Improving Met Office seasonal forecasts of Arctic sea ice using assimilation of CryoSat-2 thickness

Edward W. Blockley[1] and K. Andrew Peterson[1]

[1]Met Office, FitzRoy Road, Exeter, EX1 3PB, United Kingdom

5   *Correspondence to*: Ed Blockley (ed.blockley@metoffice.gov.uk)

Abstract. Interest in seasonal predictions of Arctic sea ice has been increasing in recent years owing, primarily, to the sharp reduction in Arctic sea ice cover observed over the last few decades, which is projected to continue. The prospect of increased human industrial activity in the region, as well as scientific interest in the predictability of sea ice, provides important motivation for understanding, and improving, the skill of Arctic predictions. Several operational forecasting

10   centres now routinely produce seasonal predictions of sea ice cover using coupled atmosphere-ocean-sea ice models. Although assimilation of sea ice concentration into these systems is commonplace, sea ice thickness observations, being much less mature, are typically not assimilated. However many studies suggest that initialisation of winter sea ice thickness could lead to improved prediction of Arctic summer sea ice. Here, for the first time, we directly assess the impact of winter sea ice thickness initialisation on the skill of seasonal summer forecasts by assimilating CryoSat-2 thickness data into the

15   Met Office's coupled seasonal forecasting system (GloSea). We show a significant improvement in predictive skill of Arctic sea ice extent and ice-edge location for forecasts of September Arctic sea ice made from the beginning of the melt season. The improvements in sea ice cover lead to further improvement of near-surface air temperature and pressure fields across the region. A clear relationship between modelled winter thickness biases and summer extent errors is identified which supports the theory that Arctic winter thickness provides some predictive capability for summer ice extent, and further highlights the

20   importance that modelled winter thickness biases can have on the evolution of forecast errors through the melt season.

1 Introduction and motivation

Arctic sea ice is one of the most rapidly, and visibly, changing components of the global climate system. The past few decades have seen a considerable reduction in the extent and thickness of Arctic sea ice (Vaughan et al., 2013; Meier et al., 2014; Lindsay and Schweiger, 2015; Kwok et al., 2009). Although the areal extent of Arctic sea ice has declined in all

25   seasons, the reduction has been most pronounced in the summer with the seasonal minimum extent hitting record low values in September 2007 and 2012 (Meier et al., 2014; Vaughan et al., 2013). This decline is projected to continue in the future in response to rising global temperatures and atmospheric CO_2 concentrations (Collins et al., 2013; Notz and Stroeve, 2016).

In response to declining sea ice cover, human activity in the Arctic is increasing with access to the Arctic Ocean becoming more important for socioeconomic reasons (Meier et al., 2014). Such activities include commercial activities like tourism, fishing, mineral and oil extraction, and shipping (Smith and Stephenson, 2013), along with activities of importance to local communities such as subsistence hunting and fishing, search and rescue, and community re-supply. Accurate forecasts of

5   Arctic sea ice are therefore becoming increasingly important for the safety of human activities in the Arctic (Eicken, 2013). Better knowledge of sea ice on seasonal timescales allows for better planning which should lead to a reduced level of risk and a reduction in operational costs for human activities in the Arctic Ocean. Regional changes in Arctic sea ice cover can also have implications for lower-latitude weather and climate (Koenigk et al. 2016; Balmaseda et al. 2010; Screen, 2013). For example, Koenigk et al. (2016) show that late summer sea ice cover can be linked to winter NAO-like patterns and

10   blocking in Western Europe. More accurate Arctic sea ice predictions therefore, can also contribute to improved forecasts, and hence longer-term planning, in mid-latitude regions.

Following the drastic reduction in Arctic sea ice extent in the summer of 2007, which led to a (then) record-low summer minimum extent being set, interest in seasonal predictions has increased. In response to this, in 2008, the Sea Ice Outlook

15   (SIO) was instigated by the Study of Environmental ARctic CHange (SEARCH) to synthesise seasonal predictions of September Arctic sea ice extent, made from late spring and early summer, using a variety of modelling, statistical, and heuristic approaches (see Stroeve et al., 2014). For seasonal forecasts to be of use to stakeholders, a thorough understanding of their predictive skill is needed. The SIO fosters collaboration of such activities across various prediction centres through the inter-comparison and common evaluation of forecasts (see https://www.arcus.org/sipn/sea-ice-outlook). There is also an

20   interesting scientific problem here to test our ability to predict sea ice on seasonal timescales. As the sea ice thins, variability in ice extent increases (Holland et al., 2011; Goosse et al., 2009) and so the problem of making seasonal Arctic sea ice predictions – particularly for the September minimum – is one that is getting more challenging and interesting as the ice cover declines.

25   Although global coupled forecasting systems have been used successfully for seasonal prediction of mid-latitude weather and climate for some time now (see for example Scaife et al., 2014), their application to Arctic sea ice prediction is much less mature. In particular, forecasts in the Arctic are hampered by the fact that observations are much less abundant and data assimilation techniques less advanced in the polar regions than at lower-latitudes (Jung et al., 2016; Bauer et al., 2016). Despite this fact, several operational forecasting centres regularly contribute to the SIO with sea ice predictions from fully

30   coupled atmosphere-sea ice-ocean modelling systems. One such system is the Met Office's Global Seasonal (GloSea) coupled ensemble prediction system (MacLachlan et al., 2014; Peterson et al., 2015) which has contributed to the SIO since 2010. The ocean and sea ice components of GloSea are initialised each day using the FOAM operational ocean-sea ice analysis of Blockley et al. (2014; 2015). FOAM routinely assimilates sea ice concentration along with various ocean quantities (satellite and in-situ SST, satellite SLA, in-situ profiles of temperature and salinity) but, in common with most

operational ocean analysis systems (Tonani et al., 2015; Martin et al., 2015; Balmaseda et al., 2015), does not assimilate sea ice thickness.

Despite the use of dynamic models for seasonal sea ice prediction being in its relative infancy, there have been several studies that have demonstrated skill in retrospective forecasts (or hindcasts) of September-mean Arctic sea ice extent from spring (e.g., Chevallier et al., 2013; Msadek et al., 2014; Peterson et al., 2015). There has also been much progress made towards understanding potential predictability of Arctic sea ice extent – often using so-called 'perfect model' approaches (Guemas et al., 2016; Tietsche et al., 2014; Day et al., 2014; Blanchard-Wrigglesworth et al., 2011). However other studies have suggested there is still much improvement that can be made to these systems before they can be considered as useful operational tools. Blanchard-Wrigglesworth et al. (2015) showed that skill of the actual forecasts submitted to the SIO is lower than that suggested by these hindcast and perfect model studies, and Stroeve et al. (2014) found that skill was only marginally better than a linear trend forecast.

Several studies have suggested that seasonal predictions could be improved by the addition of sea ice thickness. Blanchard-Wrigglesworth and Bitz (2014) found sea ice thickness anomalies in GCMs to have a timescale of between 6 and 20 months making their correct representation in model initial conditions of importance for seasonal predictions. Other modelling studies by Holland et al. (2011) and Kauker et al. (2009) have also shown that knowledge of winter ice thickness can provide some predictive capability for summer ice extent. Perfect model studies (e.g. Day et al., 2014) have also suggested that correct initialisation can lead to improved seasonal forecasts. Day et al. (2014) used the HadGEM1.2 climate model to show that memory of winter thickness conditions can persist well beyond seasonal timescales and provide predictive capability for up to 2 years. Collow et al. (2015) found considerable changes in ice concentration forecasts when changing the initial thickness in the CFSv2 seasonal prediction system. They show an improvement to seasonal forecasts when using thickness fields from the Pan-Arctic Ice-Ocean Model Assimilation System (PIOMAS) model of Zhang and Rothrock (2003). However Blanchard-Wrigglesworth et al. (2017), who also used PIOMAS to initialise many different SIO seasonal prediction models, find that there is also a strong role for model uncertainty in the evolution of seasonal forecasting errors.

Although the assimilation of sea ice concentration has been included in ocean reanalysis, operational ocean prediction and seasonal forecasting systems for several years (Stark et al., 2008; Peterson et al. 2015), sea ice thickness is not yet routinely used to initialise these systems (Martin et al., 2015; Balmaseda et al., 2015; Tonani et al., 2015). There have however, been several recent studies that have sought to improve the representation of Arctic sea ice thickness in analyses and short-range forecasts using satellite thickness products derived from Soil Moisture and Ocean Salinity (SMOS) brightness temperatures and/or from CryoSat-2 (hereafter CS2) radar freeboard measurements. Such studies have generally focused on assimilation of thickness using ensemble techniques into short-range forced ocean-sea ice models in the Topaz system (Xie et al., 2016) or using MITgcm (Yang at al., 2014; Mu et al., 2018). Although these studies showed considerable improvement to the

simulation of sea ice thickness, the impact on short-range forecasts of sea ice concentration or extent was minimal. More recently, Allard et al. (2018) used direct initialisation of CS2-derived thickness within a series of reanalyses performed with NRL's ocean-sea ice Arctic Cap Nowcast/Forecast System (ACNFS). They show that the analysed sea ice thickness is significantly improved when assimilating CS2 thickness compared against in-situ and airborne measurements. They also
5 show a good agreement between the CS2-derived ice thickness and observations from in-situ and airborne sources.

In this study, we use a nudging technique to assimilate CS2 sea ice thickness within the FOAM/GloSea reanalysis system and use these initial conditions to determine the impact of sea ice thickness initialisation on the skill of GloSea seasonal predictions. Although several studies have looked at the impact of thickness initialisation on short-term forecasts and in
10 forced ocean-sea ice models, there have been no studies exploring the impact of initialising coupled seasonal forecasts using sea ice thickness. Here we do so for the first time using the Met Office GloSea seasonal prediction system and show that sea ice thickness initialisation leads to a considerable improvement in the skill of seasonal predictions of Arctic sea ice extent and ice edge location.

15 This paper is structured as follows: Section 2 introduces the modelling systems and observations used in this study; Section 3 describes the assimilation of CS2 thickness and the impact on the ocean-sea ice reanalysis; Section 4 provides details of GloSea coupled seasonal forecast experiments performed using CS2 initialised thickness and shows improved skill for seasonal forecasts of Arctic ice cover. Section 5 provides summary discussion and an overview of proposed future work.

2 Models and observations used in this study

20 ### 2.1 Modelling systems

The model systems used in this study are taken from the Met Office suite of seamless, traceable prediction systems introduced by Brown et al. (2012) using components of the HadGEM3 coupled model architecture described by Hewitt et al. (2011). All of these HadGEM3-based modelling systems simulate the ocean and sea ice conditions using the Nucleus for European Modelling of the Ocean (NEMO) ocean model (Madec, 2008) coupled to the Los Alamos sea ice model (CICE)
25 (Hunke et al., 2015).

Within the Met Office's unified, seamless framework, seasonal forecasts are performed using the GloSea coupled prediction system (MacLachlan et al., 2014; Scaife et al., 2014). GloSea produces two 210-day seasonal forecasts every day, which, together with those from previous days, are combined to form a lagged ensemble prediction system. Meanwhile hindcasts
30 performed for previous years are used to establish errors in the model climatology for the purposes of bias correction. More details on GloSea can be found in MacLachlan et al., (2014). The ocean and sea ice components of the GloSea system are initialised each day using analyses from the Forecast Ocean Assimilation Model (FOAM) system described in Blockley et al.

(2014; 2015). FOAM is an operational ocean-sea ice analysis and forecast system run daily at the Met Office. Satellite and in-situ observations of temperature, salinity, sea level anomaly and sea ice concentration are assimilated by FOAM each day using the NEMOVAR 3D-Var FGAT scheme. Sea ice thickness is not currently assimilated in FOAM; new ice is added by the concentration assimilation at a default thickness of 0.5 metres. More details of the FOAM system can be found in

5   Blockley et al. (2014) and more about the NEMOVAR 3D-Var FGAT scheme used therein can be found in Waters et al. (2014).

As well as operational analyses and forecasts, longer reanalyses are performed with the FOAM system using surface forcing derived from the ERA-Interim atmospheric reanalysis (Dee et al., 2011). Within the GloSea seasonal prediction system,

10   hindcast experiments initialised from these reanalyses are used to bias correct the GloSea seasonal forecasts (see MacLachlan et al., (2014) for more information). As well as being used for bias correction within GloSea, these ocean reanalyses are used more widely within the ocean community and have been used to help answer a number of other scientific questions (e.g. by Roberts et al., 2013; Jackson et al., 2015).

15   Throughout this study we shall use prototype FOAM and GloSea systems based on the latest configuration of the Met Office coupled modelling system (GC3: Williams et al., 2017) which will be used as part of Met Office Hadley Centre's contribution to phase 6 of the Coupled Model Intercomparison Project (CMIP6). This GC3 coupled model version uses the GO6 ocean and GSI8 sea ice component configurations described in Storkey et al., (2018, in review) and Ridley et al., (2018) respectively and uses the extended ORCA025 grid described therein (nominally 1/4 degree horizontal resolution with

20   75 vertical levels).

2.2 Observations of sea ice thickness

Whilst observations of sea ice concentration providing large-scale coverage for both poles have been available since 1979 (Fetterer et al., 2016; Rayner et al., 2003), measurements of sea ice thickness are, relatively, much less abundant. However, satellite estimates of winter thickness have been available for a number of years using radar altimetry (Laxon et al., 2003),

25   laser altimetry (Kwok et al., 2009), and, more recently for thin ice, microwave brightness temperatures (Kaleschke et al., 2016). Although radar altimeter estimates of sea ice thickness have been around for some years their up-take into operational ocean-sea ice assimilation systems has been minimal. The main reasons for this are three-fold: owing to the orbit inclination, these datasets often have a large 'pole-hole' giving poor coverage in the central Arctic Ocean; there is considerable uncertainty associated with these estimates of ice thickness (Ricker et al., 2014); the data were not made available in near-

30   real-time for use in operational analysis systems. The problems outlined above have been ameliorated somewhat during the last few years by the launch of ESA's CryoSat-2 satellite (CS2) whose primary objective is to acquire accurate measurements of sea ice thickness. CS2 is fitted with a SIRAL altimeter and has an unusually high inclination orbit that provides observational coverage up to 88° North (Laxon et al., 2013). The processed data from CS2 is also provided in

almost near-real-time by the Centre for Polar Observation and Modelling (CPOM) (Tilling et al., 2016) making its use within operational analysis systems a realistic proposition.

2.2.1 CryoSat-2 thickness observations

In this study, we use monthly CS2 winter (Oct-Apr) thickness estimates produced by CPOM (Tilling et al., 2016) which start

5   from October 2010 until present (at time of writing). Sea ice freeboard is inferred from radar altimetry aboard the CS2 satellite and is converted to thickness by assuming that the sea ice floats in hydrostatic equilibrium and by making various assumptions about the snow loading and the relative densities of the sea ice, the ocean and the overlying snow. Details of the methods used to generate the CPOM thickness fields can be found in Laxon et al., (2013) and Tilling et al. (2015). Some more general discussion of the uncertainties involved in the calculation of sea ice freeboard and thickness using radar

10 altimetry can be found in Ricker et al. (2014).

The CPOM thickness data are provided on a 5 km polar stereographic grid having been smoothed with an averaging window of radius 25 km. We apply a further quality control (QC) to the data before use. To prevent any smearing of thickness values near the ice edge during the filtering step we impose a maximum displacement of 15 km between the average location of the

15 raw track observations and the centre of the final grid point (Andy Ridout, pers. comm., 2017). We further remove any spuriously high observations by imposing a maximum thickness threshold of 7 m. The CS2 thickness retrieval methodology is particularly sensitive for thin ice where the ice freeboard is not much higher than sea level (Ricker et al. 2014; 2017). To avoid high observational error associated with these thin measurements we impose a minimum thickness threshold of 1 m – a choice that was motivated by Figure 2b of Ricker et al. (2017). Further, to ensure that the observations are as representative

20 of the month as possible we apply the constraint that at least 10 different altimeter tracks are used to determine the monthly-mean observation. We also impose a constraint on the spread of the track observations by keeping monthly observations only when the standard deviation of the contributing individual track observations was less than 2 m. In total, application of the abovementioned QC rejected roughly 21.5% of the original observations; about 9.4% of the observations were removed by the 1m cut-off and just over 12% were rejected by the remaining constraints. An example of the thickness observations used

25 in this study can be seen in Figure 1, which shows a map of average December, January and February Arctic thickness for 2011-2015 inferred from CS2 estimates after application of the QC process described above.

Although SMOS thickness data is also available for much of the study period used here, we choose to use only CS2 data for this study. The motivation for this is that SMOS only provides information about thin ice up to about 0.5 m whilst the CS2

30 data is good for thicker ice above 1 m (Ricker et al., 2017). Here we are concerned with seasonal predictions of Arctic summer sea ice - for which ice less than 2 m thick tends to melt away completely (Keen et al., 2013). We therefore expect the assimilation of thicker ice to be more important for our needs and so use only the CS2 thickness observations for this feasibility study.

2.3 Observations of sea ice extent and concentration

Uncertainty associated with sea ice concentration and extent estimates from satellites is high (Ivanova et al., 2015) and the commonly used sea ice extent metric is nonlinear and dependent on resolution (Notz, 2014). To account for this uncertainty we include observational estimates from three different sources: extents calculated from the 1 degree gridded HadISST1.2

5   dataset of Rayner et al. (2003); the NSIDC sea ice index of Fetterer et al. (2016); and gridded sea ice concentration fields from the most recent FOAM-GloSea ocean-sea ice reanalysis. This reanalysis, performed using version 13 of the FOAM system (Blockley et al., 2015), is used within the Copernicus Marine Environment Monitoring Service (CMEMS; http://marine.copernicus.eu/) global ocean reanalyses ensemble product (ID: GLOBAL-REANALYSIS-PHY-001-026; described in http://cmems-resources.cls.fr/documents/QUID/CMEMS-GLO-QUID-001-02). Using this CMEMS reanalysis

10   has the benefit that it is performed on the same ORCA025 grid as the ocean-sea ice components of the GloSea seasonal forecasting system, which makes spatial comparisons easier. This reanalysis has also been evaluated thoroughly through the Ocean Reanalyses Inter-comparison Project (ORA-IP) (see Balmaseda et al., 2015; Chevallier et al., 2017; Uotilla et al., in review). To avoid confusion with the FOAM reanalyses performed as part of this study, and described later, we refer to this product as "CMEMS" hereafter.

The CMEMS reanalysis was performed using SSMI/S sea ice concentration data provided by EUMETSAT's Ocean and Sea Ice Satellite Application Facility (OSI-SAF). Sea ice concentration is assimilated along with ocean data sources using the NEMOVAR 3D-Var scheme (see Blockley et al., 2014; Waters et al., 2014). Prior to October 2009, OSI-SAF's Global Sea Ice Concentration Climate Data Records (OSI-409, version 1.1) product was assimilated. When the reanalysis was run, in

20   2015, these data were only available up to the end of 2009 and so the OSI-SAF near-real-time (NRT) product OSI-401a was used from 25th October 2009 onwards. These two datasets have differences in the processing of low concentration ice and near coastlines (see Section 4.2of OSI-SAF, 2017). However, this does not cause us any concern here because our study is focussed on the CS2 era from October 2010 onwards.

3 Initialisation of thickness in the ocean-sea ice reanalysis system

25   Here we use the latest development version of the FOAM-GloSea reanalysis system that has been undertaken as part of the upgrade of GloSea and FOAM to use the latest GC3 version of the Met Office coupled model architecture (Williams et al., 2017). Specifically here the ocean reanalysis system is using the GO6 ocean configuration described in Storkey et al. (2017, in review) and the GSI8 sea ice configuration described in Ridley et al. (2017, in review). We take the latest GO6+GSI8 reanalysis as our control (hereafter CTRL-RA) and modify it to include initialisation of sea ice thickness using CS2

30   observations (hereafter ThkDA-RA). The CTRL-RA reanalysis was run from 1992 to 2015 but here we only re-run the last 5 years – from October 2010 to the end of 2015 – to tie in with availability of CS2 thickness estimates.

Within the ThkDA-RA reanalysis, CS2 thickness data are assimilated using a basic nudging technique in which thickness fields are nudged towards the monthly gridded CS2 observations in a fashion akin to that employed by climatological relaxation schemes. All other data used within the control run (i.e., SST, SLA, T/S profiles, and SSMI/S concentration) are assimilated here too in the same manner as in the standard FOAM system (Blockley et al., 2014; 2015). An overview of
5   reanalysis experiments used in this study can be found in Table 1.

We use the monthly CPOM measurements introduced in Section 2.2 and map them onto the model grid using a standard binning technique. A linear interpolation is performed each day to get daily thickness observations from the nearest two months. Assimilation increments are created by taking a simple difference between these daily CS2 thickness observations
10   and the daily-mean model thickness. Where no observations are present, the increments are set to zero to ensure no thickness nudging is performed. We do things this way to avoid problems arising with the sparse data and so we can keep nudging model towards CS2 thickness.

The increments are applied within the CICE model code in a similar fashion to the sea ice concentration assimilation
15   described in Peterson et al. (2014) and Blockley et al. (2014). Thickness changes are made at each time step using the Incremental Analysis Update (IAU) method. A 5-day relaxation timescale is used and increments are only applied where the grid-cell ice concentration is above 40%. The CICE sea ice model uses multiple thickness categories to represent the sub-grid thickness distribution. To apply the thickness increments into the multi-category model we chose to nudge the grid-box-mean thickness towards observations by making changes across each of the 5 sub-grid categories - so long as there is ice
20   present there with concentration above 1% - maintaining the initial distribution of volume between the categories. We note here that this approach is similar to that employed by Allard et al. (2018) who multiply the ice volume in each category by the grid-box-mean model-observation thickness difference. However, whilst they use direct initialisation, we use the IAU approach to incorporate changes into the model in a gradual manner and limit the potential for sudden shock in the system (Bloom et al., 1996).

25 **3.1 Impact of CroySat-2 initialisation on reanalysis thickness**

Figure 2 shows the impact of the CS2 thickness initialisation on the reanalysis thickness fields; the 5-year mean differences for 1st May – at the end of winter when CS2 observations cease – and for 30th September – after 5 months of running without thickness assimilation – are shown. At the end of winter (Figure 2a) it is apparent that inclusion of CS2 thickness nudging has increased sea ice thickness across much of the Atlantic sector of the Arctic (Barents, Kara and Greenland Seas).
30   Conversely, ice thickness has been decreased in the Canadian Arctic Archipelago (CAA) and, to a lesser degree, across much of the Pacific sector (Beaufort, Chukchi and East Siberian Seas). Thickness is also increased in many of the marginal seas outside of the central Arctic such as the Bering Sea and Hudson Bay. One notable exception is the Labrador Sea/Baffin

Bay where the differences form an east-west dipole with ice thickness being reduced along the Canadian coast but increased on the Greenland side.

At the end of summer (Figure 2b) the impact of the CS2 nudging is an increase in sea ice thickness throughout much of the Arctic save for small patches in the East Siberian Sea and within the CAA. The pattern is broadly consistent with the differences seen at the end of winter in Figure 2a. Even after 5 months of running without the CS2 thickness nudging – although whilst still assimilating ice concentration and other ocean quantities – we can see the impact of initialising thickness through the winter. This is good news for the feasibility study because it tells us that the thickness changes are being retained by the model and not being rejected or washed out by the assimilation of other quantities such as ice concentration.

The general picture shown by the 5-year means in (Figure 2a) is typical of the winter thickness differences seen for each of the 5 years 2010-2015 (not shown). However, the situation is not so clear-cut for the summer case where thickness increases are much more pronounced in 2014 and 2013 (see Figure 3).

Although there is a dipole in the thickness changes, with some areas being consistently thicker and others consistently thinner, the net effect of the CS2 thickness nudging is an increase in sea ice thickness. This can be seen on the volume time series plot in Figure 3 – noting that an increase in volume here directly implies an increase in average ice thickness because, as sea ice concentration is tightly constrained by the assimilation of sea ice concentration and sea surface temperature, the ice area between the two reanalysis simulations is virtually identical (not shown). In Figure 3, sea ice volume for the CTRL reanalysis is compared with the equivalent for the ThkDA reanalysis along with, as a reference, volume estimates from the PIOMAS model of Zhang and Rothrock (2003). The volume in the CTRL run is much closer to PIOMAS than the ThkDA run. This is expected as PIOMAS has been shown to underestimate thickness/volume in the winter when compared with thickness derived from radar altimeter (Tilling et al., 2015; Laxon et al., 2013) – although has been shown to compare better with laser altimeter estimates such as ICESat (Schweiger et al., 2011). Figure 3 shows that winter volume is increased the most by the assimilation of CS2 thickness. This is perhaps not surprising given that winter is the time when the data is available. However, there is some evidence that these winter changes also affect the summer volume, which is most pronounced in 2014 and, to a lesser extent, 2013 and 2015. In all years the volume time series shows a clear kink on 1st October when the CS2 data comes back online and begins to be assimilated in the reanalysis - although this is much less pronounced in 2014 when the summer thickness was also increased.

In summary, we have shown that nudging Arctic sea ice thickness to CS2 observations within the ThkDA-RA reanalysis has the net effect of increasing sea ice volume. The differences between the two reanalyses reveal a persistent bias in the thickness distribution in the model when compared with CS2 whereby sea ice is too thick on the Pacific side and not thick

enough on the Atlantic side of the Arctic. There is evidence to suggest that the winter Arctic sea ice thickness/volume is an important precondition for evolution of ice through the melt season (in agreement with the current literature) because the effects of winter thickness changes imposed by the nudging are still evident at the end of the summer. Another important result to note here is the fact that the assimilation of thickness worked well and the increments were successfully retained by

5   the model, which bodes well for inclusion of sea ice thickness within the NEMOVAR system in the future.

4 Initialisation of thickness in the GloSea coupled seasonal prediction system

Seasonal forecasts of sea ice extent, amongst other things, are made operationally by the GloSea system each day. Hindcasts, performed from a discrete predefined set of start dates each year, are also run within the operational suite each day and used as part of the bias correction process. These hindcasts are initialised using the long GloSea ocean-sea ice reanalysis (as

10   described in Section 3) which is coupled to atmosphere initial conditions interpolated from the ERA-I reanalysis (Dee et al., 2011). In addition to being used operationally for bias correcting forecasts, seasonal hindcasts such as this are performed for testing of model configuration upgrades prior to implementation within the GloSea operational suite. A recent trial of the new GC3 coupled model configuration of Williams et al. (2017) has been performed using the GloSea system and we shall use as our control (denoted CTRL-HC). The ocean and sea ice for these hindcasts are initialised using the control reanalysis

15   (CTRL-RA) described in Section 3 and the atmosphere is initialised from the ERA-I reanalysis. As the GC3 developments include the implementation of a new multi-layer model for terrestrial snow (see Walters et al., 2017; Williams et al., 2017) the snow fields were initialised separately from the atmosphere using a standalone version of the GC3 land surface component (JULES) with ERA-I snow precipitation and data assimilation.

20   Here we wish to test the impact of initialising with CS2 sea ice thickness on the seasonal forecasts of September sea ice extent. For this purpose, an ensemble of seasonal forecasts was configured that was identical to the CTRL-HC hindcasts except that the ocean and sea ice components were initialised from the ThkDA-RA reanalysis instead of CTRL-RA. Seasonal forecasts were performed from 3 different spring start dates (25th April, 1st May and 9th May). For each of these start dates, an ensemble of 8 hindcasts was initialised from the same analysis fields with spread between the members

25   achieved by using stochastic physics (see MacLachlan et al., (2013) for more details). This methodology is identical to that used for CRTL-HC and, through a mixture of lagged and perturbed methods, provides an ensemble of 24 forecasts of September sea ice each year. These hindcasts were performed for 2011-2015 – each year that spring analyses are available from the ThkDA-RA ocean reanalysis. We denote this system of hindcasts as ThkDA-HC. Details of the GloSea coupled hindcast experiments used in this study can be found in Table 2.

4.1 Improvements to seasonal prediction of Arctic extent and ice edge location

Results from the ThkDA-HC seasonal hindcasts show that the CS2 thickness initialisation has considerably improved the skill of GloSea seasonal predictions of Arctic sea ice cover. Figure 4 shows September-mean Arctic sea ice extent (upper panel) from the GloSea control ensemble (CTRL-HC; blue) and the ensemble run with initialised thickness (ThkDA-HC; pink). Hindcasts from each of the 24 ensemble members, initialised from the 3 April/May start-dates, are depicted by the crosses; the ensemble mean is plotted with bold symbols and inter-connecting lines. Although the ThkDA-HC hindcasts only start from 2011 we plot the CTRL-HC throughout the whole period of the run from 1992-2015 to help put the, relatively short, 5-year time series into context. To assess the accuracy of the GloSea seasonal predictions, observational estimates of Arctic extent, from the CMEMS, HadISST and NSIDC sources (see Section 2.3), are plotted alongside the model hindcasts (black/grey). We note here that the difference in extent prior to 2010 between the CMEMS and the HadISST and NSIDC data sources apparent in Figure 4a is caused by the switch in OSI-SAF data products in October 2009, from OSI-409 version 1.1 to OSI-401a, described in Section 2.3 above. Being prior to the launch of CS2, this change does not have any impact on the results of our study but we include all years available from CTRL-HC in Figure 4 to build a picture of the skill in the CRTL-HC simulations.

The total extent comparisons in Figure 4a show that the run with initialised winter thickness gives improved predictions of September sea ice extent. This is particularly true for 2011 and 2012, for which the ThkDA-HC predictions of total extent are very close to the observed values. The underestimation of basin-wide extent seen throughout the CTRL-HC hindcasts has been reduced; 2011-2015 5-year-mean extent of 3.78 x106 km2 for ThkDA-HC is much closer to the observational average of 4.62 x106 km2 than is the CTRL-HC value of 2.79 x106 km2 (Figure 4a).

Basin-wide extent is not a very useful metric for assessing sea ice because, although it provides information about the amount of ice present, it does not take into account the location of the ice or the position of the ice edge - which are more useful for operational users (Notz, 2014). To assess the skill of GloSea predictions in relation to the spatial distribution of ice and ice edge location, we use the Integrated Ice Edge Error (IIEE) metric introduced by Goessling et al. (2016). This metric is essentially the area integral of all model grid cells where the forecast and observations disagree about whether sea ice is present or not (see Goessling et al., 2016 for more details). Here we use a sea ice concentration threshold of 15% to define whether ice is present or not in any particular grid cell and compare the GloSea hindcast predictions to the CMEMS reanalysis which assimilated the OSI-SAF data. The GloSea and CMEMS products are on the same ORCA025 grid and so comparisons between the two are easy and not degraded by having to remap the data between different grids. Results from the IIEE analysis can be found in Figure 4b which shows IIEE for each ensemble member of the CTRL-HC and ThkDA-HC GloSea forecasts (as in Figure 4a but for IIEE not extent).

Figure 4b shows that ice-edge error is considerably improved by the CS2 thickness initialisation with IIEE reduced from 3.21 x106 km2 for CTRL-HC down to 2.02 x106 km2 for ThkDA-HC - a reduction of 37%. The differences in both extent and IIEE shown in Figure 4 are significant at the 1% level over the whole 5-year period and for each of the individual years except for 2013. In general, the improvement in the ice edge location and IIEE is more pronounced than the improvement to

5   the basin-wide extent. This is to be expected given that the CS2 thickness initialisation changed the distribution of sea ice thickness in the Arctic as well as increasing average thickness. Figure 5 further illustrates the spatial improvement in sea ice predictions showing the probability of ice across the CTRL-HC and ThkDA-HC ensembles for each year (2011-2015) with ensemble-mean and observed ice extent (represented by 15% concentration contours) overlain. Here we calculate the probability of ice, at each grid-cell, as the proportion of ensemble members for which the ice concentration is at least 15%.

10 Consistent with the IIEE results in Figure 4b, the ice edge location in Figure 5 for the ThkDA-HC system is much better than for CTRL-HC. In particular, the ThkDA-HC ensemble-mean ice edges for 2011 and 2012 are very close to those produced by the CMEMS reanalysis. A consistent feature of Figure 5 is that the ice edge along the Atlantic sector of the Arctic is very well defined for the ThkDA predictions and is very close to the CMEMS reanalysis for all years.

15 The spatial changes in the September-mean sea ice concentration forecasts depicted in Figure 5 match well with the May mean thickness dipole shown in Figure 2a. A good illustration of this is 2012 for which the extent improvement is much smaller than the IIEE improvement (Figure 4) which is caused by the fact that much of the ice that remains in the CTRL-HC predictions is located in the Beaufort Sea rather than in the Atlantic sector (north of Fram Strait/Svalbard and east of Greenland). Figure 6 further illustrates this point by showing how thickness differences between the CTRL and ThkDA

20 experiments – for both the analysed spring initial conditions and the September-mean seasonal forecasts – relate to the eventual predictions of ice edge. The thickness dipole from the CS2 nudging matches up well with the areas of missing ice in the Atlantic sector and the areas of excess ice in the Beaufort Sea. This suggests a strong relationship, in this model at least, between wintertime thickness biases and the evolution of errors in sea ice concentration through the summer.

4.2 Wider impact of Arctic sea ice changes

25 The changes in winter ice thickness, and associated changes in the evolution of Arctic ice coverage through the melt season (described above), lead to some significant changes, and improvements, for prediction of the wider climate system. Owing to the overall increase in Arctic sea ice thickness and extent, the ThkDA predictions show a general cooling of September 2m air temperatures in the Arctic (Figure 7a). This cooling improves the model error – calculated here as the difference from the ERA-I atmospheric reanalysis – both for each individual ensemble member (Figure 8a) and for the ensemble mean (not

30 shown). The exception to this is south of the Fram Strait in ice export regions, where the 2m temperature has become too cool. We note, however, that for individual ensemble members, the disagreement with observations should not completely disappear, as the observed value represents only one possible manifestation of the range of possible realities that the ensemble attempts to model. Interestingly, this improvement is also seen over perennially ice covered regions north of

Greenland and the Canadian Arctic Archipelago, where significant improvements in air-sea fluxes would not necessarily be expected.

Significant differences are also present in the pressure fields. A general reduction in 500hPa thickness is seen in the ThkDA experiment over the Arctic Ocean (Figure 7b) – which in turn is also manifest in a decreased mean sea level pressure field over the region (not shown). This reduction leads to an improved comparison with observations (ERA-I reanalysis) over the Canadian Basin and Greenland, but slightly worse comparison with observations over the Barents Sea and Western Europe (Figure 8b). These pressure differences however are generally not significant save for a small patch over the Canadian Arctic Archipelago (Figure 8b). The 500 hPa and mean sea level decrease over the Arctic is suggestive of an increase in both the Arctic Oscillation (AO) and North Atlantic Oscillation (NAO) indices. This is consistent with other studies that have linked lower Arctic sea ice coverage with a tendency for a more meridional atmospheric jet (Francis and Vavrus, 2012), along with a tendency toward the negative phase of the NAO (Petoukhov and Semenov, 2010). However, owing to the small sample of years looked at here, it is doubtful we could establish a link with increased predictive skill of the inter-annual variability of the atmospheric mid-latitude circulation.

4.3 Impact of an improved model thickness climatology

The reanalysis comparison performed in Section 3 revealed persistent thickness distribution biases in the model relative to the CS2 derived data, whereby the ice was too thin in the Atlantic sector and too thick in the Pacific sector. As shown previously (Figure 6) these biases align very well with the ice edge errors suggesting a clear relationship between model thickness bias and forecast error. We would therefore like to understand whether the improvements we see in the GloSea seasonal forecasts are caused primarily by an improvement to the model's thickness climatology, or whether the inter-annual thickness distribution changes present in the observations are having an impact.

To try to answer this question another ensemble of seasonal hindcasts was performed, for years 2011-2014 only, using the 2015 sea ice initial conditions each year. This ensemble of hindcasts is denoted CLIM-2015. We note here that CLIM-2015 hindcasts are not performed for 2015 because they would simply be a duplication of the ThkDA-HC 2015 hindcasts. The motivation for doing things this way is to ensure that we have a dynamically self-consistent initial condition for the sea ice model. Simply averaging the initial conditions for the 5 years would potentially introduce some coupled initialisation shock that could make the results harder to analyse.

The total extent and IIEE relative to the CMEMS reanalysis for CLIM-2015 can be found in Figure 4 alongside those for the ThkDA-HC and CTRL-HC ensembles. The 2015 hindcasts for ThkDA-HC have been replicated and included as part of the CLIM-2015 experiment. As was the case for the ThkDA experiment, the CLIM-2015 hindcasts are much improved compared with the CTRL-HC experiment. The underestimation of Arctic-wide extent and the IIEE are both reduced. The

improvement seen in CLIM-2015 is similar in magnitude, but a little lower, than that seen with ThkDA-HC. Although results are worse for 2013, the extent and IIEE analysis in Figure 4 shows ThkDA-HC to be better than CLIM-2015 in 2011, 2012 and 2014. With only a short 4-year time series however, it is not possible to distinguish between the CLIM-2015 and ThkDA-HC runs statistically.

Interestingly the CLIM-2015 hindcasts show much reduced inter-annual variability when compared to those from the CTRL and ThkDA experiments and the ensemble-mean extents for each year are close. This is interesting, given that Arctic summer sea ice melt is strongly influenced by atmospheric variability (Deser et al., 2000), and suggests that the ensemble size of 24 used here is sufficient to remove atmospheric variability from the ensemble mean. It also suggests that the initial

10 Arctic thickness distribution and/or volume at the start of the melt season exhibits a controlling factor on the evolution of the ice through the melt season and the eventual September mean extent. This latter point is further supported by the fact that an additional ensemble of GloSea seasonal forecasts, performed using constant 2015 initial conditions for both the ocean and sea ice components, gave very similar results to that seen in the CLIM-2015 experiment (not shown).

5 Summary and conclusions

15 In this study, we have used nudging techniques to test the impact that initialising forecasts using CryoSat-2 (CS2) thickness measurements could have on Met Office seasonal forecasts of September sea ice extent. We have shown that initialisation of sea ice thickness significantly improves the accuracy of GloSea seasonal forecasts of summer sea ice cover. Biases in total Arctic extent are reduced as a whole and there are considerable improvements to the spatial distribution of sea ice and ice-edge location – particularly in the Atlantic sector. These improvements to the sea ice cover also lead to improvements in

20 near-surface temperature and pressure fields over the Arctic domain.

Technically the application of thickness increments within the CICE sea ice model has been shown to work well. The model is able to retain the information supplied by the thickness nudging all the way through the summer when thickness observations are absent. This is true during the GloSea coupled seasonal forecasts but also for the FOAM reanalysis in which

25 the sea ice model is also being modified by the assimilation of concentration. This result, which is also supported by the findings of Allard et al. (2018), increases our confidence that assimilating sea ice thickness using a more sophisticated and consistent approach will lead to improvements in the FOAM analyses as well as the short-range (FOAM) and seasonal (GloSea) forecasts initiated from them. Work is now underway at the Met Office, under the EU-SEDNA project, to include thickness within the fully balanced NEMOVAR 3D-Var FGAT variational scheme used in FOAM.

Confronting the sea ice thickness from the FOAM reanalysis with the CS2 satellite data has revealed a persistent bias in the modelled thickness distribution whereby the simulated Arctic sea ice is too thin on the Atlantic side and too thick in the

Beaufort Sea. This bias is most likely caused by deficiencies in the formulation of the sea ice dynamics: either the rheology – the so-called ice-ice force – or deficiencies in the momentum exchange between components in the atmosphere-ice-ocean (primarily wind drag). To ameliorate this situation we plan to experiment with the form-drag scheme and the anisotropic rheology developed for CICE by the CPOM group at University of Reading (Tsamados et al., 2013; 2014). In particular, the

5   form-drag scheme has been shown to improve the Arctic thickness distribution in standalone sea ice model experiments (D. Schroeder, pers. comm., 2017).

The clear relationship between modelled winter thickness biases and summer extent errors shown in Figure 6, along with the improved ice cover obtained using thickness initialisation (Figure 4 and Figure 5), highlights the importance that modelled

10   winter thickness biases can have on the evolution of forecast errors through the melt season. Results from the CLIM-2015 experiment further suggest that the ensemble size of 24 used here is sufficient to account for atmospheric variability and that late winter/spring sea ice thickness provides quite a strong constraint on the eventual September extent (Figure 4) in any particular year.

15   Although the addition of sea ice thickness nudging to the FOAM analysis system clearly improves the seasonal forecasts of summer sea ice, it is not clear how much of this improvement comes from initialising each year with the CS2 thickness and how much is down to the assimilation improving the model's thickness distribution climatology. The IIEE and extent analysis suggests that, for 3 out of the 4 years, using the correct thickness initialisation (ThkDA-HC) provides a better forecast of September ice edge location when compared with the run using the 2015 thickness (CLIM-2015). This is in

20   agreement with the findings of Day et al. (2014) who showed that, in the perfect model framework, that correct initialisation of Arctic thickness in the HadGEM1 climate model, led to an improved model evolution when compared with initialising the model with its own thickness climatology. In this case, however, we are not able to say this conclusively because the time series is too short to allow us to reject the null hypothesis that all the ensemble members from these two runs are taken from the same distribution.

Furthermore, the improvement shown in Figure 4 between ThkDA and CLIM-2015 is small relative to the improvement between ThkDA and CTRL. Therefore, we conclude that, certainly for the FOAM-GloSea system, improving the model thickness climatology is at least as important as initialisation of sea ice thickness for improving predictive skill of seasonal forecasts.

Acknowledgements. We acknowledge funding support from: the Joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01 101); the European Union's Horizon 2020 Research & Innovation programme through grant agreement No. 727862 (APPLICATE); the UK-China Research and Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund; and the UK Public Weather Service research

programme. We are thankful to Philip Davis (Met Office) for running and providing access to the CTRL-HC hindcast ensemble. Provision of observational data sources used within this study is also acknowledged: CryoSat2-derived thickness fields from CPOM; sea ice concentration products from OSI-SAF and NSIDC; ERA-Interim atmospheric reanalysis from ECMWF. EB would further like to thank Andy Ridout (CPOM, UCL) for providing monthly CS2 data and for his useful advice regarding aspects of the quality control.

References

Allard, R. A., Farrell, S. L., Hebert, D. A., Johnston, W. F., Li, L., Kurtz, N. T., Phelps, M. W., Posey, P. G., Tilling, R., Ridout, A., and Wallcraft, A. J.: Utilizing CryoSat-2 sea ice thickness to initialize a coupled ice-ocean modeling system, Advances in Space Research, doi:10.1016/j.asr.2017.12.030, 2018.

Balmaseda, M. A., Ferranti, L., Molteni, F. and Palmer, T. N.: Impact of 2007 and 2008 Arctic ice anomalies on the atmospheric circulation: Implications for long-range predictions. Q.J.R. Meteorol. Soc., 136: 1655-1664, doi:10.1002/qj.661, 2010.

Balmaseda, M. A., Hernandez, F., Storto, A., Palmer, M. D., Alves, O., Shi, L., Smith, G. C., Toyoda, T., Valdivieso, M., Barnier, B., Behringer, D., Boyer, T., Chang, Y.-S., Chepurin, G. A., Ferry, N., Forget, G., Fujii, Y., Good, S., Guinehut, S., Haines, K., Ishikawa, Y., Keeley, S., Köhl, A., Lee, T., Martin, M., Masina, S., Masuda, S., Meyssignac, B., Mogensen, K., Parent, L., Peterson, K. A., Tang, Y. M., Yin, Y., Vernieres, G., Wang, X., Waters, J., Wedd, R., Wang, O., Xue, Y., Chevallier, M., Lemieux, J.-F., Dupont, F., Kuragano, T., Kamachi, M., Awaji, T., Caltabiano, A., Wilmer-Becker, K., Gaillard, F.: The Ocean Reanalyses Intercomparison Project (ORA-IP), J Oper. Oceanogr. 8(S1):s80-s97, doi:10.1080/1755876X.2015.1022329, 2015.

Bauer, P., Magnusson, L., Thépaut, J.-N., and Hamill, T. M.: Aspects of ECMWF model performance in polar areas. Q.J.R. Meteorol. Soc., 142: 583-596. doi:10.1002/qj.2449, 2016.

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

Blanchard-Wrigglesworth, E. and Bitz, C. M.: Characteristics of Arctic Sea-Ice Thickness Variability in GCMs. J. Climate, 27, 8244-8258, doi:10.1175/JCLI-D-14-00345.1, 2014.

Blanchard-Wrigglesworth, E., Cullather, R. I., Wang, W., Zhang, J., and Bitz, C. M.: Model forecast skill and sensitivity to initial conditions in the seasonal Sea Ice Outlook, Geophys. Res. Lett., 42, 8042-8048, doi:10.1002/2015GL065860, 2015.

Blanchard-Wrigglesworth, E., Barthélemy, A, Chevallier, M., Cullather, R., Fučkar, N., Massonnet, F., Posey, P., Wang, W.,
5 Zhang, J., Ardilouze, C., Bitz, C. M., Vernieres, G., Wallcraft, A., and Wang, M.: Multi-model seasonal forecast of Arctic sea-ice: forecast uncertainty at pan-Arctic and regional scales, Clim. Dynam., 49, 1399-1410, doi:10.1007/s00382-016-3388-9, 2017.

Blockley, E. W., Martin, M. J., McLaren, A. J., Ryan, A. G., Waters, J., Lea, D. J., Mirouze, I., Peterson, K. A., Sellar, A.,
10 and Storkey, D.: Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts, Geosci. Model Dev., 7, 2613-2638, doi:10.5194/gmd-7-2613-2014, 2014.

Blockley, E., Lea, D., Martin, M ., McLaren, A., Mirouze, I., Ryan, A., Siddorn, J., Storkey, D., and Waters, J.: Recent development and assessment of the Met Office operational global ocean forecasting system (FOAM), in Operational
15 Oceanography for Sustainable Blue Growth. Proceedings of the Seventh EuroGOOS International Conference, 28-30 October 2014, Lisbon, Portugal. Buch E, Antoniou Y, Eparkhina D, Nolan G (Eds.) ISBN 978-2-9601883-1-8, p425-428, 2015.

Bloom, S. C., Takacs, L. L., da Silva, A. M., and Ledvina, D.: Data assimilation using incremental analysis updates, Mon.
20 Weather Rev., 124, 1256-1271, 1996.

Brown, A., Milton, S., Cullen, M., Golding, B., Mitchell, J., and Shelly, A.: Unified Modeling and Prediction of Weather and Climate: A 25-Year Journey. Bull. Amer. Meteor. Soc., 93, 1865-1877, doi:10.1175/BAMS-D-12-00018.1, 2012.

25 Chevallier, M., Salas y Mélia, D., Voldoire, A., Déqué, M., and Garric, G.: Seasonal forecasts of the pan-Arctic sea ice extent using a GCM-based seasonal prediction system, J. Clim., 26(16), 6092-6104, 2013.

Chevallier, M., Smith, G.C., Dupont, F., et al.: Intercomparison of the Arctic sea ice cover in global ocean-sea ice reanalyses from the ORA-IP project, Clim Dyn 49: 1107, doi:10.1007/s00382-016-2985-y, 2017.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A. J. and Wehner, M.: Long-term Climate Change: Projections, Commitments and Irreversibility. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to

the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Collow, T.W., Wang, W., Kumar, A., and Zhang, J.: Improving Arctic Sea Ice Prediction Using PIOMAS Initial Sea Ice Thickness in a Coupled Ocean-Atmosphere Model. Mon. Wea. Rev., 143, 4618-4630, doi:10.1175/MWR-D-15-0097.1, 2015.

Day, J. J., Hawkins, E., and Tietsche, S.: Will Arctic sea ice thickness initialization improve seasonal forecast skill?, Geophys. Res. Lett., 41, 7566-7575, doi:10.1002/2014GL061694, 2014.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137, 553-597, doi:10.1002/qj.828, doi:10.1002/qj.828, 2011.

Deser, C., Walsh, J.E., and Timlin, M.S.: Arctic sea ice variability in the context of recent atmospheric circulation trends, J. Climate, 13, 617-633, doi:10.1175/1520-0442(2000)013<0617:ASIVIT>2.0.CO;2, 2000.

Eicken, H.: Arctic sea ice needs better forecasts, Nature, 497, 431-433, doi:10.1038/497431a, 2013.

Fetterer, F., Knowles, K., Meier, W., Savoie, M., and Windnagel, A.: Sea Ice Index, Version 2, updated daily. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center, doi:10.7265/N5736NV7, 2016.

Francis, J. A., and S. J. Vavrus: Evidence linking Arctic amplification to extreme weather in mid-latitudes, Geophys. Res. Lett., 39, L06801, doi:10.1029/2012GL051000, 2012.

Goessling, H. F., Tietsche, S., Day, J. J., Hawkins, E., and Jung, T.: Predictability of the Arctic sea ice edge, Geophys. Res. Lett., 43, 1642-1650, doi:10.1002/2015GL067232, 2016.

Goosse, H., Arzel, O., Bitz, C. M., de Montety, A., and Vancoppenolle, M.: Increased variability of the Arctic summer ice extent in a warmer climate, Geophys. Res. Lett., 36, L23702, doi:10.1029/2009GL040546, 2009.

Guemas, V., Blanchard-Wrigglesworth, E., Chevallier, M., Day, J. J., Déqué, M., Doblas-Reyes, F. J., Fučkar, N. S., Germe, A., Hawkins, E., Keeley, S., Koenigk, T., Salas y Mélia, D., and Tietsche, S.: A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales. Q.J.R. Meteorol. Soc., 142: 546-561. doi:10.1002/qj.2401, 2016.

5   Hewitt, H. T., Copsey, D., Culverwell, I. D., Harris, C. M., Hill, R. S. R., Keen, A. B., McLaren, A. J., and Hunke, E. C.: Design and implementation of the infrastructure of HadGEM3: the next-generation Met Office climate modelling system, Geosci. Model Dev., 4, 223-253, doi:10.5194/gmd-4-223-2011, 2011.

Holland, M.M., Bailey, D.A., and Vavrus, S.: Inherent sea ice predictability in the rapidly changing Arctic environment of
10 the Community Climate System Model, version 3, Clim Dyn 36, 1239, doi:10.1007/s00382-010-0792-4, 2011.

Hunke, E.C., Lipscomb, W. H., Turner, A. K., Jeffery, N., Elliott, S.: CICE: the Los Alamos sea ice model documentation and software user's manual version 5.1, (LA-CC-06-012), Los Alamos National Laboratory, N.M., 2015.

15 Ivanova, N., Pedersen, L. T., Tonboe, R. T., Kern, S., Heygster, G., Lavergne, T., Sørensen, A., Saldo, R., Dybkjær, G., Brucker, L., and Shokr, M.: Inter-comparison and evaluation of sea ice algorithms: towards further identification of challenges and optimal approach using passive microwave observations, The Cryosphere, 9, 1797-1817, doi:10.5194/tc-9-1797-2015, 2015.

20 Jackson, L. C., Peterson, K. A., Roberts, C. D., and Wood, R. A.: Recent slowing of Atlantic overturning circulation as a recovery from earlier strengthening, Nat. Geosci, 9, 518-522, doi:10.1038/ngeo2715, 2015.

Jung, T., Gordon, N. D., Bauer, P., Bromwich, D. H., Chevallier, M., Day, J. J., Dawson, J., Doblas-Reyes, F., Fairall, C., Goessling, H. F., Holland, M., Inoue, J., Iversen, T., Klebe, S., Lemke, P., Losch, M., Makshtas, A., Mills, B., Nurmi, P.,
25 Perovich, D., Reid, P., Renfrew, I. A., Smith, G., Svensson, G., Tolstykh, M., and Yang, Q.: Advancing Polar Prediction Capabilities on Daily to Seasonal Time Scales. Bull. Amer. Meteor. Soc., 97, 1631-1647, doi:10.1175/BAMS-D-14-00246.1, 2016.

Kaleschke, L., Tian-Kunze, X., Maaß, N., Beitsch, A., Wernecke, A., Miernecki, M., Müller, G., Fock, B. H., Gierisch, A.
30 M., Schlünzen, K. H., Pohlmann, T., Dobrynin, M., Hendricks, S., Asseng, J., Gerdes, R., Jochmann, P., Reimer, N., Holfort, J., Melsheimer, C., Heygster, G., Spreen, G., Gerland, S., King, J., Skou, N., Søbjærg, S. S., Haas, C., Richter, F., and Casal, T.: SMOS sea ice product: Operational application and validation in the Barents Sea marginal ice zone, Remote Sensing of Environment, 180, 264–273, doi:10.1016/j.rse.2016.03.009, special Issue: ESA's Soil Moisture and Ocean Salinity Mission - Achievements and Applications, 2016.

Kauker, F., Kaminski, T., Karcher, M., Giering, R., Gerdes, R., and Voßbeck, M.: Adjoint analysis of the 2007 all time Arctic sea-ice minimum, Geophys. Res. Lett., 36, L03707, doi:10.1029/2008GL036323, 2009.

5 Keen, A.B., Hewitt, H.T. & Ridley, J.K.: A case study of a modelled episode of low Arctic sea ice, Clim. Dyn., 41, 1229, doi:10.1007/s00382-013-1679-y, 2013.

Koenigk, T., Caian, M., Nikulin, G., and Schimanke, S.: Regional Arctic sea ice variations as predictor for winter climate conditions, Clim Dyn 46: 317, doi:10.1007/s00382-015-2586-1, 2016.

Kwok, R., Cunningham, G. F., Wensnahan, M., Rigor, I., Zwally, H. J., and Yi, D.: Thinning and volume loss of the Arctic Ocean sea ice cover: 2003-2008, J. Geophys. Res., 114, C07005, doi:10.1029/2009JC005312, 2009.

Laxon, S., Peacock, N., and Smith, D.: High interannual variability of sea ice thickness in the Arctic region, Nature, 15 425(6961), 947-950, doi:10.1038/nature02050, 2003.

Laxon S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J., Haas, C., Hendricks, S., Krishfield, R., Kurtz, N., Farrell, S., and Davidson, M.: CryoSat-2 estimates of Arctic sea ice thickness and volume, Geophys. Res. Lett., 40, 732-737, doi:10.1002/grl.50193, 2013.

Lindsay, R. and Schweiger, A.: Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations, The Cryosphere, 9, 269-283, doi:10.5194/tc-9-269-2015, 2015.

MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., Gordon, M., Vellinga, M., Williams, 25 A., Comer, R. E., Camp, J., Xavier, P., and Madec, G.: Global Seasonal Fore cast System version 5 (GloSea5): a high resolution seasonal forecast system, Q. J. Roy. Meteor. Soc., doi:10.1002/qj.2396, 2014.

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

Martin, M. J., Balmaseda, M., Bertino, L., Brasseur, P., Brassington, G., Cummings, J., Fujii, Y., Lea, D. J., Lellouche, J.-M., Mogensen, K., Oke, P. R., Smith, G. C., Testut, C.-E., Waagbø, G. A., Waters, J., and Weaver, A. T.: Status and future of   data   assimilation   in   operational   oceanography,   ,   J. Oper. Oceanogr., 8, sup1, s28-s48, doi:10.1080/1755876X.2015.1022055, 2015.

Meier, W. N., Hovelsrud, G. K., van Oort, B. E. H., Key, J. R., Kovacs, K. M., Michel, C., Haas, C., Granskog, M. A., Gerland, S., Perovich, D. K., Makshtas, A., and Reist, J. D.: Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity, Rev. Geophys., 52, 185-217, doi:10.1002/2013RG000431, 2014.

Msadek, R., Vecchi, G., Winton, M., and Gudgel, R: Importance of initial conditions in seasonal predictions of Arctic sea ice extent, Geophys. Res. Lett., 41, 5208-5215, doi:10.1002/2014GL060799, 2014.

Mu, L., Yang, Q., Losch, M., Losa, S. N., Ricker, R., Nerger, L. and Liang, X.: Improving sea ice thickness estimates by
10 assimilating CryoSat-2 and SMOS sea ice thickness data simultaneously. Q.J.R. Meteorol. Soc.. doi:10.1002/qj.3225, 2018.

Notz, D.: Sea-ice extent and its trend provide limited metrics of model performance, The Cryosphere, 8, 229-243, doi:10.5194/tc-8-229-2014, 2014.

15 Notz, D. and Stroeve, J.: Observed Arctic sea-ice loss directly follows anthropogenic CO_2 emission, Science, 354, 747-750, doi:10.1126/science.aag2345, 2016.

OSI-SAF: EUMETSAT Ocean and Sea Ice Satellite Application Facility. Global sea ice concentration climate data record 1979-2015 (v2.0). Norwegian and Danish Meteorological Institutes. doi:10.15770/EUM_SAF_OSI_0008, 2017.

Peterson, K. A., Arribas, A., Hewitt, H. T., Keen, A. B., Lea, D. J., and McLaren, A. J.: Assessing the forecast skill of Arctic sea ice extent in the GloSea4 seasonal prediction system, Clim. Dynam., 1-16, doi:10.1007/s00382-014-2190-9, 2014.

Petoukhov, V., and V. A. Semenov: A link between reduced Barents-Kara sea ice and cold winter extremes over northern
25 continents, J. Geophys. Res., 115, D21111, doi:10.1029/2009JD013568, 2010.

Rayner, N. A.; Parker, D. E.; Horton, E. B.; Folland, C. K.; Alexander, L. V.; Rowell, D. P.; Kent, E. C.; Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, J. Geophys. Res.Vol. 108, No. D14, 4407 doi:10.1029/2002JD002670, 2003.

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

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

5   Ridley, J. K., Blockley, E. W., Keen, A. B., Rae, J. G. L., West, A. E., and Schroeder, D.: The sea ice model component of HadGEM3-GC3.1, Geosci. Model Dev., 11, 713-723, doi:10.5194/gmd-11-713-2018, 2018.

Roberts, C. D., Waters, J., Peterson, K. A., Palmer, M. D., McCarthy, G. D., Frajka-Williams, E., Haines, K., Lea, D. J., Martin, M. J., Storkey, D., Blockley, E. W., and Zuo, H.: Atmosphere drives recent interannual variability of the Atlantic
10   meridional overturning circulation at 26.5°N, Geophys. Res. Lett., 40, 5164-5170 doi:10.1002/grl.50930, 2013.

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., Eade, R., Fereday, D., Folland, C. K., Gordon, M., Hermanson, L., Knight, J. R., Lea, D. J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A. K., Smith, D., Vellinga, M., Wallace, E., Waters, J., and Williams, A.: Skillful long-range prediction of European and North American
15   winters, Geophys. Res. Lett., 41, doi:10.1002/2014GL059637, 2014.

Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., and Kwok, R.: Uncertainty in modeled Arctic sea ice volume, J. Geophys. Res., 116, C00D06, doi:10.1029/2011JC007084, 2011.

20   Screen, J. A.: Influence of Arctic sea ice on European summer precipitation, Environ. Res. Lett., 8(4), 044015, doi:10.1088/1748-9326/8/4/044015, 2013.

Smith, L. C., and Stephenson, S. R.: New Trans-Arctic shipping routes navigable by midcentury, Proc. Nat. Acad. Sci., 110(13), E1191-E1195, doi:10.1073/pnas.1214212110, 2013.
25
Stark, J. D., Ridley, J., Martin, M., and Hines, A.: Sea ice concentration and motion assimilation in a sea ice−ocean model, J. Geophys. Res., 113, C05S91, doi:10.1029/2007JC004224, 2008.

Storkey, D., Blaker, A.T., Mathiot, P., Megann, A., Aksenov, Y., Blockley, E. W., Calvert, D., Graham, T., Hewitt, H. T.,
30   Hyder, P., Kuhlbrodt, T., Rae, J. G. L., Sinha, B.: UK Global Ocean GO6 and GO7: a traceable hierarchy of model resolutions, Geosci. Model Dev. Discuss., in review, 2017.

Stroeve, J., Hamilton, L. C., Bitz, C. M., and Blanchard-Wrigglesworth, E.: Predicting September sea ice: Ensemble skill of the SEARCH Sea Ice Outlook 2008-2013, Geophys. Res. Lett., 41, 2411-2418, doi:10.1002/2014GL059388, 2014.

Tietsche, S., J. J. Day, V. Guemas, W. J. Hurlin, S. P. E. Keeley, D. Matei, R. Msadek, M. Collins, and E. Hawkins: Seasonal to interannual arctic sea ice predictability in current global climate models, Geophys. Res. Lett., 41, 1035-1043, doi:10.1002/2013GL058755, 2014.

Tilling, R. L., Ridout, A., Shepherd, A., and Wingham, D. J.: Increased Arctic sea ice volume after anomalously low melting in 2013, Nat. Geosci., 8, 643-646, 2015.

Tilling, R. L., Ridout, A., and Shepherd, A.: Near-real-time Arctic sea ice thickness and volume from CryoSat-2, The

10 Cryosphere, 10, 2003-2012, doi:10.5194/tc-10-2003-2016, 2016.

Tonani, M., Balmaseda, M., Bertino, L., Blockley, E. W., Brassington, G., Davidson, F., Drillet, Y., Hogan, P., Kuragano, T., Lee, T., Mehra, A., Paranathara, F., Tanajura, C. A. S. and Wang, H.: Status and future of global and regional ocean prediction systems, J. Oper. Oceanogr., 8, sup2, s201-s220, doi:10.1080/1755876X.2015.1049892, 2015.

Uotilla et al.,: An assessment of ten ocean reanalyses in the polar regions, submitted to Clim. Dyn, in review.

Vaughan, D.G., J.C. Comiso, I. Allison, J. Carrasco, G. Kaser, R. Kwok, P. Mote, T. Murray, F. Paul, J. Ren, E. Rignot, O. Solomina, K. Steffen and T. Zhang: Observations: Cryosphere. In: Climate Change 2013: The Physical Science Basis.

20 Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Walters, D., Baran, A., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J., Morcrette, C., Mulcahy, J., Sanchez, C., Smith, C., Stratton, R., Tennant, W., Tomassini, L., Van Weverberg, K., Vosper, S.,

25 Willett, M., Browse, J., Bushell, A., Dalvi, M., Essery, R., Gedney, N., Hardiman, S., Johnson, B., Johnson, C., Jones, A., Mann, G., Milton, S., Rumbold, H., Sellar, A., Ujiie, M., Whitall, M., Williams, K., and Zerroukat, M.: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-291, in review, 2017.

30 Waters, J., Lea, D. J., Martin, M. J., Mirouze, I., Weaver, A. T., and While, J.: Implementing a variational data assimilation system in an operational 1/4 degree global ocean model, Q. J. Roy. Meteor. Soc., doi:10.1002/qj.2388, 2014.

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., Hyder, P., Ineson, S., Johns, T. C., Keen, A. B., Lee, R. W., Megann, A., Milton, S. F., Rae, J. G. L., Roberts, M.

J., Scaife, A. A., Schiemann, R., Storkey, D., Thorpe, L., Watterson, I. G., Walters, D. N., West, A., Wood, R. A., Woollings, T., and Xavier, P. K.: The Met Office Global Coupled model 3.0 and 3.1 (GC3.0 & GC3.1) configurations, JAMES, doi:10.1002/2017MS001115, 2018.

5   Xie, J., Counillon, F., Bertino, L., Tian-Kunze, X., and Kaleschke, L.: Benefits of assimilating thin sea ice thickness from SMOS into the TOPAZ system, The Cryosphere, 10, 2745-2761, doi:10.5194/tc-10-2745-2016, 2016.

Zhang, J. and Rothrock, D. A.: Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates, Mon. Wea. Rev. 131(5), 681-697, 2003.

Table 1: Details of ocean-sea ice reanalysis experiments used in this study

Reanalysis	Run period	Surface forcing	Assimilated variables (3D-Var)	Thickness nudging used
CTRL-RA	01/01/1992 – 31/12/2015	ERA-Interim	SST, SLA, T&S, SIC	None
ThkDA-RA	01/10/2010 – 31/12/2015	ERA-Interim	SST, SLA, T&S, SIC	CryoSat-2

15   **Table 2: Details of GloSea coupled seasonal hindcast experiments used in this study**

Hindcasts	Run period	Years	Ensemble members	Atmosphere ICs	Ocean/Ice ICs
CTRL-HC	May – Sep	1992-2015	24 per year[a]	ERA-Interim	CTRL-RA
ThkDA-HC	May – Sep	2011-2015	24 per year[a]	ERA-Interim	ThkDA-RA
CLIM-2015	May – Sep	2011-2014	24 per year[a]	ERA-Interim	ThkDA-RA: 2015 for all years

[a] 24 members per year = 3 start dates with 8 stochastic members for each
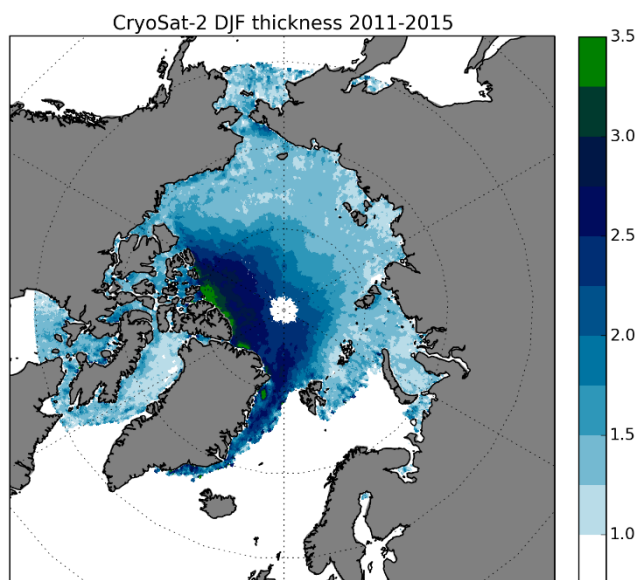
Figure 1: Example CPOM Arctic thickness used in this study derived from CryoSat-2 measurements: showing December-January-February (winter) mean thickness (m) over the period 2011-2015 after application of the QC and imposing the 1m minimum thickness threshold.
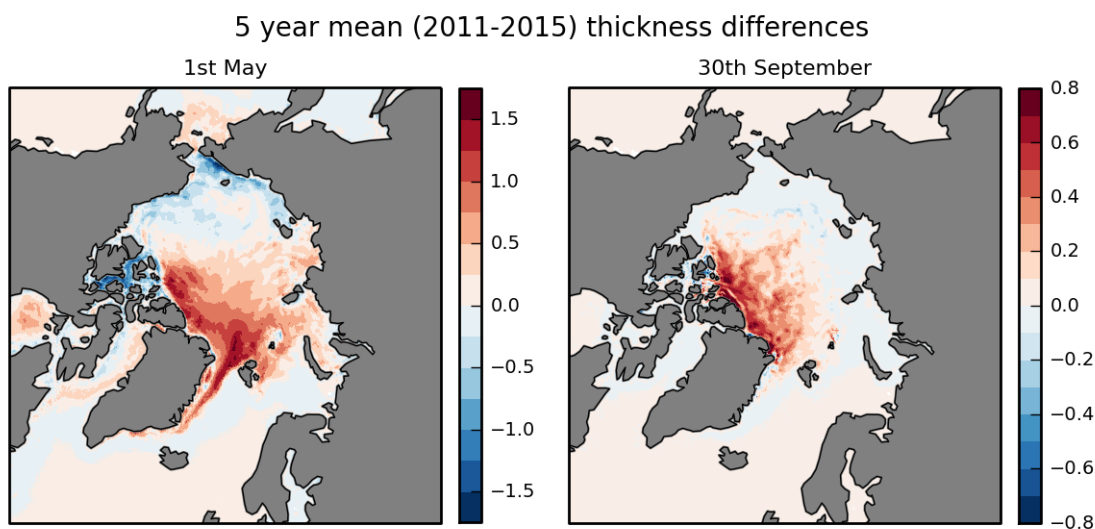
Figure 2: Mean sea ice thickness difference (m) between ThkDA-RA and CTRL-RA experiments on 1st May (left) and 30th September (right) over the full 5-year reanalysis period from 2011 to 2015. The difference is taken as ThkDA – CTRL so red (blue) implies that the CS2 nudging has increased (decreased) the thickness.
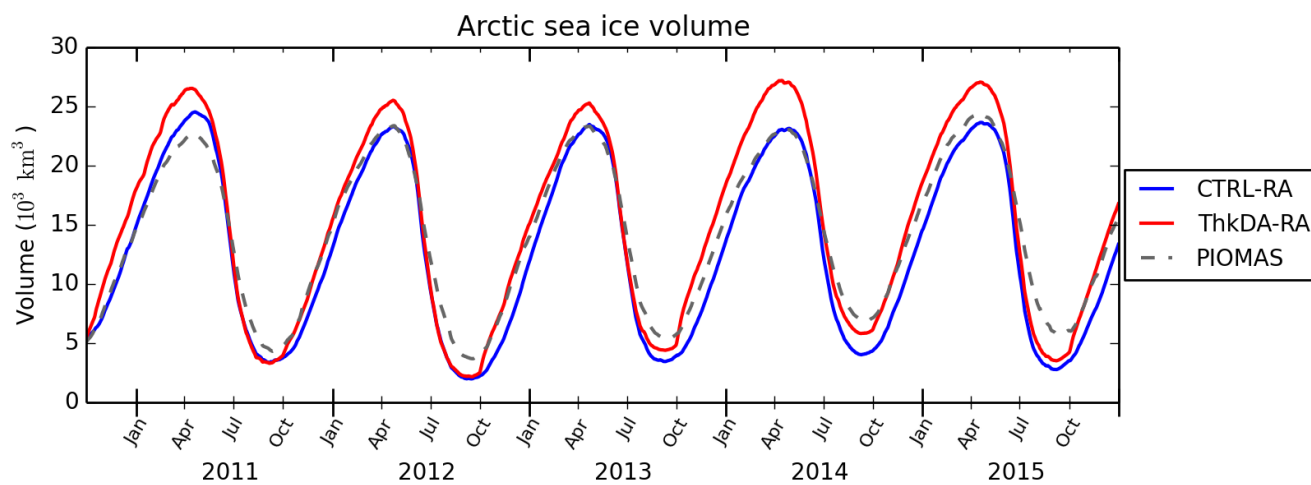
Figure 3: Reanalysis Arctic sea ice volume from 1st October 2010 to the end of December 2015 from the CTRL-RA (blue) and ThkDA-RA (red) experiments and from the PIOMAS model (grey dashed).
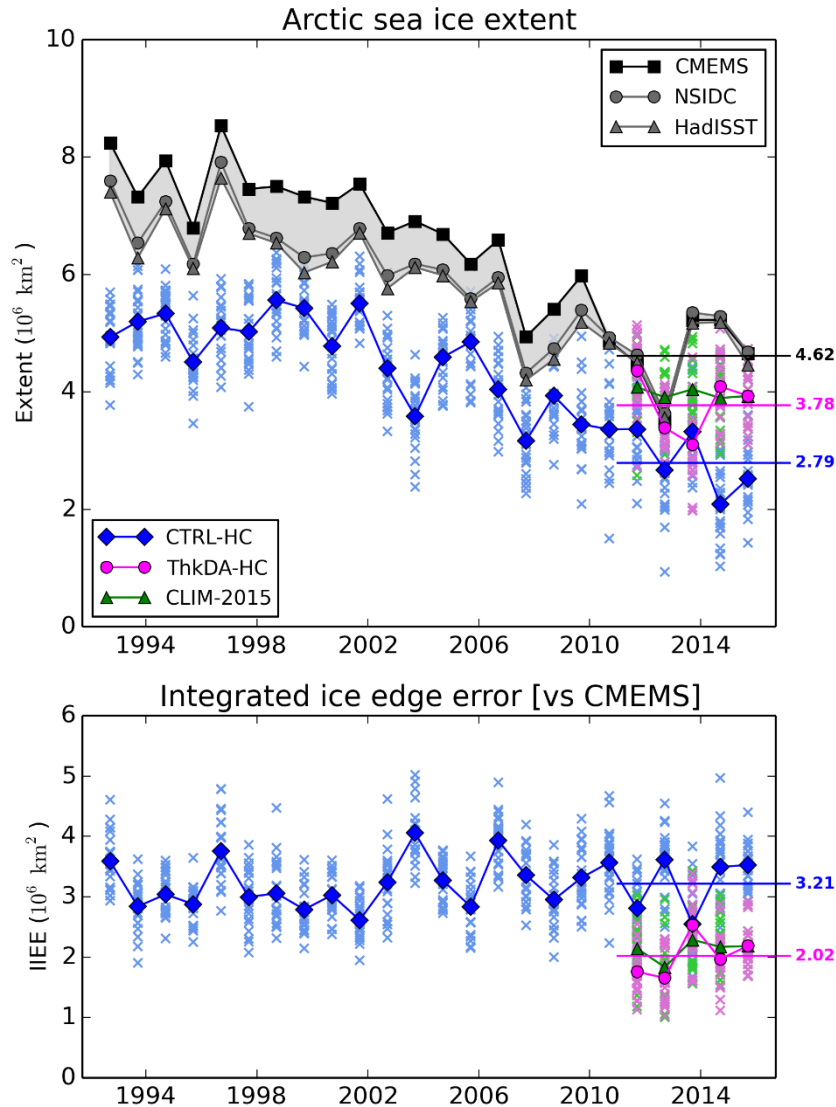
Figure 4: (upper) Arctic sea ice extent for CTRL-HC (blue), ThkDA-HC (pink) and CLIM-2015 (green) seasonal forecast experiments. Observational estimates from the CMEMS reanalysis assimilating OSI-SAF (black square), NSIDC (grey circles) and HadISST1.2 (grey triangles) are included and the area between them shaded light grey. **(lower)** Integrated Ice Edge Error (IIEE) for forecasts relative to the CMEMS reanalysis product introduced in Section 2.3. Individual ensemble members are represented by coloured crosses and ensemble means by the solid symbols and inter-connecting lines. Horizontal coloured lines depict 2011-15 mean values.
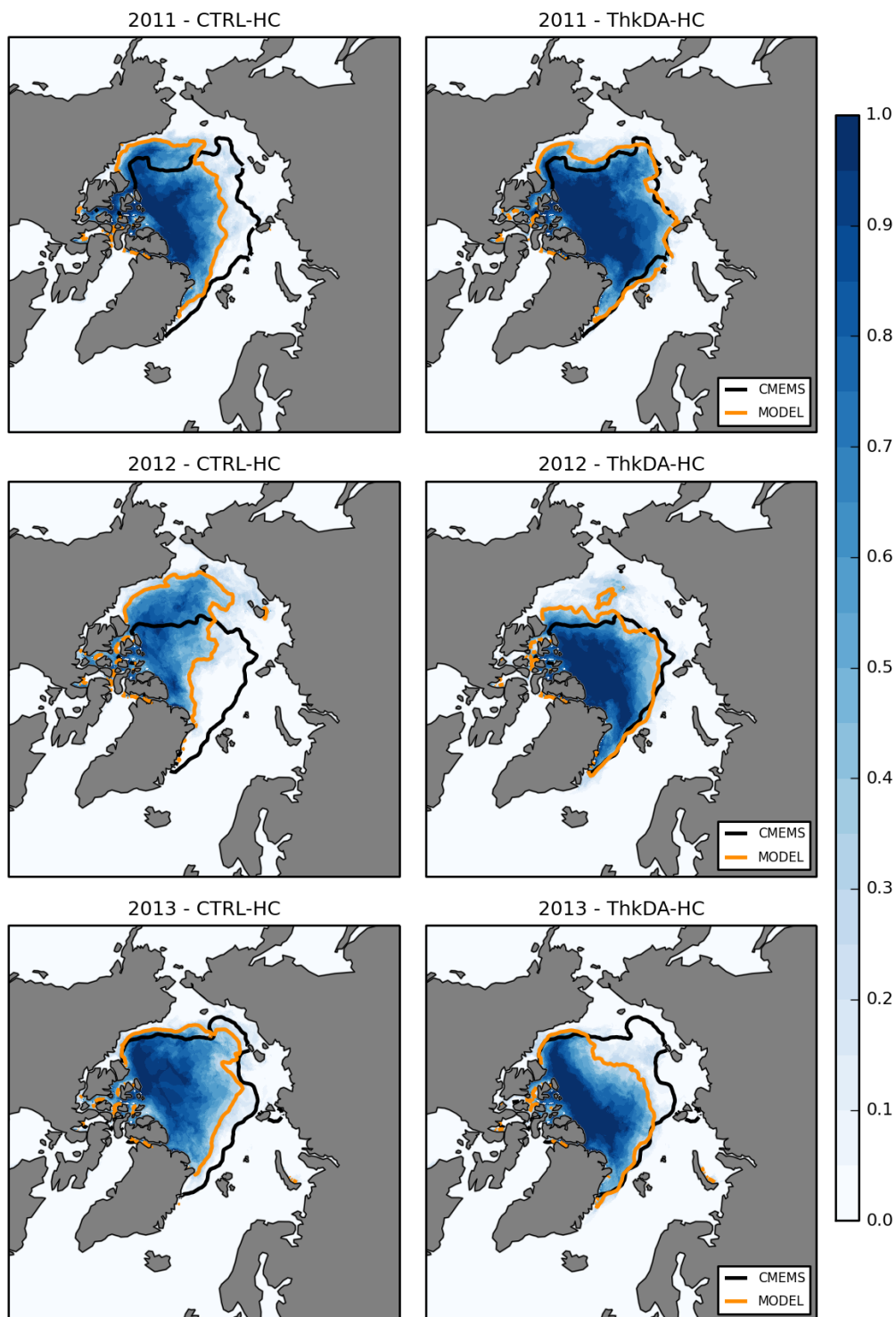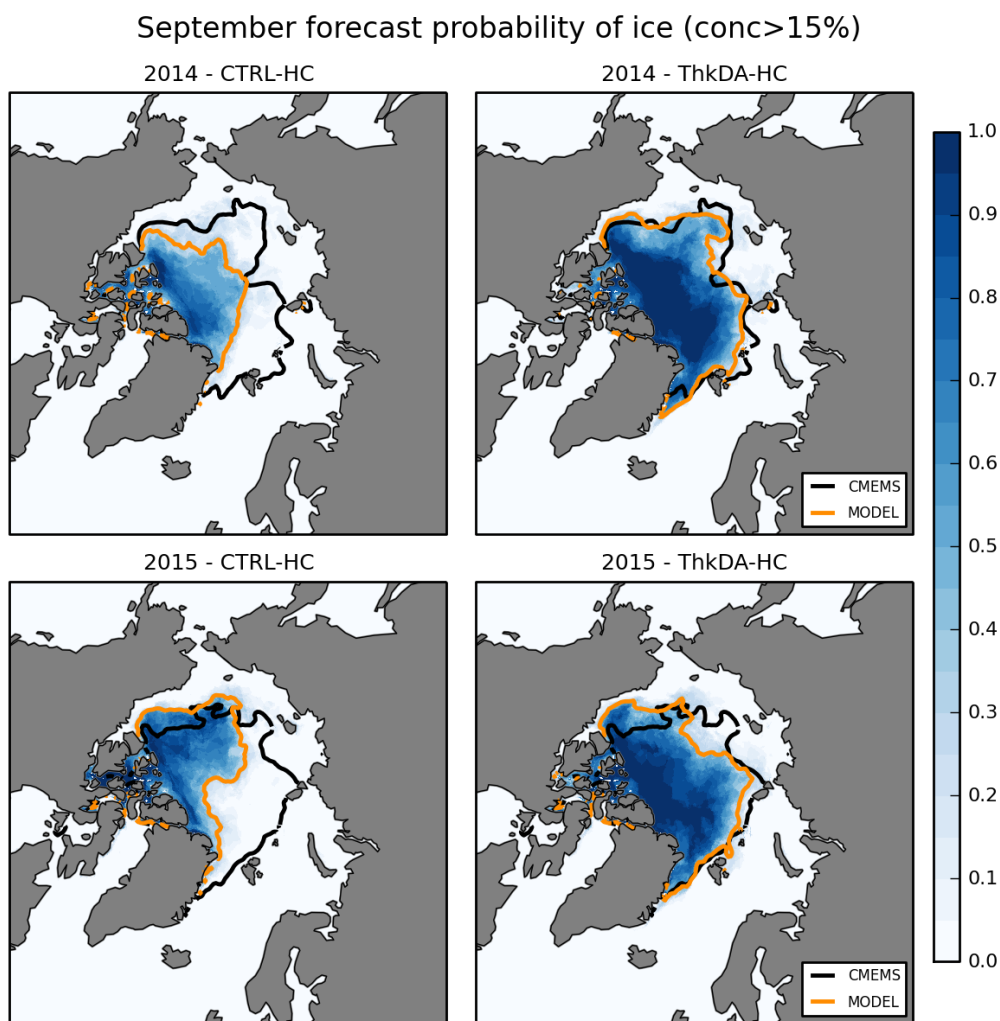
Figure 5: September mean probability of sea ice for the CTRL-HC (left) and ThkDA-HC (right) seasonal forecasts for all years from 2011 (top) to 2015 (bottom). Contours of 15% concentration are overlain to represent the sea ice edge for the forecast
5 **ensemble mean (orange) and CMEMS reanalysis product (black). Probability is defined at each point as the proportion of ensemble members that have at least 15% ice concentration.**
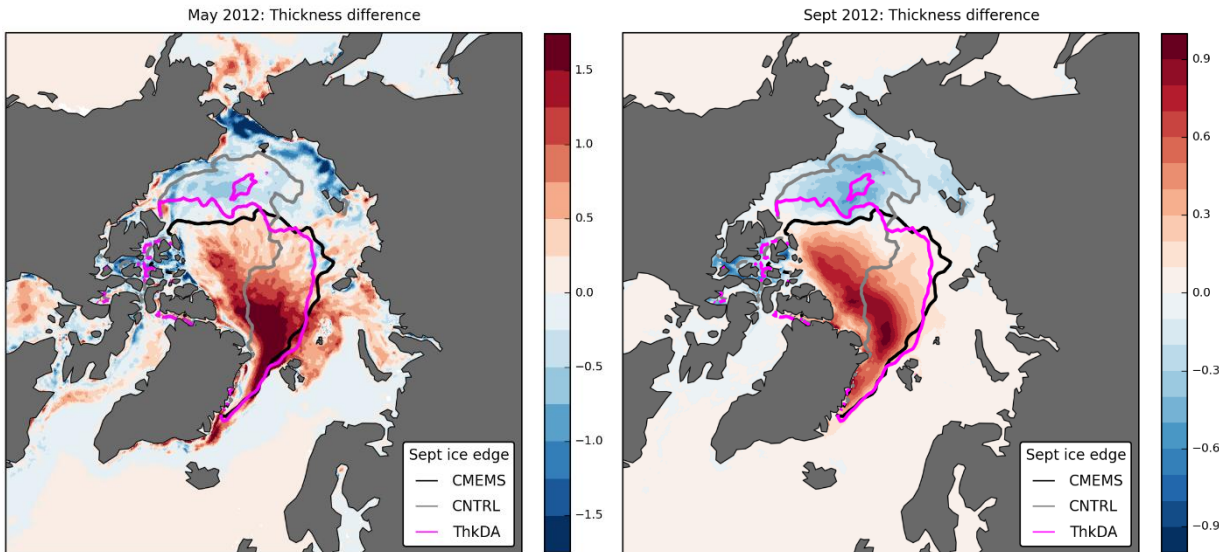
Figure 6: Thickness difference (m) between ThkDA and CTRL experiments in May and September 2012 (contour shading) with differences calculated as ThkDA – CTRL in each case. (left) shows difference between reanalyses around the start of the seasonal hindcasts on 1st May 2012. (right) shows ensemble mean forecast differences for September 2012. (Note the different scales used for the coloured shading). Overlain on both panels are September-mean contours of 15% ice concentration to represent the sea ice edge for the CTRL-HC (grey) and ThkDA-HC (pink) experiments along with the CMEMS reanalysis product (black).
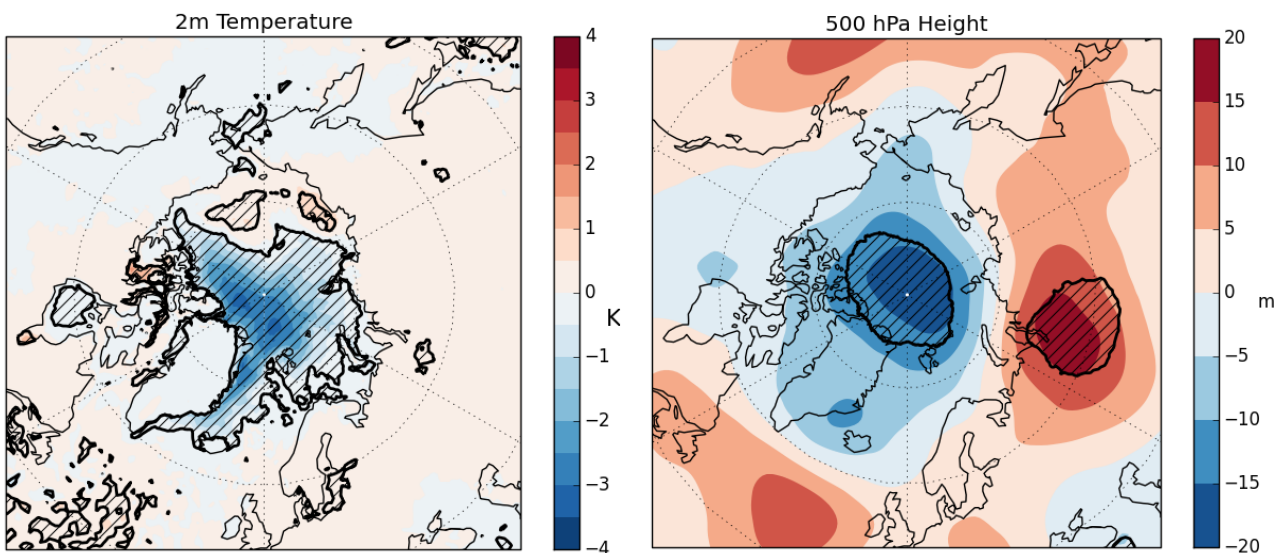
Figure 7: September differences in (left) near-surface (2m) air temperature (K), and (right) 500 hPa height (m) between the ThkDA and CTRL predictions averaged over all ensemble members for the 5-year period 2011-2015. Differences are calculated as ThkDA – CTRL meaning that areas of blue (red) denote that ThkDA predictions are lower (higher). Black contours and hatching denote areas where differences are significant at the 95% level as determined using a Mann-Whitney U-test.
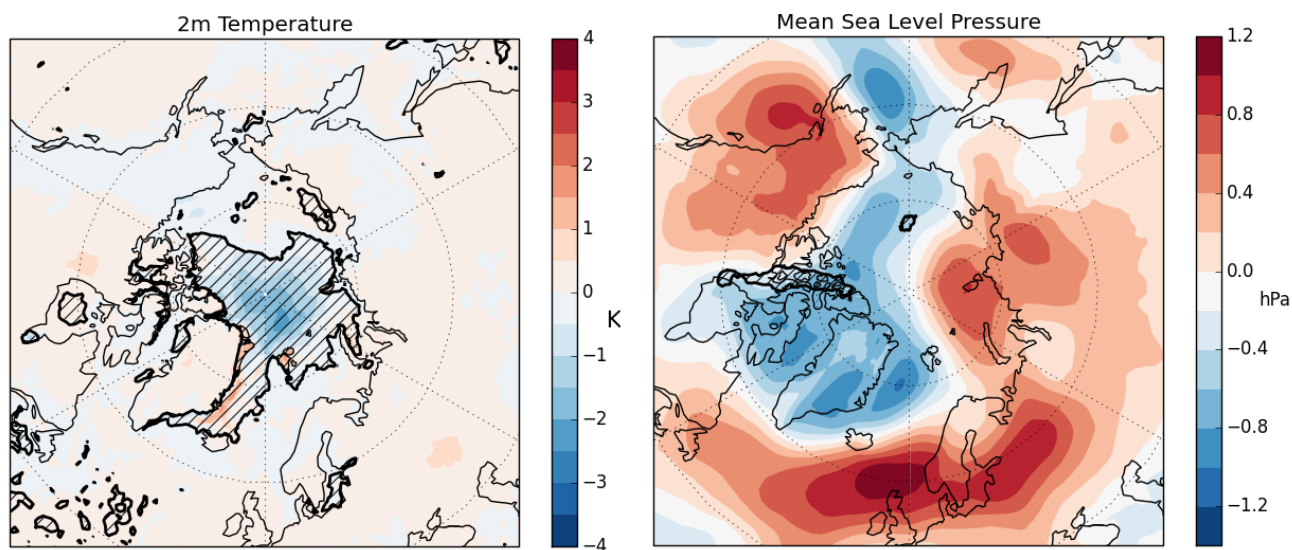
Figure 8: Difference in root-mean-square-error (RMSE) between individual ensemble members and the ERA-I reanalysis averaged over the 5-year period 2011-2015 for (left) near-surface (2m) air temperature (K), and (right) mean sea-level-pressure (hPa). Differences are calculated as ThkDA – CTRL meaning that areas of blue (red) denote that ThkDA predictions are better (worse). Black contours and hatching denote areas where differences are significant at the 95% level as determined using a Mann-Whitney U-test.