Sea-ice forecasts with an upgraded AWI Coupled Prediction System

Longjiang Mu[1], Lars Nerger[2], Jan Streffing[1], Qi Tang[2], Bimochan Niraula[3], Lorenzo Zampieri[4], Svetlana Loza[5], and Helge Goessling[6]

[1]Alfred Wegener Institute for Polar and Marine Research
[2]Alfred Wegener Institute Helmholtz Center for Polar and Marine Research
[3]Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung
[4]Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung
[5]Alfred Wener Institute for Polar and Marine Research
[6]Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research

November 24, 2022

Abstract

A new version of the AWI Coupled Prediction System is developed based on the Alfred Wegener Institute Climate Model v3.0. Both the ocean and the atmosphere models are upgraded or replaced, reducing the computation time by a factor of 5 at a given resolution. This allowed us to increase the ensemble size from 12 to 30, maintaining a similar resolution in both model components. The online coupled data assimilation scheme now additionally utilizes sea-surface salinity and sea-level anomaly as well as temperature and salinity profile observations. Results from the data assimilation demonstrate that the sea-ice and ocean states are reasonably constrained. In particular, the temperature and salinity profile assimilation has mitigated systematic errors in the deeper ocean, although issues remain over polar regions where strong atmosphere-ocean-ice interaction occurs. One-year-long sea-ice forecasts initialized on January 1st, April 1st, July 1st and October 1st from 2003 to 2019 are described. To correct systematic forecast errors, sea-ice concentration from 2011 to 2019 is calibrated by trend-adjusted quantile mapping using the preceding forecasts from 2003 to 2010. The sea-ice edge raw forecast skill is within the range of operational global subseasonal-to-seasonal forecast systems, outperforming a climatological benchmark for about two weeks in the Arctic and about three weeks in the Antarctic. The calibration is much more effective in the Arctic: Calibrated sea-ice edge forecasts outperform climatology for about 45 days in the Arctic but only 27 days in the Antarctic. Both the raw and the calibrated forecast skill exhibit strong seasonal variations.

Sea-ice forecasts with an upgraded AWI Coupled Prediction System

Longjiang Mu[1,2], Lars Nerger[1], Jan Streffing[1,3], Qi Tang[1,4], Bimochan Niraula[1], Lorenzo Zampieri[1], Svetlana N. Loza[1], Helge F. Goessling[1]

[1]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[2]Pilot National Laboratory for Marine Science and Technology, Qingdao, China
[3]Jacobs University Bremen, Bremen, Germany
[4]Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

Key Points:

- We describe an upgrade of the AWI Coupled Prediction System with new ocean and atmosphere models and more observations assimilated.
- The assimilation of ocean temperature and salinity (surface and profile observations) improves the ocean state significantly.
- Calibrated sea-ice edge forecasts outperform a climatological benchmark for around one month in both hemispheres.

Corresponding author: Longjiang Mu, `longjiang.mu@awi.de`

Abstract

A new version of the AWI Coupled Prediction System is developed based on the Alfred Wegener Institute Climate Model v3.0. Both the ocean and the atmosphere models are upgraded or replaced, reducing the computation time by a factor of 5 at a given resolution. This allowed us to increase the ensemble size from 12 to 30, maintaining a similar resolution in both model components. The online coupled data assimilation scheme now additionally utilizes sea-surface salinity and sea-level anomaly as well as temperature and salinity profile observations. Results from the data assimilation demonstrate that the sea-ice and ocean states are reasonably constrained. In particular, the temperature and salinity profile assimilation has mitigated systematic errors in the deeper ocean, although issues remain over polar regions where strong atmosphere-ocean-ice interaction occurs. One-year-long sea-ice forecasts initialized on January 1st, April 1st, July 1st and October 1st from 2003 to 2019 are described. To correct systematic forecast errors, sea-ice concentration from 2011 to 2019 is calibrated by trend-adjusted quantile mapping using the preceding forecasts from 2003 to 2010. The sea-ice edge raw forecast skill is within the range of operational global subseasonal-to-seasonal forecast systems, outperforming a climatological benchmark for about two weeks in the Arctic and about three weeks in the Antarctic. The calibration is much more effective in the Arctic: Calibrated sea-ice edge forecasts outperform climatology for about 45 days in the Arctic but only 27 days in the Antarctic. Both the raw and the calibrated forecast skill exhibit strong seasonal variations.

Plain Language Summary

Ocean data sparseness and systematic model errors pose problems for the initialization of coupled seasonal forecasts, especially in polar regions. Our global forecast system follows a seamless approach with refined ocean resolution in the Arctic. The new version presented here features higher computational efficiency and utilizes more ocean and sea-ice observations. Ice-edge forecasts outperform a climatological benchmark for about one month, comparable to established systems.

1 Introduction

With the increasing scientific and socioeconomic demands for long-term sea ice prediction (Jung et al., 2016), dynamical and statistical models are following different strategies to enhance prediction skill. When it comes to sea-ice forecasting with dynamical models, it is now common practice to assimilate remotely sensed sea ice concentration (SIC), ensuring a basic level of sea-ice forecast skill. With the advent of sea ice thickness (SIT) observations from satellites such as CryoSat-2 (Ricker et al., 2014), SMOS (Tian-Kunze et al., 2014) and ICESat-2 (Petty et al., 2020) and their assimilation into forecast systems in recent years, increased sea ice forecast skill with lead times ranging from synoptic to seasonal time scale has been reported as a result of better SIT initialization (Yang et al., 2014; Collow et al., 2015; Blockley et al., 2018; Mu et al., 2019; Liu et al., 2019). This holds in particular during the melt season when SIT anomalies are determining how long the ice can withstand the summer melt.

Perfect-model studies suggest that some predictive skill for sea-ice forecasts should be achievable even after a whole year due to the memory of sea-surface temperature (SST) and SIT anomalies (Blanchard-Wrigglesworth et al., 2011; Tietsche et al., 2014; Day et al., 2016; Goessling et al., 2016). However, a large gap between potential and actual forecast skill remains, due to the sparseness of ocean observations and because substantial systematic errors prevail in all models. Recent studies suggest that applying bias correction, commonplace in seasonal forecasting for other predictands, to sea ice can effectively remove the systematic drift and thereby substantially increase the long-term forecast skill (Director et al., 2017; Dirkson et al., 2019).

The sparseness of observations in the polar regions is another limiting factor for the skill of sea-ice predictions, and in some cases even a lack of quality of oceanic data poses problems. For instance, Xie et al. (2019) found that the satellite sea-surface salinity (SSS) observations are inaccurate in the Arctic Ocean when SSS $< 24\,psu$. T/S profile observations are rather limited with only several ice-tethered profilers each year in the central Arctic, and most are concentrated in the Beaufort Sea.

The first version of our forecast system, termed the Seamless Sea Ice Prediction System (SSIPS v1.0, Mu et al., 2020), is based on the AWI Climate Model version 1.1 (AWI-CM1.1). The ocean/ice model component, FESOM (v1.4), employs an unstructured mesh with varying resolution ranging from about 25 km in the Arctic to about 100 km at lower latitudes and uses the finite-element method. The atmosphere component ECHAM (v6.3.02p4) is a spectral model with a resolution of T63L47, corresponding to about 200 km horizontal grid-point resolution and 47 vertical levels. SSIPS v1.0 assimilates SIC, SIT, sea ice drift (SID), and SST using the Local Error Subspace Transform Kalman Filter (Nerger et al., 2012). It employs only 12 ensemble members but is still computationally rather expensive, mainly due to intricacies of the finite-element method in FESOM1 as well as the requirement to use rather short time steps in ECHAM.

The availability of more efficient ocean and atmosphere model components has been identified as a promising way to increase the ensemble size and thereby to better represent the covariance of the state vector without compromising in terms of model resolution or throughput. We have thus upgraded our forecast system - now termed the AWI Coupled Prediction System - based on AWI-CM3, which features more efficient ocean and atmosphere model components. This enables the use of higher atmospheric resolution as well as a larger ensemble without speed reduction.

Another major enhancement concerns the set of assimilated observations. Subsurface temperatures have been shown to yield significantly increased regional winter sea-ice extent forecast skill (Bushuk et al., 2017). We have thus incorporated in-situ T/S profile assimilation in the new system, alongside the assimilation of satellite-derived SSS and sea-level anomaly (SLA) observations.

The paper is structured as follows. The model components and their coupling and the data assimilation are described in Section 2. Sections 3 introduces the forecast calibration method for sea ice and the metric for evaluations in the study. Section 4 provides the forecast experiment design. In Section 5 the forecast skill is evaluated. Summary and discussions are given in Section 6.

2 The AWI Coupled Prediction System

2.1 AWI-CM

A new version of the AWI Climate Model (AWI-CM3) has recently been developed (Streffing et al., 2022). The new ocean model is FESOM version 2.0, described in Danilov et al. (2017). The most important difference between FESOM version 1.0 and 2.0 is that the discretization method has been changed from finite elements to finite volumes, leading to a speed-up of integration by a factor 3-5. For the ocean model we still use the same CORE II mesh (`https://fesom.de/models/meshessetups/`) as in Mu et al. (2020) with a resolution of around 25 km in the tropical oceans, the northern North Atlantic and Arctic Ocean, and around 100 km in the subtropics and mid-latitudes. In the vertical, the ocean model employs Arbitrary Lagrangian Eulerian (ALE) coordinates, here with 47 levels at fixed depths. The sea ice model component is the successor of the Finite-Element Sea-Ice Model FESIM (Danilov et al., 2015). A single ice thickness category, zero-layer thermodynamics, and the "standard" elastic-viscous-plastic (sEVP, Bouillon et al., 2013; Danilov et al., 2015) rheology are used. The prognostic snow layer follows Owens and Lemke (1990). Flooding happens when snow becomes thick enough to submerge.

The atmosphere component now replaces ECHAM with OpenIFS, the open version of the Integrated Forecasting System (OpenIFS) maintained by the European Centre for Medium-Range Weather Forecasts (ECMWF). The version v43r3 is used at TL159L60 resolution, which means vertically 60 levels and horizontally a triangular truncation of the spherical harmonics at wavenumber 159 for the dynamics paired with a linear reduced Gaussian grid corresponding to 110 km resolution for the physics. The model is forced by historical greenhouse-gas forcing before 2015, and follows the scenario of the Shared Socioeconomic Pathway 5 (SSP5), that is "a world of rapid and unconstrained growth in economic output and energy use" (Kriegler et al., 2017), thereafter. A fixed aerosol climatology and land-use pattern is employed, which we consider acceptable given the moderate changes that occurred during the considered period which spans the first two decades of the 21st century.

FESOM2 integrates the model with a time step of 1800 s with each run using 72 cores. OpenIFS uses 71 cores and has a large time step of 3600 s, benefiting from the Lagrangian advection scheme. As a standalone executable, a runoff mapper maps the runoff from OpenIFS to FESOM2. The two components exchange information every hour through the recently released coupler OASIS3-MCT4 (Craig et al., 2017). In total, 144 cores are used for one coupled model instance. This computation cost is a dramatic reduction compared to that used for its predecessor (480 cores) in Mu et al. (2020). The spherical resolution in the atmosphere model in the current configuration is even higher than before, whereas the grid-point resolution is slightly lower. Overall, a larger ensemble with less computational cost and faster throughput is possible with the new system.

2.2 Data assimilation

As the first version of the forecast system described in Mu et al. (2020), the new AWI-CPS implements an Ensemble Kalman Filter with the Parallel Data Assimilation Framework (PDAF) (Nerger & Hiller, 2013, http://pdaf.awi.de). We again adopt the online-coupled data assimilation (Nerger et al., 2019) feature, which provides high efficiency and full parallelization of the daily analysis and forecast steps. Again, the Local Error Subspace Transform Kalman Filter (LESTKF, Nerger et al., 2012) that preserves and projects all the ensemble information onto the error subspace is used. The ensemble size is now 30, which is 2.5 times larger than that used in the first version (12). The online data assimilation enables direct MPI communication among all coupled-model instances. The processes owned by one coupled model gather information from all the processes, conduct the analysis step, and redistribute the updated state vector back to all the processes (grey box in Figure 1).

The current state vector includes sea-ice concentration, sea-ice thickness, sea-ice velocity (u-component and v-component), 3D temperature, 3D salinity, and sea-surface height. A comparison of the state vector against the old version is shown in Table 1. The initialization of the ensemble at the very beginning of the assimilation procedure on 1st January 2002 is performed by adding perturbations generated by second-order exact sampling (Pham, 2001) onto the state that is reconstructed by the leading empirical orthogonal functions decomposed from a 30-year simulation of the free-running model.

Another notable upgrade is that more different types of observations are assimilated. Apart from the assimilation of sea-ice concentration, sea-ice thickness, sea-ice velocity and sea-surface temperature in the previous system, sea-level anomaly, sea-surface salinity and temperature/salinity profiles are further assimilated now (Table 1). For sea-ice concentration, we use the product "Interim Sea Ice Concentration Climate Data Record from EUMETSAT OSI SAF" with ID OSI-430-b (Lavergne et al., 2019) from EUMETSAT OSI SAF. The spatial resolution is 25 km on the EASE grid. Given that the OSI-430-b is the extension of OSI-450, a discontinuity and deterioration of the ensemble spread and the analysis of sea-ice concentration over the transition period, as in (Mu et al., 2020), is avoided in the new system. The maximum observation error for sea-ice concentration is set to 0.15.

For sea-ice thickness, the CS2SMOS sea-ice thickness product (Ricker et al., 2017) that merges data from CryoSat-2 and SMOS operationally is used. We use the daily product (v202) to meet the daily assimilation cycle in our system, as before. The observation errors in the product are directly used for the assimilation. In addition, the daily sea-ice thickness data on Level 2 derived from the EnviSat satellite (Paul et al., 2018) are assimilated to cover the pre-CryoSat-2 period. Note that no sea-ice thickness observations are assimilated in the Antarctic currently.

The sea-ice drift product from OSI SAF (OSI-405-c) constrains both components of the sea-ice drift. A constant uncertainty of 4.1 cm/s is prescribed. While no clear direct benefits have been reported from sea-ice drift assimilation due to the small inertia of sea-ice movements (Mu et al., 2020), we keep the assimilation of drift data because it provides benefits due to the impact on other, less volatile variables through cross-covariances.

For sea-surface temperature we use the product from the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA), as before. The Level-2 sea-surface salinity data product from SMOS provided by European Space Agency (ESA, `https://earth.esa.int/eogateway/`) is also assimilated in the system. Both the sea-surface temperature and the sea-surface salinity product include error information explicitly.

The sea-level anomaly data are taken from the Copernicus Marine Service (GLOBAL OCEAN ALONG-TRACK L3 SEA SURFACE HEIGHTS REPROCESSED (1993-ONGOING) TAILORED FOR DATA ASSIMILATION). This product provides along-track data for individual sensor from all altimeter missions. The mean dynamic topography (MDT) is calculated by averaging the sea-surface height from a 30-years simulation of the free-running model, as in Skachko et al. (2019). Like for the sea-surface salinity observations, data thinning is applied by averaging the observations that are located in the same triangle of the ocean mesh.

Lastly, the EN.4.2.1 profiles data from the Met Office Hadley Centre are also assimilated, as in Tang et al. (2020). A pre-processing step re-arranges the data along the time axis and distributes the profiles into mesh partitions. Uncertainties for the profiles decay with depth, as in Xie and Zhu (2010), with maximum values of 0.5 *degree* and 0.04 *psu* for temperature and salinity. Like in the previous version of our forecast system, no data constraints are applied in the atmospheric component here, although tests with nudging of the free-tropospheric winds are ongoing.

All these observations are quality-checked and pre-processed by applying physical limitations. Gridded data are interpolated by distance-weighted average remapping onto the unstructured mesh. To avoid the initial shock, during the post-processing step the system constrains the absolute analysis increments to not exceed twice the ensemble spread, as in Sakov et al. (2012). The localization radius is set to 200 km. Considering the largely random atmospheric states, no ensemble inflation is required. The daily data assimilation with 30 ensemble members at the chosen resolution uses a total of 4320 cores in parallel. The typical wall-clock time required for the analysis of one year is about 4.5 hours.

2.3 Forecast

After the analysis step, forecasts are started directly from the analyzed fields with the same ensemble size (30). The same coupling frequency of 3600 s is applied between OpenIFS and FESOM2. These seamlessly continuing forecasts are carried out to cover one year, initialized at different times of the year over numerous years (see Sect. 4).

3 Calibration and metrics

Post-processing is applied to the sea-ice forecasts to correct systematic errors that occur in any climate model and gain importance the longer the forecast range. It has been shown that calibration can dramatically improve Arctic sea-ice forecasts (e.g., Krikken et al., 2016). To maintain the added-value of the forecast ensemble, a calibration method suitable for probabilistic forecasts is required. Therefore, we use the Trend-Adjusted Quantile Mapping (TAQM) method (Dirkson et al., 2019) to alleviate the forecast error. The TAQM method computes the probabilities of the historical observations and the historical forecasts after removing their linear trends and then maps the historical forecasted distribution to the historical observed distribution by quantiles. Subsequently, calibrated forecasts can be obtained by applying the reversed observed distribution function. The sea-ice concentration forecasts shown after Section 5.2.1 are all after calibration, unless stated otherwise.

To evaluate the analyzed states and forecasts, we firstly consider the well-known Root Mean Squared Error (RMSE), defined as $\sqrt{\langle(x_f - x_o)^2\rangle}$, where x_f is the analyzed or forecast value and x_o is the observed value of the considered variable (e.g., sea-ice concentration), and $\sqrt{\langle\cdot\rangle}$ denotes an average over multiple analysis or forecast cases. Moreover, in order to gain a more comprehensive and integrative picture of the forecast skill with respect to the ice-edge location, we consider the Spatial Probability Score (SPS, Goessling & Jung, 2018). This metric was designed to measure ensemble-based forecast skill of the sea-ice edge location from a probabilistic viewpoint. The SPS is defined as $\int_\Omega (P[sic > 0.15]_f - P[sic > 0.15]_o)^2 \, d\Omega$, where $P[sic > 0.15]_f$ is the probability of the sea-ice concentration to exceed 15% within the ensemble, $P[sic > 0.15]_o$ is the dichotomous probability of the observed sea-ice concentration to exceed 15%, and Ω is the integration area. The 15% sea-ice concentration contour is the most widely used definition of the ice-edge location.

4 Experimental design

The model and data assimilation system are initialized on January 1st 2002, as described above, but the first year is considered as spin-up and excluded from further analyses. The considered period thus starts on January 1st, 2003 and ends on December 31st, 2019. Restart files at the end of each month have been kept as initial states for free-running forecasts.

To assess the performance of the analysis results, it is useful to compare the analysis error with the error of an unconstrained control experiment (CTRL). However, since a single control simulation contains its own weather-related variability, this independent variability adds to the CTRL error, which can result in an overestimation of the benefit from the data assimilation. We have thus simulated five realizations, with hydrography slightly perturbed in the ocean component at the very beginning, and use the ensemble mean to derive the CTRL error.

Targeting on the subseasonal-to-seasonal time scale, we design four forecast experiments per year. For each year, the system restarts at the beginning of each season, i.e., January 1st, April 1st, July 1st and October 1st, and then continues the forecasts out to one year lead time. Such experiments are carried out from 2003 to 2019. The forecast results from 2011 to 2019 are evaluated in the study. Before 2011, the sea-ice forecasts are employed as the historical forecasts required for the TAQM calibration. Figure 2 illustrates the timing of the forecast experiments from 2010 to 2019. Hereinafter, these experiments are referred to as Jan-init, Apr-init, Jul-init, and Oct-init, starting from these four months of each year.

For each target month, four different forecasts with different lead times are available (Fig. 2). The climatology of Jan-init forecasts is computed by averaging all forecasts over the 2011-2019 period starting from January 1st, and analogously for Apr-init, Jul-init and Oct-init forecasts. Note that for forecasts starting from the year preceding the target year, the forecast climatology averages the forecasts initialized from 2010 to 2018. Taking the forecast climatology in April as an example, we have four forecasts: the first one starts from July of the previous year with lead month 9 (blue arrow), the second one starts from October of the previous year with lead month 6 (light-blue arrow), the third one starts from January of the same year with lead month 3 (light-green arrow), and the last one starts

Table 1. The state vector and observation vector for current AWI-CPS and its predecessor SSIPS v1.0. SIC, SIT, SID, SSH, SLA, SST, and SSS are abbreviations for sea ice concentration, sea ice thickness, sea ice drift, sea surface height, sea-level anomaly, sea surface temperature and sea surface salinity, respectively. T and S represent the temperature and salinity, while T_mix and S_mix specifically stand for the temperature and salinity in the mixed layer.

	SIC	**SIT**	**SID**	**SSH**	**T**	**S**	
State vector	**X**	**X**	**X**	**X**	**X**	**X**	AWI-CPS v2.0
	X	**X**	**X**		T_mix	S_mix	SSIPS v1.0
Observation vector	**X**	**X**	**X**	SLA	SST&T profiles	SSS&S profiles	AWI-CPS v2.0
	X	**X**	**X**		SST		SSIPS v1.0

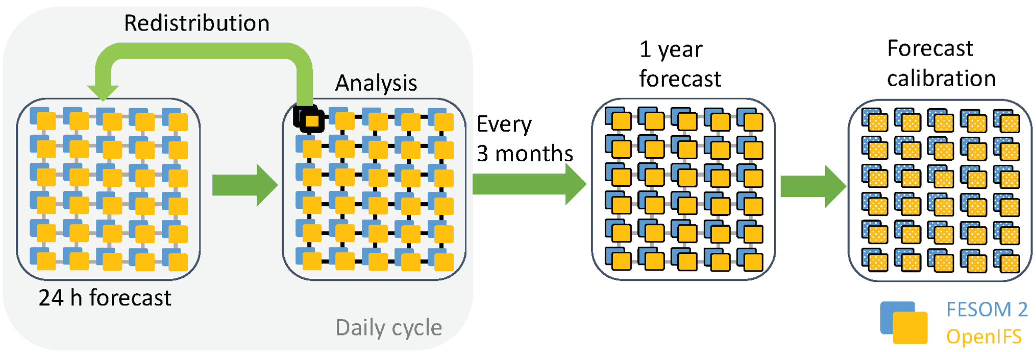

Figure 1. Schematic of the AWI Coupled Prediction System. The ensemble has 30 members with FESOM2 and OpenIFS shown in blue and orange. Communication between each coupled model instance through the Message Passing Interface (MPI) is represented by the staggered mesh in the background, which is emphasized by black color when the communication is active. The coupled model instance with black border in the 'Analysis' chip conducts the analysis and redistributes all the information to each processor. The seamless forecasts start from the analyzed states and run one year ahead. Forecast calibration is applied to correct for drift due to long-term systematic errors.

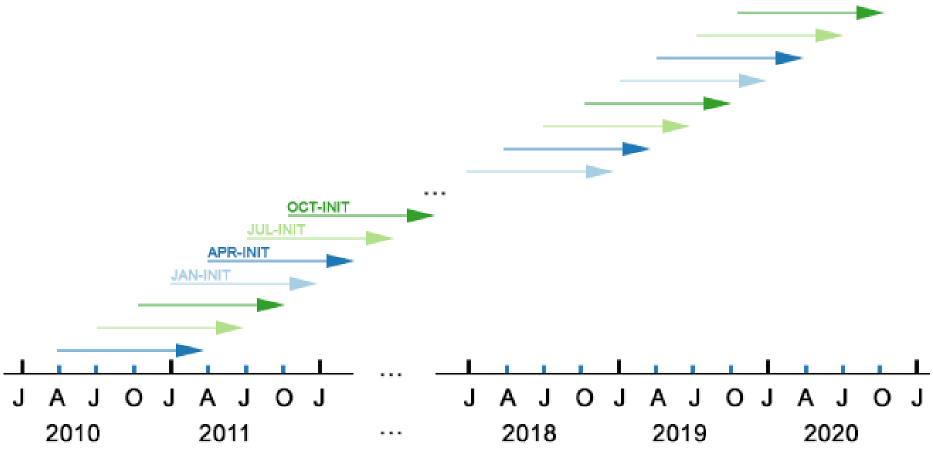

Figure 2. Timing of the forecast experiments. The light-blue arrows indicate forecasts starting from January 1st (Jan-init). The blue, light-green and green arrows indicate forecasts starting from April 1st (Apr-init), July 1st (Jul-init) and October 1st (Oct-init).

Table 2. Lead time in months of the forecasts for each month. L0-2 indicates forecasts with lead times of 0, 1, and 2 months, and so forth for L3-5, L6-8, and L9-11. The months when the forecasts are initialized are marked in bold. Taking January as an example, it consists of four forecasts with lead times of 0, 3, 6, and 9 months that start in January of the same year and October, July and April of the previous year.

	Jan	Feb	Mar	**Apr**	May	Jun	**Jul**	Aug	Sep	**Oct**	Nov	Dec
L0-2	0	1	2	0	1	2	0	1	2	0	1	2
L3-5	3	4	5	3	4	5	3	4	5	3	4	5
L6-8	6	7	8	6	7	8	6	7	8	6	7	8
L9-11	9	10	11	9	10	11	9	10	11	9	10	11

from April of the same year with lead month 0 (green arrow). In such a case, the forecast climatology with lead months 0-2 is computed with output from 2011 to 2019 since they are all started in the current year, while for the other climatologies are computed over the period 2010-2018. For convenience, a table lists the lead time for each month (Table 2).

5 Results

5.1 Analysis state

We firstly examine the performance of the data assimilation system in AWI-CPS with respect to the analyzed state. Figure 3 shows the RMSE of the sea-ice concentration, measured against the OSI SAF satellite observations, in the Arctic and the Antarctic. The free-running CTRL experiment exhibits an area-averaged RMSE around 0.15–0.3 in both polar regions. In the Arctic, particularly high errors occur in September 2007 and 2012 when record-low sea-ice extent was observed which, not surprisingly, the free-running simulations do not capture. Overall, the free model exhibits an overestimation of sea-ice concentration in the melt season and an underestimation in the freezing season in both hemispheres (not shown). The assimilation reduces the sea-ice concentration RMSE by more than 80% for both the Arctic and the Antarctic. Strong error reduction is also found for the sea-ice thickness in the Arctic with respect to the assimilated data (not shown). While not surprising, this error reduction is an important sanity check. Less obviously, the sea-ice velocity is also well constrained by the observations, as in the first version (Mu et al., 2020), despite the small inertia of the sea-ice movement.

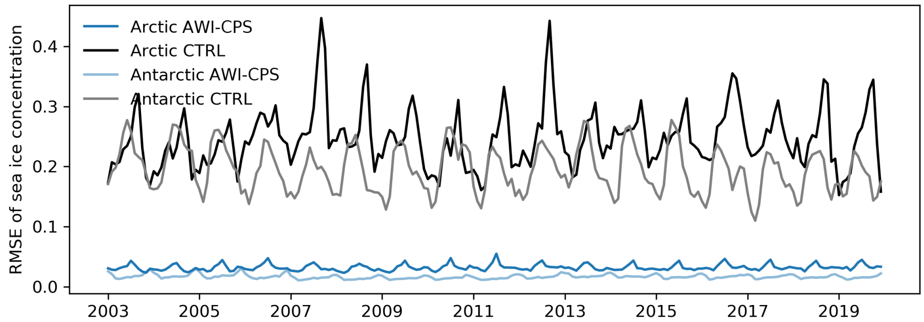

Figure 3. RMSE of monthly mean sea-ice concentration in the Arctic and the Antarctic with respect to the OSI SAF observations. The AWI-CPS analysis is shown in blue, while the CTRL ensemble run without data assimilation is shown in black.

When it comes to the global ocean surface state, the RMSE of the sea-surface temperature and salinity is strongly reduced by the assimilation (Fig. 4a). The systematic errors of sea-surface temperature exhibit remarkable seasonality for both the northern and the southern hemisphere. The average assimilation increment for temperature during the analysis step (Fig. 4b) is negative during the respective summer months in both hemispheres, and positive during winter. This indicates that the system systematically tends toward warmer sea-surface temperature in summer, which the data assimilation needs to work against. This is more pronounced in the northern hemisphere. Larger interannual variability of the RMSE is found for the salinity due to the sparser along-track SMOS sea-surface salinity data. In our system the assimilation generally increases the sea-surface salinity along the tracks, which is in accordance with the on-average positive increments of salinity brought through the cross-covariances from the sea-surface temperature assimilation (not shown). The SLA assimilation using the model-derived MDT field shows pronounced sea level increments in the sub-polar regions specifically the North Atlantic and the Southern Ocean where strong currents prevail due to large sea level gradient (Fig. 4d).

Figure 5 shows the vertical profiles of the RMSE for temperature and salinity in different oceans. The temperature profile assimilation improves the temperature state in the water column below 400 m depth in the Pacific and Indian Ocean, albeit the sparseness of the spatial distribution of the profiles. That is, the deep ocean can be reasonably constrained with a smaller density of profiles than we expected. With only sea-surface temperature assimilation, the improvements are limited to the mixed layer and do not penetrate to the deeper ocean. The simultaneous assimilation of sea-surface temperature and profiles brings these temperature errors down, which has also been reported in Tang et al. (2020). In the Atlantic Ocean and the Arctic Ocean, improvements with respect to temperature are mostly found in the upper ocean. Here, slight RMSE increases are found in the deep ocean. Even larger detrimental effect of the assimilation in AWI-CPS occur in the Southern Ocean, where the temperature RMSE reveals a substantial deterioration of the state compared to the CTRL run. In contrast, the temperature variability in the upper Southern Ocean is very effectively constrained, given that the RMSE is strongly reduced while only small changes are found for the mean state (Fig. 6). When it comes to the vertical salinity profiles, the RMSE is reduced at almost all depths in all ocean basins, besides the Southern Ocean.

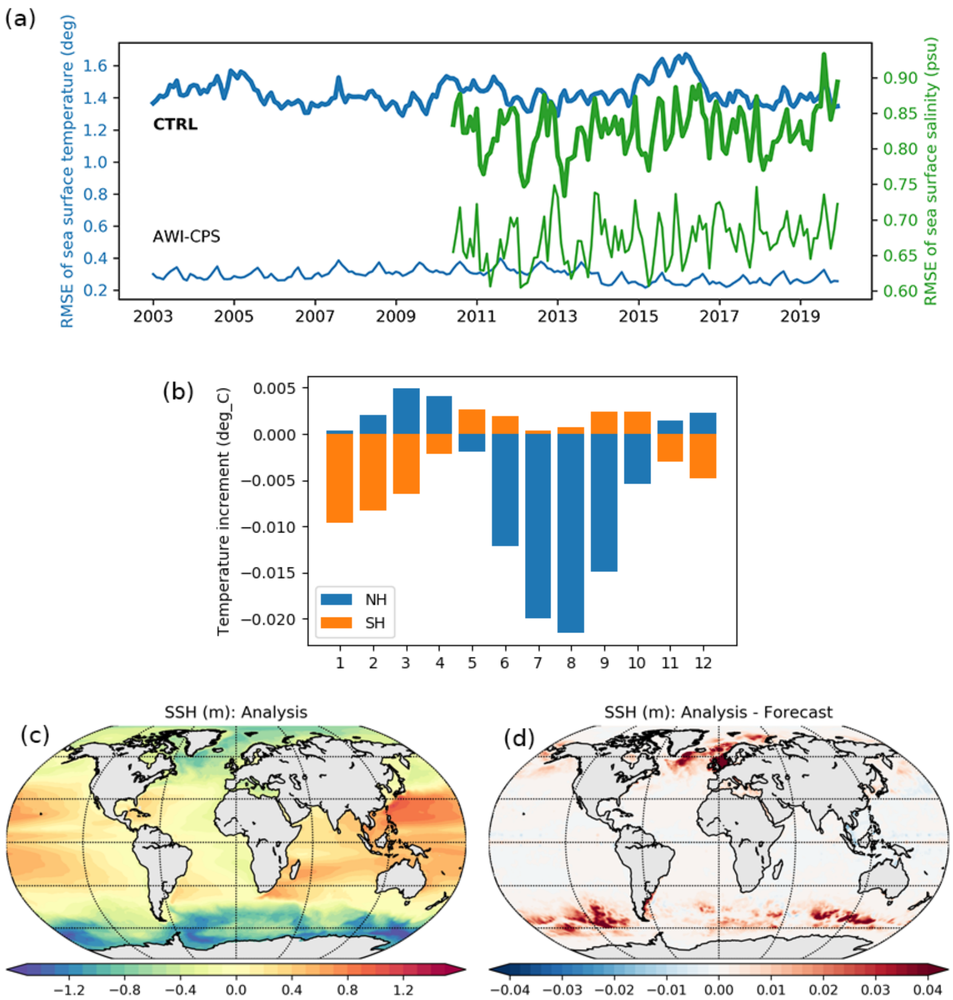

Figure 4. RMSE of monthly mean sea-surface temperature/salinity relative to the OSTIA/SMOS observations (a). AWI-CPS results are shown by thin lines, CTRL results are shown by thick lines. Annual cycle of the mean sea-surface temperature assimilation increments for the two hemispheres (b). Mean sea surface height averaged from 2011-2019 (c) and the corresponding increments (d).

Without salinity assimilation, we have not found such improvements in our previous system (SSIPS v1.0) where only sea-surface temperature and sea-ice observations are assimilated (Table 1).

5.2 Sea-ice forecasts

5.2.1 Forecast spread

The ensemble spread provides a measure for the forecast uncertainty, which we consider here in terms of the ensemble standard deviation. Figure 7 shows the forecast spread of sea-ice and ocean variables. For sea-ice concentration, sea-ice thickness, sea-surface temperature and sea-surface salinity, the state is initially rather well constrained and generally shows a gradually growing spread, superimposed by seasonal variations. The initial spread corresponds to the spread of the assimilation ensemble at the respective initial times. Taking

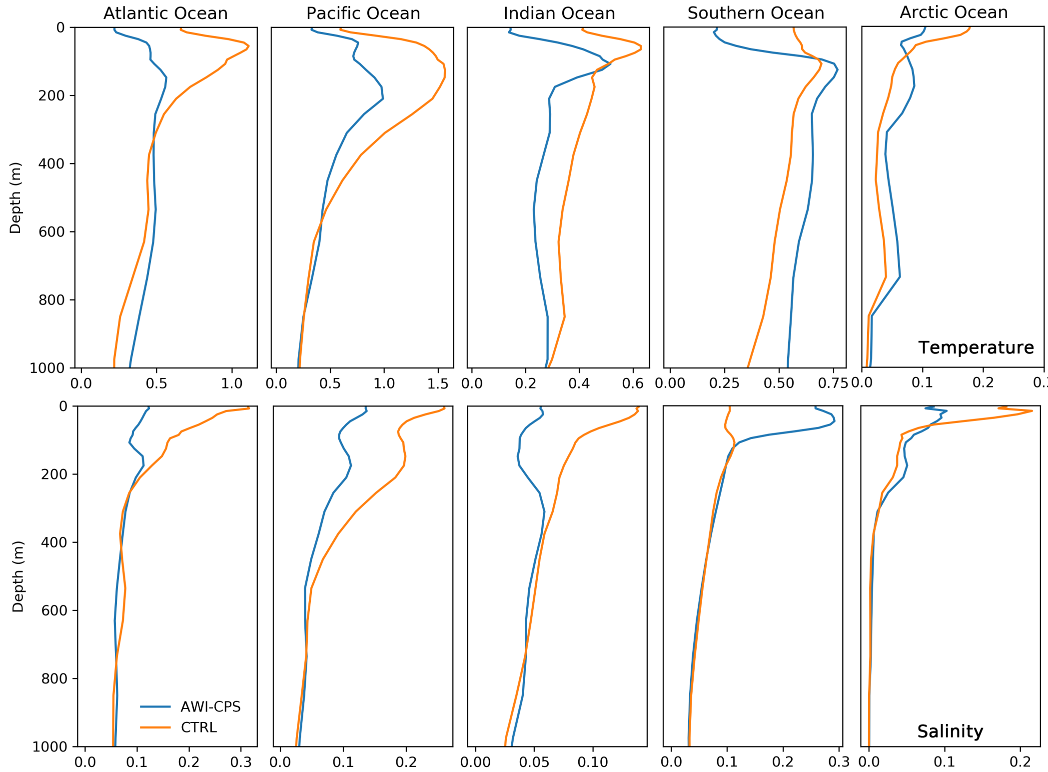

Figure 5. RMSE of temperature (degree Celsius, first row) and salinity (psu, second row) with respect to the World Ocean Atlas (WOA) 2018 data which are constructed by observations over the period 2005-2017 for different ocean basins.

sea-surface temperature as an example, the initial standard deviation (variance) is about 35% (12%) of the final standard deviation (variance) after one year of integration. The spread tends to converge toward a seasonally varying upper envelope that corresponds to the interannual spread of the unconstrained free-running model. Only for sea-surface salinity the spread after one year has not yet converged to a common envelope.

In contrast to the previously mentioned variables, snow thickness is nearly completely unconstrained by the assimilation of the other variables. This is not a surprise as we neither assimilate snow observations, nor update the snow state through the covariances with the other variables because the snow is not included in the state vector. However, we do assimilate sea-ice drift observations, but due to the short inertia of the sea-ice motion this has almost no effect on the sea-ice velocity spread, even on the respective first day. Besides the lack of a direct influence from the assimilation on snow thickness and sea-ice drift, the unaffected spread also reveals that indirect constraints that could in principle be communicated through the atmosphere, e.g. because SST patterns might influence the atmospheric circulation and humidity content, are at most very weak.

For the sea-ice variables, the seasonal cycle of the upper spread envelope generally corresponds to the seasonal cycle of the sea-ice state. The spread of the sea-ice variables shown in Figure 7 is small in the respective freezing season and gets large in summer. The seasonality is also prominent in the sea-surface temperature but not in the sea-surface salinity. The two peaks of the temperature spread per year are the result of the phase shift of the seasonal cycle between the two hemispheres.

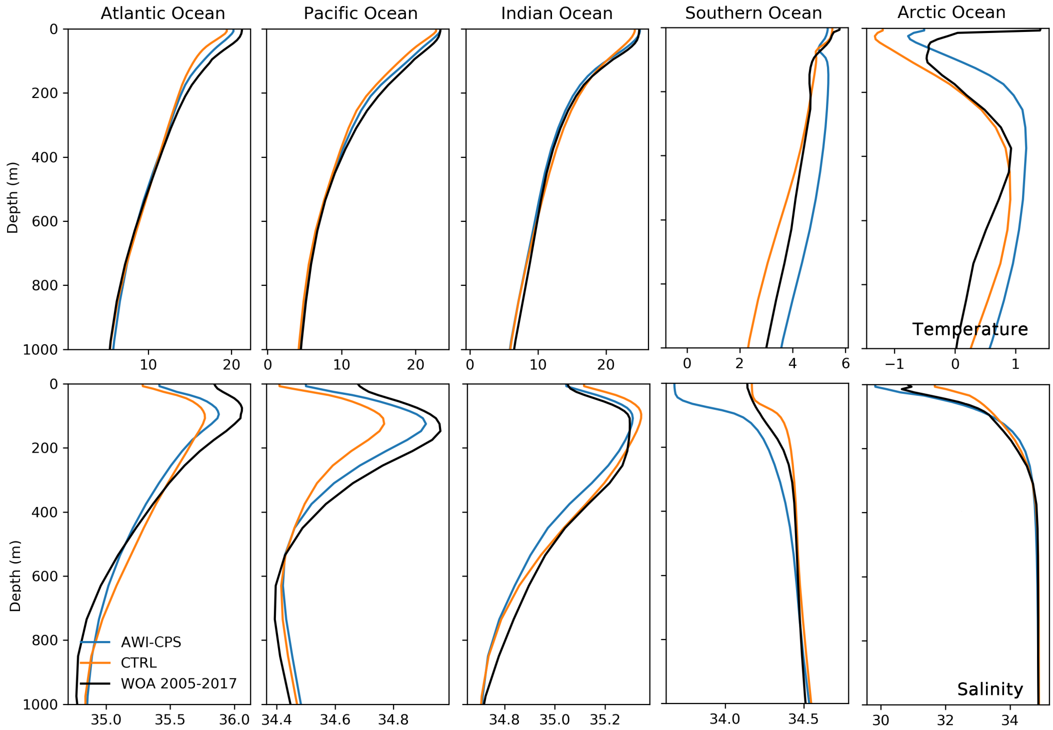

Figure 6. Mean temperature (degree Celsius, first row) and salinity (psu, second row) with respect to the WOA 2018 data (black), constructed by observations over the period 2005-2017 for different ocean basins.

5.2.2 Forecast climatology before and after calibration

Comparing the multi-annual raw forecast climatology and the observed climatology (Fig. 8) of sea-ice concentration reveals a strong positive bias in the Arctic along the marginal ice zones in both March and September already after 2 months. In particular in the East Greenland Sea and the Sea of Okhotsk the ice cover in the raw forecasts is significantly overestimated, resulting in sea-ice concentration climatology differences exceeding 0.5. In September, the regions of dense ice cover in the central Arctic however exhibit too low sea-ice concentrations in the raw forecasts. In contrast, the raw forecasts in the Antarctic tend to underestimate the sea-ice extent almost all around the continent in March, both after 2 and 8 months lead time. In September, when the ice extent here reaches its annual maximum, the bias pattern after 2 months is less uniform, with a tendency toward too loose ice closer to the continent and too dense ice along the sea-ice margins. At 8 months lead time, however, the September bias is predominantly negative around the Antarctic. This suggests that the sea ice might be transported away from the continent too quickly, resulting in a dipole pattern at first and, due to faster melting, a more uniform negative bias later.

The forecast calibration effectively reduces these systematic errors in both hemispheres (Fig. 9), which confirms that most of the trend-adjusted differences are robust between the periods used to determine the calibration (2003-2010) and for the evaluation (2011-2019). After calibration, the average forecast sea-ice concentration however still exceeds the observed values by up to about 0.2. A slight underestimation of the sea-ice concentration is still found in the central Arctic in September. The average errors of sea-ice concentration forecasts in the Antarctic are also strongly reduced, although with a slightly more prominent growth with lead time. Positive bias of the calibrated forecasts in March is found in the

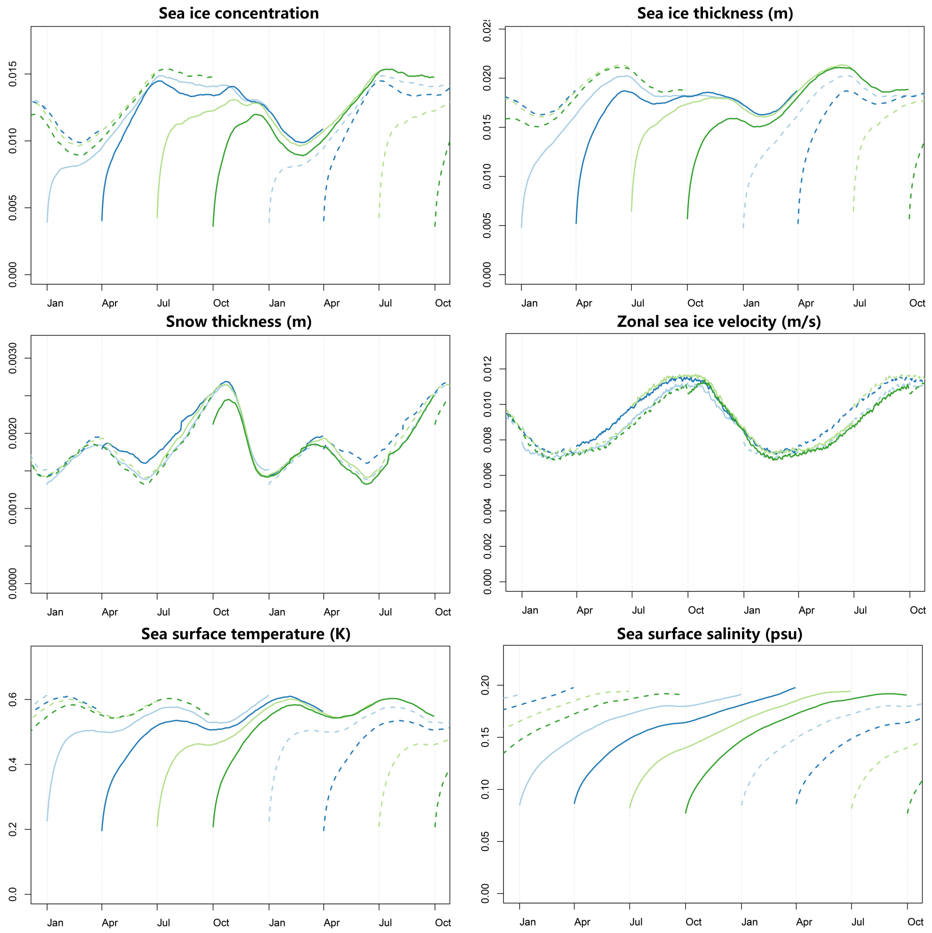

Figure 7. Evolution of the forecast ensemble spread (standard deviation) for ocean and sea-ice variables. Different colors mark the forecasts with different initial times as in Figure 2. All variables are averaged globally, which explains the small numerical values for the sea-ice variables. Note that only the zonal component is shown for the sea-ice velocity.

Ross Sea and north of the Weddell Sea, and a negative bias is present in the interior Weddell Sea.

Climate models generally have difficulties in simulating Antarctic sea ice (Shu et al., 2020; Rackow et al., 2022) due to the complex ocean-atmosphere-ice interactions in the Southern Ocean. In particular, the well-documented sea-surface temperature bias in the Southern Ocean observed nearly in all the climate models deteriorates the sea-ice simulation in the models. This also holds for AWI-CM3, where these discrepancies lead to an non-uniform error distribution along the ice edge in September that persists also after calibration.

5.2.3 Calibrated forecast skill

The RMSE of the calibrated daily sea-ice concentration forecast against the OSI SAF observations for the Arctic and the Antarctic, arranged in groups corresponding to different lead-time ranges, is shown in Figure 10. As a reference, the OSI SAF climatology forecast for each target year is computed simply using the average of the observed values in the previous 9 years. For example, to get the climatology forecast in January 2018, we average the January observations from 2009 to 2017. The climatology forecast is normally employed

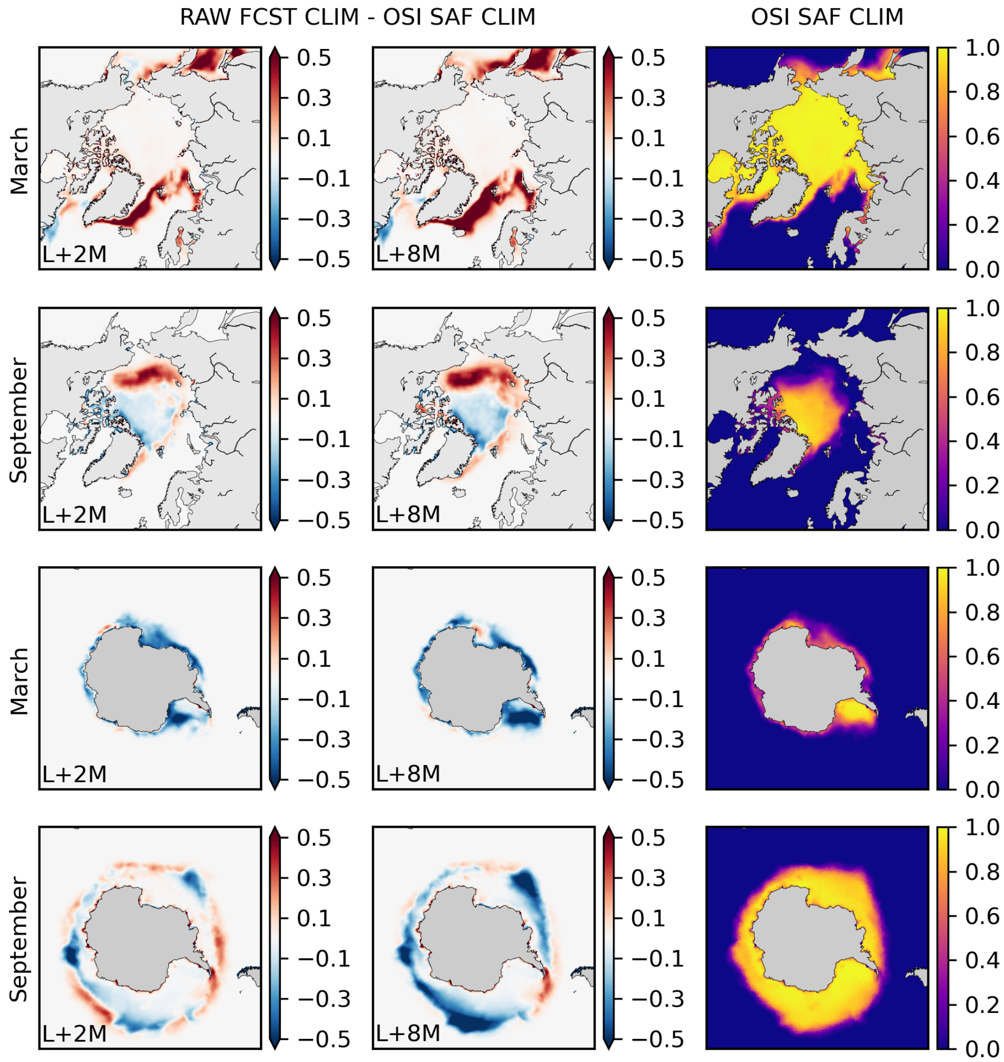

Figure 8. Climatological difference of the sea-ice concentration raw forecasts and OSI SAF satellite observations in the Arctic (first and second row) and Antarctic (third and fourth row) in March (first and third row) and September (second and fourth row), with two (left column) and eight (middle column) months lead time, averaged over the period 2011-2019. The right column shows the corresponding observed climatologies.

as a benchmark to assess when a forecast system completely loses its skill (Goessling et al., 2016; Goessling & Jung, 2018; Zampieri et al., 2018). The comparison suggests that in the Arctic, AWI-CPS forecasts averaged over lead times of 0-2 months (L0-2 in the figure, less than 90 days) have an RMSE of 0.167, thus beating the climatology forecast with a RMSE of 0.172 only slightly. Indeed, the time series reveals that the RMSE is substantially lower than the climatological error only during approximately the first half of each first season, implying that most of the L0-2 skill derives only from the first weeks. AWI-CPS forecasts with longer lead time tend to be worse than the climatology forecast, despite the forecast calibration. In the Antarctic, AWI-CPS forecast skill averaged over L0-2 does not exceed the skill of the climatology forecast (RMSE = 0.221 > 0.208). Here, the stronger biases result in errors at longer lead times exceeding the climatological error by an even larger margin.

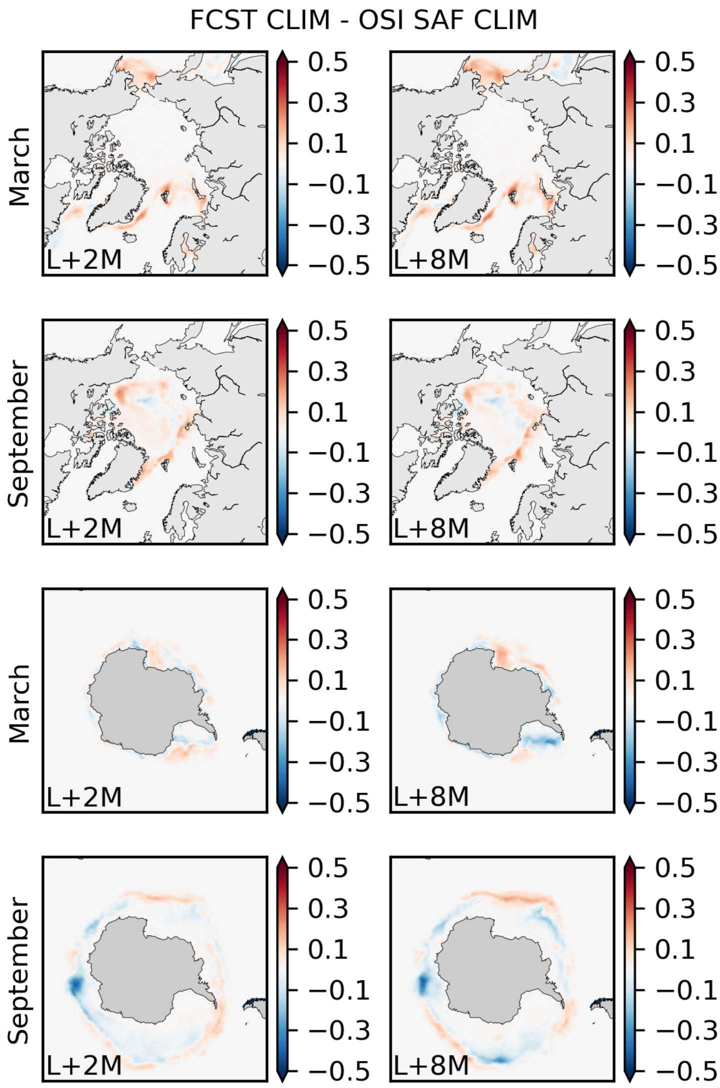

Figure 9. Same as Fig. 8, but after calibration and without the observed climatologies.

The annual cycle of the SPS for the ice-edge location of the OSI SAF climatology forecast is shown in the upper panel of Figure 11. The SPS exhibits strong seasonal variability. The lowest SPS is observed in late summer (March) in the Antarctic, corresponding to the shortest ice-edge length. In the Arctic, the lowest SPS is found in late autumn and winter (November, December, January), again corresponding to the time of the year with the shortest ice edge (Goessling & Jung, 2018; Zampieri et al., 2019). The SPS in the Antarctic is up to twice as large as that in the Arctic, largely due to the longer ice edge. In the following we therefore always compare the SPS of the forecast system to the corresponding climatological error.

The SPS for the different initial months, averaged over the evaluation period (Fig. 11), is largely consistent with the results obtained for the RMSE of the sea-ice concentration. The AWI-CPS forecasts outperform the climatological forecast during the first month (labeled L0 above) but perform worse than climatology from the third month (labeled L2 above) onward; whether or not the forecast outperforms climatology during the second month depends on the hemisphere and the time of the year.

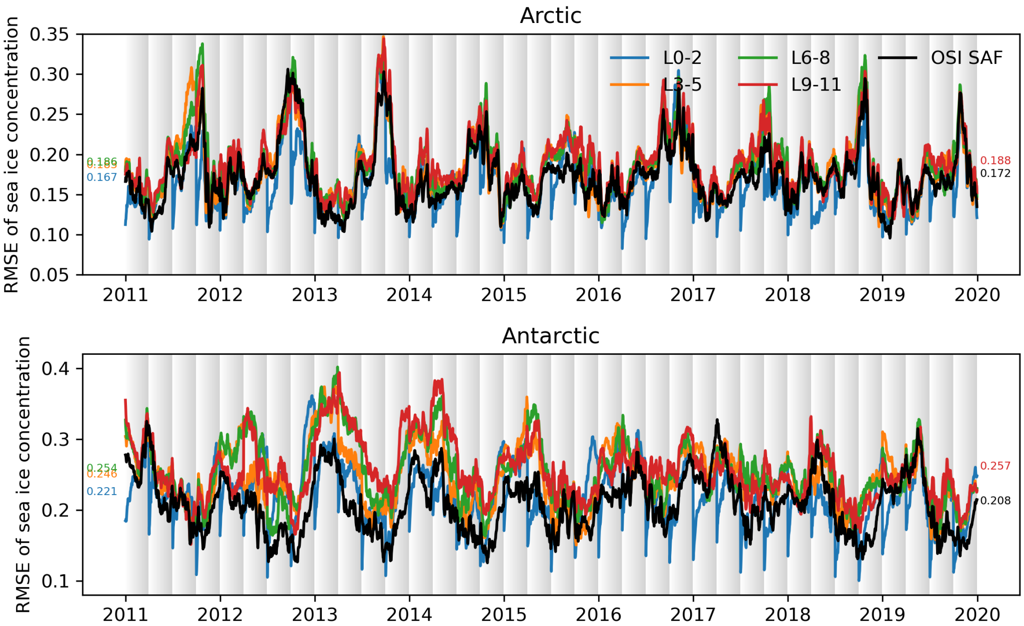

Figure 10. RMSE of the sea-ice concentration (0∼1.0) forecasts with respect to the observations in the Arctic (top) and the Antarctic (bottom). L0-2, L3-5, L6-8, and L9-11 represent forecasts with lead time of 0-2 months, 3-5 months, 6-8 months, and 9-11 months, respectively. Note that, for example, forecasts in January, February, and March on line L0-2 are all initialized on January 1st and thus have lead months of 0, 1, and 2. The lead time for the very first month of each forecast is labelled "0 months". Stronger shading in the background indicates longer lead time. The OSI SAF climatology forecast, derived from the preceding 9 years for each year individually, is shown in black. The mean RMSEs of each time series, grouped by lead time, are annotated in the corresponding color and located vertically at their value, to the left and right of the time series.

The forecasts tend to perform better in the Arctic, where in particular those initialized in April remain skillful also during the second month. It appears that Arctic sea-ice conditions in May are generally better forecast compared to other months in our system, as the SPS difference from climatology exhibits a temporal local minimum in May, independent of lead time. This minimum is very pronounced in the Arctic forecasts initialized in January; in this case the skill gain in May can be attributed partly to a phenomenon termed re-emergence of skill that occurs when the marginal ice zone returns to the same regions where it was at the time of initialization over the course of the seasonal cycle (Blanchard-Wrigglesworth et al., 2011; Goessling et al., 2016; Blanchard-Wrigglesworth et al., 2017). Re-emergence of skill is even more prominent in November for the Arctic forecasts initialized in July. Finally, the Arctic forecasts initialized in April, although maintaining skill longest initially, exhibit a peculiarity around the following September, when the Arctic sea-ice minimum occurs: Here they perform even worse than those forecasts initialized already three months earlier, in January. This is likely related to the seasonal cycle of model biases.

In the Antarctic, the ice-edge forecasts initialized around the sea-ice minimum, in January and April, exhibit reasonable skill during the first month (labeled L0 above) and neutral skill relative to climatology during the second month (labeled L1 above). In contrast, the forecasts initialized around the sea-ice maximum, in July and October, are only marginally skillful already during the first month and worse than climatology during the second. The negative skill relative to climatology is generally most severe around December, which can

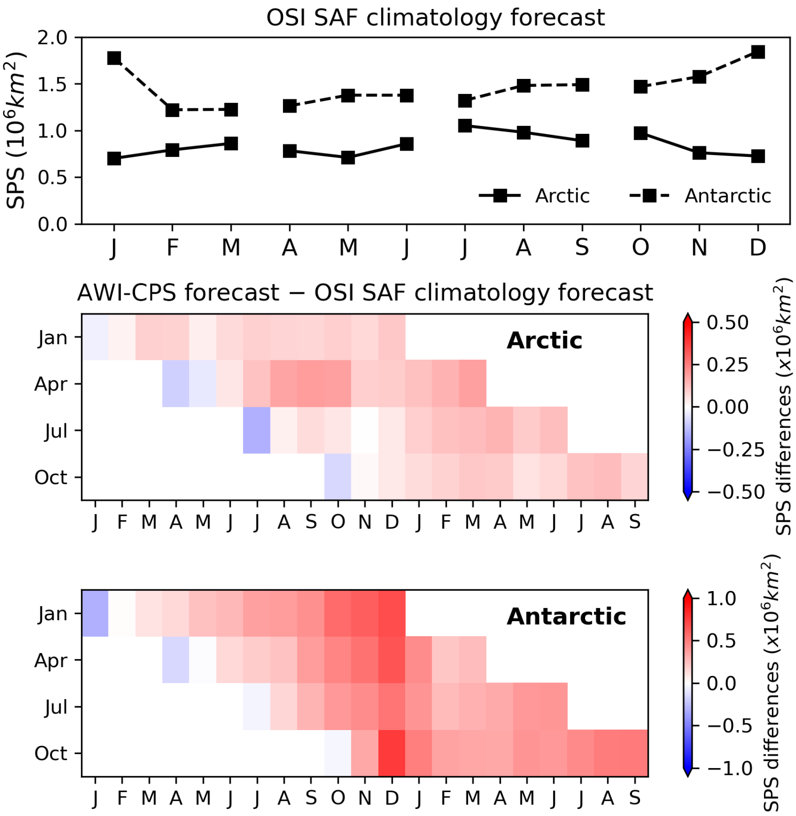

Figure 11. Monthly Spatial Probability Score (SPS) of sea-ice forecasts averaged over 2011-2019, relative to OSI SAF observations. The seasonal cycle of the absolute SPS of the OSI SAF climatology forecasts is shown in the upper panel. SPS differences between AWI-CPS forecasts and the OSI SAF climatology forecasts are shown in the middle and bottom panels.

be partly explained by the corresponding seasonal cycle of the ice-edge length (see above). Here we again observe a case where the forecasts initialized earlier, in July, perform less bad than those initialized later, in October, again presumably due to the seasonality of model biases.

The low forecast skill beyond two months lead time, remaining even after calibration, exposes plainly a gap in forecast skill compared to the potential predictability found in perfect-model studies (Tietsche et al., 2014; Day et al., 2016; Goessling et al., 2016). This is however not specific to our forecast system, but has also been reported for current operational subseasonal-to-seasonal forecast systems (Zampieri et al., 2018, 2019). In the following we thus consider forecast skill during the first three months more closely, at daily resolution, for both the raw and the calibrated forecasts, and relative not only to climatology but also to a benchmark based on persistence of the observed initial state.

Averaged over all initial times across seasons and measured by the SPS for the ice-edge location, the raw forecasts outperform climatology only for about 12 days (Fig. 12). The calibrated forecasts beat climatology out to 45 days; however, beyond two weeks the advantage relative to climatology is marginal. At short lead times up to about 6 days, the calibration in fact deteriorates the forecast skill. What is more, for up to about 8 days, both the raw and calibrated forecasts are outperformed by persistence of the observed initial state. The superiority of the initial-state persistence may be partly due to the fact that

OSI SAF observations are essentially evaluated against the same data set, albeit at different times, but still this implies that the period where the (calibrated) forecasts are better than any of the two benchmarks by more than a small margin is limited to a short time window between about 8 and 14 days.

In the Antarctic, the raw forecasts outperform the climatological benchmark for about 19 days and thus slightly longer than in the Arctic. In contrast, the calibration delays the lead time at which the climatological SPS is surpassed only by about one week. Moreover, the deterioration brought about by the calibration at short lead times is even more severe than in the Arctic, so that the calibration becomes on average beneficial only after about 14 days, just a few days before the calibrated forecast error surpasses the climatological error. On the other hand, the raw forecasts outperform the initial-state persistence already after 4 days, implying that the period of forecast skill beyond both benchmarks is broader in the Antarctic compared to the Arctic. Consistent with the monthly results based on RMSE and SPS discussed above, the Antarctic calibrated forecasts however develop even a large skill gap relative to climatology at long lead times beyond 1-2 months.

As has become apparent already from the monthly analysis, the ice-edge forecast skill is initial-date dependent (Fig. 13). The raw forecasts outperform the climatology forecast for a longer lead time when initialized in April and October in the Arctic; the same holds in January and April in the Antarctic. This seasonal contrast is partly transferred also to the calibrated forecasts. For both hemispheres, the calibrated forecasts initialized in April show the longest lead time during which they outperform climatology – more than 45 days – although still only marginally beyond about two weeks. Moreover, there are considerable seasonal variations of the effect of the calibration, in particular in the Arctic at short lead times. In January and July the calibration barely deteriorates the Arctic ice-edge forecast skill at all, but significantly improves the skill already after 1–3 days. In July this results in an extended time window during which the calibrated forecasts outperform the two benchmarks, whereas this is not the case for the January forecasts which exhibit too large errors already at initial time.

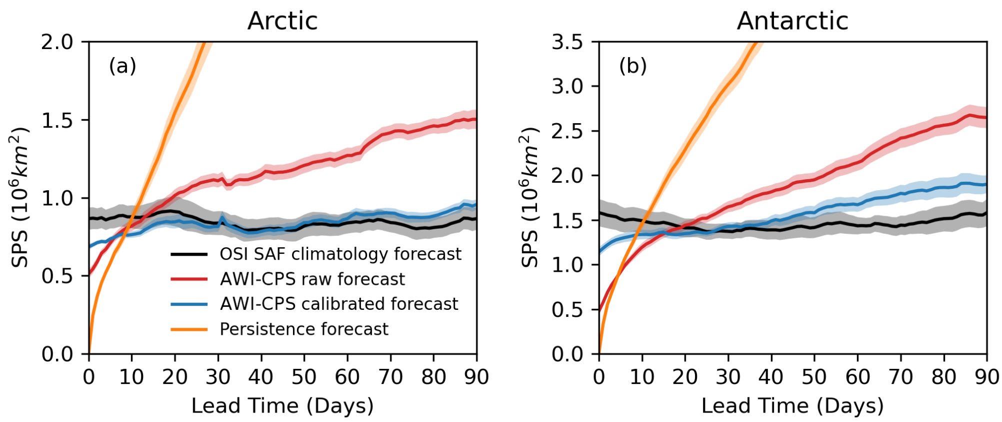

Figure 12. Daily Spatial Probability Score (SPS) of AWI-CPS forecasts as a function of lead time. The errors of forecasts with the same lead time (see Table 2) have been averaged over the period 2011–2019 and all initial seasons. The shading area indicates their standard errors (approximate 67% confidence intervals) over that period. The OSI SAF climatology forecast (OSI SAF CLIM, black) and the persistence forecast (PER), which keeps the observed state on the initial day, serve as benchmarks. AWI-CPS forecasts are evaluated before (AWI-CPS RAW) and after (AWI-CPS CALI) calibration with the TAQM method.

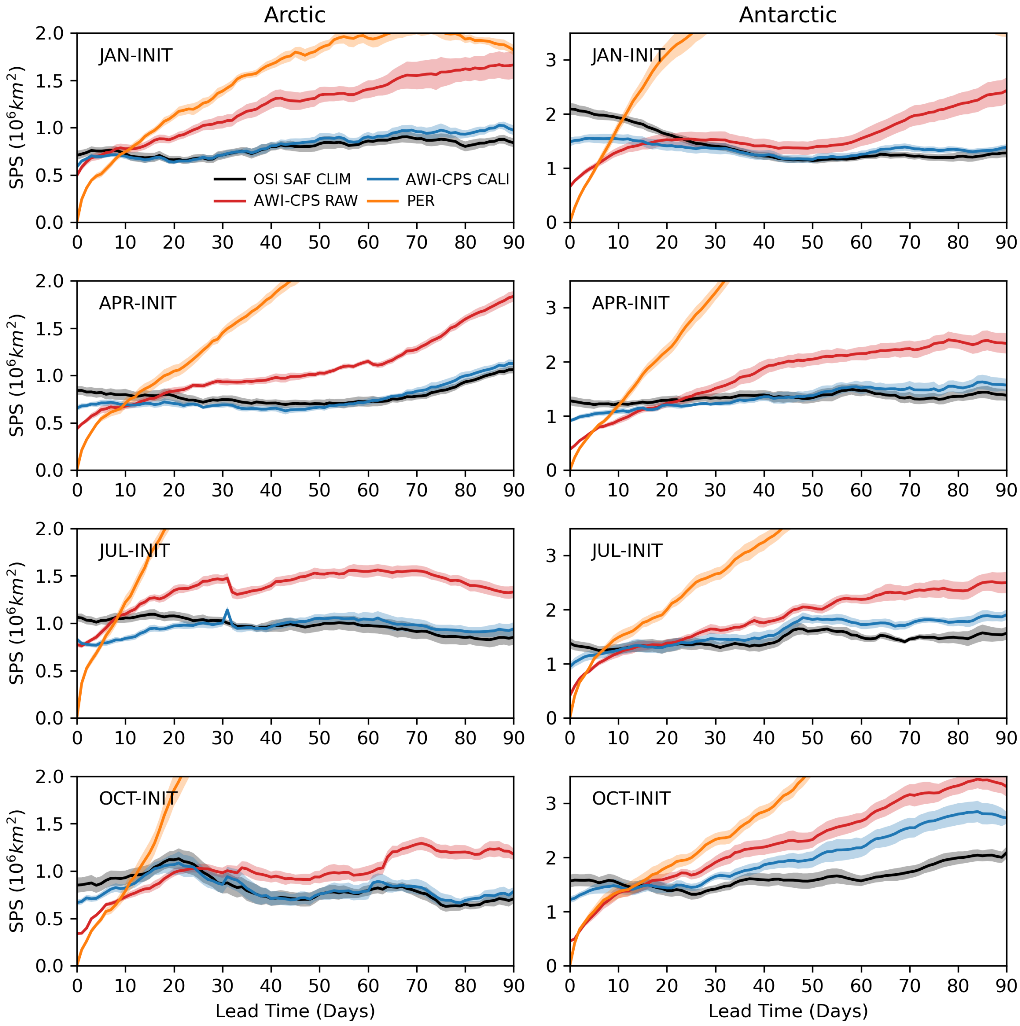

Figure 13. Same as Figure 12 but separately for the different initial seasons.

6 Summary and Discussion

A new version of the AWI Coupled Prediction System based on the AWI Climate Model (AWI-CM3.0) with ocean and sea-ice data assimilation has been developed. The main upgrades compared to the old version are: 1. the atmosphere model has been replaced by OpenIFS, with finer spectral resolution, and the ocean model now uses the finite-volume method; 2. the speed-up of both the ocean and the atmosphere model enables us to have an ensemble size more than two times larger than before while using less computing resources; 3. the online data assimilation now includes essentially all relevant large-scale observations for the ocean and sea ice. Moreover, we now use more sophisticated post-processing, namely: 4. sea-ice calibration by Trend-Adjusted Quantile Mapping (TAQM) is now applied to account for systematic errors for long-term prediction.

The assimilation results are evaluated by the metric of RMSE. We observe strong error reductions and well constrained ensemble spread after data assimilation. The seasonal cycle of mean surface temperature increments points to model discrepancies related to surface flux errors, in particular in summer. The assimilation of temperature and salinity profiles tends to improve the ocean state even in deep levels where no observations are assimilated.

However, complex processes within the ocean, atmosphere and sea-ice system in the Arctic Ocean and Southern Ocean may introduce spurious covariances. Here, further relaxation of the Gaussian error distribution assumption or a smaller localization radius are options that should be explored.

Forecast experiments over the period 2003–2019, of which we use the first eight years to derive the calibration parameters, show an overestimation of sea-ice concentration in the marginal ice zone in the Arctic in both March and September. The errors in the Antarctic are spatially less uniform, possibly due to different processes involved over different parts of the Southern Ocean. Calibrated sea-ice concentration and ice-edge forecasts in AWI-CPS outperform a climatological benchmark for about 45 days in the Arctic and about 30 days in the Antarctic, albeit with seasonal variations. The sea-ice forecast calibration performs better in the Arctic than in the Antarctic, which is possibly related to the different basin geometries, specifically the semi-closed basin in the Arctic. Our evaluation reveals that the calibration considerably deteriorates the forecasts at short lead times, implying that the raw forecast without calibration should be trusted over the calibrated one during the first week after initialization in the Arctic, and during the first two weeks in the Antarctic. This unwanted side effect is possibly related to the rather short time period used to derive the calibration parameters. It also calls for an advanced application of the calibration that gradually takes effect in a more seamless way.

In our forecast system the atmosphere is still evolving without any constraint, except for the influence of the constrained ocean and sea-ice surface states. Efforts are currently ongoing to explore how sea-ice forecasts and climate forecasts more generally can be improved by also constraining the atmosphere directly. While improvements at short, weather-related lead times are to be expected rather obviously, it is not clear to what extent such improvement can also sustain for longer times where model biases become the dominant matter of concern. Another aspect coming into play when the atmosphere is simultaneously constrained is how this affects the oceanic data assimilation, specifically the spread of the model background. Replacing the currently largely random atmospheric weather states in the different ensemble members by more coherent states might necessitate to re-introduce a forgetting factor to prevent the ocean model spread from collapsing. At the same time, the more realistic atmospheric states are expected to help drive ocean and sea-ice anomalies, reducing the corrections to be introduced by the data assimilation. These aspects should be explored in future studies.

The skill of the raw forecasts from our system is largely comparable to that from operational subseasonal-to-seasonal (S2S) forecast systems, even though the atmosphere models in these S2S systems are generally initialized by atmospheric data assimilation. This strongly indicates that constraining the atmosphere alone will not be sufficient to achieve major forecast performance gains at longer lead times. Rather, the correction of systematic errors will be critical, in general and in our forecast system specifically. The climate model used in our system, AWI-CM3, is a rather new combination of its model components and is still undergoing major tuning and even more fundamental developments at the moment. The co-development of our forecast system – the AWI Coupled Prediction System – is a major opportunity to help inform the development of the underlying climate model and, vice versa, to benefit from these developments.

Acknowledgments

This study is supported by the Federal Ministry of Education and Research of Germany in the framework of SSIP (grant01LN1701A) and the National Natural Science Foundation of China (42176235). The simulations were performed at the German Climate Computing Centre (DKRZ) and the North-German Supercomputing Alliance (HLRN). The source code for AWI-CM3 is available at `https://github.com/FESOM/fesom2/releases/tag/AWI-CM3_v3.0`. FESOM2 can be obtained from `https://github.com/FESOM/fesom2`. OpenIFS is maintained by the European Centre for Medium-Range Weather Forecasts and could be

obtained from `https://www.ecmwf.int/en/research/projects/openifs` with a license. The coupler used is the OASIS3-MCT4, which is available at `https://portal.enes.org/oasis`. PDAF is available at `http://pdaf.awi.de`. The experiment output data and model mesh are archived on Zenodo (`https://doi.org/10.5281/zenodo.6481116`). LM developed the system, conducted the experiments and drafted the paper. HG initiated and leads the development of the system. LN is the main developer of PDAF and contributed to data assimilation aspects specific to our forecast system. JS is the main developer of AWI-CM3 and provided technical support. BN helped with the sea-ice forecast calibration. QT provided support with the T/S profiles assimilation. LZ gave supports on the coupling details in the AWI-CM3. SL contributed to the discussion about the framework of the system. All authors discussed the system development together and contributed to the manuscript. The authors declare that there is no conflict of interest regarding the publication of this article. Thanks to the technical help from Patrick Scholz, Dmitry Sidorenko, and the supports from Arlan Dirkson on the TAQM calibration method.

References

Blanchard-Wrigglesworth, E., Armour, K. C., Bitz, C. M., & DeWeaver, E. (2011). Persistence and inherent predictability of arctic sea ice in a gcm ensemble and observations. *Journal of Climate*, *24*(1), 231–250.

Blanchard-Wrigglesworth, E., Barthélemy, A., Chevallier, M., Cullather, R., Fučkar, N., Massonnet, F., ... others (2017). Multi-model seasonal forecast of arctic sea-ice: forecast uncertainty at pan-arctic and regional scales. *Climate Dynamics*, *49*(4), 1399–1410.

Blockley, E. W., Peterson, K. A., & Blockley, E. (2018). Improving Met Office seasonal predictions of Arctic sea ice using assimilation of CryoSat-2 thickness. *The Cryosphere*, *12*, 3419–3438. doi: 10.5194/tc-12-3419-2018

Bouillon, S., Fichefet, T., Legat, V., & Madec, G. (2013). The elastic–viscous–plastic method revisited. *Ocean Modelling*, *71*, 2–12.

Bushuk, M., Msadek, R., Winton, M., Vecchi, G. A., Gudgel, R., Rosati, A., & Yang, X. (2017). Skillful regional prediction of arctic sea ice on seasonal timescales. *Geophysical Research Letters*, *44*(10), 4953–4964.

Collow, T. W., Wang, W., Kumar, A., & Zhang, J. (2015). Improving arctic sea ice prediction using piomas initial sea ice thickness in a coupled ocean–atmosphere model. *Monthly Weather Review*, *143*(11), 4618–4630.

Craig, A., Valcke, S., & Coquart, L. (2017). Development and performance of a new version of the oasis coupler, oasis3-mct_3. 0. *Geoscientific Model Development*, *10*(9).

Danilov, S., Sidorenko, D., Wang, Q., & Jung, T. (2017). The finite-volume sea ice–ocean model (fesom2). *Geosci. Model Dev.*, *10*, 765–789.

Danilov, S., Wang, Q., Timmermann, R., Iakovlev, N., Sidorenko, D., Kimmritz, M., ... Schröter, J. (2015). Finite-element sea ice model (fesim), version 2. *Geoscientific Model Development*, *8*(6), 1747–1761.

Day, J., Tietsche, S., Collins, M., Goessling, H., Guemas, V., Guilliory, A., ... others (2016). The arctic predictability and prediction on seasonal-to-interannual timescales (apposite) data set version 1. *Geoscientific Model Development*, *9*(6), 2255–2270.

Director, H. M., Raftery, A. E., & Bitz, C. M. (2017). Improved sea ice forecasting through spatiotemporal bias correction. *Journal of Climate*, *30*(23), 9493–9510.

Dirkson, A., Merryfield, W. J., & Monahan, A. H. (2019). Calibrated probabilistic forecasts of arctic sea ice concentration. *Journal of Climate*, *32*(4), 1251–1271.

Goessling, H., & Jung, T. (2018). A probabilistic verification score for contours: Methodology and application to arctic ice-edge forecasts. *Quarterly Journal of the Royal Meteorological Society*, *144*(712), 735–743.

Goessling, H., Tietsche, S., Day, J. J., Hawkins, E., & Jung, T. (2016). Predictability of the arctic sea ice edge. *Geophysical Research Letters*, *43*(4), 1642–1650.

Jung, T., Gordon, N. D., Bauer, P., Bromwich, D. H., Chevallier, M., Day, J. J., ... others (2016). Advancing polar prediction capabilities on daily to seasonal time scales. *Bulletin of the American Meteorological Society*, *97*(9), 1631–1647.

Kriegler, E., Bauer, N., Popp, A., Humpenöder, F., Leimbach, M., Strefler, J., ... others (2017). Fossil-fueled development (ssp5): an energy and resource intensive scenario for the 21st century. *Global environmental change*, *42*, 297–315.

Krikken, F., Schmeits, M., Vlot, W., Guemas, V., & Hazeleger, W. (2016). Skill improvement of dynamical seasonal arctic sea ice forecasts. *Geophysical Research Letters*, *43*(10), 5124–5132.

Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., ... others (2019). Version 2 of the eumetsat osi saf and esa cci sea-ice concentration climate data records. *The Cryosphere*, *13*(1), 49–78.

Liu, J., Chen, Z., Hu, Y., Zhang, Y., Ding, Y., Cheng, X., ... others (2019). Towards reliable arctic sea ice prediction using multivariate data assimilation. *Science Bulletin*, *64*(1), 63–72.

Mu, L., Liang, X., Yang, Q., Liu, J., & Zheng, F. (2019). Arctic ice ocean prediction system: evaluating sea-ice forecasts during xuelong's first trans-arctic passage in summer 2017. *Journal of Glaciology*, *65*(253), 813–821.

Mu, L., Nerger, L., Tang, Q., Loza, S. N., Sidorenko, D., Wang, Q., ... Goessling, H. F. (2020). Toward a data assimilation system for seamless sea ice prediction based on the awi climate model. *Journal of Advances in Modeling Earth Systems*, *12*(4), e2019MS001937.

Nerger, L., & Hiller, W. (2013). Software for ensemble-based data assimilation systems – implementation strategies and scalability. *Comput. Geosci.*, *55*, 110–118. doi: 10.1016/j.cageo.2012.03.026

Nerger, L., Janjić, T., Schröter, J., & Hiller, W. (2012). A unification of ensemble square root kalman filters. *Monthly Weather Review*, *140*(7), 2335-2345.

Nerger, L., Tang, Q., & Mu, L. (2019). Efficient ensemble data assimilation for coupled models with the parallel data assimilation framework: Example of awi-cm. *Geoscientific Model Development Discussions*, *2019*, 1–23. doi: 10.5194/gmd-2019-167

Owens, W. B., & Lemke, P. (1990). Sensitivity studies with a sea ice-mixed layer-pycnocline model in the weddell sea. *Journal of Geophysical Research: Oceans*, *95*(C6), 9527–9538.

Paul, S., Hendricks, S., Ricker, R., Kern, S., & Rinne, E. (2018). Empirical parametrization of envisat freeboard retrieval of arctic and antarctic sea ice based on cryosat-2: progress in the esa climate change initiative. *The Cryosphere*, *12*(7), 2437–2460. doi: 10.5194/tc-12-2437-2018

Petty, A. A., Kurtz, N. T., Kwok, R., Markus, T., & Neumann, T. A. (2020). Winter arctic sea ice thickness from icesat-2 freeboards. *Journal of Geophysical Research: Oceans*, *125*(5), e2019JC015764.

Pham, D. T. (2001). Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly weather review*, *129*(5), 1194–1207.

Rackow, T., Danilov, S., Goessling, H. F., Hellmer, H. H., Sein, D. V., Semmler, T., ... Jung, T. (2022). Delayed antarctic sea-ice decline in high-resolution climate change simulations. *Nature communications*, *13*(1), 1–12.

Ricker, R., Hendricks, S., Helm, V., Skourup, H., & Davidson, M. (2014). Sensitivity of CryoSat-2 Arctic sea ice freeboard and thickness on radar-waveform interpretation. *The Cryosphere*, *8*(4), 1607–1622. doi: 10.5194/tc-8-1607-2014

Ricker, R., Hendricks, S., Kaleschke, L., Tian-Kunze, X., King, J., & Haas, C. (2017). A weekly Arctic sea-ice thickness data record from merged CryoSat-2 and SMOS satellite data. *The Cryosphere*, *11*, 1607–1623. doi: 10.5194/tc-11-1607-2017

Sakov, P., Counillon, F., Bertino, L., Lisæter, K., Oke, P., & Korablev, A. (2012). Topaz4: an ocean-sea ice data assimilation system for the north atlantic and arctic. *Ocean Science Discussions*, *9*(2).

Shu, Q., Wang, Q., Song, Z., Qiao, F., Zhao, J., Chu, M., & Li, X. (2020). Assessment of sea ice extent in cmip6 with comparison to observations and cmip5. *Geophysical*

Research Letters, *47*(9), e2020GL087965.
Skachko, S., Buehner, M., Laroche, S., Lapalme, E., Smith, G., Roy, F., ... Garand, L. (2019). Weakly coupled atmosphere–ocean data assimilation in the canadian global prediction system (v1). *Geoscientific Model Development*, *12*(12), 5097–5112.
Streffing, J., Sidorenko, D., Semmler, T., Zampieri, L., Scholz, P., Andrés-Martínez, M., ... Jung, T. (2022). Awi-cm3 coupled climate model: Description and evaluation experiments for a prototype post-cmip6 model. *EGUsphere Discussions*, *2022*, 1–37. Retrieved from `https://egusphere.copernicus.org/preprints/egusphere-2022-32/` doi: 10.5194/egusphere-2022-32
Tang, Q., Mu, L., Sidorenko, D., Goessling, H., Semmler, T., & Nerger, L. (2020). Improving the ocean and atmosphere in a coupled ocean–atmosphere model by assimilating satellite sea-surface temperature and subsurface profile data. *Quarterly Journal of the Royal Meteorological Society*, *146*(733), 4014–4029.
Tian-Kunze, X., Kaleschke, L., Maaß, N., Mäkynen, M., Serra, N., Drusch, M., & Krumpen, T. (2014). SMOS-derived thin sea ice thickness: Algorithm baseline, product specifications and initial verification. *The Cryosphere*, *8*(3), 997–1018. doi: 10.5194/tc-8-997-2014
Tietsche, S., Day, J. J., Guemas, V., Hurlin, W., Keeley, S., Matei, D., ... Hawkins, E. (2014). Seasonal to interannual arctic sea ice predictability in current global climate models. *Geophysical Research Letters*, *41*(3), 1035–1043.
Xie, J., Raj, R. P., Bertino, L., Samuelsen, A., & Wakamatsu, T. (2019). Evaluation of arctic ocean surface salinities from the soil moisture and ocean salinity (smos) mission against a regional reanalysis and in situ data. *Ocean Science*, *15*(5).
Xie, J., & Zhu, J. (2010). Ensemble optimal interpolation schemes for assimilating argo profiles into a hybrid coordinate ocean model. *Ocean Modelling*, *33*(3-4), 283–298.
Yang, Q., Losa, S. N., Losch, M., Tian-Kunze, X., Nerger, L., Liu, J., ... Zhang, Z. (2014). Assimilating SMOS sea ice thickness into a coupled ice-ocean model using a local SEIK filter. *J. Geophys. Res.*, *119*(10), 6680–6692. doi: 10.1002/2014JC009963
Zampieri, L., Goessling, H. F., & Jung, T. (2018). Bright prospects for arctic sea ice prediction on subseasonal time scales. *Geophysical Research Letters*, *45*(18), 9731–9738.
Zampieri, L., Goessling, H. F., & Jung, T. (2019). Predictability of antarctic sea ice edge on subseasonal time scales. *Geophysical Research Letters*.