CrossMark

An assessment of ten ocean reanalyses in the polar regions

Petteri Uotila[1] · Hugues Goosse[2] · Keith Haines[3] · Matthieu Chevallier[4] · Antoine Barthélemy[2] · Clément Bricaud[5] · Jim Carton[6] · Neven Fučkar[7,8] · Gilles Garric[5] · Doroteaciro Iovino[9] · Frank Kauker[10] · Meri Korhonen[11] · Vidar S. Lien[12] · Marika Marnela[11] · François Massonnet[2,7] · Davi Mignac[3] · K. Andrew Peterson[13] · Remon Sadikni[14] · Li Shi[15] · Steffen Tietsche[16] · Takahiro Toyoda[17] · Jiping Xie[18] · Zhaoru Zhang[19]

Abstract
Global and regional ocean and sea ice reanalysis products (ORAs) are increasingly used in polar research, but their quality remains to be systematically assessed. To address this, the Polar ORA Intercomparison Project (Polar ORA-IP) has been established following on from the ORA-IP project. Several aspects of ten selected ORAs in the Arctic and Antarctic were addressed by concentrating on comparing their mean states in terms of snow, sea ice, ocean transports and hydrography. Most polar diagnostics were carried out for the first time in such an extensive set of ORAs. For the multi-ORA mean state, we found that deviations from observations were typically smaller than individual ORA anomalies, often attributed to off-setting biases of individual ORAs. The ORA ensemble mean therefore appears to be a useful product and while knowing its main deficiencies and recognising its restrictions, it can be used to gain useful information on the physical state of the polar marine environment.

Marika Marnela was formerly at Finnish Meteorological Institute.

✉ Petteri Uotila
petteri.uotila@helsinki.fi

[1] Institute for Atmospheric and Earth System Research (INAR)/Physics, University of Helsinki, Helsinki, Finland

[2] Earth and Life Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

[3] University of Reading and National Centre for Earth Observation, Reading, UK

[4] Centre National de Recherches Météorologiques, Météo France/CNRS UMR3589, Toulouse, France

[5] Mercator Océan, Toulouse, France

[6] University of Maryland, College Park, USA

[7] Barcelona Supercomputing Centre, Barcelona, Spain

[8] Environmental Change Institute, University of Oxford, Oxford, UK

[9] Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy

[10] Alfred Wegener Institute, Bremerhaven, Germany

[11] Finnish Meteorological Institute, Helsinki, Finland

[12] Institute of Marine Research, Bergen, Norway

[13] Met Office, Exeter, UK

[14] University of Hamburg, Hamburg, Germany

[15] Bureau of Meteorology, Melbourne, Australia

[16] European Centre for Medium-Range Weather Forecasts, Reading, UK

[17] Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan

[18] Nansen Environmental and Remote Sensing Center, Bergen, Norway

[19] Shanghai Jiao Tong University, Shanghai, China

1 Introduction

For years, atmospheric reanalysis products, which consist of multidecadal meteorological model simulations with assimilated observations, have become an invaluable resource for researchers representing a wide range of disciplines. Recently, similar products—ocean reanalyses (ORAs)—have been constructed by many research groups. It is likely that these products will become as valuable as their atmospheric counterparts.

Specifically, an ocean analysis describes an ocean state valid for a particular time by a set of gridded oceanographic variables. Typically an ocean analysis is generated by an analysis system consisting of a hydrodynamical or statistical model and an observation assimilation framework, for the purpose of initialising a forecast. During the analysis generation process, the forecast model background state is adjusted toward new observations. The amount of adjustment is denoted as the analysis increment, which quantify the impact of data assimilation in the analysis system (Cullather and Bosilovich 2012).

Ocean and sea ice reanalyses are analyses in the form of time series, where every analysis is generated using the same analysis system for all historical observations. Hence, they combine observations either statistically or with a hydrodynamical model, to reconstruct historical conditions and their changes in the ocean.

Global and regional ORA products are increasingly used in polar research, but their quality remains to be systematically assessed. To address this, the Polar ORA Intercomparison Project (Polar ORA-IP) has been established following on from the ORA-IP project (Balmaseda et al. 2015; Toyoda et al. 2017a, b; Chevallier et al. 2017; Tietsche et al. 2015; Karspeck et al. 2015; Shi et al. 2017; Valdivieso et al. 2017; Palmer et al. 2017; Masina et al. 2015; Storto et al. 2017).

These ORA-IP studies have looked at various aspects of global ocean hydrodynamics (steric sea level, air-sea fluxes, ocean heat and salt content among others). The only ORA-IP publication with a polar focus has been Chevallier et al. (2017), who compared the representation of the sea-ice cover in the Arctic Ocean in 14 global reanalyses. Using a variety of in-situ and satellite-based observational datasets, they investigated mean states, trends and interannual variability in these reanalyses, focusing on sea-ice concentration (with extent and area), thickness (with volume), velocity and snow depth over sea ice.

Chevallier et al. (2017) showed consistency with respect to sea-ice concentration, which is primarily due to the constraints in surface temperature imposed by atmospheric forcing, and ocean-ice data assimilation. However, they found a large spread in sea-ice and snow thicknesses within the ensemble of ORAs, due to biases in the ocean-ice model

components, and lack of observational constraint. Chevallier et al. (2017) discussed the possible role of model parameters, prescribed atmospheric forcing and data assimilation on the spread. They concluded that none of the ORAs stands superior to the others when compared with observed sea-ice thickness calculated from satellite altimetry data, and that data assimilation does not seem to improve the simulated sea-ice thickness. As a result, estimates of Arctic sea-ice volume by individual ORAs suffer large uncertainties, and the ORA multi-model ensemble mean (MMM) ice volume does not provide a more robust estimate. Most of the global reanalyses used in Chevallier et al. (2017) have now been updated and their updates are evaluated in the present paper which allows direct comparisons with their results.

In this study, we aim for a comprehensive evaluation of ten selected ORA products (C-GLORS025v5, ECDA3, GECCO2, Glorys2v4, GloSea5-GO5, MOVE-G2i, ORAP5, SODA3.3.1, TOPAZ4 and UR025.4) in the Arctic and Southern Oceans (Table 1). For these regions the diagnostics target the following topics: hydrography; ocean heat (OHC), salt content (OSC); ocean transports; mixed layer depth (MLD); sea-ice concentration (SIC) and thickness (SIT); and snow thickness over sea ice. The ORA product biases against observed reference data and their mutual spread are quantified, and possible reasons for discrepancies discussed.

The scope of our manuscript is to provide a broad state-of-the-science overview of ocean reanalyses, plus our best estimate of what the truth might look like. In this context, we will check if the MMM is a useful estimate. As we will repeatedly show, it is a set of fields which is generally most consistent with observations. This is what many users require, although it may not be best suited to analysing dynamical or physical processes, for example.

If a user does not want the MMM, but would prefer a single ORA output, for instance to understand the dynamics, this paper does not seek to tell the user which one to use, but in addition to providing a general evaluation, it is able to show which are outliers for certain variables, which can still be very useful.

We pay particular attention to the performance of the MMM compared to individual products and the identification of outliers. Notably, as the ORAs assimilate observations they are not independent of some of the reference data they are compared to. Moreover, we investigate links and co-variability between the diagnostics, such as the Arctic Ocean heat content and North Atlantic heat transport, and between the mixed layer depth, oceanic convection, the upper ocean hydrography, sea ice and snow. In this way, we try to identify physical mechanisms causing common and individual ORA biases.

Although a large majority of the existing ORA publications does not focus on polar regions, the Coordinated Ocean Reference Experiment (CORE-II; Danabasoglu et al.

Table 1 List of ten ocean reanalyses used in the study and their central characteristics

Name	C-GLORS025v5	ECDA3	GECCO2	GLORYS2v4	GloSea5-GO5	MOVE-G2i	ORAP5	SODA3.3.1	TOPAZ4	UR025.4
Institution	CMCC	GFDL/NOAA	Hamburg University	Mercator Océan	UK MetOffice	MRI/JMA	ECMWF	University of Maryland	NERSC	University of Reading
Nominal horizontal resolution	0.25°	1°	1° × 1/3°	0.25°	0.25°	1° × 0.3° − 0.5°	0.25°	0.25°	12–16 km	0.25°
Vertical resolution	50 z-levels	50 z-levels	50 z-levels	75 z-levels	75 z-levels	52 z-levels	75 z-levels	50 z-levels	28 z-isopycnal layers	75 z-levels
Top-level thickness	~1 m	10 m	10 m	~1 m	~1 m	2.25 m	~1 m	~10 m	min 3 m	~1 m
Ocean-ice model	NEMO3.2-LIM2	MOM4-SIS	MITgcm	NEMO3.1-LIM2	NEMO3.4-CICE	MRI.COM3-CICE4	NEMO3.4-LIM2	MOM5-SIS	HYCOM-EVP SI	NEMO3.2-LIM2
Time period	1980–2015	1961–2012	1948–2014	1992–2015	1993–2012	1980–2012	1979–2012	1980–2015	1991–2016	1989–2010
Initialization	Spinup	Spinup	Cold start	Cold start	Spinup	Spinup	Spinup	Spinup	Cold start	Cold start
Source of atmospheric forcing data	ERA-Interim	Coupled	NCEP RA1	ERA-Interim	ERA-Interim	JRA-55[a]	ERA-Interim	NASA MERRA2	ERA-Interim	ERA-Interim
Ocean restoring	Large scale bias correction to EN3v2a	Fully coupled	None	T, S restoring towards EN4.1.1 for z > 2000 m and lat < 60°S (τ = 20 years)	Surface Haney SSS restoring (− 33.333 mm/day/PSU), 3D T/S to ENACT3 2004–2008 climatology (τ = 1 year)	Relaxing (by IAU) T/S to merged PHC3-WOA13 climatology (τ = 5 years)	Relaxation to OSTIA/NOAA OIv2d SST	Restoring to mean T and S (τ = 10 years). Relaxation to WOA SSS (τ = 3 months)	Relaxing T/S to merged PHC3–WOA13 climatology	None
Sea-ice DA method	Nudging	None (SST)	None (SST)	Reduced order KF	3DVAR	3DVAR	3DVAR-FGAT	None (SST)	EnKF	OI
Sea-ice DA variables	SIC, Arctic SIT	–	–	SIC	SIC	SIC	SIC	–	SIC, SIV	SIC
Sea-ice DA sources	NOAA OIv2d, PIOMAS	–	–	CERSAT	OSISAFv2	MGDSST	OSTIA, NOAA OIv2d	–	OSISAF	OSISAF
Ocean DA method	3DVAR	EnKF	4DVAR (adjoint)	Reduced order KF + 3DVAR large scale bias correction to in-situ T, S	3DVAR	3DVAR	3DVAR	OI	EnKF	OI
Ocean DA variables	T, S, SSH, SST	T, S, SST	T, S, SSH, SST	T, S, SSH, SST	T, S, SST, SSH	T, S, SSH, SST	T, S, SSH, SST	T, S, SST	T, S, SSH, SST	T, S, SSH, SST, SSS

Table 1 (continued)

Name	C-GLORS025v5	ECDA3	GECCO2	GLORYS2v4	GloSea5-GO5	MOVE-G2i	ORAP5	SODA3.3.1	TOPAZ4	UR025.4
Ocean DA sources	EN3v2a, AVISO	HadSST, OISST, WOD09, GTSPP, Argo	EN3v2a, AVISO, GOCO, HadISST, AMSRE, WOA09	CMEMS, NOAA AVHRR	EN4, ICOADS, AVHRR, ATSR, AMSRE, AVISOv3	WOD13, GTSPP, AVISO, MGDSST	EN3v2a, AVISO, OSTIA	WOD, ICOADS, AVHRR, Metosat SEVIRI	NOAA Reynolds, OSTIA, CLS, Damocles	EN3v2a, OSTIA, AVISO
Reference	Storto et al. (2016)	Chang et al. (2013)	Köhl (2015)	Garric et al. (2018)	Blockley et al. (2014, 2015)	Toyoda et al. (2016)	Zuo et al. (2015), Tietsche et al. (2015)	Carton and Giese (2008)	Xie et al. (2017)	Valdivieso et al. (2014)

T temperature data, *S* salinity, *SST* sea surface temperature, *SSS* sea surface salinity, *SSH* sea surface height, *SIC* sea-ice concentration, *SIT* sea-ice thickness, *SIV* sea-ice velocity, *DA* data assimilation, *KF* Kalman filter, *EnKF* ensemble Kalman filter, *OI* objective interpolation

[a]Climatological radiation biases from GEWEX3.0 (Stackhouse Jr et al. 2011) are corrected

2014) has produced papers (Downes et al. 2015; Farneti et al. 2015; Wang et al. 2016a, b) which evaluate the polar performance of a number of state-of-the-science global ocean models. The main difference between the CORE-II model configurations and the ORAs is that the latter employ advanced data assimilation schemes using mostly the same ocean-ice observations, while CORE-II models only apply simple surface flux corrections that, for example, nudge their sea surface salinities toward climatological values. However the CORE-II protocol requires the participating modelling groups to use common atmospheric states and boundary layer parameterisations to drive their multidecadal simulations (e.g. Griffies et al. 2009; Danabasoglu et al. 2014), which is not the case for the ORAs. It is interesting to compare the relative effectiveness of the common CORE-II framework with the ORA observations in producing consistent results.

Due to these dependencies, comparisons between CORE-II and ORA results potentially enable us to estimate the role of different factors affecting the multi-model skill in the polar oceans. Similarities between CORE-II and the ORA MMM performance may reveal common issues in model physics and resolution, while discrepancies may provide information on the role of data assimilation and atmospheric forcing.

Along with CORE-II results, other relevant literature for the Arctic and Southern Oceans are discussed in the next two Sects. 2.1 and 2.2, respectively. In Sect. 3, we describe our diagnostic methods and in Sect. 4 we represent the analysis results of ten ORAs. These results are then compared with previous results, including Chevallier et al. (2017) and CORE-II, in the discussion (Sect. 5). Conclusions follow in Sect. 6.

2 Observed and simulated changes in the polar oceans

2.1 The Arctic Ocean

The Arctic sea ice has shown an unprecedented decline since the mid-1990s, which also has impacted the state of the Arctic Ocean (Comiso 2012; Polyakov et al. 2013; IPCC 2013; Polyakov et al. 2017). This dramatic change highlights the need for more comprehensive environmental data to assess the state and impacts of the Arctic in transition. However, even after a number of targeted field expeditions and improved satellite coverage, the Arctic Ocean observations remain sparse compared to the northern North Atlantic. An important reason for this is that with a few exceptions there are no Argo-buoy deployments north of 70°N to provide hydrographic observations, as the buoys cannot operate under perennial sea ice. Furthermore, international research

teams have had restricted access to the observations from the Russian Arctic which has further limited the observational coverage. Climate models appear too conservative in terms of simulating the observed Arctic sea-ice decline, although there have been some improvements, while their prediction accuracy is significantly limited by the relatively large climate variability (Stroeve et al. 2012; Jahn et al. 2016; Melia et al. 2015).

Despite the aforementioned limitations, significant progress in understanding of the physical state and evolution of the Arctic Ocean has been gained during the last decade. We briefly list some research efforts closely related to the development of ocean reanalysis products in the Arctic.

The Arctic Ocean Model Intercomparison Project (AOMIP) and its successor, the Forum for Arctic Modeling and Observational Synthesis (FAMOS), have in the last two decades identified many model shortcomings and come up with recommendations to reduce the impacts of these shortcomings (Proshutinsky et al. 2016). AOMIP and FAMOS have covered a wide range of topics from Arctic Ocean energetics to sea-ice dynamics (for example Uotila et al. 2006; Heimbach et al. 2010; Karcher et al. 2012). The first AOMIP phase proved that the co-ordinated community approach is the most effective way to address the degree of uncertainty of model results. During AOMIP, ocean-ice models with data assimilation were first introduced to the community (see for example Kauker et al. 2009). Later, FAMOS has been a very productive collaborative effort by producing more than 60 publications including a special issue in the Journal of Geophysical Research (Proshutinsky et al. 2016). The AOMIP/FAMOS modelling studies document, in addition to their scientific results, important ORA developments in the polar regions from the reanalysis methodological perspective. However, a systematic diagnostic analysis of ORA products in the Arctic is missing from the AOMIP/FAMOS studies. This is likely due to the relatively late appearance of ORAs, which have a global scope, in contrast to the regional AOMIP/FAMOS one, and to the strong process focus of AOMIP/FAMOS.

In addition to the sea-ice changes mentioned above, the upper Arctic Ocean is freshening and Rabe et al. (2014) were able to identify a freshwater flux trend of 600 ± 300 km^3 year^{-1} from 1992 to 2012. The variability of the Arctic freshwater content correlates well with the atmospheric forcing and can be closely reproduced by the regional coupled sea ice-ocean model North Atlantic Arctic Sea Ice Ocean Model (NAOSIM) simulations (Karcher et al. 2003). Rabe et al. (2014) suggest a high freshwater export through the Fram Strait until the mid-1990s, followed by lower export rates with no trend thereafter, although models may show large differences in terms of interannual variability of the liquid freshwater through the Fram Strait (Jahn et al. 2012).

Some more recent studies present results from individual polar ocean reanalyses and are worth mentioning here. For example, Xie et al. (2017) analysed multi-decadal ensemble simulations from the regional TOPAZ4 ocean-ice data assimilation system in the Arctic and found that TOPAZ4 performed better with respect to near-surface ocean variables compared to subsurface ocean and sea-ice thickness due to sparse observations. Furthermore, the TOPAZ4 skill improved as the polar observation network became denser. Specifically, TOPAZ4 has a too cold and diffuse Atlantic water (AW) layer in the Arctic leading to a cold bias of 0.3 °C at around 400 m, while the Barents Sea is too warm and saline. Although, the decadal reduction of TOPAZ4 sea-ice extent is close to the observed, its regional distribution has a dipole bias—sea-ice concentration is too low close to the ice edge and too high in the central pack, due to the missing sea-ice heat capacity of TOPAZ4 sea-ice model. Xie et al. (2017) also found that the TOPAZ4 sea ice is too thin, on average.

Lien et al. (2016) applied objective statistical methods to assess the added value of data assimilation in three ocean models, including TOPAZ4, for hydrography, volume and heat transports in the Nordic Seas (the Greenland, Iceland, Norwegian and Seas) and the Barents Sea. They found that both data assimilation and higher model resolution improved the model realism. Specifically, high model resolution in ocean and atmospheric forcing improved the representation of variables closely related to forcing, such as sea-ice concentration and sea surface temperature. Hydrographic data assimilation had a tendency to reduce hydrographic biases, but its effect on the liquid ocean transport remained limited (Zuo et al. 2011). Lien et al. (2016) found that the modelled heat transports through the Fram Strait to the Arctic Ocean were within the observational range related to generally realistic looking hydrography and currents.

Recently, a set of multidecadal ocean-ice model hindcasts generated following the CORE-II protocol has provided a wealth of information on the performance of state-of-the-science global ocean-ice models in the Arctic Ocean (Danabasoglu et al. 2014). The CORE-II atmospheric state, including the global warming trend, was used to drive the models for 60 years from 1948 to 2007. In total, CORE-II models were run for 300 years, corresponding to 5 consecutive loops of the 60-year forcing period. Wang et al. (2016a) analysed the sea-ice extent, sources of solid freshwater and the solid freshwater content of CORE-II models in the Arctic focussing on the fifth forcing cycle. They found that the models reproduced observed sea-ice variability more consistently than the mean state. The CORE-II MMM sea-ice extent was somewhat smaller than observed, in particular in summer, which resulted in a stronger than observed seasonal cycle. The CORE-II MMM overestimated the winter-to-summer sea-ice retreat rate, related to the negative summer sea-ice extent bias. Models that overestimated the sea-ice thickness,

underestimated the multidecadal decline of the Arctic summer sea-ice cover. On average, the models underestimated the observed sea-ice thinning by a factor of two Wang et al. (2016a) stated.

In terms of hydrography, Ilicak et al. (2016) found that while the CORE-II MMM appears to be relatively close to observations, there is a large inter-model temperature spread in the Arctic Ocean. Specifically, at intermediate depths, including the warm AW layer, modelled-to-observed temperature differences were large. The CORE-II MMM had a too cold AW at 400 m whose signal disappeared quickly northward away from the Fram Strait, and an overall cold and fresh bias in the Arctic interior, although its mean freshwater transports through the Arctic gateways appear realistic (Wang et al. 2016b). With respect to individual models, those with too cold intermediate depths have an excessive cold water transport to the Arctic Ocean through the St. Anna Trough, while those models with a warm Arctic have a strong inflow of warm water in the Fram Strait. As with sea ice, the CORE-II models agree on the ocean decadal variability, which is dictated by the common atmospheric forcing, more than they do on the ocean mean state. Following these findings, Ilicak et al. (2016) point out that the CORE-II ocean-ice models have a too coarse horizontal resolution, typically 1° in latitude, to realistically represent the AW inflow, and the deep water formation and currents originating from the shallow continental shelf regions.

2.2 The Southern Ocean

Over recent decades, the Antarctic sea-ice extent has remained relatively stable but with large interannual variability and a small increasing trend that strongly contrasts with the large decline in the Arctic over the same period (Parkinson and Cavalieri 2012; Maksym et al. 2012). Over the Southern Ocean the westerlies have strengthened and shifted southward, spreading the sea ice northward more effectively (Marshall 2003; Zhang 2014). Below the surface layer, the temperature has risen while a freshening is observed in many areas (Gille 2008; Schmidtko et al. 2014; de Lavergne et al. 2014). Simulations performed with coupled climate models are generally not able to adequately reproduce these trends. In particular, the majority of them display a decrease in ice extent over the last 30 years in response to anthropogenic forcing. Part of the discrepancy may relate to the large internal variability of the Southern Ocean, but systematic biases are also present in the simulations (Zunz et al. 2013; Turner et al. 2015; Jones et al. 2016). Even the ocean-ice models driven by prescribed forcing derived from atmospheric reanalyses, such as in CORE-II experiments, have trouble reproducing the mean state of the Southern Ocean. For example, CORE-II models display relatively large biases in the position of the ice edge

all year long and the CORE-II MMM sea-ice extent is lower than observed, particularly in summer (Farneti et al. 2015; Downes et al. 2015). Part of these common biases are related to the common CORE-II atmospheric forcing.

In addition to sea-ice biases, the majority of the CORE-II models underestimate the MLD in summer while some overestimate it in winter, with a clear impact on the characteristics of the intermediate water masses (Downes et al. 2015). On average, the CORE-II MMM winter mixed layer depth bias is positive and dominated by models with a deep mixed layer and more-saline-than-observed upper ocean. Models with warmer and fresher upper ocean produce shallower-than-observed winter mixed layers. Downes et al. (2015) conclude that the uniformly shallow summer mixed layers are mainly a result of the common atmospheric forcing, while in winter many other additional factors, such as sea ice, surface buoyancy fluxes and model parameterisations, affect the mixed layer depth, and result in varying biases in individual CORE-II models.

Deeper in the ocean, several CORE-II models have cold biases associated with positive MLD biases in the regions of the Antarctic Bottom Water formation. The CORE-II MMM shows warm and saline biases north of 50°S, but cool and fresh biases to the south in the upper 2000 m layer. The fresh bias south of 50°S could be linked to the low levels of brine rejection from ice to the surface ocean related to low CORE-II sea-ice extents (Downes et al. 2015). Below 2000 m depth the CORE-II MMM is biased towards a colder and fresher state than the observational WOA09 climatology.

Inter-ocean exchanges play an important role in global climate in response to variations of local or remote heat and freshwater fluxes via the global ocean circulation. This global ocean transport, coupled to global oceanic thermohaline circulation, links the full ocean volume to the climate at long time scales. The Antarctic Circumpolar Current (ACC) is the most intense current of the world ocean and by far the largest conduit for interbasin exchanges.

Farneti et al. (2015) found that the CORE-II MMM Drake passage transport was relatively high (∼ 150 Sv), due to two ensemble members, but close to the Climate Model Intercomparison Project Phase 3/5 (CMIP3/5) MMM transport. After excluding these two CORE-II models, the CORE-II MMM transport became closer to observed estimates of ∼ 130 to 150 Sv. However, as discussed in Sect. 3.5, CORE-II and CMIP ensembles underestimate more recent ACC transport estimates by Donohue et al. (2016) and de Verdiére and Ollitrault (2016).

The CORE-II mass transport time series in the ACC tends to increase during 1948–2008, although this increase flattens toward the end of the period. Interestingly, the eddy-permitting models and models with time-dependent and/or three-dimensional eddy-induced coefficients show lower transport trends than the models with a constant or absent

eddy-induced coefficients. This indicates that models which more realistically represent mesoscale eddy effects do not support long-term increases in the ACC transport, as a response to strengthening westerlies. This ACC insensitivity to the changing winds can be explained by eddy compensation effects at high resolution and advanced eddy-parameterisation models (Farneti et al. 2015).

These ACC transport trends in CORE-II models are in turn related to the upper ocean water mass structure and linked to temperature, salinity and sea-ice trends. As described by Downes et al. (2015), the CORE-II MMM shows cooling south of 60°S and warming north of the ACC core ~ 50°S in the upper 2000 m. Furthermore, the CORE-II MMM shows a general freshening which, along with the upper ocean temperature trends, can be explained by the stronger and southward moving westerlies which increase the ocean surface heat loss and enhance the atmospheric moisture transport (and therefore the precipitation). Another factor playing a role in the freshening is the redistribution of freshwater by sea ice which is often more important in the Southern Ocean than precipitation (Abernathey et al. 2016; Haumann et al. 2016). These model-produced trends bear good a resemblance to those observed.

For some variables such as the sea-ice concentration, observations with a good spatial coverage are available since 1979 from remote sensing. Despite the uncertainties related to the calibration of the satellite records (e.g. Eisenman et al. 2014), this provides valuable information on the state of the system and an essential metric for model validation. The number of subsurface observations has increased over the last decades thanks to Argo floats (Argo 2000) and sensors attached to marine mammals.

Nevertheless, these observations remain relatively scarce, especially below the sea ice (Schmidtko et al. 2014; de Lavergne et al. 2014; Roemmich et al. 2015; Roquet 2015; Pellichero et al. 2017). The amount of in-situ observations for sea-ice thickness is also relatively limited (Worby et al. 2008). Data assimilation is potentially a powerful tool to obtain estimates for variables that cannot be directly observed or have a poor spatial and temporal observational coverage such as the Antarctic sea-ice thickness (Massonnet et al. 2013), the transport of the subpolar gyres (Duan et al. 2016) and the amount and path of deep water formed close to Antarctica (van Sebille et al. 2013; Azaneu et al. 2014).

3 Material and methods

3.1 Ten selected ocean reanalyses

The ORA output data have been collected in a data base hosted by the Integrated Climate Data Center (ICDC) at

Hamburg University[1] and are freely available. Some data were already present from previous ORA-IP studies, but many products were updated and a few new ones added for this study. Ten ORAs were selected to be compared (Table 1), with the most comprehensive temporal overlap over 1993–2010 consisting of all variables required for the diagnostics. The remaining ORAs were discarded due to lack of data either in terms of temporal coverage or variables. Nine ORAs have a global coverage, while one (TOPAZ4) is a regional Arctic-North Atlantic product. Of nine global ORAs, five are of European origin (all using varying versions of the NEMO ocean), three are American and one is Japanese. All variables analysed were monthly means covering the common intercomparison period from 1993 to 2010 with a few exceptions (mentioned in particular subsections of that diagnostic).

For sea-ice diagnostics, Chevallier et al. (2017) analysed eleven ORAs of which eight are participating in this study, while three (GECCO2, SODA3.3.1 and TOPAZ4) were not previously assessed. Only three ORAs of the other eight (ECDA3, ORAP5 and UR025.4) have not been upgraded meanwhile. As the horizontal resolution of ORAs vary we interpolated all fields onto a common regular 1° × 1° latitude-longitude grid for intercomparisons.

Several observational data sets were used to estimate the product-to-observed performance. For the hydrographical analysis, three observational products were used: EN4.2.0.g10 (1993–2010; Good et al. 2013), World Ocean Atlas 2013 (WOA13, 1995–2015; Locarnini et al. 2013; Zweng et al. 2013) and the Sumata Arctic hydrography from Hiroshi Sumata at the Alfred Wegener Institute, Germany based on 1980–2015 observations (Sumata et al. 2017). Notably, the Sumata hydrography is the most comprehensive and up-to-date of the three observational products containing Arctic observations from 28 campaigns from 1980–2015. As for the ORA output, observational data were interpolated onto the common grid for intercomparisons.

3.2 Sea-ice concentration and thickness

Sea-ice concentration (SIC, the relative amount of area covered by ice, compared to some reference area) is the most well-constrained sea-ice variable although not flawless (Ivanova et al. 2014). Satellite observations using passive microwave sensors exist since 1979, available on a daily basis since 1987 at a horizontal resolution finer than 25 km. Chevallier et al. (2017) evaluated various aspects related to sea-ice concentration: the position of the ice edge, sea-ice concentration in the marginal ice zone (concentrations from 15 to 90%) and in the pack ice (concentrations > 90%),

---

[1] http://icdc.cen.uni-hamburg.de/daten/reanalysis-ocean/oraip.html.

representation of leads within the pack ice, seasonal cycles and trends of integrated Arctic sea-ice area and sea-ice extent.

We use these metrics to evaluate seasonal cycles of sea-ice concentration in both the Arctic and Southern Oceans in the new set of reanalyses. Due to the inclusion of one regional Arctic reanalysis that excludes the North Pacific, the Arctic-integrated extent and area are calculated over a reduced Arctic domain closed at the Bering Strait. We use the same observational datasets as in Chevallier et al. (2017) to assess the realism of ORAs, while taking into account observational uncertainties. Specifically, these observational sea-ice concentration products are based on the NASATeam algorithm of the National Snow and Ice Data Centre (NSIDC; Cavalieri et al. 1999), from Ifremer/CERSAT using the ARTIST algorithm, and by EUMETSAT Ocean-Sea Ice Satellite Application Facilities (OSISAF). Although these three products have resolutions finer than 25 km, all data are interpolated onto the common regular grid.

Sea-ice thickness (SIT hereafter) is a key diagnostic for assessing the performance of ORAs in the polar oceans. An unrealistic reconstruction of SIT would mean that essential thermodynamic processes controlling ice growth or melt are missing, or that the dynamics of the sea-ice pack is not captured accurately, or both. A major obstacle for the assessment of SIT is the lack of observationally-based data. Unlike sea-ice concentration no large-scale and time-homogeneous records of sea-ice thickness are available.

For the Arctic sea-ice thickness, most of our knowledge relies on collections of datasets from various sources (e.g. Lindsay 2010). Chevallier et al. (2017) used estimates of sea-ice thickness from the ICESat instruments, and estimates of sea-ice volume gathered in Zygmuntowska et al. (2014). In our study, data from the Ice Thickness Regression Procedure (ITRP) are used to analyze the ORA performance. We selected two 2-month periods (February/March and October/November) for the comparison because the ICESat data are available in these months. The ITRP combines upward looking sonar, airborne electromagnetic, NASA operation Icebridge, and ICESat remote sensed ice thickness observations, as explained in detail by Lindsay and Schweiger (2015). Despite the fact that the ITRP thickness data are a result of complex data processing, we believe that the ITRP is the best data set to compare models with. This is due to the following: it allows to calculate sea-ice thickness deviations per grid cell and to integrate total sea-ice volumes in the ITRP region. These metrics are calculated for the period of 2000–2012, with which the ORAs are compared, with the exception of UR025.4 which ends in 2010.

The most comprehensive database adapted for the purpose of evaluating the Antarctic SIT of ORAs is ASPeCt (Worby et al. 2008). This product covers the period 1981–2005 and comprises about 23,000 individual measurements made during ship voyages or helicopter campaigns in the Southern Ocean. Sea-ice thickness was estimated visually by experts onboard. It is therefore likely (1) that systematic errors are present: ships tend to circulate in thin ice, hence estimations are probably biased thin, and (2) that random measurement errors are large, due to the rather simplistic method of measurement (see Worby et al. 2008, for further discussion). The assessment of ORAs with respect to ASPeCt should therefore be conservative and made with extreme caution, in order to not discard ORAs for the wrong reasons.

Unlike the ORA-IP dataset, the ASPeCt data is not gridded and is provided as daily and not monthly values, which complicates further the assessment. We first binned the ASPeCt data in space and time by matching each of the ∼ 23,000 ASPeCt measurements to the corresponding ORA $1° \times 1°$ grid cell, year and month over 1993–2005. The number of measurements varies greatly from case to case, but is generally low: in 57% of the cases (one case means one given grid cell during one given month of one given year), less than three measurements are available. We excluded these cases with too few data from our assessment, to limit the probability of detecting a mismatch by chance. For all other cases (four ASPeCt measurements or more in a given month of a given year in a given grid cell), we tested whether the ASPeCt measurements and the ORA-IP monthly mean values could be drawn from the same statistical distribution. For each case, we claimed the ORA product to be 'compatible' with ASPeCt if the ORA estimate fell within the range of all available ASPeCt measurements. In addition we recorded for each case an 'error' equal to the difference between the reanalysed SIT and the mean value of ASPeCt measurements, and an "absolute error" equal to the absolute value of the previous metric. The choice of the threshold of at least four ASPeCt measurements to conduct the comparison does not have an impact on the conclusions (not shown here).

Note that Chevallier et al. (2017) carried out a thorough evaluation of the Arctic sea-ice drift in the ORA ensemble, which is not done here for either the Arctic or Antarctic. Sea-ice dynamics is primarily wind driven. Most of the reanalyses considered here use the same atmospheric reanalyses as in the ensemble considered by Chevallier et al. (2017), and there were no significant updates in the model physics regarding sea-ice dynamics or rheology. Thus, we can assume that our sea-ice drift results are consistent with those of Chevallier et al. (2017). Hence we refer to their findings, where necessary.

3.3 Snow depth

Current sea-ice models simulate snow on ice in rather rudimentary ways. Due to its low thermal conductivity and high albedo, snow is strongly altering the snow-ice energy

balance. Both thermal conductivity and albedo depend on the snow density which is kept constant in ORAs ($\sim$ 330 to 342 kg m^{-3}), while observations report a seasonal range of 250–320 kg m^{-3} from September to May (Warren et al. 1999; Chevallier et al. 2017). Most of the models melt all snow in a grid cell before sea ice is melted at the surface. Many snow related processes (such as precipitation, wind, ice drift and deformation, flooding, melting, evaporation and sublimation) are very uncertain and crudely parameterized in models.

Snow depth observations are very sparse in both polar regions, and in particular in Antarctica. A primary Arctic source is the snow depth climatology of Warren et al. (1999) which is based on data from drifting stations established typically on multi-year sea ice with relatively thick snow cover and collected over the past decades (1954–1991). Due to this, we keep in mind that the Warren climatology is likely overestimating the pan-Arctic average snow depth.

3.4 Mixed layer depth

The oceanic mixed layer constitutes the interface between the atmosphere and the interior of the ocean. This layer is where all dynamic, thermodynamic and biogeochemical air-sea exchanges take place, and where the world's deep water masses acquire their properties (e.g. de Boyer Montégut et al. 2004; Holte and Talley 2009). As the MLD is a relevant physical index of the vertical mixing intensity in the upper ocean (Toyoda et al. 2017a), the MLDs simulated by the ORAs are evaluated against two observation-based products. These are the Monthly Isopycnal and Mixed-layer Ocean Climatology for the Arctic (MIMOC; Schmidtko et al. 2013) and a recently published Southern Ocean mixed layer climatology (Pellichero et al. 2017).

These products are both based on temperature and salinity profiles from ship observations archived in the World Ocean Database, as well as from float data from the Argo international program. In addition, MIMOC includes data recorded by ice-tethered profilers in the Arctic Ocean, while Pellichero et al. (2017) use observations from animal-borne sensor programs in the Southern Ocean (Roquet et al. 2017). These contemporary sources provide an unprecedented data coverage of the sea-ice regions over the entire seasonal cycle. Both climatologies are constructed using an objective mapping of the MLDs computed from instantaneous profiles with the Holte and Talley (2009) algorithm. By contrast, reanalysis MLDs are obtained from monthly mean temperature and salinity fields, using a density threshold of 0.03 kg/m^3 with respect to the value at 10 m depth.

Table 2 Sections used for calculating net lateral volume, heat and freshwater exchange between the Arctic and Sub-Arctic

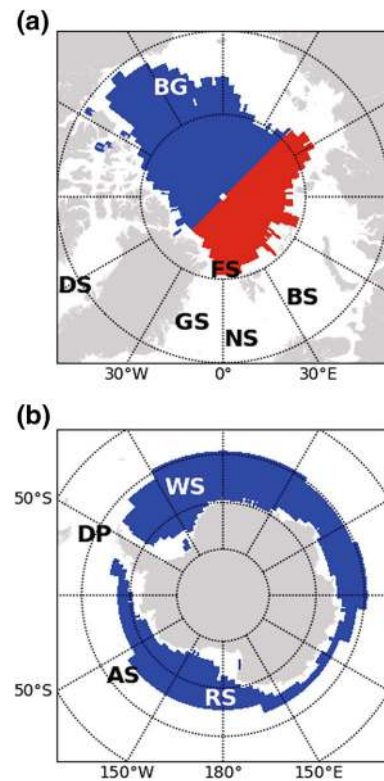| Section	Latitude	Longitude
Fram strait	N79°30′	W20°–E11°
Barents sea opening	N70°30′–N74°30′	E20°
Davis strait	N66°40′	W53°30′–W61°
Bering strait	N66°45′	W168°–W170°30′

Fig. 1 Regions used to calculate average temperature and salinity profiles. In **a** red colour shows the region of the Eurasian basin and blue colour the Amerasian basin, while in **b** blue colour shows the Antarctic open ocean. Annotations: the Fram Strait (FS), Beaufort Gyre (BG), Barents Sea (BS), Davis Strait (DS), Greenland Sea (GS), Norwegian Sea (NS), Amundsen Sea (AS), Ross Sea (RS), Weddell Sea (WS) and Drake Passage (DP)

As noted by de Boyer Montégut et al. (2004), MLDs computed from monthly, hence smoother, profiles can be underestimated approximately by 10–20 m compared to those based on instantaneous profiles. This is mostly the case in spring when rapid restratification occurs (Toyoda et al. 2017a), and needs to be kept in mind when carrying out ORA evaluation. On the other hand, Holte and Talley (2009) found that their algorithm tends to yield slightly shallower MLDs in winter than the density threshold method.

3.5 Liquid ocean transports

Lateral oceanic volume (V), heat (Q), and liquid freshwater transports are calculated through four sections nearly closing the Arctic (see Table 2; Fig. 1). The calculated values represent net transport through the openings, with positive values towards the Arctic. Heat transport is calculated relative to $T_{ref} = -0.1$ °C (Aagaard and Greisman 1975). Liquid freshwater transport is calculated relative to $S_{ref} = 34.8$ on the dimensionless practical salinity scale (Aagaard and Carmack 1989).

Observational ocean transport estimates are obtained from literature, and thus do not represent a consistent time span. Furthermore, their calculations required some assumptions due to discrete spatial sampling of observations. Hence, the observations do not fully close the Arctic Ocean transport budget.

Specifically, the oceanic flow through the Fram Strait constitutes the main volume and heat exchanges between the Arctic and the Atlantic with a complex re-circulation structure. The total northward flow is estimated as 7 Sv, while a total southward flow of ~9 Sv yields a net southward transport of ~ 2 Sv (Table S1; Fahrbach et al. 2001). The heat carried northward along the western coast of Svalbard has shown a relatively large inter-annual variability, between 26 TW (1997/98) and 50 TW (2003/04) (Schauer and Beszczynska-Möller 2009). The flow through the Barents Sea Opening (BSO) towards the Arctic has a net volume flow of 2.3 Sv with about 70 TW heat transport (Table S1; Smedsrud et al. 2013). However, most of this oceanic heat is lost to the atmosphere while en route across the shallow Barents Sea shelf upon reaching the Arctic Ocean (e.g. Gammelsrød et al. 2009).

Another connection between the Arctic and the Atlantic is through the complex channels of the Canadian Arctic Archipelago. However, most of this exchange is channelled through the Davis Strait in Baffin Bay between Greenland and Baffin Island. Here, observations show a net southward volume transport of 1.6 Sv (Table S1; Curry et al. 2014).

The only connection to the Pacific is the shallow Bering Strait. The volume transport through this passage is estimated to be 0.8 Sv directed northward (Table S1; Roach et al. 1995). However, there is a considerable seasonal cycle from 0.4 Sv in winter to about 1.2 Sv in summer (Woodgate and Aagaard 2005), in addition to a possible positive trend in the recent decade (Woodgate et al. 2012). The Bering Strait also represents the only oceanic net freshwater input to the Arctic. Due to its regional Arctic domain, TOPAZ4 model boundary is located in the Bering Strait where a volume transport of 0.7 Sv to the Arctic is prescribed. As temperature and salinity are not prescribed, we decided it is not meaningful to estimate heat and freshwater transports in the Bering Strait for TOPAZ4. Therefore these TOPAZ4 quantities, and consequently the net Arctic heat and freshwater fluxes, were excluded from the MMM.

When calculating the ocean transports from the ORA results, the Hudson Strait in the Canadian Arctic Archipelago is omitted, as is the part north of the Barents Sea Opening, i.e., the opening between Bear Island and Spitsbergen Island. These choices make the ORA data more easily compared with the observed transports across the same transects. Some of the modelled ocean transports are calculated based on aggregated data which are interpolated in space and averaged in time, excluding short-term variability. Hence, the ORA data also have some shortcomings with respect to closing budgets for the Arctic Ocean.

For the Southern Ocean transports, we present in Sect. 4.2.3 the values of volume transports across the three main transects of the ACC: the Drake Passage; a transect between South Africa and the Antarctica (Fig. 1, called "30°E"); and a transect between Australia and Antarctica (called "147°E"). We compare the values estimated from nine global ORAs to estimates from observations.

During the last three decades, the Drake Passage has been more closely monitored than the other two transects. Ganachaud and Wunsch (2000) estimate 140 Sv (± 6 Sv) using an inverse box model applied to WOCE hydrographic data. With a similar method, Lumpkin and Speer (2007) give a mean net transport of 129.7 Sv (± 6.8 Sv). The canonical value of 134 Sv (± 11.2 Sv), obtained by Cunningham et al. (2003) after reviewing ISOS data deployed from January 1979 to February 1980 (Whitworth and R. 1985), is however widely utilized by the physical oceanography community. More recent estimations with a method combining moorings and altimeter 1993–2012 measurements (Koenig et al. 2014) also give a total net transport of 140 Sv (± 10 Sv).

Recent estimations from Donohue et al. (2016), based on 2007–2011 extensive mooring measurements, and from de Verdiére and Ollitrault (2016), based on time-mean Argo float displacements and historical hydrography from the World Ocean Atlas 2009 are likely to be the most reliable ones. Compared to earlier studies, they used methods that reduce uncertainties in the barotropic flow component due to more comprehensive monitoring array and by global mass conserving mean circulation. Donohue et al. (2016) and de Verdiére and Ollitrault (2016) provide total transport estimations of 173.3 ± 10.7 and 175 Sv, respectively. These values are ~ 30% larger than the canonical value often used as the benchmark for global circulation and climate models.

3.6 Ocean heat and salt contents

Ocean heat and salt contents are denoted as OHC and OSC, respectively. They are calculated as vertical integrals from the reference depth H to the surface η:

Fig. 2 Number of ORAs per grid cell (up to 10) where their sea-ice concentration is > 15% in March (left) and in September (right) based on 1993–2010 monthly data. Black line is the 15% climatological ice edge by NSIDC NASATeam. The number of reanalyses considered here is 10. Note that the Bering Sea and the Sea of Okhotsk are not a part of the domain of TOPAZ4, so only 9 reanalyses have a solution in these areas
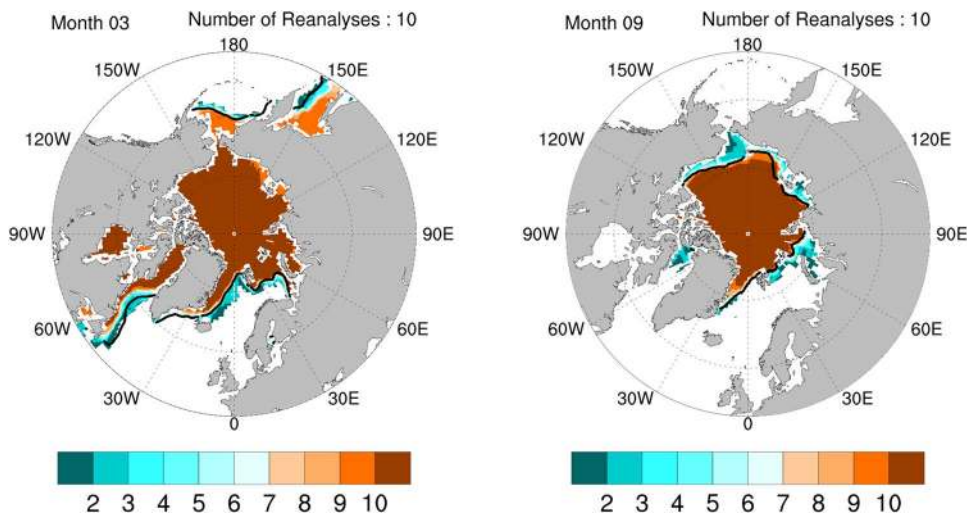
$$OHC = \int_{-H}^{\eta} \theta(z)dz, \tag{1}$$

$$OSC = \int_{-H}^{\eta} S(z)dz, \tag{2}$$

where θ and S are vertical potential temperature and salinity profiles at a horizontal ORA grid point.

The freshwater content, a common oceanographic diagnostic, is the amount of zero-salinity water required to be taken from the ocean or sea ice so that its salinity is changed to the chosen reference salinity and is closely related to OSC and therefore not presented.

3.7 Hydrography

The Antarctic and Arctic ocean basins used to calculate the hydrographic average profiles follow the definitions given in Barthélemy et al. (2015). Arctic Ocean was split into two—the Eurasian basin and the Amerasian basin, along two meridians, 135°E and 45°W, which join at the North Pole (Fig. 1). The boundary between the two basins approximately follows the Lomonosov Ridge from the East Siberian Shelf to the Lincoln Shelf north of Greenland. The reason for this division of the Arctic Ocean was to see whether product performance varies between the two main Arctic basins, for example in terms of the AW advection.

Due to the vertically integrated ORA-IP hydrographic data only waters located over deep parts of the basins are analysed, analogously to OHC and OSC diagnostics. Specifically, domain averages are limited by their depth so that in the Arctic the ocean grid points deeper than 500 m are included, while in the Antarctic the limit was 1000 m. The northern limit of the Antarctic basin is chosen as to ensure that the largest fraction of the area is covered with sea ice in

winter, and therefore represents a polar marine environment. All ten ORAs and three observational products (Sumata, WOA13 and EN4.2.0.g10) were interpolated to a common 1° horizontal latitude–longitude grid, which is identical to the WOA13 grid, before the calculation of regionally averaged hydrographic profiles. As the ORA database does not provide land-sea masks of individual ORAs, we assumed the WOA13 land-sea mask available from the WOA13 website.

First, OHC and OSC for all ORAs were calculated from five reference depths ($H = \{100, 300, 700, 1500, 3000\text{ m}\}$) to the surface ($\eta = 0$ m). After this, the mean potential temperatures and salinities $\langle X = \{\theta, S\} \rangle$ within each layer $100 \rightarrow 0$ m, $300 \rightarrow 100$ m, $700 \rightarrow 300$ m, $1500 \rightarrow 700$ m and $3000 \rightarrow 1500$ m were calculated from $OXC = \{OHC, OSC\}$ as:

$$\langle X_{L \rightarrow U} \rangle = \frac{OXC_{L \rightarrow 0} - OXC_{U \rightarrow 0}}{L - U}, \tag{3}$$

where X is either temperature or salinity, and $\langle X_{L \rightarrow U} \rangle$ its average between levels L and U. $\langle X_{L \rightarrow U} \rangle$ values where L is deeper than the ocean depth at that particular grid point were excluded from the further analysis. Finally, level averaged temperatures and salinities $\langle X_{L \rightarrow U} \rangle$ were temporally and basin-averaged.

4 Results

4.1 Arctic mean states

4.1.1 Sea ice and snow

Ten ORAs show an overall agreement in the location of the sea-ice edge in the Arctic Ocean and along its margins (Figs. 2, S1 and S2), which can be attributed to sea-ice data assimilation and the constraint by the atmospheric forcing. On average, there is a good agreement with respect to the

Fig. 3 Mean seasonal cycle (over the period of 1993–2010) of the Arctic sea-ice extent and area (upper row), and of the area covered by Marginal Ice Zone (MIZ) and pack ice (lower row), in all ORAs (colour lines) and in NSIDC, CERSAT and OSISAF observations (grey shading). Domain of integration excludes the ocean area in the North Pacific south of Bering Strait. MIZ is defined as a region where the sea-ice concentration is less than 90% and greater than 15%, while the pack ice is the region where the sea-ice concentration is higher than 90%. Units are in 10^6 km^2
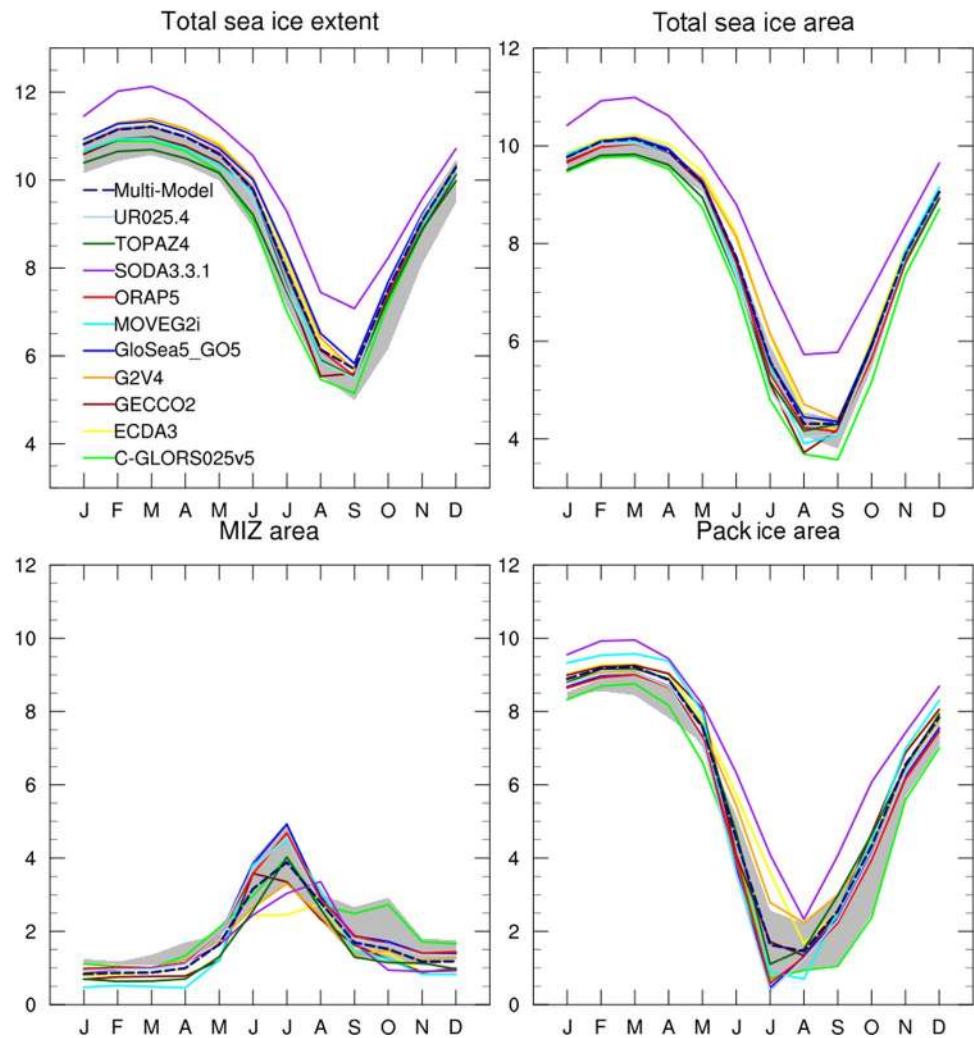
sea-ice edge in the Barents Sea, the Greenland Sea and the Bering Sea. Most reanalyses lack sea ice in the Labrador Sea and the Sea of Okhotsk, as Chevallier et al. (2017) pointed out. A few ORAs simulate too much sea ice eastward of the coasts of the Labrador Sea and the Greenland Sea: these are the ORAs that do not assimilate sea-ice concentration (Table 1; Figure S1). In summer, a number of ORAs underestimate the presence of sea ice east of Greenland, and some underestimate sea-ice melt near the shelves, in the Kara Sea and in Baffin Bay.

Figure 3 shows the seasonal cycles of Arctic sea-ice extent and area in ten ORAs. The modeled seasonal cycle is generally in phase with observations, with a maximum (minimum) sea-ice area and extent in March (September), although a few ORAs simulate sea-ice extent minima in August. SODA3.3.1 overestimates sea-ice extent and area in all months, so it is excluded from the subsequent Arctic sea-ice concentration ensemble analysis. The ensemble spread of ORA sea-ice extent, without SODA3.3.1, is limited over

the year, and is comparable to the estimated observational uncertainty. This was expected, since most reanalyses assimilate sea-ice concentration. The spread is larger during the winter months, and all ORAs align well during refreezing in autumn. A few ORAs exhibit systematic biases compared to the observations in the winter months, which is consistent with the lack of sea ice in the Labrador Sea, as noted above. In most ORAs, the simulated August–September sea-ice extents are within the observational uncertainty. Results are similar for sea-ice area, although its ensemble spread is larger in spring and summer than the sea-ice extent spread. For both sea-ice extent and area, the MMM mean without SODA3.3.1 is near the upper range of the observational estimates.

The significant spread in sea-ice area denotes differences in the distribution of sea-ice concentration within the ice cover. As in Chevallier et al. (2017), we investigate the separate contributions of Marginal Ice Zone (MIZ) and pack ice in the total area spread. In the observations, the MIZ
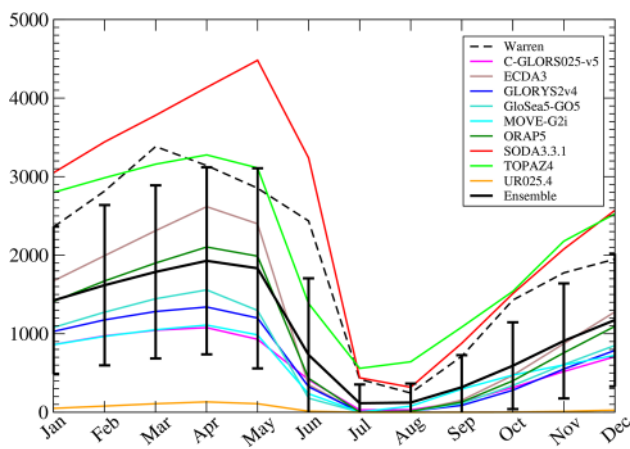
Fig. 4 Monthly climatology of the Arctic snow volume (km^3) of the ORA-IP models for the period from 1993 to 2010, its ensemble mean (black solid line—errors bars designate one standard deviation uncertainty) and the snow volume of the Warren climatology (black dashed line). The snow volume is calculated for the entire Arctic Ocean excluding regions south of the Fram Strait

area varies between 1 and 2 million km^2 from November to April, peaks in July, and decreases slowly from August to October (Fig. 3). Three observational products give consistent results, although CERSAT has a systematically smaller MIZ area in June–September. During October–December, the spread among the observational estimates is the largest, when NSIDC has a larger MIZ than the others. The pack-ice area has a seasonal cycle evolving at the same rate as total sea-ice area, although its annual minimum is reached in July–August. In the Arctic Ocean, sea ice is predominantly pack ice, except in summer when the MIZ/pack-ice area ratio is over 50%.

The ORAs reproduce these seasonal sea-ice extent and area cycles relatively well. Most ORAs are consistent with the ice product they assimilate (e.g. C-GLORS025v5 with NSIDC, GLORYS2v4 with CERSAT; Table 1). However, during winter and early spring, all ORAs simulate MIZ area lower than observed, and systematically too high pack-ice area when the assimilated ice product is taken into account (lower right panel of Fig. 3). In summer, the ensemble spread is larger, and there are a number of ORAs that align well with observational estimates. But no ORA simulates more MIZ than observed, and a few ORAs stand out with a lower-than-observed MIZ peak area: those are the products without data assimilation (Table 1). They tend to simulate very high sea-ice concentration almost all year long (not shown).

The snow volume in the ORAs varies widely—not only between the ORAs using different precipitation data sets but also between the ORAs using ERA-Interim precipitation rates (Fig. 4; Table 1). As apparent from Figs. 4 and

S3, ORAs have a thinner snow cover everywhere in the Arctic and hence smaller snow volumes than Warren et al. (1999), which is known to have a thick bias, as explained earlier (Figs. 4, S3). The maximum snow volume in the Warren climatology occurs between March and April with values around 3000 km^3. The ORA values range between > 4000 km^3 (SODA3.3.1) and < 200 km^3 (UR025.4). By inspecting the ORA ensemble mean and its standard deviation we can identify three ORAs which deviate most from the other ORAs: UR025.4 which has almost no snow at all, SODA3.3.1, driven by the MERRA2 reanalysis and associated with a high bias in sea-ice area, which exceeds the Warren climatology for all months, and TOPAZ4 which fits very closely to the Warren climatology, despite being driven by ERA-Interim. The remaining ORA snow volumes range from about 1000 km^3 (MOVE-G2i) to 2500 km^3 (ECDA3). The large variation between the ORAs driven by the same reanalysis (ERA-Interim) is surprising. This might point to large uncertainties in process parameterisations (related to for example sea-ice ridging and sublimation) which alter the snow depth.

All ORAs show a strong decrease of the snow volume from May to June (Fig. 4). This is certainly connected to the fact that ORAs first have to melt all snow off before their sea ice starts to melt. Related to this, all ORAs except SODA3.3.1 and TOPAZ4 have almost no snow on ice from July to August. Then from September to December the majority of ORAs (except UR025.4, SODA3.3.1 and TOPAZ4) show only moderate differences in the snow volume. Interestingly, differences between the ORA snow volumes grow strongly from January to April.

The mean difference of the sea-ice thickness of the ORAs relative to the ITRP data for February–March is presented in Fig. 5. Most ORAs underestimate the ice thickness north of the Canadian Arctic Archipelago, north of Greenland and the Fram Strait. Especially large deviations are found for ECDA3, MOVE-G2i, SODA3.3.1, and UR025.4 for which the deviations can amount to more then 2 m. More moderate deviations are detected for C-GLORS025v5, GECCO2, GloSea5-GO5, and TOPAZ4. ORAP5 exhibits only a minor underestimation while GLORYS2v4 overestimates the ice thickness by up to 1 m. In the Beaufort Sea, some of the ORAs overestimate the ice thickness moderately (C-GLORS025v5, GloSea5-GO5, SODA3.3.1) while ORAP5 exceeds the observed thickness by up to 1 m and GLORYS2v4 by up to 2 m. TOPAZ4 and GECCO2 show no notable deviations in the Beaufort Sea. Most of the ORAs overestimate the thickness over the Eurasian shelves. GLORYS2v4 strongly overestimates ice thickness over almost the whole Arctic Ocean. In October–November, the ORA-ITRP mean differences generally appear similar to the differences in February–March, but with a tendency towards larger underestimations of sea-ice thickness (Figure S4).
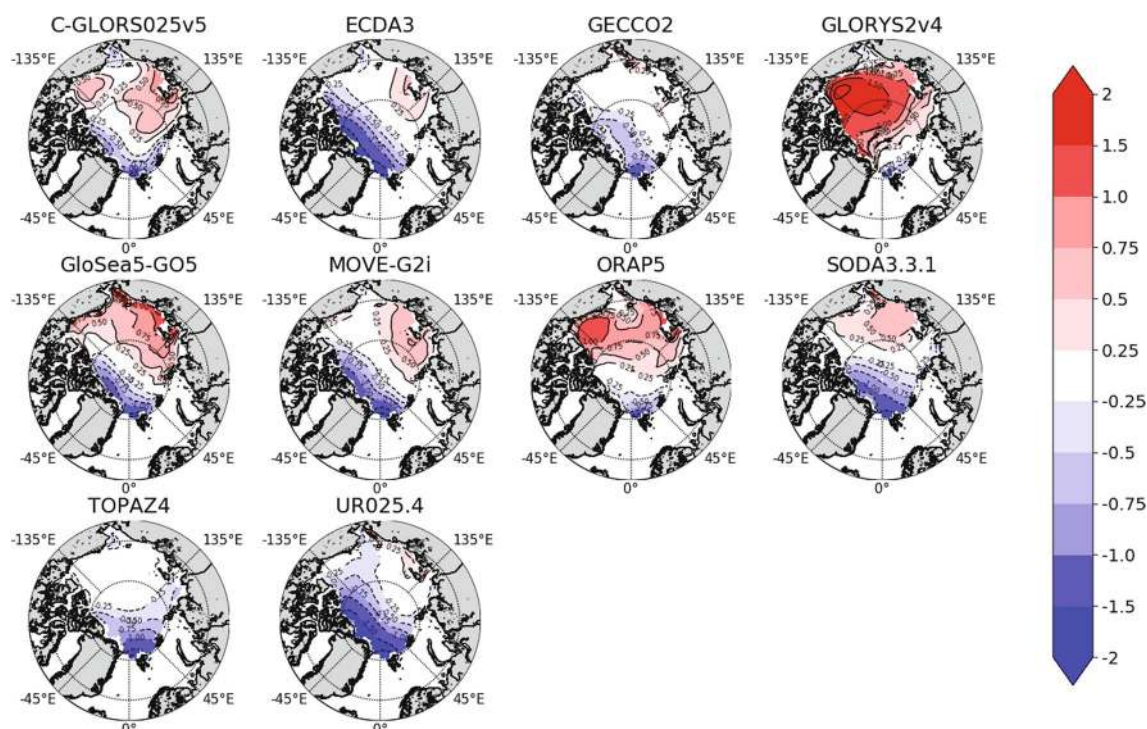
Fig. 5 The 2000–2012 mean difference of the ORAs to the ITRP sea-ice thickness (m) in February–March

In February–March the mean (period 2000 to 2012) ice volume amounts in ITRP to ~15,400 km^3 (Fig. 6a). Corresponding ORA ice volumes range between 10,500 and 12,800 km^3 with the ensemble mean of ~ 14,500 km^3. Two ORAs are very close to the ITRP value (GECCO2 and GloSea5-GO5), but this is, at least in the case of GloSea5-GO5, due to compensating regional biases (Fig. 5). In October–November the mean ITRP ice volume is about 12,400 km^3, while the ORA range is large, from 5300 to 19,200 km^3 (Fig. 6b). The average ice volume of five ORAs (C-GLORS025v5, ECDA3, GloSea5-GO5, MOVE-G2i and UR025.4) stays low—below 8000 km^3. Correspondingly, the ORA MMM ice volume is much lower than the ITRP value (about 10,000 km^3).

Figure 6c displays the mean sea-ice volume loss between February–March and October–November calculated in the ICESat domain (i.e. the difference between Fig. 6a, b). While the ITRP seasonal volume loss is about 3000 km^3, seasonal volume losses in six ORAs exceed or are close to 5000 km^3, indicating too high seasonal sea-ice volume amplitudes. Accordingly the ORA MMM volume loss is biased high (4500 km^3).

4.1.2 Mixed layer depth

There is no systematic bias in the representation of winter MLDs in the Arctic Ocean in the various ORAs (Figs. 7, S5). UR025.4, ECDA3, GloSea5-GO5 and GLORYS2v4 give the largest MLD overestimations, while MOVE-G2i and GECCO2 yield the strongest underestimations. The observed pattern, with MLDs around 40 m in the Amundsen and Makarov Basin and with shallower mixed layers in the Amerasian Basin, is not closely matched by any of the products. In the Barents Sea and south of Svalbard, all ORAs simulate deeper mixed layers than in the observation-based product (Figure S5). The difference between the ensemble mean and the climatology exceeds 400 m locally.

In summer, all ORAs underestimate MLDs (Figure S6). These shallow mixed layers are generally not due to the coarse vertical grid, because the top-level thicknesses of ocean models are mostly from 1–3 m, only ECDA3, GECCO2 and SODA3.3.1 have the top-layer thicknesses of 10 m (Table 1). The ensemble mean bias reaches as much as 20 m under sea ice (Fig. 7), although the reliability of the climatology might be questioned in the regions just north of the Canadian Arctic Archipelago and Greenland, where the ice is very thick and few hydrographic measurements exist. As a result, the MMM negative bias is of the order of 10 m in the Barents Sea and is smaller in the Greenland and Norwegian Seas.
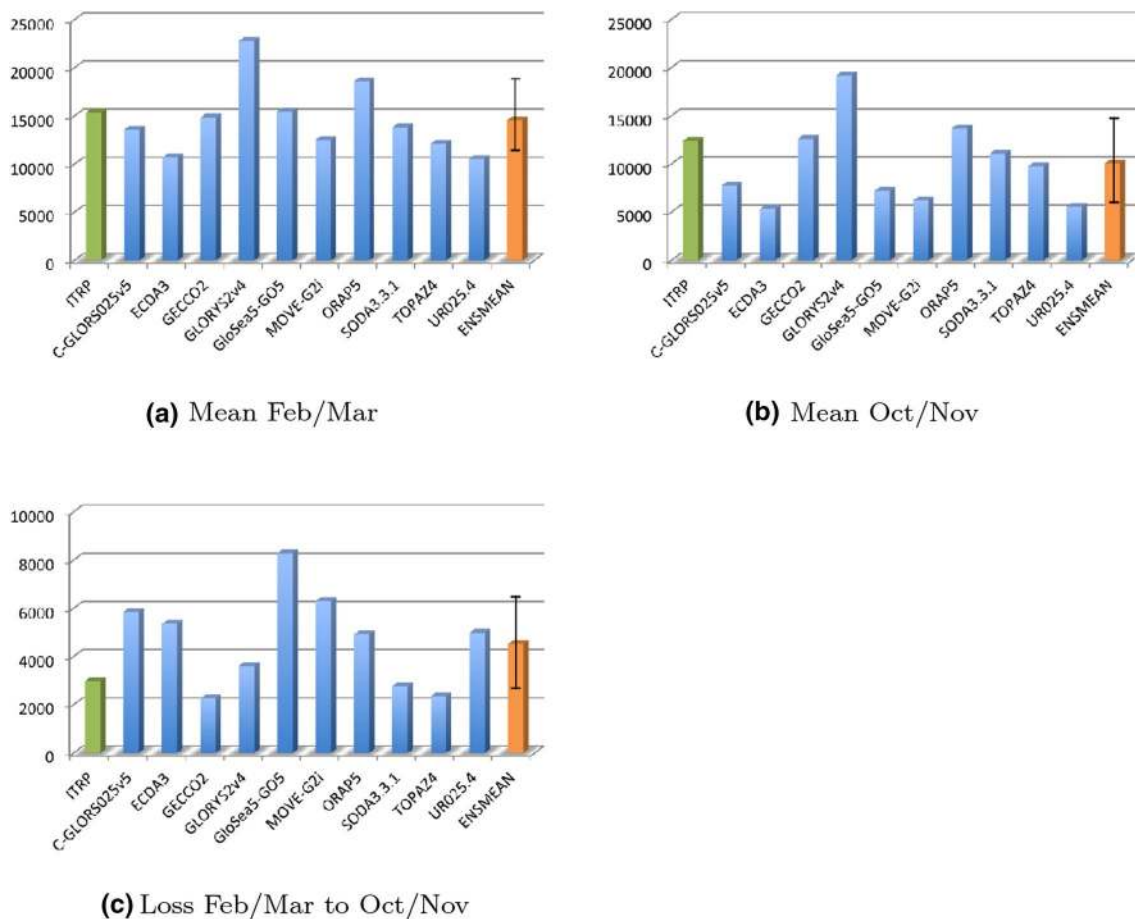
(a) Mean Feb/Mar

(b) Mean Oct/Nov

(c) Loss Feb/Mar to Oct/Nov

Fig. 6 The 2000–2012 mean sea-ice volume (km^3) in the ICESat domain in **a** February–March and **b** October–November. **c** Mean ice volume loss km^3 in the ICESat domain between February–March and October–November (the difference of **a** and **b**). The ORAs are denoted by blue bars, the ITRP by green bars and the ensemble mean (ENSMEAN) by orange bars. The error bar in ENSMEAN represents the ORA ensemble spread (standard deviation)

4.1.3 Liquid ocean transports

The net oceanic exchange between the Arctic Ocean and the sub-Arctic seas through the four major openings, the Fram Strait, Bering Strait, Davis Strait, and BSO, from ORAs and observations are summarized in Table S1 and illustrated as bar plots in Figs. 8, S7 and S8. Generally, ORA volume transports are within the observational uncertainty and for the BSO also heat transports, yet with some notable differences. All ORAs close the Arctic Ocean volume transport budget comparably to the observations. All ORAs tend to be on the low side in terms of net oceanic heat transport towards the Arctic (Fig. 8). This could imply a negative temperature bias in the ORAs. Indeed, in the Barents Sea, WOA13 and EN4.2.0.g10 are warmer than the ORA MMM, but Sumata is slightly cooler than the MMM (Figure S9). GECCO2 is an exception which also shows excessive heat transports through the BSO. Most of the ORAs also tend to be on the low side with respect to heat transport through the

Fram Strait—due to either too cold northward flowing AW or southward recirculation of too warm AW. However, some ORAs are within the uncertainty range of the observations. A caveat to this analysis is our method for computation of the heat transports. For comparison with observation-based estimates from literature, we have chosen to compute the heat transports based on net volume transports through each separate section and relative to a fixed reference temperature ($T_{ref} = -0.1\ °C$; Aagaard and Greisman 1975), which is also the common method used in the literature. However, this method has some inherent inconsistencies related to the lack of a closed volume transport budget and the actual temperature difference between the incoming and outgoing water masses. A more consistent method for the computation of ocean heat transports is discussed in Schauer and Beszczynska-Möller (2009).

The ORAs show a generally good agreement with observations with respect to freshwater export from the Arctic, except GLORYS2v4 and MOVE-G2i, which have
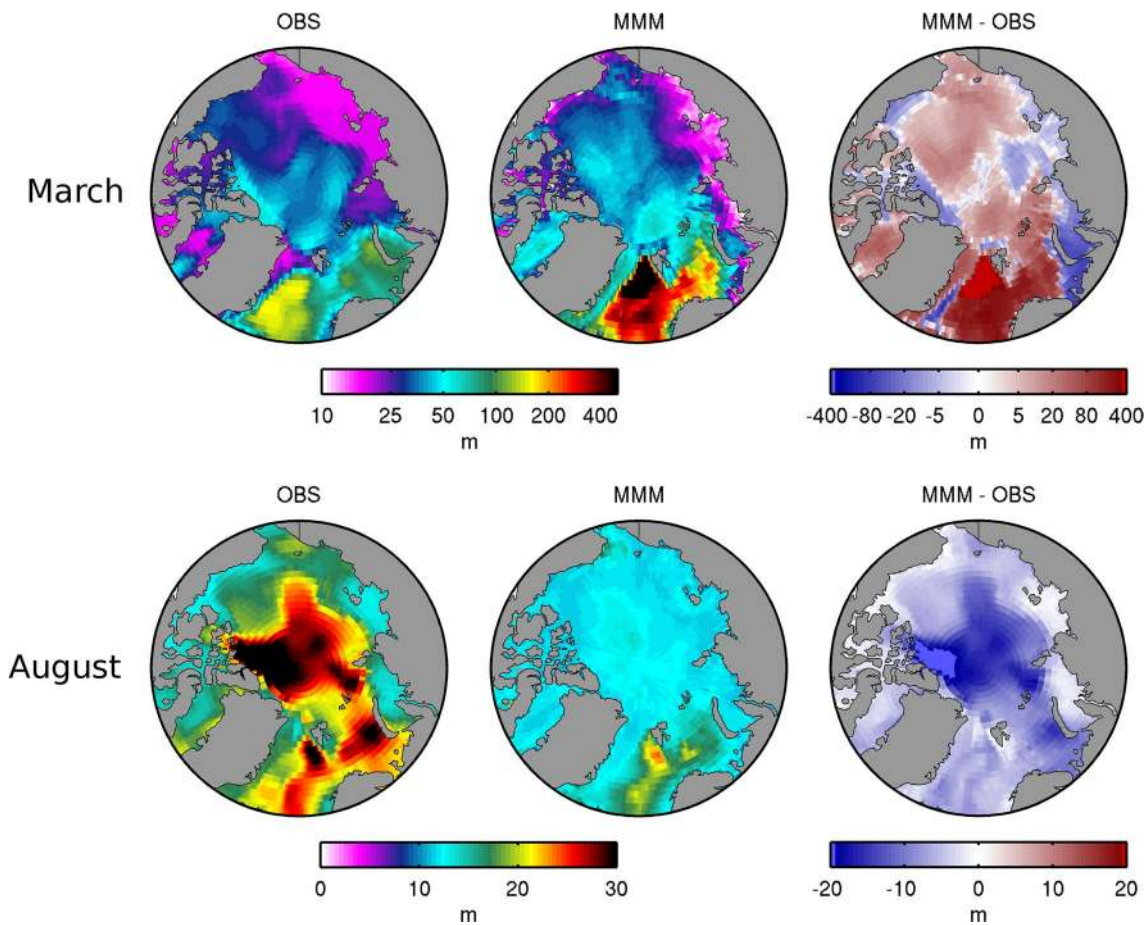
Fig. 7 Arctic mixed layer depth for March (top) and August (bottom), in the MIMOC climatology (OBS; Schmidtko et al. 2013), for the ORA ensemble mean (MMM) and the bias of the ensemble mean with respect to the climatology (MMM–OBS). Note the logarithmic colour scale used for March fields
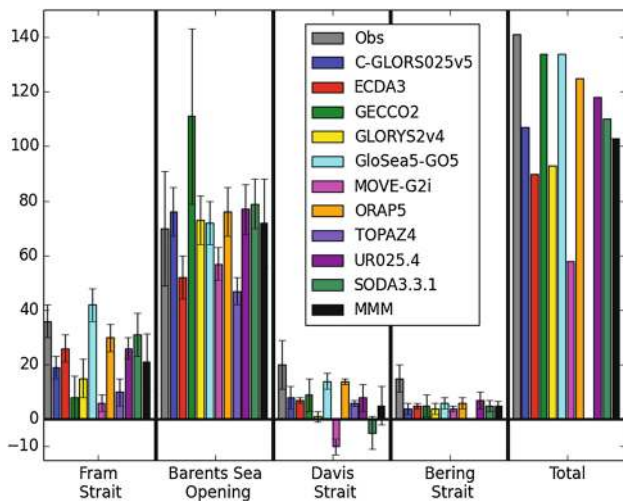

Fig. 8 Mean 1993–2010 Arctic liquid ocean heat transport through the major Arctic Ocean openings, and their net sum (total) in TW. Error bars represent standard deviations of monthly values. See Supplementary Table S1 for values of the error bars and references to the calculation of error bars in the observations

particularly low freshwater exports (Table S1; Figure S8). The reason for this discrepancy is not clear, but GLO-RYS2v4 has a positive salinity bias in the Arctic Ocean (see Sect. 4.1.4). On the other hand, C-GLORS025v5 shows a good agreement with observations in terms of total Arctic freshwater budget, but it has a different distribution with low freshwater volumes exported through the Fram Strait and an enhanced export through the Davis Strait.

The MMM represents an estimate of Arctic—sub-Arctic exchanges comparable to observed estimates (Table S1; Figure S8). The MMM freshwater transport through the Bering and Davis straits are in close agreement with the observations, while the MMM freshwater transport through the Fram Strait is on the low side, although the volume transport is comparable to the observations. Overall, the MMM is generally in closer agreement with the observations than individual ORA estimates.

In terms of heat transport variability, represented by one standard deviation based on annual averages, all ORAs overlap with the observed range of variability in the main

Fig. 9 Averaged multi-model Arctic Ocean Heat Content (OHC) in the layer of 0–1500 m (top left panel), its spread (top second left panel) and individual product departures from the multi-model mean
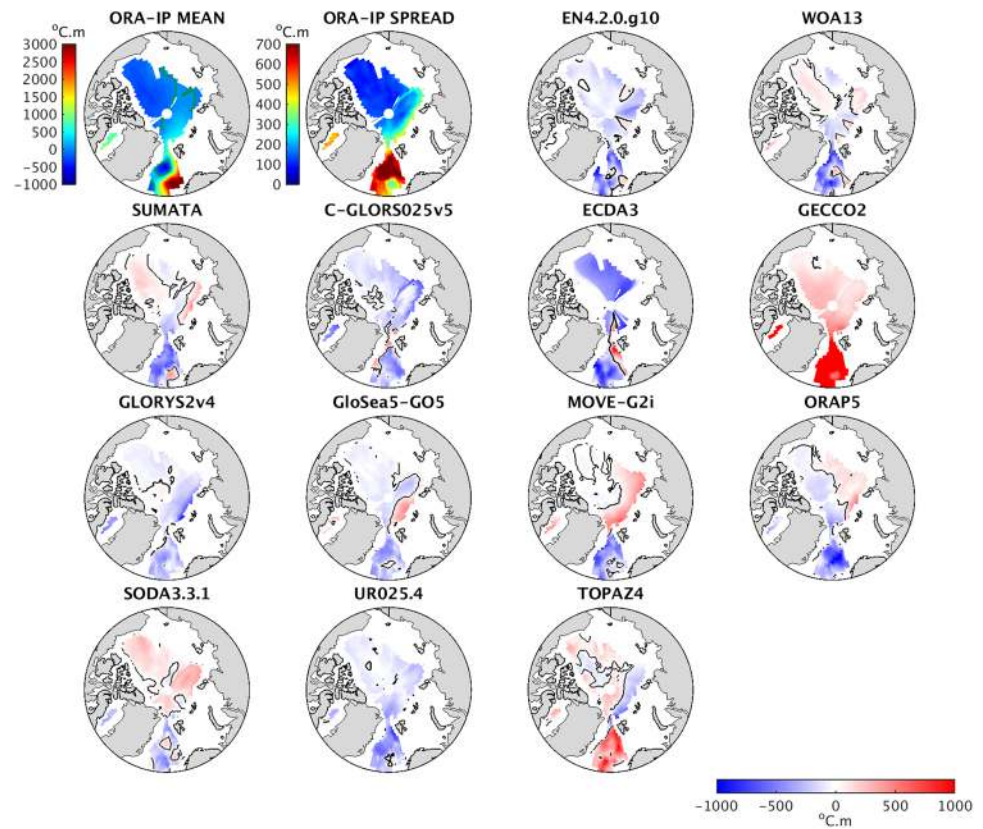
gateway to the Arctic, i.e. the BSO (Fig. 8). Through the other openings, several ORAs have means and associated ranges of variability outside the range of the observed variability, when using one standard deviation. Although the MMM is very close to the observations in terms of heat transport through the BSO, the generally lower than observed modelled heat transports through the other openings causes the MMM to be on the low side regarding overall heat transport to the Arctic. Note, however, that the model and observational periods are overlapping but not equal in length.

4.1.4 Hydrography

We begin with the integrated heat content in the top 1500 m for all products (Fig. 9), with anomalies for each product shown relative to the ORA MMM. The MMM is warmer in the Eurasian basin than the Amerasian, and much warmer in the eastern Nordic Seas than the western, reflecting the path of the warm AW northward and cold Arctic water exiting the Fram Strait. The observational products, Sumata and WOA13, are slightly warmer in the Arctic than the MMM, particularly in the Amerasian basin, but consistently colder in the Nordic Seas. The third observational product, EN4.2.0.g10 is slightly cooler in the Arctic than Sumata and WOA13, with OHC very close to the MMM. In the

Nordic Seas the MMM is clearly biased warm by 2 products, GECCO2 and TOPAZ4. GECCO2 and SODA3.3.1 are warm outliers and ECDA3 a cold outlier in the main Arctic, but other products are all fairly consistent with each other and with the MMM. Evidence of the warm Atlantic boundary current in the south Eurasian basin is very weak in the MMM and in WOA13 and EN4.2.0.g10, but shows up clearly in Sumata. Some ORAs also show it more clearly, such as GloSea5-GO5, MOVE-G2i and ORAP5, while others show its absence, such as GLORYS2v4 and TOPAZ4. This is seen as enhanced ORA spread in the boundary current region.

The integrated salt content in the top 1500 m for all products is shown in Fig. 10. The MMM shows the strong contrast between the fresh Amerasian basin, which is influenced by low salinity inflows (originating from the Pacific, Arctic rivers and precipitation) captured in the Beaufort Gyre, and the more saline Eurasian basin, dominated by Atlantic inflow from the much more saline Nordic Seas. All observational products Sumata, WOA13 and EN4.2.0.g10 show a fresher Beaufort Gyre north of Alaska, with the rest of the Arctic mainly showing more saline than the MMM. The salinity spread among ORAs is larger in the Arctic Ocean than in the Nordic Seas, especially in the Amerasian basin, in contrast to the OHC spread. However, there is considerable cancellation between individual ORA salinity anomalies, suggesting

Fig. 10 Averaged multi-model Arctic Ocean Salt Content (OSC) in the layer of 0–1500 m (top left panel), its spread (top second left panel) and individual product departures from the multi-model mean
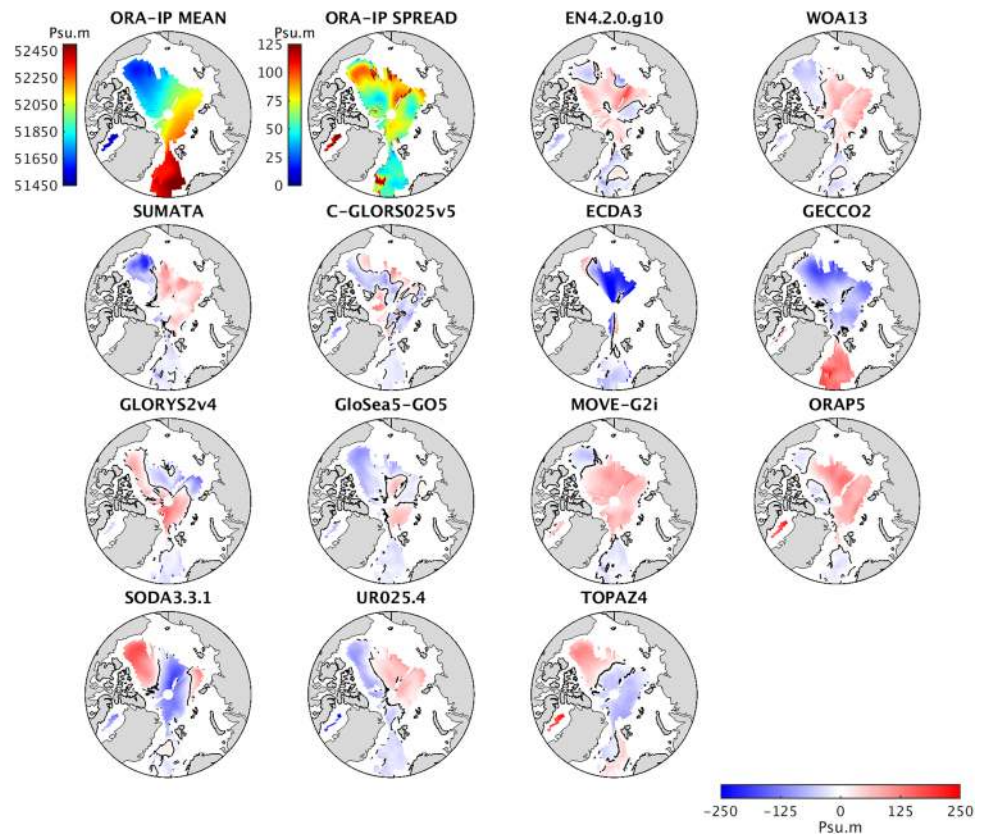

Fig. 11 Seasonal cycle of averaged monthly temperature from 1993 to 2010 in the layer of 0–100 m in **a** the Eurasian basin and **b** in the Amerasian basin. In the leftmost panels seasonal cycles for the multi-product mean (MMM) and three observational data sets (EN.4.2.0.g10, WOA13 and Sumata) are shown, while in the middle and rightmost panels individual ORA seasonal cycles are presented
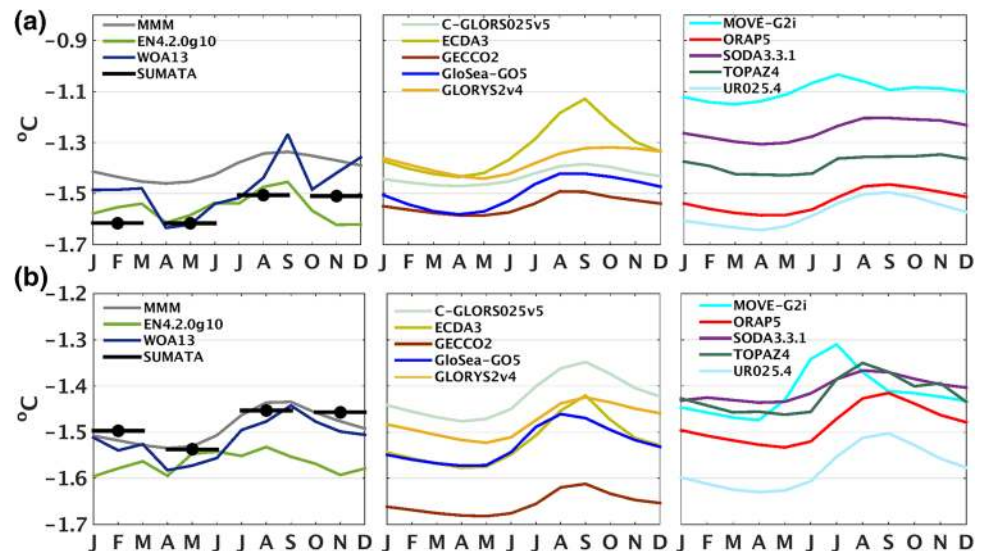
the MMM is a meaningful product. The ORAs most similar to Sumata are MOVE-G2i, ORAP5 and UR025.4, while others have clear spatial anomalies.

Seasonal cycles of OHC and OSC in the 0–100 m layer in terms of mean temperature and salinity are presented in Figs. 11 and 12, respectively. For the mean temperature, the observational products show lower values than the MMM.

The observational seasonal cycles (except Sumata) do not appear smooth (Fig. 11a, b, the leftmost panels) perhaps reflecting sparse observational coverage. The MMM shows a seasonal cycle amplitude that agrees rather well with Sumata, although the MMM appears biased warm in the Eurasian basin. Most individual ORAs also show sensible looking seasonal cycles of mean temperatures with warm

Fig. 12 Seasonal cycle of averaged monthly salinity from 1993 to 2010 in the layer of 0–100 m in **a** the Eurasian basin and **b** in the Amerasian basin. In the leftmost panels seasonal cycles for the multi-product mean (MMM) and three observational data sets (EN.4.2.0.g10, WOA13 and Sumata) are shown, while in the middle and rightmost panels individual ORA seasonal cycles are presented
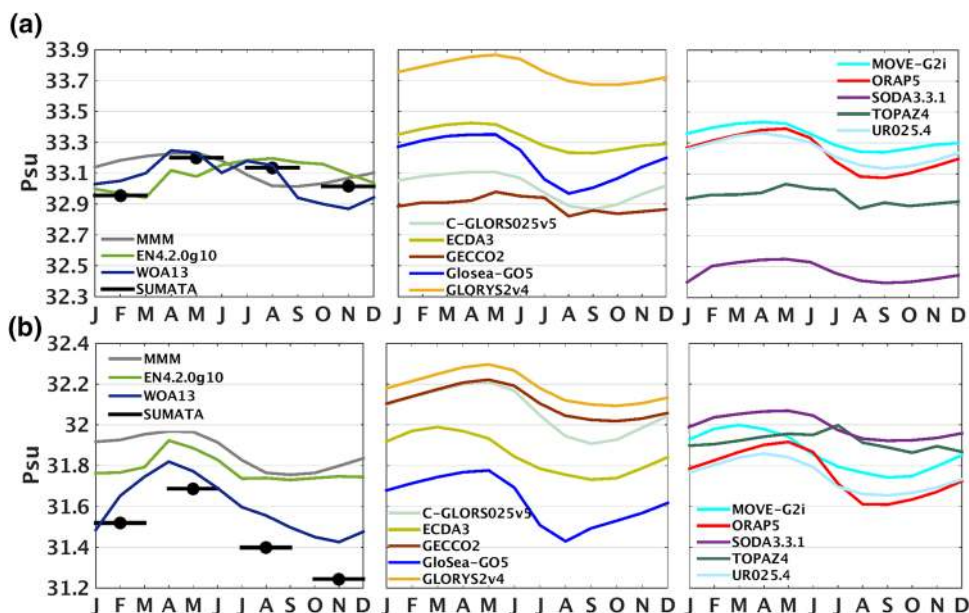
summer and autumn seasons, although their seasonal amplitudes and annual mean temperatures vary (Fig. 11a, b, the middle and rightmost panels). MOVE-G2i seems to have its warmest month earlier than other ORAs, while ECDA3 seems to have a particularly high seasonal amplitude.

For the seasonal cycle of mean salinity, the agreement between the observational products and the MMM is better in the Eurasian basin than in the Amerasian basin (Fig. 12a, b, the leftmost panels). In Sumata salinities in the 0–100 m are higher in spring, then start to decrease until winter. Given the evolution of sea ice and the timing of river runoff to the Arctic Ocean such a seasonal cycle appears sensible. The two other observational products and the MMM mainly agree with this, except the EN4.2.0.g10 in the Eurasian basin, which obtains the highest salinities in August instead of spring. Individual ORAs seem to agree with Sumata and WOA13 in seasonal amplitudes, although TOPAZ4 stands out with its relatively small annual variability.

We now look at the vertical structure of the different products in the Eurasian (above) and Amerasian (below) basin-mean temperatures shown in Fig. 13, using the layers defined in Sect. 3.7. The observational profiles are shown, along with the MMM, in the left-most plots. In the uppermost, 0–100 m layer, Sumata correctly shows the horizontal temperature difference between Eurasian and Amerasian basins with temperatures about −1.6 and −1.4 °C, respectively, reflecting the colder freezing temperatures of the more saline Eurasian basin and the warmer surface waters in the Amerasian basin. Opposite to the horizontal in Sumata's hydrography, WOA13 has the Amerasian Basin slightly colder than the Eurasian Basin and in EN4.2.0.g10 there is practically no horizontal variability with both basins having

surface temperature between −1.6 and −1.5 °C. The MMM shows similar temperatures to WOA13.

Figure 13b, d show basin-averaged errors and their standard deviations of temperature anomalies for all ORA products from Sumata, as probably the most reliable observational product. Notably the standard deviation of error is relatively large compared to mean error in top 300 m, but becomes relatively small at deeper layers indicating more systematic basin scale biases. In the Eurasian basin Sumata is warmer than the MMM from 100–700 m but slightly colder from 1500 to 3000 m (Fig. 13b). This probably reflects an AW deficit of the MMM in the both basins. In the Eurasian basin GloSea5-GO5 matches Sumata most closely in the Atlantic water layers, 100–700 m and ECDA3, TOPAZ4 and ORAP5 are cold outliers. In the 700–1500 m layer in the both basins all products except GECCO2 are inside the range of the observational products. The dominant warm layer in GECCO2 is 700–1500 m, while the dominant warmer layer in SODA3.3.1 is above 300 m in both basins, and ECDA3 has a cold bias in all layers below 100 m. In the Amerasian basin Sumata is clearly warmer than the MMM in all layers below 300 m (Fig. 13c, d). This is because the Canadian Basin Deep Water (CBDW) is known to be warmer (close to −0.5 °C) than the Eurasian Basin Deep Water (EBDW). The MMM shows a difference between the basins, but not as large as Sumata or WOA13. In the Amerasian basin, the 300–700 m layer is the AW layer and the underestimation of its temperature reflects difficulties to correctly capture the boundary current in ORAs (Fig. 13d).

Turning to the vertical structure in salinity, S of each layer in each basin are shown in (Fig. 14). At these cold

Fig. 13 Averaged temperature profiles (**a**, **c**) and their departures (errors) from the Sumata climatology (**b**, **d**) in the Eurasian (**a**, **b**) and Amerasian (**c**, **d**) basins shown in Fig. 1. In **a**, **c**, thin black lines show the non-depth averaged Sumata temperature profiles. In **b**, **d** horizontal bars indicate mean errors and black, horizontal lines their standard deviations as error ± deviation. In cases where the mean error or its standard deviation is too large to fit inside the panel, their values are indicated as numbers
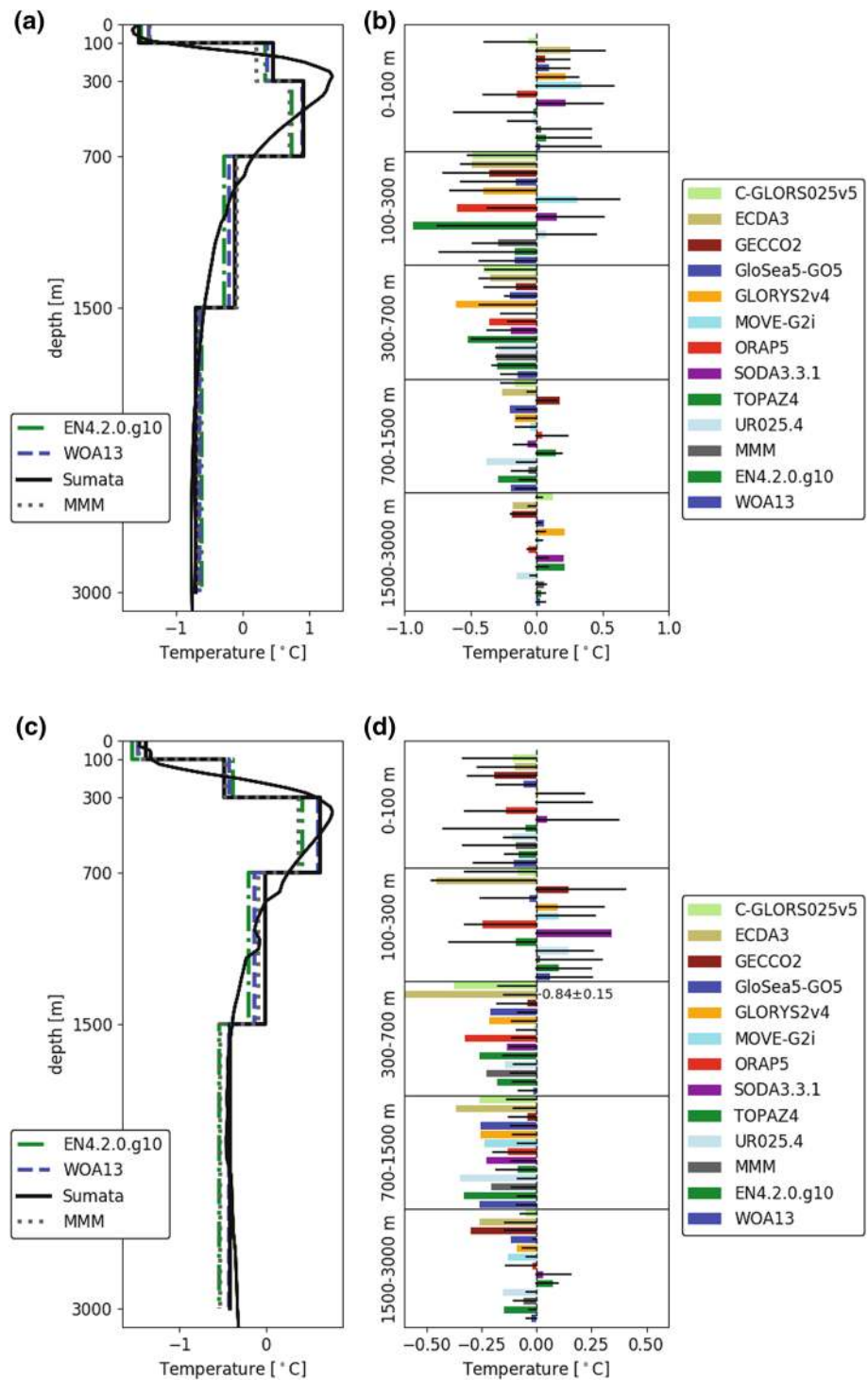
temperatures the salinity controls the potential densities. Salinities from Sumata, WOA13 and EN4.2.0.g10 agree well in all layers and in both Amerasian and Eurasian basins. Only Sumata surface salinity in the Amerasian basin is slightly lower (∼ 31.3) than in WOA13 (∼ 31.6) and EN4.2.0.g10 (∼ 31.8) perhaps capturing better the recent freshening (Proshutinsky et al. 2009). The MMM 0–100 m salinity

is 31.6, similar to WOA13 and slightly more saline than Sumata. In the Eurasian basin, the MMM 0–100 m salinity is low (∼ 32.8) compared to the observational products, ∼ 33.3, mostly due to ECDA3, with a surface salinity < 30 (Fig. 14b). However in the Eurasian AW layers, 100–700 m, the MMM is also fresher (and colder) than Sumata and WOA13, reflecting the AW deficit. GECCO2 is the fresh,

Fig. 14 Averaged salinity profiles (**a**, **c**) and their departures (errors) from the Sumata climatology (**b**, **d**) in the Eurasian (**a**, **b**) and Amerasian (**c**, **d**) basins shown in Fig. 1. In **a**, **c**, thin black lines show the non-depth averaged Sumata salinity profiles. In **b**, **d** horizontal bars indicate mean errors and black, horizontal lines their standard deviations as error ± deviation. In cases where the mean error or its standard deviation is too large to fit inside the panel, their values are indicated as numbers
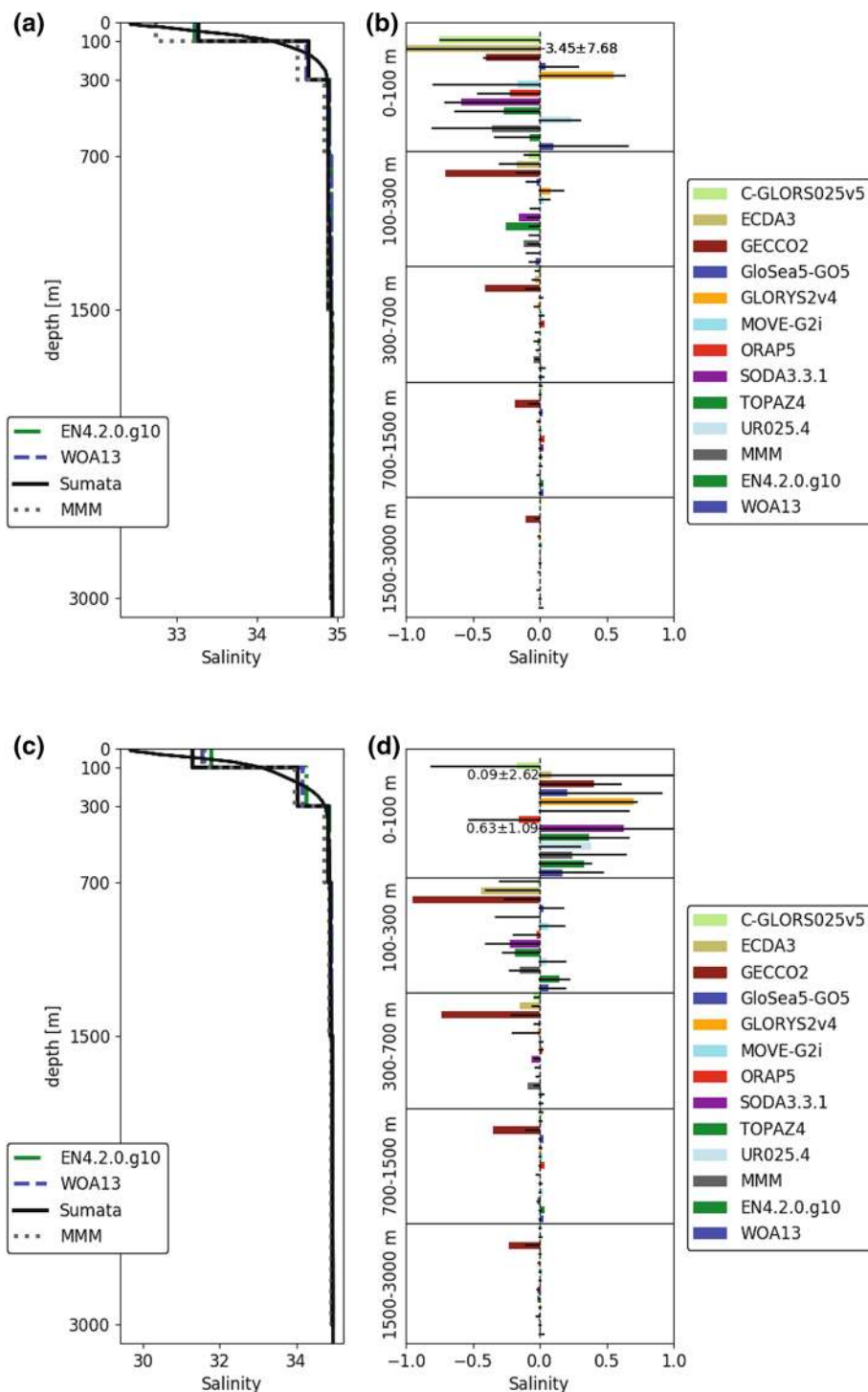
low density outlier (Fig. 14b). The individual ORA spread for 0–100m in the Amerasian basin is small with almost all ORAs between 31 and 32 (except GLORYS2v4, with > 32). Below 100 m in the Amerasian basin the MMM salinity is also fairly consistent with Sumata although GECCO2 and ECDA3 are fresh low density outliers (Fig. 14d). However, the basin mean salinities hide a lot of spatial variability that is reflected in the standard deviations.

To give a better view of spatial distributions of the MMM hydrography biases in the Arctic relative to the observational climatology (the mean of Sumata, WOA13 and EN4.2.0.g10) as a reference, Figure S10 shows the temperature and salinity anomalies (MMM–Obs) in each layer (0–100, 100–300, 300–700, and 700–1500 m). In the top 100 m the MMM in the Amerasian basin is slightly too cold and much too saline, while the Eurasian basin

Fig. 15 Number of ORAs per grid cell (up to 9) where their sea-ice concentration is >15% in February (left) and in September (right) based 1993–2010 monthly data. Number of reanalyses considered here is 9. Black line is the 15% climatological ice edge by NSIDC NASATeam
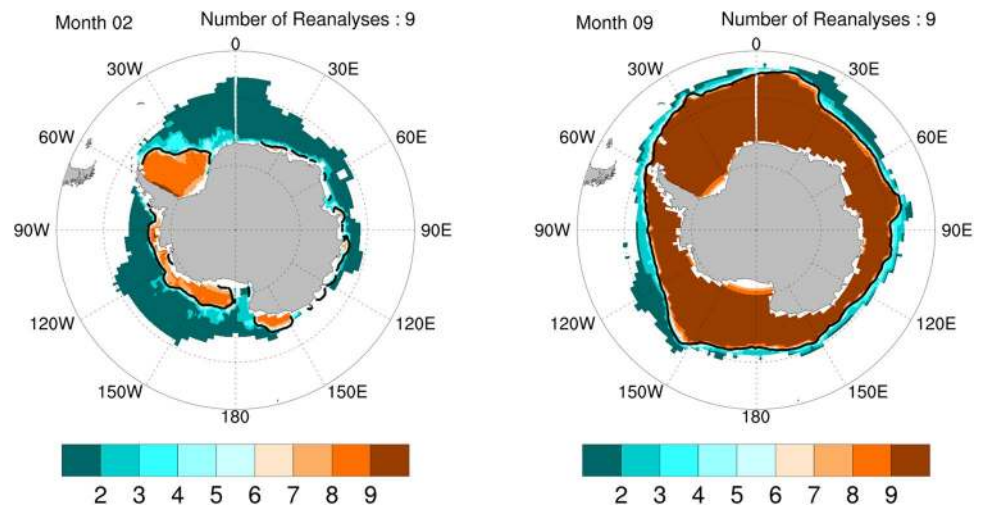
is too warm and a little too fresh. The bias changes just below (100–300 m) with the Amerasian basin now too warm, with both fresher and more saline regions, and the Eurasian basin now too cold and fresh presumably reflecting the lack of AW. Deeper layer salinities (300–1500 m) are slightly too fresh across both basins due mainly to GECCO2, ECDA3 and TOPAZ4 (Fig. 10). However, deeper temperatures are too cold in the 300–700 m layer, which looks like an inadequately resolved AW again, and too warm in the 700–1500 m layer, especially in the Eurasian basin, which is largely due to GECCO2 and TOPAZ4 as was seen in Fig. 13.

4.2 Antarctic mean states

4.2.1 Sea ice

As for the Arctic Ocean, and for the same reasons (ocean and/or sea-ice data assimilation, and atmospheric forcing) there is an overall agreement in the position of the winter sea-ice edge in the Southern Ocean (Fig. 15). The SODA3.3.1 is a major outlier extending ice too far north in the Pacific and the Indian Ocean sectors, and in all sectors during summer (Figure S11). This can be traced to extremely high snow precipitation in the MERRA2 atmospheric forcing product. In particular, thicker snow layer takes longer to melt than thin snow layer and promotes the formation of snow-ice, which increases the total ice thickness. Moreover, sea ice with thick snow on top survives longer because it retains longer a higher albedo compared to sea ice with thin snow on top. Almost all other reanalyses simulate a realistic minimum sea-ice cover and, as in the Arctic, SODA3.3.1

will be removed from multi-model estimates in further sea-ice concentration analysis.

Figure 16 is the counterpart of Fig. 3 for the Antarctic seasonal sea-ice cycle. All systems have a maximum in September, and a minimum in February. ECDA3 is the only reanalysis that loses all sea ice in summer, and stands out from November to May, but has a realistic maximum (Figure S12). The ORA ensemble spread in sea-ice extent is generally larger in winter than in summer. A possible reason is that data assimilation and atmospheric forcings provide a strong constraint on summer sea-ice extent, while there are more degrees of freedom during winter. Sea-ice area is generally overestimated during winter, as shown by the MMM, although driven by two outliers on the high side (SODA3.3.1 and GECCO2), whereas the remaining ORAs lie essentially within the range of observations.

In the observations, we see that the MIZ/pack-ice ratio in the Antarctic is much more balanced than in the Arctic. The MIZ area peaks in October–November (after the annual sea-ice area maximum) while pack-ice area peaks earlier, in June–September. Most systems that assimilate sea ice have a seasonal cycle of MIZ area consistent with that observed. However, pack-ice area in June–October is higher than observed in most systems, with the exception of ECDA3, which does not assimilate sea-ice data, and C-GLORS025v5, which assimilates sea-ice data. Therefore, sea-ice data assimilation does not seem to be sufficient to reproduce a correct MIZ/pack-ice ratio.

Figure S13 shows the ORA MMM sea-ice thickness during periods for which satellite measurements (Kurtz and Markus 2012) were taken and can be directly compared with their Fig. 2. General spatial features of the Antarctic ORA MMM sea-ice thickness distribution agree rather well with Kurtz and Markus (2012), although having a thin bias. Figure 17 shows the mean ice thickness errors assessed against the independent ASPeCt data. The results are discussed in

Fig. 16 Mean seasonal cycle (over the period of 1993–2010) of the Antarctic sea-ice extent, area (upper row) and of the area covered by Marginal Ice Zone (MIZ) and pack ice (lower row), in nine global ORAs, the multi-model mean (colour lines) and in NSIDC, CERSAT and OSISAF observations (grey shading). MIZ is defined as a region where the sea-ice concentration is less than 90% and greater than 15%, while the pack ice is the region where the sea-ice concentration is higher than 90%. Units are in 10^6 km^2
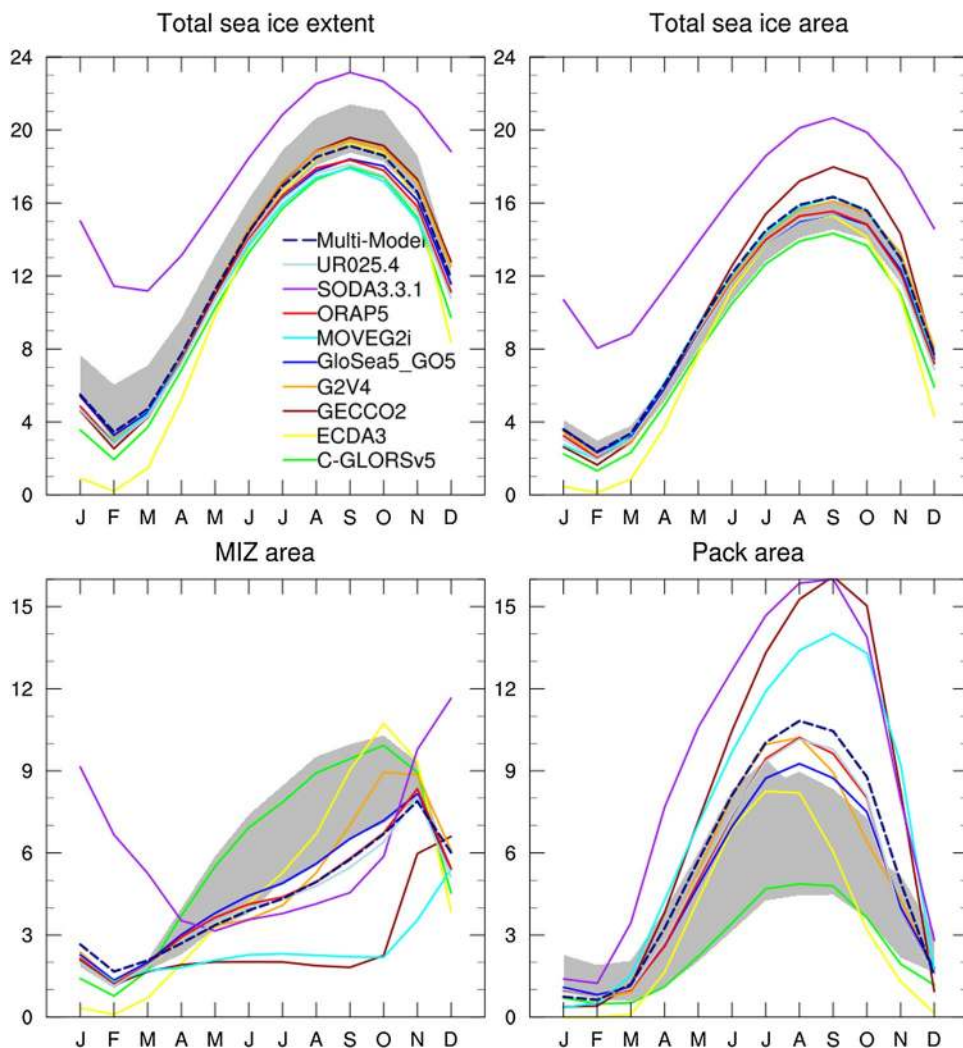
two steps: first considering each ORA reanalysis individually, and then considering ORAs as a multi-model ensemble and using the compatibility index as described in Sect. 3.2.

At the individual level, Fig. 17 shows that the ORA products perform fairly poorly. Indeed, if the ASPeCt SIT and the ORA SIT would actually be drawn from the same distribution, one would expect an incompatibility—in the sense that we defined it in Sect. 3.2—to occur by chance with probability $1/2^{N-1}$ where N is the number of ASPeCt samples, that is 12.5% or less for N greater or equal than 4, as we designed it. All ORAs are individually incompatible with ASPeCt at least more than 39, or 61% compatible, of the time, hence sampling alone cannot explain this discrepancy. This leaves two possibilities: either ORAs are truly wrong, or the observational reference has a systematic error. Given the prior information that ASPeCt is biased thin (Worby et al. 2008), we can state with confidence that the already thin ORAs (ECDA3, MOVE-G2i, UR025.4) are probably inconsistent with reality. GECCO2 can also be excluded from any realistic ensemble, given its large average absolute bias (22.0

cm) compared to the background ASPeCt SIT standard deviation (36.2 cm), and it is also an outlier in other assessments reported in this study. For the other ORAs, nothing can be said given the unclear role of the ASPeCt errors in the assessment.

For the ensemble mean, however, the agreement seems better. The ORA MMM is biased thick, but again it cannot be excluded that this is linked to the thin bias of ASPeCt. As in other parts of this study, the ensemble mean performs very well compared to the individual products, perhaps thanks to the compensation of random errors present in each product. We also considered the ensemble of nine ORA SITs and compared it to the ASPeCt data. This approach, unlike the ORA MMM, does not average SIT, it checks that the observations lie within the ensemble. We found that the ORA ensemble is only inconsistent with ASPeCt (meaning, the two ensembles do not overlap one another) 7.8% of the time, less than the significance level of 12.5%.
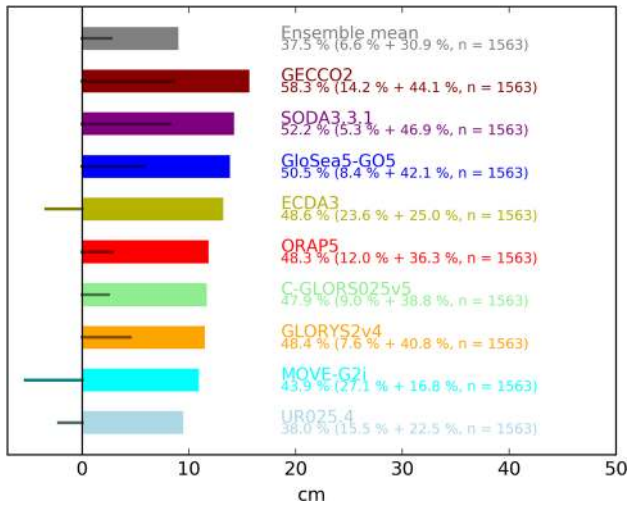
Fig. 17 Mean error (thin coloured horizontal lines) and mean absolute error (coloured rectangles) of nine ORA-IP reanalyses as well as their ensemble mean of the Antarctic sea-ice thickness. For each reanalysis, an compatibility index (in %, see Sect. 3.2 for details) is also provided: this index records the percentage of cases where the reanalysis was found to be consistent with the reference ASPeCt data set (see text for details). This index is further broken down in cases where the incompatibility comes from a thin bias and a thick bias with respect to ASPeCt (first and second number in parentheses). The total number of cases (n) on which the assessment is done is also given

4.2.2 Mixed layer depth

In the Southern Ocean, summer MLDs are underestimated south of the ACC in all ORAs except UR025.4 and MOVE-G2i (Figs. 18, S14). The band of mixed layers reaching between 60 and 80 m within the ACC is well represented in C-GLORS025v5 and UR025.4, but the ORA MMM is biased low compared to observations, as in the Arctic. In winter, the ocean destabilizes to depths down to several hundred meters in a narrow band on the equatorward edge of the ACC (e.g. Sallée et al. 2013). While the observed pattern is well captured by the MMM, most ORAs strongly overestimate these deep MLDs (Figure S15). Observations show that mixed layers also reach depths close to 200 m on the continental shelves of the Ross and Weddell Seas and along the coast of East Antarctica. Reanalyses tend to underestimate MLDs around East Antarctica, whereas an ORA ensemble mean positive bias of more than 200 m is seen in the Ross and Weddell Seas. Pellichero et al. (2017) note however that the Ross Sea sector is the least well observed region so the climatology must be used with caution in this area. Of individual ORAs, MOVE-G2i and, to a lesser extent, GLORYS2v4 and GECCO2, show signs of open ocean deep convection in the Weddell Sea (Figure S15).

Fig. 18 Antarctic mixed layer depth for January (top) and August (bottom), in an observation-based climatology (OBS; Pellichero et al. 2017), for the ensemble mean of the ORAs (MMM) and bias of the ensemble mean with respect to the climatology (MMM–OBS)
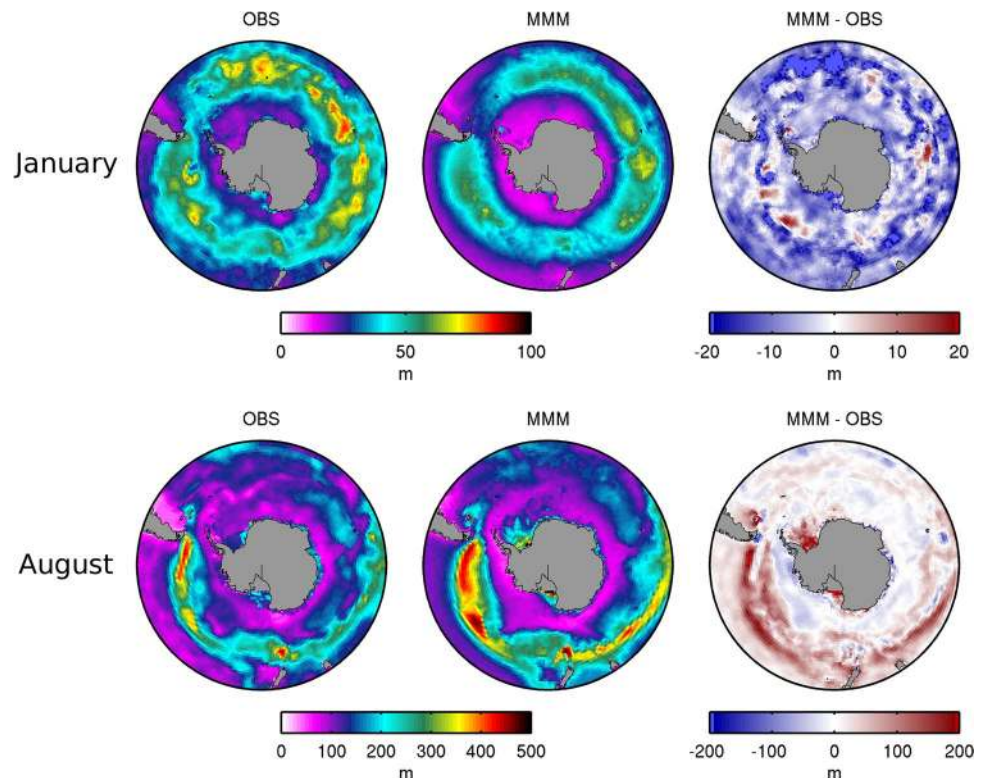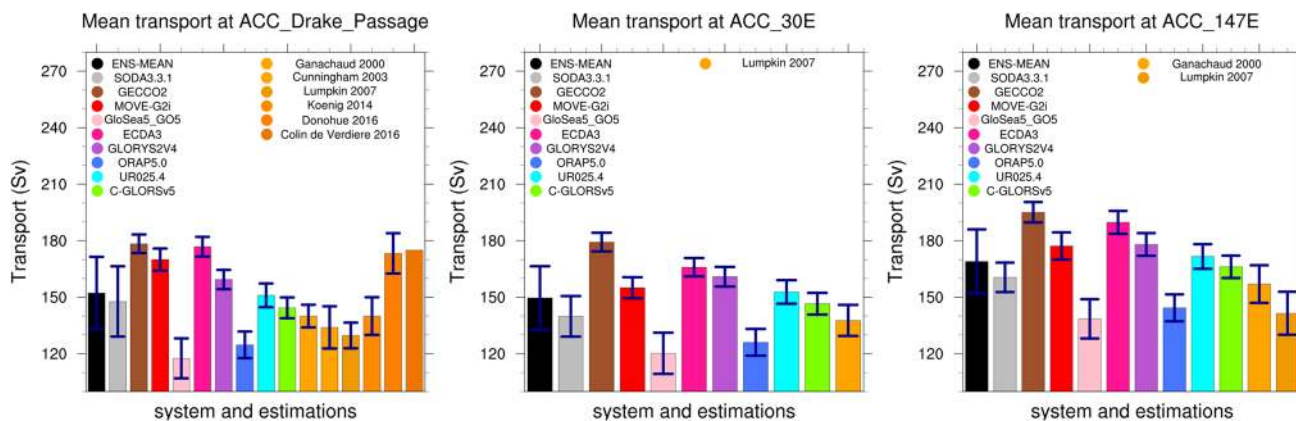
Fig. 19 Antarctic Circumpolar Current (ACC) liquid ocean volume transport through the Drake Passage (left), the "30°E" section (middle) and the "147°E" section (right). Units are in Sv (10^6 m^3 s^{-1}). Standard deviation, measuring the interannual variability, is given for each of the nine ORA products together with the standard deviation among these nine members for the ensemble mean (ENS). Estimations from different observed measurements are also given, see text for details

Fig. 20 Averaged multi-model Antarctic Ocean Heat Content (OHC) in the layer of 0–1500 m (top left panel), its spread (top second left panel) and individual product departures from the multi-model mean
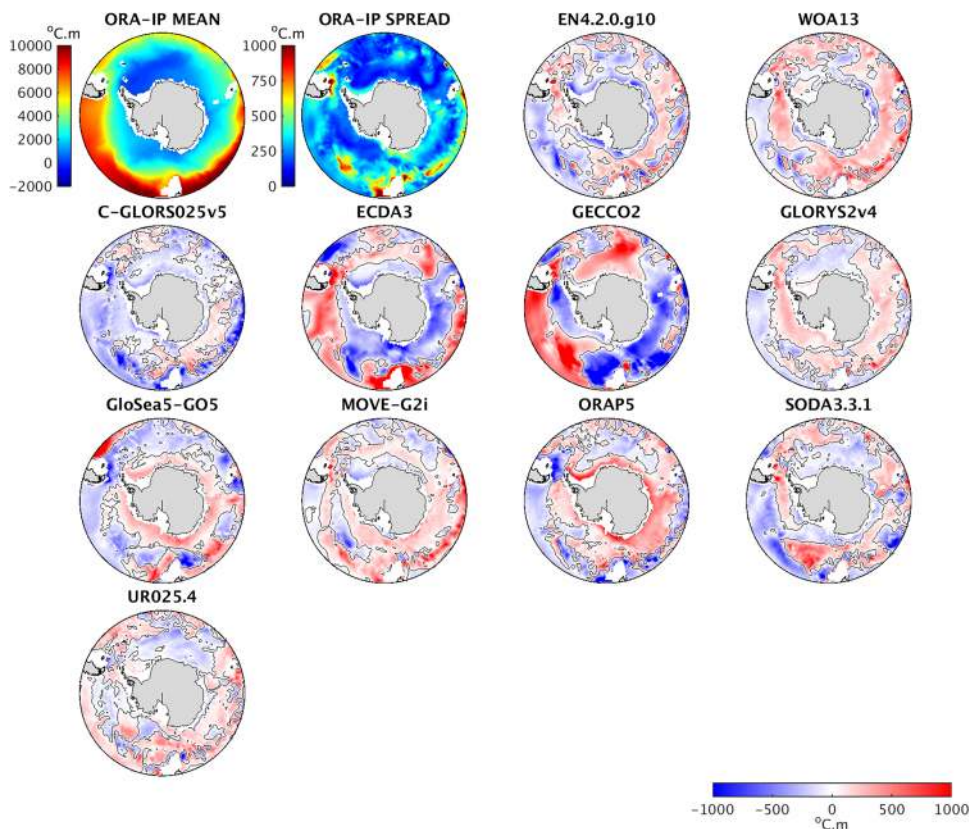
4.2.3 Liquid ocean transports

Among the ORAs, ORAP5.0 and GloSea5-GO5 simulate the weakest transport along the ACC, with SODA3.3.1 being weak on the "30°E" section (see Fig. 19). GECCO2 and ECDA3 have the strongest transports. The ensemble mean of ORAs for the Drake Passage is 152 Sv (±19.2 Sv),

149 Sv (±16.8 Sv) for "30°E" and 169 Sv (±17 Sv) for 147°E. Upper bounds of the ORA MMM transport estimate in the Drake Passage match the mean of the most realistic observed estimates (Donohue et al. 2016; de Verdiére and Ollitrault 2016). However, the ORA MMM mean value remains less than the mean minus one standard deviation bound estimate by Donohue et al. (2016). Only GECCO2,

Fig. 21 Averaged multi-model Antarctic Ocean Salt Content (OSC) in the layer of 0–1500 m (top left panel), its spread (top second left panel) and individual product departures from the multi-model mean
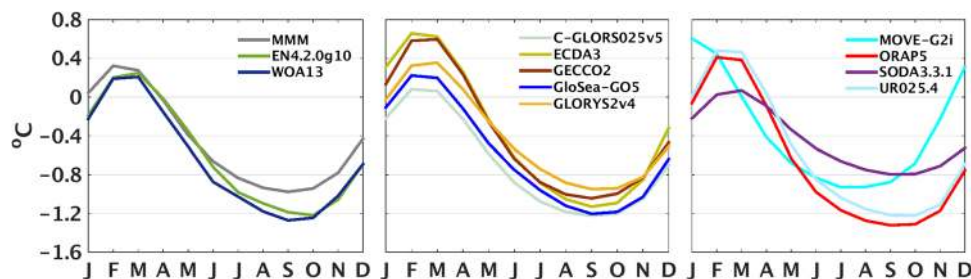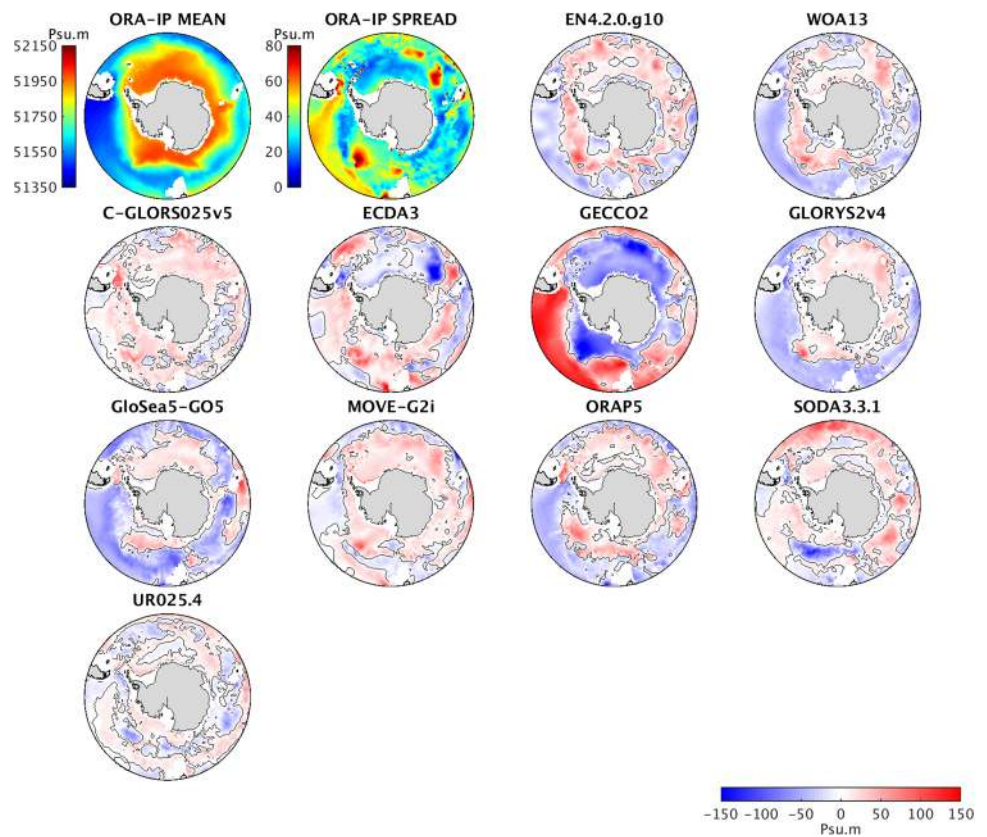

Fig. 22 Seasonal cycle of averaged monthly temperature from 1993 to 2010 in the layer of 0–100 m in the Southern Ocean. In the left-most panels seasonal cycles for the multi-product mean (MMM) and two observational data sets (EN.4.2.0.g10, WOA13) are shown, while in the middle and rightmost panels individual ORA seasonal cycles are presented

Move-G2i and ECDA3 have their mean values within this range. The amplitude of the seasonal cycle in all ORAs is weak compared to the absolute value (see Figure S16, right panel). All the ORAs exhibit insignificant trends in the three transects (see Figure S16, left panel for the Drake Passage).

We also analyzed surface ocean currents. The Ocean Surface Current Analyses Real-time (OSCAR) provides surface currents averaged over the top 30 m on a 1/3° grid every 5 days (Dohan and Maximenko 2010; Bonjean and Lagerloef 2002). The pattern comparison of different ORAs and the ORA MMM with OSCAR

dataset in the Southern Ocean indicate a good match at monthly frequency (not shown) which is reflected in well contained mean errors of zonal and meridional components, south of 50°S, on the order of 1 cm s^{-1} or less (Figure S17).

4.2.4 Hydrography

The OHC in the top 1500 m for all products is first studied (Fig. 20). The MMM is warmer along the Antarctic continent but generally colder to the north than observational products EN4.2.0.g10 and WOA13. WOA13 appears
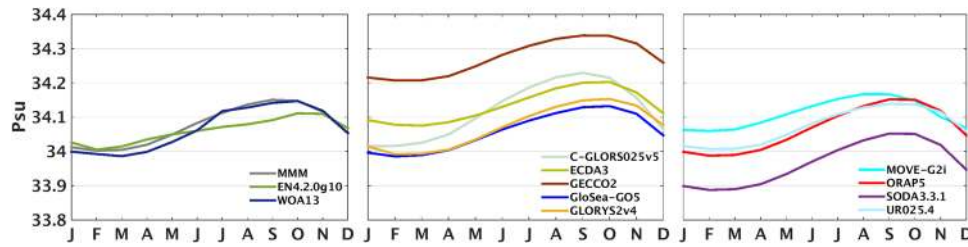
Fig. 23 Seasonal cycle of averaged monthly salinity from 1993 to 2010 in the layer of 0–100 m in the Southern Ocean. In the leftmost panels seasonal cycles for the multi-product mean (MMM) and two observational data sets (EN.4.2.0.g10, WOA13) are shown, while in the middle and rightmost panels individual ORA seasonal cycles are presented

Fig. 24 Averaged temperature profiles (**a**) and their departures (errors) from the WOA13 climatology (**b**) in the Antarctic region shown in Fig. 1. In **a**, thin, dashed blue line shows the non-depth averaged WOA13 temperature profiles. In **b** horizontal bars indicate mean errors and black, horizontal lines their standard deviations as error ± deviation. In cases where the mean error or its standard deviation is too large to fit inside the panel, their values are indicated as numbers
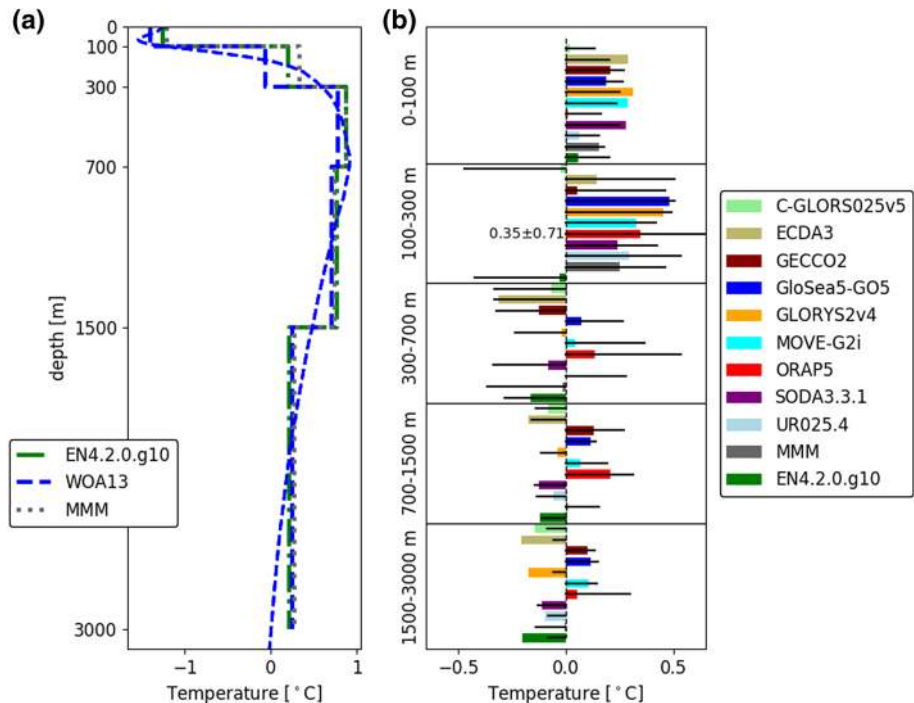
warmer than EN4.2.0.g10 with larger differences to the MMM. ECDA3 and GECCO2 have largest anomalies from the MMM. The individual ORA OHC anomalies are generally larger than those of the two observational products, especially for example ECDA3 and GECCO2.

In terms of OSC in the top 1500 m, shown in Fig. 21, EN4.2.0.g10 and WOA13 closely agree. The MMM OSC indicates somewhat fresher water around Antarctica (positive values in the EN4.2.0.g10 and WOA13 difference plots) than the two observational products on average, despite negative values immediately adjacent to Antarctica. Again Fig. 21 shows that observational products (EN4.2.0.g10 and WOA13) have smaller anomalies with respect to the MMM than most ORAs. GECCO2 is still by far the biggest outlier, while GLORYS2v4, GloSea5-GO5 and ORAP5 all broadly show opposite anomalies to GECCO2. These anomaly patterns also largely agree

with anomalies in the observational products EN4.2.g10 and WOA13. Because the MMM is relatively close to the observations despite the large departures of individual ORAs, there is clear evidence that the MMM is averaging out individual ORA biases.

Seasonal cycles of mean temperature in the 0–100 m layer for the Antarctic region are presented in Fig. 22. Observational products and the ORAs generally agree well, being warmest in summer (February–March) and coldest in winter (September) with comparable amplitudes. MOVE-G2i has a somewhat anomalous seasonal cycle of mean temperature with maximum in January and minimum in July. SODA3.3.1 has the lowest seasonal amplitude of upper ocean temperature, associated with its excessive sea-ice cover.

Seasonal cycles of mean salinity in the 0–100 m layer for the Antarctic region are presented in Fig. 23. Observational products and the MMM generally agree well

Fig. 25 Averaged salinity profiles (**a**) and their departures (errors) from the WOA13 climatology (**b**) in the Antarctic region shown in Fig. 1. In **a**, thin dashed blue line shows the non-depth averaged WOA13 salinity profiles. In **b** horizontal bars indicate mean errors and black, horizontal lines their standard deviations as error ± deviation. In cases where the mean error or its standard deviation is too large to fit inside the panel, their values are indicated as numbers
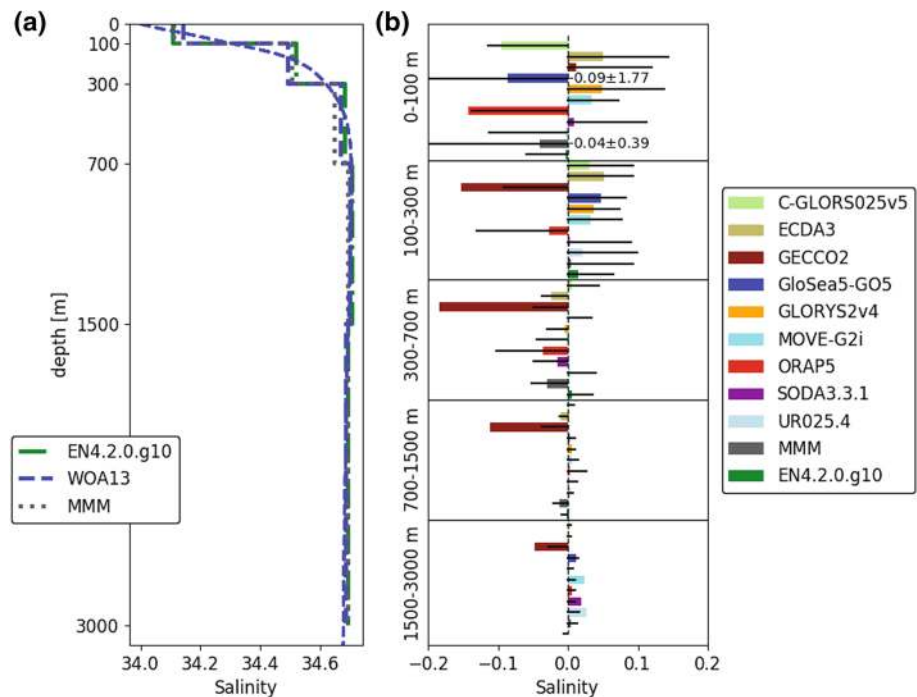
with increasing monthly mean salinities during the freeze up period (autumn–winter) which then decrease towards the summer during the ice and snow melt season. The MMM agrees somewhat better with WOA13 than with EN4.2.0.g10. Although the annual mean salinities of individual ORAs vary to some extent, their seasonal cycles of upper ocean salinity agree and have similar shapes to the MMM. C-GLORS025v5 has a relatively higher seasonal amplitude with a steep decline of salinity in spring. MOVE-G2i, on the other hand, has a relatively small seasonal amplitude.

The vertical profiles of temperature from the reanalysis products and observational datasets that are averaged over the Antarctic region (Fig. 1) are evaluated against the WOA13 temperature profiles in Fig. 24. As noted in Sect. 3.7, the water column is divided into five layers: 0–100, 100–300, 300–700, 700–1500 and 1500–3000 m, and mean temperatures for these layers were computed and compared between the separate datasets and WOA13.

In Fig. 24a the EN4.2.0.g10 and the MMM temperatures are in fairly good agreement with the WOA13 temperature below 700 m. Between 300 and 700 m, where the water column is the warmest due to the presence of the Upper Circumpolar Deep Water (UCDW), the EN4.2.0.g10 temperature is almost identical to the MMM, while the MMM temperature is higher than WOA13 by ∼ 0.1 °C. The largest temperature differences between the MMM and observational datasets occur between 100 and 300 m, where the MMM has a warm bias of 0.4 and 0.2 °C relative to WOA13 and EN4.2.0.g10, respectively. This possibly indicates a

larger fraction of winter water in the observations. In the 0–100 m layer, where the water is the coldest, the MMM is also warmer than observed values, but the deviations are not as large as in the layer below. In the 0–100 m layer the WOA13 and EN4.2.0.g10 temperatures are colder than the MMM temperature by 0.25 and 0.1 °C, respectively.

Figure 24b shows the vertical profiles of basin-wide mean and standard deviation of temperature differences between individual ORAs, MMM and EN4.2.0.g10 from the WOA13 temperature values. The ORA mean difference range and their standard deviations are relatively small in the surface layer with all the individual reanalyses differing by < 0.2 °C from the MMM. The sub-surface layer (100–300 m), where the MMM has a clear warm bias compared to observations, is characterised by a larger scatter with differences reaching 0.4 °C. Although one ORA is close to WOA13 (C-GLORS025v5), all others are warmer than this observational product. GLORYS2v4, GloSea5-GO5 and UR025.4 are much warmer than observations and the MMM. Below 300 m, observations and the MMM agree well due to compensation between the different reanalyses, with the majority being within 0.2 °C of the MMM. In agreement with Fig. 20, GloSea5-GO5 and ORAP5 display the largest systematic positive anomalies compared to the WOA13 but others also have large positive or negative differences in many layers, such as GECCO2, ECDA3 and SODA3.3.1. The largest biases, reaching nearly 0.4 °C are found in GloSea5-GO5 which has a warm bias over the whole water column.

The vertical salinity profiles for the reanalysis products are shown in Fig. 25. Overall, the MMM salinity is close to WOA13 and EN4.2.0.g10 in all layers, in particular in layers deeper than 700 m. WOA13 and EN4.2.0.g10 agree rather well, although in the 100–300 m layer WOA13 is slightly fresher with relatively large standard deviation of basin-wide differences. In the 300–700 m layer, the MMM salinity is smaller than WOA13 and EN4.2.0.g10 by ∼ 0.03. The majority of the reanalyses are also close to observations, except in the 0–100 m layer where mean differences can reach more than 0.01 and standard deviations are large. Below 100 m, GECCO2 has a considerable fresh bias which reaches a maximum in the 300–700 m layer, but remains large in the deeper layers. This makes GECCO2 an exception in the OSC patterns of ORAs for the upper 1500 m, as shown in Fig. 21. In the 100–300 m layer, noticeable deviations in salinity from the WOA13 occur also in ECDA3, GloSea5-GO5, GLORYS2v4 and ORAP5.

The surface layer (0–100 m) is characterized by low temperature and salinity and is composed of the Antarctic Surface Water. Among the reanalyses, GECCO2, SODA3.3.1 and UR025.4 have water mass properties that are closest to the MMM values, while MOVE-G2i and ORAP5 have the largest deviations from the MMM values, with their potential density higher/lower than the MMM by ∼ 0.06 and 0.1 kg m^{-3}, respectively.

In the subsurface layer (100–300 m), the MMM water mass properties are much closer to EN4.2.0.g10 than to WOA13, due to the lower temperature in WOA13 leading to a higher density. The largest deviations in water mass properties occur for GloSea5-GO5 and GLORYS2v4, which have considerably higher temperature than the MMM and observations, and for GECCO2 which has a much lower potential density than the MMM, by 0.12 kg m^{-3}, associated with the exceptionally low salinity in this layer.

In the 300–700 m layer, where the presence of the Upper Circumpolar Deep Water (UCDW) leads to the highest temperatures and salinities, the agreement is good between EN4.2.0.g10, WOA13 and the MMM, with a slightly higher potential density in WOA13. In this layer, GECCO2 is the only dataset that has considerable potential density deviation from the MMM, due to its low salinity. In the 700–1500 m layer, mainly occupied by the Lower Circumpolar Deep Water (LCDW), and in the deep layers below 1500 m, the densities are even closer between EN4.2.g10, WOA13 and the MMM, as noted for temperature and salinity profiles. The range among ORAs is also smaller, with GECCO2 still standing out due to its large potential density deviation from the MMM.

5 Discussion

After presenting individual diagnostics in the previous sections, we concentrate here on processes connecting the results. Table 3 shows ORA departures, or errors, from observed data as their basin-wide means and standard deviations for all diagnostics at a glance. This approach lets readers to understand if the errors are really a problem or within what is deemed acceptable.

5.1 Sea ice and snow

The ORAs agree well on the location of the sea-ice edge in the Arctic and the Southern Ocean. In winter, individual ORAs show both positive and negative sea-ice concentration biases in the Labrador Sea, the Greenland Sea, and the sea of Okhotsk. On average, the sea-ice area and concentration tend to be overestimated both in the Arctic Ocean and the Southern Ocean (Table 3). The Antarctic sea ice is a mix of ice with medium and high sea-ice concentration, at a more balanced ratio than in the Arctic. Thus, the general positive sea-ice area bias appears more clearly in the Antarctic with only a narrow MIZ present near the ice edge. The sea-ice edge itself is well constrained, even in summer at the Antarctic sea-ice minimum, when it is controlled by small-scale coastal processes and not well represented in rather coarse resolution free running ocean-ice models without data assimilation (Downes et al. 2015).

Although a CORE-II comparison is meaningful for the ocean, it is less so for sea ice, as the ORA sea-ice concentration is rather strongly constrained by data assimilation. Therefore, we do not compare the ORA sea ice with CORE-II in the following discussion.

Chevallier et al. (2017) noted that the sea-ice concentration is one of the most consistent features of the sea-ice cover amongst the reanalyses due to constraints imposed by direct assimilation of ocean and sea-ice concentration observations, and to the strong restoring towards near-surface air temperatures through the atmospheric reanalyses. All reanalyses in the present study, except the coupled ECDA3, are driven by a prescribed atmosphere through bulk formulae. In addition to missing sea-ice data assimilation, the vanishing Antarctic summer sea ice in ECDA3 is possibly related to the stronger surface shortwave radiation compared to prescribed atmospheric forcing. Other ORAs also have more correct marginal versus pack-ice area/extent ratios.

The fact that ORAs realistically reproduce the seasonal cycle of total sea-ice area (Figs. 3, 16) and agree in the location of the sea-ice edge (Figs. 2, 15), is owing to the compensation of errors in their simulation of two sea-ice regimes, too small MIZ area and respectively too large pack-ice area. This is consistent with the results of Chevallier

Table 3 A summary of diagnostic performance of ten ORAs and their mean (MMM) in the Arctic and Antarctic

Arctic	SIA-M	SIA-S	SND	SIT	MLD-S	MLD-W	OHT	ST	SS	OHC	OSC	AW
C-GLORS025v5	−0.24 ± 0.06	−0.60 ± 0.09	−0.15 ± 0.17	−0.16 ± 0.67	−8.9 ± 5.6	39.0 ± 150.5	−34 ± 5	−0.17 ± 0.70	−0.09 ± 0.86	−133 ± 242	−9.3 ± 75.9	−0.40 ± 0.38
ECDA3	0.20 ± 0.14	−0.01 ± 0.34	−0.09 ± 0.13	−0.79 ± 1.01	−8.3 ± 5.1	66.0 ± 133.6	−51 ± 5	−0.35 ± 0.77	−0.36 ± 5.57	−331 ± 289	−125.9 ± 198.6	−0.35 ± 0.43
GECCO2	−0.02 ± 0.16	−0.45 ± 0.58		−0.37 ± 0.71	−8.0 ± 5.1	10.9 ± 65.8	−7 ± 5	−0.08 ± 0.79	−0.36 ± 0.76	250 ± 198	−72.6 ± 71.6	−0.16 ± 0.39
GLORYS2v4	0.13 ± 0.06	0.24 ± 0.07	−0.13 ± 0.16	0.70 ± 1.08	−9.3 ± 5.1	55.3 ± 178.8	−48 ± 5	−0.10 ± 0.75	0.11 ± 1.15	−112 ± 218	−12.7 ± 153.7	−0.60 ± 0.43
GloSea5– GO5	0.14 ± 0.07	0.18 ± 0.05	−0.14 ± 0.17	−0.38 ± 0.84	−9.1 ± 5.1	25.4 ± 86.5	−7 ± 5	−0.10 ± 0.54	0.02 ± 1.27	−8 ± 171	14.8 ± 73.2	−0.20 ± 0.23
MOVE– G2i	0.08 ± 0.10	−0.09 ± 0.12	−0.13 ± 0.16	−0.52 ± 0.80	−5.6 ± 5.2	29.0 ± 112.8	−83 ± 5	0.03 ± 0.72	0.01 ± 0.75	110 ± 208	45.0 ± 41.1	−0.01 ± 0.27
ORAP5	0.25 ± 0.08	−0.02 ± 0.05	−0.08 ± 0.12	0.07 ± 0.66	−8.6 ± 5.7	57.2 ± 228.0	−16 ± 5	−0.20 ± 0.61	−0.02 ± 0.52	−112 ± 129	41.8 ± 40.3	−0.35 ± 0.22
SODA3.3.1	0.97 ± 0.37	1.60 ± 0.72	0.05 ± 0.13	−0.34 ± 0.69	−10.6 ± 5.1	8.0 ± 60.0	−31 ± 5	0.04 ± 0.71	−0.03 ± 1.15	97 ± 162	−71.7 ± 120.2	−0.20 ± 0.37
TOPAZ4	−0.19 ± 0.11	0.13 ± 0.09	0.01 ± 0.12	−0.43 ± 0.69	−5.5 ± 5.2	49.4 ± 125.8		−0.14 ± 0.99	−0.03 ± 0.75	5 ± 232	−53.5 ± 85.2	−0.52 ± 0.49
UR025.4	0.04 ± 0.09	−0.07 ± 0.05	−0.20 ± 0.24	−0.85 ± 1.06	−6.5 ± 5.0	56.4 ± 127.7	−23 ± 5	−0.14 ± 0.49	0.06 ± 0.92	−112 ± 129	−4.5 ± 42.4	−0.28 ± 0.30
MMM	0.11 ± 0.08	0.09 ± 0.10	−0.10 ± 0.16	−0.31 ± 0.84	−7.8 ± 5.0	34.0 ± 108.3	−38 ± 5	−0.11 ± 0.64	−0.06 ± 0.88	−25 ± 124	−27.8 ± 65.4	−0.30 ± 0.31
WOA13								−0.07 ± 0.59	0.04 ± 0.59	34 ± 113	16.3 ± 35.0	−0.14 ± 0.27
EN4.2.g10								−0.12 ± 0.68	0.05 ± 0.83	−118 ± 160	9.6 ± 127.0	−0.30 ± 0.33

Antarctic	SIA– S	SIA– F	SIT	MLD– S	MLD– W	OVT	ST	SS	OHC	OSC
C– GLORS025v5	−0.90 ± 0.16	−1.08 ± 0.09	0.03 ± 0.24	−7.9 ± 9.4	52.3 ± 73.1	−28.9 ± 10.9	−0.06 ± 0.62	−0.01 ± 0.15	−131 ± 235	7.1 ± 24.7
ECDA3	0.05 ± 0.58	−2.28 ± 0.33	−0.03 ± 0.31	−18.5 ± 9.1	5.4 ± 53.4	3.5 ± 10.7	−0.05 ± 0.65	0.01 ± 0.18	−277 ± 317	−0.7 ± 30.7
GECCO2	1.97 ± 0.39	0.22 ± 0.39	0.09 ± 0.31	−14.3 ± 7.4	−9.2 ± 64.0	5.1 ± 11.0	0.07 ± 0.69	−0.10 ± 0.17	−190 ± 503	−53.5 ± 53.8
GLORYS2v4	0.86 ± 0.29	−0.35 ± 0.08	0.05 ± 0.23	−10.0 ± 7.1	3.1 ± 54.1	−13.9 ± 11.0	0.10 ± 0.62	0.02 ± 0.16	11 ± 220	−9.6 ± 73.5
GloSea5–GO5	0.14 ± 0.22	−0.08 ± 0.09	0.06 ± 0.29	−4.6 ± 10.0	24.0 ± 67.4	−55.7 ± 13.5	0.19 ± 0.66	−0.00 ± 1.78	34 ± 291	−7.8 ± 38.2
MOVE-G2i	1.01 ± 0.76	−0.46 ± 0.19	−0.05 ± 0.23	7.7 ± 12.0	49.4 ± 91.5	−3.3 ± 11.1	0.17 ± 0.65	0.02 ± 0.12	39 ± 261	7.4 ± 21.4
ORAP5	0.31 ± 0.37	−0.40 ± 0.11	0.03 ± 0.25	2.6 ± 12.4	41.1 ± 67.9	−48.6 ± 11.3	0.15 ± 1.00	−0.04 ± 0.22	138 ± 388	1.8 ± 39.5
SODA3.3.1	5.43 ± 0.70	5.64 ± 1.17	0.08 ± 0.27	−19.8 ± 9.0	−30.0 ± 52.2	−25.5 ± 14.0	0.04 ± 0.63	0.00 ± 0.15	−63 ± 221	−5.8 ± 30.4
UR025.4	0.13 ± 0.30	−0.47 ± 0.08	−0.02 ± 0.21	−1.0 ± 8.3	74.2 ± 67.6	−22.3 ± 11.2	0.04 ± 0.64	0.01 ± 0.16	−57 ± 255	−7.6 ± 28.4
MMM	−0.61 ± 0.19	−0.17 ± 0.15	0.03 ± 0.19	−7.4 ± 6.7	23.8 ± 52.2	−21.1 ± 11.0	0.08 ± 0.65	−0.01 ± 0.40	−55 ± 208	−7.7 ± 21.8
EN4.2.g10							−0.09 ± 0.56	0.00 ± 0.10	−96 ± 179	4.9 ± 20.7

For each diagnostics mean errors and their standard deviations, calculated from differences between the ORAs and observed reference data, are presented, where applicable. The observed reference data are described in text. Abbreviations of diagnostics: sea-ice area in February (SIA-F, 10^6 km^2), sea-ice area in March (SIA-M, 10^6 km^2), sea-ice area in September (SIA-S, 10^6 km^2), snow depth (SND, cm), sea-ice thickness (SIT, m), mixed-layer depth in winter (MLD-W, m) and in summer (MLD-S, m), liquid ocean heat transport (OHT, TW), liquid ocean volume transport (OVT, Sv), sea temperature (ST, °C), sea salinity (SS), ocean heat content (OHC, °C m), ocean salt content (OSC, psu m), and the temperature of the warm Atlantic Water (AW at 300–700 m layer, °C) in the Eurasian basin

et al. (2017), and shows that in spite of recent physics and data assimilation improvements, ORAs still tend to simulate too high pack-ice concentration in winter.

Snow volume differences among ORAs grow during autumn and winter. They are linked to precipitation biases, and deviations in ice formation, the timing of freeze-up, melt, flooding and sublimation. Lindsay et al. (2014) showed that the rate of precipitation in the polar regions is highly uncertain between reanalyses. Most of the ORAs analyzed here are using the ERA-Interim reanalysis with the exceptions of ECDA3 (which has a coupled ocean-atmosphere, with the atmosphere relaxed towards the NCEP-NCAR reanalysis), MOVE-G2i (forced by the JRA-55 reanalysis) and SODA3.3.1 (forced by the NASA MERRA2 reanalysis), see Table 1. Differences in atmospheric forcing result in discrepancies in the air–ice surface energy balance, ice growth and melt, and upper ocean characteristics. In addition, sea-ice dynamics impacts where the ice and snow drifts to, and therefore their spatial distributions. For example, sea-ice dynamics includes the process where ORAs with larger open water fraction have lower snow volume because in these products more snow melts in the open water. To some extent, the ice-snow relationships are affected by the sea-ice concentration and sea surface temperatures, both controlled by data assimilation. For instance, many ORAs have higher than observed sea-ice concentration in the pack-ice region. Accordingly, in these ORAs data assimilation tends to reduce the sea-ice area and correspondingly increase snow and ice melt. Due to these reasons varying physical parametrisations and data assimilation schemes affect the evolution of snow volume.

Despite the differences in atmospheric forcing, all ORAs have a dipole bias of the Arctic sea-ice thickness, with too thick ice in the Beaufort Gyre and too thin in the Eurasian basin north of the Fram Strait. A consequence of the dipole bias is that their total Arctic sea-ice volumes agree rather well with the observed estimates. Similarly, when compared to Arctic ITRP observations, the ORA sea-ice volumes tend to have cancelling positive and negative biases and, as a result, the MMM sea-ice volume is surprisingly close to the ITRP sea-ice volume (Table 3).

Perhaps analogously, the ORA ensemble sea-ice thickness cannot be deemed inconsistent with the ASPeCt data in the Antarctic, while most individual ORAs are themselves inconsistent with this observational data set. This points towards the important role of model error in the misrepresentation of sea-ice thickness. In other words, the spread of the ORAs appears large enough to reflect what is uncertain in the experimental setup and in the models used, but this data set is not sufficient to provide trustful reconstructions and estimations of past Antarctic sea-ice thickness. We argue that the global ensemble diagnostics provides realistic insights to the ocean state, especially when compared to global-individual and regional-ensemble diagnostics results. These insights remain limited due to the large ORA spread.

The MMM also has a rather realistic looking snow and ice volume. This is because most individual ORAs have snow/ice volumes and ice areas close to observations, indicating that their ice and snow related processes are linked rather similarly by physical mechanisms. An exception is SODA3.3.1 which has a very large ice area and thick snow, although its Arctic sea-ice thickness looks reasonable. SODA3.3.1 ice and snow biases are related to delayed ice melt due to the thicker snow. We checked that for SODA3.3, the NASA MERRA2 forcing produces too much ice, ERA-Interim clearly less and JRA-55 produces ice volumes between those two in the Southern Hemisphere. These findings are consistent with Chevallier et al. (2017). The very thin snow cover in UR025.4 may also explain its thin ice in the Arctic and Antarctic.

5.2 Mixed layer

In agreement with our findings in Table 3, an overestimation of average winter Arctic and Antarctic MLDs was also found from the forced ocean simulations conducted in the CORE-II framework (Downes et al. 2015; Ilicak et al. 2016). The sea-ice biases noted above may be linked to this winter MLD bias. The MMM sea-ice concentration has a positive bias in winter, but is close to observed in summer which implies that the simulated sea ice must have a larger than observed seasonal amplitude and indicates increased ice formation and melt. Hence, during the freezing period more salt is rejected to the upper ocean destabilising it more in ORAs than in observations. In winter, the CORE-II MMM has a too extensive deep mixed layer in the Weddell Sea and a very limited region of deep convection in the Ross Sea (Downes et al. 2015). These spatial patterns of deep convection are less realistic that in the ORA MMM. Accordingly, it seems that in winter in the Antarctic the representation of mixed layers in ORAs are improved.

On average, the ORA MMM tends to have a close to the observed amount of ice in the Arctic and too little ice in the Antarctic in summer. The ORAs also uniformly have too shallow summer mixed layers (Table 3). This is the only diagnostics where the ORAs systematically show similar biases with basin-wide means larger than their standard deviations. As explained in the previous paragraph, the amplified seasonal sea-ice cycle may cause large surface freshwater fluxes, reducing the summer MLD. Given the vertical model resolution, the summer MLD may even be at the lowest possible model level for a couple of ORAs, for example at 10 m, although most ORAs have upper layer thicknesses in the order of 1 m. The fact that the largest

MLD discrepancies between ORAs and observations are found in ice-covered regions suggests that the MLD biases could also arise from issues in the ice-ocean coupling or in the vertical mixing of melt water by high-frequency wind events. The shallow MLD is therefore possibly caused by missing or poor representation of some mixing processes such as surface waves, Langmuir circulation and sub-mesoscale eddies. Shallow summer MLD biases are a common issue of current coupled and forced models (Huang et al. 2014; Barthélemy et al. 2015). For example, the shallow summer bias has been found in CMIP5 models (Sallée et al. 2013).

It is worth mentioning that the quality of observations may be poor in the Arctic Ocean and, for example, under thick ice MLD is not even observed. Importantly, the MLD differences between ORAs and observations appear large and robust, and it is likely that the results are not qualitatively affected by the methodological issues related to the different calculation procedures of ORA MLDs compared to observation based MLDs (Toyoda et al. 2017a). In short, the mean mixed layer biases in the ORAs, including the MMM, are similar in both hemispheres: they have a shallow mean MLD bias in summer, which is larger than the standard deviation of the bias (see MLD-S and MLD-W in Table 3. In winter the ORAs, generally a deep bias in winter, but the standard deviation of the winter bias is comparably large.

5.3 Ocean transports and hydrography

Observational products Sumata, WOA13 and EN4.2.g10 show somewhat different hydrographies. Sumata is from 1980 to 2015, but with few observations early on, WOA13 from 1995 to 2015 and EN4.2.g10 from 1993 to 2010. To some extent, deviations between these three observational data sets are probably due to oceanic decadal variability. The 2010–2015 period missing from EN4.2.g10 may be the main cause for its differences from WOA13. Because of the most recent observations, the Arctic Ocean in Sumata appears warmer than in the other two observational products. In the Southern Ocean, where Sumata is not available, EN4.2.g10 is somewhat colder than WOA13, but it is hard to say which one is more realistic. The MMM is closer to EN4.2.g10 probably because many ORAs are adjusted towards it, or its earlier versions, instead of WOA13 (Table 1).

The MMM reveals salient biases among ORAs. Its MLD is too deep in the northern North Atlantic where its warm AW cools and consequently its heat flux to the Arctic Ocean is reduced. This low heat transport, mainly through the Fram Strait, impacts the hydrography in the Arctic Ocean and results in a colder than observed Atlantic layer. Moreover, in the Barents Sea, the winter heat loss is high and many ORAs unrealistically convect to the bottom. Despite this, most ORAs have a realistic location of the sea-ice edge in the Barents–Kara Sea being apparently unaffected by the heat transport anomalies. A similar relationship is seen on the Pacific side of the Arctic, where the ORA heat transport through the Bering Strait is low but the sea-ice edge location is realistic, on average. The prescribed atmospheric forcing, and assimilated sea-ice concentration and sea surface temperature data constrain the location of the sea-ice edge, obscuring the link between it and the oceanic heat transports. However, in many ORAs in the Greenland Sea the sea-ice edge is more to the east than observed (Fig. 2) and the maximum MLD is shifted to the east as a result (Fig. 7). This could, partially at least, block the AW transport through FS.

In CORE-II, the total Atlantic ocean heat transport to the Arctic varies with some models having positive and some negative biases through the Fram Strait. These biases co-vary with the amount of cold water entering the Arctic through the St. Anna Trough (Wang et al. 2016b). In CORE-II models with a cold Arctic, the warm AW layer is eroded due to excessive cold water transport to the Arctic in the St Anna Trough. The excess of cold water is a result of negative sea-ice biases near shelf regions exposing the ocean to the cold atmosphere.

These relationships are not apparent in the ORAs—products with a low Fram Strait transport also have high BSO transports, but these products do not necessarily have a cold Arctic Ocean. The MMM heat transport through the Fram Strait is lower than observed, its heat transport through the BSO is close to the observed and its AW layer is too cold. This indicates that in the ORAs, probably due to their better resolution compared to CORE-II, and to data assimilation, the transport over the St. Anna Trough is more realistic while, on average, their Fram Strait heat fluxes remain too cold resulting in a cold Arctic Ocean.

Of individual products, GECCO2 and MOVE-G2i show excessive transports through the BSO but very low in the Fram Strait (Fig. 8). However, their Arctic Oceans are much warmer than observed. For example, the MOVE-G2i Eurasian basin mean temperature is the highest among the ORAs in the top 700 m (Fig. 13). In GECCO2 the warm Arctic can be explained by its extremely warm Nordic Seas and high BSO heat transport. MOVE-G2i, on the other hand, has a low average heat transport in the Fram Strait and rather cold Nordic Seas indicating that its warm Arctic Ocean must be of different origin than the one of GECCO2. MOVE-G2i heat content in the top 100 m is very high in the Barents Sea, compared to other ORAs and observational data, pointing to a major heat pathway to the Arctic in MOVE-G2i (Figure S9). It is probable that this exceptional heat transport pattern results in a particular heat and temperature distribution in the Eurasian basin. However, a realistic looking heat transport alone does not guarantee the correct Arctic hydrography. For

example, GloSea5-GO5 and ORAP5 capture the transports from the Atlantic well, but are still too cold in the Arctic.

There is a possible link between the ORA MLD and the cold biases in the top 100 m layer. As shown, the MMM MLD is too shallow in summer and too deep in winter in association with cool biases in the Arctic Ocean and the coastal Antarctic waters (Table 3). These biases could be related in the following way: in summer a too shallow mixed layer ($\ll$ 100 m) reduces the flux of the atmospheric heat to the ocean and the top 100 m layer stays cooler, and possibly fresher due to more winter ice melt in ORAs, as noted earlier. The shallow mixed layer loses heat faster and becomes colder, allowing for an earlier freeze-up in autumn, and the increased salt flux from ice contributes to the deepening of mixed layer in ORAs. In winter, there is always at least narrow open water areas even in compact ice fields due to ice dynamics so that a deeper mixed layer loses more heat to the cold atmosphere, which in non-coupled ORAs acts as an infinite heat sink, generating a cooler and denser upper ocean in the Arctic Ocean and the coastal Antarctic waters. Denser upper ocean waters sink and further deepen the winter mixed layer.

In the Southern Ocean, the ORAs are generally too warm in the upper 300 m which, combined with too fresh or close to observed salinity in the upper 100 m (Figs. 24, 25). This results in a stratification that is more stable than observed and is associated with low MLDs in summer. At deeper levels the ORA hydrography is close to WOA13 and EN4.2.g10i with the exception of GECCO2 due to its low salinity. Compared to the CORE-II models, which tend to have too cold Southern Ocean south of 50 °S (Downes et al. 2015), the ORA MMM shows a warm bias in the upper 700 m. As in CORE-II, the ORA hydrographic biases are likely associated with differences in oceanic transports, for instance in the ACC. GloSea5-GO5 and ORAP5 have the lowest volume transports in the ACC linked to significant cold biases in the Drake Passage (Table 3). In contrast, the ORAs with a higher than MMM ACC (ECDA3, GECCO2, GLORYS2v4, and MOVE-G2i) have positive heat biases in the Drake Passage. The model resolution seems to matter as the low resolution ORAs (ECDA3, GECCO2 and MOVE-G2i) have high ocean transports in the ACC, while the high resolution, eddy permitting ORAs have volume transports matching rather well with observational estimates. This is consistent with Farneti et al. (2015) who found that the better representation of ocean eddies among the CORE-II models resulted in a more realistic ACC. One might expect that the realistic ACC due to higher resolution would also reproduce realistic temperatures, but this is not apparent from our diagnostics results (Table 3).
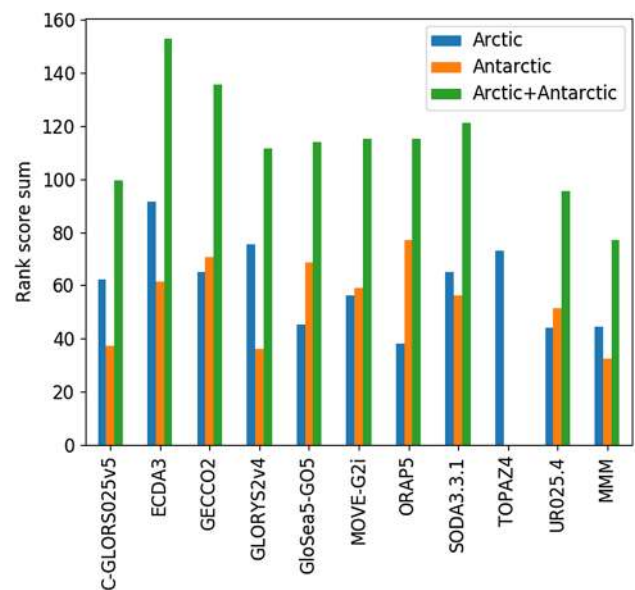

Fig. 26 Sums of rank scores of various diagnostic variables for individual ORAs and their mean product (MMM). The ranking is based on standard deviations of differences between the ORAs and observational data in the Arctic (blue bars), Antarctic (orange bars) and both added together (green bars) shown in Table 3. A smaller sum indicates a closer agreement with the observational data and a better performance. The Arctic snow depth and net liquid ocean heat tranport diagnostic rank scores were excluded from the sums as GECCO2 did not provide snow depth and the ocean transport diagnostic provided the mean values of net transports only

5.4 Synthesis of diagnostics

As has been seen, basin-wide mean errors are sensitive to opposing local biases and may give small values when, for example, large negative and positive biases cancel out. The basin-wide standard deviation of errors, on the other hand, is not similarly affected by spatial errors and obtains a large value in the aforementioned case. Therefore the products are ranked based on their standard deviations of errors in such way that for each diagnostics the product with the smallest standard deviation obtains a score of one, the product with the next smallest standard deviation a score of two and so forth. Hence, the products with relatively small standard deviations get a sum of scores smaller than those with larger standard deviations and can be assumed to perform better. However, our ranking approach is somewhat sensitive to the selection of the observational reference data and to which diagnostics enter the ranking. Hence, relatively small differences between ranking scores do not necessarily indicate significant differences in performance. However, even though a single score might be questioned we think that the general picture emerging from the sum of rank scores and linkages between diagnostics are not. Rank scores summed

across individual diagnostics in the Arctic and Antarctic are illustrated in Fig. 26 for each ORA and the MMM.

In the Arctic, only ORAP5 and GloSea5-GO5 have lower sums of rank scores than MMM, while UR025.4 is close to the MMM (Fig. 26). The remaining ORA rank score sums are clearly higher than the MMM one. In the Antarctic, the good performance of the MMM is even clearer as it obtains the lowest sum of rank scores. Closest to it are now GLORYS2v4 and C-GLORS025v5, then UR025.4 and SODA3.3.1, followed by the other ORAs. ORAP5, which is the best performer in the Arctic has the largest sum of rank scores in the Antarctic. ECDA3, the only coupled reanalysis, stands out as having the highest global sum of rank scores, followed by GECCO2 and SODA3.3.1. Globally, the MMM has the lowest sum of rank scores, followed by UR025.4 and C-GLORS025v5. In addition to the ranking based on standard deviation of errors, another ranking was carried out using absolute values of basin-wide mean errors (not shown). This ranking produced overall similar results to the ones based on standard deviations—the sum of MMM rank scores was smaller than ones for individual ORAs, globally. Results from the rankings support the good performance of the MMM and its usefulness in describing polar ocean states.

6 Conclusion

We have analysed several aspects of ten ocean reanalysis products in the Arctic and Antarctic. In this paper we concentrate on comparing the mean states of the ORAs. This is the first step towards more comprehensive analyses of interannual variability and co-variability between different fields as proposed for instance in the ongoing European H2020 APPLICATE project. The biases identified, and their potential linkages, will assist the developers to improve their products and inform the users of product quality. We emphasise that this paper is a snapshot of a moving target, because these products are constantly evolving and being updated regularly, quite often in response to ORA intercomparison efforts. Nonetheless, the performance may remain representative for a while, as was found when comparing the Arctic sea-ice diagnostics results in our study with Chevallier et al. (2017). Additionally, most of our polar diagnostics, with the exception of some related to Arctic sea ice, were carried out for the first time for a such an extensive set of ORAs.

In addition to interannual variability, future studies could focus on assessing the sea-ice dynamics in these reanalyses. A thorough evaluation of sea-ice drift in both polar oceans, similar to that performed in Chevallier et al. (2017) for the Arctic sea-ice cover, could be carried out, as sea-ice advection is one of the main mechanisms linking the polar regions with lower latitudes. An important outcome from such a

study would be a more comprehensive understanding of how the atmosphere–ocean energy transfer is represented in the current ocean reanalyses, and the role of sea ice in controlling this. As climate models presently fail to realistically replicate the global sea-ice trends, such an understanding is needed to enhance climate prediction skill (Turner and Comiso 2017; Rosenblum and Eisenman 2017). However, such a study would require more detailed diagnostics than currently available in the ORA-IP database, otherwise the results would not significantly differ from those of Chevallier et al. (2017).

Ideally, we would like to separate the impact of data assimilation from, for example, model physics, which would require the use of analysis increments as a measure. Carrying out such an approach can be considered as an in-depth assessement of a few ORAs, but is clearly beyond the scope of our study. In general, the earlier ORA-IP studies, except one, did not investigate the impact of analysis increments presumably for similar reasons. Based on six ORAs, where data were available for calculations, Valdivieso et al. (2014) found that the analysis increments were compensating for the inadequacies of the atmospheric reanalysis used to force the ORA, but the ORA ensemble still gave the best estimates of the oceanic and sea-ice quantities. Their results support our approach to pay attention to the performance of the MMM.

For the ORA ensemble mean state, we found that deviations from observational estimates were typically smaller than individual ORA anomalies, a well known characteristic of many climate model ensembles, often attributed to offsetting biases of individual ORAs. While this interpretation may be challenged (Rougier 2016), the ORA ensemble appears to be a useful product and, while knowing its anomalies and recognising its restrictions, it can be used to gain useful information on the physical state of the polar marine environment.

References

Aagaard K, Carmack E (1989) The role of sea ice and other fresh water in the Arctic circulation. J Geophys Res 94(C10):14485–14498

Aagaard K, Greisman P (1975) Toward new mass and heat budgets for the Arctic Ocean. J Geophys Res 80:3821–3827

Abernathey RP, Cerovecki I, Holland PR, Newsom E, Mazloff M, Talley LD (2016) Water-mass transformation by sea ice in the upper branch of the Southern Oceanoverturning. Nat Geosci 9:596. https://doi.org/10.1038/ngeo2749 **(ISSN 1752-0894)**

Argo (2000) Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE

Azaneu M, Kerr R, Matai MM (2014) Assessment of the representation of Antarctic Bottom Water properties in the ECCO2 reanalysis. Ocean Sci 10:923–946

Balmaseda MA, Hernandez F, Storto A, Palmer MD, Alves O, Shi L, Smith GC, Toyoda T, Valdivieso M, Barnier B, Behringer D, Boyer T, Chang Y-S, Chepurin GA, Ferry N, Forget G, Fujii Y, Good S, Guinehut S, Haines K, Ishikawa Y, Keeley S, Köhl A, Lee T, Martin MJ, Masina S, Masuda S, Meyssignac B, Mogensen K, Parent L, Peterson KA, Tang YM, Yin Y, Vernieres G, Wang X, Waters J, Wedd R, Wang O, Xue Y, Chevallier M, Lemieux J-F, Dupont F, Kuragano T, Kamachi M, Awaji T, Caltabiano A, Wilmer-Becker K, Gaillard F (2015) The Ocean Reanalyses Intercomparison Project (ORA-IP). J Oper Oceanogr 8(sup1):s80–s97. https://doi.org/10.1080/1755876X.2015.1022329

Barthélemy A, Fichefet T, Goosse H, Madec G (2015) Modeling the interplay between sea ice formation and the oceanic mixed layer: limitations of simple brine rejection parameterizations. Ocean Modell 86:141–152. https://doi.org/10.1016/j.ocemod.2014.12.009

Blockley EW, Martin MJ, McLaren AJ, Ryan AG, Waters J, Lea DJ, Mirouze I, Peterson KA, Sellar A, Storkey D (2014) Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts. Geosci Model Dev 7(6):2613–2638. https://doi.org/10.5194/gmd-7-2613-2014

Blockley E, Lea D, Martin M, McLaren A, Mirouze I, Ryan A, Siddorn J, Storkey D, Waters J (2015) Recent development and assessment of the Met Office operational global ocean forecasting system (FOAM). In: Buch E, Antoniou Y, Eparkhina D, Nolan G (eds) Operational oceanography for sustainable blue growth, Proceedings of the seventh EuroGOOS international conference, 28–30 October 2014, Lisbon, Portugal **(ISBN 978-2-9601883-1-8)**

Bonjean F, Lagerloef G (2002) Diagnostic model and analysis of the surface currents in the tropical Pacific Ocean. J Phys Oceanogr 32:2938–2954

Carton JA, Giese BS (2008) A reanalysis of ocean climate using simple ocean data assimilation (SODA). Mon Weather Rev 136(8):2999–3017. https://doi.org/10.1175/2007MWR1978.1 **(ISSN 0027-0644)**

Cavalieri DJ, Parkinson CL, Gloersen P, Comiso JC, Zwally HJ (1999) Deriving long-term time series of sea ice cover from satellite passive-microwave multisensor data sets. J Geophys Res 104:15803–15814. https://doi.org/10.1029/1999JC900081

Chang Y-S, Zhang S, Rosati A, Delworth TL, Stern WF (2013) An assessment of oceanic variability for 1960–2010 from the GFDL ensemble coupled data assimilation. Clim Dyn 40(3–4):775–803. https://doi.org/10.1007/s00382-012-1412-2 **(ISSN 0930-7575)**

Chevallier M, Smith GC, Dupont F, Lemieux J-F, Forget G, Fujii Y, Hernandez F, Msadek R, Peterson KA, Storto A, Toyoda T, Valdivieso M, Vernieres G, Zuo H, Balmaseda M, Chang Y-S, Ferry N, Garric G, Haines K, Keeley S, Kovach RM, Kuragano T, Masina S, Tang Y, Tsujino H, Wang X (2017) Intercomparison of the Arctic sea ice cover in global ocean-sea ice reanalyses from the ORA-IP project. Clim Dyn 49:1107–1136. https://doi.org/10.1007/s00382-016-2985-y **(ISSN 0930-7575)**

Comiso JC (2012) Large decadal decline of the arctic multiyear ice cover. J Clim 25(4):1176–1193. https://doi.org/10.1175/JCLI-D-11-00113.1 **(ISSN 0894-8755)**

Cullather RI, Bosilovich MG (2012) The energy budget of the polar atmosphere in MERRA. J Clim 25(1):5–24. https://doi.org/10.1175/2011JCLI4138.1 **(ISSN 08948755)**

Cunningham SA, Alderson SG, King BA, Brandon MA (2003) Transport and variability of the antarctic circumpolar current in drake passage. J Geophys Res 108:8084. https://doi.org/10.1029/2001JC001147

Curry B, Lee CM, Petrie B, Moritz RE, Kwok R (2014) Multiyear volume, liquid freshwater, and sea ice transports through Davis Strait, 2004–2010. J Phys Oceanogr 44:1244–1266

Danabasoglu G, Yeager SG, Bailey DA, Behrens E, Mats Bentsen D, Bi AB, Böning C, Bozec A, Canuto VM, Cassou C, Chassignet EP, Coward AC, Danilov S, Diansky N, Drange H, Farneti R, Fernandez E, Fogli PG, Forget G, Yosuke Fujii SM, Griffies AG, Heimbach P, Howard A, Jung T, Maxwell Kelley WG, Large AL, Jianhua L, Madec G, Marsland SJ, Masina S, Antonio Navarra AJ, Nurser G, Pirani A, Mélia DS, Samuels BL, Scheinert M, Sidorenko D, Treguier AM, Tsujino H, Uotila P, Valcke S, Voldoire A, Wang Q (2014) North Atlantic simulations in coordinated ocean-ice reference experiments phase II (CORE-II) Part I: mean states. Ocean Model 73:76–107. https://doi.org/10.1016/j.ocemod.2013.10.005 **(ISSN 14635003)**

de Boyer Montégut C, Madec G, Fischer AS, Lazar A, Iudicone D (2004) Mixed layer depth over the global ocean: an examination of profile data and a profile-based climatology. J Geophys Res 109(12):C12003. https://doi.org/10.1029/2004JC002378 **(ISSN 0148-0227)**

de Lavergne C, Palter JB, Galbraith ED, Bernardello R, Marinov I (2014) Cessation of deep convection in the open Southern Ocean under anthropogenic climate change. Nat Clim Chang 4:278–282

de Verdiére C, Ollitrault M (2016) A direct determination of the world ocean barotropic circulation. J Phys Ocean. https://doi.org/10.1175/JPO-D-15-0046.1

Dohan K, Maximenko N (2010) Monitoring ocean currents with satellite sensors. Oceanography 23(4):94–103. https://doi.org/10.5670/oceanog.2010.08

Donohue KA, Tracey KL, Watts DR, Chidichimo MP, Chereskin TK (2016) Mean antarctic circumpolar current transport measured in drake passage. Geophys Res Lett 43:11760–11767. https://doi.org/10.1002/2016GL070319

Downes SM, Farneti R, Uotila P, Griffies SM, Marsland SJ, Bailey D, Behrens E, Bentsen M, Bi D, Biastoch A, Böning C, Bozec A, Canuto VM, Chassignet E, Danabasoglu G, Danilov S, Diansky N, Drange H, Fogli PG, Gusev A, Howard A, Ilicak M, Jung T, Kelley M, Large WG, Leboissetier A, Long M, Jianhua L, Masina S, Mishra A, Antonio Navarra AJ, Nurser G, Patara L, Samuels BL, Sidorenko D, Spence P, Tsujino H, Wang Q, Yeager SG (2015) An assessment of southern ocean water masses and sea ice during 1988–2007 in a suite of interannual core-ii simulations. Ocean Modell 94:67–94. https://doi.org/10.1016/j.ocemod.2015.07.022 **(ISSN 1463-5003)**

Duan Y, Liu H, Yu W, Hou Y (2016) The mean properties and variations of the Southern Hemisphere subpolar gyres estimated by

simple ocean data assimilation (SODA) products. Acta Oceanol Sin 35:8–13

Eisenman I, Meier WN, Norris JR (2014) A spurious jump in the satellite record: has Antarctic sea ice expansion been overestimated? Cryosphere 8:1289–1296

Fahrbach E, Meincke J, Østerhus S, Rohardt G, Schauer U, Tverberg V, Verduin J (2001) Direct measurements of volume transports through Fram Strait. Polar Res 20(2):217–224

Farneti R, Downes SM, Griffies SM, Marsland SJ, Behrens E, Mats Bentsen D, Bi AB, Böning C, Bozec A, Canuto VM, Eric Chassignet G, Danabasoglu SD, Diansky N, Drange H, Fogli PG, Gusev A, Hallberg RW, Howard A, Ilicak M, Jung T, Maxwell Kelley WG, Large AL, Long M, Jianhua L, Masina S, Mishra A, Antonio Navarra AJ, Nurser G, Patara L, Samuels BL, Sidorenko D, Hiroyuki Tsujino P, Uotila QW, Yeager SG (2015) An assessment of Antarctic Circumpolar Current and Southern Ocean Meridional Overturning Circulation during 1958–2007 in a suite of interannual CORE-II simulations. Ocean Model 93:84. https://doi.org/10.1016/j.ocemod.2015.07.009 **(ISSN 14635003)**

Gammelsrød T, Leikvin V, Lien Ø, Budgell WP, Loeng H, Maslowski W (2009) Mass and heat transports in the NE Barents sea: observations and models. J Mar Sys 75:56–69

Ganachaud A, Wunsch C (2000) Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. Nature 408:453–457

Garric G, Parent L, Greiner E, Drévillon M, Hamon M, Lellouche JM, Régnier C, Desportes C, Le Galloudec O, Bricaud C, Drillet Y, Hernandez F, Dubois C, Le Traon P-Y (2017) Performance and quality assessment of the global ocean eddy-permitting physical reanalysis glorys2v4. In: Buch E, Fernandez V, Nolan G, Eparkhina D (eds) Operational oceanography serving sustainable marine development. Proceedings of the eight EuroGOOS international conference, 2018, 3–5 October 2017, Bergen, Norway **(ISBN 978-2-9601883-3-2)**

Gille ST (2008) Decadal-scale temperature trends in the Southern Hemisphere ocean. J Clim 21:4749–4765

Good SA, Martin MJ, Rayner NA (2013) En4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. J Geophys Res Oceans 118(12):6704–6716. https://doi.org/10.1002/2013JC009067 **(ISSN 2169-9291)**

Griffies SM, Biastoch A, Böning C, Bryan F, Danabasoglu G, Chassignet EP, England MH, Gerdes R, Haak H, Hallberg RW, Hazeleger W, Jungclaus J, Large WG, Madec G, Pirani A, Samuels BL, Scheinert M, Gupta AS, Severijns CA, Simmons HL, Treguier AM, Winton M, Yeager SG, Yin J (2009) Coordinated ocean-ice reference experiments (COREs). Ocean Model 26(1–2):1–46. https://doi.org/10.1016/j.ocemod.2008.08.007 **(ISSN 14635003)**

Haumann FA, Gruber N, Münnich M, Frenger I, Kern S (2016) Sea-ice transport driving Southern Ocean salinity and its recent trends. Nature 537(7618):89–92. https://doi.org/10.1038/nature19101 **(ISSN 0028-0836)**

Heimbach P, Menemenlis D, Losch M, Campin JM, Hill C (2010) On the formulation of sea-ice models. Part 2: Lessons from multi-year adjoint sea-ice export sensitivities through the Canadian Arctic Archipelago. Ocean Model 33(12):145–158. https://doi.org/10.1016/j.ocemod.2010.02.002 **(ISSN 1463-5003)**

Holte J, Talley L (2009) A new algorithm for finding mixed layer depths with applications to argo data and subantarctic mode water formation. J Atmos Ocean Technol 26(9):1920–1939. https://doi.org/10.1175/2009JTECHO543.1 **(ISSN 0739-0572)**

Huang CJ, Qiao F, Dai D (2014) Evaluating CMIP5 simulations of mixed layer depth during summer. J Geophys Res Oceans 119(4):2568–2582. https://doi.org/10.1002/2013JC009535 **(ISSN 2169-9291)**

Ilicak M, Drange H, Wang Q, Gerdes R, Aksenov Y, Bailey D, Bentsen M, Biastoch A, Bozec A, Böning C, Cassou C, Chassignet E, Coward AC, Curry B, Danabasoglu G, Danilov S, Fernandez E, Fogli PG, Fujii Y, Griffies SM, Iovino D, Jahn A, Jung T, Large WG, Lee C, Lique C, Jianhua L, Simona Masina AJ, Nurser G, Roth C, David Salas Y, Mlia BL, Samuels PS, Tsujino H, Valcke S, Voldoire A, Wang X, Yeager SG (2016) An assessment of the arctic ocean in a suite of interannual core-II simulations. Part III: Hydrography and fluxes. Ocean Modell 100:141–161. https://doi.org/10.1016/j.ocemod.2016.02.004 **(ISSN 1463-5003)**

IPCC (2013) Climate change 2013: the physical science basis. Contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781107415324 **(ISBN 978-1-107-66182-0)**

Ivanova N, Johannessen OM, Pedersen LT, Tonboe R (2014) Retrieval of arctic sea ice parameters by satellite passive microwave sensors: a comparison of eleven sea ice concentration algorithms. IEEE Trans Geosci Remote Sens 52(11):7233–7246

Jahn A, Aksenov Y, de Cuevas B, de Steur L, Häkkinen S, Hansen E, Herbaut C, Houssais MN, Karcher M, Kauker F, Lique C, Nguyen A, Pemberton P, Worthen D, Zhang J (2012) Arctic Ocean freshwater: how robust are model simulations? J Geophys Res 117:C00D16. https://doi.org/10.1029/2012JC007907 **(ISSN 0148-0227)**

Jahn A, Kay JE, Holland MM, Hall DM (2016) How predictable is the timing of a summer ice-free Arctic? Geophys Res Lett 43:9113. https://doi.org/10.1002/2016GL070067 **(ISSN 00948276)**

Jones JM et al (2016) Assessing recent trends in high-latitude Southern Hemisphere surface climate. Nat Clim Chang 6:917–926. https://doi.org/10.1038/NCLIMATE3103

Karcher MJ, Gerdes R, Kauker F, Köberle C (2003) Arctic warming: evolution and spreading of the 1990s warm event in the Nordic seas and the Arctic Ocean. J Geophys Res 108(C2):3034. https://doi.org/10.1029/2001JC001265

Karcher M, Smith JN, Kauker F, Gerdes R, Smethie WM (2012) Recent changes in arctic ocean circulation revealed by iodine-129 observations and modeling. J Geophys Res Oceans 117(C8):C08007. https://doi.org/10.1029/2011JC007513 **(ISSN 2156-2202)**

Karspeck AR, Stammer D, Köhl A, Danabasoglu G, Balmaseda M, Smith DM, Fujii Y, Zhang S, Giese B, Tsujino H, Rosati A (2015) Comparison of the Atlantic meridional overturning circulation between 1960 and 2007 in six ocean reanalysis products. Clim Dyn 49:957. https://doi.org/10.1007/s00382-015-2787-7 **(ISSN 0930-7575)**

Kauker F, Kaminski T, Karcher M, Giering R, Gerdes R, Vobeck M (2009) Adjoint analysis of the 2007 all time Arctic sea-ice minimum. Geophys Res Lett 36(3):L03707. https://doi.org/10.1029/2008GL036323 **(ISSN 1944-8007)**

Koenig Z, Provost C, Ferrari R, Sennéchael N, Rio M-H (2014) Volume transport of the Antarctic Circumpolar Current: production and validation of a 20 year long time series obtained from in situ and satellite observations. J Geophys Res Oceans 119:5407–5433. https://doi.org/10.1002/2014JC009966

Köhl A (2015) Evaluation of the GECCO2 ocean synthesis: transports of volume, heat, and freshwater in the Atlantic. Q J R Meteorol Soc 141(686):166–181

Kurtz NT, Markus T (2012) Satellite observations of Antarctic sea ice thickness and volume. J Geophys Res 117(C8):1–9. https://doi.org/10.1029/2012JC008141 **(ISSN 0148-0227)**

Lien VS, Hjøllo SS, Skogen MD, Svendsen E, Wehde H, Bertino L, Counillon F, Chevallier M, Garric G (2016) An assessment of the added value from data assimilation on modelled Nordic Seas hydrography and ocean transports. Ocean Model 99:43–59. https://doi.org/10.1016/j.ocemod.2015.12.010 **(ISSN 14635003)**

Lindsay R (2010) New unified sea ice thickness climate data record. Eos Trans Am Geophys Union 91(44):405–406. https://doi.org/10.1029/2010EO440001 **(ISSN 2324-9250)**

Lindsay R, Schweiger A (2015) Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations. Cryosphere 9(1):269–283

Lindsay R, Wensnahan M, Schweiger A, Zhang J (2014) Evaluation of seven different atmospheric reanalysis products in the Arctic. J Clim 27(7):2588–2606

Locarnini RA, Mishonov AV, Antonov JI, Boyer TP, Garcia HE, Baranova OK, Zweng MM, Paver CR, Reagan JR, Johnson DR, Hamilton M, Seidov D (2013) World Ocean Atlas 2013, Volume 1: temperature, S. Levitus, Ed., A. Mishonov Technical Ed. NOAA Atlas NESDIS 73:40

Lumpkin R, Speer K (2007) Global ocean meridional overturning. J Phys Oceanogr 37:2550–2562. https://doi.org/10.1175/JPO3130.1

Maksym T, Stammerjohn S, Ackley S, Massom R (2012) Antarctic sea ice—a polar opposite? Oceanography 25(3):140–151. https://doi.org/10.5670/oceanog.2012.88

Marshall GJ (2003) Trends in the southern annular mode from observations and reanalyses. J Clim 16(24):4134–4143. https://doi.org/10.1175/1520-0442(2003)016<4134:TITSAM>2.0.CO;2

Masina S, Storto A, Ferry N, Valdivieso M, Haines K, Balmaseda M, Zuo H, Drevillon M, Parent L (2015) An ensemble of eddy-permitting global ocean reanalyses from the MyOcean project. Clim Dyn 49:813. https://doi.org/10.1007/s00382-015-2728-5 **(ISSN 1432-0894)**

Massonnet F, Mathiot P, Fichefet T, Goosse H, König Beatty C, Vancoppenolle M, Lavergne T, Lavergne T (2013) A model reconstruction of the Antarctic sea ice thickness and volume changes over 1980–2008 using data assimilation. Ocean Model 64:67–75

Melia N, Haines K, Hawkins E (2015) Improved Arctic sea ice thickness projections using bias-corrected CMIP5 simulations. Cryosphere 9(6):2237–2251. https://doi.org/10.5194/tc-9-2237-2015 **(ISSN 1994-0424)**

Palmer MD, Roberts CD, Balmaseda M, Chang Y-S, Chepurin G, Ferry N, Fujii Y, Good SA, Guinehut S, Haines K, Hernandez F, Köhl A, Lee T, Martin MJ, Masina S, Masuda S, Peterson KA, Storto A, Toyoda T, Valdivieso M, Vernieres G, Wang O, Xue Y (2017) Ocean heat content variability and change in an ensemble of ocean reanalyses. Clim Dyn 49:909–930. https://doi.org/10.1007/s00382-015-2801-0 **(ISSN 1432-0894)**

Parkinson CL, Cavalieri DJ (2012) Antarctic sea ice variability and trends, 1979–2010. Cryosphere 6:871–880

Pellichero V, Sallée JB, Schmidtko S, Roquet F, Charrassin J-B (2017) The ocean mixed layer under southern ocean sea-ice: seasonal cycle and forcing. J Geophys Res Oceans 122(2):1608–1633. https://doi.org/10.1002/2016JC011970 **(ISSN 2169-9291)**

Polyakov IV, Bhatt US, Walsh JE, Povl Abrahamsen E, Pnyushkov AV, Wassmann PF (2013) Recent oceanic changes in the Arctic in the context of long-term observations. Ecol Appl 23(8):1745–1764. https://doi.org/10.1890/11-0902.1 **(ISSN 10510761)**

Polyakov IV, Pnyushkov AV, Alkire MB, Ashik IM, Baumann TM, Carmack EC, Goszczko I, Guthrie J, Ivanov VV, Kanzow T, Krishfield R, Kwok R, Sundfjord A, Morison J, Rember R, Yulin A (2017) Greater role for Atlantic inflows on sea-ice loss in the Eurasian Basin of the Arctic Ocean. Science (80) 356(6335):285–291. https://doi.org/10.1126/science.aai8204 **(ISSN 0036-8075)**

Proshutinsky A, Krishfield R, Timmermans M-L, Toole J, Carmack E, McLaughlin F, Williams WJ, Zimmermann S, Itoh M, Shimada K (2009) Beaufort gyre freshwater reservoir: state and variability from observations. J Geophys Res Oceans 114(C1):C00A10. https://doi.org/10.1029/2008JC005104 **(ISSN 2156-2202)**

Proshutinsky A, Steele M, Timmermans M-L (2016) Forum for Arctic modeling and observational synthesis (FAMOS): past, current, and future activities. J Geophys Res Ocean 121(6):3803–3819. https://doi.org/10.1002/2016JC011898 **(ISSN 21699275)**

Rabe B, Karcher M, Kauker F, Schauer U, Toole JM, Krishfield RA, Pisarev S, Kikuchi T, Su J (2014) Arctic Ocean basin liquid freshwater storage trend 1992–2012. Geophys Res Lett 41(3):961–968. https://doi.org/10.1002/2013GL058121 **(ISSN 00948276)**

Roach AT, Aagaard K, Pease CH, Salo SA, Weingartner T, Pavlov V, Kulakov M (1995) Direct measurements of transport and water properties through the Bering Strait. J Geophys Res 100(C9):18443–18457

Roemmich D et al (2015) Unabated planetary warming and its ocean structure since 2006. Nat Clim Chang 5:240–245

Roquet F et al (2015) A Southern Indian Ocean database of hydrographic profiles obtained with instrumented elephant seals. Sci Data 1:140028. https://doi.org/10.1038/sdata.2014.28

Roquet F, Boehme L, Fedak MA, Block B, Charrassin JB, Costa D, Hückstädt LA, Guinet C, Harcourt RG, Hindell MAA, McMahon CR, Woodward B (2017) Ocean observations using tagged animals. Oceanography

Rosenblum E, Eisenman I (2017) Sea ice trends in climate models only accurate in runs with biased global warming. J Clim 30(16):6265–6278. https://doi.org/10.1175/JCLI-D-16-0455.1 **(ISSN 0894-8755)**

Rougier J (2016) Ensemble averaging and mean squared error. J Clim 29(24):8865–8870. https://doi.org/10.1175/JCLI-D-16-0012.1 **(ISSN 0894-8755)**

Sallée JB, Shuckburgh E, Bruneau N, Meijers AJS, Bracegirdle TJ, Wang Z (2013) Assessment of southern ocean mixed-layer depths in CMIP5 models: historical bias and forcing response. J Geophys Res Oceans 118(4):1845–1862. https://doi.org/10.1002/jgrc.20157 **(ISSN 2169-9291)**

Schauer U, Beszczynska-Möller A (2009) Problems with estimation and interpretation of oceanic heat transport—conceptual remarks for the case of Fram Strait in the Arctic Ocean. Ocean Sci 5:487–494

Schmidtko S, Heywood KJ, Thompson AF, Aoki S (2014) Multidecadal warming of Antarctic waters. Science 346:1227–1231

Shi L, Alves O, Wedd R, Balmaseda MA, Chang Y, Chepurin G, Ferry N, Fujii Y, Gaillard F, Good SA, Guinehut S, Haines K, Hernandez F, Lee T, Palmer M, Peterson KA, Masuda S, Storto A, Toyoda T, Valdivieso M, Vernieres G, Wang X, Yin Y (2017) An assessment of upper ocean salinity content from the Ocean Reanalyses Inter-comparison Project (ORA-IP). Clim Dyn 49:1009–1029. https://doi.org/10.1007/s00382-015-2868-7 **(ISSN 0930-7575)**

Smedsrud LH, Esau I, Ingvaldsen RB, Eldevik T, Haugan PM, Li C, Lien VS, Olsen A, Omar AM, Otterå OH, Risebrobakken B, Sandø AB, Semenov VA, Sorokina SA (2013) The role of the Barents Sea in the climate system. Rev Geophys 51:415–449

Stackhouse PW Jr, Gupta SK, Cox SJ, Zhang T, Mikovitz JC, Hinkelman LM (2011) 24.5-year srb date set released. GEWEX News 21:10–12

Storto A, Masina S, Navarra A (2016) Evaluation of the CMCC eddy-permitting global ocean physical reanalysis system (C-GLORS, 19822012) and its assimilation components. Q J R Meteorol Soc 142:738–758. https://doi.org/10.1002/qj.2673

Storto A, Masina S, Balmaseda M, Guinehut S, Xue Y, Szekely T, Fukumori I, Forget G, Chang Y-S, Good SA, Köhl A, Vernieres G, Nicolas Ferry K, Peterson A, Behringer D, Ishii M, Masuda S, Fujii Y, Toyoda T, Yin Y, Valdivieso M, Barnier B, Boyer T, Lee T, Jérome Gourrion O, Wang PH, Rosati A, Kovach R, Hernandez F, Martin MJ, Kamachi M, Kuragano T, Mogensen

K, Alves O, Haines K, Wang X (2017) Steric sea level variability (1993–2010) in an ensemble of ocean reanalyses and objective analyses. Clim Dyn 49(3):709–729. https://doi.org/10.1007/s00382-015-2554-9 **(ISSN 1432-0894)**

Stroeve J, Kattsov VM, Barrett A, Serreze M, Pavlova T, Holland MM, Meier WN (2012) Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations. Geophys Res Lett 39(16):1–7. https://doi.org/10.1029/2012GL052676 **(ISSN 0094-8276)**

Sumata H, Kauker F, Karcher M, Rabe B, Timmermans M-L, Behrendt A, Gerdes R, Schauer U, Shimada K, Cho K-H, Kikuchi T (2017) Decorrelation scales for arctic ocean hydrography Part I: Amerasian basin. Ocean Sci Discuss. https://doi.org/10.5194/os-2017-70

Sunke Schmidtko GC, Johnson JML (2013) Mimoc: a global monthly isopycnal upper-ocean climatology with mixed layers. J Geophys Res Oceans 118(4):1658–1672. https://doi.org/10.1002/jgrc.20122 **(ISSN 2169-9291)**

Tietsche S, Balmaseda MA, Zuo H, Mogensen K (2015) Arctic sea ice in the global eddy-permitting ocean reanalysis ORAP5. Clim Dyn. https://doi.org/10.1007/s00382-015-2673-3 **(ISSN 0930-7575)**

Toyoda T, Fujii Y, Yasuda T, Usui N, Ogawa K, Kuragano T, Tsujino H, Kamachi M (2016) Data assimilation of sea ice concentration into a global ocean-sea ice model with corrections for atmospheric forcing and ocean temperature fields. J Oceanogr 72(2):235–262. https://doi.org/10.1007/s10872-015-0326-0 **(ISSN 0916-8370)**

Toyoda T, Fujii Y, Kuragano T, Kamachi M, Ishikawa Y, Masuda S, Sato K, Awaji T, Hernandez F, Ferry N, Guinehut S, Martin MJ, Andrew Peterson K, Good SA, Valdivieso M, Haines K, Storto A, Masina S, Khl A, Zuo H, Balmaseda M, Yin Y, Shi L, Alves O, Smith G, Chang Y-S, Vernieres G, Wang X, Forget G, Patrick Heimbach O, Wang IF, Lee T (2017a) Intercomparison and validation of the mixed layer depth fields of global ocean syntheses. Clim Dyn 49(3):753–773. https://doi.org/10.1007/s00382-015-2637-7 **(ISSN 1432-0894)**

Toyoda T, Fujii Y, Kuragano T, Kosugi N, Sasano D, Kamachi M, Ishikawa Y, Masuda S, Sato K, Awaji T, Hernandez F, Ferry N, Guinehut S, Matthew Martin K, Peterson A, Good SA, Valdivieso M, Haines K, Storto A, Masina S, Köhl A, Yin Y, Shi L, Alves O, Smith G, Chang Y-S, Vernieres G, Wang X, Forget G, Patrick Heimbach O, Wang IF, Lee T, Zuo H, Balmaseda M (2017b) Interannual-decadal variability of wintertime mixed layer depths in the North Pacific detected by an ensemble of ocean syntheses. Clim Dyn 49(3):891–907. https://doi.org/10.1007/s00382-015-2762-3 **(ISSN 1432-0894)**

Turner J, Comiso J (2017) Solve Antarctica's sea-ice puzzle. Nature 547(7663):275–277. https://doi.org/10.1038/547275a **(ISSN 0028-0836)**

Turner J, Hosking JS, Bracegirdle TJ, Marshall GJ, Phillips T (2015) Recent changes in Antarctic Sea Ice. Philos Trans R Soc A 373:20140163

Uotila P, Holland DM, Morales Maqueda MA, Häkkinen S, Holloway G, Karcher M, Kauker F, Michael Steele N, Yakovlev JZ et al (2006) An energy-diagnostics intercomparison of coupled ice-ocean Arctic models. Ocean Model 11(1–2):1–27. https://doi.org/10.1016/j.ocemod.2004.11.003 **(ISSN 1463-5003)**

Valdivieso M, Haines K, Zuo H, Lea D (2014) Freshwater and heat transports from global ocean synthesis. J Geophys Res Oceans 119:394–409. https://doi.org/10.1002/2013JC009357

Valdivieso M, Haines K, Balmaseda M, Chang Y-S, Drevillon M, Ferry N, Fujii Y, Köhl A, Storto A, Toyoda T, Wang X, Waters J, Xue Y, Yin Y, Barnier B, Hernandez F, Kumar A, Lee T, Simona Masina K, Peterson A (2017) An assessment of air-sea heat fluxes from ocean and coupled reanalyses. Clim Dyn 49(3):983–1008. https://doi.org/10.1007/s00382-015-2843-3

van Sebille E, Spence P, Mazloff MR, England MH, Rintoul SR, Saenko OA (2013) Abyssal connections of Antarctic Bottom Water in a Southern Ocean State estimate. Geophys Res Lett 40:2177–2182. https://doi.org/10.1002/grl.50483

Wang Q, Ilicak M, Gerdes R, Drange H, Aksenov Y, Bailey DA, Bentsen M, Biastoch A, Bozec A, Böning C, Cassou C, Chassignet E, Coward AC, Curry B, Danabasoglu G, Danilov S, Fernandez E, Fogli PG, Fujii Y, Griffies SM, Iovino D, Jahn A, Jung T, Large WG, Lee C, Lique C, Lu J, Masina S, Nurser AJG, Rabe B, Roth C, Mélia DS, Samuels BL, Spence P, Tsujino H, Valcke S, Voldoire A, Wang X, Yeager SG (2016b) An assessment of the Arctic Ocean in a suite of interannual CORE-II simulations Part II: Liquid freshwater. Ocean Model 99:86–109. https://doi.org/10.1016/j.ocemod.2015.12.009 **(ISSN 14635003)**

Wang Q, Ilicak M, Gerdes R, Drange H, Aksenov Y, Bailey DA, Bentsen M, Biastoch A, Bozec A, Böning C, Cassou C, Chassignet E, Coward AC, Beth Curry G, Danabasoglu SD, Fernandez E, Fogli PG, Yosuke Fujii SM, Griffies DI, Jahn A, Thomas Jung WG, Large CL, Lique C, Jianhua L, Simona Masina AJ, Nurser G, Rabe B, Roth C, Mélia SD, Samuels BL, Spence P, T Hiroyuki, Valcke S, Voldoire A, Wang X, Yeager SG (2016a) An assessment of the Arctic Ocean in a suite of interannual CORE-II simulations. Part I: Sea ice and solid freshwater. Ocean Model 99:110–132. https://doi.org/10.1016/j.ocemod.2015.12.008 **(ISSN 14635003)**

Warren SG, Rigor IG, Untersteiner N, Radionov VF, Bryaz-gin NN, Aleksandrov YI, Colony R (1999) Snow depth on Arctic sea ice. J Clim 12:18141829

Whitworth T, Peterson R (1985) Volume transport of the antarctic circumpolar current from bottom pressure measurements. J Phys Oceanogr 15:810–816

Woodgate RA, Aagaard K (2005) Revising the Bering Strait freshwater flux into the Arctic Ocean. Geophys Res Lett 32:L02602

Woodgate RA, Weingartner TJ, Lindsay R (2012) Observed increases in Bering Strait oceanic fluxes from the Pacific to the Arctic from 2001 to 2011 and their impacts on the Arctic Ocean water column. Geophys Res Lett 39:L24603

Worby AP, Geiger CA, Paget MJ, Woert MLV, Ackley SF, DeLiberty TL (2008) Thickness distribution of Antarctic sea ice. J Geophys Res 113:C05S92

Xie J, Bertino L, Counillon F, Lisæter KA, Sakov P (2017) Quality assessment of the TOPAZ4 reanalysis in the Arctic over the period 1991–2013. Ocean Sci 13(1):123–144. https://doi.org/10.5194/os-13-123-2017 **(ISSN 1812-0792)**

Zhang J (2014) Modeling the impact of wind intensification on antarctic sea ice volume. J Clim 27(1):202–214. https://doi.org/10.1175/JCLI-D-12-00139.1

Zunz V, Goosse H, Massonnet F (2013) How does internal variability influence the ability of CMIP5 models to reproduce the recent trend in Southern Ocean sea ice extent? Cryosphere 7:451–468

Zuo H, Mugford RI, Haines K, Smith GC (2011) Assimilation impacts on Arctic Ocean circulation, heat and freshwater budgets. Ocean Model 40(2):147–163. https://doi.org/10.1016/j.ocemod.2011.08.008 **(ISSN 14635003)**

Zuo H, Balmaseda MA, Mogensen K (2015) The new eddy-permitting ORAP5 ocean reanalysis: description, evaluation and uncertainties in climate signals. Clim Dyn https://doi.org/10.1007/s00382-015-2675-1

Zweng MM, Reagan JR, Antonov JI, Locarnini RA, Mishonov AV, Boyer TP, Garcia HE, Baranova OK, Johnson DR, Seidov D, Biddle MM (2013) World Ocean Atlas 2013, Volume 2: Salinity. S. Levitus, Ed., A. Mishonov Technical Ed. NOAA Atlas NESDIS 74:39

Zygmuntowska M, Rampal P, Ivanova N, Smedsrud LH (2014) Uncertainties in Arctic sea ice thickness and volume: new estimates and implications for trends. Cryosphere 8:705–720. https://doi.org/10.5194/tc-8-705-2014