Assessing the forecast skill of Arctic sea ice extent in the GloSea4 seasonal prediction system

K. Andrew Peterson · A. Arribas · H. T. Hewitt ·
A. B. Keen · D. J. Lea · A. J. McLaren

Abstract An assessment of the ability of the Met Office seasonal prediction system, GloSea4, to accurately forecast Arctic sea ice concentration and extent over seasonal time scales is presented. GloSea4 was upgraded in November 2010 to include the initialization of the observed sea ice concentration from satellite measurements. GloSea4 is one of only a few operational seasonal prediction systems to include both the initialization of observed sea ice followed by its prognostic determination in a coupled dynamical model of sea ice. For the forecast of the September monthly mean ice extent the best skill in GloSea4, as judged from the historical forecast period of 1996–2009, is when the system is initialized in late March and early April near to the sea ice maxima with correlation skills in the range of 0.6. In contrast, correlation skills using May initialization dates are much lower due to thinning of the sea ice at the start of the melt season which allows ice to melt too rapidly. This is likely to be due both to a systematic bias in the ice-ocean forced model as well as biases in the ice analysis system. Detailing the forecast correlation skill throughout the whole year shows that for our system, the correlation skill for ice extent at five to six months lead time is highest leading up to the September minimum (from March/April start dates) and leading up to the March maximum (from October/November start dates). Conversely, little skill is found for the shoulder seasons of November and May at any lead time.

Keywords Arctic sea ice · Seasonal forecasting · Ocean and sea ice analysis · Data assimilation · Ice concentration

K. A. Peterson (✉) · A. Arribas · H. T. Hewitt · A. B. Keen
Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK
e-mail: drew.peterson@metoffice.gov.uk

A. Arribas
e-mail: alberto.arribas@metoffice.gov.uk

H. T. Hewitt
e-mail: helene.hewitt@metoffice.gov.uk

A. B. Keen
e-mail: ann.keen@metoffice.gov.uk

D. J. Lea · A. J. McLaren
Met Office, FitzRoy Road, Exeter EX1 3PB, UK
e-mail: dan.lea@metoffice.gov.uk

A. J. McLaren
e-mail: alison.mclaren@metoffice.gov.uk

1 Introduction

The trend towards lower sea ice extent in the Arctic (Stroeve et al. 2007) has piqued interest in predictions of Arctic sea ice cover at seasonal timescales. Currently, the vast majority of sea ice predictions are forecast with the use of statistical models (Kim and North 1998, 1999; Tivy et al. 2007). A natural extension for coupled seasonal forecast systems would be to forecast sea ice extents, particularly forecasts of the critical September minimum in sea ice. Seasonal prediction is a rapidly advancing field that seeks to obtain maximum forecast skill given intertwined sources of predictability coming from initialization of all components (atmosphere, ocean, sea ice, soil moisture) of the coupled system (Kumar et al. 2007), as well as external sources of predictability such as solar (Woollings et al. 2010; Ineson et al. 2011) or volcanic (Marshall et al. 2009). Whilst accurate initial conditions and the subsequent correct dynamical evolution leads to enhanced predictability in the tropics, and to a lesser extent, in the extra-tropics, seasonal forecasts are affected by inherently unpredictable

internal non-linear variability, particularly within the atmosphere (James and James 1989). Of the many possible reasons for reduced predictability in the extra-tropics (Barsigli and Battisti 1998; Balmaseda et al. 2007, 2010; Fereday et al. 2012), recent interest has been drawn to the effect of polar sea ice on extratropical predictability (Francis and Vavrus 2012; Francis et al. 2009; Overland and Wang 2010; Budikova 2009). That work has suggested that sea ice, or more precisely, the absence of sea ice, can lead to changes in the extra-tropical circulation and opens up the possibility of enhanced predictability of the North Atlantic Oscillation (Defant 1924; Walker and Bliss 1932), or Northern Annual Mode (Thompson and Wallace 1998) when sea ice is properly initialized. Given the large trends in sea ice extent over the last few decades, predictability of the sea ice component of seasonal forecast systems could heavily influence the predictability of the system in general.

Meanwhile, numerous studies have shown that there is predictability of the sea ice at seasonal timescales by looking at the value of persistence (or lagged correlations) between sea ice extent in different months, or between sea ice extent and earlier estimates of ice volume, Arctic Ocean heat content, and atmospheric circulation indices. In particular, Lindsay et al. (2008) found a link between Arctic Ocean heat content and September sea ice extent at leads up to 9 months in the context of an ocean and sea ice analysis of the Arctic. Blanchard-Wrigglesworth et al. (2011a, b) on the other hand, found re-emergence of lagged correlation peaks between ice concentration and ice volumes at certain times of the year for differing lead times in the context of a free running coupled model experiment together with observational evidence. Whilst this is encouraging, these studies were both in the context of having perfect knowledge of the ocean and sea ice state. Given a scarcity of both ocean temperature observations in the Arctic regions and sea ice thickness, there is no guarantee that the correct dynamical information could be properly assimilated into an ocean and sea ice analysis to adequately initialize seasonal forecasts and utilize this potential predictability.

The use of initialized sea ice in operational seasonal forecast systems has not been widely exploited yet, and is often completely ignored. The sea ice component of coupled seasonal prediction systems ranges from not being represented at all; included as part of the coupled system, but not initialized to observations; to being fully initialized from a sea ice analysis using observations as part of the analysis system. The GloSea4 system is one of the first operational seasonal forecast systems to include proper sea ice initialization (by assimilating sea ice concentration) with subsequent dynamical and thermodynamical evolution of the sea ice. Of the contributors to the WMO Seasonal Forecast Producing Centres (http://www.wmo.int/pages/prog/wcp/wcasp/clips/producers_forecasts.html), there are only two other centres, the Canadian Seasonal to Interannual Prediction System (CanSIP) (Merryfield et al. 2012; Sigmond et al. 2013) and NCEP Climate Forecast System (CFSv2) (Wang et al. 2013), which include the initialization of sea ice to observations in their operational systems.

In this paper, we will give a brief summary of the sea ice initialization in the GloSea4 system, and then show how this has been used to forecast the September minimum Arctic ice extent. Further work is addressing how the ice initialization has enhanced the overall predictability of the GloSea4 system as well as investigations into the impact of ice initialization on atmospheric circulation. As it will be shown, despite some obvious deficiencies in being able to accurately initialize ice thickness (to date too poorly observed to be used in an ice analysis system), our forecast system has skill to forecast ice coverage, at least in an integrated fashion such as in the prediction of total Arctic ice extent. In what follows we describe the upgrade in November 2010 to include the assimilation of sea ice concentration in the Met Office seasonal prediction system, GloSea4. A brief description of the GloSea4 seasonal prediction system, including a detailed discussion of the assimilation of sea ice concentration in the GloSea4 ocean and sea ice analysis are given in Sect. 2. Section 3 assesses the GloSea4 sea ice analysis, while Sect. 4 examines the ability of the GloSea4 system to forecast Arctic sea ice. Conclusions are presented in Sect. 5.

2 Description of Met Office seasonal prediction system: GloSea4

2.1 Background

Although the Met Office Seasonal Prediction System, GloSea4, was implemented in 2009 (Arribas et al. 2011), this work will deal with a later upgrade to the system which introduced increased vertical resolution in both the ocean (to enhance the resolution of the diurnal cycle) and atmosphere (fully resolved stratosphere). This particular upgrade, implemented in October 2010, also introduced the assimilation of satellite observations of sea ice concentration into the ocean and sea ice analysis undertaken as part of the system, which was then used to initialize both the ocean and sea ice components of the coupled atmosphere/ocean/sea ice and soil moisture forecast model. Previous to this, the ocean and sea ice analysis was created using only ocean observations, with only the ocean component being passed onto the coupled model as initial conditions, the sea ice initial conditions being prescribed by a seasonally varying model climatology. The atmospheric and soil

moisture initial conditions in turn are taken from either the Met Office NWP analysis (Clayton et al. 2013) for the forecast, or the ECMWF interim atmospheric re-analysis (ERAI) (Dee et al. 2009) for the re-forecast, or hindcast.

The dynamical coupled atmosphere/ocean/sea ice/land model used within GloSea4 is HadGEM3 (Hewitt et al. 2011) version 3.0. This version has a horizontal resolution of approximately 120 km at mid-latitudes (N96) with 85 vertical levels in the atmosphere, and nominally 1° horizontal resolution with 75 vertical levels in the ocean [ORCA1 tripolar grid; Madec (2008)]. The results described in this paper correspond to the system and model configurations just described. It should be noted that another major upgrade was implemented in January 2013 to increase the horizontal model resolution to 50 km in the atmosphere and 0.25° in the ocean (GloSea5) (MacLachlan et al. 2014), with the ocean and sea ice analysis also being upgraded to a three dimensional variational system (FOAM V12) (Blockley et al. 2014).

2.2 The initial sea ice state: GloSea4 ocean and sea ice analysis

The ocean and sea ice initial states are provided by the ocean and sea ice analysis and are referred to as the GloSea4 Ocean and Sea Ice Analysis. The sea ice assimilation used in GloSea4 follows the same approach as that used in the Met Office Forecasting Ocean Assimilation Model (FOAM) (Storkey et al. 2010; Stark et al. 2008; Martin et al. 2007), but improved for the multi-category ice used in the HadGEM3 coupled model (Hewitt et al. 2011) through the sea ice model CICE (Hunke and Lipscomb 2010). McLaren et al. (2006) have shown that the sea ice physics as implemented in GloSea4 is capable of reproducing the observed mean state and variability of the sea ice. The FOAM system was upgraded in January 2013 to use CICE and the same sea ice assimilation scheme presented here (Blockley et al. 2014).

GloSea4 assimilates the sea ice concentration data from Scanning Multi-channel Microwave Radiometer (SSMR) and Special Sensor Microwave/Imager (SSM/I) reprocessed data provided by the EUMETSAT Ocean Sea Ice Satellite Application Facility (OSI-SAF), which is available from 1978 through 2007[1] (OSI-SAF 2011). From 2008 onward, a transition was made to the near real time SSM/I data quality controlled against a climatological ice concentration field. This is the same data used for the real time operational sea ice analysis (hereafter referred to as forecast analysis) that was performed on a daily basis between 1 Nov 2010 and 22 July 2013 in preparation for the coupled forecast described below in Sect. 2.3. Furthermore, the

reprocessed SSMR and SSM/I data were averaged over 25 adjacent points from the original 10 km by 10 km resolution for an effective resolution of 50 km by 50 km more suitable for the resolution of the nominally 1° ORCA1 configuration. The data assimilation techniques employed in the GloSea4 analysis require an assumption that observation error is uncorrelated. Input of a large number of gridded observations with correlated error could lead to overfitting of the data. The averaging process, by thinning the number of observations with correlated error reduces this possibility (Butterworth et al. 2002). No such averaging process is used with the near-real time data used during 2008 and 2009 of the hindcast analysis as well as during the GloSea4 forecast analysis, but it is not thought that this will have a large effect on the assimilation of the sea ice. A discernible discontinuity does exist in the hindcast between 2007 and 2008, which is related to the differing quality control process used in the OSI-SAF re-analysis and the OSI-SAF real time products.

The GloSea4 Ocean and Sea Ice Analysis assimilates concentration of sea ice by incremental analysis. Increments are calculated from estimates of observations minus background, from a first guess at appropriate time run of the NEMO/CICE system, using an anomaly correction method (Martin et al. 2007). Positive ice concentration increments are always added to the thinnest category of ice (consisting of ice up to 0.6 m thick) (Stark et al. 2008) while negative increments are first removed from the thinnest available category until it reaches zero concentration, and then progressively removed from thicker categories. When ice is removed, a volume of ice associated with the grid point average thickness of ice for that category and change in concentration is removed, while new ice is added with a thickness of 0.5 m which is thicker than the thickness of frazil ice (ice added due to freezing of sea water; 0.2 m) to prevent immediate melting of new ice. There is a lack of symmetry in this process, as ice is removed from the lowest available category (which might be high category thick ice), but only added to the 1st category ice (0–0.6 m) which leads to an inherent thinning of the ice by the assimilation process and can be seen as a bias in the system. If sea ice thickness data was available in real time, and could be adequately assimilated into the ice thickness field, this would likely produce a significant reduction in this bias.

2.3 Forecast and hindcast suites

The GloSea4 system runs both a forecast and a hindcast suite. The forecast suite is an ensemble seasonal forecast that is updated on a weekly basis. The hindcast suite is broadly a parallel version of the forecast for historical dates. One of the purposes of the hindcast is to provide a

---

[1] This data has since been reprocessed though 2009.

baseline for the system against observations, which allows bias corrections to be calculated. The forecast can be improved by the application of bias corrections.

In the forecast suite, atmosphere, land surface, ocean and sea ice initial states are calculated daily and two ensemble members are completed every day. Every Monday, a 42-member lagged ensemble is created by pulling together all forecast members available from the previous three weeks.

For consistency with the forecast, initial start-dates for the hindcast are spread throughout the month, but for simplicity, fixed calendar dates (1, 9, 17 and 25 of every month) are used. Initial states for these fixed calendar dates are calculated off-line. For the hindcast, the relevant start-dates from the off-line analysis is fed into the coupled model and a total of 42 hindcast simulations (three members for each year in the 1996–2009 period) are run every week.

3 Results from the GloSea4 sea ice analysis

The dataset presented here consists of the GloSea4 sea ice analysis performed for the period 1989–2009, extended through 2010 using a quasi-operation version of the GloSea4 ocean and sea ice analysis, and then from 2011 onward by the GloSea4 operational ocean and sea ice analysis. The GloSea4 operational analysis was terminated in July 2013, but we only show results through 2012 here. We shall refer to the 2010 through 2012 analysis as the forecast analysis to differentiate it from the hindcast analysis for the 1989–2009 period, primarily to identify a change in external forcing applied during this period. The forecast analysis uses direct flux forcing diagnosed from the Met Office Numerical Weather Prediction (NWP) analysis (Clayton et al. 2013) using observed SST, while the hindcast analysis uses interactive CORE bulk formula forcing (Large and Yeager 2009) with the atmospheric data from the ERAI analysis (Dee et al. 2009). This is very contrary to the purpose of maintaining consistency between hindcast and forecast, but was necessary due to data availability. It is possible that this difference in the external surface forcing could lead to a bias in the sea ice analysis between the hindcast and the forecast systems, which may particularly affect the ice thickness. However, the lack of overlapping data means that it is not possible to quantify the impact of this change. The correction of this inconsistency was one of the main considerations for future upgrades to the GloSea ocean and sea ice analysis system. In addition to the change in fluxes, the forecast analysis and the final 2 years (2008/2009) of the hindcast analysis transition to the real-time OSI-SAF SMM/I realtime sea ice
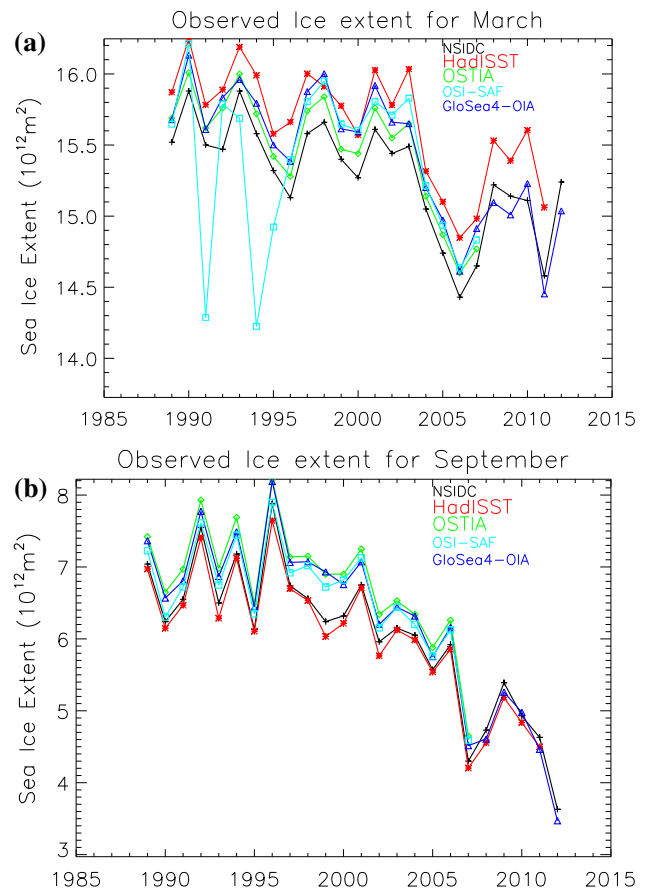
Fig. 1 Time series of Arctic ice extent for **a** March and **b** September from various analysis. Points on the NSIDC analysis (Fetterer et al. 2002, updated 2011) are denoted by *plus symbol* and *coloured black*, the HadISST analysis Rayner et al. (2003) are denoted by *asterisk* and *coloured red*, the OSTIA analysis (Roberts-Jones et al. 2012) are denoted by *diamond* and *coloured green*, the OSI-SAF values are denoted by *square* and *coloured cyan*, and our GloSea4 analysis are denoted by *triangle* and *coloured blue*. The OSI-SAF sea ice extents have been calculated using the same strict QC that was used for processing of data for the GloSea4 analysis, and in particular does not account for missing data over the pole. *Note* The low March values in the OSI-SAF ice extends in March 1991, 1994 and 1995 are due to missing observations, as the ice extent calculation does not account for this possibility

observations discussed in Sect. 2.2, along with SST observations from the Group for High Resolution SST (GHRSST) satellite products (AVHRR, AATSR, AMSRE). The hindcast analysis uses only the NOAA Pathfinder satellites for AVHRR SST observations (Casey et al. 2010). The change to GHRSST satellite products for the operation analysis was done in order to align ourselves as closely as possible with the 0.25° operation FOAM ocean analysis and make use of all the SST observations available to us, but still maintain as much consistency as possible with the hindcast (1989–2009) analysis.

3.1 Ice extent

The most directly measured, and for the case of the GloSea4 analysis, the only directly assimilated sea ice field, is the sea ice concentration. It should therefore not be surprising that sea ice extent is simulated with some skill by the GloSea4 analysis system. Ice extent is defined as the area occupied by grid point ice concentrations above 15 %. Figure 1 shows the time series of March and September sea ice extents from five separate analysis: The GloSea4 analysis, the Met Office Hadley Centre Ice and SST (HadISST) re-analysis (Rayner et al. 2003), the Met Office Operational Sea Surface Temperature and Sea Ice (OSTIA) re-analysis (Roberts-Jones et al. 2012), the National Snow and Ice Data Center (NSIDC) ice analysis (Fetterer et al. 2002, updated 2011), and the OSI-SAF (OSI-SAF 2011) re-analysis used as observations in this study. Note that no data filling has been used with the OSI-SAF analysis, so the calculated ice extent neglects the North Pole observation hole, and any other missing data. Thus the noticeable dips in March sea ice extent for the OSI-SAF product in 1991, 1994 and 1995 are due solely to missing data. Despite the missing polar hole (approximately $0.31 \times 10^{12}\,\mathrm{m}^2$), the OSI-SAF ice extents are very close to those quoted for the other analysis. This is due to the fact that the OSI-SAF ice extents are calculated as the monthly mean of ice extents from daily observed ice concentrations, which due to the non-linear nature of ice extent, gives higher values than the ice extent of monthly mean ice concentrations and seems to fortuitously compensate for the lack of consideration of ice over the North Pole. Although all the analysis differ subtly in the absolute value of ice concentration, by and large, the interannual variability is the same in all the analysis, with the GloSea4 analysis correlating with NSIDC at 0.94 in March and at 0.98 in September for the twenty year period 1989–2009.

The OSTIA and GloSea4 ice extents shown here are based on the OSI-SAF sea ice analysis with its underlying retrieval algorithm which uses a combination of the Bootstrap (Comiso et al. 1997) and Bristol (Smith 1996) algorithms. HadISST and the NSIDC ice extents are both broadly based on the Goddard Space Flight Center data set (Cavalieri et al. 1996, updated yearly, 1999) and its near real time equivalents which use only the Bootstrap algorithm. Differences between these two underlying data sets would therefore be expected. In principle, the only difference between the GloSea4 and OSTIA sea ice analysis and the OSI-SAF analysis on which they are based would be the dynamical nature of the GloSea4 analysis versus OSTIA's method of accounting for missing data and conflicting SST and sea ice data. However, there are differences in the quality control (QC) flags used to accept or reject the observations, with GloSea4 taking a stricter flag that excluded data obtained by the filling of the polar hole and

data obtained through a coastal correction method. Additionally, OSTIA included (A)ATSR satellite SST observations to augment the Pathfinder AVHRR SST observations. Given that the polar hole is successfully filled by the sea ice dynamics, these differences primarily seem to be manifest in coastal areas, with additional complications of differing land sea masks. OSTIA is on a 0.16° regular grid, although, what is shown in Fig. 1 has been regridded to a 0.25° grid consistent with the land mask used in the FOAM 0.25° global analysis. Thus the OSTIA results are largely identical to the GloSea4 results in September, but differ slightly in March when more sea ice can be found in coastal areas, but still much less than the differences between HadISST and NSIDC. Nevertheless, it is perhaps worth re-iterating, the differences in the value of ice extent are well within differences resulting from different methods of ice concentration retrieval and analysis.

One issue with the GloSea4 analysis is the change in bias introduced in 2008: While the year to year variability looks reasonable, the gain in ice extent from September 2007 to September of 2008 is not as large as in the other analysis. This is most likely due to a change in observing systems at the start of 2008, when a switch was made from the OSI-SAF re-analysis to the OSI-SAF realtime observations. The most likely cause of the change when the switch took place relates to how coastal ice is handled; after 2008 there is a reduction in the number of land points which are misidentified as coastal ice. It would appear that post 2008 the GloSea4 September analysis (Fig. 1b) is more in line with the HadISST analysis, whereas prior to 2008 it was slightly high. Similarly, the GloSea4 March analysis (Fig. 1a) appears to drop down to the lower NSIDC analysis after 2008. Judging solely by the difference between 2007 and 2008 in the various analysis, the GloSea4 value would appear to have a discontinuity downward of about $0.3 \times 10^{12}\,\mathrm{m}^2$ in September and $0.4 \times 10^{12}\,\mathrm{m}^2$ in March between 2007 and 2008. However, due to a lack of overlapping coverage period at the time when this analysis was performed, it was impossible to accurately quantify it. Subsequent work using the latest 0.25° GloSea5 sea ice analysis system suggests that the OSI-SAF re-analysis produces higher ice extents by about $0.75 \times 10^{12}\,\mathrm{m}^2$ in September and $0.45 \times 10^{12}\,\mathrm{m}^2$ in March, mostly due to coastal ice differences in the Canadian Archipelago, Baltic Sea, and Gulf of St. Lawrence. Due to the differing resolution, however, those numbers may not be completely appropriate here.

For the remainder of this paper, we will use the NSIDC (Fetterer et al. 2002, updated 2011) as the validation data set for sea ice extent, as this represents an independent and external source. However, for geographic details, such as the location of the ice edge, we will use the GloSea4 analysis.

Model: Climatology

Obs: Climatology
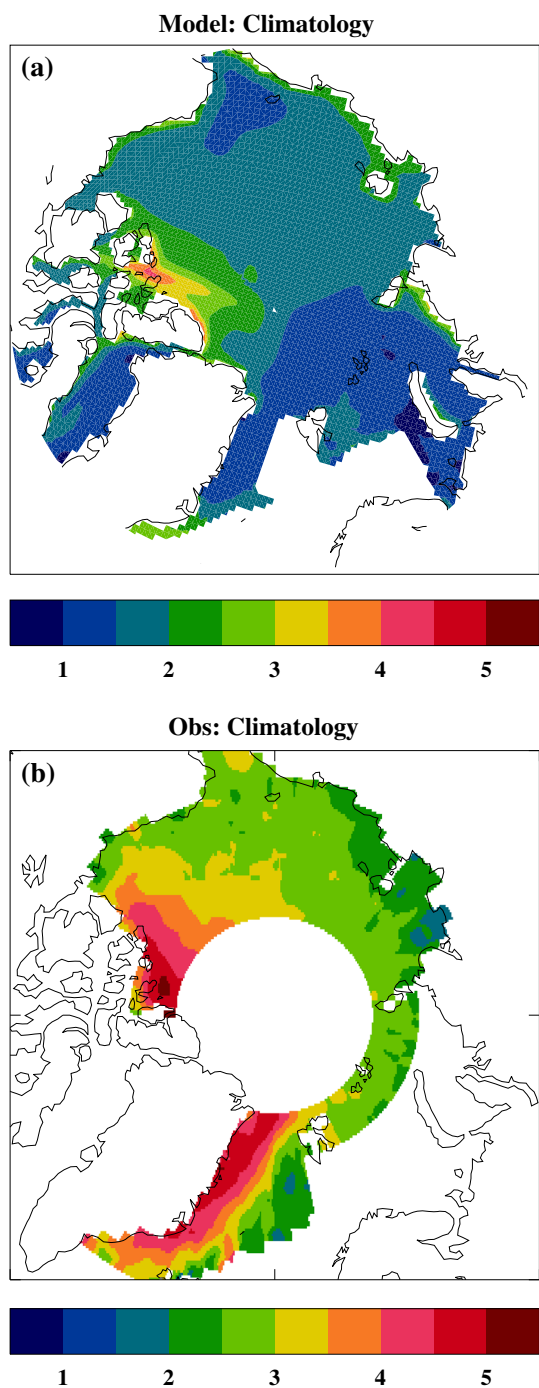
Fig. 2 GloSea4 analysis ice thickness (a) versus observational (Laxon et al. 2003) estimates of ice thickness (b) for winter (Jan–Mar) of 1994–2001. Thickness is in metres (m)

3.2 Ice thickness and volume

In order to assess sea ice thickness and volume, two products have been used; the radar altimetry ice thickness observations of Laxon et al. (2003) and the modelled ice volume from the PIOMAS system (Schweiger et al. 2011). Ice volume is not a directly measurable quantity, however,

the re-analysis estimates of volume in PIOMAS (which assimilates ice concentration but not thickness) have been well validated against independent ice thickness observations and are considered to be the best available estimates of ice volume.

The thickness of ice (and snow) in the GloSea4 sea ice analysis are purely prognostic variables. There are no assimilative constraints on ice thickness and therefore we rely solely on the ability of the system to correctly model its evolution. However, it is likely that some skill in simulating ice thickness is possible due to assimilation of the underlying ice concentration. Unlike the treatment in PIOMAS no consideration is given to the non-gaussian nature of sea ice concentration observations and therefore sea ice concentration observations near to the ice edge are weighted equally to sea ice observations inside the ice pack. Evidence suggests that treating ice observations within the ice pack equally to observations near the ice edge can have detrimental effects on the ice thickness (Lindsay and Zhang 2006).

Figure 2 shows the average sea ice thickness over the winter months (Jan–Mar) of 1994–2001 in the GloSea4 analysis and in the radar altimetry based estimates of Laxon et al. (2003). Clearly, the GloSea4 analysis is not thick enough during this period.

The model does correctly pack the ice over the Canadian Archipelago, but does have trouble with having ice thick enough to survive passage through the Fram Strait and southward along the east coast of Greenland.

Figure 3a shows the (1989–2009) climatological seasonal cycle of ice volume in the GloSea4 analysis and PIOMAS. GloSea4 has significantly lower volumes than PIOMAS especially in the summer months. Also shown on Fig. 3a are the 2011 ice volumes. Due to the already depleted summer volumes in the GloSea4 analysis, there is significantly less difference between the 1989–2009 climatology and the 2011 ice volume in GloSea4 than there is in PIOMAS.

Figure 3b, c shows estimates of monthly averaged March/September Arctic ice volumes in GloSea4 and PIOMAS. As Fig. 3a suggests, estimates of ice volume are considerably below those of PIOMAS for all years. In particular, the $0.50 \times 10^{12} m^3$/year downward trend in September ice volume found in PIOMAS from 1992 to 2012 is reduced to only $0.12 \times 10^{12} m^3$/year in the GloSea4 analysis. A smaller reduction in the trend is seen for March reducing from $0.40 \times 10^{12} m^3$/year in PIOMAS to $0.31 \times 10^{12} m^3$/year in the GloSea4 analysis. The period 1992–2012 was chosen to eliminate an approximate three year spin-up of the ice volume from the initial conditions at the start of the ocean and sea ice analysis run. The inter-annual variability is fairly well modelled with correlations
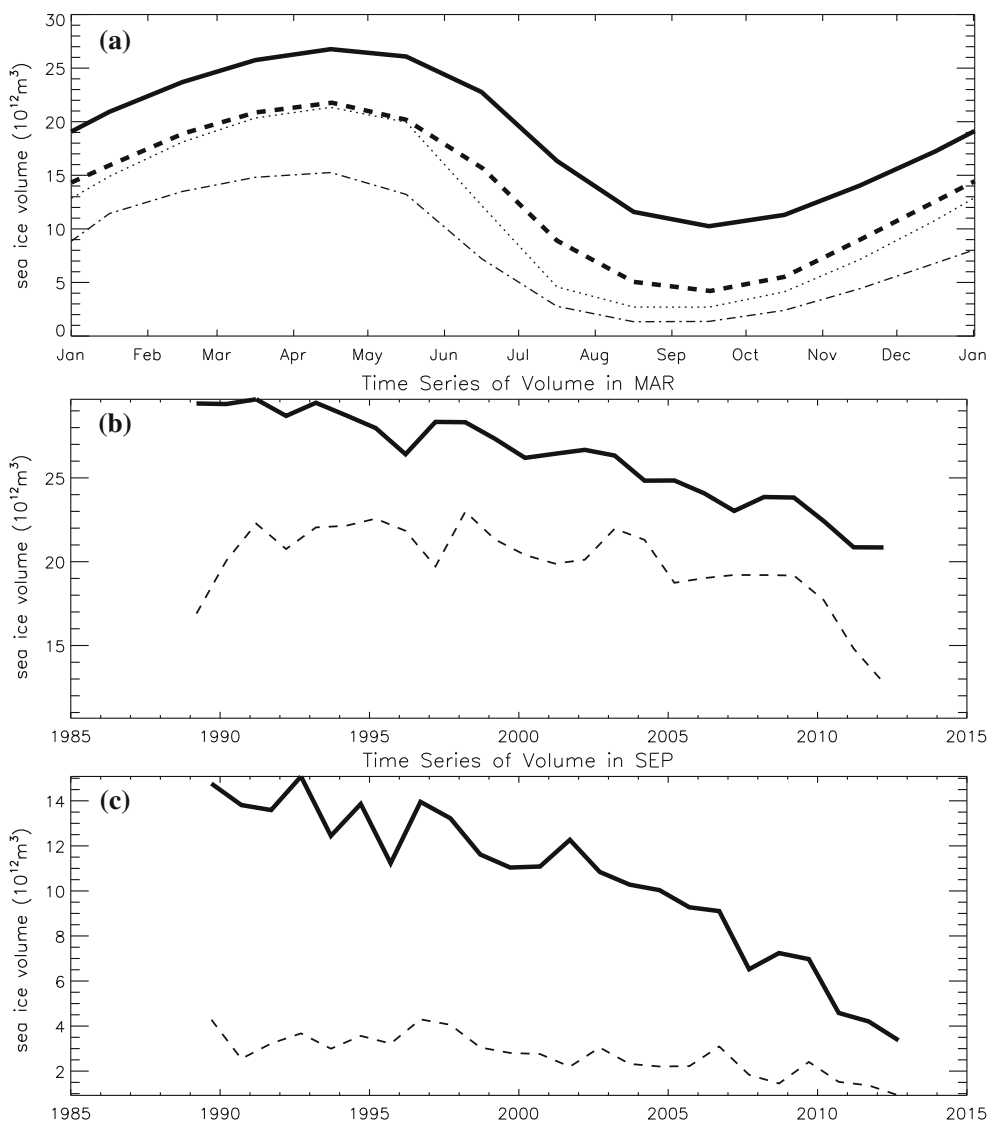
Fig. 3 a Seasonal cycle of northern hemisphere sea ice volume in PIOMAS (*thick solid line*) and GloSea4 analysis (*thin dotted line*). Also plotted are the seasonal cycle for 2011 in PIOMAS (*thick dashed line*) and the GloSea4 analysis (*thin dash-dotted line*). Time series of northern hemisphere sea ice volume in PIOMAS (*thick solid line*) and GloSea4 analysis (*thin dashed line*) for **b** March, **c** September

between the 1989–2009 GloSea4 and PIOMAS monthly anomalies of 0.38 for March and 0.77 for September and correlations of the detrended monthly anomalies of 0.36 for March and 0.47 for September. Eliminating the spin-up period and considering the 1992–2012 period the correlations go up to 0.85 for March and 0.88 for September, with detrended correlations of 0.53 and 0.48.

4 Sea ice concentration predictive skill

4.1 September ice minimum prediction

We will now investigate the performance of our 1996–2009 hindcasts (or re-forecasts) to evaluate the performance of the seasonal prediction system's ability to accurately forecast September ice extent. These hindcasts are important from the standpoint of measuring skill— shown below through the anomaly correlation coefficient, but also necessary to calibrate the system biases.

Figure 4 shows the time series of September ice extent forecasts from late March start dates in the hindcast and forecasts. The 1996–2009 hindcasts are 6 ensemble members each from 17, 25 March and 1 April start dates, and the 2011 and 2012 forecasts have 42 ensemble members initialized from 12 March to 1 April. The climatological ice extent of the hindcast is $6.8 \times 10^{12}\,\mathrm{m}^2$, which is approximately 10 % above the NSIDC climatological average for this period of $6.1 \times 10^{12}\,\mathrm{m}^2$, and the GloSea4 hindcast ice analysis average of $6.3 \times 10^{12}\,\mathrm{m}^2$. This bias has been
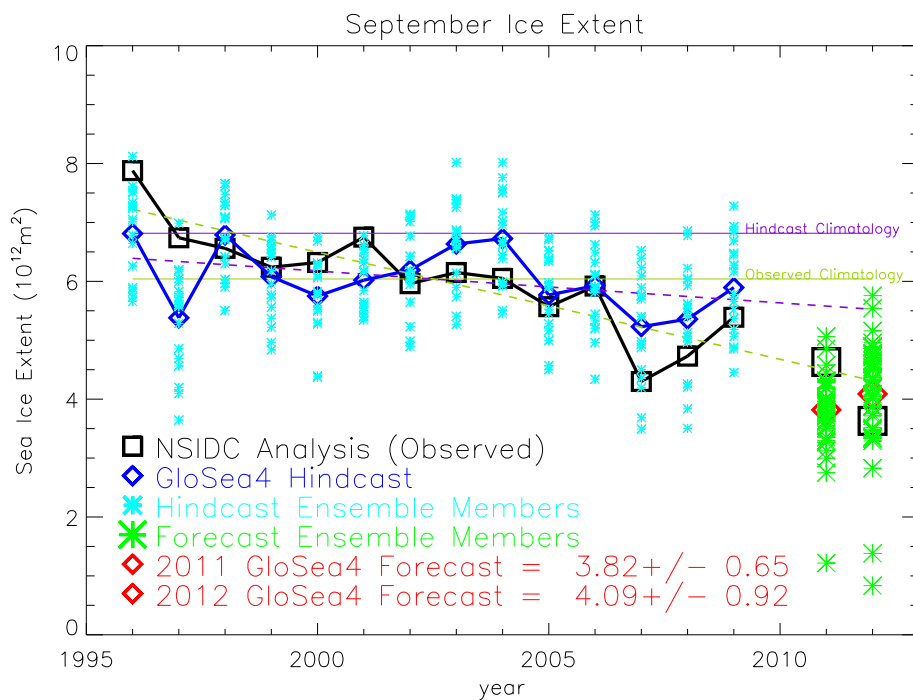
Fig. 4 Bias corrected forecast of the September ice extent in the hindcast *cyan* (*asterisk*). 2011 and 2012 forecast members are slightly larger *green asterisk's*. The ensemble mean ice extent are labelled with a *diamond* and joined with a *solid blue line*. The *square's* joined by a *solid black line* are the observed ice extents from NSIDC (Fetterer et al. 2002, updated 2011). The observed (*yellowish green*) and hindcast (*magenta*) climatologies are both denoted by the labelled *horizontal lines* on the plot, and thus the *vertical* separation between the two lines (0.8×10^{12} m^2) is the amount by which the forecast has been bias corrected downward. Further *dashed lines* are indicative of the trends in the observations (*yellow green*) and forecast (*magenta*)

accounted for in the combined hindcast and forecast time series by subtracting 0.7×10^{12} m^2 (indicated by the vertical separation between the climatologies in Fig. 4) to put their climatological values in line with the NSIDC analysis. The fact that the forecast gives an accurate estimate of September ice climatologies when initialized with the March ice concentration and implied thickness, seems to suggest that the coupled model can accurately integrate the ice forward in time given the initial concentration and adequate ice thickness.

The 1996–2009 hindcast has a correlation of 0.62 with the NSIDC (Fetterer et al. 2002, updated 2011) observed September ice extents also plotted in Fig. 4. This decreases slightly to 0.56 if trends are removed from the timeseries. These correlations are significantly different from zero at the 95 % c.l. after accounting for serial correlations. The correlations with the GloSea4 analysis are 0.62 and 0.63 detrended. Given the smaller gain in ice extent for 2008 over 2007 for both the GloSea4 forecast value and the GloSea4 analysis value as compared to the NSIDC estimate, it would appear the forecast is integrating forward the smaller ice extents that are seen in the analysis post 2008. This might explain the slightly better detrended correlations against the GloSea4 analysis versus the NSIDC analysis. It would furthermore suggest that the

subsequent forecasts for 2011 and 2012 would also be biased somewhat low compared to other estimates—the work with the GloSea5 system, suggesting it could be nearly as much as the forecast has been biased corrected downward. Finally, it appears from the time series that the hindcast is able to better capture the year-to-year variability in the sea ice extent than it is able to capture the trend. The trend during the hindcast is 0.5×10^{12} m^2/decade as opposed to a much larger 1.8×10^{12} m^2/decade in the NSIDC analysis ice extents, or the 2.2×10^{12} m^2/decade in the GloSea4 analysis ice extents over the same 1996–2009 period.[2] Since the GloSea4 analysis already has low ice volumes (see Fig. 3), there is only a weak trend in the time series of ice volume in the hindcast compared to estimates from PIOMAS. This results in a much smaller trend in the ice extent as the amount of ice area depletion in the summer is very dependent on the underlying ice thickness with thinner ice more easily lost (Notz 2009). Due to the already small low ice volumes in GloSea4, there is not much change in the summer ice melt over time. In spite of this, the significant skill in capturing the interannual variability

---

[2] The GloSea4 analysis and NSIDC trends are closer together if one considers only the 1996–2007 period where GloSea4 has a consistent set of sea ice observations.

Fig. 5 Plume plots of northern hemisphere ice volume in 2012 from the 12 March to 1 April start dates. The *thick line* is the daily PIOMAS (Schweiger et al. 2011) ice volume with the monthly averaged values overlaid with *diamond symbols*. The *thinner solid line* is the daily GloSea4 sea ice analysis with the monthly averaged values overlaid with *square symbols*. *Dotted lines with asterisk symbols* are the monthly averaged sea ice volume from each ensemble member of the GloSea4 forecast between 12 March and 1 April
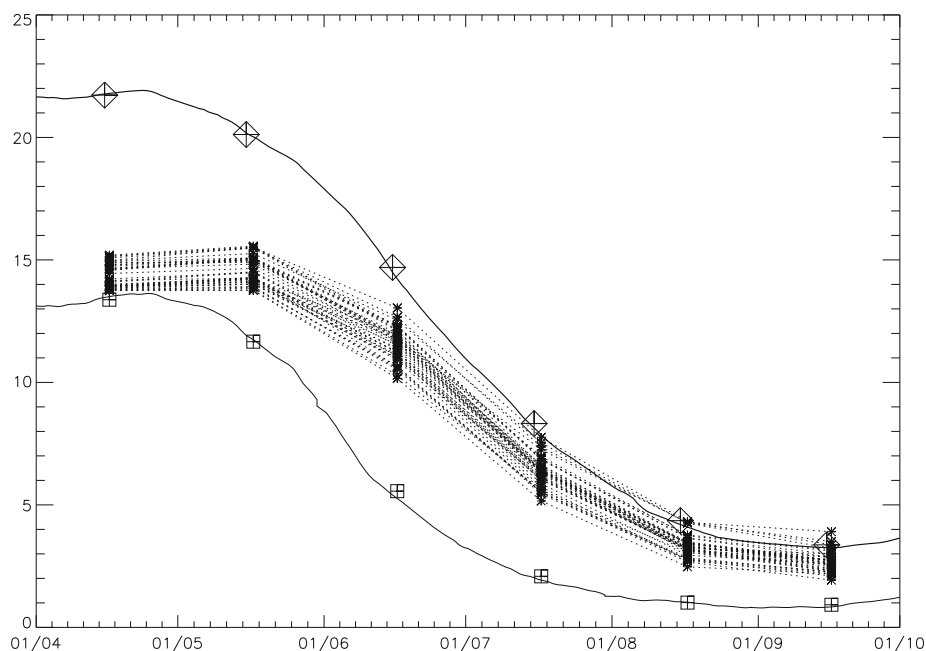

indicates that GloSea4 is adequately capturing the variability of the climatic patterns in the Arctic.

Plume plots of ice volume (Fig. 5) demonstrate that much of the degradation in ice volume is due to the ice analysis (initialization), as the forecast plumes regularly show a better realization of volume (as compared with PIOMAS) than does the analysis. As the coupled model has a better sea ice volume climatology than the externally forced analysis, this represents a drift from the forced analysis to the coupled model climatology.

After subtracting the mean ice extent bias correction of $0.8 \times 10^{12}\,\text{m}^2$, the ensemble mean ice extent forecasts for 2011 and 2012 are $(3.8 \pm 0.65) \times 10^{12}\,\text{m}^2$ and $(4.1 \pm 0.92) \times 10^{12}\,\text{m}^2$ respectively compared to the NSIDC observed values of $4.6 \times 10^{12}\,\text{m}^2$ and $3.6 \times 10^{12}\,\text{m}^2$, respectively. The forecast values are the mean value of the individual ensemble member ice extents (rather than the ice extent of the ensemble mean ice concentration) and the quoted errors are the standard deviation of the ice extents of the individual ensemble members. Although the 2011 and 2012 ensemble mean forecasts do not capture the dramatic drop in the ice cover between 2011 and 2012, it should be noted that both observations fit well inside the main envelope of the ensemble predictions, with the observed 2012 value falling inside the second lowest category of five equal quintile probabilities and the 2011 forecast falling inside the highest quintile category. Specifically, the closest bias corrected ice extent to the observed value amongst the 2012 ensemble members corresponds to the 10th ranked (smallest to largest) value of 42, whilst the closest bias corrected ice extent to the 2011 observed value ranks 36th of 38. The ensemble range is

larger for the 2012 forecast than for the 2011 forecast, which may be indicative of a further destabilization of the ice cover leading up to the initialization of the forecast in March 2012, but further investigation outside the scope of this paper is required to confirm this.

The choice of March as the initialization date for the September forecast was chosen due to an apparent degradation of forecast skill as the initialization date moved further into the melt season. This degradation is due to the ice analysis producing ice which is too thin. Thin ice leads to more rapid loss of ice area, as the same amount of heat input with identical loss of ice volume will lead to a larger loss of ice area and extent for thinner ice (Notz 2009). Estimates of correlation skill for September in the hindcast are shown in Fig. 6a, while the amount of bias correction for September, along with the raw model and bias corrected ice extent forecasts are shown in Fig. 6b. The correlation skill in the detrended hindcast (Fig. 6a asterisk symbols) quickly decreases as the amount of bias correction (6b triangle symbols joined by dashed line) increases and the uncorrected forecast (Fig. 6b asterisk symbols joined by dotted line) decreases. As the ice extent in the uncorrected 2012 forecast decreases to unphysically small amounts, much of the intrinsic variability is lost which removes any skill in the system. Indeed from June initialization dates onward, the bias correction dominates the forecast of ice extent. This is not only a problem for the sea ice forecast, but, given the evidence suggesting large scale atmospheric circulation effects from low sea ice (Francis et al. 2009; Francis and Vavrus 2012), it poses a potential problem for extratropical predictability of the system.
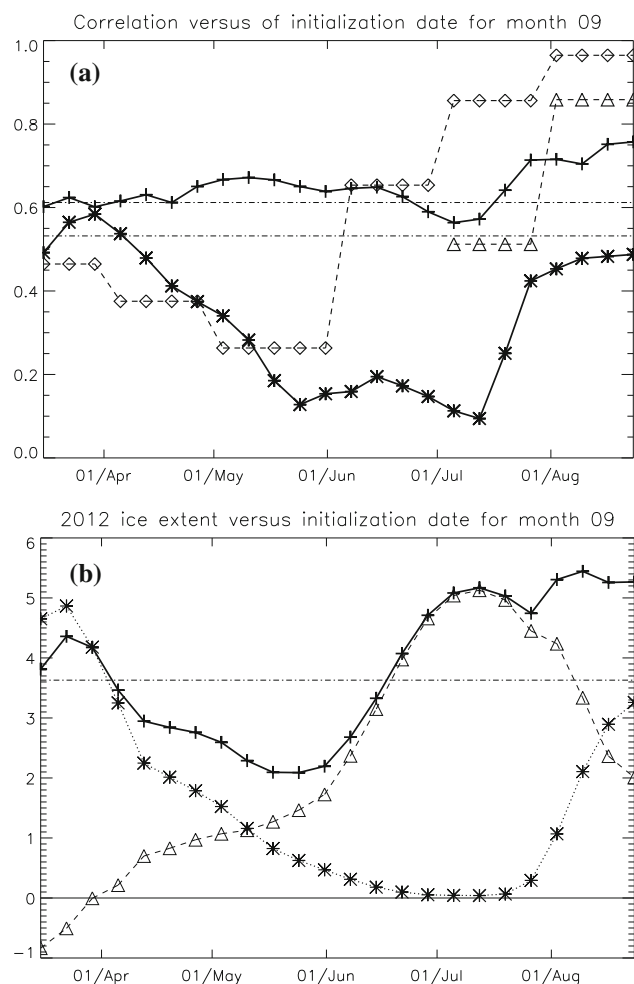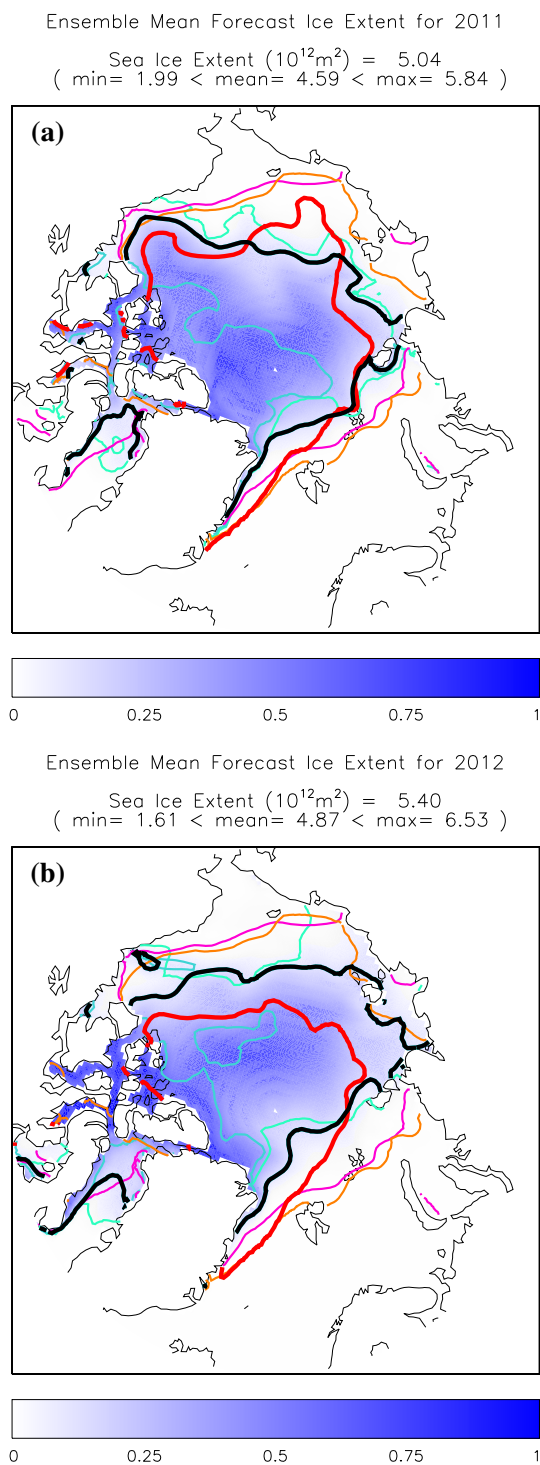
Fig. 6 a Correlations between hindcast and observed (NSIDC) September ice extent as a function of forecast start date (*solid line with plus symbols*). Also shown are the detrended correlations (*solid line with asterisk symbols*) and the correlations for a persistence forecast (*dash line with diamond symbols*; *triangle symbols* are positive values for detrended persistence). The *horizontal dashed-dotted lines* are the 95 and 98 % confidence levels that the correlation is non-zero respectively. b The 2012 September bias corrected sea ice extent forecast as a function of start date (*solid line with plus symbols*) along with the the uncorrected forecast (*dotted line with asterisk symbols*) and bias correction term (*dashed line with triangle symbols*). The 2nd point in both plots represents the forecast being presented here for a start date centred on 22 March (12 March to 1 April). This forecast has a slightly smaller detrended correlation and slightly larger bias correction than the forecast centred on 29 March, and a larger (post season) error then the forecast centred on 5 April. The *horizontal dashed-dotted line* is the 2012 observed extent

Figure 7 shows the ice extent of the ensemble mean ice concentration of the two "true"[3] ice forecasts for September of 2011 and 2012 initialized between 12 March and the 1 April. Individual plots of all the ensemble members that constituted this ensemble mean are shown in supplementary figures 1 and 2 for 2011 and 2012 respectively.

---

[3] In the sense that these forecasts were performed prior to September 2011 and 2012.

Note that the area of this ensemble mean ice extent will not be identical to the ensemble average of the area of extent for each member owing to the non-linear nature of the extent (0.15 cutoff). In general, the ice extent of the ensemble mean ice concentration will be slightly larger then the ensemble mean of ice extents by approximately 10 %. The same effect can also be seen in monthly averages, with the monthly average of daily ice extent being larger than the ice extent of the monthly average ice concentration—one of the reasons why the OSI-SAF ice extents (monthly average of daily values) in Fig. 1 were in line with other estimates of ice extent, despite omitting ice not observed due to the polar hole in the observing system of $0.3 \times 10^{12}\,\mathrm{m}^2$. Also included on the graph is the observed GloSea4 analysis ice extent in red, the 1996–2009 climatological value of the GloSea4 analysis in orange, and the 1996–2009 climatological value of the GloSea4 forecast in magenta. Finally, the ensemble members with the maximum and minimum extent (by area) are plotted in cyan, showing the large variation in this quantity. This is further quantified in supplementary figures 1 and 2 showing the ice edge of each ensemble member. The contraction of ice edge on 1 April compared to late March values was due to an error in the monthly averaging program, which would have included an increasing number of daily October ice concentrations into the September average value for late March start dates. In both 2011 and 2012, the one characteristic of the extent not captured is the observed southward extension of the ice edge down the eastern coast of Greenland. This is also seen in the relationship between climatology of the hindcast (magenta line in Fig. 7) and the observed climatology (orange line in Fig. 7). In fact, the observed southward progression of the ice edge in both 2011 and 2012 is virtually identical to that seen in climatology. Given the thinness of the ice along the Greenland coast, it is doubtful that it can be successfully advected far enough south to match the observed climatology.

The 2011 ensemble mean September ice concentration forecast appears to fairly closely resemble the observed value, but the September 2012 forecast overestimates the observed ice concentration, particularly in the eastern Arctic. In 2011 (Fig. 7a), the range of possible extents, as shown by the minimum and maximum ensemble members, as well as in the supplementary material, envelopes the observed value. This is also true in 2012 (Fig. 7b), although biased to the low extent end of the range. Note, however, that the ice concentrations are not bias corrected, as it is problematic to correct non-guassian fields such as ice concentration. The forecast member from 31 March (ensemble member #34) in 2011 (Supplementary Figure 1) and the forecast member from 18 March (ensemble member #13) (Supplementary Figure 2) have bias corrected ice

Ensemble Mean Forecast Ice Extent for 2011

Sea Ice Extent $(10^{12}m^2)$ = 5.04
(min= 1.99 < mean= 4.59 < max= 5.84)

(a)

Ensemble Mean Forecast Ice Extent for 2012

Sea Ice Extent $(10^{12}m^2)$ = 5.40
(min= 1.61 < mean= 4.87 < max= 6.53)

(b)

◄**Fig. 7** Plot of ensemble mean ice concentration with the black line being threshold for ice extent (ice concentration = 0.15) in **a** 2011 and **b** 2012. *Note* The area enclosed by the *black line*, the ice extent of the ensemble mean ice concentration **a** 2011: 5.04×10^{12} m² **b** 2012: 5.40×10^{12} m²), will not be identically equal to the ensemble mean of the individual member's ice extents **a** 2011: 4.59×10^{12} m² **b** 2012: 4.87×10^{12} m² – before bias correction of 0.78×10^{12} m²) which are plotted in Fig. 4 owing to the non-linear (threshold) nature of ice extent. The *red line* is the observed ice extent as determined by the GloSea4 ice analysis. The two *cyan lines* are the ice extents of the ensemble members with minimum and maximum ice extents. The *magenta* and *orange lines* are the climatological mean ice concentration extents between 1996–2009 for the hindcast and analysis respectively. Individual ensemble members can be seen in the supplementary figures 1 (2011) and 2 (2012). The ice extents quoted at the *top of the figures* have not been biased corrected

August (see http://www.nasa.gov/topics/earth/features/arctic-storm.html and http://nsidc.org/arcticseaicenews/2012/08/a-summer-storm-in-the-arctic), the GloSea4 system is able to capture the full range of the natural variability in the climate system.

4.2 Skill in ice extent prediction throughout the year

Although the September ice extent minimum is of primary interest, how the system performs throughout the year is also important, particularly since that might have an impact on the prediction skill for patterns of atmospheric teleconnections related to ice extent (Francis et al. 2009; Francis and Vavrus 2012). It is also important from the aspect of choosing the best dates for testing the forecast system: Often seasonal forecast systems are initially tested using only sample starts dates in November (for winter forecasts) and May (for summer forecasts), while additional start dates in February (spring) and August (autumn) might also be tested. These may not represent the most skillful months to initialize the model for all interesting forecast variables. Figure 8 shows a quantification of forecast correlation skill as a function of target month and lead time, and Fig. 9 shows a similar quantification of forecast skill as a function of start date and lead time. In both figures, the top plot (a) is correlation between detrended NSIDC observations and detrended forecast, while the bottom plot (b) is the correlation before trends are removed (full anomalies). Thus Fig. 8a is directly comparable with figure 1 of Merryfield et al. (2013) and figure 5a of Wang et al. (2013), while figure 8b is directly comparable to figure 2b of Sigmond et al. (2013) and figure 5b of Wang et al. (2013), although restricted to the 6 month model runs of the GloSea4 operational setup. These figures show the long lead (6 month) correlation skill for the target months of July through September and December through March (with a degradation in January centered around 5.5 months, corresponding to 1 August start dates when the

extents that most closely match the observed value for 2011 and 2012 respectively. In both cases, there is too little ice on the Atlantic side of the Arctic and too much ice on the Pacific side. The fact that the ensemble spread does roughly incorporate the observed value demonstrates, that while the ice extent of 2012 was undoubtedly partially due to events such as the extreme Arctic storm that appeared in
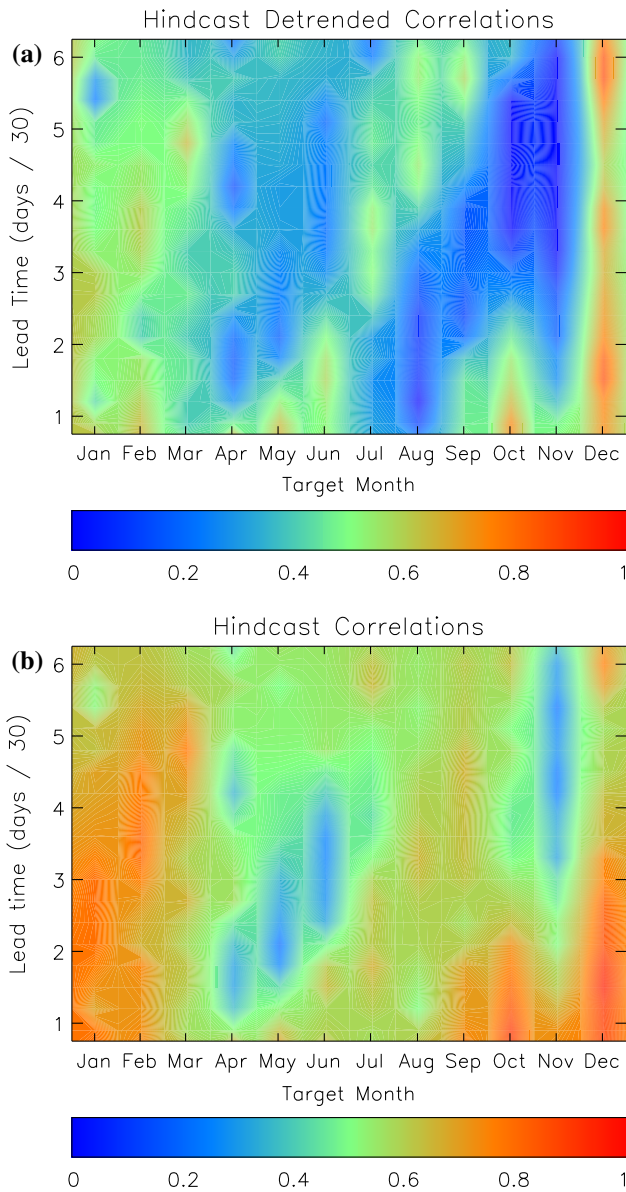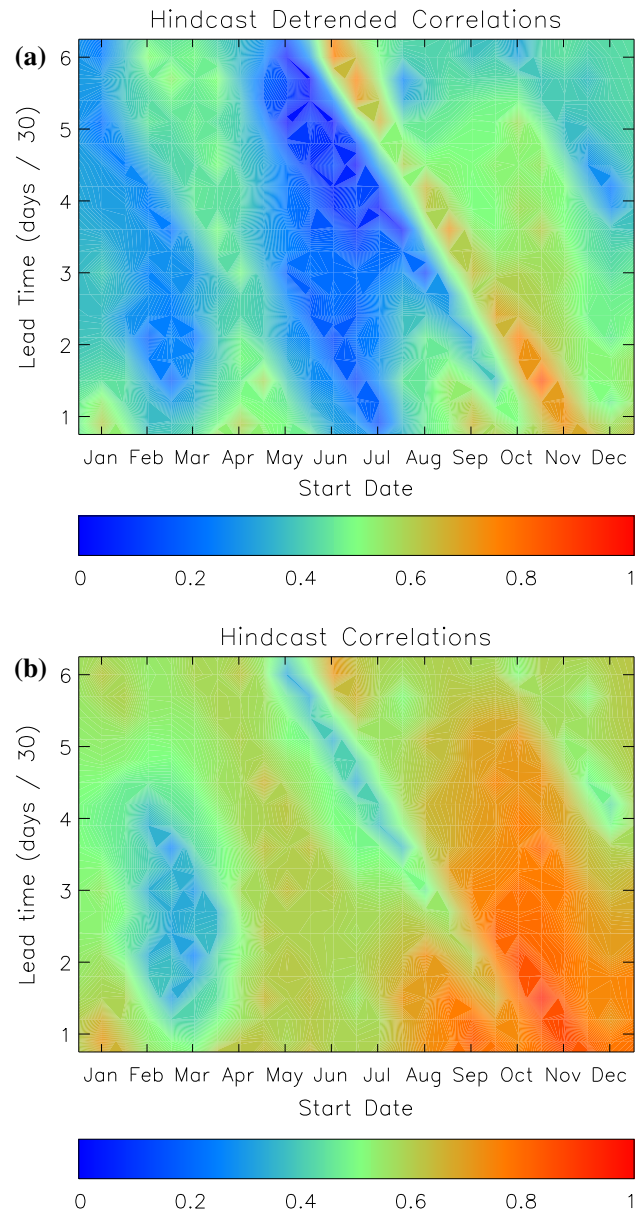
Fig. 8 Correlation skill as a function of target month (*horizontal axis*) and lead time (*vertical axis*). Correlations above 0.6 would be significantly different from zero at the 95 % confidence level. Correlations between detrended forecast and NSIDC observations are given in **a** (*top*) while **b** (*bottom*) shows correlations of the full anomalies prior to trend removal. *Note* Small sub-month scale features are due to errors introduced by regridding and interpolation of irregularly gridded data and should be disregarded

Fig. 9 Correlation skill as a function of start date (*horizontal axis*) and lead time (*vertical axis*). Correlations above 0.6 would be significantly different from zero at the 95 % confidence level. Correlations between detrended forecast and NSIDC observations are given in **a** (*top*) while **b** (*bottom*) shows correlations of full anomalies prior to trend removal. Note regarding small sub-month scale features from Fig. 8 applies here too

biases in the initial sea ice thickness field are at its largest. December in particular is substantially skillful at all lead times up to 6 months. In all cases, the full anomalies (Fig. 8b) give higher correlation values then the detrended anomalies (Fig. 8b). This is due to the strong trend in the observations as pointed out in Sigmond et al. (2013), despite the fact that GloSea4 forecast values do not show near as large a trend, at least for the September ice extents

shown in Fig. 4. It should also be noted that although the full anomalies have enhanced correlations, they may not be any more significantly different from zero correlations than the detrended correlations, owing to the significant autocorrelation of the time series inherent with a strong trend. The target months of July through September are interesting in that the shorter lead time forecasts are noticeably less skillful for the detrended correlations (Fig. 8b), again

owing to the too thinly initialized ice during the summer months. In contrast, the months April through May in the spring and October and November in the autumn are inherently not skillful at all but the shortest lead times. The issue of which initialization months are best for skillful predictions can be seen in Fig. 8 by tracing back along particularly high correlation diagonals, but is much better seen directly in Fig. 9 which shows skill as a function of start date. It is apparent from both those figures that March and April in the spring, along with October and November in the fall are both particularly good months to initialize the GloSea4 system for the prediction of sea ice. This is seen both in the detrended correlations (Fig. 9a) and in the full anomaly correlations (Fig. 9b). Conversely, January and February, along with May through July (except for the December predictions at the longer lead times) are particularly poor months to initialize the system. This is particularly relevant, as seasonal forecast systems are traditionally tested using initialization dates relevant for each of the four seasons, namely 1 November, 1 February, 1 May and 1 August. Of these, only the November start date is particularly skillful for ice prediction in GloSea4.

Both Sigmond et al. (2013) and Wang et al. (2013) found similar increases in predictive skill as demonstrated in figure 1 of their combined ensemble (Merryfield et al. 2013), but at subtlety different times to those seen here. The CanSIP [Fig. 2b of Sigmond et al. (2013)] and CFSv2 [Fig. 5b of Wang et al. (2013)] systems see much better full anomaly (non-detrended) correlation than those seen in our Fig. 8b. This is likely due to the lack of strong trends in ice extent forecasts seen in GloSea4 which result from the too thin ice analysis. For correlation skill of detrended anomalies, the results are much more comparable. They too found enhanced predictability of target months in the autumn and winter, the increase in September, but more so October, predictability was roughly the same in the both the CanSIP and the CFSv2 systems (Merryfield et al. 2013) peaking one month later than in the GloSea4 system. More notably, they did not see a large drop off in September and October predictability when initializing in the summer months. Again, this was probably due to the particularly bad performance of the GloSea4 sea ice analysis during the summer, with ice volume being significantly below other estimates (Schweiger et al. 2011). Their increase in winter predictability is also later than that in the GloSea4 system, occurring in January through March, without the large increase in December predictability seen in the GloSea4 system. Bitz et al. (2005) speculate the enhanced winter predictability is due to the location of the winter sea ice edge being closely related to the convergences of ocean heat fluxes with long associated time scales, and thus relatively predictable. In terms of start months, Merryfield et al. (2013) found enhanced interannual predictability as

demonstrated in their plots of detrended correlations (their figure 1) stemming from initializing the system in March for the CanSIP system and May for the CFSv2 system, and for both systems when initialized in November. The peak in predictability for initialization in March is in line with an enhanced value of March damped persistence as documented in Merryfield et al. (2013) and also seen in our Fig. 6. The enhanced predictability seen in the CFSv2 system for May, but not seen in either the CanSIP system or in the GloSea4 system described here, actually precedes the peak value of damped persistence, which appears to peak for July start dates. However, Chevallier et al. (2013) also saw increased predictability for May start dates. This May start date predictability is notable in that it appears to extent out through the maximum nine month lead time available in the CFSv2 system. This peak is definitely not seen in the GloSea4 system, and if anything June and July represent a low in start date predictability, owing to the poor summer ice volume seen in the GloSea4 system. Finally, an increase in November start date predictability is seen in all three systems, although it may be slightly later, more towards December, in the CanSIP and CFSv2 sytems. This increased late autumn, early winter start date predictability seems to substantially exceed the value of damped persistence for these start dates. Again, Chevallier et al. (2013) also sees good predictability for November start dates. All the systems seem to show a link between properly initializing the maximum ice extents and being subsequently able to forecast the next minimum, which is supported by the large value of March damped persistence (Merryfield et al. 2013). Being ensemble prediction systems, the GloSea4, CanSIP and CFSv2 systems will be better able to capture events that are cumulative in nature, such as the sea ice maximum and minimums, where the probabilistic events happening over long periods can be properly integrated (Hasselmann 1976). Transitional seasons, which can be largely influenced by a single event can only be readily forecast in a probabilistic sense, but not in the ensemble mean.

5 Summary and conclusions

With a decreasing trend in the amount of sea ice extent in the Arctic over the past decades (Stroeve et al. 2007), the accurate prediction of the minimum sea ice extent several months in advance has many societal and commercial implications. Seasonal prediction systems, with their ensemble prediction methods, offer some of the best hope to accomplish this in a manner that correctly portrays the probabilistic nature of the climate system on these time scales. To date, many seasonal forecast systems do not include dynamical initialization and forward time

integration of the sea ice component of the climate system. The work described here using the GloSea4 system represents one of the first attempts to do this, producing comparable results to the two other known operational systems that have sea ice initialized from observations (Sigmond et al. 2013; Wang et al. 2013; Merryfield et al. 2013), as well as other experimental systems (Chevallier et al. 2013; Guemas et al. 2014).

Investigation of the GloSea4 sea ice analysis and sea ice forecast has demonstrated the usefulness of properly incorporating initialization and integration of sea ice into the system. The system shows a remarkable ability to accurately forecast the September sea ice extent throughout the hindcast period (1996–2009) when forecast from late March and early April start dates (a 6 month lead time), with correlations with the observed extent being 0.62, decreasing only slightly to 0.56 when removing the trend and considering only the interannual variability. While two years of forecast experience does not allow for a quantifiable identification of skill in the actual forecast, both observed sea ice extents of 2011 and the extreme low of 2012 fell inside the envelope of ensemble predictions, and thus within the range of possibilities exhibited by the system.

While the sea ice analysis has deficiencies in its ability to prognostically determine the sea ice thickness, the analysis is able to provide a suitable starting point for the forecasting of the sea ice environment. The issues with the analysis are likely due to a number of factors; the assimilation of ice concentration, the atmospheric forcing used to produce the analysis and underlying biases in the ocean-ice model. Despite all the deficiencies documented in this paper, initialization of the sea ice has both improved the seasonal forecasting of ice and indeed has improved seasonal predictability in general (see Maidens et al. 2012; MacLachlan et al. 2014 as an example).

Further investigation of the predictability of sea ice in the system throughout the year has also been undertaken. This predictability has potential implications for the large scale atmospheric circulation (Francis et al. 2009; Francis and Vavrus 2012). Although much of this predictability will depend on the performance of the sea ice analysis system at the time of initialization, the GloSea4 system shows an increased predictability for July through September target months from late March early April start dates, and general increased predictability for the winter months, specifically for December, and particularly for the start dates in October and November. The results of Merryfield et al. (2013) for the CanSIP and CFSv2 systems are broadly similar, but do differ subtly, presumably due to the differing biases in the systems, although all are broadly consistent with the value of persistence at the start date. This could have consequences for the testing of model performance, as not all systems perform equally for the traditional summer seasonal initialization on 1 May. Furthermore, despite an enhanced value of persistence in February, no system has particularly good forecast skill of ice extent for this traditional spring start date.

References

Arribas A, Glover M, Maidens A, Peterson K, Gordon M, MacLachlan C, Graham R, Fereday D, Camp J, Scaife AA, Xavier P, McLean P, Colman A, Cusack S (2011) The GloSea4 ensemble prediction system for seasonal forecasting. Mon Weather Rev 139(6):1891–1910. doi:10.1175/2010MWR3615.1

Balmaseda M, Anderson D, Vidard A (2007) Impact of Argo on analyses of the global ocean. Geophys Res Lett 34(6):L16605. doi:10.1029/2007GL030452

Balmaseda M, Ferranti L, Molteni F, Palmer T (2010) Impact of 2007 and 2008 Arctic ice anomalies on the atmospheric circulation: implications for long-range predictions. QJRMS 136(652):1655–1664. doi:10.1002/qj.661

Barsigli J, Battisti D (1998) The basic effects of atmosphere–ocean thermal coupling on midlatitude variability. J Atmos Sci 55:477–493. doi:10.1175/1520-0469(1998)055<0477:TBEOAO>2.0.CO;2

Bitz CM, Holland MM, Hunke EC, Moritz RE (2005) Maintenance of the sea-ice edge. J Clim 18:29032921. doi:10.1175/JCLI3428.1

Blanchard-Wrigglesworth E, Armour KC, Bitz CM, DeWeaver E (2011a) Persistence and inherent predictability of Arctic sea ice in a GCM ensemble and observations. J Clim 24(1):231–250. doi:10.1175/2010JCLI3775.1

Blanchard-Wrigglesworth E, Bitz CM, Holland MM (2011b) Influence of initial conditions and climate forcing on predicting Arctic sea ice. Geophys Res Lett 38(18). doi:10.1029/2011GL048807

Blockley EW, Martin MJ, McLaren AJ, Ryan AG, Waters J, Lea DJ, Mirouze I, Peterson KA, Sellar A, Storkey D (2014) Recent development of the met office operational ocean forecasting system: an overview and assessment of the new global foam forecasts. Geosci Model Dev Discuss 6(4):6219–6278. doi:10.5194/gmdd-6-6219-2013

Budikova D (2009) Role of Arctic sea ice in global atmospheric circulation: a review. Glob Planet Change 68(3):149–163. doi:10.1016/j.gloplacha.2009.04.001

Butterworth P, English S, Hilton F, Whyte K (2002) Investigation into optimal observation errors for satellite winds. Tech. Rep. NWPSAF_MO_TR_007, Met Office, UK. http://research.metoffice.gov.uk/research/interproj/nwpsaf/satwind_report/trialsdir/nwpsaf_mo_tr_007.pdf

Casey K, Brandon T, Cornillon P, Evans R (2010) The past, present and future of the AVHRR Pathfinder SST program. In: Barale V, Gower J, Alberotanza L (eds) Oceanography from space: revisited. Springer, Berlin. doi:10.1007/978-90-481-8681-5_16

Cavalieri D, Parkinson C, Gloersen P, Zwally HJ (1996, updated yearly) Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I passive microwave data. http://nsidc.org/data/nsidc-0051.html

Cavalieri DJ, Parkinson CL, Gloersen P, Comiso JC, Zwally HJ (1999) Deriving long-term time series of sea ice cover from satellite passive-microwave multisensor data sets. J Geophys Res 104:15,803–15,814

Chevallier M, y Mélia DS, Voldoire A, Déqué M, Garric G (2013) Seasonal forecasts of the pan-Arctic sea ice extent using a GCM-based seasonal prediction system. J Clim 26:60926104. doi:10.1175/JCLI-D-12-00612.1

Clayton AM, Lorenc AC, Barker DM (2013) Operational implementation of a hybrid ensemble/4d-var global data assimilation system at the met office. Q J R Meteorol Soc 139(675):1445–1461. doi:10.1002/qj.2054

Comiso J, Cavalieri D, Parkinson C, Gloersen P (1997) Passive microwave algorithms for sea ice concentration: a comparison of two techniques. Remote Sens Environ 60:357–384

Dee D, Berrisford P, Poli P, Fuentes M (2009) ERA-interim for climate monitoring. ECMWF Newslett 119:5–6

Defant A (1924) Die schwankungen der atmosphärischen zirkulation über dem nordatlantischen ozean im 25-jährigen zeitraum 1881–1905. Geogr Ann 6:13–41

Fereday DR, Maidens A, Arribas A, Scaife AA, Knight JR (2012) Seasonal forecasts of northern hemisphere winter 2009/10. Environ Res Lett 7(3):034031. http://stacks.iop.org/1748-9326/7/i=3/a=034031

Fetterer F, Knowles K, Meier W, Savoie M (2002, updated 2011) Sea ice index. Boulder: National Snow and Ice Data Center. Digital Media. http://nsidc.org/data/g02135.html

Francis JA, Vavrus SJ (2012) Evidence linking Arctic amplification to extreme weather in mid-latitudes. Geophys Res Lett L06801. doi:10.1029/2012GL051000

Francis JA, Chan W, Leathers DJ, Miller JR, Veron DE (2009) Winter northern hemisphere weather patterns remember summer Arctic sea ice extent. Geophys Res Lett 36(L07):503. doi:10.1029/2009GL037274

Guemas V, Doblas-Reyes F, Mogensen K, Keeley S, Tang Y (2014) Ensemble of sea ice initial conditions for interannual climate predictions. Clim Dyn 1–17. doi:10.1007/s00382-014-2095-7

Hasselmann K (1976) Stochastic climate models part I. theory. Tellus 28(6):473–485. doi:10.1111/j.2153-3490.1976.tb00696.x

Hewitt HT, Copsey D, Culverwell ID, Harris CM, Hill RSR, Keen AB, McLaren AJ, Hunke EC (2011) Design and implementation of the infrastructure of HadGEM3: the next-generation Met Office climate modelling system. Geosci Model Dev 4(2):223–253. doi:10.5194/gmd-4-223-2011

Hunke EC, Lipscomb WH (2010) Cice: the Los Alamos sea ice model documentation and software users manual, version 4.1. LA-CC-06-012, Los Alamos National Laboratory

Ineson S, Scaife AA, Knight JR, Manners JC, Dunstone NJ, Gray LJ, Haigh JD (2011) Solar forcing of winter climate variability in the northern hemisphere. Nat Geosci 4:753–757. doi:10.1038/ngeo1282

James IN, James P (1989) Ultra-low-frequency variability in a simple atmospheric circulation model. Nature 342:53–55. doi:10.1038/342053a0

Kim KY, North GR (1998) EOF-based linear prediction algorithm: theory. J Clim 3046–3056: doi:10.1175/1520-0442(1998)011<3046:EBLPAT>2.0.CO;2

Kim KY, North GR (1999) EOF-based linear prediction algorithm: examples. J Clim 2076–2092. doi:10.1175/1520-0442(1999)012<2076:EBLPAE>2.0.CO;2

Kumar A, Bhaskar J, Zhang Q, Bounoua L (2007) A new methodology for estimating the unpredictable component of seasonal atmospheric variability. J Clim 20:3888–3901

Large W, Yeager S (2009) The global climatology of an interannually varying airsea flux data set. Clim Dyn 33(2–3):341–364. doi:10.1007/s00382-008-0441-3

Laxon S, Peacock N, Smith D (2003) High interannual variability of sea ice thickness in the Arctic region. Nature 425:947–950. doi:10.1038/nature02050

Lindsay RW, Zhang J (2006) Assimilation of ice concentration in an ice-ocean model. J Atmos Oceanic Technol 23:742749. doi:10.1175/JTECH1871.1

Lindsay RW, Zhang J, Schweiger AJ, Steele MA (2008) Seasonal predictions of ice extent in the Arctic ocean. J Geophys Res Oceans 113(C2). doi:10.1029/2007JC004259

MacLachlan C, Arribas A, Peterson K, Maidens A, Fereday D, Scaife A, Gordon M, Vellinga M, Williams A, Comer RE, Camp J, Xavier P (2014) Description of GloSea5: the Met Office high resolution seasonal forecast system. in preparation for QJRMS

Madec G (2008) NEMO ocean engine. Tech. Rep. Note du Pole de modélisation No 27, ISSN No 1288–1619, Institut Pierre-Simon Laplace (IPSL), France

Maidens A, Arribas A, Scaife A, MacLachlan C, Peterson D, Knight J (2012) Predictability of the North Atlantic Oscillation in early winter 2010–2011. Submitted to Monthly Weather Review

Marshall AG, Scaife AA, Ineson S (2009) Enhanced seasonal prediction of European winter warming following volcanic eruptions. J Clim 22:6168–6180. doi:10.1175/2009JCLI3145.1

Martin M, Hines A, Bell M (2007) Data assimilation in the FOAM operational short-range ocean forecasting system: a description of the scheme and its impact. QJR Meteorol Soc 133:981–995

McLaren AJ, Banks HT, Durman C, Gregory J, Johns T, Keen A, Ridley J, Roberts M, Lipscomb W, Connolley W, Laxon S (2006) Evaluation of the sea ice simulation in a new coupled atmosphere–ocean climate model (HadGEM1). J Geophys Res 111(C12):014. doi:10.1029/2005JC003033

Merryfield W, Lee WS, Boer GJ, Kharin VV, Scinocca JS, Flato G, Ajayamohan RS, Fyfe JC, Tang Y, Polavarapu S (2012) The Canadian seasonal to interannual prediction system: part I. Models and initialization. Submitted to Monthly Weather Review

Merryfield WJ, Lee WS, Wang W, Chen M, Kumar A (2013) Multi-system seasonal predictions of Arctic sea ice. Geophys Res Lett 40(8):1551–1556. doi:10.1002/grl.50317

Notz D (2009) The future of ice sheets and sea ice: between reversible retreat and unstoppable loss. Proc Natl Acad Sci 106(49):20590–20595. doi:10.1073/pnas.0902356106

OSI-SAF (2011) EUMETSAT ocean and sea ice satellite application facility. Global sea ice concentration reprocessing dataset 1978–2009 (v1.1, 2011). http://osisaf.met.no

Overland JE, Wang M (2010) Large-scale atmospheric circulation changes are associated with the recent loss of Arctic sea ice. Tellus A 62(1):1–9. doi:10.1111/j.1600-0870.2009.00421.x

Rayner NA, Parker D, Horton E, Folland C, Alexander L, Rowell D, Kent E, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J Geophys Res 108:4407. doi:10.1029/2002JD002670

Roberts-Jones J, Fiedler EK, Martin MJ (2012) Daily, global, high-resolution SST and sea ice reanalysis for 1985–2007 using the OSTIA system. J Clim 25:6215–6232. doi:10.1175/JCLI-D-11-00648.1

Schweiger A, Lindsay R, Zhang J, Steele M, Stern H, Kwok R (2011) Uncertainty in modeled Arctic sea ice volume. J Geophys Res Oceans 116(C8). doi:10.1029/2011JC007084

Sigmond M, Fyfe JC, Flato GM, Kharin VV, Merryfield WJ (2013) Seasonal forecast skill of Arctic sea ice area in a dynamical forecast system. Geophys Res Lett. doi:10.1002/grl.50129

Smith DM (1996) Extraction of winter sea ice concentration in the Greenland and Barents seas from SSM/I data. Int J Remote Sens 17:2625–2646

Stark J, Ridley J, Martin M, Hines A (2008) Sea ice concentration and motion assimilation in a sea ice ocean model. J Geophys Res

113:C05S91. URL http://www.agu.org/pubs/crossref/2008/2007JC004224.shtml

Storkey D, Blockley E, Furner R, Guiavarc'h C, Lea D, Martin M, Barciela R, Hines A, Hyder P, Siddorn J (2010) Forecasting the ocean state using NEMO: the new FOAM system. J Oper Oceanogr 3:3–15. http://www.ingentaconnect.com/content/imarest/joo/2010/00000003/00000001/art00001

Stroeve J, Holland MM, Meier W, Scambos T, Serreze M (2007) Arctic sea ice decline: faster than forecast. Geophys Res Lett 34(9). doi:10.1029/2007GL029703

Thompson DWJ, Wallace JM (1998) The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. Geophys Res Lett 25:1297–1300

Tivy A, Alt B, Howell S, Wilson K, Yackel J (2007) Long-range prediction of the shipping season in Hudson Bay: a statistical approach. Weather Forecast 22:1063–1075. doi:10.1175/WAF1038.1

Walker GT, Bliss EW (1932) World weather V. Mem Roy Meteor Soc 4(36):53–84

Wang W, Chen M, Kumar A (2013) Seasonal prediction of Arctic sea ice extent from a coupled dynamical forecast system. Mon Weather Rev 141:1375–1394. doi:10.1175/MWR-D-12-00057.1

Woollings T, Lockwood M, Masato G, Bell C, Gray L (2010) Enhanced signature of solar variability in Eurasian winter climate. Geophys Res Lett 37(L20):805