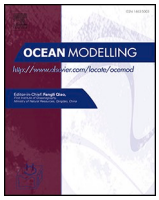
Ocean Modelling 197 (2025) 102565

Contents lists available at ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod



Challenges in improving Arctic freshwater simulations: An evaluation of CMIP6 models in the Beaufort Gyre region

Yu Zhang [a,*], Zhou Ye [a], Feifan Chen [a], Changsheng Chen [b], Guoping Gao [a], Robert C. Beardsley [c], Deshuai Wang [d], Jianhua Qi [b], Danya Xu [e], Yi Zhou [f]

[a] College of Oceanography and Ecological Science, Shanghai Ocean University, Shanghai 201306, China
[b] School for Marine Science and Technology, University of Massachusetts Dartmouth, New Bedford, MA 02744, USA
[c] Department of Physical Oceanography, Woods Hole Oceanographic, Woods Hole, MA 02543, USA
[d] First Institute of Oceanography and Key Laboratory of Marine Science and Numerical Modeling, Ministry of Natural Resources, Qingdao 266061, China
[e] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China
[f] School of Oceanography, Shanghai Jiao Tong University, Shanghai 200030, China

ARTICLE INFO

Keywords:
Freshwater content
Beaufort Gyre
CMIP6
FVCOM
Evaluation

ABSTRACT

The performance of CMIP6 models in simulating freshwater content (FWC) in the Beaufort Gyre remains unclear. This study evaluated 17 CMIP6 models using both observational and reanalysis datasets. Additionally, a global ice-ocean coupled model based on Finite Volume Community Ocean Model (Global-FVCOM) was incorporated for reference. The results revealed a significant inter-model spread among the CMIP6 models in spatiotemporal variations of FWC, with discrepancies relative to the evaluation data that were larger than those exhibited in Global-FVCOM. These discrepancies were primarily attributed to simulation errors of the salinity structure within the CMIP6 models. Over half of the models indicated that the primary source of FWC error originated from the layers above the base of halocline, where most models underestimated FWC, while others suggested the error originated from the layers between the base of the halocline and the 34.8 psu isohaline, where models tended to overestimate FWC. Based on an overall evaluation using observational and reanalysis datasets, EC-Earth3, MRI-ESM2-0, and FIO-ESM-2-0 showed better performance relative to other CMIP6 models. However, these three models, along with the multi-model mean, exhibited larger errors than Global-FVCOM, suggesting that current CMIP6 models still face challenges in FWC simulation relative to some ice-ocean coupled models. The main aspects contributing to the errors, including discrepancies in uncertainties induced by internal variability, numerical configurations, vertical mixing schemes, model resolutions, freshwater inputs, and atmospheric forcings were further discussed in this study. This study enhances understandings of CMIP6 models' capabilities to simulate FWC in the Beaufort Gyre region, providing valuable insights for future model improvements.

1. Introduction

The freshwater system of the Arctic Ocean regulates both physical and biogeochemical processes, affecting the structure of ocean currents, sea ice formation and melting, as well as the stability of marine ecosystems (White et al., 2007; Carmack et al., 2016). A key area within this system is the Beaufort Gyre, which functions as the largest reservoir, accounting for roughly 25 % of the Arctic's total freshwater storage (Haine et al., 2015). Numerous observations have been conducted in this region, with the most prominent being the Beaufort Gyre Exploration Project (BGEP) launched in 2003. Freshwater accumulation in the Beaufort Gyre region began in the 1990s. From 2003 to 2007, the interannual variation of freshwater content (FWC) displayed a pronounced positive trend, with approximately 1.7 m/year (Proshutinsky et al., 2009). The trend in liquid freshwater accumulation between 1992 and 2012 was estimated at around 600 $\pm$ 300 km^3/year, corresponding to a 30 % increase in storage (Rabe et al., 2014). The volume of liquid freshwater in the Beaufort Gyre region increased by >6400 km^3 between 2003 and 2018, representing a 40 % growth relative to the climatology of the 1970s (Proshutinsky et al., 2019). Lin et al. (2023) indicated that the FWC has plateaued from 2012 to 2019.

Although extensive observations have been conducted, including

* Corresponding author.
E-mail address: yuzhang@shou.edu.cn (Y. Zhang).

https://doi.org/10.1016/j.ocemod.2025.102565
Received 17 November 2024; Received in revised form 22 May 2025; Accepted 26 May 2025
Available online 26 May 2025

Conductivity, Temperature, Depth (CTD) profiles, moorings, and buoys, the spatiotemporal continuity of observational data in the Beaufort Gyre region remains limited. This limitation restricts accurate estimations of the Arctic's closed freshwater budget (Lique et al., 2016) and limits a clear understanding of freshwater variability. Thus, the comprehensive spatiotemporal series generated by numerical models can help mitigate this limitation and facilitate a deeper understanding of complex oceanic processes. Additionally, models assist in clarifying the major mechanisms governing freshwater variability. Current studies have highlighted the importance of wind variability and eddy-induced processes in shaping interannual freshwater changes in the Beaufort Gyre. The FWC is regulated by a balance between Ekman pumping and mesoscale eddy fluxes, with eddies acting to counteract halocline deepening induced by wind-driven convergence (Manucharyan et al., 2016; Hochet et al., 2024). Furthermore, several studies suggested that accelerated sea ice melting (McPhee et al., 2009), variations in the freshwater flux (FWF) through the Bering Strait (Rabe et al., 2013), shifts in the Eurasian river inflow pathways (Morison et al., 2012), and increased North American river discharge (Macdonald et al., 1999) can also significantly contribute to freshwater accumulation.

With climate change, if the high-pressure system over the Beaufort Sea weakens in the future, the anticyclonic circulation and Ekman pumping are expected to be reduced, which will diminish the capacity to retain large volumes of freshwater, ultimately leading to a significant release of freshwater from the Beaufort Sea. The freshwater in the Beaufort Gyre region that flows into the North Atlantic may potentially reduce regional salinity, thereby altering the hydrography, increasing stratification, weakening the Atlantic Meridional Overturning Circulation, and hindering the flow of warm Atlantic water to higher latitudes, thereby weakening the global thermohaline circulation (Dickson et al., 1988; Zhang and Steele, 2007; Buckley and Marshall, 2016). Therefore, accurately predicting freshwater changes in the Beaufort Gyre region is critical for a deeper understanding of the Arctic freshwater response to climate change and its impact on the North Atlantic. Currently, projection studies primarily rely on climate models, with the latest CMIP6 models playing a key role in exploring and predicting various climate scenarios. CMIP6 integrates a variety of models, but considerable inter-model spread in simulation capabilities remains among the different CMIP6 models. Several studies have been conducted to assess the freshwater simulation capabilities of CMIP6 models. Zanowski et al. (2021) analyzed 7 CMIP6 models across historical and future scenarios, revealing variations in freshwater storage and fluxes among the models, with a general consensus on the projection of increased liquid freshwater in the future. Wang et al. (2022) conducted a comparative study of liquid FWC, sea surface salinity, and the freshwater budget in the Arctic Ocean by analyzing differences between 31 CMIP6 models and 39 CMIP5 models. However, these studies primarily focused on broad freshwater variations in the Arctic Ocean, and they largely emphasized the model ensemble mean, resulting in a less detailed analysis of the inter-model spread. As a result, the performance of CMIP6 models in simulating freshwater in the Beaufort Gyre region remains uncertain. Furthermore, as climate models encompass complex processes and interactions among various spheres, how their performance in simulating freshwater compares to models that exclusively consider ice-ocean interactions requires further exploration.

Therefore, in this study, we analyzed historical data from 17 CMIP6 models, focusing on the Beaufort Gyre region. Model performance was evaluated based on FWC and salinity profiles, using a range of reference datasets, including observational and reanalysis data. Additionally, we explored the primary sources of error between the models and these reference datasets. Moreover, a global ice-ocean coupled model based on Finite Volume Community Ocean Model (Global-FVCOM) was included for comparison with the CMIP6 models. We also conducted a detailed analysis of the factors contributing to the differences between the CMIP6 models and the ice-ocean coupled model. This study enhances the understanding of climate models' performance in simulating freshwater, provides new insights into freshwater characteristics, and contributes to the further improvement of climate models.

The rest of the paper is organized as follows. The models, observational data, reanalysis products, and methodologies used in our evaluation of CMIP6 simulations are described in Section 2. Section 3 presents an analysis of CMIP6 model performance in simulating freshwater, focusing on FWC and salinity profile intercomparisons, along with a quantitative evaluation using various reference datasets. Section 4 discusses the key factors contributing to freshwater simulation errors. Finally, in Section 5, the conclusion about the major findings is summarized.

2. Data and methods

2.1. CMIP6 models

In this study, we utilized the historical salinity dataset provided by CMIP6. Taking into account the overlapping time period of observational data, reanalysis data, and CMIP6 historical simulations, the evaluation period was set from 2003 to 2014. Some previous studies on Arctic Ocean freshwater and salinity (Zanowski et al., 2021; Wang et al., 2022; Langehaug et al., 2023) have shown that the internal variability among ensemble members within the same model is relatively small compared to the spread across different models. Based on these findings, we focused on the single ensemble member of r1i1p1f1 (r: realization, i: initialization, p: physics, f: forcing) variant from each model. We collected the available models and selected 17 models that can generally reproduce the fundamental patterns of freshwater distribution according to our preliminary analysis. Detailed information about the selected CMIP6 models is provided in Table 1.

2.2. Global-FVCOM

In this study, we employed Global-FVCOM as a representative of ice-ocean coupled models for comparison with CMIP6 models. The ocean component is a prognostic, unstructured-grid, 3-D primitive equation, Finite Volume Community Ocean Model (FVCOM), operating within the spherical coordinate framework (Chen et al., 2003, 2006, 2007). The sea ice component is based on the Los Alamos Community Ice Code (CICE) model, which has been converted from its original structured-grid form into an unstructured-grid finite volume version (UG-CICE) and subsequently coupled with FVCOM. In this model, the grid covers the global ocean, with horizontal resolution ranging from approximately 1 km in coastal areas to 50 km in the interior regions, and 40 km in the Arctic. A hybrid terrain-following coordinate system with 45 vertical layers was utilized. In regions with water depth exceeds 225 m, the s-coordinate system was employed, featuring 10 uniform layers near the surface and 3 near the bottom, each with a thickness of 5 m. In shallow continental and coastal regions with depth $<$225 m, the σ-coordinate system was applied, maintaining the same total number of vertical layers. A transition between the two coordinate systems occurs at the 225-m isobath, where all layers have a uniform thickness of 5 m. The related information about this model is shown in Table 1.

We used data from Global-FVCOM simulations covering the period from 2003 to 2014, under conditions without data assimilation. The Global-FVCOM simulation has been running since 1978, utilizing atmospheric forcings including surface wind stress, net heat flux, shortwave irradiance, air pressure gradients, precipitation minus evaporation from the CORE-v2 dataset for the years 1978–2009, since CORE-v2 is only available up to 2009. For subsequent years, NCEP/NCAR reanalysis datasets were used (Zhang et al., 2016a, 2016b). Since CORE-v2 is generated from the NCEP/NCAR reanalysis data and both datasets share the same spatial resolution, any potential discontinuities between the two datasets are considered minimal. Global-FVCOM includes 766 rivers from various sources. Given the focus on Arctic freshwater variation, data for major rivers in the Arctic region were derived from daily or

Table 1
Information of the 17 CMIP6 models and Global-FVCOM used in this study. We adopt a nomenclature for vertical mixing schemes that is similar to that described by Heuzé et al. (2023): KPP = *K*-profile parameterization scheme (Large et al., 1994); TKE = Turbulent kinetic energy scheme (Gaspar et al., 1990); TC = turbulence closure scheme (Canuto et al., 2002, 2001); PP = Pacanowski et al. (1981); DL = Decloedt and Luther (2010); MY-2.5 = Mellor and Yamada (1982) level 2.5 turbulent closure model.

Model	Number of grid points	Number of levels	Vertical mixing	Ocean model	Sea ice model
ACCESS-CM2	360 × 300	50	KPP	ACCESS-OM2	CICE5.1.2
ACCESS-ESM1-5	360 × 300	50	KPP	ACCESS-OM2	CICE4.1
BCC-CSM2-MR	360 × 232	40	KPP	MOM4	SIS
CanESM5	361 × 290	45	TKE	NEMO3.4.1	LIM2
CESM2	320 × 384	60	KPP	POP2	CICE5
CMCC-ESM2	362 × 292	50	TKE	NEMO3.6	CICE4
EC-Earth3	362 × 292	75	TKE	NEMO3.6	LIM3
FGOALS-f3-L	360 × 218	30	TC	LICOM3.0	CICE4
FGOALS-g3	360 × 218	30	TC	LICOM3.0	CICE4
FIO-ESM-2-0	320 × 384	60	KPP	POP2-W	CICE4
HadGEM3-GC31-LL	360 × 330	75	TKE	NEMO-HadGEM3-GO6.0	CICE5.1
IPSL-CM6A-LR	362 × 332	75	TKE	NEMO-OPA	LIM3.6
MCM-UA-1-0	192 × 80	18	PP	MOM1.0	Thermodynamic ice model
MPI-ESM1-2-HR	802 × 404	40	PP	MPIOM1.63	Integrated within MPIOM and ECHAM
MRI-ESM2-0	360 × 364	61	DL	MRI.COM4.4	MRI.COM4.4
NESM3	384 × 362	46	TKE	NEMO v3.4	CICE4.1
UKESM1-0-LL	360 × 330	75	TKE	NEMO-HadGEM3-GO6.0	CICE5.1.2
Global-FVCOM	351,153 unstructured grid	45	MY-2.5	FVCOM3.1	CICE4

monthly real-time observations collected by the Arctic Great Rivers Observatory Project, the United States Geological Survey, and the Canada Water Agency. When real-time discharge records were unavailable, the climatological monthly mean data from the Global Navy Coastal Ocean Model (GNCOM) were employed (Barron and Smedstad, 2002). Additionally, astronomical tidal forcing was incorporated, accounting for eight constituents: M_2, S_2, N_2, K_2, K_1, P_1, O_1, and Q_1. The model was initialized with a 50-year spin-up field output, which was generated under climatologic meteorological forcing, river discharge, and astronomical tidal conditions (Gao et al., 2011; Chen et al., 2016). The Global-FVCOM has been extensively validated in the Arctic Ocean, including tidal elevation (Chen et al., 2009), current circulation (Chen et al., 2016; Zhang et al., 2016a; Deng et al., 2019, 2021), sea ice (Gao et al., 2011; Zhang et al., 2016b; Shen et al., 2021), and surface waves (Zhang et al., 2020).

2.3. Observational data

In this study, observational data from CTD and Ice-Tethered Profiler (ITP) measurements were employed to evaluate the CMIP6 models' performance. CTD, collected from the Beaufort Gyre Exploration Project conducted by the Woods Hole Oceanographic Institution, were derived from ship-based observations collected between July and October. This study specifically focused on the period from 2003 to 2014. ITP, deployed on sea ice, gathered temperature and salinity (T/S) profile data from beneath the ice and were supplied by the Ice-Tethered Profiler Program at the Woods Hole Oceanographic Institution. Level 3 data was utilized in this study. The data collection commenced in 2004, and the dataset used in this study spans from 2004 to 2014. The spatial distribution of the CTD and ITP data is shown in Fig. 1.

2.4. Reanalysis data

Due to the spatiotemporal limitations of the observational data, we additionally incorporated two reanalysis datasets to support the assessment.

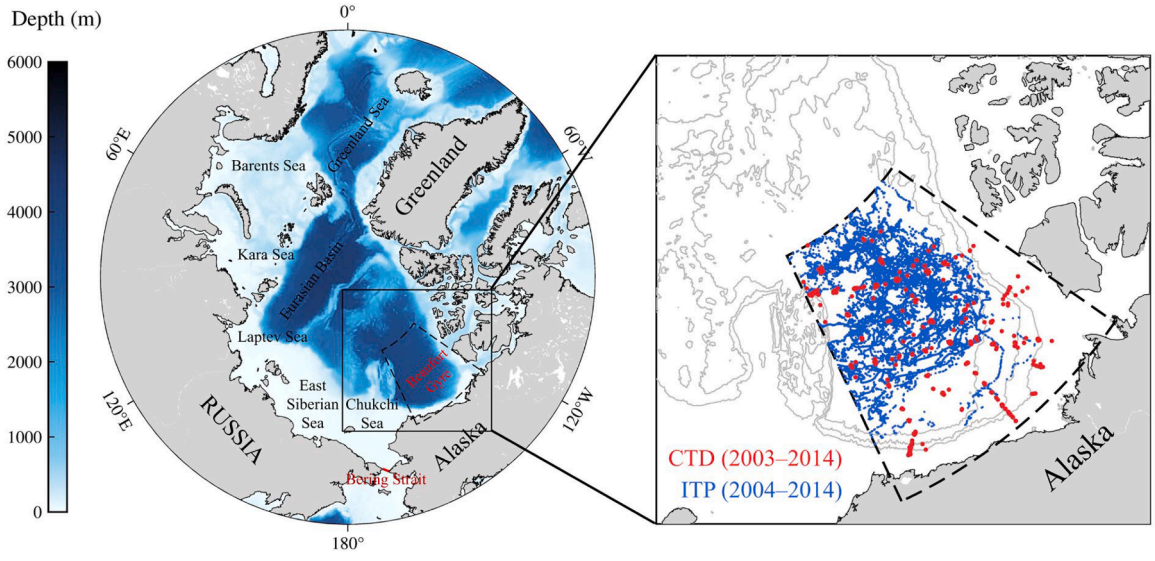


Fig. 1. Illustration of the main study area (black dashed line) and the distribution of observational data (CTD: red dots; ITP: blue dots). The red line indicates the section used to calculate the freshwater flux in the Bering Strait.

2.4.1. ORAS5

Ocean Reanalysis System version 5 (ORAS5) is utilized as an evaluation dataset to assess the performance of the CMIP6 models, with salinity data from 2003 to 2014 being employed for this analysis. ORAS5 is a global ocean reanalysis system developed and operated by the European Centre for Medium-Range Weather Forecasts (ECMWF). The system assimilates data from sea surface temperature, sea ice concentration, sea level anomaly, and in situ T/S profiles (Zuo et al., 2019). The horizontal resolution is 0.25°, and the vertical structure is divided into 75 layers, with vertical spacing increasing from 1 m at the surface to 200 m in the deep sea. ORAS5 comprises a total of five ensemble members, including one control member without perturbations and four members with perturbations in initial conditions, assimilated observations, and surface forcing fields. This study analyzed them by computing their ensemble mean.

2.4.2. TOPAZ4

Similarly, salinity data from Towards an Operational Prediction System for the North Atlantic European Coastal Zones 4 (TOPAZ4) were also used as an additional validation dataset (Sakov et al., 2012). TOPAZ4 is an ocean forecasting and reanalysis system used by the Arctic Marine Forecasting Centre (ARC-MFC), which assimilates data from sea surface temperature, sea level anomaly, sea ice concentration, sea ice thickness, sea surface salinity, and in situ T/S profiles. It bases on the Hybrid Coordinate Ocean Model (HYCOM). The resulting Arctic reanalysis product features a horizontal grid resolution of 12–16 km and consists of 50 hybrid vertical layers.

2.5. Methods of freshwater content calculation and assessment

FWC was computed using established methodologies from most previous literature including Aagaard and Carmack (1989) and Serreze et al. (2006). As shown in Eq. (1)

$$FWC = \int_{z=0}^{h} \frac{S_{ref} - S}{S_{ref}} dz \tag{1}$$

where S is the salinity of the seawater at depth z, and S_{ref} indicates the reference salinity, which is defined as 34.8 psu, a typical salinity for Atlantic Water, to maintain consistency with previous studies. h represents the depth of seawater with a salinity of 34.8 psu.

In this study, we calculated the FWC in the Beaufort Gyre region, integrating both reanalysis and observational data for a comparative assessment. The location of the Beaufort Gyre region is defined according to Polyakov et al. (2018), as shown in Fig. 1. Given the differences in spatial resolution between the models (CMIP6 and Global-FVCOM) and the reference datasets, we applied inverse distance weighting (IDW) interpolation to project the model data onto the exact grid points of the observational and reanalysis datasets. This ensures consistency in the evaluation and enables a direct point-to-point comparison. Additionally, the multi-model mean (MMM), computed as the mean of the 17 selected CMIP6 models, was used as a reference benchmark. Although the MMM does not fully disentangle model biases from internal variability, it provides a representative baseline for evaluating systematic errors across models. In Section 3, the MMM was evaluated alongside the individual CMIP6 models.

2.6. Halocline base depth estimation

Salinity profiles and halocline structures are crucial aspects for evaluating the models' capability in simulating freshwater. The base of the halocline was determined by the ratio of the density gradient caused by temperature and salinity, a method initially proposed by Bourgain and Gascard (2011) and subsequently used in several studies (Polyakov et al., 2018; Metzner and Salzmann, 2023). The ratio R is defined as follows:

$$R = \frac{\alpha \frac{\partial \theta}{\partial z}}{\beta \frac{\partial S}{\partial z}} \tag{2}$$

where θ and Srepresent potential temperature and practical salinity, respectively.α represents the thermal expansion coefficient and β denotes the haline contraction coefficient. These coefficients were calculated using the GSW Oceanographic Toolbox of TEOS-10 (Feistel et al., 2010).

As a general guideline, searching downward to a depth where the value of R exceeds 0.05 indicates the base of the halocline. The halocline in the Canada Basin exhibits complex vertical structure (Shimada et al., 2005), with some profiles affected by noise potentially leading to multiple occurrences of R values exceeding 0.05 above the base depth of the halocline. To address the issue of potentially shallow halocline base depths in such situations, this study identified the depth at which R first exceeds 0.05, searching from 300 m upwards, as the base depth of the halocline.

2.7. Primary source of FWC error

To analyze the primary source of FWC error, we divided the water column above the isohaline of 34.8 psu into two layers based on the base depth of the halocline. These two layers are referred to as the upper layer and the lower layer. The upper layer is defined as the depth between surface and base depth of the halocline, while the lower layer is defined as the depth between base depth of the halocline and the isohaline of 34.8 psu. We defined the primary source of FWC error as the layer that contributes most to the total error. This is determined by whether the overestimation or underestimation of FWC in a given layer aligns with the total FWC error. If both layers exhibit consistent overestimation or underestimation with the total FWC error, the layer contributing more is identified as the primary source.

We analyzed the primary source of FWC error across 144 months from 2003 to 2014, calculating the percentage of months in which the upper and lower layers were the dominant error source. The layer with the higher percentage was considered the primary source.

2.8. Evaluation metrics

In this study, we used the correlation coefficient (CC), bias, and root mean square error (RMSE) as statistical metrics. The equation for bias is defined as follows:

$$bias = \frac{1}{n} \sum_{i=1}^{n} (m_i - r_i) \tag{3}$$

where i indicates the i th grid point, and n represents the total number of grid points in the study area. m and r denotes model and reference data, respectively. These three statistical metrics are employed to quantitatively evaluate the models' accuracy from various perspectives (Hu et al., 2019).

Additionally, Distance between Indices of Simulation and Observation (DISO) was introduced as a comprehensive index to assess the simulation performance of the models. DISO integrates various performance metrics (Nie et al., 2023; Chen et al., 2024; Zhang et al., 2024), and it is based on the Euclidean Distance, providing a flexible approach to determining statistical metrics and their quantities (Hu et al., 2022). Here, we employed CC, bias, and RMSE to calculate DISO. The equation for DISO is defined as follows:

$$DISO_i = \sqrt{(NCC_i - 1)^2 + NBias_i^2 + NRMSE_i^2} \tag{4}$$

where i=0, 1, …, m, with i=0 corresponding to the validation data and i=m indicating the total number of models. The abbreviations *NCC*, *NBias*, and *NRMSE* refer to the normalized forms of CC, bias, and RMSE,

respectively. These metrics are normalized to a range between 0 and 1, according to the following normalization formula:

$$NS_i = \frac{S_i - \min(S)}{\max(S) - \min(S)} \quad (5)$$

where S denotes metrics such as CC. A lower DISO value indicates superior performance.

3. Result

In this section, we first intercompared the spatiotemporal variations in FWC among the CMIP6 models, including comparisons with an ice-ocean coupled model and reanalysis data. Similarly, further analysis focused on salinity profiles and their contributions to FWC errors. Finally, a quantitative evaluation of FWC was performed to evaluate the models in comparison with both reanalysis and observational data.

3.1. Freshwater content intercomparisons

Generally, there was significant inter-model spread among the 17 CMIP6 models in both the temporal evolution and spatial distribution of FWC, with spatiotemporal mean FWC ranging from 10 to 48 m (Fig. 2a). The mean FWC of MMM exceeded the spatiotemporal mean FWC of the two reanalysis datasets (ORAS5 and TOPAZ4) by 0.54 m and 1.51 m, respectively. Among these models, 9 surpassed the FWC values of ORAS5, and 11 exceeded those of TOPAZ4. As a reference, the ice-ocean coupled model Global-FVCOM underestimated FWC, exhibiting differences of −1.42 m and −0.44 m relative to ORAS5 and TOPAZ4. Regarding the FWC of MMM, it was closer to ORAS5, while Global-FVCOM showed greater alignment with TOPAZ4.

Further analysis was conducted to explore the primary spatiotemporal sources contributing to the differences in FWC among the CMIP6 models. For each model, the FWC bias relative to ORAS5 and TOPAZ4 exhibited little variation across months or years, indicating that the performance of CMIP6 models remained relatively stable over different timescales (Fig. 2b and Fig. 2c). The biases of 17 CMIP6 models relative to the two reanalysis datasets ranged from −9.13 m to 27.93 m and −8.16 m to 28.91 m, respectively. In terms of monthly variation, most CMIP6 models exhibited similar patterns to ORAS5 and TOPAZ4. Both ORAS5 and TOPAZ4 showed that the minimum FWC occurred in April–May, while their estimates of the maximum FWC differed, with ORAS5 showing it in September–October and TOPAZ4 in October–November. 9 CMIP6 models, along with Global-FVCOM, exhibited the pattern consistent with ORAS5, while none of the models successfully reproduced the FWC cycle of TOPAZ4. Regarding interannual variation, ORAS5 showed that FWC reached its maximum in 2008 and its minimum in 2004, while TOPAZ4 indicated that the maximum and minimum FWC occurred in 2009 and 2004, respectively. However, none of the CMIP6 models fully captured this pattern. As a reference, Global-FVCOM exhibited the characteristic consistent with ORAS5. The comparison results revealed that inter-model spreads in spatiotemporal mean FWC were not primarily attributed to a specific time period, but rather manifested as consistent patterns maintained by individual models throughout the study period. This persistence reflects a systematic bias inherent to the models, with MCM-UA-1-0 producing the largest FWC estimates and FGOALS-g3 having the smallest.

In terms of spatial distribution, the two reanalysis datasets indicated

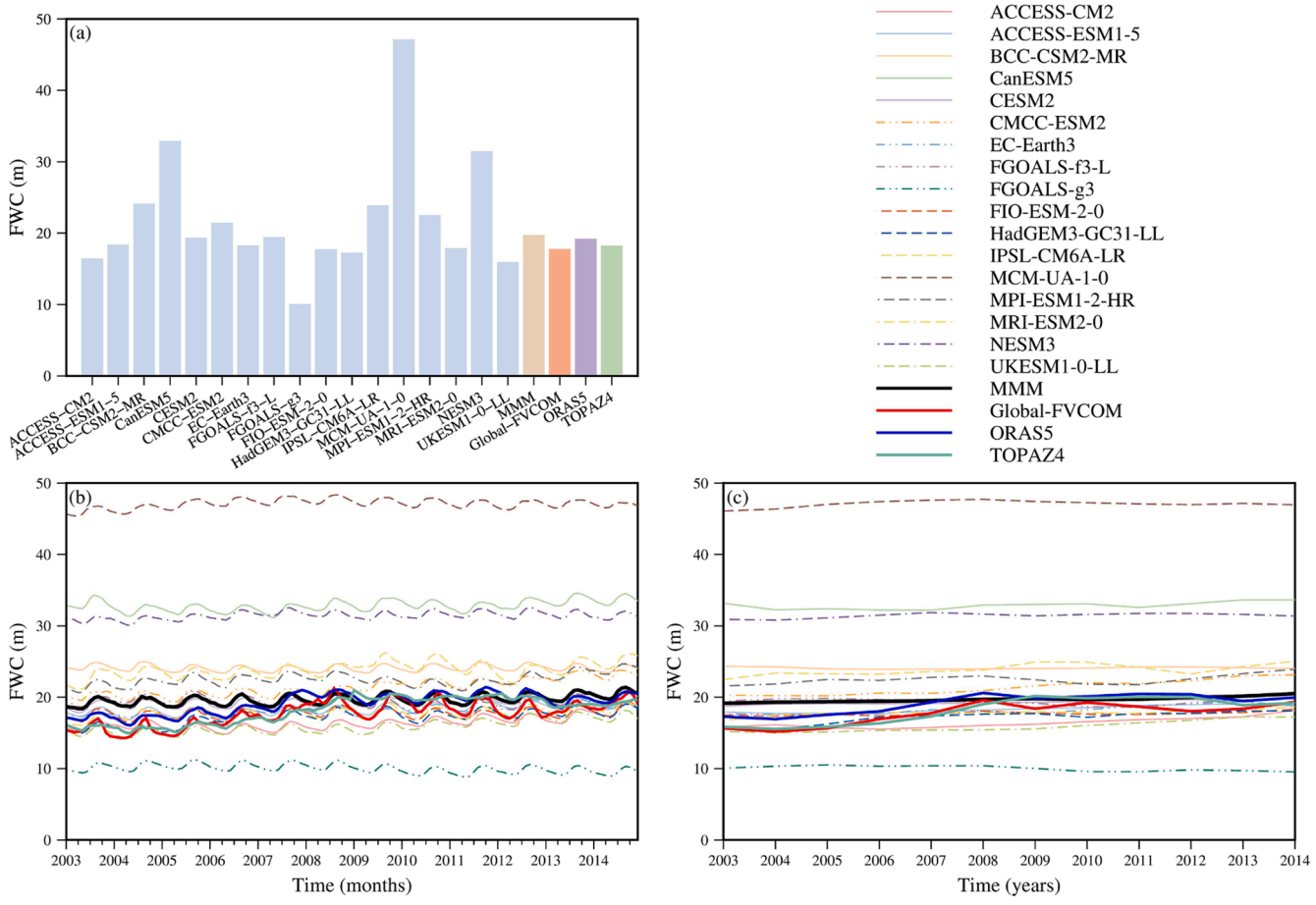


Fig. 2. Comparisons of (a) spatiotemporal mean, (b) monthly variation, and (c) yearly variation of FWC among 17 individual CMIP6 models, the MMM, the ice-ocean coupled model and two reanalysis data from 2003 to 2014.

that regions with higher FWC were located in the southwestern part of the study area, with maximum values ranging from 21 to 23 m (Fig. 3). The simulation error in FWC from 17 CMIP6 models were primary attributed to differences in the spatial distribution characteristics, as most models exhibited higher FWC in the northwestern regions. CanESM5, MCM-UA-1-0, and NESM3 displayed significantly overestimated FWC values, exceeding 30 m across much of the area. In contrast, FGOALS-g3 exhibited the lowest values across the entire region, remaining below 14 m. The MMM retained the characteristics of most CMIP6 models, including a high FWC region in the northwest, where the high values were comparable to those in ORAS5. Unlike the CMIP6 models, Global-FVCOM showed a FWC distribution similar to the reanalysis data, with high values of 20 m located in the southwest.

3.2. Salinity profile intercomparisons

The disparities in FWC among the 17 CMIP6 models are attributed to differences in the vertical salinity structure. Both two reanalysis datasets (ORAS5 and TOPAZ4) primarily indicated pronounced vertical salinity gradients within the 0–500 m depth range, exhibiting relatively strong stratification and thin mixed layer in the near surface layer (Fig. 4). The isohaline of 34 psu was approximately located at a depth of 210 m. Most CMIP6 models also displayed noticeable vertical salinity gradients, but the stratification of most models was generally weaker than that observed in the reanalysis data, except for IPSL-CM6A-LR, and 15 of the CMIP6 models exhibited a deeper isohaline at the 34 psu depth compared to the reanalysis data. The salinity profile characteristics of the MMM were similar to those of the majority of CMIP6 models. Global-FVCOM also exhibited notable vertical gradients at the surface layer. Compared to most CMIP6 models, its salinity gradients and the depth of the isohaline of 34 psu were more comparable to those of the two reanalysis datasets.

A quantitative comparison of salinity profile errors between the CMIP6 models and the Global-FVCOM, relative to the two reanalysis datasets, was further presented (Figs. 5 and 6). In the comparison with ORAS5, all models except IPSL-CM6A-LR overestimated surface layer salinity (Fig. 5a). Above 400 m, significant variations in salinity discrepancies between the models and reanalysis data were observed with increasing depth. Above 50 m, most models exhibited an overestimation of salinity. From 50 to 600 m in depth, underestimations were dominated, indicating a shift from positive to negative salinity discrepancies. Below 600 m, the discrepancies gradually stabilized, with errors remaining within the range of −0.5 to 0.5, and the MMM exhibited a slight underestimation, with errors less than −0.06. Global-FVCOM also exhibited an error shift from positive to negative with depth. Compared to the CMIP6 models, the differences between Global-FVCOM and ORAS5 began to stabilize at a shallower depth of approximately 550 m, with the errors less than −0.05.

Vertical depth of notable salinity variation were primarily located above and within the halocline, which are key regions influencing FWC changes. In this study, the FWC errors from the upper layer above the base depth of the halocline and the lower layer between the base depth of the halocline and the isohaline at the 34.8 psu depth are assessed for each model relative to the reanalysis data, and the primary sources of FWC error are identified.

The total error in the water column above the isohaline of 34.8 psu varied across months, with 4 CMIP6 models consistently overestimating and another 4 models persistently underestimating (Fig. 5b). In the upper layer, 14 CMIP6 models and Global-FVCOM underestimated FWC, with the maximum underestimation reaching approximately −11 m (Fig. 5c). In the lower layer, CMIP6 models, Global-FVCOM, and the MMM generally overestimated FWC, with the maximum overestimation reaching up to 20 m (Fig. 5d). Regarding the primary sources of FWC error, 11 CMIP6 models and FVCOM indicated that the primary source of error originated from the upper layer, while 6 CMIP6 models and the MMM showed that the lower layer dominated (Fig. 5e). Among the models where the upper layer was the dominant source, 10 CMIP6 models and Global-FVCOM underestimated FWC, whereas in those models where the lower layer was dominant, all the models and the MMM overestimated FWC. This suggested that FWC overestimation was primarily driven by the lower layer, whereas underestimation was predominantly influenced by the upper layer.

In comparison with TOPAZ4, the magnitude of salinity difference transitioned from positive to negative more sharply across the models (Fig. 6a). All models except IPSL-CM6A-LR overestimated near surface salinity, this overestimation was mainly concentrated within the upper 20 m. Within the depth range of 20 to 450 m, most models exhibited underestimation. At depths below 450 m, the differences gradually stabilized, with errors confined within the range of −0.43 to 0.47. The MMM displayed error characteristics similar to those of the 17 CMIP6 models, with underestimation stabilizing at approximately −0.12 starting from around 450 m. Global-FVCOM performed notably well, exhibiting a more pronounced change in error from positive to negative with depth, and the discrepancy with TOPAZ4 began to stabilize around 300 m, with errors gradually approaching zero.

The primary sources of error relative to TOPAZ4 were further analyzed. The temporal variation of FWC errors in the total water column above the isohaline of 34.8 psu, as well as in the upper layer and lower layer, was similar to that of ORAS5 (Fig. 6b–d). In a similar manner, 9 CMIP6 models and Global-FVCOM identified the upper layer as the primary source of error, whereas 8 CMIP6 models and the MMM indicated that the lower layer was dominant. The comparison with TOPAZ4 also indicated that the overestimation of FWC was primarily attributed to the lower layer, while the underestimation was chiefly influenced by the upper layer (Fig. 6e).

3.3. Quantitative evaluation

The above section primarily highlighted the similarities and intermodel spread among CMIP6 models through intercomparison, as well as their discrepancies with the ice-ocean coupled model. A quantitative analysis was conducted to further evaluate the performance ranking of these 17 CMIP6 models, with a focus on listing the top three models that demonstrated superior performance.

The evaluation of the 17 CMIP6 models' and Global-FVCOM's performance against observations using CTD data is shown in Fig. 7. Significant performance variability was observed among the 17 CMIP6 models, with CC ranging from a minimum of −0.15 to a maximum of 0.61, RMSE varying from 3.39 m to 31.54 m, and bias spanning from −9.91 m to 30.24 m. The models performed differently across various metrics: NESM3, CanESM5, and FIO-ESM-2-0 achieved relatively high CC, CESM2, EC-Earth3 and MRI-ESM2-0 demonstrated smaller RMSE, and CESM2, FGOALS-f3-L and ACCESS-ESM1-5 exhibited comparatively small bias. Bias analysis indicated that 10 CMIP6 models tended to underestimate FWC. The MMM displayed a lower CC but exhibited relatively small RMSE and bias. Global-FVCOM outperformed all CMIP6 models in terms of both CC and RMSE, with only 4 models and the MMM exhibiting a bias superior to that of Global-FVCOM. When assessing the 17 CMIP6 models using ITP data, the models demonstrated considerable variability in metric values (Fig. 8). Most models did not achieve higher CC values with ITP data compared to those with CTD data. More than half of the models displayed reduced RMSE, and the majority showed smaller bias. Similar to the evaluation with CTD data, NESM3 and FIO-ESM-2-0 maintained relatively high CC, CESM2, EC-Earth3, and MRI-ESM2-0 continued to exhibit small RMSE, and CESM2 and ACCESS-ESM1-5 retained small bias. Furthermore, EC-Earth3 achieved both a higher CC and smaller bias. The MMM showed a low CC but relatively small RMSE and bias. Global-FVCOM exhibited superior performance compared to all CMIP6 models in both CC and RMSE, with only the MMM showing a smaller bias than Global-FVCOM.

In addition to the observational data, reanalysis data was utilized to further assess their performance. When evaluated against ORAS5, the 17

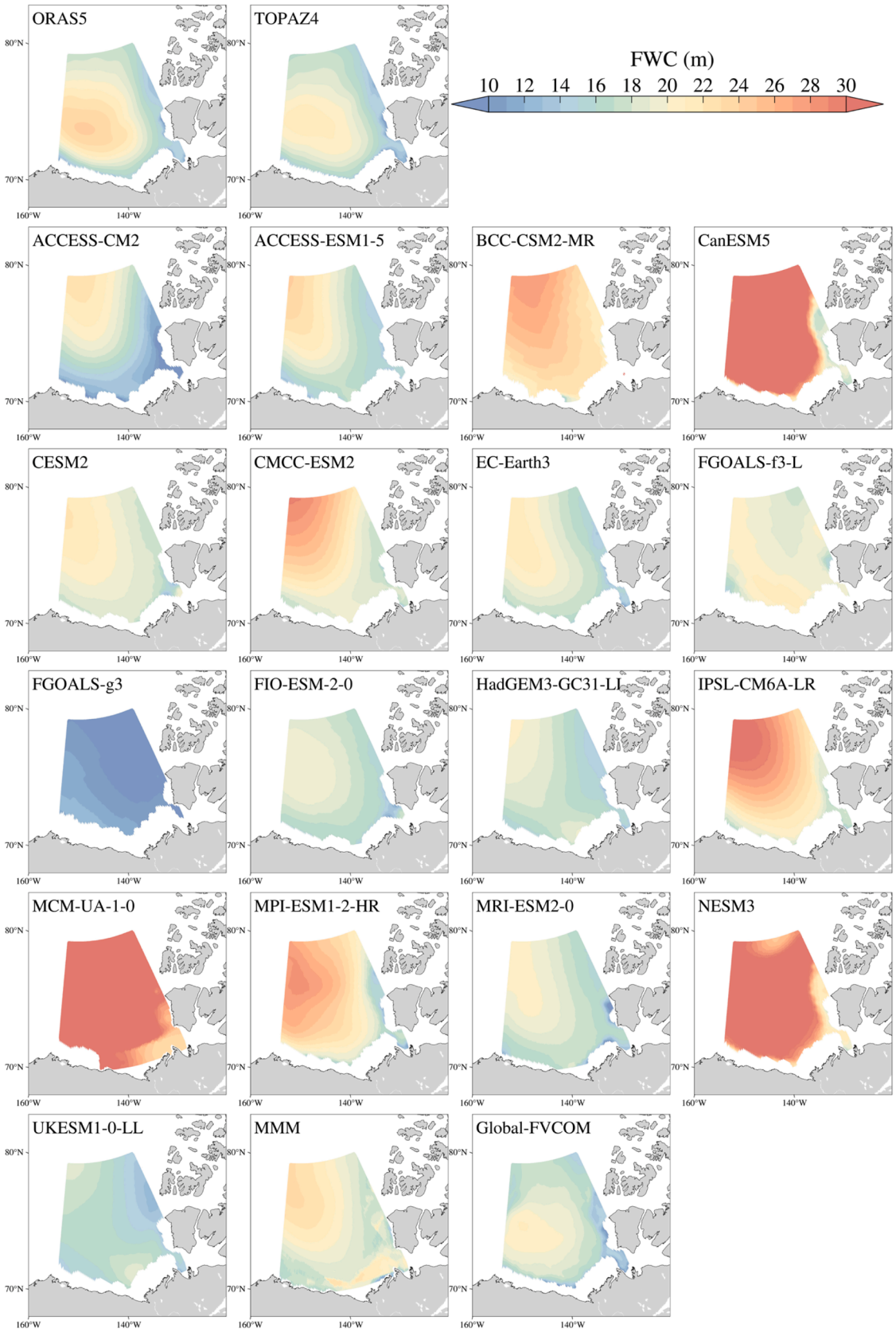


Fig. 3. Spatial distribution of the multi-year mean FWC in the Beaufort Gyre over the period 2003–2014 among two reanalysis datasets, 17 individual CMIP6 models, the MMM and the ice-ocean coupled model.

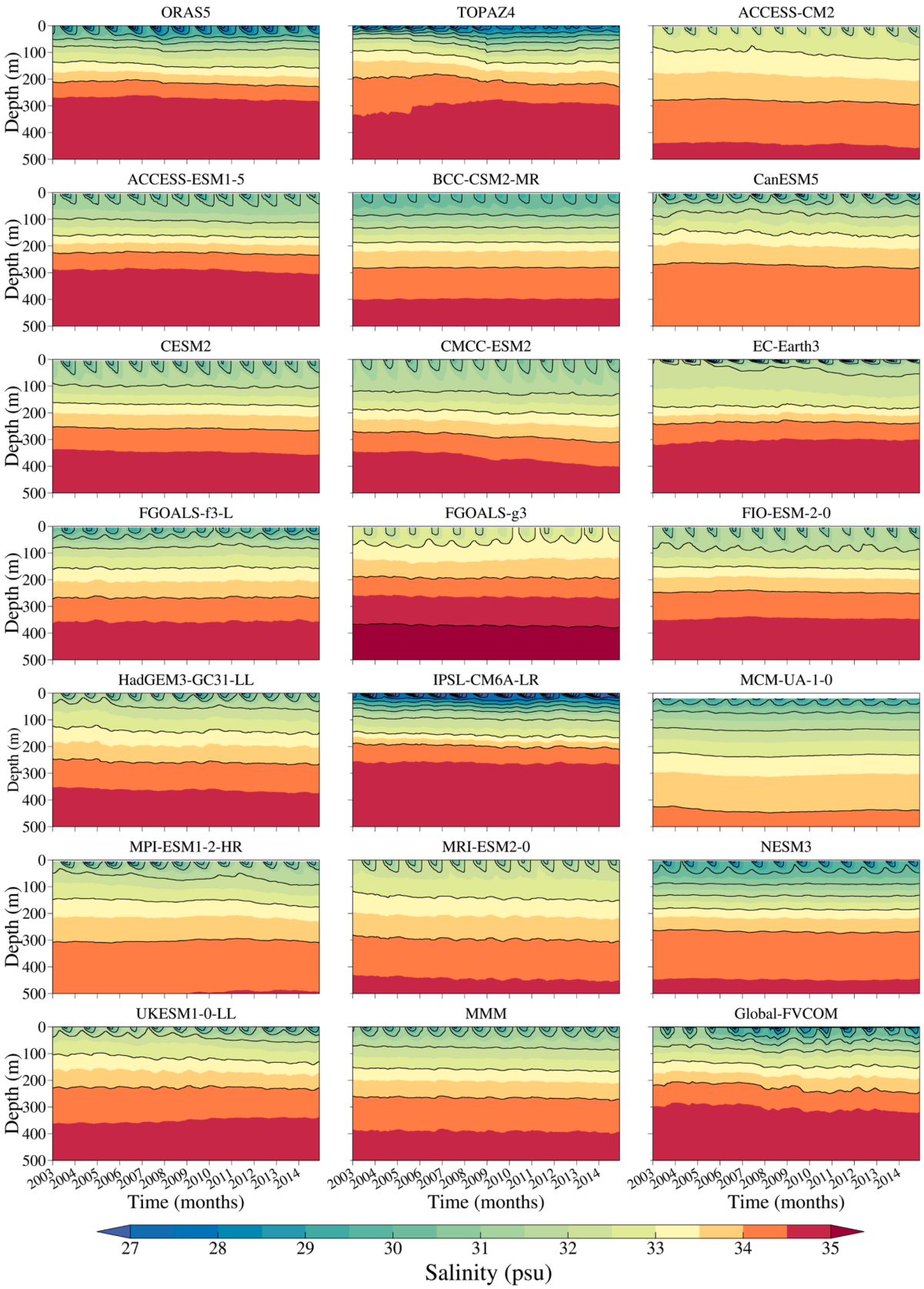


Fig. 4. Salinity contour diagrams (depth vs. time) at various depths for two reanalysis datasets, 17 individual CMIP6 models, the MMM and the ice-ocean coupled model.

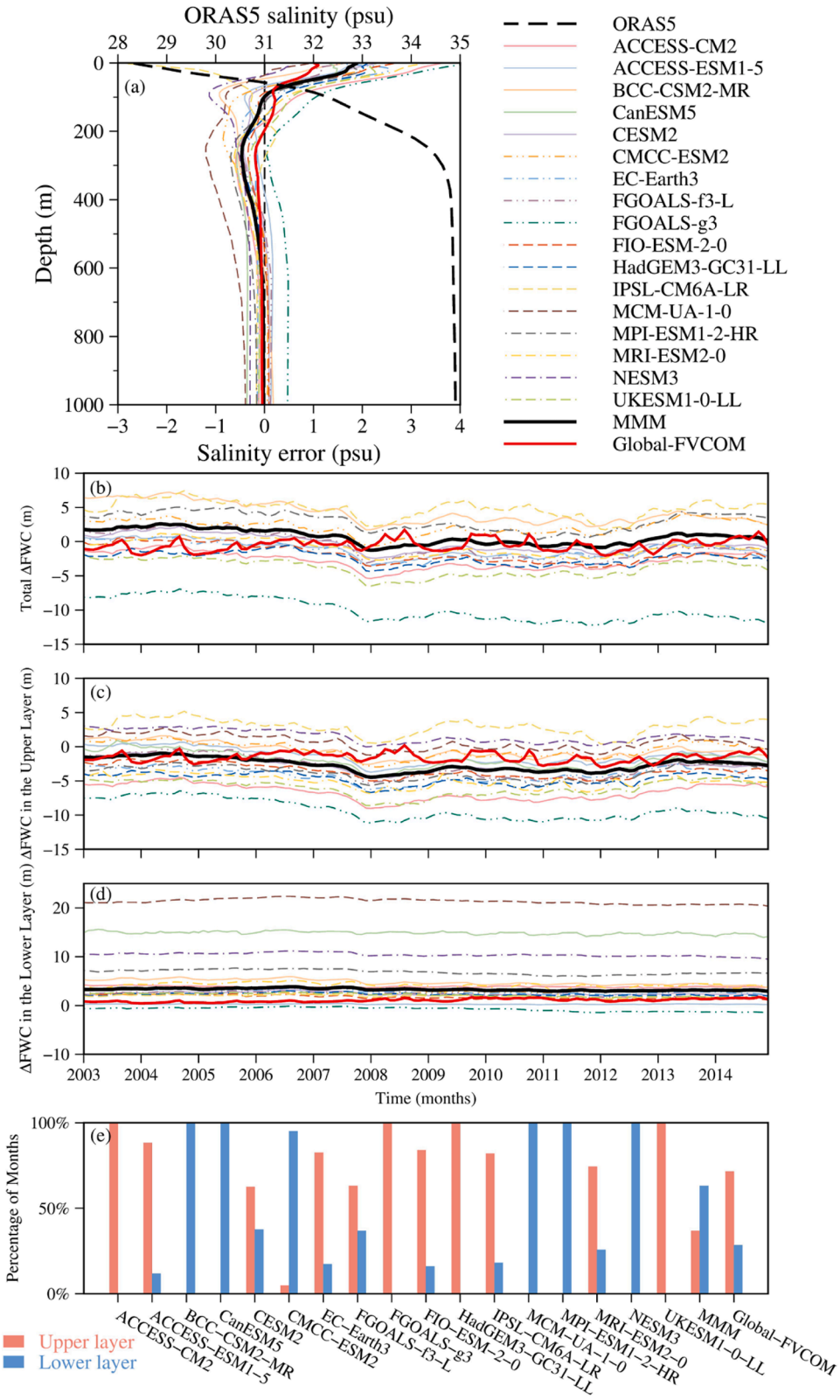


Fig. 5. (a) The average salinity profiles of the Beaufort Gyre region from 2003 to 2014 among ORAS5, 17 individual CMIP6 models, the MMM and the ice-ocean coupled model. The thick black dashed line represents the salinity of ORAS5. The other lines indicate the salinity error for each model defined as the model salinity minus the ORAS5 salinity. (b–d) Errors in FWC for CMIP6 models and Global-FVCOM relative to ORAS5 in the total water column above the isohaline of 34.8 psu, upper layer and lower layer, respectively. (e) Percentage of months where the upper layer (orange) and lower layer (blue) were the primary source of error, respectively.

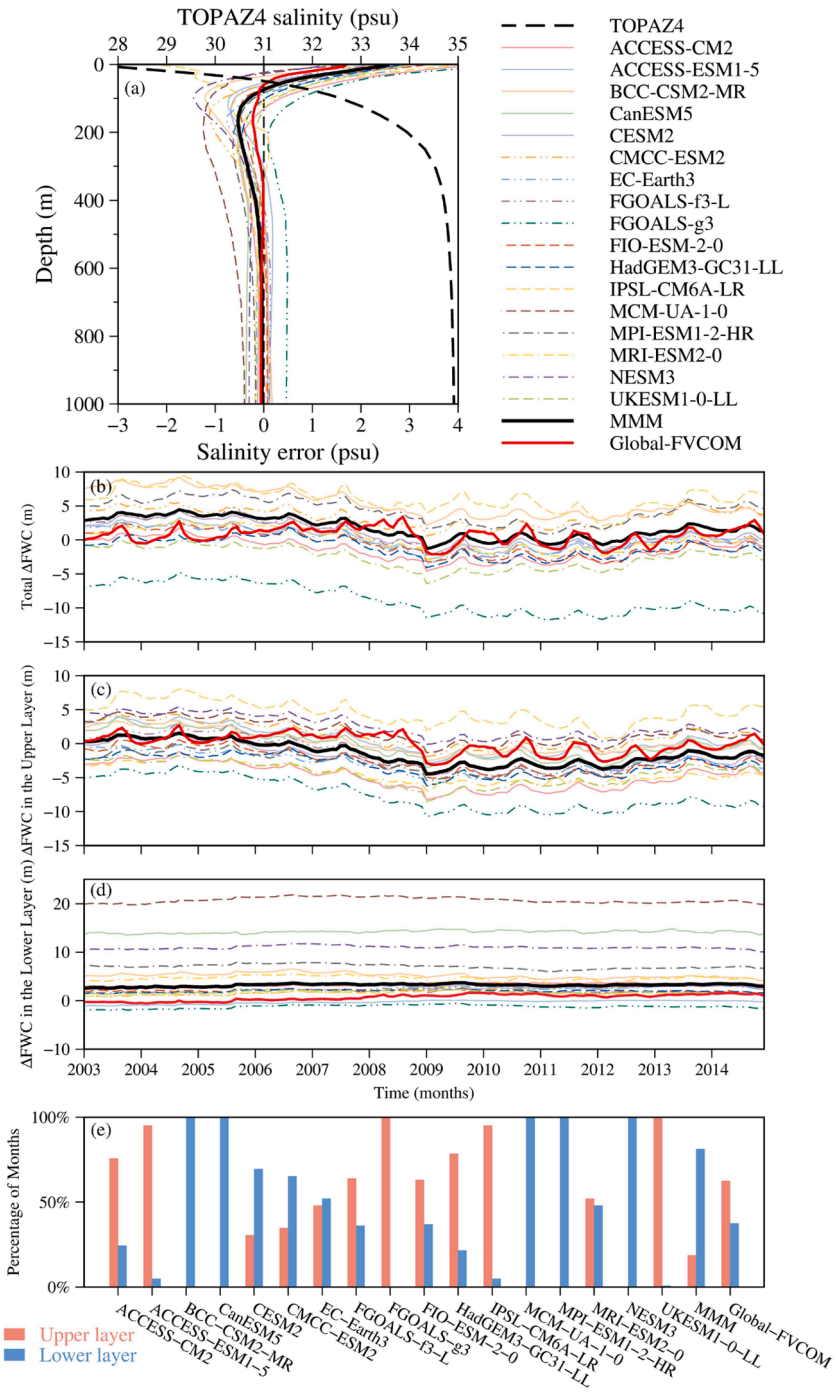


Fig. 6. Same as Fig. 5 except with TOPAZ4 instead of ORAS5.

CMIP6 models exhibited CC ranging from 0.19 to 0.71, with 4 models showing RMSE and 3 models displaying bias exceeding 10 m (Fig. 9). EC-Earth3, CESM2, and MRI-ESM2-0consistently displayed smaller RMSE. EC-Earth3, MRI-ESM2-0, and NESM3 demonstrated relatively high CC, while CESM2, FGOALS-f3-L, and EC-Earth3 exhibited smaller bias. The MMM exhibited lower CC than most CMIP6 models but showed smaller RMSE and bias. Compared to Global-FVCOM, all 17 models had lower CC, with only EC-Earth3 showing RMSE and bias, along with the MMM's bias, smaller than Global-FVCOM.

When evaluating against TOPAZ4, the models exhibited some differences in performance compared to ORAS5, with most models showing lower CC, while the 3 models with higher CC remained consistent with the results observed in the ORAS5 evaluation (Fig. 10). EC-Earth3, MRI-ESM2-0, and FIO-ESM-2-0 exhibited smaller RMSE, while MRI-ESM2-0, ACCESS-ESM1-5, and EC-Earth3 demonstrated smaller bias. Additionally, FIO-ESM-2-0 showed a relatively small RMSE, while MRI-ESM2-0 and ACCESS-ESM1-5 exhibited smaller bias. The MMM still demonstrated low CC but achieved relatively small RMSE and bias. Relative to Global-FVCOM, all 17 models had lower CC and larger RMSE, with only 4 CMIP6 models showing smaller bias.

Given the significant differences in CC, RMSE, and bias among the models, this study employed DISO for a comprehensive performance evaluation across different reference datasets. Among the CMIP6 models, EC-Earth3 and MRI-ESM2-0 consistently ranked within the top three across evaluations using various reference datasets (Fig. 11). In the overall performance assessment, EC-Earth3 ranked highest, followed by

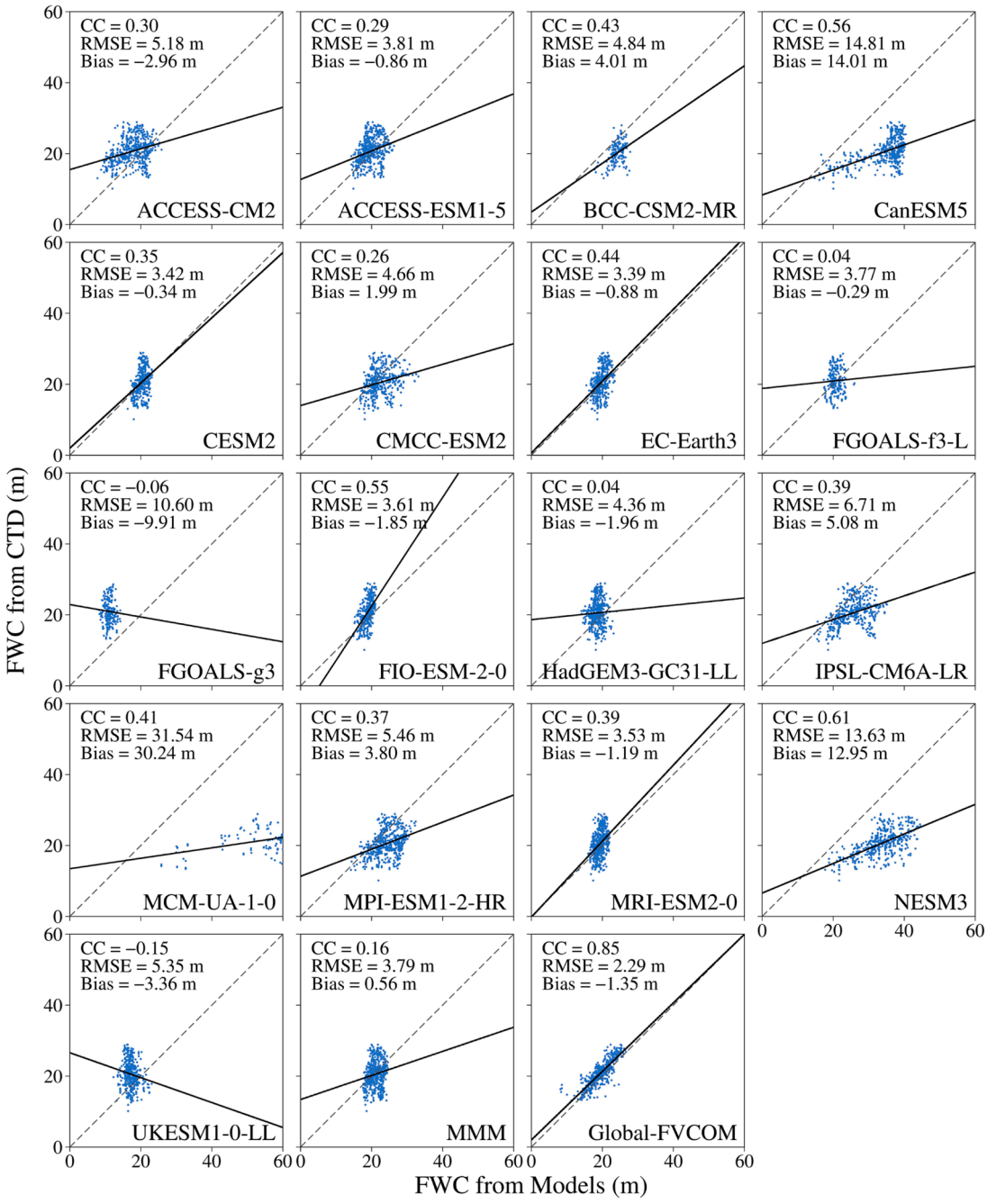


Fig. 7. Evaluation of FWC on CTD observations among the CMIP6 models, the MMM, and the ice-ocean coupled model. The black line indicates the optimal fit, while the red line represents the fit between the FWC of the models and CTD.

MRI-ESM2-0 and FIO-ESM-2-0. However, their DISO scores remained higher across all evaluations when compared with Global-FVCOM, indicating that Global-FVCOM exhibited relatively smaller errors than the CMIP6 models. These findings suggested that the CMIP6 models' ability to simulate freshwater in the Beaufort Gyre remained limited and required further improvement.

4. Discussion

The evaluation above revealed that the current CMIP6 models exhibited considerable errors when compared to the ice-ocean coupled model, Global-FVCOM, which served as a reference. We further discussed the major factors contributing to these errors. Generally, the freshwater simulation errors arise from multiple interdependent factors, including inherent uncertainties induced by internal variability among ensemble members within the same model, structural differences in numerical configurations, divergent parameterization of vertical mixing schemes, varying model resolutions, inconsistencies in freshwater inputs, and differences in internally simulated atmospheric conditions.

Although some previous studies (Zanowski et al., 2021; Wang et al., 2022; Langehaug et al., 2023) have validated the effectiveness of using a single ensemble member to characterize model performance, the quantitative impact of uncertainty induced by internal variability across ensemble members on model evaluation remains a critical focus. To address this, we systematically selected multiple ensemble members from the models employed in this study. Notably, eight models include <6 ensemble members, while the remaining nine models provide >10 ensemble members. To ensure analytical consistency and mitigate

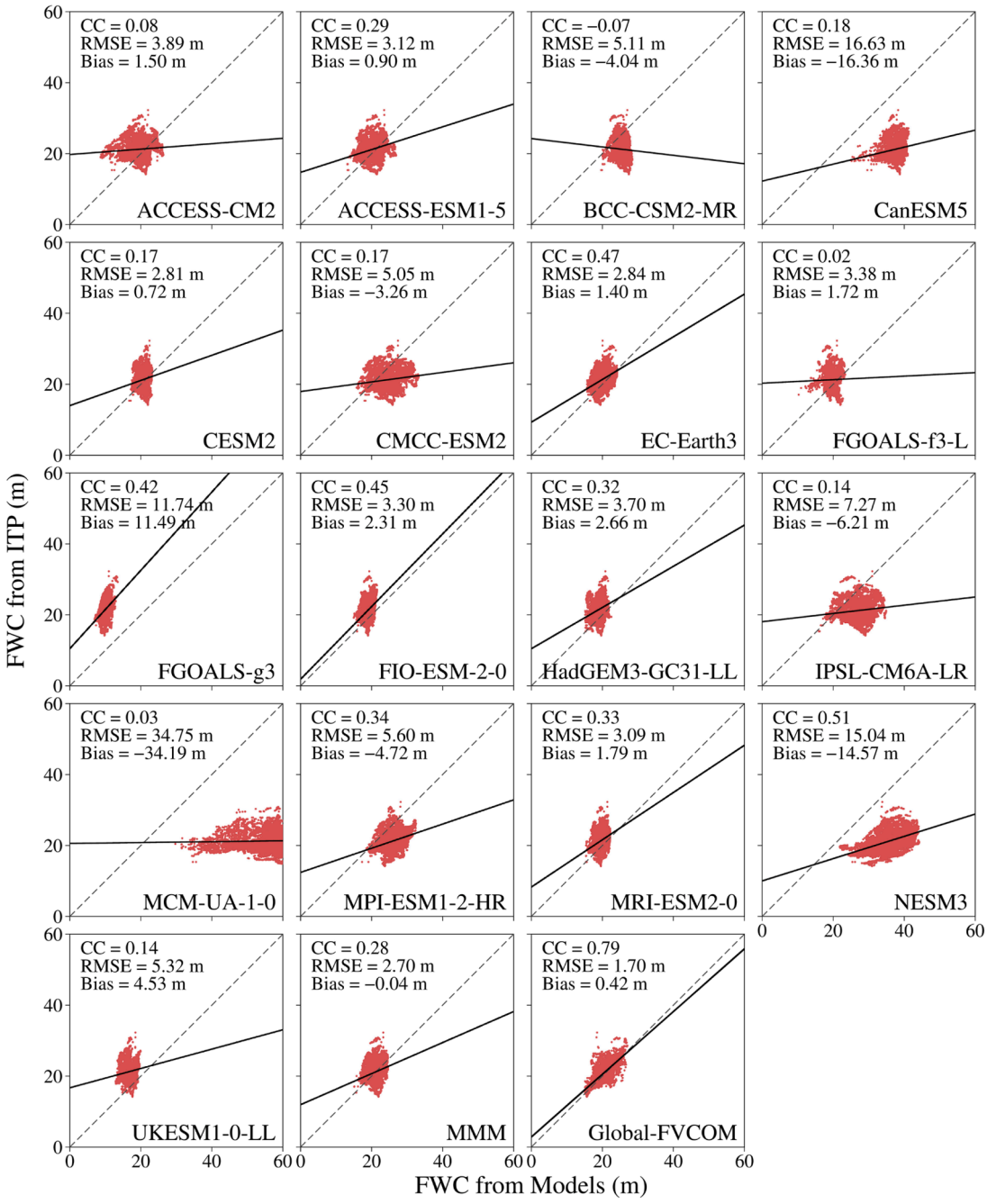


Fig. 8. Evaluation of FWC on ITP observations among the CMIP6 models, the MMM, and the ice-ocean coupled model. The black line indicates the optimal fit, while the red line represents the fit between the FWC of the models and ITP.

potential biases arising from differences between small- ensemble and large-ensemble simulations, we set a uniform threshold of 10 ensemble members per model. For models with >10 members, we selected first 10 representative realizations, while models with <10 members were excluded from the analysis (Table 2). The comparative analysis demonstrates that although internal variability among ensemble members introduces some differences in FWC simulations, the variations among ensemble members within the same model are indeed much smaller than the spreads across different models (Fig. 12). For all models, the ensemble mean statistically remains very similar to single ensemble member. This indicates that discrepancies between models and observations predominantly arise from systematic errors inherent in each model, rather than internal variability. Consequently, a single representative ensemble member suffices to capture a model's general performance in the Arctic freshwater simulation.

Additionally, the choice of ocean model plays a critical role in simulating Arctic FWC, and systematic errors within the numerical configuration can substantially influence simulation outcomes. Pan et al. (2023) highlighted that climate models employing the same ocean model family often exhibit common biases, and inherent limitations in ocean model frameworks can significantly affect the MMM results. Among the 17 CMIP6 models, the primary ocean model families represented were NEMO, POP, MOM, and LICOM (Table 1). In this study, the comparative results indeed revealed some commonalities. For instance, the LICOM and MOM ocean model family consistently underestimated and overestimated FWC, respectively. Both the ACCESS-OM2 family and 4 models within the NEMO family (CMCC-ESM2, EC-Earth, HadGEM3-GC31-LL, IPSL-CM6A-LR) showed a region of high FWC in the

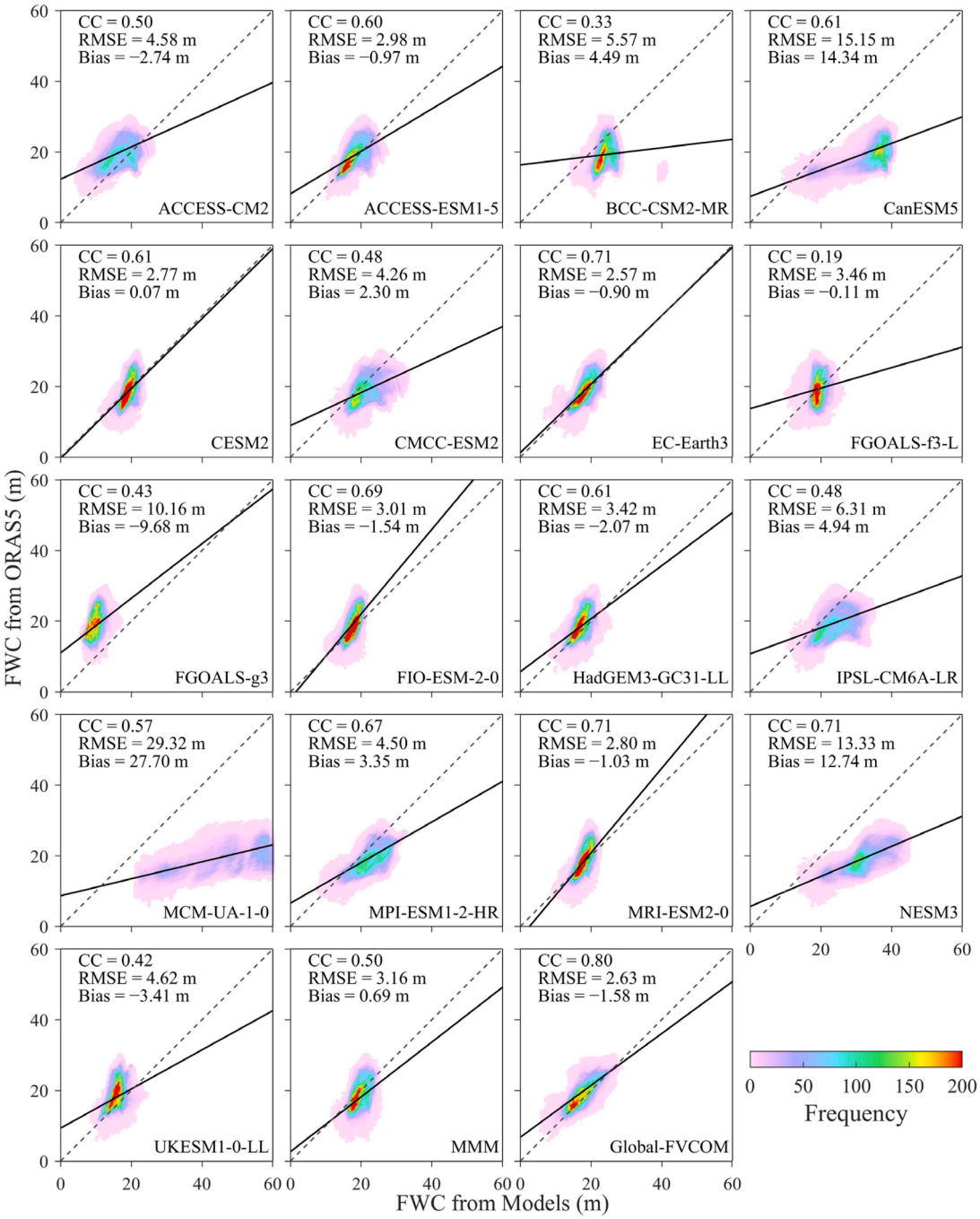


Fig. 9. Evaluation of FWC on ORAS5 among the CMIP6 models, the MMM, and the ice-ocean coupled model. The black line indicates the optimal fit, while the blue line represents the fit between the FWC of the models and ORAS5.

northwest. In addition, the LICOM and ACCESS-OM2 ocean model families identified the upper layer as the primary source of error, whereas the MOM family attributed the dominant errors to the lower layer.

The selection of vertical mixing schemes is essential for the accuracy of freshwater simulation. In this study, the 17 CMIP6 models incorporate various mixing schemes such as KPP, TKE, TC, PP, and DL, whereas Global-FVCOM employs the MY-2.5 mixing scheme (Table 1). These vertical mixing schemes lead to variations in both the mixed layer and the halocline among different models, thereby influencing the overall structure of the salinity profile and, consequently, the simulation of FWC.

Model resolution is also a critical factor that can potentially affect the accuracy of its simulation results. Coarse horizontal resolution may fail to accurately simulate intricate oceanic processes and capture meso- and sub-mesoscale features. For instance, among the 17 CMIP6 models, MCM-UA-1-0, FGOALS-f3-L, FGOALS-g3, and BCC-CSM2-MR, which had relatively low horizontal resolution (Fig. 13), showed the largest errors in the overall assessment (Fig. 11). The horizontal resolution of other CMIP6 models is relatively low compared to Global-FVCOM, and they also exhibited greater errors in freshwater simulation. In terms of vertical resolution, the 4 CMIP6 models with the weakest performance also feature the lowest vertical resolution. Moreover, the top three CMIP6 models in overall performance, EC-Earth3, MRI-ESM2-0, and FIO-ESM-2-0, had higher vertical resolution compared to other models with similar horizontal resolution, highlighting the critical importance

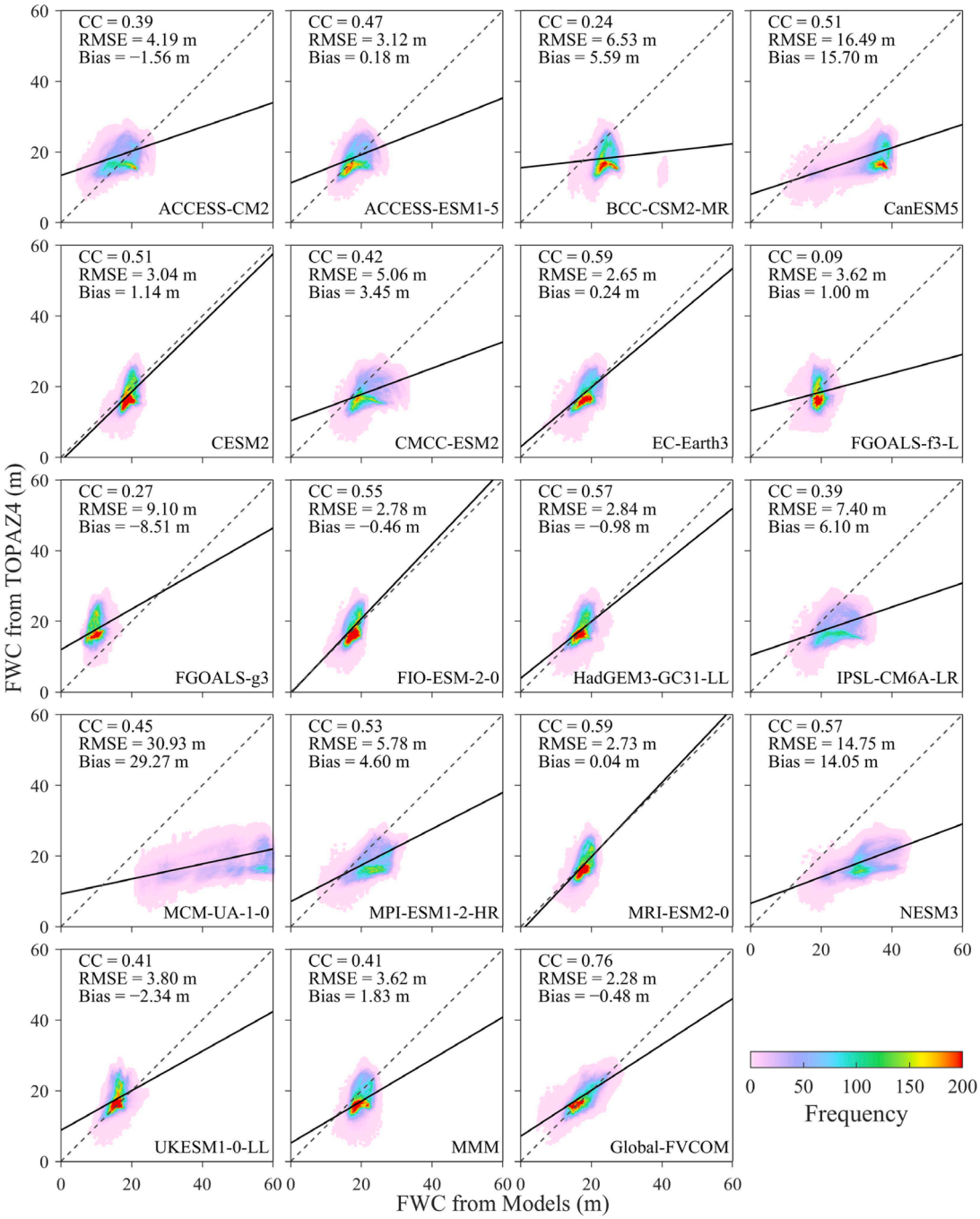


Fig. 10. Evaluation of FWC on TOPAZ4 among the CMIP6 models, the MMM, and the ice-ocean coupled model. The black line indicates the optimal fit, while the blue line represents the fit between the FWC of the models and TOPAZ4.

of vertical resolution.

Furthermore, differences in freshwater input can significantly affect the simulation of freshwater in the Beaufort Gyre. Freshwater input consists of solid and liquid freshwater, with sea ice representing the source of solid freshwater. The formation and melting of sea ice can alter the salinity of the surface layer. Sea ice represents a crucial source of freshwater. The sea ice models in the 17 CMIP6 models primarily utilize CICE, SIS, and LIM, with CICE being the most widely adopted (Table 1). The specific versions of CICE employed by these models also vary. Differences in sea ice formation and melting processes exist across various sea ice models (Long et al., 2021) and even among different versions of the same model. The studies by Shu et al. (2020), Watts et al. (2021), Chen et al. (2023), and Wang et al. (2023) all indicated that, although CMIP6 models exhibit similarities in simulating historical sea ice variations, significant inter-model spread persists. It is suggested that this spread leads to differences in the contribution of sea ice as a freshwater input in the Beaufort Gyre region. For Global-FVCOM, some studies have been conducted to validate sea ice, demonstrating good performance in the climatological means and real-time series of sea ice concentration, extent, thickness, and velocity (Gao et al., 2011; Zhang et al., 2016b; Shen et al., 2021). Therefore, this accuracy in sea ice simulation facilitates a more precise estimation of the freshwater contribution from sea ice.

The freshwater inflow through Bering Strait is another important source of freshwater input. We calculated the FWF through the Bering Strait in each model and compared it with the observational FWF from

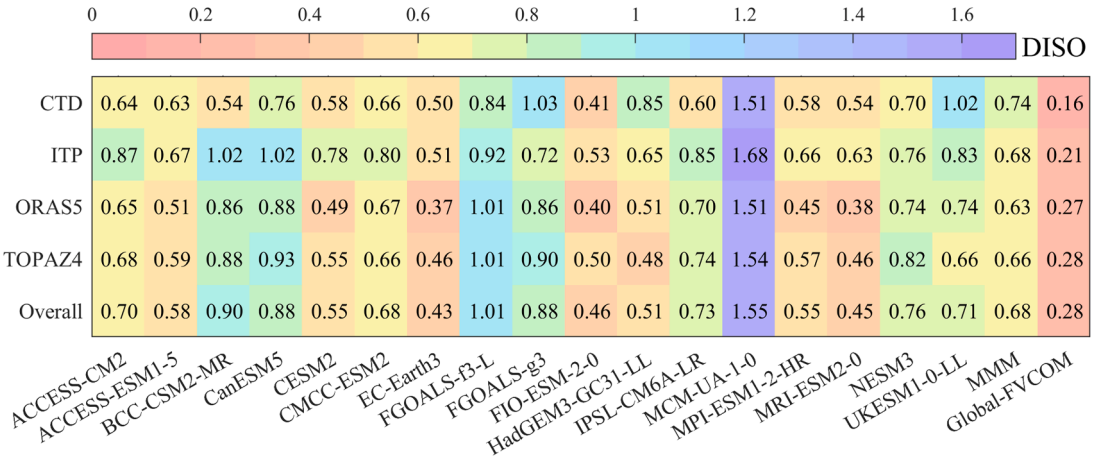


Fig. 11. Comprehensive assessments of models based on DISO values against various reference datasets. "Overall" denotes the overall performance with the combination of various reference datasets.

Table 2
Summary of selected ensemble members in the CMIP6 models.

Model	Number of members	Information of ensemble members
ACCESS-CM2	10	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r8i1p1f1, r9i1p1f1, r10i1p1f1
ACCESS-ESM1-5	10	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r8i1p1f1, r9i1p1f1, r10i1p1f1
CanESM5	10	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r8i1p1f1, r9i1p1f1, r10i1p1f1
CESM2	10	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r8i1p1f1, r9i1p1f1, r10i1p1f1
EC-Earth3	10	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r9i1p1f1, r10i1p1f1, r11i1p1f1
HadGEM3-GC31-LL	10	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r11i1p1f1, r12i1p1f1, r13i1p1f1, r14i1p1f1, r15i1p1f1
IPSL-CM6A-LR	10	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r8i1p1f1, r9i1p1f1, r10i1p1f1
MPI-ESM1-2-HR	10	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r8i1p1f1, r9i1p1f1, r10i1p1f1
UKESM1-0-LL	10	r1i1p1f2, r2i1p1f2, r3i1p1f2, r4i1p1f2, r5i1p1f3, r6i1p1f3, r7i1p1f3, r8i1p1f2, r9i1p1f2, r10i1p1f2

mooring site A3 released by Woodgate (2018) (Fig. 14). The FWF from the CMIP6 models exhibited larger errors compared to that of Global-FVCOM. The CMIP6 models exhibited RMSE values ranging from 31.54 mSv to 127.55 mSv, which is larger than the 21.96 mSv estimated in Global-FVCOM. Additionally, the CC for FWF variability in the CMIP6 models ranged from −0.07 to 0.38, which are notably lower than the 0.80 calculated in Global-FVCOM. It is important to note that this analysis is based on a single ensemble member for each model. The internal variability across different ensemble members plays a crucial role in the estimation and comparison of FWF, potentially contributing to the discrepancies observed among CMIP6 models. A more comprehensive assessment that incorporates multiple ensemble members would provide deeper insights into the uncertainties associated with FWF simulations and should be considered in future studies.

In addition, river runoff also represents a critical source of liquid freshwater. In CMIP6 models, river runoff is primarily simulated within the models. However, the complexities of ocean-land-atmosphere interactions can lead to inaccuracies in the estimation of river runoff into the Arctic region. A similar issue applies to net evaporation and precipitation, which also contribute to liquid freshwater input. In contrast, the ice-ocean coupled model incorporates river runoff, evaporation, and precipitation as external forcings, primarily sourced from observational and reanalysis data, which generally leads to more accurate freshwater input estimates.

Moreover, Ekman pumping, driven by both air-ocean stress and ice-ocean stress, is recognized as a key factor influencing FWC. To isolate the impact of wind field differences in CMIP6 models and avoid potential effects from sea ice and ocean surface currents, we chose to focus solely on the component of Ekman pumping induced by wind stress. In CMIP6 models, wind stress is computed internally, whereas the ice-ocean coupled model obtains this input from reanalysis data. Excluding the effects of sea ice and surface currents, a comparative study was conducted to evaluate only the wind-induced Ekman pumping velocity estimated by CMIP6 and Global-FVCOM models, with the Ekman pumping velocity derived from ERA5 (ECMWF Reanalysis v5) wind data. The Ekman pumping velocity associated with wind stress is computed as follows:

$$W_{ekman} = \frac{1}{\rho f}\left(\frac{\partial \tau_y}{\partial x} - \frac{\partial \tau_x}{\partial y}\right) \tag{6}$$

where W_{ekman} indicates Ekman pumping velocity, f represents the Coriolis parameter, and ρ is the seawater density. τ_x and τ_y are the zonal and meridional components of the wind stress, respectively.

The results revealed that, compared to the Ekman pumping velocity estimated by Global-FVCOM, the estimates from CMIP6 models exhibited substantial differences from those derived from ERA5 (Fig. 15). The mean Ekman pumping velocity from ERA5 between 2003 and 2014 was −0.59 cm/day, indicating prevalent downwelling. Global-FVCOM produced a similar downwelling pattern, with a mean velocity of −0.51 cm/day. However, among the 17 CMIP6 models, 10 exhibited prevalent upwelling, contradicting the ERA5 results. The remaining 7 models simulated downwelling, but their biases were larger than that of FVCOM relative to ERA5. The RMSE of the CMIP6 models ranged from 1.94 cm/day to 5.47 cm/day, significantly exceeding the 0.33 cm/day found in FVCOM. Furthermore, the CC between the Ekman pumping velocity estimates from CMIP6 models and ERA5 ranged from −0.14 to 0.24, considerably lower than the CC of 0.93 between FVCOM and ERA5.

It can be concluded that climate models, which are capable of coupling multiple systems including land, atmosphere, and ocean, enable the simulation of a comprehensive set of variables. However, this capability also necessitates managing extensive parameter settings and

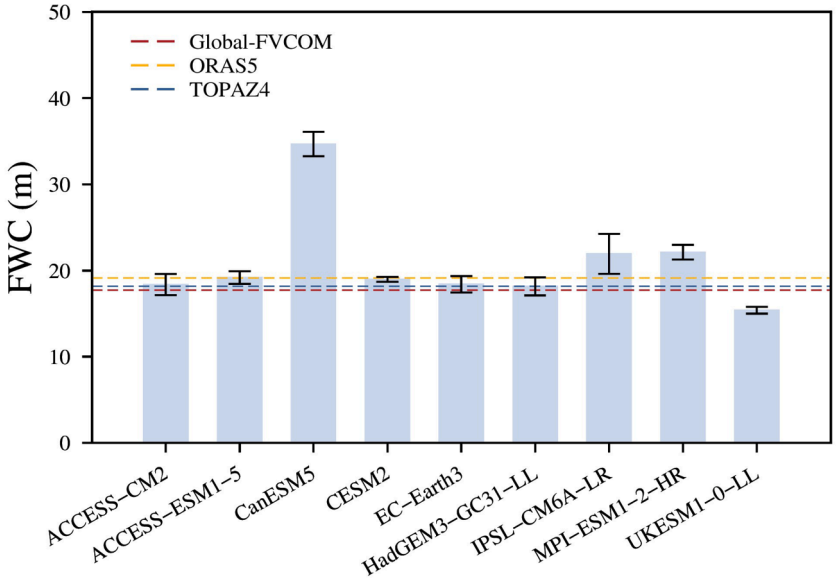


Fig. 12. Comparison of FWC internal variability among ensemble members within each model. Blue and black bar indicates the ensemble mean averaged over selected members and the standard deviation for each model from 2003 to 2014, respectively. Dashed lines represent the mean FWC of the ice-ocean coupled model and two reanalysis data.

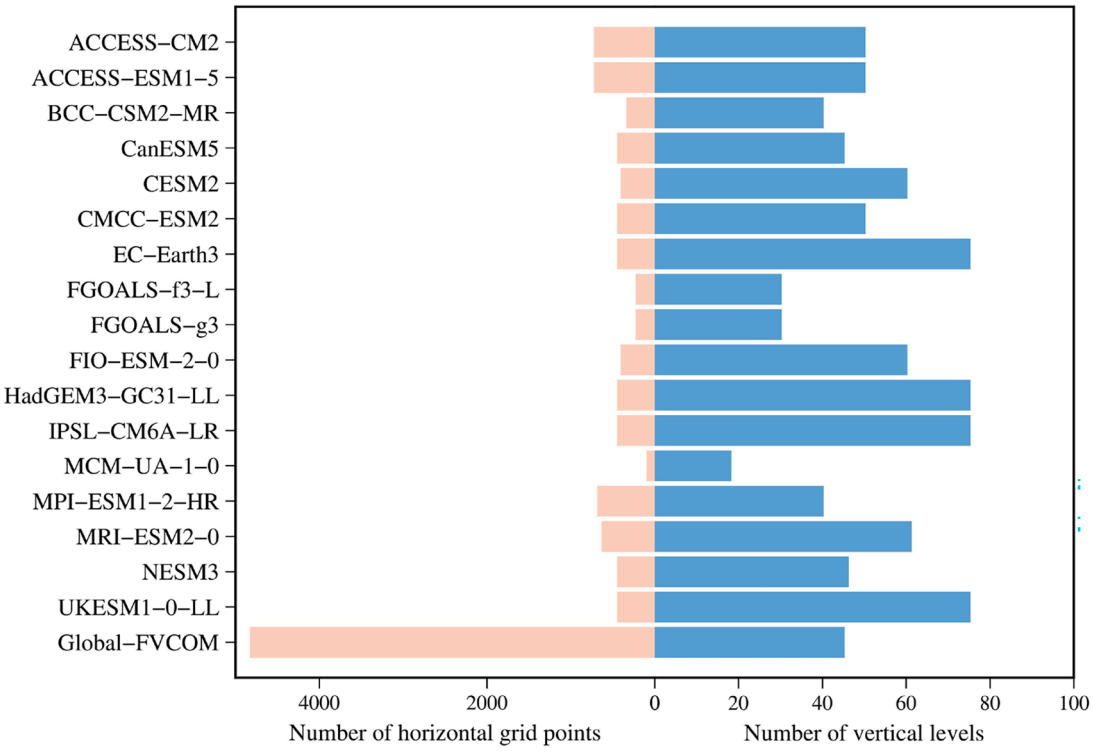


Fig. 13. Horizontal grid points and vertical levels of 17 CMIP6 models and the ice-ocean coupled model in the Beaufort Gyre region.

addressing complex interactions, potentially limiting their performance in simulating FWC in the Beaufort Gyre region compared to specialized ice-ocean coupled models. Therefore, further improvements in CMIP6 models are necessary.

5. Conclusion

The primary objective of this study is to assess the capability of 17 CMIP6 models in simulating freshwater in the Beaufort Gyre region of the Arctic. To achieve this, we evaluated their performance by comparing them with two reanalysis datasets, as well as observational data. For reference, the ice-ocean coupled model, Global-FVCOM, was also included in the comparison. The evaluation focused on FWC and the structure of salinity profiles.

The intercomparison results revealed that, compared to Global-FVCOM, the CMIP6 models exhibited greater discrepancies in spatio-temporal averages of FWC when compared to reanalysis data. More than half of the 17 CMIP6 models overestimated the spatiotemporal mean FWC. The errors in FWC simulation were not attributed to any time period but rather reflects the general inter-model spread in the spatial

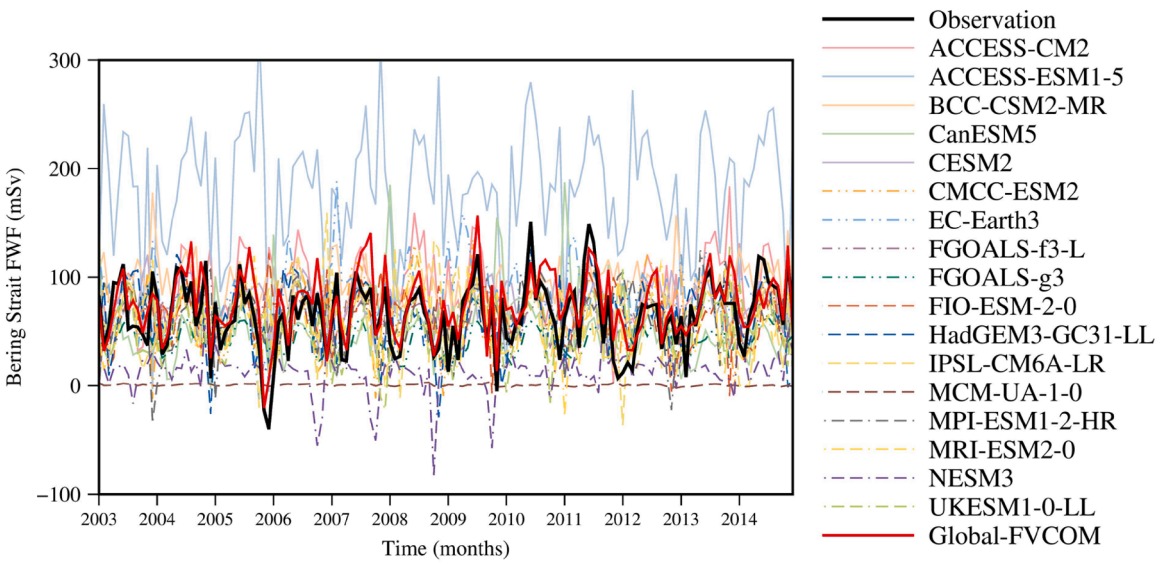


Fig. 14. Monthly variation of FWF through Bering Strait from the observation, 17 CMIP6 models, and the ice-ocean coupled model over the period 2003–2014.

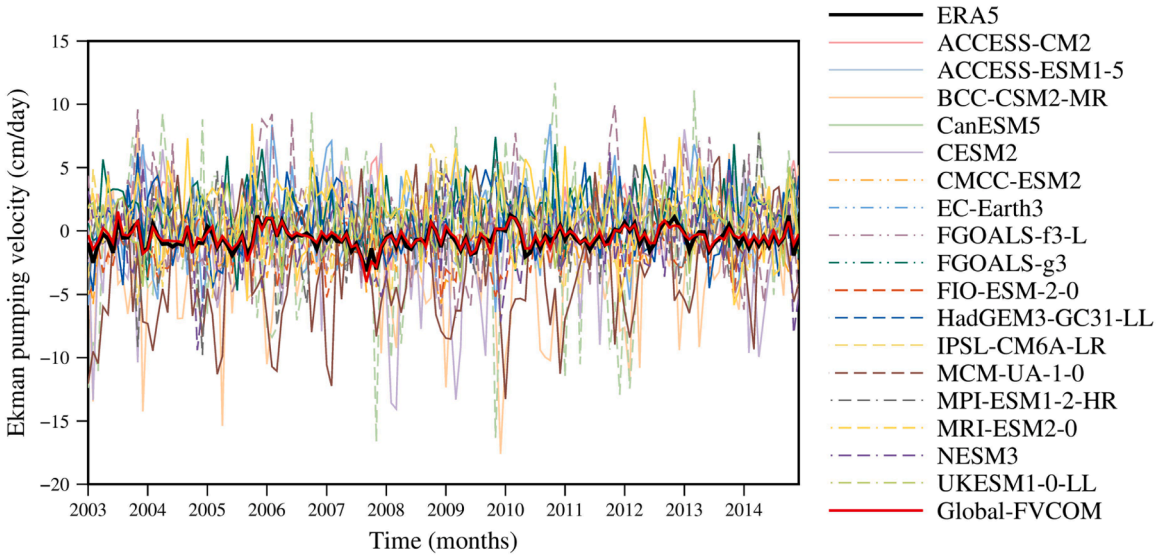


Fig. 15. Monthly variation of wind-induced Ekman pumping velocity from the ERA5 reanalysis, 17 CMIP6 models, and the ice-ocean coupled model over the period 2003–2014.

distribution of FWC. The MMM reproduced the primary characteristics of the CMIP6 models in terms of spatiotemporal averages of FWC.

Additionally, the 17 CMIP6 models exhibited inter-model spread in salinity profiles at different depths, with larger errors compared to Global-FVCOM. Only a few models reasonably captured the stratification near the surface. Most CMIP6 models overestimated salinity near the surface. The vertical salinity errors in these models typically transition from positive at the surface to negative with increasing depth, eventually stabilizing.

The water column with salinity above the isohaline of 34.8 psu was stratified into two layers, separated by the base of the halocline. The primary sources of FWC error within two layers were further investigated. The findings showed that, in more than half of the CMIP6 models, the primary source of FWC error originated from the upper layer. In models where the upper layer was the primary source of error, FWC was primarily underestimated, whereas in those models where the lower layer was the primary source, FWC was typically overestimated.

The evaluation results based on observational and reanalysis data indicated that the simulation errors of FWC in the CMIP6 models and the MMM were greater than those of Global-FVCOM. Among the 17 CMIP6 models, EC-Earth3, MRI-ESM2-0, and FIO-ESM-2-0 exhibited the best overall performance, demonstrating relatively high CC and relatively small RMSE and bias. Inter-model spreads in Arctic freshwater simulations predominantly reflect systematic errors, as intra-model variations induced by internal variability remain substantially smaller than cross-model differences. Several aspects, including discrepancies in numerical configuration, vertical mixing schemes, model resolution, freshwater inputs, and atmospheric forcings, have the potential to significantly influence the simulation results of FWC.

The evaluation results of this study indicate that the current CMIP6 models still face challenges in simulating FWC in the Beaufort Gyre region. Compared to the ice-ocean coupled model, the simulation capabilities of CMIP6 models need further improvement. These findings contribute to the selection and enhancement of CMIP6 models for future projections of freshwater variability.

Data availability

The CMIP6 data can be downloaded at https://esgf-node.llnl.gov/search/cmip6/. The Global-FVCOM data is available from the corresponding author upon reasonable request. The ORAS5 data is collected from https://www.cen.uni-hamburg.de/en/icdc/data/ocean/easy-init-ocean/ecmwf-oras5.html. The TOPAZ4 data is available at https://data.marine.copernicus.eu/product/ARCTIC_MULTIYEAR_PHY_002_003/description. The CTD data is downloaded from https://www2.whoi.edu/site/beaufortgyre/data/ctd-and-geochemistry/. The ITP data is obtained from https://www2.whoi.edu/site/itp/data/data-products/. The observational freshwater flux data is obtained from https://pscfiles.apl.uw.edu/woodgate/BeringStraitArchive/BeringStraitMooringData/BeringStraitProducts/. The wind data of ERA5 is available at https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels-monthly-means?tab=download.

CRediT authorship contribution statement

Yu Zhang: Writing – original draft, Validation, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Zhou Ye:** Writing – review & editing, Visualization, Validation, Software, Methodology, Formal analysis, Data curation. **Feifan Chen:** Writing – original draft, Visualization, Validation, Software, Formal analysis, Data curation. **Changsheng Chen:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization. **Guoping Gao:** Writing – review & editing, Validation, Project administration, Methodology, Conceptualization. **Robert C. Beardsley:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Deshuai Wang:** Writing – review & editing, Validation, Software, Data curation. **Jianhua Qi:** Writing – review & editing, Methodology. **Danya Xu:** Writing – review & editing, Data curation. **Yi Zhou:** Writing – review & editing, Software.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No.42376231), National Key Research and Development Program of China (No.2019YFA0607001), Natural Science Foundation of Shanghai (No.22ZR1427400), National Natural Science Foundation of China (No.42130402 and No.42076240), Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No.311022006), and U.S. National Science Foundation (No.1603000).

References

Aagaard, K., Carmack, E.C., 1989. The role of sea ice and other fresh water in the Arctic circulation. J. Geophys. Res. 94, 14485–14498. https://doi.org/10.1029/JC094iC10p14485.

Barron, C., Smedstad, L., 2002. Global river inflow within the Navy Coastal Ocean Model, paper presented at Oceans 2002 MTS. In: IEEE Conference, pp. 29–31.

Bourgain, P., Gascard, J.C., 2011. The Arctic Ocean halocline and its interannual variability from 1997 to 2008. Deep Sea Res. Part I 58, 745–756. https://doi.org/10.1016/j.dsr.2011.05.001.

Buckley, M.W., Marshall, J., 2016. Observations, inferences, and mechanisms of the atlantic meridional overturning circulation: a review. Rev. Geophys. 54, 5–63. https://doi.org/10.1002/2015RG000493.

Canuto, V.M., Howard, A., Cheng, Y., Dubovikov, M.S., 2002. Ocean turbulence. Part II: vertical diffusivities of momentum, heat, salt, mass, and passive scalars. J. Phys. Oceanogr. 32, 240–264. https://doi.org/10.1175/1520-0485(2002)032<0240:OTPIVD>2.0.CO;2.

Canuto, V.M., Howard, A., Cheng, Y., Dubovikov, M.S., 2001. Ocean turbulence. Part I: one-point closure model—Momentum and heat vertical diffusivities. J. Phys. Oceanogr. 31, 1413–1426. https://doi.org/10.1175/1520-0485(2001)031<1413:OTPIOP>2.0.CO;2.

Carmack, E.C., Yamamoto-Kawai, M., Haine, T.W.N., Bacon, S., Bluhm, B.A., Lique, C., Melling, H., Polyakov, I.V., Straneo, F., Timmermans, M.-L., Williams, W.J., 2016. Freshwater and its role in the Arctic Marine System: sources, disposition, storage, export, and physical and biogeochemical consequences in the Arctic and global oceans. J. Geophys. Res. Biogeosc. 121, 675–717. https://doi.org/10.1002/2015JG003140.

Chen, C., Beardsley, R.C., Cowles, G., 2006. An unstructured grid, finite-volume coastal ocean model (FVCOM) system. Oceanography 19, 78–89. https://doi.org/10.5670/oceanog.2006.92.

Chen, C., Gao, G., Qi, J., Proshutinsky, A., Beardsley, R.C., Kowalik, Z., Lin, H., Cowles, G., 2009. A new high-resolution unstructured grid finite volume Arctic Ocean model (AO-FVCOM): an application for tidal studies. J. Geophys. Res. 114, 2008JC004941. https://doi.org/10.1029/2008JC004941.

Chen, C., Gao, G., Zhang, Y., Beardsley, R.C., Lai, Z., Qi, J., Lin, H., 2016. Circulation in the Arctic Ocean: results from a high-resolution coupled ice-sea nested Global-FVCOM and Arctic-FVCOM system. Prog. Oceanogr. 141, 60–80. https://doi.org/10.1016/j.pocean.2015.12.002.

Chen, C., Huang, H., Beardsley, R.C., Liu, H., Xu, Q., Cowles, G., 2007. A finite volume numerical approach for coastal ocean circulation studies: comparisons with finite difference models. J. Geophys. Res. 112, 2006JC003485. https://doi.org/10.1029/2006JC003485.

Chen, C., Liu, H., Beardsley, R.C., 2003. An unstructured grid, finite-volume, three-dimensional, primitive equations ocean model: application to Coastal Ocean and estuaries. J. Atmos. Oceanic Technol. 20, 159–186. https://doi.org/10.1175/1520-0426(2003)020<0159:AUGFVT>2.0.CO;2.

Chen, F., Wang, D., Zhang, Y., Zhou, Y., Chen, C., 2024. Intercomparisons and evaluations of satellite-derived arctic sea ice thickness products. Remote Sens (Basel) 16. https://doi.org/10.3390/rs16030508.

Chen, L., Wu, R., Shu, Q., Min, C., Yang, Q., Han, B., 2023. The Arctic sea ice thickness change in CMIP6's historical simulations. Adv. Atmos. Sci. 40, 2331–2343. https://doi.org/10.1007/s00376-022-1460-4.

Decloedt, T., Luther, D.S., 2010. On a simple empirical parameterization of topography-catalyzed diapycnal mixing in the abyssal ocean. J. Phys. Oceanogr. 40, 487–508. https://doi.org/10.1175/2009JPO4275.1.

Deng, Y., Gao, G., Zhang, Y., Chen, C., 2019. Seasonal and interannual variability of Bering Strait throughflow from AO-FVCOM and observation. J. Ocean Univ. China 18, 615–625. https://doi.org/10.1007/s11802-019-3906-6.

Dickson, R.R., Meincke, J., Malmberg, S.-A., Lee, A.J., 1988. The "great salinity anomaly" in the Northern North Atlantic 1968–1982. Prog. Oceanogr. 20, 103–151. https://doi.org/10.1016/0079-6611(88)90049-3.

Feistel, R., Wright, D.G., Jackett, D.R., Miyagawa, K., Reissmann, J.H., Wagner, W., Overhoff, U., Guder, C., Feistel, A., Marion, G.M., 2010. Numerical implementation and oceanographic application of the thermodynamic potentials of liquid water, water vapour, ice, seawater and humid air – Part 1: background and equations. Ocean Sci. 6, 633–677. https://doi.org/10.5194/os-6-633-2010.

Gao, G., Chen, C., Qi, J., Beardsley, R.C., 2011. An unstructured-grid, finite-volume sea ice model: development, validation, and application. J. Geophys. Res. 116, C00D04. https://doi.org/10.1029/2010JC006688.

Gaspar, P., Grégoris, Y., Lefevre, J., 1990. A simple eddy kinetic energy model for simulations of the oceanic vertical mixing: tests at station Papa and long-term upper ocean study site. J. Geophys. Res. 95, 16179–16193. https://doi.org/10.1029/JC095iC09p16179.

Haine, T.W.N., Curry, B., Gerdes, R., Hansen, E., Karcher, M., Lee, C., Rudels, B., Spreen, G., De Steur, L., Stewart, K.D., Woodgate, R., 2015. Arctic freshwater export: status, mechanisms, and prospects. Glob Planet Change 125, 13–35. https://doi.org/10.1016/j.gloplacha.2014.11.013.

Heuzé, C., Zanowski, H., Karam, S., Muilwijk, M., 2023. The deep arctic ocean and fram strait in CMIP6 models. J. Clim. 36, 2551–2584. https://doi.org/10.1175/JCLI-D-22-0194.1.

Hochet, A., Lique, C., Sévellec, F., Llovel, W., 2024. Drivers of interannual salinity variability in the Arctic Ocean. J. Geophys. Res. Oceans 129, e2023JC020852. https://doi.org/10.1029/2023JC020852.

Hu, Z., Chen, D., Chen, X., Zhou, Q., Peng, Y., Li, J., Sang, Y., 2022. CCHZ-DISO: a timely new assessment system for data quality or model performance from Da Dao Zhi Jian. Geophys. Res. Lett. 49, e2022GL100681. https://doi.org/10.1029/2022GL100681.

Hu, Z., Chen, X., Zhou, Q., Chen, D., Li, J., 2019. DISO: a rethink of Taylor diagram. Intl. J. Climatol. 39, 2825–2832. https://doi.org/10.1002/joc.5972.

Langehaug, H.R., Sagen, H., Stallemo, A., Uotila, P., Rautiainen, L., Olsen, S.M., Devilliers, M., Yang, S., Storheim, E., 2023. Constraining CMIP6 estimates of Arctic Ocean temperature and salinity in 2025-2055. Front. Mar. Sci. 10, 1211562. https://doi.org/10.3389/fmars.2023.1211562.

Large, W.G., McWilliams, J.C., Doney, S.C., 1994. Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. Rev. Geophys. 32, 363–403. https://doi.org/10.1029/94RG01872.

Lin, P., Pickart, R.S., Heorton, H., Tsamados, M., Itoh, M., Kikuchi, T., 2023. Recent state transition of the Arctic Ocean's Beaufort Gyre. Nat. Geosci. 16, 485–491. https://doi.org/10.1038/s41561-023-01184-5.

Lique, C., Holland, M.M., Dibike, Y.B., Lawrence, D.M., Screen, J.A., 2016. Modeling the Arctic freshwater system and its integration in the global system: lessons learned and future challenges. J. Geophys. Res. Biogeosci. 121, 540–566. https://doi.org/10.1002/2015JG003120.

Long, M., Zhang, L., Hu, S., Qian, S., 2021. Multi-aspect assessment of CMIP6 models for Arctic sea ice simulation. J. Climate 34, 1515–1529. https://doi.org/10.1175/JCLI-D20-0522.1.

Macdonald, R.W., Carmack, E.C., McLaughlin, F.A., Falkner, K.K., Swift, J.H., 1999. Connections among ice, runoff and atmospheric forcing in the Beaufort Gyre. Geophys. Res. Lett. 26, 2223–2226. https://doi.org/10.1029/1999GL900508.

Manucharyan, G.E., Spall, M.A., Thompson, A.F., 2016. A theory of the wind-driven Beaufort Gyre variability. J. Phys. Oceanogr. 46, 3263–3278. https://doi.org/10.1175/JPO-D-16-0091.1.

McPhee, M.G., Proshutinsky, A., Morison, J.H., Steele, M., Alkire, M.B., 2009. Rapid change in freshwater content of the Arctic Ocean. Geophys. Res. Lett. 36, 2009GL037525. https://doi.org/10.1029/2009GL037525.

Mellor, G.L., Yamada, T., 1982. Development of a turbulence closure model for geophysical fluid problems. Rev. Geophys. 20, 851–875. https://doi.org/10.1029/RG020i004p00851.

Metzner, E.P., Salzmann, M., 2023. Technical note: determining Arctic Ocean halocline and cold halostad depths based on vertical stability. Ocean Sci. 19, 1453–1464. https://doi.org/10.5194/os-19-1453-2023.

Morison, J., Kwok, R., Peralta-Ferriz, C., Alkire, M., Rigor, I., Andersen, R., Steele, M., 2012. Changing Arctic Ocean freshwater pathways. Nature 481, 66–70. https://doi.org/10.1038/nature10705.

Nie, Y., Lin, X., Yang, Q., Liu, J., Chen, D., Uotila, P., 2023. Differences between the CMIP5 and CMIP6 antarctic sea ice concentration budgets. Geophys. Res. Lett. 50, e2023GL105265. https://doi.org/10.1029/2023GL105265.

Pacanowski, R.C., Philander, S.G.H., 1981. Parameterization of vertical mixing in numerical models of tropical oceans. J. Phys. Oceanogr. 11, 1443–1451.

Pan, R., Shu, Q., Wang, Q., Wang, S., Song, Z., He, Y., Qiao, F., 2023. Future Arctic climate change in CMIP6 strikingly intensified by NEMO-Family climate models. Geophys. Res. Lett. 50, e2022GL102077. https://doi.org/10.1029/2022GL102077.

Polyakov, I.V., Pnyushkov, A.V., Carmack, E.C., 2018. Stability of the arctic halocline: a new indicator of arctic climate change. Environ. Res. Lett. 13, 125008. https://doi.org/10.1088/1748-9326/aaec1e.

Proshutinsky, A., Krishfield, R., Timmermans, M., Toole, J., Carmack, E., McLaughlin, F., Williams, W.J., Zimmermann, S., Itoh, M., Shimada, K., 2009. Beaufort Gyre freshwater reservoir: state and variability from observations. J. Geophys. Res. 114, 2008JC005104. https://doi.org/10.1029/2008JC005104.

Proshutinsky, A., Krishfield, R., Toole, J.M., Timmermans, M.-L., Williams, W., Zimmermann, S., Yamamoto-Kawai, M., Armitage, T.W.K., Dukhovskoy, D., Golubeva, E., Manucharyan, G.E., Platov, G., Watanabe, E., Kikuchi, T., Nishino, S., Itoh, M., Kang, S.-H., Cho, K.-H., Tateyama, K., Zhao, J., 2019. Analysis of the Beaufort Gyre freshwater content in 2003–2018. J. Geophys. Res. Oceans 124, 9658–9689. https://doi.org/10.1029/2019JC015281.

Rabe, B., Karcher, M., Kauker, F., Schauer, U., Toole, J.M., Krischfield, R.A., Pisarev, S., Rudels, B., Kikuchi, T., Zhao, J.P., 2013. Arctic Ocean liquid freshwater storage trend 1992 - 2012. In: EGU General Assembly Conference Abstracts, EGU General Assembly Conference Abstracts, pp. EGU2013–EGU8213.

Rabe, B., Karcher, M., Kauker, F., Schauer, U., Toole, J.M., Krishfield, R.A., Pisarev, S., Kikuchi, T., Su, J., 2014. Arctic Ocean basin liquid freshwater storage trend 1992–2012. Geophys. Res. Lett. 41, 961–968. https://doi.org/10.1002/2013GL058121.

Sakov, P., Counillon, F., Bertino, L., Lisæter, K.A., Oke, P.R., Korablev, A., 2012. TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. Ocean Sci. 8 (4), 633–656. https://doi.org/10.5194/os-8-633-2012.

Serreze, M.C., Barrett, A.P., Slater, A.G., Woodgate, R.A., Aagaard, K., Lammers, R.B., Steele, M., Moritz, R., Meredith, M., Lee, C.M., 2006. The large-scale freshwater cycle of the Arctic. J. Geophys. Res. 111, 2005JC003424. https://doi.org/10.1029/2005JC003424.

Shen, X., Zhang, Y., Chen, C., Hu, S., Xu, D., Shao, W., Chang, L., Feng, G., 2021. Arctic sea ice variation in the Northwest Passage in 1979–2017 and its response to surface thermodynamics factors. Adv. Climate Change Res. 12, 563–580. https://doi.org/10.1016/j.accre.2021.08.004.

Shimada, K., Itoh, M., Nishino, S., McLaughlin, F., Carmack, E., Proshutinsky, A., 2005. Halocline structure in the Canada Basin of the Arctic Ocean. Geophys. Res. Lett. 32, 2004GL021358. https://doi.org/10.1029/2004GL021358.

Shu, Q., Wang, Q., Song, Z., Qiao, F., Zhao, J., Chu, M., Li, X., 2020. Assessment of sea ice extent in CMIP6 with comparison to observations and CMIP5. Geophys. Res. Lett. 47, e2020GL087965. https://doi.org/10.1029/2020GL087965.

Wang, S., Wang, Q., Wang, M., Lohmann, G., Qiao, F., 2022. Arctic Ocean freshwater in CMIP6 coupled models. Earth's Future 10, e2022EF002878. https://doi.org/10.1029/2022EF002878.

Wang, X., Lu, R., Wang, S.-Y., Chen, R.-T., Chen, Z.-Q., Hui, F.-M., Huang, H.-B., Cheng, X., 2023. Assessing CMIP6 simulations of Arctic sea ice drift: role of near-surface wind and surface ocean current in model performance. Adv. Climate Change Res. 14, 691–706. https://doi.org/10.1016/j.accre.2023.09.005.

Watts, M., Maslowski, W., Lee, Y.J., Kinney, J.C., Osinski, R., 2021. A spatial evaluation of Arctic sea ice and regional limitations in CMIP6 historical simulations. J. Clim. 34, 6399–6420. https://doi.org/10.1175/JCLI-D-20-0491.1.

White, D., Hinzman, L., Alessa, L., Cassano, J., Chambers, M., Falkner, K., Francis, J., Gutowski, W.J., Holland, M., Holmes, R.M., Huntington, H., Kane, D., Kliskey, A., Lee, C., McClelland, J., Peterson, B., Rupp, T.S., Straneo, F., Steele, M., Woodgate, R., Yang, D., Yoshikawa, K., Zhang, T., 2007. The arctic freshwater system: changes and impacts. J. Geophys. Res. 112, 2006JG000353. https://doi.org/10.1029/2006JG000353.

Woodgate, R.A., 2018. Increases in the Pacific inflow to the Arctic from 1990 to 2015, and insights into seasonal trends and driving mechanisms from year-round Bering Strait mooring data. Prog. Oceanogr. 160, 124–154. https://doi.org/10.1016/j.pocean.2017.12.007.

Zanowski, H., Jahn, A., Holland, M.M., 2021. Arctic Ocean freshwater in CMIP6 ensembles: declining sea ice, increasing Ocean storage and export. J. Geophys. Res. Oceans 126, e2020JC016930. https://doi.org/10.1029/2020JC016930.

Zhang, J., Steele, M., 2007. Effect of vertical mixing on the Atlantic Water layer circulation in the Arctic Ocean. J. Geophys. Res. 112, 2006JC003732. https://doi.org/10.1029/2006JC003732.

Zhang, Y., Chen, C., Beardsley, R.C., Gao, G., Lai, Z., Curry, B., Lee, C.M., Lin, H., Qi, J., Xu, Q., 2016a. Studies of the Canadian Arctic Archipelago water transport and its relationship to basin-local forcings: results from AO-FVCOM. J. Geophys. Res. Oceans 121, 4392–4415. https://doi.org/10.1002/2016JC011634.

Zhang, Y., Chen, C., Beardsley, R.C., Gao, G., Qi, J., Lin, H., 2016b. Seasonal and interannual variability of the arctic sea ice: a comparison between AO-FVCOM and observations: numerical study on the arctic sea ice. J. Geophys. Res. Oceans 121, 8320–8350. https://doi.org/10.1002/2016JC011841.

Zhang, Y., Chen, C., Shen, X., Xu, D., Shao, W., Beardsley, R.C., Chang, L., Feng, G., 2021. Role of sea level pressure in variations of the Canadian Arctic Archipelago throughflow. Adv. Climate Change Res. 12, 539–552. https://doi.org/10.1016/j.accre.2021.07.009.

Zhang, Y., Li, G., Li, H., Chen, C., Shao, W., Zhou, Y., Wang, D., 2024. A spatiotemporal comparison and assessment of multisource satellite derived sea ice thickness in the Arctic thinner ice region. IEEE J. Selected Topics Appl. Earth Observat. Remote Sens. 17, 8710–8723. https://doi.org/10.1109/JSTARS.2024.3390618.

Zhang, Y., Chen, C., Beardsley, R.C., Perrie, W., Gao, G., Zhang, Y., Qi, J., Lin, H., 2020. Applications of an unstructured grid surface wave model (FVCOM-SWAVE) to the Arctic Ocean: the interaction between ocean waves and sea ice. Ocean Modell. 145, 101532. https://doi.org/10.1016/j.ocemod.2019.101532.

Zuo, H., Balmaseda, M.A., Tietsche, S., Mogensen, K., Mayer, M., 2019. The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment. Ocean Sci. 15, 779–808. https://doi.org/10.5194/os-15-779-2019.