Spatial Damped Anomaly Persistence of the sea-ice edge as a benchmark for dynamical forecast systems

Bimochan Niraula[1] and Helge Goessling[2]

[1]Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung
[2]Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research

November 22, 2022

Abstract

Accelerated loss of the sea-ice cover and increased human activities in the Arctic emphasize the need for skillful prediction of sea-ice conditions at sub-seasonal to seasonal (S2S) time scales. To assess the quality of predictions, dynamical forecast systems can be benchmarked against reference forecasts based on present and past observations of the ice edge. However, the simplest types of reference forecasts –persistence of the present state and climatology– do not exploit the observations optimally and thus lead to an overestimation of forecast skill. For spatial objects such as the ice-edge location, the development of damped-persistence forecasts that combine persistence and climatology in a meaningful way poses a challenge. We have developed a probabilistic reference forecast method that combines the climatologically derived probability of ice presence with initial anomalies of the ice-edge location, both derived from satellite sea-ice concentration data. No other observations, such as sea-surface temperature or sea-ice thickness, are used. We have tested and optimized the method based on minimization of the Spatial Probability Score. The resulting Spatial Damped Anomaly Persistence forecasts clearly outperform both simple persistence and climatology at sub-seasonal timescales. The benchmark is thus about as skilful as the best-performing dynamical forecast system in the S2S database. Despite using only sea-ice concentration observations, the method provides a challenging benchmark to assess the added value of dynamical forecast systems.

Spatial Damped Anomaly Persistence of the sea-ice edge as a benchmark for dynamical forecast systems

Bimochan Niraula[1], **Helge F. Goessling**[1]

[1]Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung

Key Points:

- We have developed a new method that combines climatological sea-ice probability and initial-state anomaly to forecast sea-ice presence.
- Ice-edge forecasts derived from this method can outperform climatological benchmarks at lead times of up to 2 months.
- Spatial damped anomaly persistence forecasts have a higher predictive skill than most models from the sub-seasonal to seasonal database.

Corresponding author: Bimochan Niraula, `bimochan.niraula@awi.de`

Abstract

Accelerated loss of the sea-ice cover and increased human activities in the Arctic emphasize the need for skillful prediction of sea-ice conditions at sub-seasonal to seasonal (S2S) time scales. To assess the quality of predictions, dynamical forecast systems can be benchmarked against reference forecasts based on present and past observations of the ice edge. However, the simplest types of reference forecasts – persistence of the present state and climatology – do not exploit the observations optimally and thus lead to an overestimation of forecast skill. For spatial objects such as the ice-edge location, the development of damped-persistence forecasts that combine persistence and climatology in a meaningful way poses a challenge. We have developed a probabilistic reference forecast method that combines the climatologically derived probability of ice presence with initial anomalies of the ice-edge location, both derived from satellite sea-ice concentration data. No other observations, such as sea-surface temperature or sea-ice thickness, are used. We have tested and optimized the method based on minimization of the Spatial Probability Score. The resulting Spatial Damped Anomaly Persistence forecasts clearly outperform both simple persistence and climatology at sub-seasonal timescales. The benchmark is thus about as skilful as the best-performing dynamical forecast system in the S2S database. Despite using only sea-ice concentration observations, the method provides a challenging benchmark to assess the added value of dynamical forecast systems.

Plain Language Summary

The Arctic is becoming more ice free and seeing more human activities, which means it is important to have reliable forecasts of sea-ice conditions weeks to months ahead. The accuracy of a forecast system is typically compared against reference forecasts based on present and past observations of the ice-edge location. However, the most widely used references either simply maintain the current state or consider states at the same time of the year during previous years. Such simple benchmarks can lead to an overestimation of how "skilful" a forecast system is considered. In the case of sea-ice edge, creating a better reference forecast combining both historical and current observation information can be challenging. We have addressed this challenge and developed a method that combines the historical probability of ice presence with the current location of the ice edge. The new method clearly outperforms the simpler methods and remains slightly better than historical-based forecasts even 2 months ahead. Despite using only observed sea-ice concentration data (like the simpler benchmarks), the new benchmark is about as good as the best modern model-based forecast system. The method therefore provides a good reference to study how well the latest forecast systems are actually performing.

1 Introduction

Accelerated sea ice loss and the possibility of ice-free summers in the Arctic has increased the interest in potential human activities in the far North (Stephenson et al., 2011). To address the planning and safety concerns associated with this, government and private agencies need better predictions of sea-ice at sub-seasonal to seasonal time-scales (Jung et al., 2016). Over the past few years, many operational centers are already starting to provide such forecasts with longer lead times, although the skill of these forecasts – and how to assess the skill in the first place – is still under question (Smith et al., 2015).

There are numerous metrics to measure and quantify the accuracy of a forecast against observations, or 'true' conditions, depending on the variable in question (Wilks, 2019). Whether or not forecasts are considered skillful depends not only on the metric to measure the forecast error, but also what benchmark is used to measure skill against. The skill of the forecast produced by a particular forecast system can be compared against that of an earlier version of the same system (e.g., Balan-Sarojini et al., 2019), a different forecast system (e.g., Zampieri et al., 2018, 2019), and also against a simpler reference forecast for benchmark (e.g., Woert et al., 2004; Pohlmann et al., 2004) .

Model outputs are commonly compared against two observation-based reference forecasts: Climatology and Persistence. Climatology is based on the historical records for the given time of the year. An ensemble constructed from previous years can give a probabilistic estimate of the target variable. In the presence of a significant seasonal cycle and for lead times longer than just a few days, a climatological forecast needs to change with lead time according to the evolving time of the year. Persistence, on the other hand, is maintaining the initial state of the variable - thus giving a constant output of the variable. A persistence forecast can also be constructed such that the seasonal evolution, estimated from previous years, is taken into account, resulting in an anomaly persistence forecast. For either of these benchmark approaches, a secular trend can be taken into account to derive a trend-adjusted variant of such a forecast (Van den Dool et al., 2006). By design, persistence has better forecast skill at shorter lead times, whereas climatology is more skillful at longer lead times when errors approach a saturation level due to chaotic error growth (Woert et al., 2004). Finally, a damped anomaly persistence forecast attempts to combine persistence and climatology in such a way that it gradually transitions from the persisted to the climatological state, thereby optimizing the skill of the forecast at intermediate lead times (Van den Dool et al., 2006).

Anomaly persistence and damped anomaly persistence can be applied easily to continuous quantities on a grid-cell per grid-cell basis and have been used as benchmark forecasts for quantities such as sea-surface temperature on decadal timescale (e.g., Pohlmann et al., 2004) and sea-ice concentration on seasonal timescale (e.g., Blanchard-Wrigglesworth et al., 2011) . However, it is not trivial to transfer the concept of (damped) anomaly persistence to spatial objects such as the ice-edge location that corresponds to a binary, rather than a continuous, gridded field. For such a quantity, it can be difficult to establish meaningful autocorrelations at longer time scales; anomalies in this case tend to migrate spatially with the seasonal cycle (e.g., Goessling et al., 2016) , which means that initial anomalies at one location may not be relevant at the same location after some time, but they might still hold information about anomalies at a different nearby location. Properly utilising these anomalies for prediction thus requires consideration for how to transfer information across the domain and not just at increasing lead time.

In case of the ice edge, an important variable for marine activities in the Arctic, the mismatch between the predicted and 'true' ice edge can be quantified using the Spatial Probability Score (SPS; Goessling & Jung, 2018). The SPS (and its deterministic counterpart, the Integrated Ice Edge Error; see Goessling et al., 2016) determines the area where ice is either under-forecast or over-forecast in comparison to the true outcome. Palerme et al. (2019) compared the SPS to the (modified) Hausdorff distance (MHD; Dukhovskoy et al., 2015), another commonly used verification metric, and determined that the SPS is more robust and less affected by isolated patches of ice. Therefore, we will primarily be using SPS as the comparison metric in this paper, but will also use MHD to test whether our results are robust with respect to the choice of the metric.

In terms of lead-times, Zampieri et al. (2018, 2019) have shown that some operational subseasonal-to-seasonal (S2S) forecast systems are more skillful than climatology in predicting the location of the ice edge several weeks ahead. After 1.5 months, however, even the most skilful forecast system is not performing better than climatology. This can be compared against perfect-model studies that suggest that the ice-edge position can be predictable up to 6 months ahead (Goessling et al., 2016), suggesting that forecast calibration (not applied in Zampieri et al., 2018, 2019) and/or forecast system improvements should in principle be possible. Moreover, simple initial-state persistence tends to outperform these dynamical systems for at least the first 3-4 days, leaving a temporal window between about 4 to 45 days where the best system can be considered skilful beyond the simple benchmark methods (Zampieri et al., 2018, 2019). However, given the simplicity of strict initial-state persistence and climatology, the term skilful might be complacent in the sense that a less naive, but still simple, benchmark forecast method building on the concept of damped anomaly persistence may exhibit similar skill or even outperform existing dynamical forecast systems.

In this context, we have developed a method for predicting the location of the sea-ice edge that combines the climatologically derived probability of ice presence with initial (present) anomalies of the ice edge. Here we describe the method and an assessment of the forecasts produced. This paper is structured as follows: Data used for creating the forecasts are presented in section 2, followed by a description of the applied verification metrics in section 3. The Spatial Damped Anomaly Persistence (SDAP) method is described in detail in section 4. In section 5, we go through the results from verifying the forecasts produced with our method compared against other traditional references as well as against the performance of model-based S2S forecast systems. The paper concludes with a short discussion in section 6.

2 Data

2.1 Sea-ice concentration observations

We have used the Global Sea Ice Concentration Climate Data Record from OSI SAF to determine the climatological and initial sea-ice edge. The data is labelled as OSI-450 for years 1979 to 2015 and is based on satellite microwave measurements. From 2016 onwards, OSI-450 was extended as OSI-430b and is available with a 16 days latency. Alongside the microwave measurements, these products also use operational analyses and forecasts from ECMWF for atmospheric corrections. A near real time version of this record without the additional corrections is also available, as OSI-430. All of the data is freely available on the OSI SAF website and further details regarding the processing are described by Lavergne et al. (2019). OSI-450 (and OSI-430b) records are given on a Lambert Azimuthal Equal Area polar projection, also known as the EASE2 grid. The two hemispheres are separated and the horizontal grid spacing is 25km. Following this setup, our forecasts are also produced on the EASE2 grid at 25km resolution for each hemisphere.

2.2 Subseasonal-to-seasonal (S2S) forecast data

We compare the performance of our forecasts against the performance of the models from the Subseasonal to Seasonal (S2S) Prediction Project. The S2S database contains forecasts and reforecasts from several major operational centers. We have measured the concentration forecast of 5 dynamical sea ice forecasting system - the European Centre for Medium- Range Weather Forecasts (ECMWF), the Korea Meteorological Administration (KMA), Météo-France(MF), the National Centers for Environmental Prediction (NCEP), and the UK Met Office (UKMO). Alongside, we also used concentration forecasts from an older version of ECMWF (here named 'ECMWFpres') where the sea ice state is prescribed using initial persistence for the first 15 days and then relaxed toward climatology. Further description regarding the S2S project is given by Vitart et al. (2017). For a consistent comparison, all forecasts have been interpolated to a common grid with 1.5 degree horizontal resolution and only the period between 1999 and 2010 are considered in the analysis, in line with the analysis provided in Zampieri et al. (2018, 2019).

3 Verification Metrics

3.1 Spatial Probability Score

We are primarily using the Spatial Probability Score (SPS; Goessling & Jung, 2018) for verification of the sea-ice forecasts, which is defined as the spatial integral of the squared probability difference (i.e., the half-Brier Score). Since the S2S models provide ensemble forecasts of sea-ice concentration, we can derive a continuous (non-binary) probability of ice-presence. The resulting score, which is in area units, quantifies the area of mismatch between forecast and observation. Measuring SPS of a deterministic, or binary, forecast results in the Integrated Ice Edge Error (IIEE; Goessling et al., 2016). We also use the SPS in our method for empirically optimising the weights by which our initial binary (anomaly persistence) forecast of the ice-edge location is damped towards climatology, resulting in a probabilistic (damped anomaly persistence) forecast, as detailed in Sect. 3.

3.2 Modified Hausdorff Distance

Given that we have used the SPS not only for evaluation but also for the empirical estimation of optimal damping weights (detailed below), we have also used a second verification metric to validate the skill of our method. The Modified Hausdorff Distance (MHD; Dukhovskoy et al., 2015; Palerme et al., 2019) measures the distance between two contours and the resulting score is in distance units. MHD can not consider non-binary probabilities, and therefore the probabilistic forecasts have been converted to a binary equivalent for measurement of MHD using the 50% contour of sea-ice probability. Since SPS and MHD measure forecast skill quite differently, using both metrics can reveal whether or not the forecasts from our method are skillful independent of the verification metric used.

4 Spatial Damped Anomaly Persistence (SDAP) Method

The steps towards creating the SDAP forecast of the ice-edge location can be divided into two parts. In the initialisation phase, the initial anomaly in ice-edge location is first derived from the climatological Sea Ice Probability (SIP; the probability of sea-ice concentration exceeding 15%) at points constituting the initially observed ice edge and then propagated (inherited) from the initial ice edge to each point of a (quasi-)global grid. In the forecasting phase, the inherited SIP and the climatology valid for the forecast target date are compared to determine first a deterministic ice-edge forecast by specifying "ice" versus "no ice" at each grid point, corresponding to an undamped anomaly persistence forecast, and second a probabilistic forecast by relaxing the deterministic forecast toward climatology, resulting in a damped anomaly persistence forecast.

Our method, detailed in the following, propagates the initial anomalies in local sea-ice extent in space by spatial inheritance, and then uses the spatial information to propagate the information in time based on the seasonal evolution of the climatological sea-ice probability. The rationale of this approach is based on the assumption that anomalies especially in the ice and ocean state associated with an ice-edge anomaly have a certain spatial and temporal correlation length. For example, if the ice edge extends further than usual into the ocean, this might have been caused thermodynamically by colder-than-usual regional temperatures (e.g., due to anomalous atmospheric circulation) that would also have caused thicker-than-usual and/or denser floes in the adjacent pack ice and colder-than-usual sea-surface temperatures (SST) in the adjacent open ocean. If this situation occurred during the melt season, the thicker-than-average ice would hamper the ice-edge retreat; if it occurred during the freeze season, the colder-than-average SST would facilitate the ice-edge advance.

4.1 Initialisation phase

For a given initial time, the fraction of years in the preceding 10 years with ice present (sea-ice concentration ¿ 15%) at the same time of the year determines the raw climatological sea-ice probability (SIP) at each gridpoint. A larger number of years is not used in order to minimize the influence of long-term trends. The limited sample size however can result in spurious small-scale variations in the raw climatological SIP. Therefore a Gaussian filter with a radius of 220 km is applied to smooth the climatological probability field (exemplary result shown in figure 1). This results in smoother SIP contour lines which are advantageous for the subsequent steps of spatial inheritance.

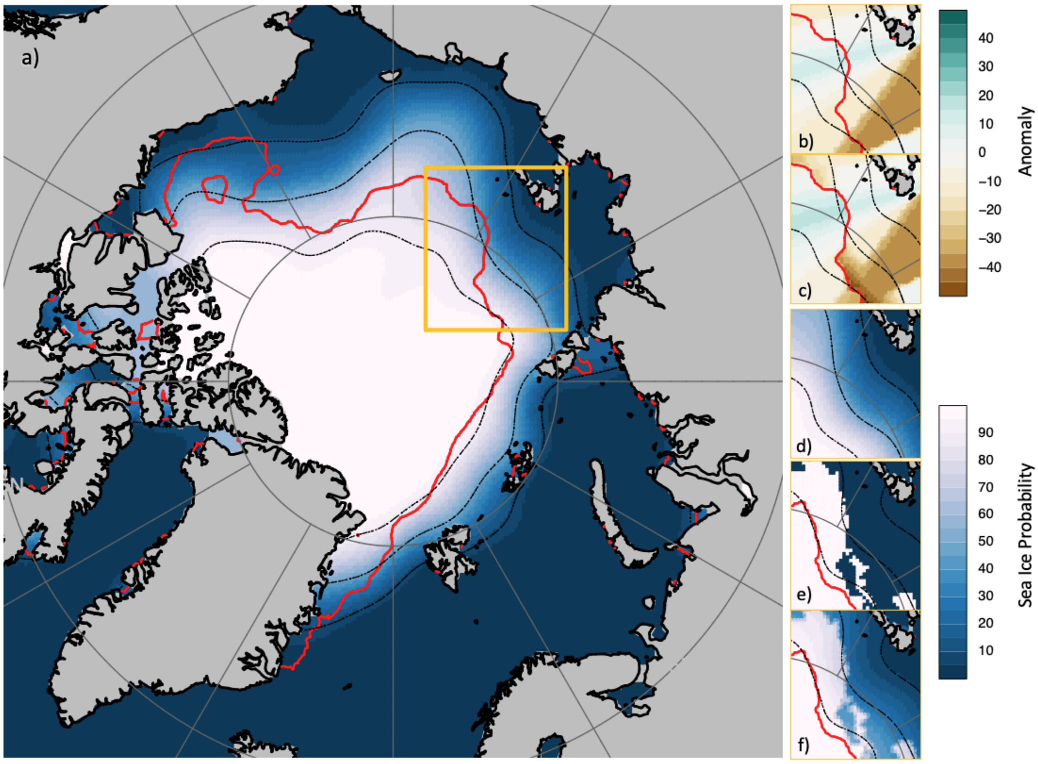

Figure 1. Map of the Arctic domain showing the climatological probability field for 1st of September 2020, and steps of the forecasting method (in insets) for the region selected. Insets 'b' and 'c' show the inherited and adjusted anomaly fields, 'd' shows the climatological probability field for the target date (30th of September, 2020), 'e' shows the Spatial Anomaly Persistence forecast and 'f' shows the Spatial Damped Anomaly Persistence forecast for the target date. In each panel, the red line denotes the 15% contour of the observed ice concentration for the relevant date while the dashed black lines show the 10%, 50% and 90% contour of the climatological probability field. Maps 'b' and 'c' use the Anomaly colour scale while all other maps use the Sea Ice Probability colour scale. Further details are in the text.

The initial ice edge, defined as the 15% contour of ice concentration on the initialisation date, is overlaid on the climatological probability. For every point along the contour, the climatological SIP at the nearest grid cell is used to compute the SIP anomaly in comparison to the median probability (SIP = 50%) along the initial ice edge. If the ice edge is in a region of high (or low) climatological probability, the anomaly is negative (or positive). In other words, if the ice extent is higher than the median, the anomaly is positive and the ice edge should be in a region of low climatological probability. Next, this anomaly is mapped from the initial edge to the climatological median contour (corresponding to 50% ice-presence probability). This intermediate step has been introduced to avoid biased patterns of spatial inheritance that otherwise occur due to geometrical effects (not shown). The anomaly is then passed (spatially "inherited") to the full grid using a nearest-neighbour match of the median contour from each grid cell, resulting in a map as shown in figure 1b.

In some cases, a grid cell with (without) ice in the initial observation might inherit a low (high) anomaly from the median, which causes the initial forecast at day 0 to not have (have) ice in the grid-cell. This initial mismatch generally occurs in cases where the observational gradient of ice presence is locally reversed compared to the climatological gradient of ice presence probability. The problem is simply solved by checking the initial day forecast and replacing the anomaly inherited from the climatological median with the anomaly at the grid cell (i.e., the climatological probability of ice presence minus 50%) for grid cells with an initial mismatch. The result is a corrected probability anomaly map that can be persisted in time and used for forecasting.

4.2 Forecasting phase

Based on the corrected anomaly map from the initialisation phase (Fig. 1c), a forecast for any lead time can be created by simply adding the anomaly to the climatological SIP at each grid cell for the day in question. For example, if we initialise our forecast on September 1 and want to forecast the sea-ice edge 29 days later, then we add the anomaly from September 1 to the climatological SIP for September 30 (Fig. 1d). For every gridpoint, if the resulting probability is 50% or more, then we assign that point to be ice-present (as shown in inset ‘d’ of figure 1) . This gives a map of 1 (ice) and 0 (no-ice) which is our deterministic anomaly persistence forecast of ice-presence (Fig. 1e).

The binary anomaly persistence forecast, corresponding to a single sharp ice edge, does not take into account that the spatial and temporal correlation scales of anomalies are limited. For example, an initial ice-edge anomaly in March obviously carries much less information about ice-edge anomalies in the following September, when (in most Arctic regions) the ice edge will be far away from the initial location, and sufficient time will have passed to turn initial anomalies into largely uncorrelated anomalies. Therefore, we create a probabilistic ice forecast by damping our deterministic forecast towards climatology. For each lead-time, a damped probabilistic forecast is obtained as the weighted average of the deterministic prediction and the climatological probability. The optimal weights are determined by using a training set of forecasts, where we compute the SPS for anomaly weights between 0 and 1 in steps of 0.05, for each combination of lead time and initial time of the year, and find the weights minimising the error. The optimised anomaly weights (shown in section 3.3) can then be used for other, independent, years to get the probabilistic Spatial Damped Anomaly Persistence (SDAP) forecast (Fig 1 inset ‘f’). A schematic overview on the whole forecasting procedure is provided in Fig. S1.

In our study, we initialise the forecasts at the start of each month between 1989 and 2020 and make predictions of the ice edge for the following year. The first 10 years of the period are used for empirical training to determine the anomaly weights and are thus excluded from the analysis. The remaining 22 years (1999 to 2020) have been used for the evaluation shown in the results below. For comparing to the forecasts from dynamical models in the S2S dataset (section 3.4), the forecasts from our method were interpolated to the coarse 1.5° S2S grid, and only years 1999 to 2010 were used, as in ZZampieri et al. (2018, 2019).

5 Results

Here, we present the evaluation of the forecasts from our method compared against other traditional benchmarks for ice-edge forecast, as well as against forecasts from dynamical models in the S2S dataset. Alongside climatological probability (hereafter referred to as CLIM) and initial-state persistence (PERS), the 50% contour of the climatological probability has also been used to generate a binary forecast - the climatological median (CLIM-MED). Similarly, the 50% contour of the spatial damped anomaly persistence (SDAP) has been used to generate a median damped anomaly persistence forecast (SDAP-MED). Note that this is different from the binary Spatial Anomaly Persistence forecast (SAP) described above.

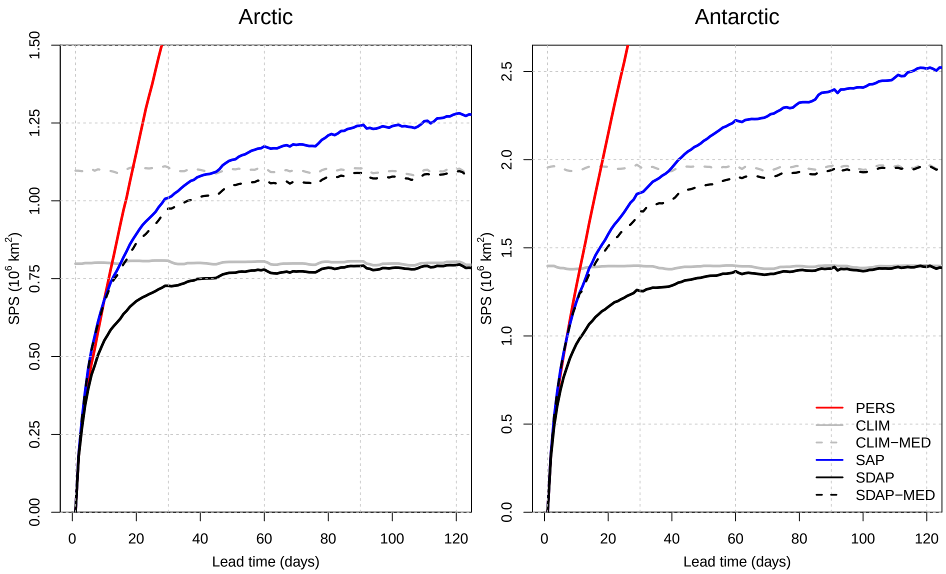

Figure 2. Spatial Probability Score (SPS) at increasing lead times (in days) for all forecasts averaged across all forecasts initialised between 1999 and 2020. The two dotted lines (CLIM-MED and SDAP-MED) are binary forecasts derived by assigning the ice edge at the 50% contour of the probabilistic forecasts (CLIM and SDAP). Further details are in the text.

5.1 Comparison against traditional benchmarks

The SPS result of all forecasts, averaged between 1999 and 2020 and across all seasons (Fig. 2), reveals that the performance of the SAP forecast is better than CLIM up to day 15 and better than CLIM-MED up to day 39 in both hemispheres. The performance of SDAP is better than SAP, outperforming CLIM at 2 months of lead time by an average of 0.03 million km2 in the Arctic (0.04 million km2 in the Antarctic). While SAP forecasts show an improvement over simple persistence from day 6, SDAP is better from day 3. The method design enables this forecast to perform at least as well as CLIM even at long lead times, leading it to be the most skillful forecast in this comparison set at all lead times except the first 2-3 days where PERS is marginally better.

Since the SPS considers the probabilistic information in a forecast, both CLIM and SDAP have a clear advantage over their binary counterparts. CLIM-MED has a near constant skill loss of 0.3 million km2 in the Arctic (0.5 million km2 in the Antarctic) compared to CLIM. The error of SDAP-MED is low initially (similar to the other anomaly forecasts) and converges towards the error of CLIM-MED, similar to how the error of SDAP converges towards the error of CLIM.

Given that our methodology uses the SPS to optimise damping weights, it is possible that using the SPS also as a verification metric leads to an overestimation of the skill. We thus also use the Modified Hausdorff Distance (MHD) as an independent verification metric, although the MDH can be applied only to the binary (median-based) forecast variants. While this verification method measures the forecast skill quite differently compared to SPS (Palerme et al., 2019), repeating the evaluation of the binary forecast variants with the MHD instead of the SPS provides overall a very similar picture: The SDAP-MED forecast outperforms the CLIM-MED forecast throughout the lead-time range, the undamped (SAP) forecast outperforms the CLIM-MED forecast up to 60 days lead time in the Arctic (43 days in the Antarctic), and simple persistence outperforms the other forecasts only slightly during

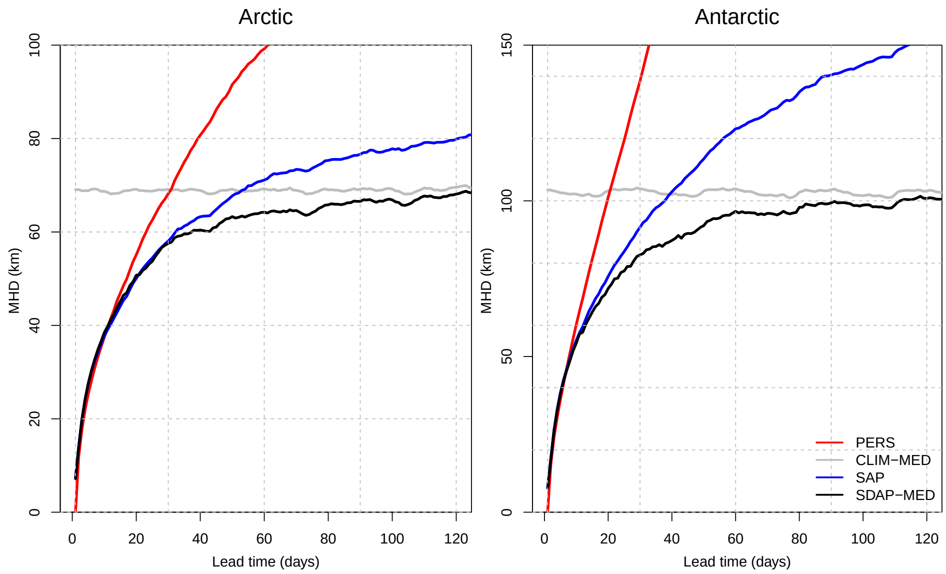

Figure 3. Same as figure 2, but for Modified Hausdorff Distance (MHD) of the binary forecasts compared to the observed ice edge.

the first few days (Fig. 4). The confirmation of the overall results with an independent metric provides additional confidence in our reference forecast method.

The previous analysis was based on outputs averaged over the entire time period. Given the strong seasonal cycle of Arctic sea ice and the processes that dominate the sea-ice budget, we now assess the performance of our method as a function of season. The forecast errors generally increase with lead time for all initialisation months, yet the actual score differs based on the month (Fig. 4). The errors are on average higher for the summer period in both hemispheres, and also during the late winter in the Arctic, when the sea ice is at its largest extent. These seasonal cycles can be linked to corresponding variations of the ice-edge length, although the December/January SPS maximum in the Antarctic is more likely related to enhanced interannual lateral ice-edge variability (Goessling et al., 2016).

For each initialisation month, the SDAP scores approach and then follow the SPS of climatology. Forecasts initialised in some months perform worse than climatology at long lead times, such as the Arctic forecasts initialised in June, despite the method design to merge the SDAP forecasts at long lead times to the climatological forecast. This is likely a result of sampling uncertainty and using different years for training the anomaly weights and for evaluating the forecast results, as detailed in the following.

5.2 Damping weights

As mentioned above, the deterministic anomaly forecasts are damped by empirically determined weights and added to an inversely weighted climatology to derive the probabilistic anomaly forecast. Here, the empirical fitting was done only over the optimisation period of 1989 to 1998 and a constant set of weights was then used for all forecasts within the evaluation period of 1999 to 2020. The weights, derived from the optimization period for each hemisphere separately (Fig. 4), reveal the timescale at which information from the initial anomalies is lost and climatology becomes more informative. By day 30, the initial-anomaly weights decrease to 50% for most initialisations, but most forecasts have not been completely damped to climatology even at 3 months of lead time. In both hemispheres,

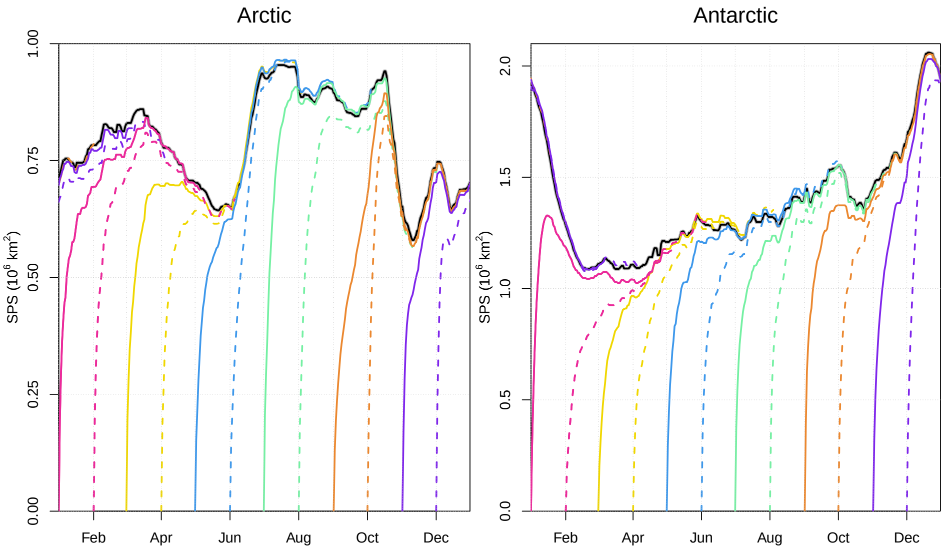

Figure 4. SPS for the first 120 days of SDAP forecasts initialised at the start of each month (as represented by the coloured lines), alongside CLIM for each date (black continuous line), with results averaged between 1999 and 2020.

the anomaly weight stays high for a longer period at the end of the summer melt season - August/September in the Arctic and February in the Antarctic.

The weight of the anomaly is not always decreasing monotonically (see figure 5), although in general the influence of initial conditions should decrease with time. While some intermittent fluctuations can be caused by sampling uncertainty due to the limited optimisation period, we argue that some of this can be linked to what is known as reemergence of sea-ice anomalies (Blanchard-Wrigglesworth et al., 2011): When over the course of the seasonal cycle the (climatological) ice edge first migrates away from the initial location and then returns to a similar location later, the initial ice-edge anomalies may carry more information for that later state than for the earlier - but more remote - intermediate state. This seems to be the case in particular for the Arctic forecasts initialised at the beginning of August, which exhibit a local minimum of the anomaly weight around day 35-40 (around the sea-ice minimum) and higher weights again thereafter until around day 70 (Fig. 5). The increased anomaly weight acts to keep the associated forecast error below the error of climatology for longer compared to other initialisation months (Fig. 4).

5.3 September 2020 as an illustrative example

To illustrate our forecast method, we now consider the example of the Arctic ice edge with the initial condition on the 1st of September, 2020, as shown in Fig. 1. Sea-ice extent in the Arctic during September 2020 was the second lowest in satellite records and during October 2020 was the lowest October ice extent measured ((NSIDC, 2020). Persistent offshore winds from the Siberian coast, associated with a positive phase of the Arctic Oscillation (AO) during the preceding winter, meant that the Eurasian parts of the Arctic were mostly ice-free and had strong negative anomalies (as seen in Fig. 1b).

SDAP forecasts in this particular case are better in the American part of the Arctic than in the Eurasian part. The forecasts correctly suggest that positive anomalies would persist in

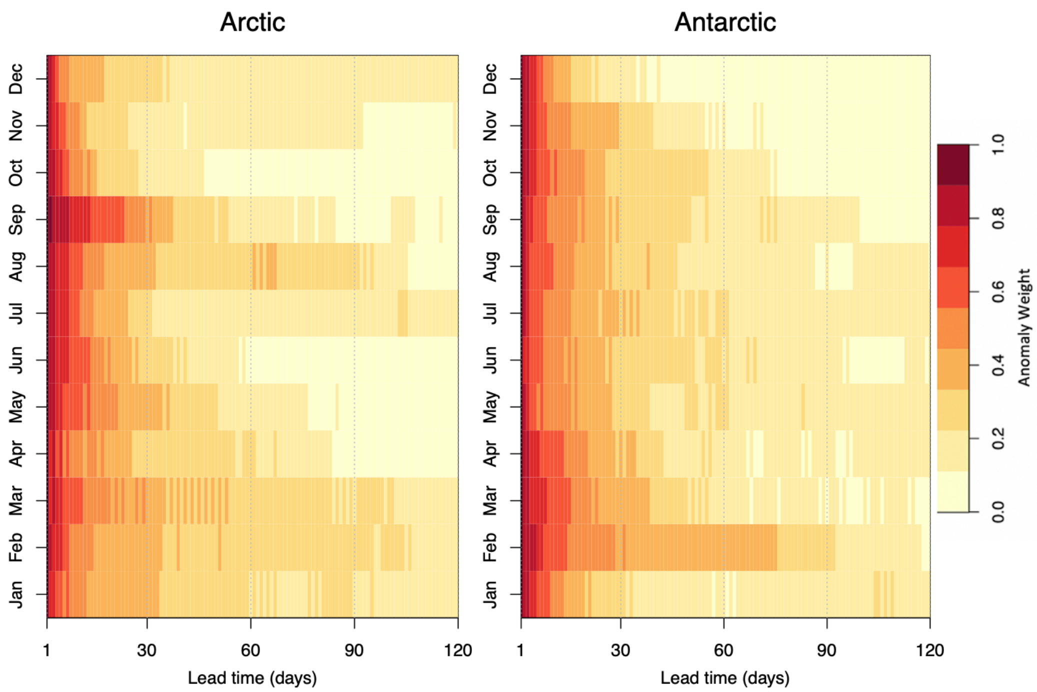

Figure 5. Anomaly weights for the SDAP forecast as a function of initialisation month and lead time (first 120 days). These weights determine the ratio of SAP forecast to climatology in the SDAP forecast and are measured using optimization of SPS as described in the text.

the southeastern Beaufort Sea (on the left in each panel of Fig. 6) until the end of the month, and even the persistence of the ice-free patch within the ice cover of the Beaufort Sea is correctly forecast, although the position is shifted. In the Greenland Sea and Fram strait, the forecasts are fairly accurate at day 15 and 30, and still better than the climatological median at day 45. North of the Laptev Sea, the forecast retains the initial negative anomalies - as dictated by the concept of anomaly persistence - and thus shows a similarity in shape to the initial ice edge. While a small patch remains accurately ice-free at day 30, the overall forecast in this region shows an expansion of the high-probability areas toward the coast following the seasonal evolution of the climatological probabilities. By contrast, the actual ice edge does not follow the usual seasonality but remains largely unchanged, thereby developing even stronger negative anomalies. The SDAP forecast remains better than climatology, but the anomaly intensification is not captured. This highlights the limitation of the SDAP approach, given that it is based on the persistence of anomalies.

North of Franz Joseph Land, the initial anomaly is strongly negative, yet there is ice present along the island coasts (in a positive anomaly region), allowing some parts of the median edge to inherit positive anomalies. Depending on the exact position of a grid cell in relation to the median and ice-edge contours, it might inherit slightly different anomalies and predict different ice conditions. This can be seen in the SDAP forecast for day 45, where some grid cells of the ocean in this region have a lower probability of ice-presence than the surrounding region. We discuss this issue further in section 6.

5.4 Comparison against S2S dataset

The main motivation for this study is to use the damped anomaly forecast as a reference benchmark for evaluating dynamical sea-ice models. Therefore, here we compare the performance of the forecasts from this method against those from the S2S Prediction database

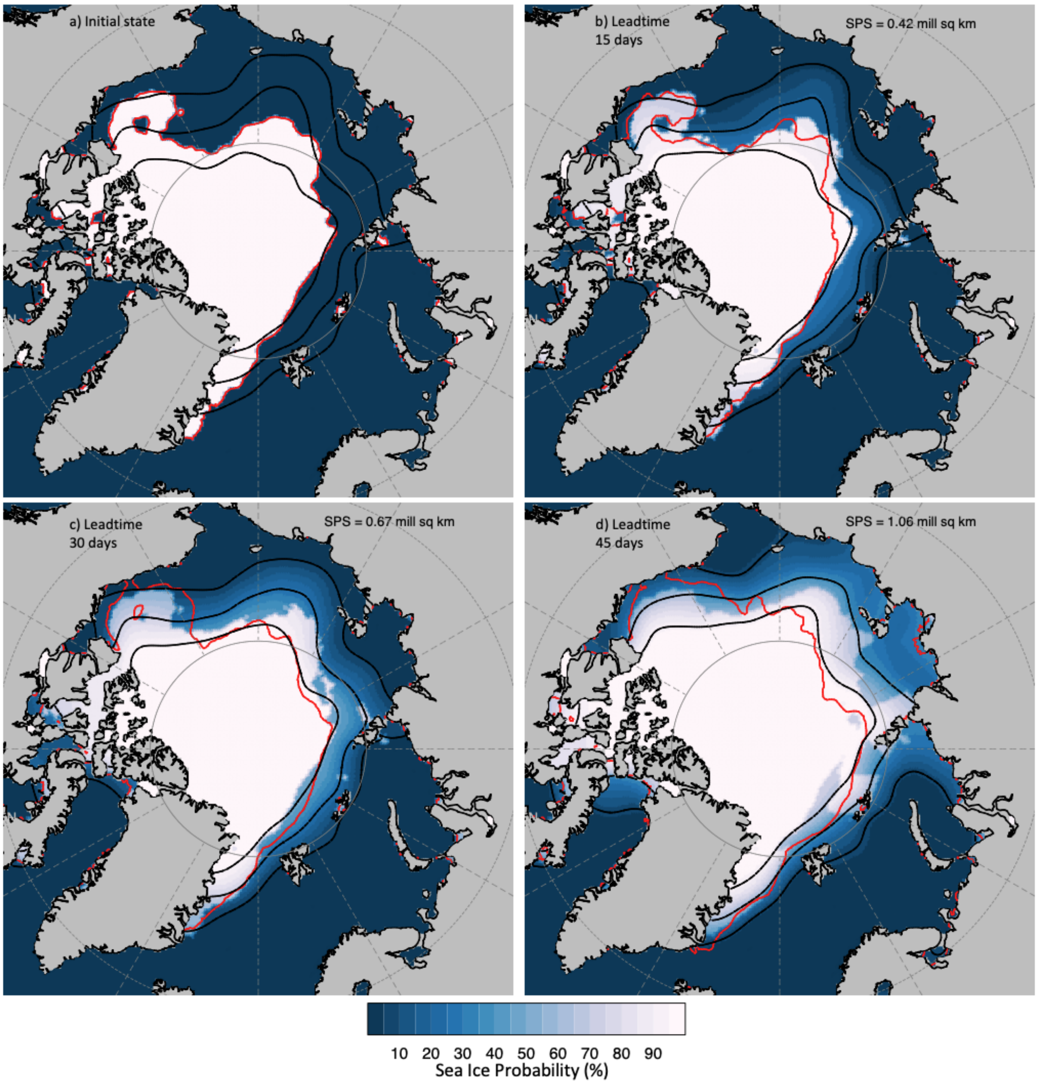

Figure 6. Observed sea ice conditions in the Arctic for 1st of September, 2020 and the corresponding SDAP forecasts at lead times of 15, 30 and 45 days. In each of the panels, the green lines show the contours of climatological probability (10, 50 and 90%), while the red line shows the actual ice edge on the respective date.

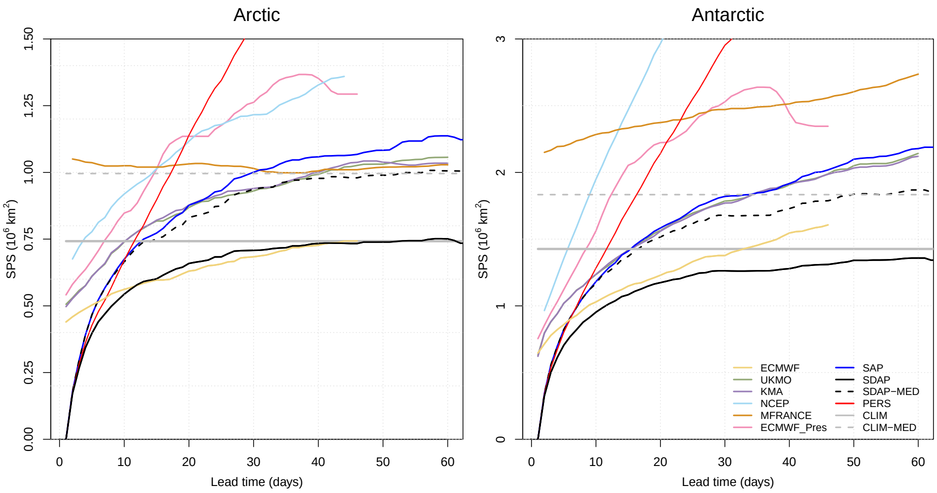

Figure 7. SPS for forecasts from the S2S dataset, alongside climatology (CLIM), initial state persistence (PERS), Spatial Anomaly Persistence (SAP) and Spatial Anomaly Damped Persistence (SDAP) forecasts, averaged across all seasons and years between 1999 and 2010.

(Vitart et al., 2017), in line with the analyses presented in Zampieri et al. (2018, 2019). The results for the S2S models shown here are similar to those found by Zampieri et al. (2018, 2019), although the scores have increased throughout due to the use of a larger sea-mask. The SDAP forecasts for years 1999 to 2010 were remapped to the common 1.5° S2S grid for this analysis and this has led to an increase in forecast error relative to climatology - in contrast to the evaluation on the OSI-SAF grid (Fig. 2), the SDAP error now reaches and slightly surpasses the clmatological error around 60 days lead time in the Arctic (Fig. 7 left; see also Fig. S3).

The undamped (thus binary) SAP forecast has a similar forecast skill as UKMO and KMA in both hemispheres (Fig. 7), despite the fact that these forecast systems have the advantage of providing ensemble-based probabilities rather than a binary ice edge. The damped (thus probabilistic) SDAP forecast in the Arctic clearly outperforms UKMO and KMA and is about as skillful as ECMWF. The latter is the best-performing model in the S2S set (Zampieri et al., 2018, 2019) and (without calibration) the only one that is more skillful than climatology beyond day 15 in the annual average. In the Antarctic, our SDAP method even outperforms the ECMWF ensemble, in particular beyond 20 days lead time when the ECMWF system appears to develop biases so that climatology provides a better forecast beyond 32 days lead time.

To assess the robustness of our results and to avoid a skill overestimation for our benchmark method due to the use of the SPS for the weight optimisation, we now compare the performance of the SDAP forecasts against those from the S2S dataset using Modified Hausdorff Distance (MHD; Fig. S2). Due to the differences in the method, including that MHD can be applied only to binary forecasts (based on the median ice edge where applicable), MHD measurements do not precisely mirror the results of the SPS measurements (for all models). Yet the ranking of the models is similar. The anomaly forecasts have a high skill compared to most other models or climatology. The average MHD of the ECMWF forecasts is again the lowest in the S2S set, except at short lead times below 12 days. The SDAP forecasts outperform the ECMWF forecasts in the short range and roughly match the skill of ECMWF at longer lead times, particularly in the Antarctic.

6 Summary and Conclusion

This paper describes a novel method for forecasting the presence and extent of ice in the Arctic or Antarctic based on persistence and damped persistence of probabilistic anomalies. The method requires only historical and initial ice-presence information to make the predictions, yet remains more skillful than climatological forecasts at month-long lead times. This is in contrast to most of the models from the S2S database, of which (without calibration) only one model performs better than climatology beyond 12 days of lead time.

The SDAP method uses the probabilistic anomaly from the initial date, which is spatially distributed ("inherited") by a nearest-neighbour search, and adds it to the climatology of the target date to generate a deterministic anomaly persistence forecast. With increasing lead time the probabilistic anomaly is damped to generate a probabilistic forecast. At longer lead times, one can expect that dynamical and thermodynamical processes cause the initial anomalies to be less informative. Therefore, the method has been designed to increase the damping with time and converge to climatology at long lead times. The damping weights, determined empirically using the reforecasts between 1989 and 1998, show that initial anomalies remain highly informative (weight $> 30\%$) for about 20-30 days in most months, and even longer in late summer. It is possible that in the Arctic, as the climatological ice-edge shifts with the continued decrease in ice extent, the optimal damping weights might also change.

We note that there are instances of sharp spatial transitions in the anomalies inherited to the grid (as described in section 5.3) due to the spatial distribution of the initial ice edge. Spatially smoothing the anomaly before passing it to the grid could smoothen the transition. Explicitly adding a spatial component to the damping might also be better than using a pan-Arctic damping weight evolving only with lead time. This might also have the potential to implicitly capture the re-emergence of anomalies when the ice edge returns to a location over the course of the seasonal cycle after the extent has reached its maximum or minimum. Nevertheless, the empirical approach used here, while simplistic, gives a good estimation of the overall decrease in the information content of the anomalies.

While our results show that the damped anomaly forecast outperforms most of the models in the S2S dataset, it must be noted that the skill of the dynamical sea-ice models could be higher than shown after bias correction or other forms of forecast calibration, which is standard for forecasts of other predictands at subseasonal-to-seasonal timescales. Forecast calibration remains challenging for sea ice, although promising approaches have recently been suggested (e.g., Dirkson et al., 2019; Director et al., 2017). Moreover, the resolution of the common S2S grid is low, and forecast skill was found to deteriorate after interpolation into this grid (Fig S3). It is possible that measuring the performance of the S2S models on their native grid would have resulted in a higher skill. The models output several variables, whereas our method is designed to only forecast ice-presence. Combining these outputs, or simply using a different concentration threshold, could also lead to more skill in the predictions, as shown by Zampieri et al. (2019).

The SDAP method, applied here for predicting ice-presence equivalent to 15% or more sea-ice concentration, could also be used for predicting other binary fields. Considering sea-ice concentration, Mizuta et al. (2008) proposed a different probabilistic method to estimate ice concentration by combining individual predictions for different concentration thresholds; While the method is quite different, a similar framework for our method can be used for estimating ice concentration or thickness by using several binary levels. Furthermore, the damped anomaly prediction method might also have applications in other fields not related to sea ice.

To conclude, the method proposed here is on average as skillful as the ECMWF forecast system, which is the most skillful one in the S2S database. Comparing only against persistence and climatology can give the impression that sea-ice forecasts from some of the

S2S forecast systems can already be regarded as "skillful" and thus of potential value for users. However, using a more challenging benchmark such as the spatial damped anomaly persistence (SDAP) method introduced here reveals that the forecast systems still have a way to go until they can generate substantial value beyond much simpler methods. We hope that, by including more challenging benchmark forecast methods such as ours in their evaluation workflow, other researchers and forecasting centers can build a better basis to improve their sea-ice forecast systems.

Acknowledgments

All data analyzed here are openly available. The OSI-SAF sea ice concentration product can be retrieved from the MET Norway FTP server at ftp://osisaf.met.no/reprocessed/ice/. The S2S forecasts data can be retrieved from the ECMWF data portal at http://apps.ecmwf.int /datasets/data/s2s/levtype=sfc/type=cf/. We are grateful to the World Weather Research Program (WWRP) and to the World Climate Research Program (WCRP), to operational forecast centers and individuals that contributed to the S2S database, as we are grateful to all those involved in implementing and maintaining the database. We acknowledge the OSI-SAF consortium for making their sea ice concentration products available. A special thanks to Lorenzo Zampieri and Barbara Casati for providing helpful comments and discussions regarding our work. This paper is a contribution to the Year of Polar Prediction (YOPP), a flagship activity of the Polar Prediction Project (PPP), initiated by the World Weather Research Programme (WWRP) of the World Meteorological Organisation (WMO). We acknowledge the WMO WWRP for its role in coordinating this international research activity. We acknowledge the financial support by the Federal Ministry of Education and Research of Germany in the framework of SSIP (grant 01LN1701A).

References

Balan-Sarojini, B., Tietsche, S., Mayer, M., Alonso-Balmaseda, M., & Zuo, H. (2019). *Towards improved sea ice initialization and forecasting with the ifs.* Retrieved 2021-07-01, from `https://www.ecmwf.int/en/elibrary/18918-towards-improved-sea-ice-initialization-and-forecasting-ifs`

Blanchard-Wrigglesworth, E., Bitz, C. M., & Holland, M. M. (2011, 09). Influence of initial conditions and climate forcing on predicting arctic sea ice. *Geophysical Research Letters*, *38*, n/a-n/a. doi: 10.1029/2011gl048807

Director, H. M., Raftery, A. E., & Bitz, C. M. (2017, 12). Improved sea ice forecasting through spatiotemporal bias correction. *Journal of Climate*, *30*, 9493-9510. doi: 10.1175/jcli-d-17-0185.1

Dirkson, A., Merryfield, W. J., & Monahan, A. H. (2019, 02). Calibrated probabilistic forecasts of arctic sea ice concentration. *Journal of Climate*, *32*, 1251-1271. doi: 10.1175/jcli-d-18-0224.1

Dukhovskoy, D. S., Ubnoske, J., Blanchard-Wrigglesworth, E., Hiester, H. R., & Proshutinsky, A. (2015, 09). Skill metrics for evaluation and comparison of sea ice models. *Journal of Geophysical Research: Oceans*, *120*, 5910-5931. doi: 10.1002/2015jc010989

Goessling, H. F., & Jung, T. (2018, 04). A probabilistic verification score for contours: Methodology and application to arctic ice-edge forecasts. *Quarterly Journal of the Royal Meteorological Society*, *144*, 735-743. doi: 10.1002/qj.3242

Goessling, H. F., Tietsche, S., Day, J. J., Hawkins, E., & Jung, T. (2016, 02). Predictability of the arctic sea ice edge. *Geophysical Research Letters*, *43*, 1642-1650. doi: 10.1002/2015gl067232

Jung, T., Gordon, N. D., Bauer, P., Bromwich, D. H., Chevallier, M., Day, J. J., … Yang, Q. (2016, 09). Advancing polar prediction capabilities on daily to seasonal time scales. *Bulletin of the American Meteorological Society*, *97*, 1631-1647. doi: 10.1175/bams-d-14-00246.1

Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., … Pedersen,

L. T. (2019, 01). Version 2 of the eumetsat osi saf and esa cci sea-ice concentration climate data records. *The Cryosphere*, *13*, 49-78. doi: 10.5194/tc-13-49-2019

Mizuta, R., Adachi, Y., & Kusunoki, S. (2008). *Estimation of the future distribution of sea surface temperature and sea ice using the cmip3 multi-model ensemble mean.* Retrieved 2020-07-01, from https://www.mri-jma.go.jp/Publish/Technical/DATA/VOL_56/tec_rep_mri_56.pdf

NSIDC. (2020, 10). *October — 2020 — arctic sea ice news and analysis.* Retrieved 2021-01-01, from https://nsidc.org/arcticseaicenews/2020/10/

Palerme, C., Müller, M., & Melsom, A. (2019, 05). An intercomparison of verification scores for evaluating the sea ice edge position in seasonal forecasts. *Geophysical Research Letters*, *46*, 4757-4763. doi: 10.1029/2019gl082482

Pohlmann, H., Botzet, M., Latif, M., Roesch, A., Wild, M., & Tschuck, P. (2004, 11). Estimating the decadal predictability of a coupled aogcm. *Journal of Climate*, *17*, 4463-4472. doi: 10.1175/3209.1

Smith, G. C., Roy, F., Reszka, M., Surcel Colan, D., He, Z., Deacu, D., . . . Lajoie, M. (2015, 05). Sea ice forecast verification in the canadian global ice ocean prediction system. *Quarterly Journal of the Royal Meteorological Society*, *142*, 659-671. doi: 10.1002/qj.2555

Stephenson, S. R., Smith, L. C., & Agnew, J. A. (2011, 05). Divergent long-term trajectories of human access to the arctic. *Nature Climate Change*, *1*, 156-160. doi: 10.1038/nclimate1120

Van den Dool, H. M., Peng, P., Johansson, A., Chelliah, M., Shabbar, A., & Saha, S. (2006, 12). Seasonal-to-decadal predictability and prediction of north american climate—the atlantic influence. *Journal of Climate*, *19*, 6005-6024. doi: 10.1175/jcli3942.1

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., . . . Zhang, L. (2017, 01). The subseasonal to seasonal (s2s) prediction project database. *Bulletin of the American Meteorological Society*, *98*, 163-173. doi: 10.1175/bams-d-16-0017.1

Wilks, D. S. (2019). Forecast verification. *Statistical Methods in the Atmospheric Sciences*, 369-483. doi: 10.1016/b978-0-12-815823-4.00009-2

Woert, M. L. V., Zou, C.-Z., Meier, W. N., Hovey, P. D., Preller, R. H., & Posey, P. G. (2004, 06). Forecast verification of the polar ice prediction system (pips) sea ice concentration fields. *Journal of Atmospheric and Oceanic Technology*, *21*, 944–957. Retrieved 2021-07-01, from https://journals.ametsoc.org/view/journals/atot/21/6/1520-0426_2004_021_0944_fvotpi_2_0_co_2.xml doi: 10.1175/1520-0426(2004)021⟨0944:FVOTPI⟩2.0.CO;2

Zampieri, L., Goessling, H. F., & Jung, T. (2018, 09). Bright prospects for arctic sea ice prediction on subseasonal time scales. *Geophysical Research Letters*, *45*, 9731-9738. doi: 10.1029/2018gl079394

Zampieri, L., Goessling, H. F., & Jung, T. (2019, 08). Predictability of antarctic sea ice edge on subseasonal time scales. *Geophysical Research Letters*, *46*, 9719-9727. doi: 10.1029/2019gl084096

JGR-Oceans

Supporting Information for

Challenging Dynamical Forecast Systems with Spatial Damped Anomaly Persistence for the Sea-Ice Edge

Bimochan Niraula, Helge F. Goessling

Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung

Contents of this file

Figures S1 to S3

Introduction

This file contains three figures. The first figure shows a schematic showing the steps taken in deriving the Spatial Damped Anomaly Forecast from initial and climatological sea-ice concentration data. The second figure is the result of the average Modified Hausdorff Distance (MHD) for forecasts from the S2S database, as well as the traditional reference and anomaly forecasts. The third figure shows the change in mean SPS for CLIM and SDAP forecast after interpolation into the common S2S grid and after applying the common mask. Each figure has an attached caption with additional information.

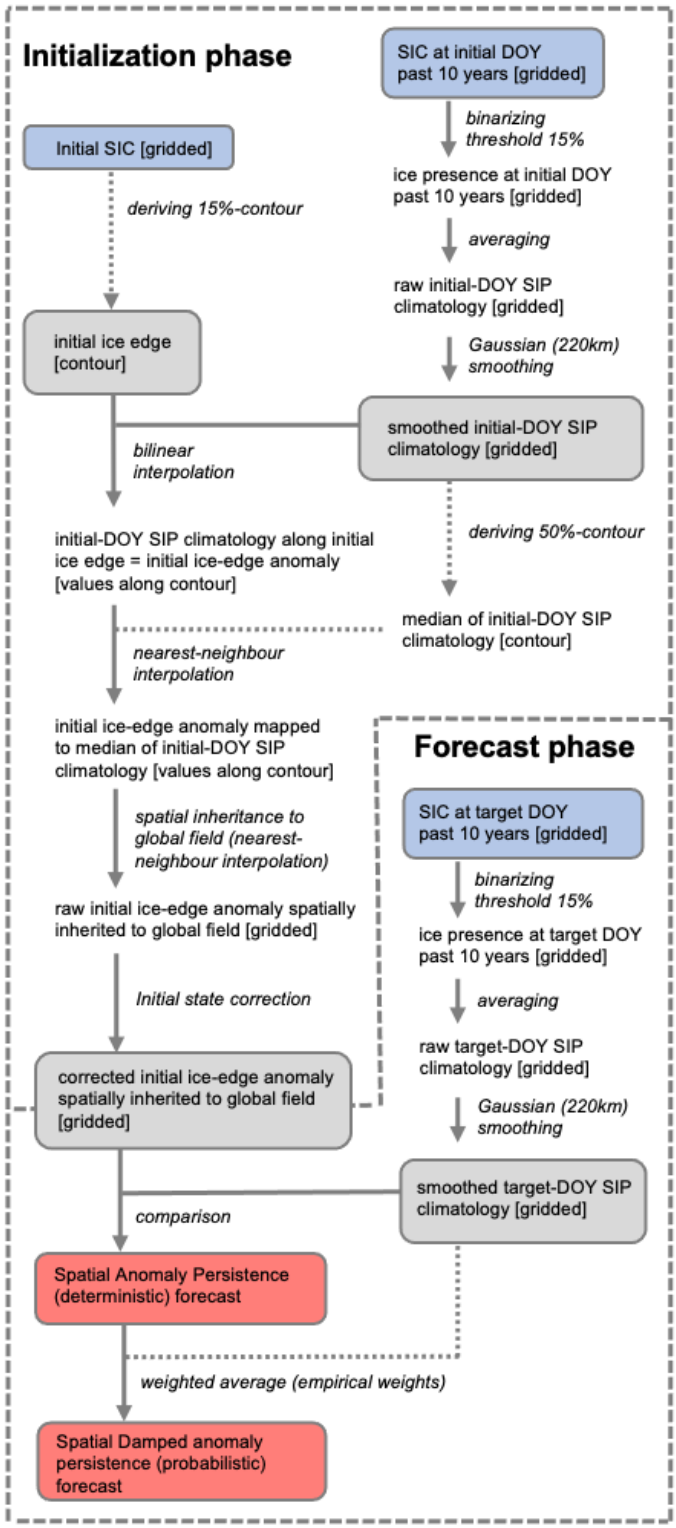

Figure S1. Schematic of the Spatial Damped Anomaly Persistence method as described in the main publication. The only data used as input for the method is SIC at initial date and climatological SIC, found using the SIC for the same DOY for the past 10 years. Further information is in the main text.

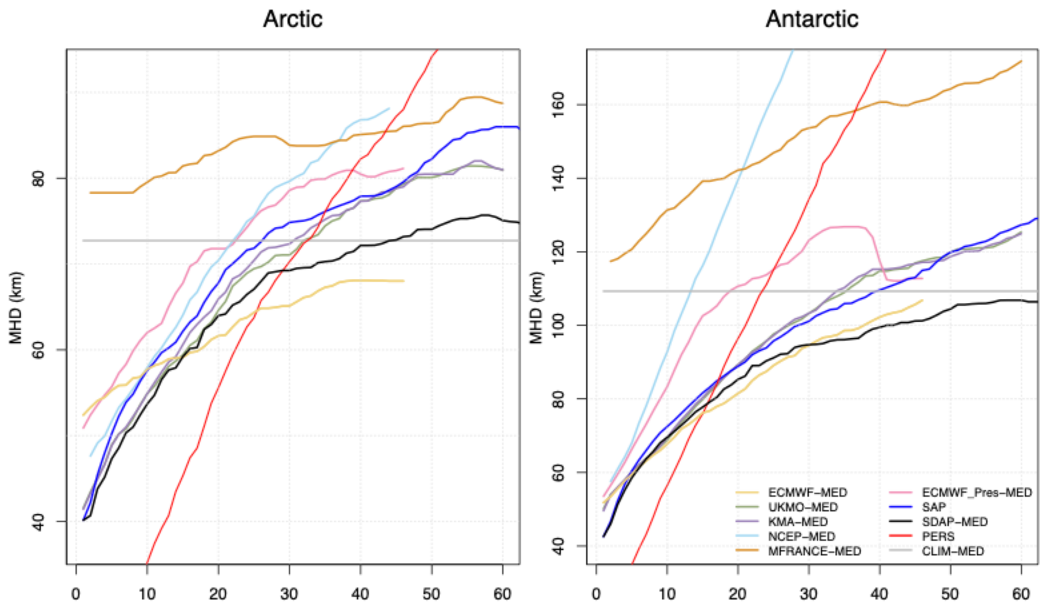

Figure S2. Same as figure 7, but for Modified Hausdorff Distance (MHD), using the 50% probability threshold to create the binary forecasts. For the probabilistic forecasts, the median probability of ice presence is used to determine the predicted ice edge. Due to the resolution of the S2S grid and the disconnected segments of the ice-edge, the MHD measurements do not precisely mirror the results of the SPS measurements. Nevertheless, the ranking of the models is similar to that found using SPS and we can confirm that the SDAP forecasts perform better than all models in the S2S database except ECMWF.

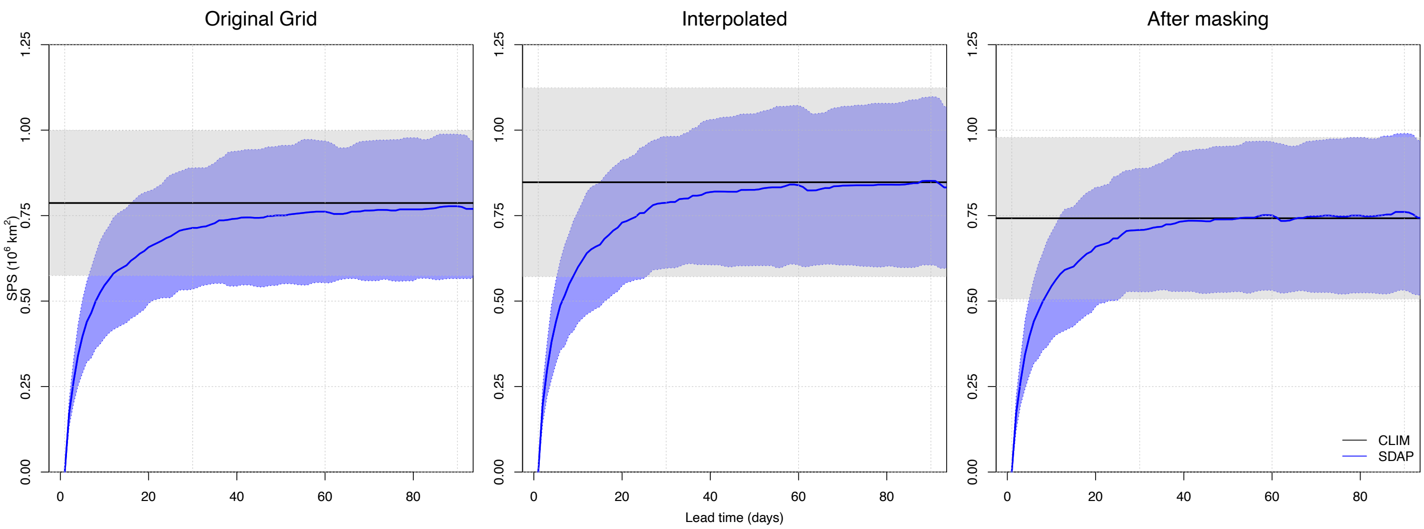

Figure S3. A comparison between the SPS of climatological forecast (CLIM) and the Spatial Damped Anomaly Persistence forecast (SDAP) in the original grid, after interpolation to the common S2S grid and after applying a common mask. The mask is derived by finding all grid-cells, where each of the forecast systems being compared has either ice or ocean. Results show the average of all forecasts for years 1999 to 2010, while the shaded areas show +/- 1 standard deviation.