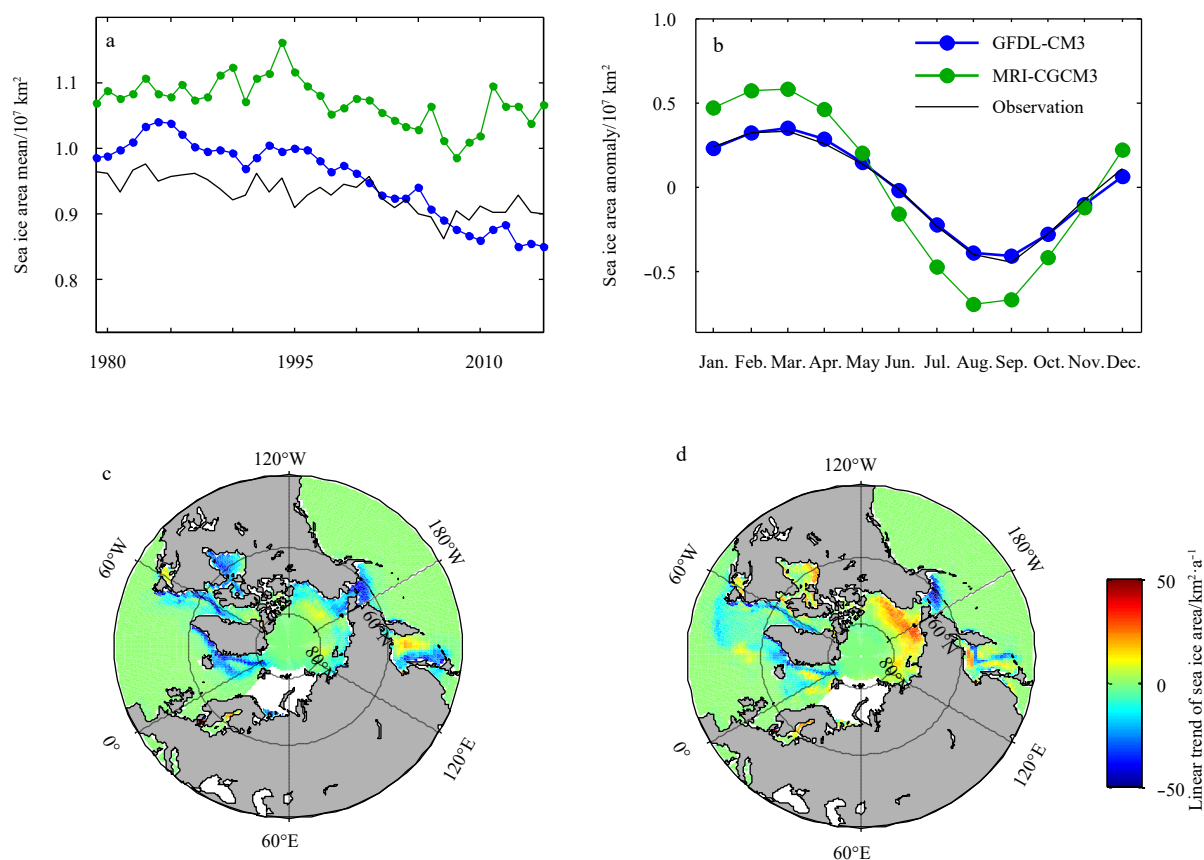


Fig. 8. Case study of the simulations of Model 13 GRDL-CM3 in Group I and Model 26 MRI-CGCM3 in Group III. a. Time series of the annual-mean sea ice area (SIA) of the two selected models and the observational data, b. the seasonal cycle of SIA anomalies after subtracting the long-term mean, and c and d. the linear trend distribution for the exBK region for Models 13 and 26, respectively.

in greenhouse gas concentrations in four RCPs, although it is still calculated offline (Eyring et al., 2013). The average scores of four metrics and their weighted mean were calculated and are displayed in Fig. 9a. For the weighted mean, the scores were -0.22 for NOCHEM models and 0.41 for CHEM models. The strong contrast of scores between both categories suggests the superiority of interactive or semi-offline ozone chemistry over the prescribed ozone representation. Comparing model score groups in Table 2 and ozone representations in Table 3 showed that 8 out of the last 10 models (the lowest score models) applied the prescribed ozone representation (NOCHEM). Nevertheless, the CHEM and NOCHEM models were well-matched in the top group (3 CHEM models vs. 3 NOCHEM models in Group I and 7 CHEM models vs. 3 NOCHEM models in top 10 models). The BK anomaly skill scores were consistent with the weighted mean scores. In contrast, the linear trend scores were not so sensitive to the choice of ozone representation. This suggests that the application of the interactive and semi-offline ozone chemistry may greatly improve the model performance of Arctic sea ice simulation within interannual to decadal time scales, but exert little effect on the linear trend simulation.

To further explore the effect of ozone on model performance, the models with prescribed ozone representation and semi-offline ozone chemistry were subdivided according to the ozone datasets they used. The applied analysis examined whether various ozone datasets affect the simulation of sea ice. The ozone datasets used by 18 NOCHEM models include C¹ (10 models), P⁵ (2 models), P⁷ (2 models), C_{modA}² (2 models), C_{modA}³ (1 model),

and C_{modB}⁴ (1 model). The latter two datasets were excluded because their sample sizes were too small. The ozone datasets used by semi-offline chemistry models include P² (3 models) and P⁶ (3 models). Furthermore, the scores of the interactive ozone chemistry models (4 models) were estimated in comparison to other models although they did not have any prescribed ozone data. The average scores of sub-group models with regard to four metrics (BK trend, exBK trend, BK anomaly and exBK anomaly) and their weighted mean scores were calculated and the results are shown in Fig. 9b. For weighted mean scores, the C_{modA}² models (score=1.2) were far superior to the other three categories in the NOCHEM group and were even better than the CHEM models. Moreover, the C_{modA}² models were superior to the other models and achieved high and balanced scores in all four metrics. In contrast to the P⁶ models, the P² models in the semi-offline group achieved the second-highest score of 0.74. Like interactive ozone models (represented by “1” in Table 3), the P² model performed well when simulating sea ice anomalies but unsatisfactorily when simulating linear trends. The model preference can be demonstrated by comparing the ozone data for the top 10 models (Group I) and the lowest 10 models (Group III). Both sets of models with the prescribed C_{modA}² ozone data and models with semi-offline P² ozone data ranked among top 10 models, while none of the models with these two ozone datasets ranked in the lowest 10 models. It seems that model performance depends more on the quality of the ozone dataset than on whether ozone is prescribed or interactive.

Although the models showed remarkable bias due to the

Table 3. Ozone chemistry of CMIP5 simulations

Model	O ₃ chemistry		Prescribed ozone dataset
	Troposphere	Stratosphere	
ACCESS1-0	P	P	C ¹
ACCESS1-3	P	P	C ¹
BCC-CSM1-1-M	P	P	C ¹
BCC-CSM1-1	P	P	C ¹
CanESM2	P	P	C _{modB} ⁴
CCSM4	SO	SO	p ²
CESM1-CAM5	SO	SO	p ²
CMCC-CM	P	P	C _{modA} ³
CNRM-CM5	I	I	—
CSIRO-Mk3-6-0	P	P	C ¹
FGOALS-g2	P	P	C ¹
FGOALS-s2	P	P	C ¹
GFDL-CM3	I	I	—
GFDL-ESM2G	P	P	C ¹
GFDL-ESM2M	P	P	C ¹
GISS-E2-H p1	P	P	p ⁵
GISS-E2-R p1	P	P	p ⁵
IPSL-CM5A-LR	SO	SO	p ⁶
IPSL-CM5A-MR	SO	SO	p ⁶
IPSL-CM5B-LR	SO	SO	p ⁶
MIROC-ESM-CHEM	I	I	—
MIROC-ESM	P	P	p ⁷
MIROC5	P	P	p ⁷
MPI-ESM-LR	P	P	C _{modA} ²
MPI-ESM-MR	P	P	N/A
MRI-CGCM3	P	P	C ¹
NorESM1-M	SO	SO	p ²
CESM1-WACCM	I	I	—
HadGEM2-CC	P	P	C _{modA} ²
HadGEM2-ES	I	P	—/ C _{modA} ²

Note: — means the models using the interactive ozone chemistry have no prescribed ozone dataset.

choice of ozone datasets, other factors may also influence model preference. Therefore, the possible impact of model resolution on the performance of models was also investigated. However, no evidence was found for the superiority of high-resolution models over low-resolution models (data not shown). The scores of high-resolution models can be either higher or lower than those with low resolution, and even models with the same resolution can show entirely different sea ice simulation ability (e.g.,

model No. 27 NorESM1-M and model No. 28 CESM1-WACCM). Nevertheless, it is difficult to conclude the role of spatial resolution in sea ice simulation due to the small sample size. Too many other parameters will likely affect model performance. Thus, it remains an open question which factors dominate the model capability to simulate Arctic sea ice.

4 Conclusion

In this study, the Arctic sea ice simulation of models that participated in CMIP5 was evaluated. Four metrics including the long-term trends and SIC anomalies in the interannual and decadal time scales in the BK as well as in the exBK were integrated with different weighted coefficients. Models were divided into three groups according to weighted scores. The feasibility and robustness of this assessment method was verified for various metrics. Finally, several parameters (mainly ozone representation) of the models were investigated to explain the discrepancies between models.

In general, the MMEM values of three groups were well distinguished in the annual cycle, linear trends, and interannual variability of the SIA, which demonstrates the rationality of the utilized evaluation criterion and the feasibility of constraining model uncertainty by selecting models in this way. With regard to individual models, a number of high-ranking models were superior to low-ranking models, not in all metrics, but in the core metrics such as BK trends and anomalies. This highlights the necessity of the weighted method in model assessment. Previous studies evaluated the model performance on sea ice simulation (Maslowski et al., 2012; Semenov et al., 2015; Shu et al., 2015). Climate models in general reproduce the sea ice retreating trend in the Arctic during the 20th century and simulate further sea ice area loss during the 21st century in response to anthropogenic forces. However, these models still suffer from large biases and model dispersions. Most previous studies thus applied the MMEM in order to abate the individual model bias and they mostly focused more on the sea ice tendency or long-term trend (Zhang and Walsh, 2006; Arzel et al., 2006). However, simulations vary from model to model and the sea ice variability is an important factor for atmospheric circulation. The presented assessment highlights the capability to simulate the BK sea ice trends and anomalies, due to their significance for driving the recent climate change both in the Arctic and in the mid-latitudes for each model. In addition, a preliminary discussion is presented regarding the reasons for model dispersion. Several parameters of models including resolution and ozone representation

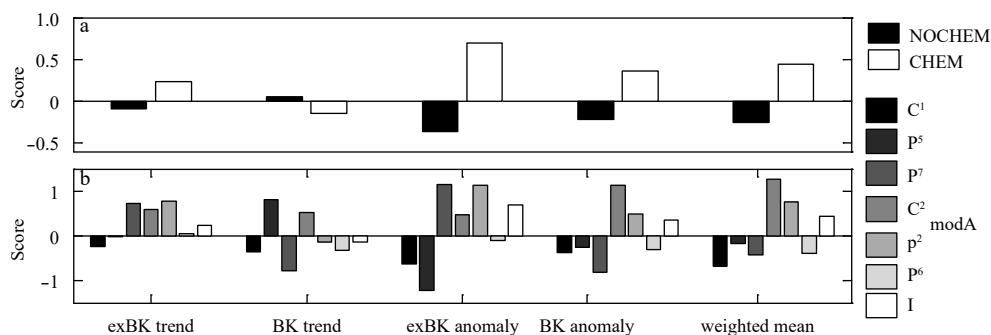


Fig. 9. Bar graph of four metrics (N_RREs and N_Sks) and their weighted mean scores (NF-scores). a. Scores of models with two different stratospheric ozone representations. Black bar denotes the score of models using the prescribed ozone representation. White bar denotes the score of models using the interactive and semi-offline chemistry ozone representation. b. Scores of models that use different ozone representation datasets.

were investigated. The obtained results show that the model resolution does not significantly impact model performance, which is in accordance with Rosenblum and Eisenman (2016). Instead, the ozone datasets that models used were found to be an important factor. For NOCHEM model developers this indicates that the use of C_{modA}^2 ozone data seems to be preferable.

Although this study highlights the importance of the BK sea ice simulation for this assessment system, it should be regarded only as a necessary but not a sufficient condition for climate models toward the prediction of future climate change. Further research is required to explore the detailed physical processes and mechanisms with which ice anomalies exert their influences on both local and remote climate variability.

Acknowledgements

We acknowledge the climate modelling groups, the World Climate Research Programme's (WCRP) Working Group on Coupled Modelling (WGCM), and the Met Office Hadley Centre's sea ice for the open datasets used in this study.

References

- Alexeev V A, Langen P L, Bates J R. 2005. Polar amplification of surface warming on an aquaplanet in "ghost forcing" experiments without sea ice feedbacks. *Climate Dynamics*, 24(7–8): 655–666, doi: [10.1007/s00382-005-0018-3](https://doi.org/10.1007/s00382-005-0018-3)
- Arzel O, Fichefet T, Goosse H. 2006. Sea ice evolution over the 20th and 21st centuries as simulated by current AOGCMs. *Ocean Modelling*, 12(3–4): 401–415, doi: [10.1016/j.ocemod.2005.08.002](https://doi.org/10.1016/j.ocemod.2005.08.002)
- Bengtsson L, Semenov V A, Johannessen O M. 2004. The early twentieth-century warming in the Arctic—a possible mechanism. *Journal of Climate*, 17(20): 4045–4057, doi: [10.1175/1520-0442\(2004\)017<4045:TETWIT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<4045:TETWIT>2.0.CO;2)
- Chapman W L, Walsh J E. 2007. Simulations of arctic temperature and pressure by global coupled models. *Journal of Climate*, 20(4): 609–632, doi: [10.1175/jcli4026.1](https://doi.org/10.1175/jcli4026.1)
- Deser C, Tomas R, Alexander M, et al. 2010. The seasonal atmospheric response to projected Arctic sea ice loss in the late twenty-first century. *Journal of Climate*, 23(2): 333–351, doi: [10.1175/2009JCLI3053.1](https://doi.org/10.1175/2009JCLI3053.1)
- Eyring V, Arblaster J M, Cionni I, et al. 2013. Long-term ozone changes and associated climate impacts in CMIP5 simulations. *Journal of Geophysical Research: Atmospheres*, 118(10): 5029–5060, doi: [10.1002/jgrd.50316](https://doi.org/10.1002/jgrd.50316)
- Francis J A, Chan Wei-han, Leathers D J, et al. 2009. Winter Northern Hemisphere weather patterns remember summer Arctic sea-ice extent. *Geophysical Research Letters*, 36(7): L07503, doi: [10.1029/2009gl013724](https://doi.org/10.1029/2009gl013724)
- Francis J A, Vavrus S J. 2012. Evidence linking Arctic amplification to extreme weather in mid-latitudes. *Geophysical Research Letters*, 39(6): L06801, doi: [10.1029/2012gl0151000](https://doi.org/10.1029/2012gl0151000)
- Graversen R G, Mauritsen T, Tjernström M, et al. 2008. Vertical structure of recent Arctic warming. *Nature*, 451(7174): 53–56, doi: [10.1038/nature06502](https://doi.org/10.1038/nature06502)
- Holland M M, Bitz C M. 2003. Polar amplification of climate change in coupled models. *Climate Dynamics*, 21(3–4): 221–232, doi: [10.1007/s00382-003-0332-6](https://doi.org/10.1007/s00382-003-0332-6)
- Huang Fei, Zhou Xiao, Wang Hong. 2017. Arctic sea ice in CMIP5 climate model projections and their seasonal variability. *Acta Oceanologica Sinica*, 36(8): 1–8, doi: [10.1007/s13131-017-1029-8](https://doi.org/10.1007/s13131-017-1029-8)
- Kim K Y, Hamlington B D, Na Hanna, et al. 2016. Mechanism of seasonal Arctic sea ice evolution and Arctic amplification. *The Cryosphere*, 10(5): 2191–2202, doi: [10.5194/tc-10-2191-2016](https://doi.org/10.5194/tc-10-2191-2016)
- Lamarque J F, Bond T C, Eyring V, et al. 2010. Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application. *Atmospheric Chemistry and Physics*, 10(15): 7017–7039, doi: [10.5194/acpd-10-4963-2010](https://doi.org/10.5194/acpd-10-4963-2010)
- Lamarque J F, Kyle G P, Meinshausen M, et al. 2011. Global and regional evolution of short-lived radiatively-active gases and aerosols in the Representative Concentration Pathways. *Climate Change*, 109: 191–212, doi: [10.1007/s10584-011-0155-0](https://doi.org/10.1007/s10584-011-0155-0)
- Li Dawei, Zhang Rong, Knutson T R. 2017. On the discrepancy between observed and CMIP5 multi-model simulated Barents Sea winter sea ice decline. *Nature Communications*, 8: 14991, doi: [10.1038/ncomms14991](https://doi.org/10.1038/ncomms14991)
- Liu Jiping, Song Mirong, Horton R M, et al. 2013. Reducing spread in climate model projections of a September ice-free Arctic. *Proceedings of the National Academy of Sciences of the United States of America*, 110(31): 12571–12576, doi: [10.1073/pnas.1219716110](https://doi.org/10.1073/pnas.1219716110)
- Maslowski W, Clement K J, Higgins M, et al. 2012. The future of Arctic sea ice. *Annual Review of Earth and Planetary Sciences*, 40: 625–654, doi: [10.1146/annurev-earth-042711-105345](https://doi.org/10.1146/annurev-earth-042711-105345)
- Massonnet F, Fichefet T, Goosse H, et al. 2012. Constraining projections of summer Arctic sea ice. *The Cryosphere*, 6(6): 1383–1394, doi: [10.5194/tc-6-1383-2012](https://doi.org/10.5194/tc-6-1383-2012)
- Overland J E, Adams J M, Bond N A. 1997. Regional variation of winter temperatures in the Arctic. *Journal of Climate*, 10(5): 821–837, doi: [10.1175/1520-0442\(1997\)010<0821:RVOWTI>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<0821:RVOWTI>2.0.CO;2)
- Petoukhov V, Semenov V A. 2010. A link between reduced Barents-Kara sea ice and cold winter extremes over northern continents. *Journal of Geophysical Research: Atmospheres*, 115(D21): D21111, doi: [10.1029/2009jd013568](https://doi.org/10.1029/2009jd013568)
- Rayner N A, Parker D E, Horton E B, et al. 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, 108(D14): 1063–1082, doi: [10.1029/2002JD002670](https://doi.org/10.1029/2002JD002670)
- Rogers K G, Goodbred S L Jr, Mondal D R. 2013. Monsoon sedimentation on the 'abandoned' tide-influenced Ganges-Brahmaputra delta plain. *Estuarine, Coastal and Shelf Science*, 131: 297–309, doi: [10.1016/j.ecss.2013.07.014](https://doi.org/10.1016/j.ecss.2013.07.014)
- Rosenblum E, Eisenman I. 2016. Faster Arctic sea ice retreat in CMIP5 than in CMIP3 due to volcanoes. *Journal of Climate*, 29(24): 9179–9188, doi: [10.1175/JCLI-D-16-0391.1](https://doi.org/10.1175/JCLI-D-16-0391.1)
- Ruggieri P, Buizza R, Visconti G. 2016. On the link between Barents-Kara sea ice variability and European blocking. *Journal of Geophysical Research: Atmospheres*, 121(10): 5664–5679, doi: [10.1002/2015jd024021](https://doi.org/10.1002/2015jd024021)
- Screen J A, Simmonds I. 2010a. The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature*, 464(7293): 1334–1337, doi: [10.1038/nature09051](https://doi.org/10.1038/nature09051)
- Screen J A, Simmonds I. 2010b. Increasing fall-winter energy loss from the Arctic Ocean and its role in Arctic temperature amplification. *Geophysical Research Letters*, 37(16): L16707, doi: [10.1029/2010GL044136](https://doi.org/10.1029/2010GL044136)
- Semenov M A. 2008. Simulation of extreme weather events by a stochastic weather generator. *Climate Research*, 35(3): 203–212, doi: [10.3354/cr00731](https://doi.org/10.3354/cr00731)
- Semenov V A, Bengtsson L. 2003. Modes of the wintertime Arctic temperature variability. *Geophysical Research Letters*, 30(15): 1787, doi: [10.1029/2003gl017112](https://doi.org/10.1029/2003gl017112)
- Semenov M A, Halford N G. 2009. Identifying target traits and molecular mechanisms for wheat breeding under a changing climate. *Journal of Experimental Botany*, 60(10): 2791–2804, doi: [10.1093/jxb/erp164](https://doi.org/10.1093/jxb/erp164)
- Semenov V A, Martin T, Behrens L K, et al. 2015. Arctic sea ice area in CMIP3 and CMIP5 climate model ensembles - variability and change. *The Cryosphere Discussions*, 9(1): 1077–1131, doi: [10.5194/tcd-9-1077-2015](https://doi.org/10.5194/tcd-9-1077-2015)
- Serreze M, Barrett A, Stroeve J, et al. 2009. The emergence of surface-based Arctic amplification. *The Cryosphere*, 3: 11–19, doi: [10.5194/tc-3-11-2009](https://doi.org/10.5194/tc-3-11-2009)
- Shu Qi, Song Zhenya, Qiao Fangli. 2015. Assessment of sea ice simulations in the CMIP5 models. *The Cryosphere*, 9(1): 399–409, doi: [10.5194/tc-9-399-2015](https://doi.org/10.5194/tc-9-399-2015)
- Sigmond M, Fyfe J C. 2010. Has the ozone hole contributed to in-

- creased Antarctic sea ice extent?. *Geophysical Research Letters*, 37(18): L18520, doi: [10.1029/2010gl044301](https://doi.org/10.1029/2010gl044301)
- Smedsrud L H, Esau I, Ingvaldsen R B, et al. 2013. The role of the Barents Sea in the Arctic climate system. *Review of Geophysics*, 51(3): 415–449, doi: [10.1002/rog.20017](https://doi.org/10.1002/rog.20017)
- Stroeve J C, Notz D. 2015. Insights on past and future sea-ice evolution from combining observations and models. *Global and Planetary Change*, 135: 119–132, doi: [10.1016/j.gloplacha.2015.10.011](https://doi.org/10.1016/j.gloplacha.2015.10.011)
- Stroeve J C, Serreze M C, Holland M M, et al. 2012. The Arctic's rapidly shrinking sea ice cover: a research synthesis. *Climatic Change*, 110(3–4): 1005–1027, doi: [10.1007/s10584-011-0101-1](https://doi.org/10.1007/s10584-011-0101-1)
- Taylor K E, Stouffer R J, Meehl G A. 2012. An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4): 485–498, doi: [10.1175/BAMS-D-11-00094.1](https://doi.org/10.1175/BAMS-D-11-00094.1)
- Turner J, Comiso J C, Marshall G J, et al. 2009. Non-annular atmospheric circulation change induced by stratospheric ozone depletion and its role in the recent increase of Antarctic sea ice extent. *Geophysical Research Letters*, 36(8): L08502, doi: [10.1029/2009GL037524](https://doi.org/10.1029/2009GL037524)
- Venegas S A, Mysak L A. 2000. Is there a dominant timescale of natural climate variability in the Arctic?. *Journal of Climate*, 13(19): 3412–3434, doi: [10.1175/1520-0442\(2000\)013<3412:ITADTO>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<3412:ITADTO>2.0.CO;2)
- Walsh J E. 2014. Intensified warming of the Arctic: Causes and impacts on middle latitudes. *Global and Planetary Change*, 117: 52–63, doi: [10.1016/j.gloplacha.2014.03.003](https://doi.org/10.1016/j.gloplacha.2014.03.003)
- Warner J C, Geyer W R, Lerczak J A. 2005. Numerical modeling of an estuary: A comprehensive skill assessment. *Journal of Geophysical Research: Oceans*, 110(C5): C05001, doi: [10.1029/2004JC002691](https://doi.org/10.1029/2004JC002691)
- Willmott C J. 1981. On the validation of models. *Physical Geography*, 2(2): 184–194, doi: [10.1080/02723646.1981.10642213](https://doi.org/10.1080/02723646.1981.10642213)
- Yang Xiaoyi, Yuan Xiaojun. 2014. The early winter sea ice variability under the recent Arctic climate shift. *Journal of Climate*, 27(13): 5092–5110, doi: [10.1175/jcli-d-13-00536.1](https://doi.org/10.1175/jcli-d-13-00536.1)
- Yang Xiaoyi, Yuan Xiaojun, Ting Mingfang. 2016. Dynamical link between the Barents-Kara sea ice and the Arctic Oscillation. *Journal of Climate*, 29: 5103–5122, doi: [10.1175/JCLI-D-15-0669.1](https://doi.org/10.1175/JCLI-D-15-0669.1)
- Zhang Xiangdong, Walsh J E. 2006. Toward a seasonally ice-covered Arctic Ocean: Scenarios from the IPCC AR4 model simulations. *Journal of Climate*, 19(9): 1730–1747, doi: [10.1175/JCLI3767.1](https://doi.org/10.1175/JCLI3767.1)
- Zunz V, Goosse H, Massonnet F. 2013. How does internal variability influence the ability of CMIP5 models to reproduce the recent trend in Southern Ocean sea ice extent?. *The Cryosphere*, 7(2): 451–468, doi: [10.5194/tc-7-451-2013](https://doi.org/10.5194/tc-7-451-2013)