# Improved Multimodel Superensemble Forecast for Sea Ice Thickness using Global Climate Models

Wang Yangjun1, Liu Kefeng1*, Zhang Ren1, Qian Longxia2, Zhang Yu3

1 Institute of Meteorology and Oceanology, National University of Defense Technology, Nanjing 211101, China

5  2 School of Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

3 PLA 95746 troops, Sichuan 611530, China

*Correspondence to*: Liu Kefeng (lkfnudt@sina.com)

**Abstract.** This paper aims to find an ensemble method that combines the global climate models, providing an accurate forecast of sea ice thickness (SIT). The conventional multimodel superensemble method is widely used in the atmospheric, oceanic, and other fields, but it does not perform well in SIT simulations. Hence, an adaptive forecasting through exponential re-weighting algorithm is adopted to improve the conventional multimodel superensemble method. The results demonstrate, through a multi-criteria evaluation, that our proposed method performs better than any other mainstream ensemble method. The proposed method is used to predict future SIT in 2020–2049, and its potential biases are discussed.

## 1 Introduction

Since the 1980s, global climate models (GCMs) have developed rapidly with enormous improvements in computer technologies and global climate observation systems, which are important tools for climate predictions (Somerville, 2011). For enhanced data sharing and model comparison, the fifth phase of the Coupled Model Intercomparison Project (CMIP5) has been developed to gather the most advanced GCMs for Arctic sea ice simulations (Taylor et al., 2012). The results show that Arctic sea ice will likely melt and thin in the 21st century (Taylor et al., 2012). However, there are still large uncertainties in GCM future sea ice thickness (SIT) simulations (Shu et al., 2015; Stroeve et al., 2014) that are determined by the initial conditions, physical processes, and resolutions (Taylor et al., 2012). Thus, future local SIT is predicted at a low level of confidence with a single model (Stephenson and Smith, 2015). These GCMs, albeit with a high uncertainty, can still be applied to the prediction of future Arctic sea ice after careful evaluation (Wang and Overland, 2009).

To improve GCM forecasting accuracy, scholars began to gather the GCM sub candidates into an ensemble forecast, taking the initial error and model uncertainty into consideration (Hou et al., 2001). The idea was first proposed by Epstein (1969) and Leith (1974), and is now widely used in many countries (Du, 2002). The ensemble mean is the simplest and most commonly used method, demonstrating a performance superior  to that of most single models (e.g. Xu et al., 2010). There is a shortcoming in the ensemble mean method where the models with poor skills carry the same weight as those with high ones.

The Cryosphere
Discussions

The bias-removed ensemble mean method (e.g. Zhi et al., 2011; Zhu, 2011), assumes that the poorest model can be made

30　equivalent to the best model with a bias correction. Then, the bias-removed models can be assigned the same weights.

Krishnamurti et al. (1999) adopted a new method called the multimodel superensemble, which utilises linear regression to

minimise the errors between GCMs and observations at the grid level during the training period, and different GCM weights

are obtained and transferred to the forecast phase. This method can effectively reduce the local biases in space and time and

those of vast parameters on different models, as it is far more particular in its weight assignments compared with that of the

35　other two methods (Krishnamurti et al., 2016). Existing studies illustrate that the multimodel superensemble has been widely

applied in weather and climate (e.g. Derber and Wu, 1998; Kazumori et al., 2008; Leutbecher, 2003; Mahfouf and Rabier,

2000), ocean (e.g. Kantha et al., 2008; Lenartz et al., 2010a, 2010b), hurricane (e.g. Munsell et al., 2015; Rios-Berrios et al.,

2014; Sato and Xue, 2013; Xue et al., 2013), and other forecasting, significantly reducing the prediction error. Additionally,

ensemble forecasting based on artificial intelligence technology has been gradually developed (e.g. Zhi ying et al., 2004; Gui

40　and Zhao, 2013; Shi, 2013; Zhang et al., 2018).

However, there is a lack of research on the ensemble forecasting of SIT. Compared with the sea ice concentration (SIC), SIT

is a complex variable that is difficult to observe (Haas, 2010) without sufficient and reliable large-scale satellite observations

(e.g. Laxon et al., 2013; Tilling et al., 2016). Both statistical methods (e.g. Lindsay et al., 2008) and numerical models (e.g.

Holland et al., 2011) show that there is a strong correlation between SIT and sea ice extent (SIE). The change in SIT is more

45　significant and can provide more information than SIC, especially in central areas (Melia et al., 2015c). The reduction in SIT

extends the navigation season, making high-latitude sailing possible (Smith and Stephenson, 2013) and facilitating the

exploration of abundant natural resources, impacting the Arctic ecosystem and mid-latitude climate (Francis and Vavrus, 2012).

However, experiments have illustrated that the conventional multimodel superensemble based on linear regression cannot

simulate SIT well due to the sparsity of temporal and spatial SIT data. Moreover, Yang (2001) pointed out that complicated

50　ensemble methods can lead to unstable weights and inferior performances than those of the best candidates. Therefore, an

ensemble method with an improved forecasting capacity is required for further investigations.

To fill this research gap, this study incorporates 12 high-performing GCMs for different scenarios and initial conditions that

were evaluated by Wang and Overland (2015), amounting to 101 ensemble candidates in total, and a new method called

adaptive forecasting through exponential re-weighting (AFTER) is adopted to improve the conventional multimodel

55　superensemble method. Monthly SIT data from 2006–2017 were used in the training phase, while monthly data from 2018

were used in the test phase. A multi-criteria evaluation, including root mean square error (RMSE), correlation coefficient (CC),

structural similarity index measure (SSIM), empirical orthogonal function (EOF) analysis, and sea ice volume (SIV), is

incorporated in this study to examine the validation of the proposed method and other mainstream ensemble methods, e.g. the

ensemble mean, bias-removed ensemble mean, multimodel superensemble, and artificial neural network. The results show that

The Cryosphere
Discussions
Open Access
EGU

60　the improved multimodel superensemble algorithm has a superior performance to that of the other algorithms. Finally, a new

method is adopted for SIT projection from 2020 to 2049.

The reminder of this paper is organised as follows. A data description is presented in Section 2. Section 3 introduces the

methodology, followed by the model validation test in Section 4. Finally, Section 5 provides the future SIT predictions and

summary.

## 2 Observations and climate simulations

### 2.1 Pan-Arctic Ice-Ocean Modelling and Assimilation System data

Spatial consistency, temporality, and completeness are key factors in data evaluation (Melia et al., 2015a). The Pan-Arctic Ice-

Ocean Modelling and Assimilation System (PIOMAS) sea ice reanalysis data, which assimilated the atmospheric reanalysis

from the National Centres for Environment Prediction, consists of SIC satellite (Lindsay and Zhang, 2006) and sea surface

70　temperature observations (Schweiger et al., 2011); this dataset was selected for use in this study (Zhang and Rothrock, 2003).

The quality of PIOMAS was evaluated by Schweiger et al. (2011), demonstrating biases in PIOMAS of 0.26 m in autumn and

0.1 m in spring, compared with the ICESat data (Zwally et al., 2002). Although uncertainty exists in the PIOMAS data, current

satellite observations (i.e. ICESat or CryoSat-2) have limited spatial and temporal coverage, restricting their ability to evaluate

models. Moreover, the largest discrepancy between PIOMAS and ICESat data is found in the north of Greenland and the

75　Canadian Archipelago, the thickest sea ice areas; the PIOMAS data have fewer discrepancies with the situ data than that of

ICESat due to the difference in the satellite inversion methods (Schweiger et al., 2011). Labe et al. (2018) pointed out that the

spatial patterns, seasonal cycles, and SIT trends are sufficiently reproduced by the PIOMAS data. Therefore, PIOMAS has

been widely used to represent observations in various studies (e.g. Shu et al., 2015; Labe et al., 2018).

### 2.2 Global climate models

80　This study incorporates 12 GCMs from CMIP5 that were evaluated by Wang & Overland (2015) for a combined forecast, with

a total of 101 ensemble candidates for four emission scenarios called representative concentration pathways (RCPs) 2.6, 4.5,

6.0, and 8.5 (van Vuuren et al., 2011).

The basic characteristics of the selected ensemble candidates are displayed in Table 1. Regarding the discrepancies in spatial

resolution, all the model candidates and PIOMAS were interpolated into the same $1^{\circ} \times 1^{\circ}$ resolution.

85　For each candidate of the 12 GCMs, monthly data for 2006–2017 were utilised in the training phase because RCPs were first

used in 2006. Then, monthly data from 2018 were used in the test phase to validate different ensemble methods. Finally,

monthly SIT ensemble data was forecasted for 2019–2050.

The Cryosphere
Discussions
Open Access
EGU

**Table 1. List of models used in the CMIP5 subset**

| Number | Model Name | Spatial Resolution | Ensemble candidates (RCP) | | | | Reference |
|---|---|---|---|---|---|---|---|
| | | | 26 | 45 | 60 | 85 | |
| 1 | ACCESS1.0 | tripolar, $1° × 1°$, refinement at the equator | | 1 | | 1 | (Bi et al., 2013) |
| 2 | ACCESS1.3 | tripolar, $1° × 1°$, refinement at the equator | | 1 | | 1 | (Bi et al., 2013) |
| 3 | CCSM4 | dipolar, $1.11° × (0.27 − 0.54)°$, NP in Greenland | 5 | 6 | 6 | 6 | (Gent and Danabasoglu, 2011) |
| 4 | CESM1 | dipolar, $1.11° × (0.27 − 0.54)°$, NP in Greenland | 3 | | | 1 | (Gent and Danabasoglu, 2011) |
| 5 | EC-EARTH | tripolar, $1° × 1°$, refinement at the equator | 2 | 10 | | 10 | (Fichefet and Maqueda, 1999) |
| 6 | HadGEM2-ES | $(1 − 0.3)° × 1°$ | 4 | 4 | 4 | 5 | (Mclaren et al., 2006) |
| 7 | HadGEM2-CC | $(1 − 0.3)° × 1°$ | 1 | 1 | 1 | 1 | (Mclaren et al., 2006) |
| 8 | HadGEM2-AO | $(1 − 0.3)° × 1°$ | | 1 | | 3 | (Mclaren et al., 2006) |
| 9 | MIROC-ESM | $~1.4° × 1°$ | 1 | 1 | 1 | 2 | (Watanabe et al., 2011) |
| 10 | MIROC-ESM-CHEM | $~1.4° × 1°$ | 1 | 1 | 1 | 1 | (Watanabe et al., 2011) |
| 11 | MPI-ESM-LR | $~1.5° × 1.5°$ | 3 | 3 | | 3 | (Notz et al., 2013) |
| 12 | MPI-ESM-MR | $~0.4° × 0.4°$ | 1 | 3 | | 1 | (Notz et al., 2013) |
| Sum | | | 21 | 32 | 13 | 35 | 101 |

**3 Ensemble forecast methodology**

90    The ensemble forecast aims to improve the model projection accuracy by making full use of multiple information sources from

the GCMs and constraining they GCMs using observations. We have tested the performances of different ensemble forecast

methods that are seldomly used in SIT projection (Section 4). SIT simulations based on the conventional multimodel

superensemble, an advanced method that can significantly improve predictions in other areas, exhibit large observation biases

due to the linear regression overfitting. Therefore, a new weight determination method was adopted to improve the

95    conventional multimodel superensemble. The ideas behind this improved method and other ensemble forecast methods used

in this study are introduced in this section. The mathematical notation for the following equations is in Table 2.

**Table 2. Notation key**

| Notation | Description |
|---|---|
| $M_i$ | The ith candidate of the total ensemble candidates $i \in N$ |
| $L$ | Latitude and longitude information of each grid |
| $O_h$ | PIOMAS data over the training period (2006–2017) |
| $x_h$ | $x$ over the training period (2006–2017) |
| $\bar{x}$ | Time mean of $x$ over the training period |
| $\langle x \rangle$ | Ensemble mean of $x$ |
| $\hat{x}$ | Temporally detrended $x$ over the training period |
| $\tilde{x}$ | Temporally trend $x$ over the training period |

| Notation | Description |
|---|---|
| $x_t$ | $t$ over the test period (2018) |
| $\sigma$ | Standard deviation |
| MEAN | Ensemble mean method |
| BIAS | Bias-removed ensemble mean method |
| MARVIC | MARVIC method |
| ANN | Ensemble forecast with ANN |
| SUPER | Conventional multimodel superensemble method |
| AFTER.L1 | Improved multimodel superensemble method with L1-norm AFTER |
| AFTER.L2 | Improved multimodel superensemble method with L2-norm AFTER |

**3.1 Ensemble mean**

The ensemble mean method was widely used in the fifth report from Intergovernmental Panel on Climate Change to predict

100   atmospheric, oceanic, and cryospheric variables. This approach averages all the ensemble candidates, $\langle M_{i,t} \rangle$, regardless of

model discrepancy.

$$SIT_{MEAN} = \langle M_{i,t} \rangle. \tag{1}$$

**3.2 Bias-removed ensemble mean**

Bias-removed ensemble mean methods attempt to correct the models with poor accuracies before averaging them. The

conventional approach corrects the time mean by subtracting the biases between each ensemble candidate, $\bar{M}_{i,h}$ , and

105   observation, $\bar{O}_h$, during the training period at the grid level.

$$SIT_{BIAS} = \langle M_{i,t} - (\bar{M}_{i,h} - \bar{O}_h) \rangle. \tag{2}$$

The mean and variance correction ensemble mean (MAVRIC), first proposed by Nathanael Melia et al. (2015), attempts to

consider both mean and variance in the bias correction. This study incorporates MAVRIC using the ratio of the temporal

standard deviation of the detrended observations, $\sigma_{\tilde{O}_h}$, to the standard deviation of each detrended candidate, $\sigma_{\tilde{M}_{i,h}}$, over the

training period (Eq. 3). Each model is detrended using the linear time series trend during the training period. The multiplicative

110   correction is first detrended, the variance is then corrected, and the trend is re-applied.

$$SIT_{MAVRIC} = \langle (M_{i,t} - \tilde{M}_{i,t}) \frac{\sigma_{\tilde{O}_h}}{\sigma_{\tilde{M}_{i,h}}} + \tilde{M}_{i,t} \frac{\bar{O}_h}{\bar{M}_{i,h}} \rangle. \tag{3}$$

**3.3 Ensemble forecast via artificial neural network**

The ensemble forecast with an artificial neural network (ANN) is biologically motivated, imitating the abilities of the human

brain including information storing, learning, and training to minimise the difference between multi-models and observations.

The algorithm structure can be seen in Eq. (4) and Figure 1. The input layer consists of all the candidates and their geographic

115   information (latitude and longitude), and each grid with a single time clip during the training period is treated as a training

sample, 905,616 in total with 103 dimensions. The output layer consists of related observations. The network can be obtained

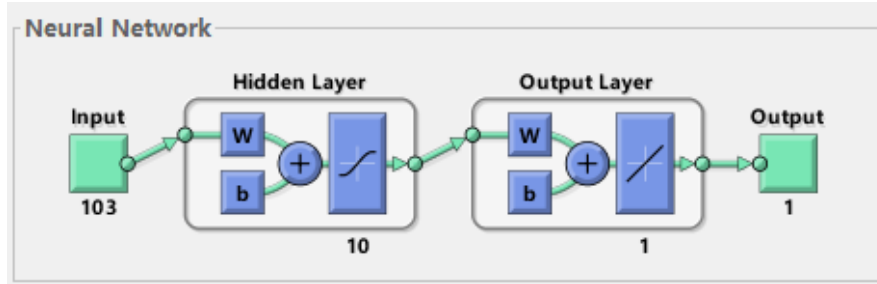through the training phase and used for predictions.

**Figure 1:** ANN network.

$$SIT_{ANN} = net(M_{i,t}, L), \tag{4}$$

### 3.4 Multimodel superensemble forecasting and its improvement

The conventional multimodel superensemble method is a type of regression-improved forecast that provides weights for each grid (Eq. 5). The weights can be obtained by minimising the errors between all the candidates and observations during the training phase, and a linear regression is most commonly used (Eq. 6).

$$SIT_{SUPER} = \overline{O_h} + \langle a_i(M_{i,t} - \overline{M}_{i,h}) \rangle, \tag{5}$$

$$a_i \in argmin \sum_{j=1}^{h} (M_{i,j} - O_j)^2. \tag{6}$$

In our experiment, this superensemble method contributes to large systematic errors in SIT projection (Section 4). As a solution for this, we adopted the AFTER algorithm in the superensemble structure to improve the weight calculations, maintaining all positive weights to avoid overfitting and instability. The improved algorithm was first proposed by Yang (2001b) and can be presented as follows:

$$SIT_{AFTER} = \overline{O_h} + \langle w_i(M_{i,t} - \overline{M}_{i,h}) \rangle, \tag{7}$$

$$w_i = \frac{E_i}{\sum_{j=1}^{n} E_j}, \tag{8}$$

$$E_i = \prod_{k=1}^{n} \hat{s}_{i,k}^{-n/2} exp(-\lambda L_i). \tag{9}$$

Two weighting forms are proposed, and the L1-norm AFTER algorithm is as follows.

Step 1. Split the data of each candidate for the training period in two parts, $Z^{(1)} = (M_{i,h1}, O_{i,h1})$, $1 \le h_1 \le n/2$ and $Z^{(2)} = (M_{i,h2}, O_{i,h2})$, $n/2 + 1 \le h_2 \le n$.

Step 2. Based on $Z^{(1)}$, compute the mean absolute prediction error $\hat{s}_{k,j} = (2/n) \sum_{h_1=1}^{n/2} |M_{i,h1} - O_{h1}|$ for the $ith$ candidate model.

Step 3. For the $ith$ model, the loss function, $L_i$, from Eq. 9 can be written as

$$L_i = \sum_{h2=n/2+1}^{n} |M_{i,h2} - O_{h2}|. \tag{10}$$

Step 4. Compute the convex weight for the $ith$ model following wed by Eqs. (8) and (9).

Step 5. Randomly permute the order of the data N-1 times. Repeat Steps 2–5.

Note that a tuning parameter $\lambda$ is used to control the effect of weighting on the forecast performance (normally, $\lambda = 1$).

The workflow of the L2-norm AFTER algorithm is similar to that of L1-norm AFTER, expect for the function in Step 3, which should be rewritten as

$$L_i = \sum_{h2=n/2+1}^{n} \frac{(M_{i,h2} - O_{h2})^2}{2\hat{s}_{h2}^2}. \tag{11}$$

## 4 Method validation

140   In this study, we analysed the performance of the ensemble forecast methods mentioned above using a statistical multi-criteria evaluation approach. Both univariate and multivariate techniques including RMSE, CC, SSIM, and EOF analysis are adopted to capture the reliability and nature of the ensemble models. The analysis is performed both spatially and temporally on the ensemble forecast datasets during the testing phase. The temporal scale analysis was used to understand the prediction ability of different ensemble forecast methods for the SIT variation trends, testing whether the methods can perform well after being

145   fully trained. Gridded data were used to analyse the discrepancies between the diverse ensemble forecast models in different regions. Finally, these ensemble datasets were adopted to reproduce the monthly variations of the SIV in 2018. The results have provided various statistical properties for these methods.

### 4.1 RMSE & CC test

The RMSE of the datasets measures the deviation between the simulation and observation. The CC of the datasets refers to

150   the degree of linear correlation between them, combining the concepts of mean, standard deviation, and regression line. In this study, both the RMSE and CC of different SIT ensemble forecasts were calculated using the spatial (Figures 2 and 4) and temporal means (Figures 3 and 5).

Figure 2 illustrates that the two improved multimodel superensemble methods have the minimum spatial average RMSEs, during each month in the testing phase, showing the least amount of bias compared with the observation trends. In this

155   experiment, the bias-removed ensemble mean algorithm and ensemble forecast with the ANN algorithm can also perform better than any single ensemble candidate, improving the simulation accuracy of the conventional models. Combining the four algorithms, the largest RMSE occurs in August during the testing period, which is consistent with that of the greatest SIV anomalies driven by a positive feedback loop between the SIT and ice-albedo (Bushuk et al., 2017). That kind of feedback is affected by melt pond information, snowfall, and sea ice concentration, which cannot be sufficiently simulated by the current

160   GCMs from the CMIP5 (Stroeve et al., 2014), restricting the effectiveness of the ensemble forecast in that month.

Figure 3 reflects the temporally averaged RMSEs between the ensemble models and observations at the grid level. Two superensemble methods modified by AFTER, and the bias-removed ensemble mean method captured less of the RMSE than the other selected methods in most parts of the Arctic region. In view of the spatial distribution, the highest biases arose along

The Cryosphere
Discussions
Open Access
EGU

the coastlines, especially in the East Siberian Sea, north of the Canadian Arctic Archipelago, and coastlines of Greenland

165 extending to the islands of Svalbard, the highest thickness variability area (Blanchard-Wrigglesworth and Bitz, 2014).

Figure 4 illustrates that the best three ensemble forecast models based on averaged spatial CC tests are the two improved

multimodel superensemble methods and the bias-removed ensemble mean method, where the values are approximately 0.9 for

all the months. The latter method performs better from July to November, while our proposed methods have an advantage in

the remaining months of 2018, and all of the models together have a higher correlation with the observations than that of any

170 single ensemble candidate. Moreover, these methods and the ANN method conclude that the least relevant occurs in August,

while that appears in September for other methods, providing more evidence for model evaluation. From the spatial distribution

perspective, we can see that the datasets from all the listed ensemble methods in Figure 5 are highly correlated with the

observation data for most Arctic regions, except for an area in the north of the Canadian Arctic Archipelago, from the Fram

Strait to the Norwegian Sea, and part of the Barents Sea, which demonstrated a discrepancy compared to the RMSE test results.
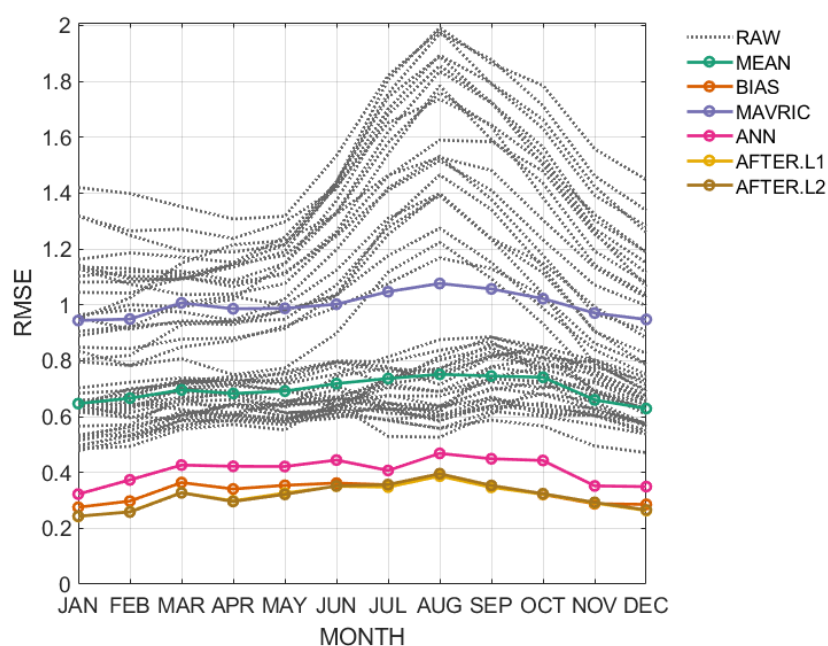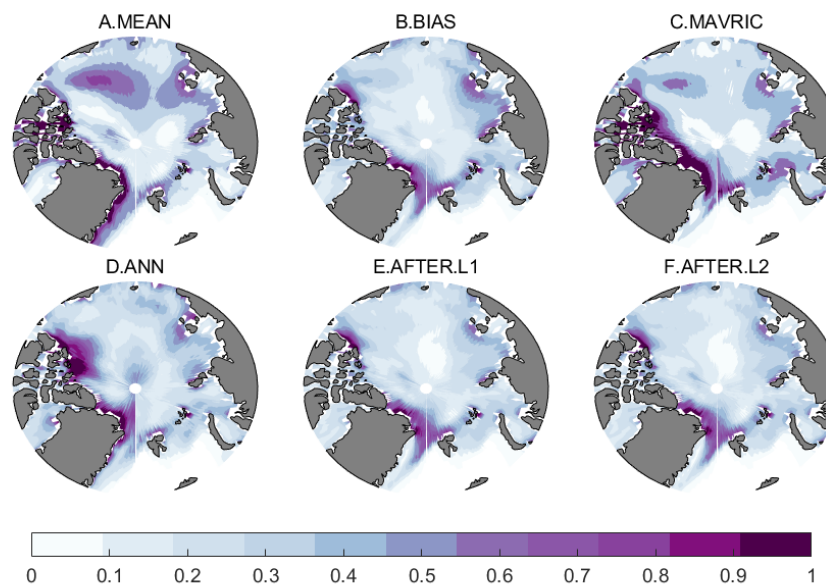


175

**Figure 2: Spatially averaged SIT RMSEs in multiple datasets based on different ensemble forecast methods and observations in 2018. Results of all 101 candidates are depicted using grey dashed lines, which are marked as raw data. All model abbreviations are the same as those provided in Table 2. Note that the average RMSEs based on the conventional multimodel superensemble method are larger than $10^6$ and are not plotted in this figure.**
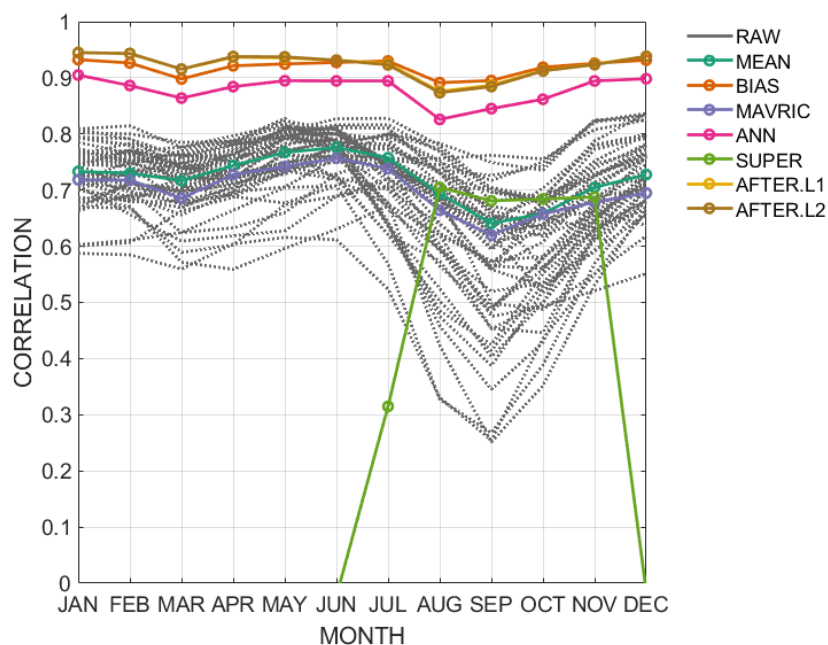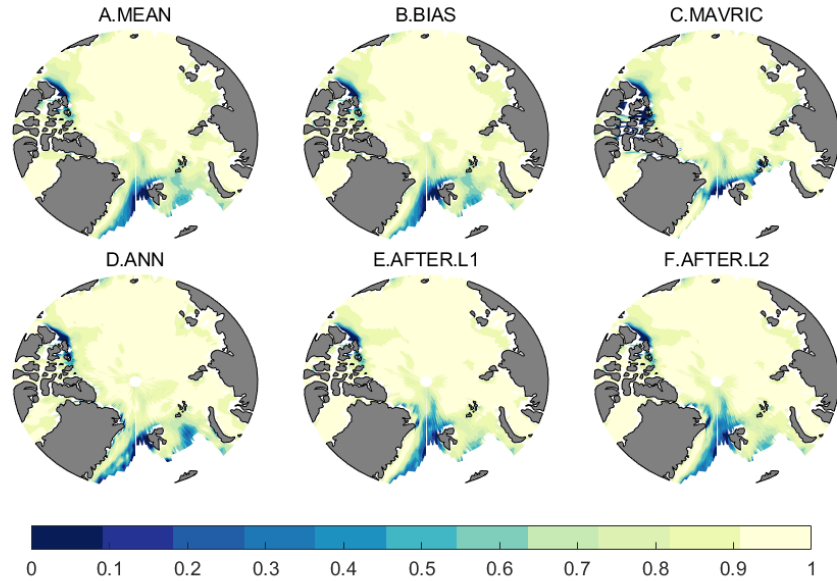
The Cryosphere
Discussions
Open Access
EGU

180



**Figure 3: Temporally averaged SIT RMSEs in multiple datasets are based on different ensemble forecast methods and gridded observations for 2018. All model abbreviations are the same as those provided in Table 2. Note that the average RMSEs based on the conventional multimodel superensemble method are larger than $10^5$ and are not plotted in this figure.**



185

**Figure 4: Spatially averaged SIT CCs for multiple datasets based on different ensemble forecast methods and observations for 2018. Results of all 101 candidates are depicted using grey dashed lines, which are marked as raw data. All model abbreviations are the same as those provided in Table 2.**

The Cryosphere

Discussions

Open Access

EGU



**Figure 5: Temporally averaged SIT CCs in multiple datasets based on different ensemble forecast methods and gridded observations for 2018. All model abbreviations are the same as those provided in Table 2. Note that the average CCs based on the conventional multimodel superensemble method perform poorly and are not plotted in this figure.**

### 4.2 Structural similarity index measure (SSIM)

The SSIM is used to evaluate the performance of different ensemble models by adopting a single index that can measure overall spatial patterns, which is a type of objective test method (Wang et al., 2004). Compared to traditional error methods (e.g. RMSE), this method can better depict the spatial distribution differences between model outputs and observations. The formula of this SSIM can be presented as follows:

$$SSIM(X_{mod}, X_{obs}) = l(X_{mod}, X_{obs})^{\alpha} \cdot c(X_{mod}, X_{obs})^{\beta} \cdot s(X_{mod}, X_{obs})^{\gamma}, \tag{12}$$

$$l(X_{mod}, X_{obs}) = \frac{\left(2\mu_{x_{mod}}\mu_{x_{obs}} + C_1\right)}{\left(\mu_{x_{mod}}^2 + \mu_{x_{obs}}^2 + C_1\right)}, \tag{13}$$

$$c(X_{mod}, X_{obs}) = \frac{2\sigma_{x_{mod}}\sigma_{x_{obs}} + C_2}{\sigma_{x_{mod}}^2 + \sigma_{x_{obs}}^2 + C_2}, \tag{14}$$

$$s(X_{mod}, X_{obs}) = \frac{\sigma_{x_{mod}x_{obs}} + C_3}{\sigma_{x_{mod}}\sigma_{x_{obs}} + C_3}, \tag{15}$$

where the variations in mean value, $l(X_{mod}, X_{obs})$, deviation, $c(X_{mod}, X_{obs})$, and structure, $s(X_{mod}, X_{obs})$ are combined together; $\mu$ and $\sigma$ are the mean value and standard deviation of $X$, respectively; and $C_1, C_2,$ and $C_3$ are the constants to prevent system instability. Generally, if $\alpha = \beta = \gamma = 1, C_3 = C_2/2$ , then Eq. (12) can be rewritten as follows:

$$SSIM(X_{mod}, X_{obs}) = \frac{\left(2\mu_{x_{mod}}\mu_{x_{obs}} + C_1\right)\left(2\sigma_{x_{mod}}\sigma_{x_{obs}} + C_2\right)}{\left(\mu_{x_{mod}}^2 + \mu_{x_{obs}}^2 + C_1\right)\left(\sigma_{x_{mod}}^2 + \sigma_{x_{obs}}^2 + C_2\right)}. \tag{16}$$

A sample model with a random matrix A of $10 \times 10$ that ranges from 0 to 1 is provided to verify the advantage of SSIM.

Matrix B can be obtained if 10 is added to the last element, while Matrix C can be obtained if we add 1 or -1 randomly to each

element. Both the SSIM and RMSE are calculated to test the difference between B and A, and that between C and A. The

results show that the RMSE of these matrices are the same, while the SSIM calculated for Matrix B is larger than that of Matrix

205    C, showing that the SSIM provides more information than the RMSE (Figure 6).

Figure 7 illustrates that the improved multimodel superensemble methods have higher scores than those of any other single

candidate every month in the SSIM, matching the spatial distributions with the observations. The improved methods, together

with the bias-removed and ANN methods, demonstrate that the largest SIT spatial distribution biases arise in August, which

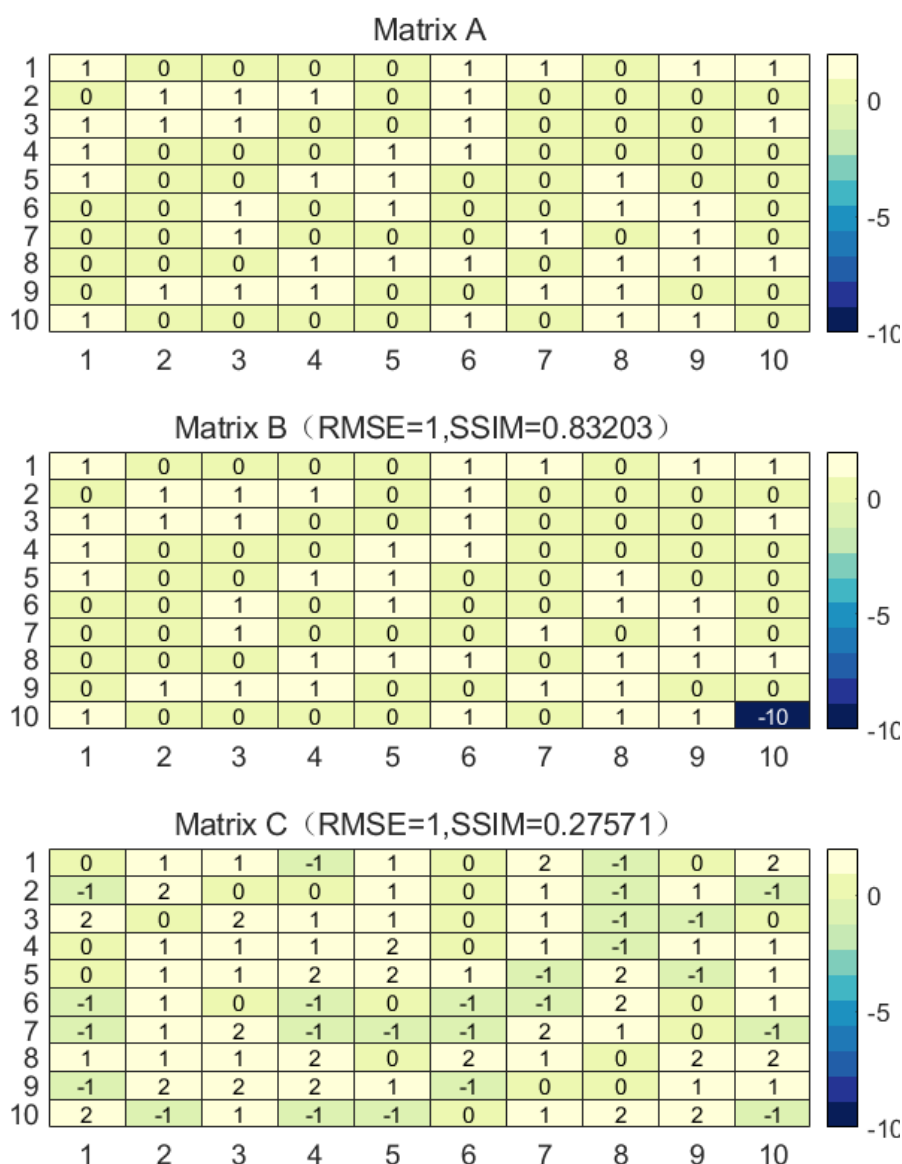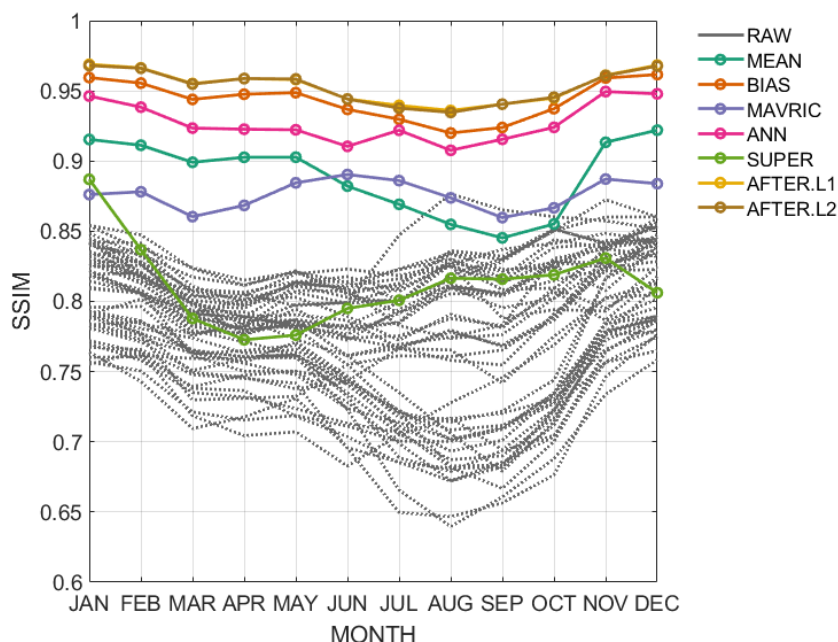the other ensemble methods cannot capture.



210

**Figure 6: An example of the structural similarity analysis.**

**Figure 7: The SSIM of SIT in multiple datasets based on different ensemble forecast methods and observations in 2018. Results of all 101 candidates are depicted using grey dashed lines and are marked as raw data. All model abbreviations are the same as those provided in Table 2.**

215

**4.3 EOF analysis**

The EOF analysis is introduced to further evaluate the model performances to find a reliable ensemble model to reproduce a realistic sea ice climatology, consistently capturing the major spatial modes and their related 1D principal components (PC). In this study, raw datasets from seven ensemble models and observations are processed by the EOF analysis, where the

220 cumulative total variance of the first four leading EOF patterns is over 99% (Figures 8 and 9). The first spatial pattern reflects the monthly mean SIT fields during the testing phase, and all the ensemble methods except for the conventional superensemble method have similar EOF1 patterns, with the thickest ice north of Greenland and the Canadian Arctic Archipelago, and thinner ice east of the Arctic Basin. The corresponding PC1 patterns of the improved superensemble forecast methods match those of the observations the best, while others have smaller fluctuations. The EOF2 patterns are similar in most ensemble datasets

225 (except for the pattern from the conventional superensemble method), exhibiting a high positive centre of variation in the northern Arctic Basin and a negative centre north of Greenland. Their corresponding PC patterns in Figure 9 consistently depict the SIT downtrend from February to September and uptrend from June to October. In Figure 8, for each EOF3 pattern, only the bias-removed ensemble mean, ANN, and the two improved superensemble methods match the observations, showing a "positive–negative–positive" spatial structure from the west to the east of the Arctic regions. When combined with their

230 corresponding PC3 patterns in Figure 9, the ANN method becomes the only ensemble model to match the observations. The

12

The Cryosphere
Discussions

EOF4 patterns of most of the datasets exhibit similar structure, leading to disorders in their corresponding PC patterns, while only the ANN method can capture a similar trend and amplitude as that of the observations in the fourth PC pattern, meaning that this method displays more detailed results.
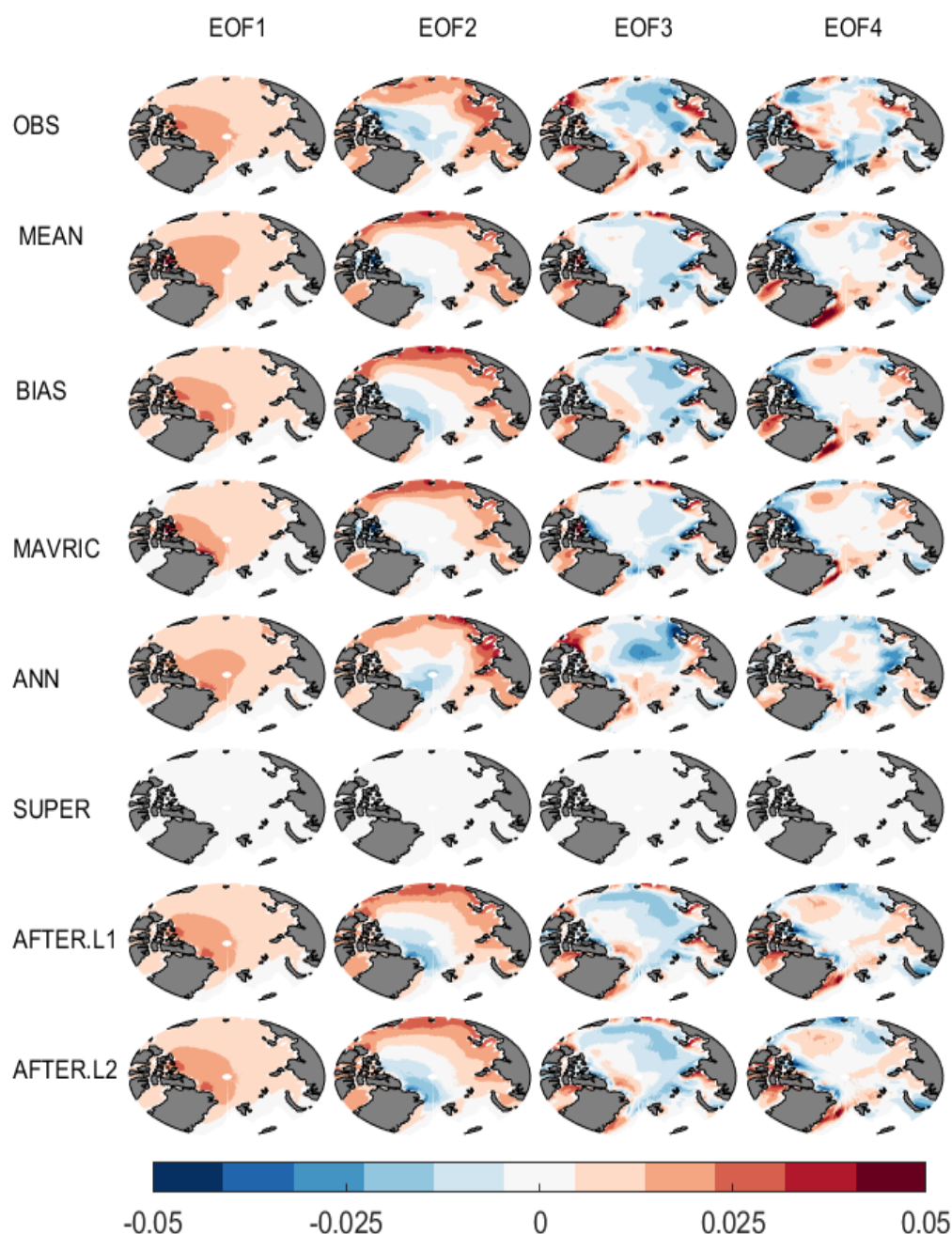


Figure 8: The first four EOF patterns of the seven ensemble forecast models and the observations. Note: areas in red indicate the highest positive anomalies, while areas in blue indicate the highest negative anomalies.
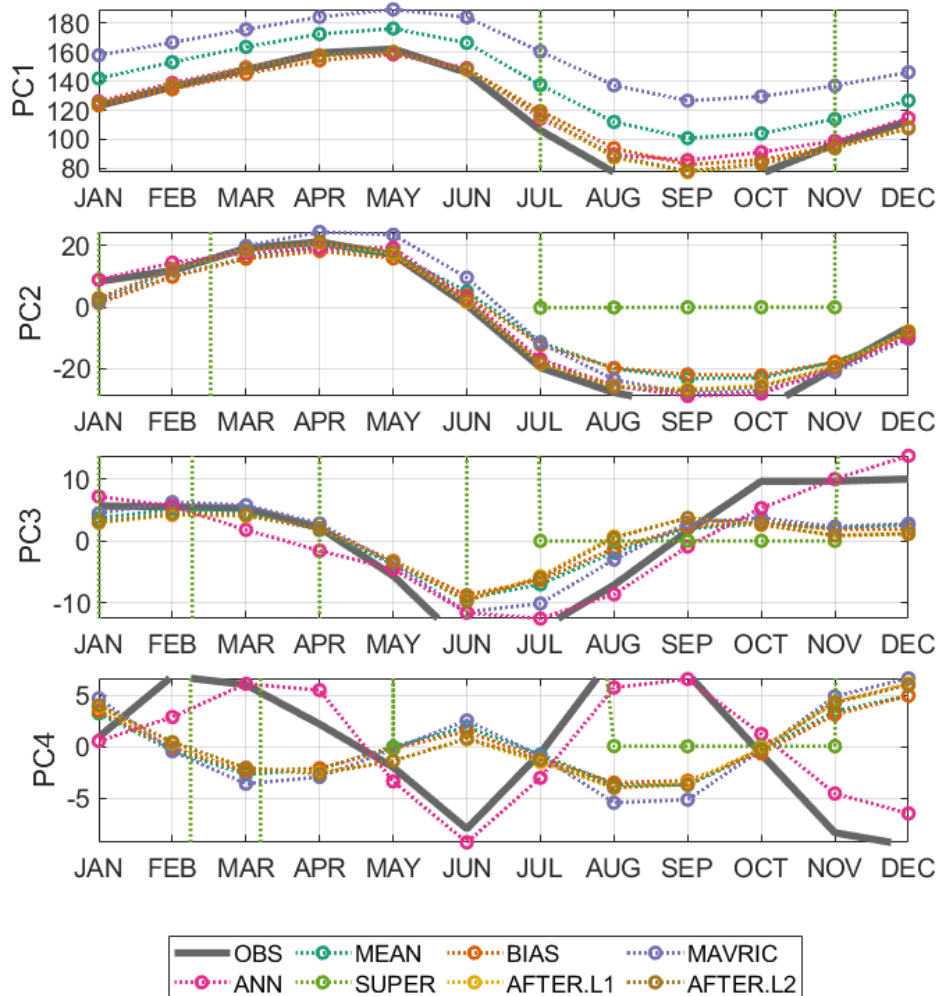
**Figure 9: PC patterns (time coefficients) of the first four leading modes.**

**4.4 Sea ice volume**

240   Sea ice volume (SIV) is an important index in the assessment of sea ice simulations, combining SIT, SIC, and grid area into

consideration. Here, SIC datasets from PIOMAS and the GCMs listed in Table 2 are adopted to calculate SIV (Eq. 17).

$$SIV = \sum_{lon} \sum_{lat} SIT(lon, lat) \cdot SIC(lon, lat) \cdot \left( sin\left( \frac{lat+1}{180} \cdot \pi \right) - sin\left( \frac{lat}{180} \cdot \pi \right) \cdot \frac{2\pi r^2}{360} \right). \qquad (17)$$

Figure 10 illustrates that SIV datasets from the improved superensemble methods have the most similar variations in trend and

amplitude compared to those of the observations. SIV simulations from the other methods, such as the bias-removed ensemble

mean and ANN, are lower than those of the observations in most months, but higher from August to October during the testing

245   phase.
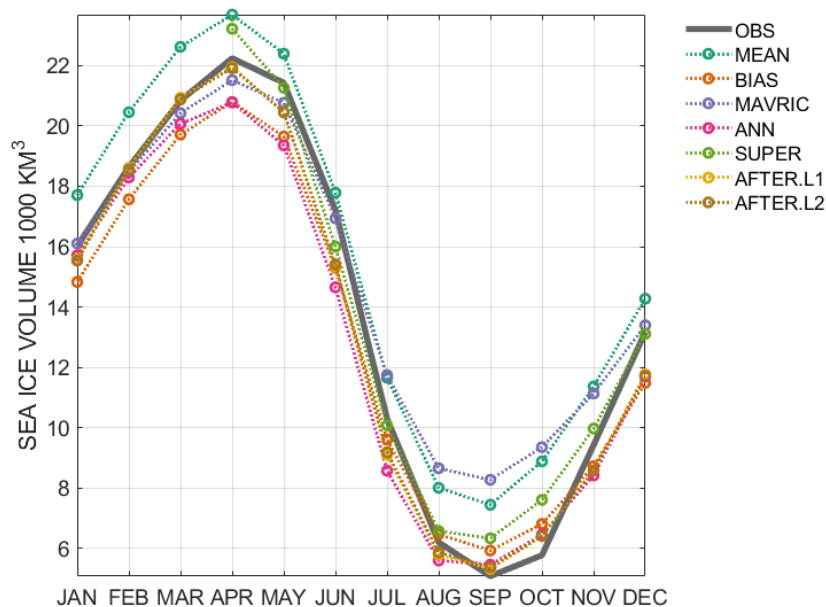
**Figure 10: SIV in multiple datasets based on different ensemble forecast methods and observations for 2018. All model abbreviations are the same as those provided in Table 2.**

## 3.4 Ranking

250   Ranking is provided for the ensemble methods listed in Table 2 based on a multi-criteria evaluation. The best model is given the top ranking, while the worst model is ranked seventh. Table 3 shows that our proposed methods perform better than the other selected ensemble methods for most evaluation techniques. Additionally, the ANN ensemble method can capture more details of the observation dataset in the EOF analysis owing to the advantage of artificial intelligence in data mining. Even the simplest method, the ensemble mean, is highly correlated with the observation dataset in the spatial distribution. To summarise,

255   our proposed methods can greatly improve the accuracy of the ensemble forecast in SIT, which is followed by the conventional bias-removed ensemble mean algorithm and ANN ensemble methods, while the remaining three methods score the lowest.

15

**Table 3. Performance of ensemble methods as evaluated against the observation dataset for each metric based on spatially averaged and temporally averaged RMSE and CC, the SSIM mean, the absolute discrepancy of EOF of the spatial and PC temporal means, and the standard deviation of SIV.**

|  | MEAN | BIAS | MAVRIC | ANN | SUPER | AFTER.L1 | AFTER.L2 |
|---|---|---|---|---|---|---|---|
| RMSE (space) | 0.4040 | 0.2849 | 0.4999 | 0.3470 | $>10^5$ | 0.2690 | 0.2696 |
| RMSE (time) | 0.6968 | 0.3326 | 0.9997 | 0.4065 | $>10^6$ | 0.3136 | 0.3155 |
| CC (space) | 0.8583 | 0.8578 | 0.8447 | 0.8513 | 0.2606 | 0.8597 | 0.8546 |
| CC (time) | 0.7208 | 0.9186 | 0.7001 | 0.8791 | 0.2439 | 0.9226 | 0.9221 |
| SSIM | 0.8894 | 0.9436 | 0.8763 | 0.9275 | 0.8121 | 0.9535 | 0.9531 |
| EOFs (10e-4) | 1.574 | 0.8421 | 2.148 | 0.3165 | 4.210 | 0.5762 | 0.6029 |
| PCs (weight) | 0.8614 | 1.180 | 1.886 | 0.4033 | $>10e6$ | 0.7310 | 0.7378 |
| SIV (std: 10e5) | 1.664 | 1.145 | 1.244 | 1.116 | $>10^5$ | 0.7018 | 0.6979 |
| RANK | 7 | 6 | 5 | 4 | 3 | 2 | 1 |

## 4 Prediction and summary

Based on the study results, the improved superensemble methods are the best ensemble models for simulating SIT, and the improved model with L1-norm AFTER performs better than the other one. Hence, the L1-norm AFTER superensemble method is adopted to predict future variations of the September mean SIT during three periods (2020–2029, 2030–2039, and 2040–2049), where the bias-removed ensemble mean is used as a control group. Figure 11 illustrates that the sea ice will continue to melt in the next three decades and the two selected ensemble models exhibit a similar overall spatial distribution, i.e. thicker sea ice in the west and nearly ice-free in the east.

However, even these two high performing methods still exhibit discernible differences that increase with the prolongation of time. Compared to the results of the bias-removed ensemble mean, the SIT distribution exhibits a "decrease–increase–decrease" belt from the north of Canadian Arctic Archipelago and Greenland Sea to the Barents and Kara Sea in the first decade of the simulation. Then the SIT in the centre of the Arctic Basin increases, while the SIT decrease in the Barents and Kara Sea disappears during 2030–2039. By the middle of the 21st century, the discrepancy arises in the Arctic Basin, while the "two opposite poles" still exist along the west of the Arctic region, showing large uncertainties in SIT prediction in these areas.

The issues related to how many ensemble candidates should be combined to improve the model performance and whether the results change with space are investigated in Figure 12, where the Arctic area is separated into 10 regions. The results show

275   that large spatial differences based on model selection exist for the Arctic regions, combining less than 30 ensemble candidates in the Lincoln Sea, Greenland Sea, and Canadian Arctic Archipelago, and greater number of candidates (over 80) in the Barents Sea, Kara Sea, Chukchi Sea, and Baffin Bay, and combining nearly 60 models for other regions. Figure 13 reflects the weight distributions of each region, showing that the roles of different ensemble candidates vary in different regions. Therefore, the multimodel superensemble structure is far more selective in its assignment of weights.

280   In summary, this study has incorporated an improved weight-determined algorithm in the multimodel superensemble structure to predict the SIT. A multi-criteria evaluation was used to validate the model. The study insights are summarised below.

- The AFTER algorithm can effectively avoid overfitting and instability in the conventional superensemble forecasting method, demonstrating better SIT simulation performance than that of the other mainstream ensemble forecast methods through a multi-criteria evaluation.

285  - Large biases in the SIT simulations between the dataset from the improved ensemble method and the observations were found along the coastline in the west and in August, which was in accordance with the largest SIT anomaly in time and space. This result is restricted by the limited simulations of internal variability and external forcing of SIT from all CMIP5 selected ensemble candidates and can be improved by further developing the GCMs.

- This method was used to forecast the September mean SIT in the next three decades, where the bias-removed ensemble
290   mean method was used as a control group. The results from these two methods exhibited a consistent spatial pattern of a continuous thinning trend in the west and an expanded ice-free area in the east. However, differences between these two high-performing ensemble methods still exist in the Canada Arctic Archipelago, Greenland Sea, and central Arctic Basin, which are enhanced over time.
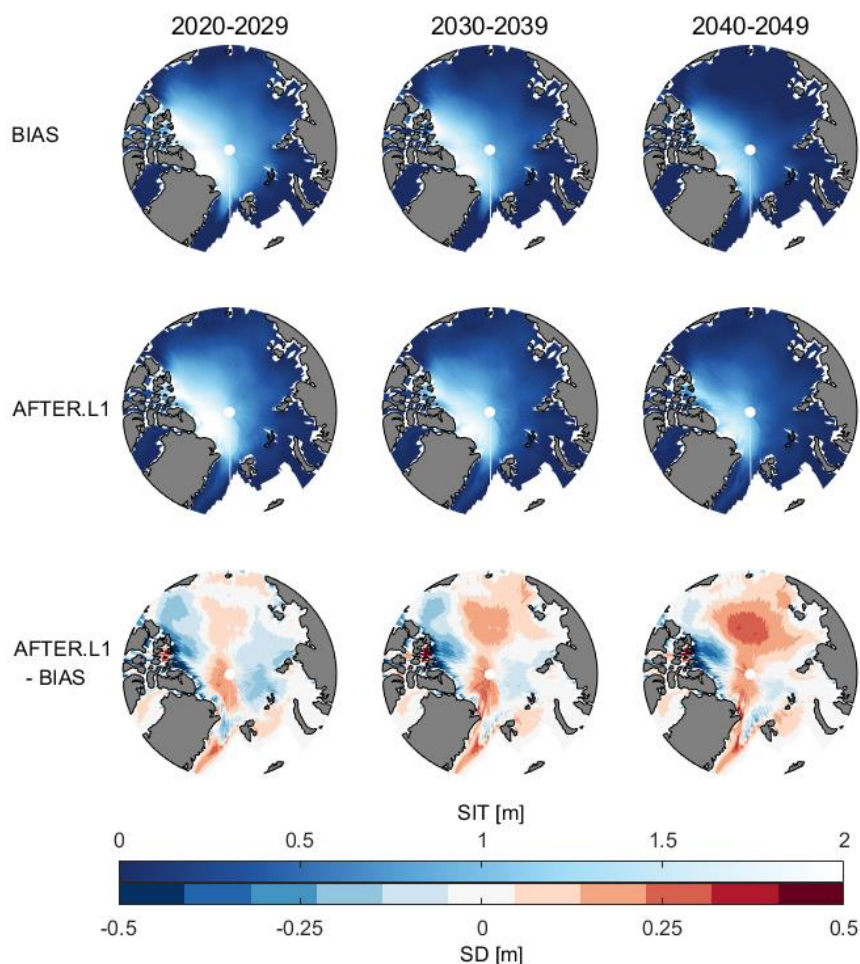
**Figure 11: Ensemble forecast of the September mean SIT derived from all the ensemble candidates listed in Table 1 during 2020–2049, using the bias-removed ensemble mean (top row) and the improved superensemble method with L1-norm AFTER (middle row). The bottom row displays the results of the AFTER.L1 – BIAS method; hence, cold colour areas are where AFTER.L1 has a reduced SIT, and the warm colour areas are where AFTER.L1 has an increased SIT.**
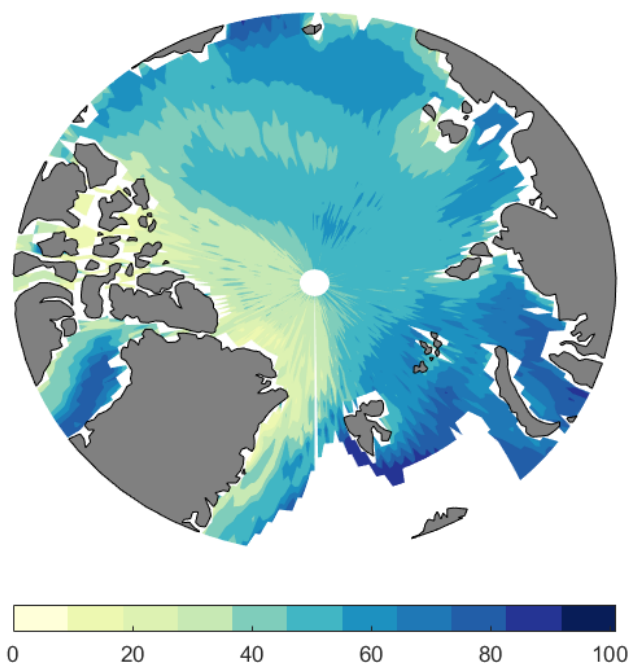
**Figure 12: Number of ensemble candidates used in the improved superensemble method with L1-norm AFTER, where 95% information can be explained. AREAS: (1) Central Arctic Ocean, (2) Lincoln and Greenland Sea, (3) Canadian Arctic Archipelago, (4) Beaufort Sea, (5) Chukchi Sea, (6) East Siberian Sea, (7) Laptev Sea, (8) Kara Sea and Barents Sea, (9) Norwegian Sea, and (10) Baffin Bay.**
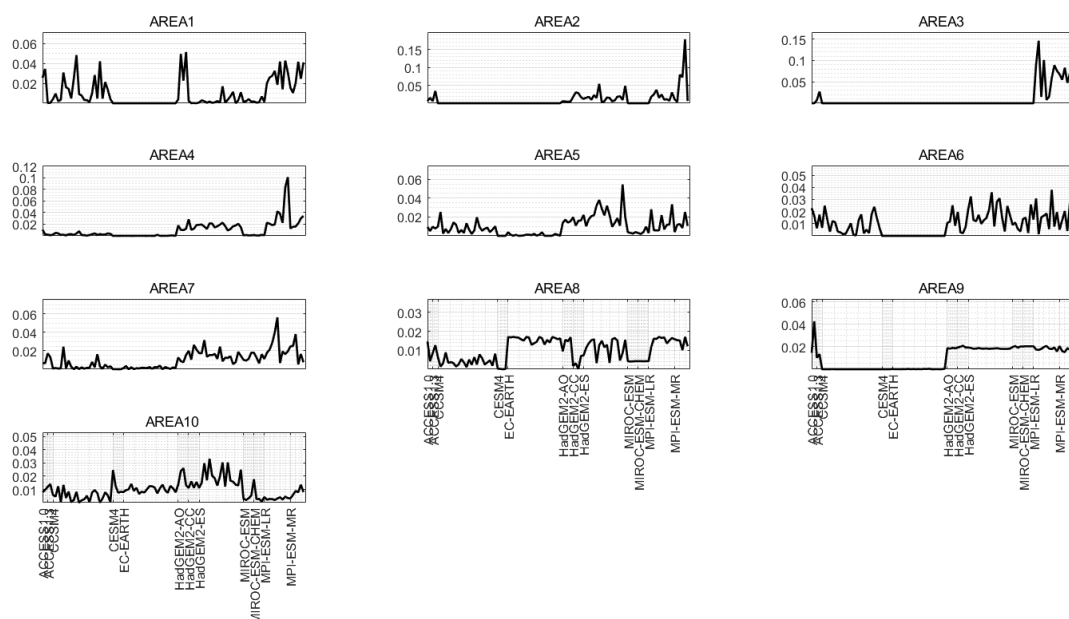


**Figure 13: Weight of each candidate model derived from the improved superensemble method with L1-norm AFTER in different regions. Areas 1–10: (1) Central Arctic Ocean, (2) Lincoln Sea and Greenland Sea, (3) Canadian Arctic Archipelago, (4) Beaufort Sea, (5) Chukchi Sea, (6) East Siberian Sea, (7) Laptev Sea, (8) Kara Sea and Barents Sea, (9) Norwegian Sea, and (10) Baffin Bay.**

310 **Conflicts of Interest:** The authors declare no conflicts of interest.

**Reference**

Bi, D., Dix, M., Marsland, S., O'Farrell, S., Rashid, H., Uotila, P., Hirst, A., Kowalczyk, E., Golebiewski, M., Sullivan, A., Yan, H., Hannah, N., Franklin, C., Sun, Z., Vohralik, P., Watterson, I., Zhou, X., Fiedler, R., Collier, M., Ma, Y., Noonan, J., Stevens, L., Uhe, P., Zhu, H., Griffies, S., Hill, R., Harris, C. and Puri, K.: The ACCESS coupled model: description, control

315 climate and evaluation, Aust. Meteorol. Oceanogr. J., doi:10.22499/2.6301.004, 2013.

Blanchard-Wrigglesworth, E. and Bitz, C. M.: Characteristics of Arctic sea-ice thickness variability in GCMs, J. Clim., 27(21), 8244–8258, doi:10.1175/JCLI-D-14-00345.1, 2014.

Bushuk, M., Msadek, R., Winton, M., Vecchi, G. A., Gudgel, R., Rosati, A. and Yang, X.: Summer enhancement of Arctic sea ice volume anomalies in the September-ice zone, J. Clim., doi:10.1175/JCLI-D-16-0470.1, 2017.

320 Derber, J. C. and Wu, W. S.: The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system, Mon. Weather Rev., doi:10.1175/1520-0493(1998)126<2287:TUOTCC>2.0.CO;2, 1998.

Du, J.: PRESENT SITUATION AND PROSPECTS OF ENSEMBLE NUMERICAL PREDICTION, Q. J. Appl. Meteorlolgy, 13(1), 16–28, 2002.

Epstein, E. S.: A Scoring System for Probability Forecasts of Ranked Categories, J. Appl. Meteorol., doi:10.1175/1520-

325 0450(1969)008<0985:assfpf>2.0.co;2, 1969.

Fichefet, T. and Maqueda, M. A. M.: Modelling the influence of snow accumulation and snow-ice formation on the seasonal cycle of the Antarctic sea-ice cover, Clim. Dyn., 15(4), 251–268, 1999.

Francis, J. A. and Vavrus, S. J.: Evidence linking Arctic amplification to extreme weather in mid-latitudes, Geophys. Res. Lett., doi:10.1029/2012GL051000, 2012.

330 Gent, P. R. and Danabasoglu, G.: Response to Increasing Southern Hemisphere Winds in CCSM4, J. Clim., 24(24), 4992–4998, 2011.

Gui-, T. and Zhao-, S. U. N.: Ensemble prediction of summer rainfall patterns over eastern China based on artificial neural networks, , doi:10.13878/j.cnki.dqkxxb.2013.01.001, 2013.

Haas, C.: Dynamics Versus Thermodynamics: The Sea Ice Thickness Distribution, in Sea Ice: Second Edition., 2010.

335 Holland, M. M., Bailey, D. A. and Vavrus, S.: Inherent sea ice predictability in the rapidly changing Arctic environment of the Community Climate System Model, version 3, Clim. Dyn., doi:10.1007/s00382-010-0792-4, 2011.

Hou, D., Kalnay, E. and Droegemeier, K. K.: Objective verification of the SAMEX '98 ensemble forecasts, Mon. Weather Rev., doi:10.1175/1520-0493(2001)129<0073:OVOTSE>2.0.CO;2, 2001.

The Cryosphere
Discussions

Kantha, L., Carniel, S. and Sclavo, M.: A note on the multimodel superensemble technique for reducing forecast errors, Nuovo

340     Cim. della Soc. Ital. di Fis. C, doi:10.1393/ncc/i2008-10288-2, 2008.

Kazumori, M., Liu, Q., Treadon, R. and Derber, J. C.: Impact study of AMSR-E radiances in the NCEP global data assimilation

system, Mon. Weather Rev., doi:10.1175/2007MWR2147.1, 2008.

Krishnamurti, T. N., Kishtawal, C. M., LaRow, T. E., Bachiochi, D. R., Zhang, Z., Eric Williford, C., Gadgil, S. and Surendran,

S.: Improved weather and seasonal climate forecasts from multimodel superensemble, Science (80-. ).,

345     doi:10.1126/science.285.5433.1548, 1999.

Krishnamurti, T. N., Kumar, V., Simon, A., Bhardwaj, A., Ghosh, T. and Ross, R.: A review of multimodel superensemble

forecasting for weather, seasonal climate, and hurricanes, Rev. Geophys., 54(2), 336–377, doi:10.1002/2015RG000513, 2016.

Labe, Z., Magnusdottir, G. and Stern, H.: Variability of Arctic sea ice thickness using PIOMAS and the CESM large ensemble,

J. Clim., 31(8), 3233–3247, doi:10.1175/JCLI-D-17-0436.1, 2018.

350     Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J., Haas,

C., Hendricks, S., Krishfield, R., Kurtz, N., Farrell, S. and Davidson, M.: CryoSat-2 estimates of Arctic sea ice thickness and

volume, Geophys. Res. Lett., doi:10.1002/grl.50193, 2013.

Leith, C. E.: Theoretical Skill of Monte Carlo Forecasts, Mon. Weather Rev., doi:10.1175/1520-

0493(1974)102<0409:tsomcf>2.0.co;2, 1974.

355     Lenartz, F., Mourre, B., Barth, A., Beckers, J. M., Vandenbulcke, L. and Rixen, M.: Enhanced ocean temperature forecast

skills through 3-D super-ensemble multi-model fusion, Geophys. Res. Lett., doi:10.1029/2010GL044591, 2010a.

Lenartz, F., Beckers, J. M., Chiggiato, J., Mourre, B., Troupin, C., Vandenbulcke, L. and Rixen, M.: Super-ensemble

techniques applied to wave forecast: Performance and limitations, Ocean Sci., doi:10.5194/os-6-595-2010, 2010b.

Leutbecher, M.: A reduced rank estimate of forecast error variance changes due to intermittent modifications of the observing

360     network, J. Atmos. Sci., doi:10.1175/1520-0469(2003)060<0729:ARREOF>2.0.CO;2, 2003.

Lindsay, R. W. and Zhang, J.: Assimilation of ice concentration in an ice-ocean model, J. Atmos. Ocean. Technol., 23(5), 742–

749, doi:10.1175/JTECH1871.1, 2006.

Lindsay, R. W., Zhang, J., Schweiger, A. J. and Steele, M. A.: Seasonal predictions of ice extent in the Arctic Ocean, J.

Geophys. Res. Ocean., doi:10.1029/2007JC004259, 2008.

365     Mahfouf, J. F. and Rabier, F.: The ECMWF operational implementation of four-dimensional variational assimilation. II:

Experimental results with improved physics, Q. J. R. Meteorol. Soc., doi:10.1256/smsqj.56415, 2000.

Mclaren, A. J., Banks, H. T., Durman, C. F., Gregory, J. M. and Laxon, S. W.: Evaluation of the sea ice simulation in a new

coupled atmosphere-ocean climate model (HadGEM1), J. Geophys. Res., 111(C12), C12014, 2006.

Melia, N., Haines, K. and Hawkins, E.: Improved Arctic sea ice thickness projections using bias-corrected CMIP5 simulations, ,

370     (3), 2237–2251, doi:10.5194/tc-9-2237-2015, 2015a.

Melia, N., Hawkins, E. and Haines, K.: Improved Arctic sea ice thickness projections using bias-corrected CMIP5 simulations, The Cryosphere,9,6(2015-12-04), 9(4), 3821–3857, 2015b.

Melia, N., Haines, K. and Hawkins, E.: Improved Arctic sea ice thickness projections using bias corrected CMIP5 simulations, Cryosph. Discuss., 9(4), 3821–3857, doi:10.5194/tcd-9-3821-2015, 2015c.

375   Munsell, E. B., Sippel, J. A., Braun, S. A., Weng, Y. and Zhang, F.: Dynamics and predictability of Hurricane Nadine (2012) evaluated through convection-permitting ensemble analysis and forecasts, Mon. Weather Rev., doi:10.1175/MWR-D-14-00358.1, 2015.

Notz, D., Haumann, F. A., Haak, H., Jungclaus, J. H. and Marotzke, J.: Arctic sea-ice evolution as modeled by Max Planck Institute for Meteorology's Earth system model, J. Adv. Model. Earth Syst., 5(2), 173–194, 2013.

380   Rios-Berrios, R., Vukicevic, T. and Tang, B.: Adopting Model Uncertainties for Tropical Cyclone Intensity Prediction, Mon. Weather Rev., doi:10.1175/mwr-d-13-00186.1, 2014.

Sato, T. and Xue, Y.: Validating a regional climate model's downscaling ability for East Asian summer monsoonal interannual variability, Clim. Dyn., doi:10.1007/s00382-012-1616-5, 2013.

Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H. and Kwok, R.: Uncertainty in modeled Arctic sea ice volume, J.

385   Geophys. Res. Ocean., doi:10.1029/2011JC007084, 2011.

Shi, X.: A new neural network ensemble forecast method based on KPCA for typhoon intensity, J. Meteorol. Sci., 33(2), 184–189, 2013.

Shu, Q., Song, Z. and Qiao, F.: Assessment of sea ice simulations in the CMIP5 models, , (February), doi:10.5194/tc-9-399-2015, 2015.

390   Smith, L. C. and Stephenson, S. R.: New Trans-Arctic shipping routes navigable by midcentury, , 2013(13), 6–10, doi:10.1073/pnas.1214212110, 2013.

Somerville, R. C. J.: The co-evolution of climate models and the Intergovernmental Panel on Climate Change, Dev. Atmos. Gen. Circ. Model. Complexity, Synth. Comput., 2(4), 51–75, 2011.

Stephenson, S. R. and Smith, L. C.: Influence of climate model variability on projected Arctic shipping futures, Earth's Futur.,

395   3(11), 331–343, doi:10.1002/2015EF000317, 2015.

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

Taylor, K. E., Stouffer, R. J. and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bull. Am. Meteorol. Soc., 93(4), 485–498, 2012.

400   Tilling, R. L., Ridout, A. and Shepherd, A.: Near-real-time Arctic sea ice thickness and volume from CryoSat-2, Cryosphere, doi:10.5194/tc-10-2003-2016, 2016.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J. F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J. and Rose, S. K.: The representative concentration pathways: An overview, Clim. Change, doi:10.1007/s10584-011-0148-z, 2011.

405 Wang, M. and Overland, J. E.: A sea ice free summer Arctic within 30 years?, Geophys. Res. Lett., 36(7), doi:10.1029/2009GL037820, 2009.

Wang, M. and Overland, J. E.: Projected future duration of the sea-ice-free season in the Alaskan Arctic, Prog. Oceanogr., 136(January), 50–59, doi:10.1016/j.pocean.2015.01.001, 2015.

Wang, Z., Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P.: Image quality assessment: From error visibility to structural

410 similarity, IEEE Trans. Image Process., 13(4), 600–612, doi:10.1109/TIP.2003.819861, 2004.

Watanabe, M., Chikira, M., Imada, Y. and Kimoto, M.: Convective control of ENSO simulated in MIROC, J. Clim., 24(2), 543–562, 2011.

Xu, Y., Gao, X. and Giorgi, F.: Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections, Clim. Res., doi:10.3354/cr00835, 2010.

415 Xue, M., Schleif, J., Kong, F., Thomas, K. W., Wang, Y. and Zhu, K.: Track and intensity forecasting of hurricanes: Impact of convection-permitting resolution and global ensemble Kalman filter analysis on 2010 Atlantic season forecasts, Weather Forecast., doi:10.1175/WAF-D-12-00063.1, 2013.

Yang, Y.: Adaptive regression by mixing, J. Am. Stat. Assoc., 96(454), 574–588, doi:10.1198/016214501753168262, 2001a.

Yang, Y.: Combining forecasting procedures: Some theoretical results, Econom. Theory, 20(1), 176–222,

420 doi:10.1017/S0266466604201086, 2001b.

Zhang, J. and Rothrock, D. A.: Modeling Global Sea Ice with a Thickness and Enthalpy Distribution Model in Generalized Curvilinear Coordinates, Mon. Weather Rev., doi:10.1175/1520-0493(2003)131<0845:mgsiwa>2.0.co;2, 2003.

Zhang, Y., Fu, K., Sun, H., Sun, X., Zheng, X. W. and Wang, H.: A multi-model ensemble method based on convolutional neural networks for aircraft detection in large remote sensing images, Remote Sens. Lett., 9(1), 11–20,

425 doi:10.1080/2150704X.2017.1378452, 2018.

Zhi, X., Zhang, L. and Bai, Y.: Application of the multi-model ensemble forecast in the QPF, 2011 Int. Conf. Inf. Sci. Technol. ICIST 2011, 657–660, doi:10.1109/ICIST.2011.5765333, 2011.

Zhi ying, L. U., Zhao, Z. C., Hao, W., Lin, K. Y. and Liu, H. Z.: Multi-model Ensemble Forecast Method Based on ANN, Comput. Appl., 2004.

430 Zhu, H.: Application of Bias-removed Ensemble Mean in the Gale Forecast over the Yellow Sea and Bohai, J. Anhui Agric. Sci., 2011.

Zwally, H. J., Schutz, B., Abdalati, W., Abshire, J., Bentley, C., Brenner, A., Bufton, J., Dezio, J., Hancock, D., Harding, D., Herring, T., Minster, B., Quinn, K., Palm, S., Spinhirne, J. and Thomas, R.: ICESat's laser measurements of polar ice, atmosphere, ocean, and land, J. Geodyn., doi:10.1016/S0264-3707(02)00042-X, 2002.

435

**Code and Data availability**

**Code:**

Adaptive Forecasting Through Exponential Re-weighting algorithm:

440 Yang, Y., 2001: Combining forecasting procedures: Some theoretical results. Econom. Theory, 20, 176–222, https://doi.org/10.1017/S0266466604201086.

Structural Similarity Index Measure algorithm:

**Data:**

ACCESS1.0, ACCESS1.3 datasets:

445 Bi, D., and Coauthors, 2013: The ACCESS coupled model: description, control climate and evaluation. Aust. Meteorol. Oceanogr. J., https://doi.org/10.22499/2.6301.004.

CCSM4, CESM1 datasets:

Gent, P. R., and G. Danabasoglu, 2011: Response to Increasing Southern Hemisphere Winds in CCSM4. J. Clim., 24, 4992–4998.

450 EC-EARTH datasets:

Fichefet, T., and M. A. M. Maqueda, 1999: Modelling the influence of snow accumulation and snow-ice formation on the seasonal cycle of the Antarctic sea-ice cover. Clim. Dyn., 15, 251–268.

HadGEM2-ES, HadGEM2-CC, HadGEM2-AO datasets:

Mclaren, A. J., H. T. Banks, C. F. Durman, J. M. Gregory, and S. W. Laxon, 2006: Evaluation of the sea ice simulation in a 455 new coupled atmosphere-ocean climate model (HadGEM1). J. Geophys. Res., 111, C12014.

MIROC-ESM, MIROC-ESM-CHEM datasets:

Watanabe, M., M. Chikira, Y. Imada, and M. Kimoto, 2011: Convective control of ENSO simulated in MIROC. J. Clim., 24, 543–562.

MPI-ESM-LR，MPI-ESM-MR datasets：

460 Notz, D., F. A. Haumann, H. Haak, J. H. Jungclaus, 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.

PIOMAS datasets:

Zhang, J., and D. A. Rothrock, 2003: Modeling Global Sea Ice with a Thickness and Enthalpy Distribution Model in Generalized Curvilinear Coordinates. Mon. Weather Rev., https://doi.org/10.1175/1520-465 0493(2003)131<0845:mgsiwa>2.0.co;2.

**Author contribution**

| | |
|---|---|
| Wang Yangjun: | Methodology; Software; Writing- Original draft preparation |
| Liu Kefeng: | Writing - Reviewing and Editing |
| 470 Zhang Ren: | Supervision; Conceptualization |
| Qian Longxia: | Data curation; Validation |
| Zhang Yu | Language Editing |

**Competing interests**

475   The authors declare that no competing interest exist.