Authors' final response

We thank the referees for their review of our manuscript. Answers to the specific comments in the review are listed inline below. The review comments are in *black italics*, our response is in red, and details of manuscript changes are in blue.

Page and line numbers given here refer to the revised manuscript version with full mark-up.

Response to review #1

While we have taken these comments seriously, it is clear that many of the comments stem from a misinterpretation of the goal of this paper. Our goal here is to show the impact, on coupled seasonal predictions, of including CPOM CS2 sea ice thickness observations within the initialisation of the GloSea coupled model. Whilst the production of a sea ice analyses, for use as initial conditions, was therefore crucial to this stated goal, it was not, unlike for Allard et al. (2018), the focus of this manuscript.

Having shown the usefulness of the sea ice thickness initialisation on the seasonal forecast system, we have laid the groundwork and motivation for a proper data assimilation treatment to use sea ice thickness observations in an ocean and sea ice analysis. That work would require the sort of data comparison with which many of the reviewer's comments are concerned. However, an in-depth evaluation of the thickness analyses, and comparison, of several sea ice thickness observational estimates, is outside the scope of the current manuscript.

Changes to the manuscript: The 2 sentences in the penultimate paragraph of the Introduction have been expanded to form a new hypothesis paragraph. (P4 L17+) Additionally more information is provided in the Summary section about what would be required to implement SIT assimilation fully within the GloSea initialisation using the NEMOVAR 3D-Var assimilation scheme (P17 L11+)

A general gripe of this study and similar ones: I find it really challenging to understand what you mean by the use of forecast and hindcast throughout the paper and they seem to be used interchangeably. If you are using prescribed atmospheric forcings that have assimilated real data then my view is anything using that is a hindcast not a forecast. Prediction is the more general term that could be appropriate but I see nothing in this paper that resembles a true forecast (no future knowledge), despite the title.

We would disagree about the definition of 'hindcast' here. A hindcast is simply a retrospective forecast - performed under the same conditions as a true forecast but done when the result is already known. For example, when we perform hindcasts to test the forecast skill of the FOAM ocean-ice only forecasting system (Blockley et al., 2014), this is done using atmospheric forcing that has not assimilated data. However, this distinction is not relevant here because we do not use any atmospheric forcing for the coupled atmosphere-ocean-sea ice-land seasonal hindcasts performed within this study. What is important here, is that a hindcast is used to test the expected skill of a real forecast – and must be done in a fashion that does not use **any further**, subsequent, observational data after initialisation so as to invalidate that expectation.

As well as defining whether the prediction is made for a known past state (hindcast) or an unknown future state (forecast), the terms hindcast and forecast are also used within the

GloSea seasonal prediction system (see MacLachlan et al. 2014) as a technical distinction. GloSea forecasts are carried out each day (2 members per day for the 210-day forecasts), whereas hindcasts – used for testing purposes and for the bias correction – are carried out for only 4 start dates per month but with more (8) ensemble members. This technical distinction has not been explained in this manuscript (although it is in the cited material) because it is not relevant to understanding our results.

Changes to the manuscript: use of 'forecast' and 'hindcast' have been replaced with 'prediction' where appropriate throughout the manuscript (including the title). However, the use of 'forecast' as an adjective (e.g. 'forecast skill') has remained. We have also made it clearer that the GloSea seasonal predictions made here (and generally within the hindcast context) are done without any knowledge of future observations and so are indeed 'true forecasts'. (P5 L14; P11 L27-29)

Main comments

It's really not clear from the motivation what it is you are trying to achieve by assimilating sea ice thickness in this study. In some cases you say the ice area/extent impact is negligible (as you ignore lower thickness ice) but other times it seems you highlight big improvements in your ice edge 'forecast'. You should really present a hypothesis you are testing in this kind of study (i.e. which metrics you are assessing). In general I would think assimilating thin ice should be especially important for seasonal (spring/summer) forecasting as you want to correctly incorporate this into your model as this is the ice most likely to melt out. If you get that wrong, you get the summer melt wrong. It seems like you used the opposite reasoning to justify not using the AWI/SMOS data but your results suggest the opposite to be true if you care about both the summer ice thickness and ice edge.

Changes to the manuscript: As mentioned above, the 2 sentences in the penultimate paragraph of the Introduction have been expanded to form a new hypothesis paragraph. (P4 L17+) We have also dropped the additional motivation/justification for using CS2 rather than SMOS from Section 2.2.1 to avoid confusion. The story is now clearer: we are interested in the impact of thick ice initialisation on the skill of seasonal predictions of Arctic September sea ice cover and hypothesise that including CS2 within the initialisation will improve skill.

I had a lot of issues with Section 2.2 (describing the CryoSat-2 data):

 - You need to make clearer the various thickness datasets available and how a number of groups are now routinely producing sea ice thickness estimates, e.g. NASA/AWI/CPOM for CryoSat-2. - CPOM and NASA data were used in the study of Allard 2018. You need to make this clearer. It was in general unclear how your study differed from Allard and I think you need more discussion of their approach and results.

Allard et al, (2018), further backed up by Stroeve et al, (2018), show the differences between the various datasets are not particularly major. Differences in the CS2 products arising from the retracking and processing algorithm differences are very small in comparison to the differences that nudging to CPOM CS2 data has on our sea ice thickness initial conditions.

As our focus here is on the large-scale impact of initialising thick Arctic sea ice from CS2 within a fully coupled seasonal prediction system (GloSea). This is very different from the work of Allard et al. whose focus is on a reanalysis of sea ice thickness made using an externally forced ocean-sea ice model. An in-depth evaluation and comparison of sea ice thickness

observational estimates, is outside the scope of our manuscript but is comprehensively addressed in Rick's. Having no desire to repeat the work already done in the Allard et al. (& now Stroeve et al.) studies, we have **chosen** here to test/show the impact of assimilating CPOM CS2 data in a seasonal prediction system.

Changes to the manuscript: We have expanded the introduction of Allard et al. (2018) to make it clear that they run an externally forced ocean-sea ice model and look at analyses and short-range forecasts. We further state that Allard et al. use multiple thickness datasets processed with different algorithms and perform an in-depth analysis of the data. We also explicitly state that what has not been previously investigated is the impact that assimilation of sea ice thickness may have in longer (>90 days), forecasts made using fully coupled systems. (P4 L7+) When introducing the CS2 data in Section 2.2.1, we have added a sentence stating that other centres provide CS2-derived Arctic sea ice thickness products. An additional citation to Allard et al. (2018) is included (who compare the different products) as well as a citation to Stroeve et al. (2018). (P7 L7-9) The new hypothesis/motivation paragraph in the Introduction, will tell the reader that our study is focussed on initialising thick sea ice using CS2 and testing the impact on seasonal predictions of September sea ice cover (and not a summary, or comparison, of available sea ice thickness datasets). (P4 L17+)

 - Is the pole hole and the data uncertainty really why these data haven't been used? It seems like any reasonable assimilation scheme shouldn't need complete coverage and can factor in data uncertainties. My guess is that the main reason was data availability, the fact these thickness products were in their infancy, and inertia.

Yes data availability is also important but so is the pole-hole. Prior to CS2 the pole hole was very large. For ERS 1&2 it essentially covered all the central Arctic basin. A large pole-hole is not a big problem for the SIC assimilation because, historically at least, the ice is, fairly uniformly, close to 100% concentration near the North Pole. However for SIT, the situation is not so simple because thickness gradients across this region are quite large, meaning that any DA scheme attempting to spread the information would be heavily reliant on models to get this right. Coupled with the observational uncertainty issues this makes for a daunting problem.

Changes to the manuscript: The order that these 3 reasons are listed in has been switched to put data availability first.

 - On that note, how do you treat the fact you are unlikely to have complete coverage from the CryoSat-2 data?

Using the nudging approach, we only modify the thickness fields where we have data (P9 L11-12) and no attempt is made to spread observational data. This is fine for our feasibility study with monthly binned data, but when we move on to assimilating the raw altimeter tracks in a full variational scheme, more work will be required to specify observational and model errors, covariances, etc.

Changes to the manuscript: More information is provided in the Summary section detailing what would be required to implement SIT assimilation fully within the GloSea initialisation using the NEMOVAR 3D-Var assimilation scheme (P17 L11+)

- After listing problems associated with generic thickness data you then say this is improved by the availability of CS-2 thickness data. This doesn't make much sense as it is written.

Changes to the manuscript: Section 2.2 has been modified to explicitly state how CS2 reduces the 3 problems we highlight (timeliness, pole-hole, accuracy). A new reference to Guerreiro et al. (2017) has been included. (P6 L25-28)

- You should cite the relevant studies regarding uncertainty estimates, not just the Ricker/AWI reference and apply that to CS-2 data derived by other groups. There are strong differences in the retracking procedures which may have impacts on respective data uncertainties across the products.

As we have previously stated, we are not concerned about the differences between these observational products in this study. Instead our motivation is the impact of SIT initialisation on seasonal predictions of Arctic sea ice cover (i.e., extent and ice-edge location). We use the – really quite comprehensive – studies of Ricker et al. here purely to provide approximate bounds below which the CS2 is likely to be of no value. However when we come to do this as a data assimilation problem – rather than a seasonal coupled prediction problem – we will be very interested in the observational uncertainty/properties.

Changes to the manuscript: A sentence has been added to Section 2.2.1 that states that other centres provide CS2-derived Arctic sea ice thickness products. An additional citation to Allard et al. (2018) is included (who compare the different products) as well as a citation to Stroeve et al. (2018). (P7 L7-9)

- I don't get what this extra quality checking of the CPOM CS-2 data is. The fact you have included a personal comms from one of the data producers of that dataset makes it seem like this is something they do too? What exactly do you mean by smearing?

This is related to the regridding and binning of the data performed by CPOM to ensure that high spatial gradients are not smeared (averaged) out. The approach taken here was that recommended by CPOM and so should be considered part of their observational processing.

Changes to the manuscript: This text has been removed from the revised version of the manuscript to prevent it causing a distraction to the reader.

Why not use daily along-track CryoSat-2 data? I thought this was the whole purpose of CPOM releasing the daily along-track data? Instead they grid the data, then you grid the data, then you interpolate to a get a daily thickness?

Yes, you are correct. Assimilating altimeter tracks of thickness (or more likely the raw freeboard) is the ultimate goal for SIT assimilation in our systems. However much work is required to do this. Observational errors need to be quantified (including representativeness), and model/observation covariances and correlations are required to spread the data from the tracks. Finally, balancing with ice concentration and other aspects of the model (SST, SSS) are required. Doing all this is a considerable undertaking and so, before doing this, we wanted to be sure that the SIT initialisation would have an impact on the model – hence this feasibility study using gridded data.

Changes to the manuscript: More information is now provided in the Summary section about what would be required to fully initialise GloSea sea ice thickness using the NEMOVAR 3D-Var assimilation scheme – which would use along-track altimeter data. (P17 L11+) The new hypothesis paragraph towards the end of the Introduction also includes mention that we are motivated here by wanting to assess the feasibility of SIT initialisation within GloSea (P4 L17+).

Hard to tell what this volume comparison really means. You compare with PIOMAS but then say that data is biased low (which I'm not actually sure is true when you look at more than one CS-2 estimate) so it is actually good that you are further from that data? You say this was expected but this seems like a hindsight statement to me. I agree PIOMAS data can provide useful context but I don't agree with how you've used it. I think you should just show the CS-2 data and say look, the assimilation does what it is supposed to do.

Yes, it was expected that the winter sea ice volume would be higher than PIOMAS. This was an obvious and logical expectation given that the CPOM CS2 data we are assimilating has higher volume than PIOMAS (as documented by the Laxon et al. and Tilling et al. CPOM studies). However, the key point here is that this doesn't actually matter. We only include PIOMAS comparison in our evaluation as a reference because it is well understood and widely used for this purpose. We do not use it for verification.

Changes to the manuscript: The manuscript has been changed to make it clear that we include PIOMAS as a reference only. In line with the final piece of advice here, Figure 1 has been expanded to show the impact of assimilating CS2 thickness data - 3 more panels have been added to show the modelled Oct-Apr thickness from the CTRL and ThkDA reanalyses, along with the difference between these two experiments (ThkDA-CTRL). An assessment of the impact of SIT initialisation on the sea ice volume between the CTRL and ThkDA experiments is now provided immediately after the discussion of Figure 1. To enable this, the order of Figure 2 and Figure 3 have been switched. (P9 L26+) No mention of PIOMAS is made until after this discussion and we qualify its inclusion by stating that it is used here purely as a reference because it is well understood and widely used for the purpose. (P10 L9+) Finally, the text stating that PIOMAS has been shown to underestimate thickness/volume in the winter has been changed to specifically state when compared with CPOM CS2 data. (P10 L12-15)

P10 L10-20 and elsewhere: Very confusing to me if these are hindcasts or genuine forecasts. You use both labels interchangeably. How could you move forward to produce genuine forecasts?

As stated earlier, these are hindcast or retrospective forecasts that are identical to genuine forecasts in everything except the fact that they are performed in the past (i.e., with the result already known). A 'hindcast' is a prediction made for a known past state whereas a 'forecast' is made now for an unknown future state. All other aspects of the prediction are essentially identical. To produce genuine forecasts we would just need initial conditions for now and then wait 4 months to see how well they did.

Changes to the manuscript: use of 'forecast' and 'hindcast' have been replaced with 'prediction' where appropriate throughout the manuscript. We have also added text to make it clearer that the GloSea seasonal predictions made here (and generally within the hindcast context) are done without any knowledge of future observations. (P5 L14; P11 L27-29)

Specific comments

In the abstract: really the 'first time, we directly assess the impact of winter sea ice thickness initialisation on the skill of seasonal summer forecasts'? Do you mean in the Met Office model framework? I think Allard and others have done this and I also don't think you do assess forecast skill

Yes, this is the first documented study to assess the impact of using satellite sea ice thickness data to initialise a fully coupled seasonal prediction system. This fact is confirmed by the comments of referee #2 who states, "This study represents the first known fully coupled atmosphere-ocean-ice forecast system to utilize CryoSat-2 ice thickness data for seasonal forecasts".

The study of Allard et al (2018) is very different because they use a forced ocean-sea ice model to perform long ocean/sea ice analyses. They also perform a thorough assessment of the thickness analyses produced and consider short-range, uncoupled, forecasts. Here we are only interested in the analysis in the context of providing sea ice initial conditions for our seasonal predictions – made using the GloSea seasonal prediction system.

We are a little confused about the referee's comment "I also don't think you do assess forecast skill". Here we run 24 retrospective seasonal forecasts per year, for each of 5 years (120 in total), to produce seasonal predictions of September Arctic sea ice using initial conditions in April/May. These predictions are performed under forecast conditions using an initialised coupled climate model in the same way as we do operationally in the GloSea Seasonal Prediction System. We then evaluate the quality of September-mean Arctic sea ice predictions by comparing basin-wide extent against observational estimates (from NSIDC, HadISST & OSI-SAF – the latter using the CMEMS reanalysis), and by calculating integrated ice edge error (IIEE) against the latter dataset. This is the standard methodology for examining skill in a seasonal forecast system, where the accumulation of sufficient evidence from "real time" forecasts would require substantial delays to the provision of information required to update forecast systems within a realistic period.

Changes to the manuscript: We have expanded the Introduction to be more explicit about our motivation for this study and to provide more details about what was done in Allard et al. (2018). In particular we now explicitly mention that previous studies have looked at analyses and short-range forecasts using externally forced ocean-sea ice models rather than seasonal forecasts using fully coupled models (as we do here). (P4 L7+) We have also added text to make it clearer that the GloSea seasonal predictions made here (and generally within the hindcast context) are done without any knowledge of future observations. (P5 L14; P11 L27-29)

Introduction - In general you need more updated references. A lot of this discussion is a bit outdated now. i.e.:
 - Drop the Vaughan/IPCC refs and use the more specific refs. Try Serezze & Stroeve 2015 for a more recent seasonal sea ice trends citation?
- The Collins/IPCC is also a bit outdated. I think you can add some of the more recent references to sea ice projections - e.g. Jahn 2016, 2018.

We believe the most appropriate citations are those given. The intention here is to motivate the fact that Arctic sea ice has declined/is in decline and that is projected to continue.

The most robust evidence of this is provided by these multi-author IPCC references, which are created by a multi-disciplinary (multi-centre, multi-country, etc.) team of authors. (This is also true for the very comprehensive, multi-author study of Meier et al. (2014).)

P2 L1-11 - I think this is not useful information as it is not that relevant to seasonal forecasting and a lot of the references and discussion are outdated. Either update/improve or drop.

We disagree and think this is extremely relevant material. Substantial resources are being invested in seasonal planning in the Arctic (e.g. within the EU's Horizon 2020 programme, the ARCUS Sea Ice for Walrus Outlook (SIWO), in support of projects endorsed by the WMO's Year Of Polar Prediction (YOPP) & MOSAiC), with more such investment by both government and private organisations likely in the future. In particular, there are a number of projects endorsed by YOPP which focus on sea ice seasonal prediction (you can see an overview of YOPP-endorsed projects at https://apps3.awi.de/YPP/endorsed/projects, noting in particular: https://apps3.awi.de/YPP/pdf/stream/79, https://apps3.awi.de/YPP/pdf/stream/100, https://apps3.awi.de/YPP/pdf/stream/106, and https://apps3.awi.de/YPP/pdf/stream/172).

We do not believe that, as a general rule, citations should have a "best before date". So if no further information has been published in the meantime to contradict the findings of these papers, which clearly have had a large impact on the funding agencies, then they are entirely appropriate.

P2 L13 - change sentence ordering.

Changes to the manuscript: We have changed this to: "Interest in seasonal predictions has increased 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." (P2 L13-14)

P2 L17 - the predictive skill sentence is confusing. SIPN haven't really assessed that.

Yes it is true that SPIN, the US project, has not done this itself. However, our point here is that the existence of SIPN has caused this as a secondary effect.

Changes to the manuscript: The text has been changed to make it clear that the community that has been built up around the SIO has enabled collaborative activities addressing such issues, rather than SIPN itself having done this. (P2 L18)

P2 L19-23 - I don't think you've really said why it is interesting though! Either make a clearer point regarding its scientific interest (e.g. what the predictability/memory of the system is compared to other components of the climate system).

Changes to the manuscript: we have included some text to make it clear that we are talking about the fact that seasonal forecasts are considerably longer than the (typically 1-2 week) limit, beyond which, the chaotic nature of the atmosphere and ocean inhibit traditional deterministic forecasting. We have included a new reference (Slingo and Palmer, 2011) which documents this well. (P2 L21-22)

Also I don't think it is clear that sea ice is now necessarily harder to predict. Having some enhanced variability may be useful. Your figure 4 doesn't show an increase in ice edge error for instance!

In fact Figure 4 does have an increasing trend! The IIEE in the control predictions (Fig 4b) is increasing over the period 1992-2015 with a small slope of approx. 0.0087 million square km per year (or 8700 square km per year). Although this is small relative to the long-term mean, it is statistically significant (p-value < 0.016). Of course, this in itself does not prove that forecasting is more challenging with less/thinner/more variable sea ice – and we are not arguing here that it does. However thinner sea ice will be more heavily influenced by the non-linear, chaotic atmospheric circulation (both dynamically and thermodynamically), which would undoubtedly be less predictable.

That prediction of sea ice becomes harder as the ice thins is one of the results of Holland et al (2011) who show that "ice area in a thicker sea ice regime generally exhibits higher potential predictability for a longer period of time". Furthermore Stroeve et al. (2014) also support this stating: "The reduced predictive skill as the winter ice cover thins has been noted in some of the contributions to the SIO and appears to be coincident with the rapid thinning of the ice cover."

Changes to the manuscript: We have added additional citations to Holland et al. (2011) and Stroeve et al. (2014) at the end of this sentence. (P2 L25)

P2 L27 - why exactly does a lack of observations make the forecasts harder? Less to assimilate in models or to validate? Below you list a number of observations that are available in the poles...

Changes to the manuscript: we have added "…meaning that the initial conditions used for forecasts in the Arctic are less accurate than for lower latitudes." To emphasise how a lack of observations would make forecasting harder. (P2 L31)

P2 L35 - include acronym definitions.

Changes to the manuscript: definition of acronym FOAM = "Forecast Ocean Assimilation Model has been brought forward from later in the manuscript.

P3 L4 onwards - this is a bit of a confusing paragraph to me. What is the point you are trying to make? In general my view is that there is hope for dynamical models being used for skillful sea ice forecasting based on some of the perfect model studies that you cited. However the SIO has really shown that they are not currently performing much better than the linear trend in many cases (as shown by Stroeve et al., 2014). It still seems that dynamical models are lagging behind more simple statistical methods (e.g. Schroeder et al., 2014, Petty et al., 2017).

I think you need to add in some comments on the different forecast methods available, merge with the following paragraph about improving dynamical models and make clearer what the motivation of this study is! This should be the key paragraph of the introduction.

Changes to the manuscript: We have replaced this paragraph to better motivate our interest in seasonal predictions using dynamical models. (P3 L7+) Additionally the penultimate

paragraph of the Introduction now contains a hypothesis statement to make it clear what the motivation of the study is. In particular we hypothesise that seasonal predictions of late-summer (September) sea ice cover made using our fully coupled dynamical model, will be improved by initialising sea ice thickness in early spring (May) using observations of thick sea ice derived from CS2. (P4 L17+)

P3 l 24 - not sure how this point links to the above.

The motivation here is that the previous points all list examples that suggest that winter sea ice thickness would be important for seasonal forecasting of Arctic summer sea ice. This final point notes that SIT initialisation is not the complete story because model uncertainty (model structural uncertainty and model choices/parameters) is also important for the evolution of forecast errors – and likely dominate over initial condition uncertainty.

Changes to the manuscript: We have changed this paragraph to make the story clearer. We start by stating that several studies have shown that winter sea ice thickness provides important preconditioning for the evolution of Arctic sea ice through the summer melt season, and then say that these studies suggest that SIT could improve seasonal forecasts of summer sea ice. We then note that, on seasonal timescales, model uncertainty is likely to dominate the evolution of errors through the forecast. (P3 L19-33)

P4 L1-4 - but in the abstract you imply you are the first to do this?! I guess you meant in your fully coupled Met Office forecast framework. You need to make that clearer.

Yes we are. As stated previously this is the first use of satellite thickness data to initialise seasonal coupled predictions of Arctic sea ice. The previous studies listed here are using forced ocean-sea ice models and performing reanalyses and/or short-range forecasts. In contrast we are performing seasonal predictions (5-month forecasts) using a fully coupled atmosphere-ocean-sea ice-land model with the GloSea seasonal prediction system.

Changes to the manuscript: We have expanded the Introduction to be more explicit about our motivation for this study and to provide more details about what was done in Allard et al. (2018). In particular we now explicitly mention that previous studies have looked at analyses and short-range forecasts using externally forced ocean-sea ice models rather than seasonal forecasts using fully coupled models (as we do here). (P4 L17+)

P5 L11-13 - why just mention the ocean reanalysis component here? Would be more understandable if you referred to GloSea as a reanalysis.

No! GloSea is a coupled seasonal prediction system (see MacLachlan et al (2014)). Referring to it as a reanalysis would be more misleading.

The motivation in this section of the manuscript is to introduce the fact that long reanalyses are performed using an offline analogue of the FOAM ocean analysis system. These long reanalyses are primarily used within the GloSea seasonal prediction system to initialise hindcast (or retrospective forecast) experiments. However they have also been utilised by several studies to help further understanding of the ocean and how the ocean has changed over the satellite period (e.g. within ORA-IP). As the ocean reanalysis is used within the GloSea seasonal prediction system, it is sometimes (erroneously/unfortunately) referred to as

the GloSea ocean re-analysis. However, GloSea is a coupled global seasonal prediction system (see MacLachlan et al., 2014).

Changes to the manuscript: We have changed several instances of "GloSea" to "GloSea seasonal prediction system" to help make it clearer that GloSea is much larger than an ocean reanalysis.

P5 L23 - I don't think these are the correct citations here. Link to relevant passive microwave concentration datasets instead or recent papers describing that long-term record (e.g. Parkinson/Comiso papers).

The motivation behind these citations is two-fold: 1) to provide references for the HadISST and NSIDC data sources that we use within our forecast evaluation; 2) to show that we have long-term sea ice concentration observations from multiple sources. The citations given here are those recommended/requested by the data providers for the NSIDC sea ice index (https://nsidc.org/data/g02135) and for the HadISST dataset (https://www.metoffice.gov.uk/hadobs/hadisst/). Therefore, their inclusion is required for our objective #1 and we re-use them for our objective #2 to avoid including lots of similar citations.

P5 L25 - should reference Kwok and Cunningham 2008 instead.

We shall change this from the 2009 reference to the Kwok and Cunningham (2008) reference.

Changes to the manuscript: Kwok and Cunningham (2008) manuscript has been added to the references and is cited here in place of Kwok et al. (2009). (P6 L18)

P6 L16-17 - what do you mean by sensitive here? I think you mean uncertain/challenging? Again, is there nothing in one of the CPOM papers that highlights this issue?

When freeboard is very low it is difficult to distinguish from SSH fluctuations and gravity waves. This issue is well documented in the comprehensive Ricker et al. citations that we have used. Given that the issue is related to the whole process of deriving freeboard from satellite altimetry – in particular CS2 SIRAL radar altimeter which penetrates the snow and into the upper layer of the ice – rather than the centre who happen to be processing the data, we are happy that these citations cover the issue adequately.

Changes to the manuscript: The word 'sensitive' has been replaced with 'uncertain'. (P7 L15)

P6 L30 - I think this is a big guess. Do we really know much ice < 2 m melts away each summer? That Keen modeling study (Fig 2?) suggests some 40% of the ice less than 2 m (including ice and snow in that thickness) does not melt through in summer. A lot will have to do with how much snow there was on the ice and where/when melt onset occurs. Even if it does melt away, this seems to be crucial information for determining solar absorption that can drive SST increases and further sea ice melt. Also need to make the point here (and earlier) that AWI do produce a merged product!

Changes to the manuscript: We do not undertake a summary, or comparison, of available sea ice thickness datasets in this study. To avoid this sort of confusion, the abovementioned

paragraph has been removed in the revised manuscript. We now only refer to CS2 thickness observations in Section 2.2.1 ("CryoSat-2 thickness observations"). The new hypothesis/motivation paragraph in the Introduction, will tell the reader that our interest is initialising thick sea ice using CS2 and testing the impact on seasonal predictions of September sea ice cover.

Section 2.3 - Why do you need extent and concentration? Surely you are just assimilating sea ice concentration? This needs to be made much clearer here.

There appears to be some confusion that this section is referring only to data used for assimilation. Section 2 is about all the models and datasets used in this study – both for assimilation and for evaluation. Section 2.3 covers the validation datasets in paragraph 1 and the assimilation data in paragraph 2.

Changes to the manuscript: Section 2.3 has been modified to include an introductory sentence that explains that data is used for assimilation and evaluation purposes. (P7 L27-28) New subsection headings (2.3.1 and 2.3.2) are also added to split the introduction of observations according to whether they are used for evaluation (2.3.1), or assimilation (2.3.2). (P7 L29; P8 L12)

- The use of CMEMS 'data' seems very confusing to me. Why not use observed ice concentration?!

It is very common to use an analysis product for evaluation purposes. For example, many people use the OSTIA analysis for evaluating SST or ERA-Interim for evaluating atmospheric variables. Likewise, people are now starting to use the blended CS2SMOS product (Ricker et al. 2017) to evaluate sea ice thickness (which itself is an SIT analysis created using O/I assimilation methods). Although we would never assimilate analysis products like these, we do use them for the purpose of model evaluation.

In this study we validate using the ice concentration from the CMEMS reanalysis, which is an analysis made using the OSI-SAF sea ice concentration data. The main reason for using this is that it is already on the correct model grid. We state that: "Using this CMEMS reanalysis 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."

It is important to note that, except for the differences due to the spatial resolution, the sea ice concentrations in the OSI-SAF observed data and the CMEMS analysis are virtually identical. The CMEMS product can therefore be thought of as a dynamically consistent re-gridding. It is of course very important to compare extent using products on the same grid because extent, as a metric, is very much dependent on grid/resolution (Notz, 2014).

- The NSIDC sea ice index is just a monthly index of total ice extent. This isn't what you use, right?

Yes we use the NSIDC single number extent for the purposes of validating our September-mean ice extent predictions (in Fig 4). However we also use extent derived from HadISST sea ice concentration, and from the CMEMS reanalysis (i,e., OSI-SAF).

P8 L6 - what is the size of your model grid (in kilometers?) how does this translate to the 5 km CPOM data?

The ORCA025 tripolar grid was created to avoid the singularity associated with the convergence of meridians at the North Pole, which it achieves by defining two distinct north poles over Canada & Siberia. The ocean points in between form a variable resolution grid with highest resolution nearest the two poles i.e., in the Canadian Arctic Archipelago and the Laptev Sea. The resolution in the Arctic Ocean ranges from ~9km up to ~15km.

Changes to the manuscript: Section 2.1 has been expanded to include more details of the sea ice component of the model (as requested by Referee #2). We have included details of the grid resolution within this text as follows: "ranging from 8.9 km to 15.5 km in the Arctic Ocean basin" (P6 L5-6)

P8 L14-24 - Think you should list out the CICE thickness categories. How else could you have done this? I think it would be worth presenting more sophisticate approaches for future work, however I get that you started with this simple approach.

We use standard WMO categories that are one of default options within the CICE model. These are listed in the supporting Blockley et al., (2014), and Ridley et al. (2018) references.

Changes to the manuscript: Section 2.1 has been expanded to include more details of the sea ice component of the model (as requested by Referee #2). We have included details of the model thickness categories within this text as follows: "(lower bounds: 0, 0.6, 1.4, 2.4 and 3.6 m)". (P6 L10) More details are provided as future plans for sea ice thickness assimilation in the Summary section (P17 L11+)

Figure 2 - I think you should also show (maybe in the supplementary info?) what the pre nudged, and nudged thickness fields are. Could just do this for the mean October-April thickness and also update Figure 1 to show this longer season too. If the mean thickness was way off before it makes sense that assimilating the thickness will improve things.

Changes to the manuscript: Figure 1 has been expanded to show the impact of assimilating CS2 thickness data on the sea ice initial conditions. In addition to the original plot showing the CS2 thickness data, 3 more panels have been added to show the modelled thickness from the control (CTRL) and thickness initialised (ThkDA) reanalyses, along with the difference between these two experiments (ThkDA-CTRL). In line with the referee's suggestion, the time-period has been increased from DJF to October-April mean. A new paragraph has been added at the start of Section 3.1 to describe the general impact of CS2 initialisation on the winter thickness. (P9 L26+)

Why the different start dates for the forecasts? Pretty confusing.

This is how the GloSea seasonal prediction system works. GloSea runs every day at the Met Office and produces 2 forecasts of length 210-days. These are used with forecasts from previous days to create a large, lagged ensemble of forecasts. For the hindcasts/reforecasts performed here, an 8-member ensemble is run for fewer distinct start dates. This is done to

Figure 4 is hard to see. Maybe box plots of the recent years showing the variability in the different estimates in the different years?

Referee #2 also asked for the layout of Figure 4 to be modified and so it has been overhauled.

Changes to the manuscript: Figure 4 has been changed to make it easier to read. More space on the plot is now devoted to the key time-period of 2011-2015, and spacing has been used to ensure that the individual ensemble member values are now distinguishable from each other. The figure caption has been modified to make it clear that these changes to the spacing are done purely for ease of viewing. (P32)

P11 L10 - CMEMS isn't really an observational estimate, right? Based on the assimilation of OSI-SAF...

Changes to the manuscript: this text has been modified to make it clear that we compare with NSIDC and HadISST observational datasets as well as the CMEMS reanalysis dataset. (P12 L22-24)

PII L14 - not sure what you mean by 'building a picture' here. I see no value in showing that earlier data.

Changes to the manuscript: Text has been added to better paint a picture of the errors in the control simulations. Discussion of the extent predictions has been expanded to details how the control experiment predictions of extent are consistently biased low. (P12 L28-30) An additional sentence has been added to the IIEE discussions to describe that the ice edge error is consistent through the 24 year period covered by the control simulations. (P13 L15-17)

P11 L17 what is close

Changes to the manuscript: "close" has been quantified as "within 0.12 x10^6 km^2". (P13 L1)

Figure 5 - include numbers on map. Plot the IEEE as a time series.

Changes to the manuscript: Figure 5 has been modified so that each panel includes a box containing the observed and modelled extent along with the IIEE. The figure caption has been modified accordingly. (P33-34)

Figure 7 and 8 - I don't understand these maps. How exactly is the data shown in Figure 8 calculated? Also is Figure 8 averaged over the entire year, but figure 7 is September? Why are the pressure units different? Perhaps better to show Figure 7a and 8a together, then 8a and 8b
The discussion of using ensemble members instead of the ensemble means and how this relates to assessing model bias was confusing and needs more description/clarification.

Changes to the manuscript: Figures 7 and 8 have been merged together (new Figure 8 - P36) to include both differences in the mean fields (left) and the forecast errors (right). We now also show z500 and MSLP for both. Section 4.2 has been modified to provide better motivation for Figure 8. In particular, and as requested, Section 4.2 paragraph 2 provides a thorough overview of what is plotted in Figure 8. (P14 L22+)

(NB. in the original manuscript "September" was accidentally left off the Figure 8 caption, which prompted the question about whether this plot was the entire year. This has been fixed as part of the changes described here.)

Should either enhance the CLIM analysis or drop. i.e. repeat with using a fixed (think FIXED_IC would be a better acronym) for all years of data available.

Changes to the manuscript: "CLIM-2015" is changed to "FIXED-IC" throughout the manuscript.

How do these results compare with Allard?

Our study is very different from Allard et al. As stated previously, we are initialising a fully-coupled atmosphere-ocean-sea ice-land (AOIL) model with CryoSat-2 (CS2) sea ice thickness and using it to perform an ensemble of seasonal predictions from May through to September. We find that initialising with CS2 SIT gives us a considerable improvement in our seasonal predictions of Arctic sea ice extent and ice edge location. We also show that memory of winter thickness changes in the initialisation carry through to the end of summer.

Meanwhile Allard et al. (2018) performed an 18-month reanalysis using an externally forced ocean-sea ice model forced by atmospheric analyses. They assimilated CS2 thickness (from multiple sources) during the winter when the data is available. In contrast to our study, and by virtue of the fact they run a reanalysis, they continue assimilating all other variables throughout the year. They show a considerable improvement in their winter thickness analyses when using CS2.

What the two studies have in common is that:

1) we both show that assimilating sea ice thickness provides an improvement (us: in coupled seasonal forecasts; them: in their externally forced reanalysis);
2) we both show that memory of initialised winter thickness is still present in the summer (us: after a 5-month free-running coupled model forecast; them: after 5 months of assimilating everything except SIT in an externally forced reanalysis).

Both studies show that sea ice thickness initialisation is beneficial – albeit in very different setups.

Changes to the manuscript: We have expanded the Introduction to be more explicit about our motivation for this study and to provide more details about what was done in Allard et al. (2018). In particular we now explicitly mention that previous studies have looked at analyses and short-range forecasts using externally forced ocean-sea ice models rather than seasonal forecasts using fully coupled models (as we do here). (P4 L17+)

P14 L21 - not sure what you mean by work well. Maybe not gone wrong?

Technically, the application of thickness increments within the CICE sea ice model while performing a sea ice analysis has worked exactly as expected. The increments have been retained by the model and produce thickness initial conditions, for initialising the seasonal forecasts, that are much closer to the CS2 data. There is also the matter of the exceptional improvement to sea ice location (IIEE) when initialising with SIT.

Changes to the manuscript: An extra sentence has been added here to quantify this further. We now state that the winter thickness initial conditions, generated using the sea ice thickness nudging, are much closer to the CS2 thickness observations as illustrated by the new panels in Figure 1. We also state that these thickness changes lead to considerable improvement in skill when used to initialise GloSea seasonal predictions. (P17 L1-5) The existing text, which stated that (importantly) the model is able to retain the thickness information, has also been retained.

P14 L1 Why is this from dynamics not thermodynamics?

The distribution of thickness across the Arctic is caused by a series of dynamic processes. The ice motion, primarily driven by the winds, consists of a re-circulation around the Beaufort Gyre with a transpolar drift which drives across from Siberia to the north coast of Greenland and north of Svalbard. This causes thick ice to pile up north of Greenland. The fact that we have thicker ice dragged around into the Beaufort Sea and not enough thin ice north of Fram Strait suggests that the ice is too mobile. Both the proposed EAP rheology and form-drag changes would reduce the ice speed and should reduce the bias.

There is of course a small chance that this is thermodynamically driven, which is why we say this is "most likely" caused by deficiencies in the dynamics.

Also drop 'so-called ice-ice force'

Changes to the manuscript: this text has been removed.

Response to review #2

Specific Comments

Page 3 Line 1: Add Cummings and Smedstad (2014) for another coupled ocean-ice modeling system with data assimilation, here is a full reference: Cummings, J. A. and O. M. Smedstad, 2014: Ocean Data Impacts in Global HYCOM, Journal of Atmospheric and Oceanic Technology, 31, doi:10.1175/JTECH-D-14-00011.1

The articles cited here (i.e., Tonani et al., Martin et al., Balmaseda et al.) are all large multi-centre papers which describe all of the global ocean/ice analysis + forecasting systems that contribute to the GODAE OceanView project. This includes HYCOM-based systems such as that documented in Cummings and Smedstad (2014). Jim Cummings is part of the author list for one of these, and several Cummings papers are cited. However, this exact paper is not one of those referenced in the papers we cite here.

Changes to the manuscript: Cummings and Smedstad (2014) citation added.

Page 7 line 31: why couldn't a longer period, say from 2010-2017 or at least 2010-2016 be used?

This work was started in 2016 and so the analysis was performed up to September 2015, which was the most recent summer period at that time. This is true for all the control runs that we used in this project as well as the CS2 runs that we performed ourselves. Whilst undertaking this work, we were only able to perform such a large ensemble of seasonal predictions because the delay finalising our model configuration for CMIP6 meant that there was a good chunk of computational resource available. This is no longer the case and CMIP6 runs are in full swing now – meaning that we do not have the computational resources available to extend this to 2016 or 2017 at this time.

Page 8 line 14: which version of CICE (v4.0, 4.1) is used. Are melt ponds used in this study?

CICE vn5.1.2 is used here and the configuration includes topographic melt ponds. The coupled model version used here (GC3) is documented in Williams et al. (2017) and the sea ice component therein (GSI8) is documented in more detail in Ridley et al. (2018). Although we do not intend to include lots of technical information about the sea ice model used, we do agree with the referee that more information is required here.

Changes to the manuscript: Section 2.1 has been expanded to include more details of the sea ice component of the model. This text includes details about CICE model version and key features of the sea ice model component such as multi-layer thermodynamics, 5 thickness categories, and prognostic melt-ponds. We also included details of the grid resolution and category bounds as requested by Referee #1. (P6 L5+)

Page 14 line 17-18: can you quantify bias reduction with some range or percentages?

We have done this in the revised manuscript. However, we feel that this sits better at the end of page 9 where the bias/difference is first discussed rather than on page 14.

Changes to the manuscript: In the revised manuscript, the change in long-term mean thickness has been quantified by specifying percentage changes and explicit values. This is done for the whole Arctic as well as separately for the Atlantic and Pacific sectors (P11 L2-6) Additionally, Figure 1 has been expanded to show the impact of assimilating CS2 thickness data on the sea ice initial conditions. In addition to the original plot showing the CS2 thickness data, 3 more panels have been added to show the modelled thickness from the control (CTRL) and thickness initialised (ThkDA) reanalyses, along with the difference between these two experiments (ThkDA-CTRL). In line with the referee's suggestion, the time-period has been increased from DJF to October-April mean. A new paragraph has been added at the start of Section 3.1 to describe the general impact of CS2 initialisation on the winter thickness. (P9 L26+)

Figure 4: Please add an inset for both plots showing a blow-up for the period 2010-2015? It's a bit difficult to see with the longer data record shown.

Referee #1 also asked for the layout of Figure 4 to be modified and so it has been overhauled.

Changes to the manuscript: Figure 4 has been changed to make it easier to read. More space on the plot is now devoted to the key time-period of 2011-2015, and spacing has been used to ensure that the individual ensemble member values are now distinguishable from each other. The figure caption has been modified to make it clear that these changes to the spacing are done purely for ease of viewing. (P32)

There is no mention of ice drift in the paper. Could you analyze IABP ice drift data (Pan-Arctic domain) to determine the impact of assimilating CS2 data into the seasonal forecasts? This would complement your existing study. SIDFex is presently examining modeling center's skill in making long-term ice drift trajectory forecasts.

We have compared the May and September sea ice velocity fields from our two main experiments (CTRL and thkDA). We find that the May velocities are virtually indistinguishable in each of the 5 years (2011-2015). This is consistent with the findings of Allard et al. (2018) who show little impact on ice drift in their reanalysis comparisons. The September velocity fields are also very similar. Although slight differences arise from the differing ice coverage, the ice drift is broadly/qualitatively unchanged in the experiment using sea ice thickness initialisation.

Changes to the manuscript: We have included a small paragraph to state that the drift is unchanged (not shown) and that this is consistent with the findings of Allard et al. (2018). (P11 L9-11)

I would like to see the ice edge error metric used to examine the regional differences seen from use of the CS2 data. Can it be divided into the following (or similar) basins (Beaufort/Chukchi/Bering Sea, Canadian Archipelago, Greenland Sea, Laptev Sea, Barents Sea, East Siberian Sea)?

Changes to the manuscript: A new figure (Figure 6) has been added which includes details of the observational extent (from the CMEMS reanalysis), the model predictions of extent from both the CTRL and ThkDA experiments, and the ide edge error (IIEE). This is done for 3 domains which bisect the Arctic Ocean: the central Arctic, the Siberian shelf (combined Laptev, Kara + East Siberian Seas), and the combined Beaufort + Chukchi Seas. Also reported are the same statistics for the Atlantic and Pacific sectors of the Arctic. A few sentences of text has been added to Section 4.1 to discuss these results - in particular that improvement is most notable in the Atlantic sector. (P13 L31+)

No comparisons are made against ice thickness observations from either ice mass balance buoys and/or moored ULS data. I recommend inclusion of some time series plots of the modeled ice thickness beginning with the Apr/May initializations through September for 2010-2015, with the control run included. The ensemble spread can be shown as well. This should clearly show the impact of the inclusion of the CS2 data.

While comparison with ULS would be interesting, we feel it is out of scope for this study where we are heavily focussed on improvement to sea ice cover (extent and edge location). However, this sort of comparison is something we would wish to do before implementing a proper 3D-Var sea ice thickness assimilation scheme – and in fact is currently being undertaken within the H2020 SEDNA project.

Lots of acronyms are used without spelling them out. A partial list is shown below. Perhaps a list or table of acronyms would be useful.

We apologise for this oversight. The revised manuscript will ensure that acronyms are spelled out at the point of first use. We do not think that this manuscript is "acronym heavy" enough to need a glossary appendix/table though. However, we can include one if the referee (or the journal typesetters) feels strongly about this.

Changes to the manuscript: many changes have been made to define acronyms at 1st point of use.

Technical Corrections

Page 2 line 6: replace "Better knowledge" with "Improved knowledge"

Page 2 line 34: define SLA here

Page 3 line 22: define CFSv2

Page 3 line 25: replace "find" with "found"

Page 3 Line 34: Yang et al. reference not listed in References

Page 4 line 3: spell out NRL (Naval Research Laboratory)

Page 5 line 3: spell out FGAT

Page 5 lines 12-14: "used" appears in sentence 3 times. Perhaps change second mention of this word to "utilized"

Page 5 line 26: reword statement "have been around for some years"

Page 5 line 27: replace "main" with "primary"

Page 5 line 32: spell out SIRAL

Page 6 line 22: replace "was" to "is"

Page 7 line 16: spell out SSMI/S

Page 7 line 28: Ridley reference says (2017, in review); reference section states 2018

Page 9 line 24: reword to "although IT has".

Page 10 line 4: delete "the fact"

Page 10 line 17: Williams reference in references section says 2018

Page 10 line 25: MacLachlan reference says 2014 in references section

Page 12 line 2: delete "down"

Page 13 line 26: reword "doing things this way"

Page 14 line 28: spell out SEDNA

Page 15 line 4: Neither Tsamodos reference is listed in References section

Page 15 line 17: replace "down" with "due"

Page 21: Peterson reference should be 2015 not 2014

Many thanks for providing such a thorough list of technical changes.

Changes to the manuscript: all of these minor corrections have been made in the revised manuscript.

Page 10 line 7: reword "amongst other things" what things?

Essentially the "other things" mentioned here are all of the model prognostics and diagnostics which are far too many to list.

Changes to the manuscript: We have removed mention of "amongst other things" to prevent it causing confusion.

Improving Met Office seasonal predictions 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

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 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 summer seasonal predictions by assimilating CryoSat-2 thickness data into the Met Office's coupled seasonal prediction 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 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 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 Arctic sea ice are therefore becoming increasingly important for the safety of human activities in the Arctic (Eicken, 2013). Improved 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 North Atlantic Oscillation (NAO)-like patterns and 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.

Interest in seasonal predictions has increased 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. In response to this, in 2008, the Sea Ice Outlook (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 community that has been built up around the SIO has enabled collaborative activities addressing such issues 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 interesting scientific problem here to test our ability to predict sea ice on seasonal timescales that are considerably longer than the (typically 1-2 week) limit, beyond which, the chaotic nature of the atmosphere and ocean inhibit traditional deterministic forecasting (Slingo and Palmer, 2011). 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 (Holland et al., 2011; Stroeve et al., 2014).

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), meaning that the initial conditions used for forecasts in the Arctic are less accurate than for lower latitudes. Despite this, several operational forecasting centres regularly contribute to the SIO with sea ice predictions from fully 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

Deleted: Better

Deleted: F

Deleted: , interest in seasonal predictions has increased

Deleted: foster

Deleted: s

Deleted: on

Deleted: of such

Deleted: fact

of GloSea are initialised each day using the Forecast Ocean Assimilation Model (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 Sea Surface Temperature (SST), satellite Sea Level Anomaly (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; Cummings and Smedstad, 2014), does not assimilate sea ice thickness.

The use of dynamical models for seasonal sea ice prediction is in its relative infancy. Still, there have been several studies that demonstrate skill in retrospective forecasts (or hindcasts) of September-mean Arctic sea ice extent made from spring (e.g. Sigmond et al., 2013; Wang et al., 2013; Chevallier et al., 2013; Msadek et al., 2014; Peterson et al., 2015). However, none of these were able to match the potential skill found in idealised "perfect model" studies (Guemas et al., 2016; Tietsche et al., 2014; Day et al., 2014; Blanchard-Wrigglesworth et al., 2011), where all the initial conditions, but in particular, the sea ice thickness, are known precisely. Furthermore, when applied to a real-time forecast, as submitted to SIO, the skill was found to be even lower than the hindcast skill (Blanchard-Wrigglesworth et al., 2015), and only marginally better than a linear trend forecast (Stroeve et al., 2014). Clearly, there is potential for improvement in the dynamical models if more complete initial conditions are known – with an even greater need, as demonstrated by the deteriorated performance of the real-time forecasts, for more accurate real-time initial conditions. None of the systems mentioned above initialise the sea ice using observed thickness measurements.

Several studies have shown that winter sea ice thickness provides important preconditioning for the evolution of Arctic sea ice through the summer melt season. Blanchard-Wrigglesworth and Bitz (2014) found sea ice thickness anomalies in general circulation models (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 of sea ice thickness 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 coupled forecast system model version 2 (CFSv2) seasonal prediction system. They showed 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). These studies suggest that seasonal (> 90 days) predictions of Arctic summer sea ice, made with dynamical models, could be improved by correctly initialising sea ice thickness. However, we note that, although uncertainty in the initial conditions plays a crucial role for seasonal predictions of Arctic sea ice, model uncertainty is likely to dominate the evolution of seasonal forecast errors (Blanchard-Wrigglesworth et al., 2017).

3

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, externally 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 sea ice thickness, using 2 different datasets processed with different algorithms, within a series of reanalyses performed with the Naval Research Laboratory's forced 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 perform an in-depth assessment of the thickness data and analyses, and show a good agreement between the CS2-derived ice thickness and observations from in-situ and airborne sources.

As noted above, several studies (Yang at al., 2014; Xie et al., 2016; Mu et al., 2018; Allard et al., 2018) have looked at the impact of sea ice thickness initialisation on analyses and short-range forecasts produced with externally forced ocean-sea ice models. What has not been investigated is the impact that assimilation of sea ice thickness may have in longer (> 90 days), forecasts made using fully coupled dynamical models. Here we do so for the first time using the Met Office GloSea coupled seasonal prediction system. For accurate seasonal predictions of September sea ice cover, it is important to model ice that will persist throughout the summer season, and so an improved representation of the location of thick sea ice within the initialisation should be advantageous. We hypothesise that GloSea seasonal predictions of late-summer (September) sea ice cover will be improved by initialising sea ice thickness in early spring (May), using observations of thick sea ice derived from CS2. In this study, we use a simple nudging technique to test this hypothesis, and evaluate the feasibility of including sea ice thickness initialisation within the operational GloSea seasonal prediction system. We assimilate CS2 sea ice thickness within the FOAM ocean-sea ice reanalysis and use these analyses as initial conditions for an ensemble of seasonal (5-month) coupled forecasts to determine the impact of sea ice thickness initialisation on the skill of GloSea seasonal predictions. We 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.

This paper is structured as follows: Section 2 introduces the modelling systems and observations used in this study; Section 3 describes the initialisation of CS2 thickness within the ocean-sea ice reanalysis and the generation of initial conditions for seasonal predictions; Section 4 provides details of GloSea coupled seasonal prediction experiments performed using CS2

4

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

2.1 Modelling systems

5 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 Hadley Centre Global Environment Model version 3 (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) (Hunke et al., 2015).

10

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 – retrospective forecasts performed for previous years using true forecast conditions – are used to establish errors in the model

15 climatology for the purposes of bias correction, and to estimate forecast skill. More details on the GloSea seasonal prediction system 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 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 First Guess at

20 Appropriate Time (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 Blockley et al. (2014) and more about the NEMOVAR 3D-Var FGAT scheme used therein can be found in Waters et al. (2015).

As well as the abovementioned operational analyses and forecasts, longer reanalyses are performed with the FOAM system

25 using surface forcing derived from the ERA-Interim atmospheric reanalysis (Dee et al., 2011). Within the GloSea seasonal prediction system, 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 utilised more widely within the ocean community (Balmaseda et al., 2015; Chevallier et al., 2017; Uotilla et al., 2018) and have been used to help answer a number of other scientific questions (e.g. by Roberts et al., 2013; Jackson et al.,

30 2015).

5

Deleted: Forecast Ocean Assimilation Model (

Deleted:)

Deleted: used

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) and Ridley et al., (2018) respectively and uses the extended ORCA025 tripolar grid described therein – with nominal 1/4° horizontal resolution, ranging from 8.9 km to 15.5 km in the Arctic Ocean basin, and 75 vertical levels. The sea ice component of the model is based upon CICE vn5.1.2 and uses the standard CICE elastic–viscous–plastic (EVP) rheology for modelling the sea ice dynamics (Hunke et al., 2015). Growth and melt of the sea ice is calculated using a multi-layer thermodynamics scheme with 4 layers of ice and 1 layer of snow. At each model grid point, the sub-grid scale ice thickness distribution is modelled by partitioning the ice into five thickness categories (lower bounds: 0, 0.6, 1.4, 2.4 and 3.6 m), with an additional ice-free category for open water areas. The impact of surface meltwater on the sea ice albedo is explicitly represented by the prognostic evolution of melt ponds using the topographic formulation. Further details about the sea ice component, and the wider coupled model used here, can be found in Ridley et al., (2018) and Williams et al., (2017) respectively.

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), laser altimetry (Kwok and Cunningham, 2008), and, more recently for thin ice, microwave brightness temperatures (Kaleschke et al., 2016). Although radar altimeter estimates of sea ice thickness have been available for many years now, their up-take into operational ocean-sea ice assimilation systems has been minimal. The primary reasons for this are three-fold: the data were not made available in near-real-time for use in operational analysis systems; 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 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 has an unusually high inclination orbit that provides observational coverage up to 88°N, which has considerably reduced the size of the pole-hole (Laxon et al., 2013). CS2 is also fitted with a Synthetic Aperture Interferometric Radar Altimeter (SIRAL) instrument that has a higher accuracy, and along-track resolution, than was previously available from the ENVISAT and ERS-1/2 radar altimeters (Guerreiro et al., 2017). 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 initialise the model using thick ice from CS2, which are accurate for ice thicker than 1m (Ricker et al., 2017). We use monthly CS2 winter (Oct-Apr) thickness estimates produced by CPOM (Tilling et al., 2016) which start 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). Several different centres, including CPOM, are now producing CS2-derived estimates of sea ice thickness. More details on the differences between these observational estimates can be found in Stroeve et al. (2018) and Allard et al. (2018). Some more general discussion of the uncertainties involved in the calculation of sea ice freeboard and thickness using radar 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. The CS2 thickness retrieval methodology is particularly uncertain 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 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 is less than 2 m. Finally, we remove any spuriously high observations by imposing a maximum thickness threshold of 7 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 in this study can be seen in Figure 1a, which shows a map of average October-April Arctic thickness for 2011-2015 inferred from CS2 estimates after application of the QC process described above.

2.3 Observations of sea ice extent and concentration

Within this study, observations of sea ice concentration and extent from several sources are used both for evaluating seasonal predictions of Arctic sea ice, and for assimilation within the reanalyses used for initialisation of these seasonal predictions.

2.3.1 Sea ice concentration and extent datasets used for evaluation

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

---

Deleted: In this study, we

Deleted: 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 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.

Deleted: sensitive

Deleted: was

Deleted: December, January and February

Deleted: ¶
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 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.¶

we include observational estimates from three different sources: extents calculated from the 1° gridded Hadley Centre sea Ice and Sea Surface Temperature (HadISST1.2) dataset of Rayner et al. (2003); the National Snow and Ice Data Center (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 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., 2018). To avoid confusion with the FOAM reanalyses performed as part of this study, and described later, we refer to this product as "CMEMS" hereafter.

2.3.2 Sea ice concentration datasets used for assimilation

The CMEMS reanalysis, and the reanalyses performed in this study, were performed using Special Sensor Microwave Imager/Sounder (SSMI/S) sea ice concentration data provided by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) 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., 2015). 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 2014, 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 4.2 of 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

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. (2018) and the GSI8 sea ice configuration described in Ridley et al. (2018). 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 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.

8

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). The sea ice concentration observations assimilated are the same as used for the CMEMS reanalysis described in Section 2.3.2 above (i.e., OSI-401a). An overview of 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 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 described in Peterson et al. (2015) 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 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).

3.1 Impact of CryoSat-2 initialisation on reanalysis thickness

Figure 1 illustrates the general impact of including CS2 assimilation within the ThkDA-RA reanalysis by showing the mean Arctic sea ice thickness, for the months when CS2 data is available (October-April), over the whole ThkDA-RA reanalysis (2010-2015). The difference plot in Figure 1d shows that the inclusion of CS2 nudging generally acts to increase the thickness of the Arctic sea ice – in particular in the Atlantic Sector north of Fram Strait, and to the north of Greenland. Comparison with the observations in Figure 1a shows that the thickness in ThkDA-RA is much more closely aligned with the CS2 data than is the case for CTRL-RA.

9

A comparison of sea ice volume for the CTRL-RA and ThkDA-RA reanalyses in Figure 2 confirms that the net effect of CS2 thickness nudging is an increase in sea ice thickness. We note 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). Figure 2 shows

5 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 Figure 2, sea ice volume for the CTRL-

10 RA and ThkDA-RA reanalyses are also compared with volume estimates from the PIOMAS model of Zhang and Rothrock (2003). The PIOMAS volume is included here purely as a reference because it is well understood and widely used for this purpose. The volume in the CTRL-RA run is much closer to PIOMAS than the ThkDA-RA run. This is expected as PIOMAS has been shown to underestimate thickness/volume in the winter compared to the CPOM CS2-derived thickness (Tilling et al., 2015; Laxon et al., 2013) – although it has been shown to compare better with laser altimeter estimates such as ICESat

15 (Schweiger et al., 2011).

Figure 3a shows the impact of the CS2 thickness initialisation on the reanalysis end-of-winter thickness fields – that will be used in this study to initialise GloSea seasonal predictions – with the 5-year mean differences for 1st May at the end of winter when CS2 observations cease. At the end of winter it is apparent that inclusion of CS2 thickness nudging has increased sea ice

20 thickness across much of the Atlantic sector of the Arctic (Barents, Kara and Greenland Seas). 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.

25

Figure 3b shows the 5-year-mean difference in the reanalyses thickness fields at the end of summer (30th September) after 5 months of running without thickness assimilation. 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 3a. Even after 5 months of running without the CS2 thickness nudging –

30 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.

10

The general picture shown by the 5-year mean in Figure 3a is typical of the end of winter thickness differences seen for each of the 5 years 2011-2015 (not shown). Mean sea ice thickness across the Arctic Ocean basin has been increased by around 14% (from 2.00m to 2.27m). This increase is most pronounced in the Atlantic Sector of the Arctic (30°W-140°E) where thickness increased by around 33% (from 1.44m to 1.91m). Although mean thickness in the combined Beaufort and Chukchi

5 Seas has decreased by 7% (from 2.32m to 2.15m), the net effect over the whole Pacific Sector of the Arctic (140°E-20°W) is an increase of 6.6% (from 2.33m to 2.49m). However, the situation is not so clear-cut for the summer case (Figure 3b) where thickness increases are much more pronounced in 2014 and 2013 (see Figure 2).

The impact of the thickness changes on the large-scale sea ice motion is negligible with monthly-mean velocities in the two

10 experiments being virtually identical throughout the 2011-2015 period (not shown). This is consistent with the findings of Allard et al. (2018) who show little impact on ice drift in their reanalysis comparisons.

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

15 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 that the assimilation of thickness worked well and the increments were successfully retained by the model, which bodes well

20 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 are made operationally by the GloSea system each day. Hindcast predictions, 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 hindcast predictions are initialised using the long GloSea ocean-sea ice reanalysis (as described

25 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. As these hindcasts are used to test the expected skill of a real forecast, they are performed in a fashion that does not use any subsequent observational data after initialisation, so as not to invalidate that expectation. A recent trial of the new GC3 coupled model configuration of Williams

30 et al. (2017) has been performed using the GloSea seasonal prediction system, which we shall use as our control (denoted CTRL-HC). The ocean and sea ice for these hindcasts are initialised using the control reanalysis (CTRL-RA) described in Section 3 and the atmosphere is initialised from the ERA-I reanalysis. As the GC3 developments include the implementation

11

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 (Joint UK Land Environment Simulator; JULES) with ERA-I snow precipitation and data assimilation.

5 Here we wish to test the impact of initialising with CS2 sea ice thickness on the seasonal predictions of September sea ice extent. For this purpose, an ensemble of seasonal prediction experiments was configured that was identical to the CTRL-HC experiment except that the ocean and sea ice components were initialised from the ThkDA-RA reanalysis instead of CTRL-RA. Seasonal predictions 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 seasonal predictions was initialised from the same analysis fields with spread between the

10 members achieved by using stochastic physics (see MacLachlan et al., (2014) 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 predictions were performed for 2011-2015 – each year that spring analyses are available from the ThkDA-RA ocean reanalysis. We denote this system of predictions as ThkDA-HC. Details of the GloSea coupled prediction experiments used in this study can be found in Table 2.

15 **4.1 Improvements to seasonal prediction of Arctic extent and ice edge location**

Results from the ThkDA-HC experiment 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). Predictions from each of the 24 ensemble members, initialised from the 3 April/May start-dates, are depicted by the crosses; the ensemble

20 mean is plotted with bold symbols and inter-connecting lines. Although the ThkDA-HC predictions 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 and reanalysis estimates of Arctic extent, from the CMEMS reanalysis, and the HadISST and NSIDC datasets (see Section 2.3), are plotted alongside the model predictions (black/grey). We note here that the difference in extent prior to 2010 between the CMEMS reanalysis and the

25 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 of the CRTL-HC predictions made without sea ice thickness initialisation. Figure 4 illustrates that, throughout the CTRL-HC experiment, the seasonal predictions of sea ice extent are consistently biased low. The mean extent over the full

30 time series (1992-2015) of 4.20 x10^6 km^2 is between 1.53-2.21 x10^6 km^2 below that for the 3 observational datasets.

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

12

within 0.12 x10^6 km^2 of the observed values. The underestimation of basin-wide extent seen throughout the CTRL-HC predictions has been reduced; 2011-2015 5-year-mean extent of 3.78 x10^6 km^2 for ThkDA-HC is much closer to the observational average of 4.62 x10^6 km^2 than is the CTRL-HC value of 2.79 x10^6 km^2 (Figure 4a).

5   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 seasonal 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

10 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 seasonal 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 seasonal

15 predictions (as in Figure 4a, but for IIEE not extent). The IIEE is virtually flat across the length of the full time series (1992-2015) illustrating that, as for extent, the model without sea ice thickness assimilation is consistently biased throughout this 24-year period.

Figure 4b shows that ice-edge error is considerably improved by the CS2 thickness initialisation with the 2011-2015 mean

20 IIEE reduced from 3.20 x10^6 km^2 for CTRL-HC to 2.02 x10^6 km^2 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 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

25 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%. 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

30 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. These improvements are further illustrated in Figure 6, which shows, for several different Arctic Ocean regions, the ice extent predicted by the CTRL and ThkDA experiments, along with the extent from the CMEMS reanalysis and the corresponding IIEE. The predictions made using CS2 initialisation (ThkDA) have lower extent in the Beaufort and Chukchi Seas and higher extent everywhere else. In

13

all regions, the ThkDA extent predictions are closer to the CMEMS reanalysis and the corresponding IIEE is lower. Improvements are most notable in the central Arctic region – and particularly the Atlantic sector.

The spatial changes in the September-mean sea ice concentration predictions depicted in Figure 5 match well with the May
5   mean thickness dipole shown in Figure 3a. 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 7 further illustrates this point by showing how thickness differences between the CTRL and ThkDA experiments – for both the analysed spring initial conditions and the September-mean seasonal predictions – relate to the eventual predictions of ice
10   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

We now consider how the abovementioned sea ice improvements affect the wider GloSea seasonal September predictions.
15   With the changes in winter ice thickness, and in the evolution of Arctic ice coverage through the melt season described above, one would expect to see both fast changes to the local Arctic surface boundary layer (Semmler et al, 2016), as well as longer timescale changes to the wider atmospheric circulation. While much of the recent work on large-scale circulation has focused on changes to winter circulation (Koenigk et al., 2016; Vilma, 2014), studies have shown increased Northern European summer (Screen, 2013; Wu et al, 2013) and East Asian summer monsoon precipitation (Guo et al., 2014) in association with reduced
20   sea ice.

Figure 8 shows the difference between the ThkDA and CTRL predictions of September-mean near-surface air temperature (T2M), mean sea-level pressure (MSLP), and 500 hPa geopotential height (z500). The left-hand panels show the mean difference, over all ensemble members and all years (2011-2015), between the ThkDA predictions and the CTRL predictions.
25   Meanwhile the right-hand panels show the mean difference in root-mean-square error (RMSE) between the ThkDA predictions and the CTRL predictions. Here RMSE is calculated for each ensemble member relative to the ERA-I atmospheric reanalysis, which are then averaged over all ensemble members and all years (2011-2015) before differencing. Defining the error with respect to individual ensemble members in this manner, as opposed to looking at the ensemble mean error, provides a sufficiently large distribution of values to allow us to test statistical significance – which we do using a Mann-Whitney test
30   with the null hypothesis that all errors (or differences) are drawn from the same distribution. The resulting error difference fields calculated using this method are qualitatively the same as considering the difference between the RMSE of each ensemble mean relative to ERA-I (not shown), however the errors here will be larger as there will be no cancellation of errors caused by averaging across ensemble members

14

We first focus on the local temperature changes, for which Figure 8 shows that, owing to the overall increase in Arctic sea ice thickness and extent, the ThkDA predictions show a general cooling of September T2M, which is significant at the 95% level over most of the Arctic Ocean. This cooling improves the model error relative to the ERA-I atmospheric reanalysis, over the

5 majority of this area (Figure 8). The exception to this is south of the Fram Strait in ice export regions, where the T2M has become too cool. We hypothesise that this small increase in error is likely due to the model simulating too much sea ice transport south through the Fram Strait. 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. On the Pacific side of the Arctic Ocean, where T2M in the ThkDA experiment is higher than for the CTRL

10 experiments, very little improvement (or degradation) of the T2M is seen.

We next consider the longer timescale quasi-equilibrium response (Semmler et al, 2016) to the pressure fields (MSLP and z500). A significant decrease in MSLP and z500 is seen in the ThkDA experiment over the Arctic Ocean with an accompanying increase over Siberia (significantly so for z500), and with small non-significant increases over the North Atlantic and Pacific

15 (Figure 8). This reduction leads to a decrease in error over the Canadian Basin and Greenland, but slightly worse comparison with observations over the Barents Sea and Western Europe. These differences in error however are generally not significant save for a small patch of improved MSLP over the Canadian Arctic Archipelago (Figure 8).

The z500 and MSLP decrease over the Arctic is suggestive of an increase in both the Arctic Oscillation (AO) and North

20 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). It is also broadly consistent with the lower Arctic z500 and wave-train nature of pressure anomalies over Eurasia observed in Wu and Zhang (2013) and Screen et al. (2013) for summertime circulation patterns related to above average sea ice areal coverage. However, owing to the small sample of years looked at here, it is doubtful we

25 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 7) these biases align very well with the ice edge errors suggesting a clear relationship between model thickness bias

30 and forecast error. We would therefore like to understand whether the improvements we see in the GloSea seasonal predictions 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.

15

To try to answer this question another ensemble of seasonal predictions was performed, for years 2011-2014 only, using the 2015 sea ice initial conditions each year. This ensemble of predictions is denoted FIXED-IC. We note here that FIXED-IC predictions are not performed for 2015 because they would simply be a duplication of the ThkDA-HC 2015 predictions. The motivation for adopting this approach is to ensure that we have a dynamically self-consistent initial condition for the sea ice
5   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 FIXED-IC can be found in Figure 4 alongside those for the ThkDA-HC and CTRL-HC ensembles. The 2015 predictions for ThkDA-HC have been replicated and included as part of the
10 FIXED-IC experiment. As was the case for the ThkDA experiment, the FIXED-IC predictions 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 FIXED-IC 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 FIXED-IC in 2011, 2012 and 2014. With only a short 4-year time series however, it is not possible to distinguish between the FIXED-IC and ThkDA-HC runs
15 statistically.

Interestingly the FIXED-IC predictions 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
20 used here is sufficient to remove atmospheric variability from the ensemble mean. It also suggests that the initial 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 predictions, performed using constant 2015 initial conditions for both the ocean and sea ice components, gave very similar results to that seen in the FIXED-IC experiment (not shown).

25 **5 Summary and conclusions**

In this study, we have used nudging techniques to test the impact that initialising sea ice thickness using CryoSat-2 (CS2) measurements could have on Met Office seasonal predictions of September sea ice extent. We have shown that initialisation of sea ice thickness significantly improves the accuracy of GloSea seasonal predictions 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
30 ice-edge location – particularly in the Atlantic sector. These improvements to the sea ice cover also lead to improvements in near-surface temperature and pressure fields over the Arctic domain.

16

Deleted: hindcasts
Deleted: hindcasts
Deleted: CLIM-2015
Deleted: CLIM-2015
Deleted: hindcasts
Deleted: hindcasts
Deleted: doing things this way

Deleted: CLIM-2015
Deleted: hindcasts
Deleted: CLIM-2015
Deleted: CLIM-2015
Deleted: hindcasts
Deleted: CLIM-2015
Deleted: CLIM-2015
Deleted: CLIM-2015

Deleted: CLIM-2015
Deleted: hindcasts

Deleted: forecasts
Deleted: CLIM-2015

Deleted: forecasts
Deleted: thickness
Deleted: forecasts
Deleted: forecasts

Technically the application of thickness increments within the CICE sea ice model has been shown to work well. The winter thickness initial conditions, generated using the sea ice thickness nudging, are much closer to the CS2 thickness observations, and lead to considerable improvement in skill when used to initialise GloSea seasonal predictions. 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 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) predictions initiated from them.

The motivation for using a simple assimilation approach in this study, using monthly gridded CS2 observations and a nudging technique (as outlined in Section 3), is that it provides a relatively simple way for us to test our hypothesis – that CS2 thickness initialisation will improve seasonal predictions of September Arctic sea ice. The results of this study, made using this approach, suggest that sea ice thickness assimilation within the FOAM ocean-sea ice analysis is feasible and could have a positive impact on the skill of GloSea seasonal predictions. Motivated by the findings of this study, work is now underway at the Met Office, under the EU-SEDNA project ("Safe maritime operations under extreme conditions: the Arctic case"), to include sea ice thickness assimilation within the NEMOVAR 3D-Var FGAT scheme used in FOAM (Waters et al., 2015), in combination with the sea-ice concentration assimilation already in place. This work will require prescription of observational errors (including instrument, algorithm, and representativeness errors), and the development of methods to represent appropriate model background errors. It will also involve using raw (L2) satellite tracks, from as many observational platforms as possible (including both CS2 and SMOS), with information being spread through the model using spatial and inter-variable error correlations.

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, 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 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 7, along with the improved ice cover obtained using thickness initialisation (Figure 4 and Figure 5), highlights the importance that modelled winter thickness biases can have on the evolution of forecast errors through the melt season. Results from the FIXED-IC

17

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.

5 Although the addition of sea ice thickness nudging to the FOAM analysis system clearly improves the seasonal predictions 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 due 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 prediction of September ice edge location when compared with the run using the 2015 thickness (FIXED-IC). This is in agreement with the

10 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.

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

References

30 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.

18

Deleted: forecasts

Deleted: down

Deleted: forecast

Deleted: CLIM-2015

Deleted: CLIM-2015

Deleted: FOAM-

Deleted: *hindcast*

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.

5

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, 10 M., Lemieux, J.-F., Dupont, F., Kuragano, T., Kamachi, M., Awaji, T., Caltabiano, A., Wilmer-Becker, K., and 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. 15 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.

20 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.

25

Blanchard-Wrigglesworth, E., Barthélemy, A, Chevallier, M., Cullather, R., Fučkar, N., Massonnet, F., Posey, P., Wang, W., 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.

30

Blockley, E. W., Martin, M. J., McLaren, A. J., Ryan, A. G., Waters, J., Lea, D. J., Mirouze, I., Peterson, K. A., Sellar, A., 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.

19

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 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.

5

Bloom, S. C., Takacs, L. L., da Silva, A. M., and Ledvina, D.: Data assimilation using incremental analysis updates, Mon. 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
10   Climate: A 25-Year Journey. Bull. Amer. Meteor. Soc., 93, 1865-1877, doi:10.1175/BAMS-D-12-00018.1, 2012.

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.

15   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
20   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
25   Thickness in a Coupled Ocean-Atmosphere Model. Mon. Wea. Rev., 143, 4618-4630, doi:10.1175/MWR-D-15-0097.1, 2015.

Cummings, J. A. and Smedstad, O. M.: Ocean Data Impacts in Global HYCOM, J Atmos Ocean Tech, 31, doi:10.1175/JTECH-D-14-00011.1, 2014

30   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.

Guerreiro, K., Fleury, S., Zakharova, E., Kouraev, A., Rémy, F., and Maisongrande, P.: Comparison of CryoSat-2 and ENVISAT radar freeboard over Arctic sea ice: toward an improved Envisat freeboard retrieval, The Cryosphere, 11, 2059-2073, doi:10.5194/tc-11-2059-2017, 2017.

Guo, D., Gao, Y., Bethke, I. et al.: Mechanism on how the spring Arctic sea ice impacts the East Asian summer monsoon, Theor. Appl. Climatol. 115: 107, doi:10.1007/s00704-013-0872-6, 2014.

21

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.

5 Holland, M.M., Bailey, D.A., and Vavrus, S.: Inherent sea ice predictability in the rapidly changing Arctic environment of 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., and 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.

10

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.

15

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.,
20 Goessling, H. F., Holland, M., Inoue, J., Iversen, T., Klebe, S., Lemke, P., Losch, M., Makshtas, A., Mills, B., Nurmi, P., 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.

25 Kaleschke, L., Tian-Kunze, X., Maaß, N., Beitsch, A., Wernecke, A., Miernecki, M., Müller, G., Fock, B. H., Gierisch, A. 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 -
30 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.

22

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. and Cunningham, G. F.: ICESat over Arctic sea ice: Estimation of snow depth and ice thickness, J. Geophys. Res., 113, C08010, doi:10.1029/2008JC004753, 2008.

5

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.

10   Laxon, S., Peacock, N., and Smith, D.: High interannual variability of sea ice thickness in the Arctic region, Nature, 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
15   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.

20   MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., Gordon, M., Vellinga, M., Williams, A., Comer, R. E., Camp, J., Xavier, P., and Madec, G.: Global Seasonal Forecast 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
25   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.
30
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 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.

Notz, D. and Stroeve, J.: Observed Arctic sea-ice loss directly follows anthropogenic CO2 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, 2015.

Petoukhov, V. and V. A. Semenov: A link between reduced Barents-Kara sea ice and cold winter extremes over northern 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., and 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.

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.

24

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 meridional overturning circulation at 26.5°N, Geophys. Res. Lett., 40, 5164-5170 doi:10.1002/grl.50930, 2013.

5 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 winters, Geophys. Res. Lett., 41, doi:10.1002/2014GL059637, 2014.

10 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.

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.
15

Screen, J. A., Simmonds, I., Deser, C., and Tomas, R.: The Atmospheric Response to Three Decades of Observed Arctic Sea Ice Loss, J. Climate, 26, 1230–1248, doi:10.1175/JCLI-D-12-00063.1, 2013.

Semmler, T., Jung, T., and Serrar, S.: Clim. Dyn. 46, 1015, doi:10.1007/s00382-015-2629-7, 2016.
20

Sigmond, M., Fyfe, J. C., Flato G. M., Kharin, V. V., and Merryfield, W. J.: Seasonal forecast skill of Arctic sea ice area in a dynamical forecast system, Geophys Res Lett., doi:10.1002/grl.50129, 2013.

Slingo, J. and Palmer, T.: Uncertainty in weather and climate prediction, Phil. Trans. R. Soc. A 2011 369 4751-4767;
25 doi:10.1098/rsta.2011.0161, 2011.

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.

30 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., Hyder, P., Kuhlbrodt, T., Rae, J. G. L., and Sinha, B.: UK Global Ocean GO6 and GO7: a traceable hierarchy of model resolutions, Geosci. Model Dev., 11, 3187-3213, doi:10.5194/gmd-11-3187-2018, 2018.

5   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., Day, J. J., Guemas, V., Hurlin, W. J., Keeley, S. P. E., Matei, D., Msadek, R., Collins, M., and Hawkins, E.: Seasonal to interannual arctic sea ice predictability in current global climate models, Geophys. Res. Lett., 41, 1035-1043, 10   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.

15   Tilling, R. L., Ridout, A., and Shepherd, A.: Near-real-time Arctic sea ice thickness and volume from CryoSat-2, The 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 20   systems, J. Oper. Oceanogr., 8, sup2, s201-s220, doi:10.1080/1755876X.2015.1049892, 2015.

Tsamados, M., Feltham, D. L., and Wilchinsky, A. V.: Impact of a new anisotropic rheology on simulations of Arctic sea ice, J. Geophys. Res. Oceans, 118, 91-107, doi:10.1029/2012JC007990, 2013.

25   Tsamados, M., Feltham, D. L., Schroeder, D., Flocco, D., Farrell, S. L., Kurtz, N., Laxon, S. W., and Bacon, S.: Impact of Variable Atmospheric and Oceanic Form Drag on Simulations of Arctic Sea Ice, J. Phys. Oceanogr., 44, 1329-1353, doi:10.1175/JPO-D-13-0215.1, 2014.

Uotila, P., Goosse, H., Haines, K., et al. Clim Dyn , doi:10.1007/s00382-018-4242-z, 2018.

30

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. 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.

Vihma, T.: Effects of Arctic Sea Ice Decline on Weather and Climate: A Review, Surv Geophys, 35:1175, doi:10.1007/s10712-014-9284-0, 2014.

5 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., 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., doi:10.5194/gmd-2017-10 291, in review, 2017.

Waters, J., Lea, D. J., Martin, M. J., Mirouze, I., Weaver, A., 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, 2015.

15 Wang W., Chen M., and Kumar, A.: 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, 2013.

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. 20 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, 2017.

Wu, B., Zhang, R., D'Arrigo, R., and Su, J.: On the Relationship between Winter Sea Ice and Summer Atmospheric Circulation 25 over Eurasia, J. Climate, 26, 5523–5536, doi:10.1175/JCLI-D-12-00524.1, 2013.

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.

30 Yang, Q., Losa, S. N., Losch, M., Tian-Kunze, X., Nerger, L., Liu, J., Kaleschke, L., and Zhang, Z.: Assimilating SMOS sea ice thickness into a coupled ice-ocean model using a local SEIK filter, J. Geophys. Res. Ocean 119: 6680–6692, doi:10.1002/2014JC009963, 2014.

Deleted: 2018

27

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

Experiment name	Renalaysis 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	CPOM CryoSat-2

5   **Table 2: Details of GloSea coupled seasonal prediction experiments (or 'hindcasts') used in this study**

Experiment name	Prediction lead time	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
FIXED-IC	May – Sep	2011-2014	24 per year[a]	ERA-Interim	ThkDA-RA: fixed 2015 for all years

[a] 24 members per year = 3 start dates with 8 stochastic members for each
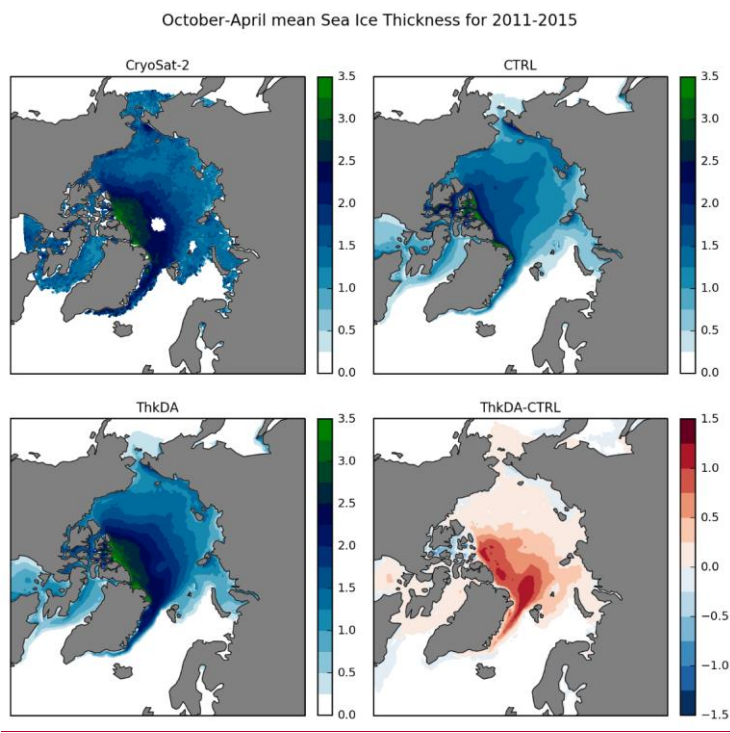
Figure 1: **Mean winter (October to April) Arctic sea ice thickness (m) from October 2010 to April 2015. Showing data from the CPOM CryoSat-2 measurements (after application of the QC and imposing the 1m minimum thickness threshold), along with modelled thickness from the CTRL and ThkDA reanalyses. The lower-right panel shows the difference between the ThkDA and CTRL experiments.**
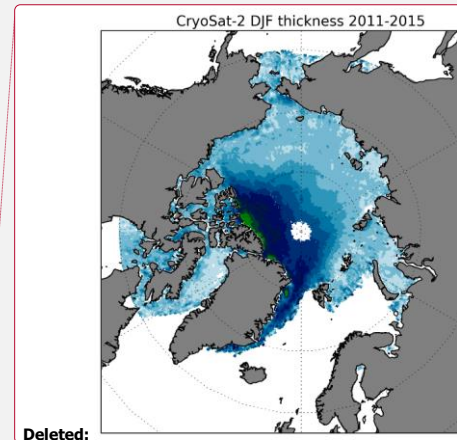
Deleted:

Deleted: Example

Deleted: Arctic thickness used in this study derived from

Deleted: : showing December-January-February (winter) mean thickness (m) over the period 2011-2015

Deleted: .
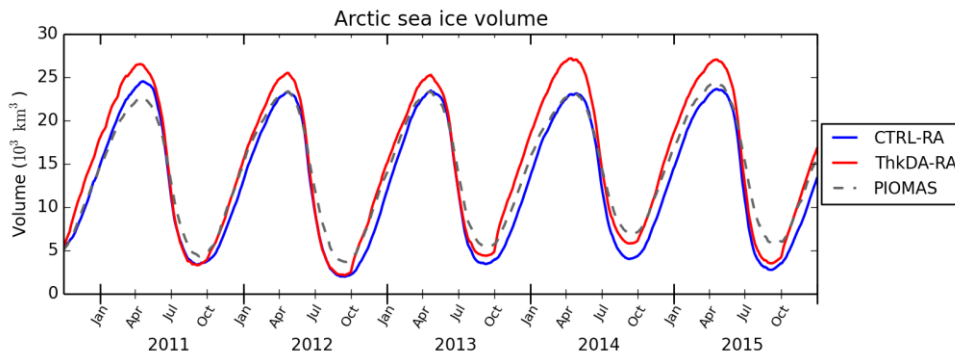
Figure 2: Reanalysis Arctic sea ice volume from 1st October 2010 to the end of December 2015 from the CTRL-RA (blue) and ThkDA-RA (red) reanalysis experiments. Sea ice volume from the PIOMAS model (grey dashed) is included as a reference.

5
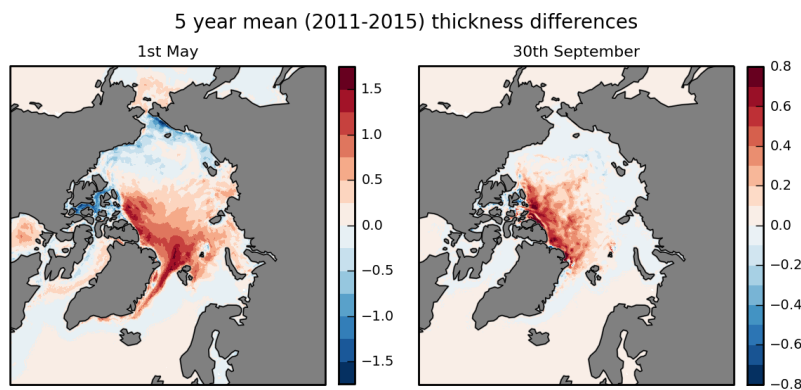
Figure 3: Mean sea ice thickness difference (m) between ThkDA-RA and CTRL-RA experiments over the full 5-year reanalysis period from 2011 to 2015. Showing differences for (left) the end-of-winter on 1st May used for the initialisation of summer seasonal forecasts, and for (right) the end-of-summer on 30th September. The difference is taken as ThkDA – CTRL so red (blue) implies that the CS2 nudging has increased (decreased) the thickness.

10
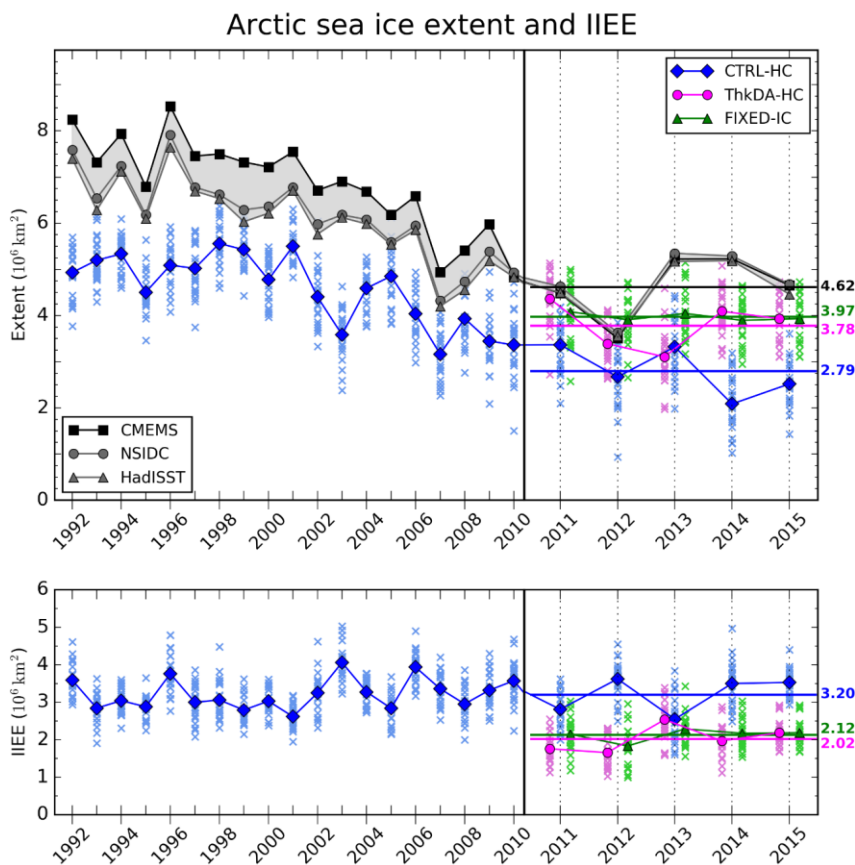
Figure 4: (upper) September-mean Arctic sea ice extent from the CTRL-HC (blue), ThkDA-HC (pink) and FIXED-IC (green) seasonal prediction experiments. Observational estimates from the CMEMS reanalysis assimilating OSI-SAF (black square), NSIDC (grey circles) and HadISST1.2 (grey triangles) are included and the area between them shaded light grey. **(lower)** Integrated Ice Edge Error (IIEE) for seasonal predictions relative to the CMEMS reanalysis product introduced in Section 2.3. In both panels, 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. For ease of viewing, the ThkDA-HC (pink) and FIXED-IC (green) experiments are plotted with a small offset relative to the CTRL-HC (blue) experiment, and the CS2 period (2011-2015) is plotted with an increased x-axis scale, approximately twice that for the early period 1992-2010, with the transition indicated by a vertical black line.
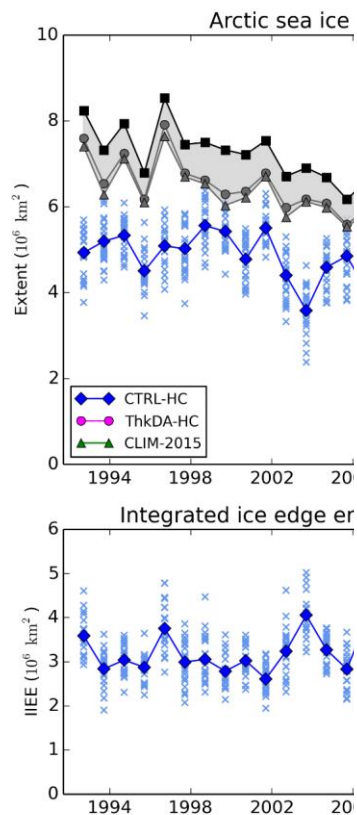
Deleted: or
Deleted: CLIM-2015
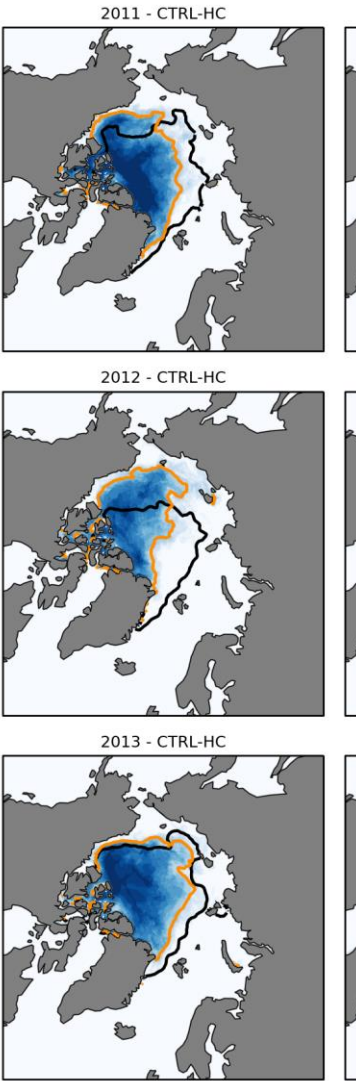Deleted: forecast
Deleted: forecasts

32

September forecast probability of ice (conc>15%)

2011 - CTRL-HC

2011 - ThkDA-HC

OBS = 4.46
MOD = 3.35
IIEE = 2.80

OBS = 4.46
MOD = 4.34
IIEE = 1.75

2012 - CTRL-HC

2012 - ThkDA-HC

OBS = 3.48
MOD = 2.66
IIEE = 3.60

OBS = 3.48
MOD = 3.37
IIEE = 1.65

2013 - CTRL-HC

2013 - ThkDA-HC

OBS = 5.20
MOD = 3.31
IIEE = 2.54

OBS = 5.20
MOD = 3.09
IIEE = 2.53

September forecast prob

2011 - CTRL-HC

2012 - CTRL-HC

2013 - CTRL-HC
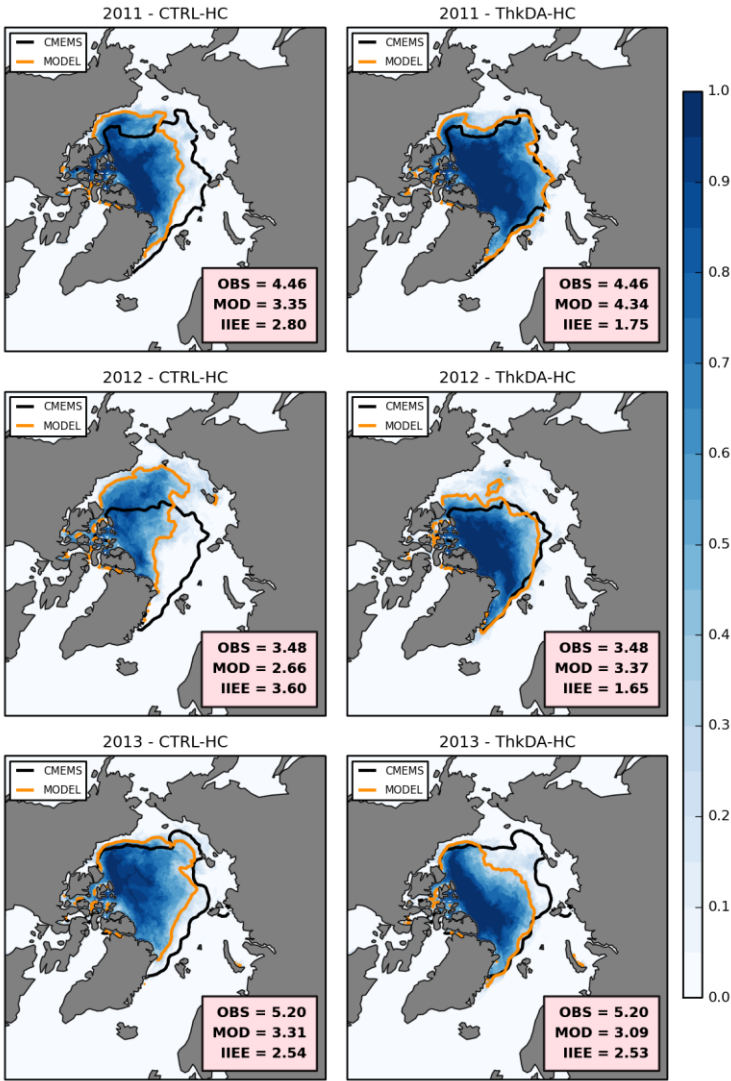
Deleted:

September forecast probability of ice (conc>15%)

2014 - CTRL-HC

2014 - ThkDA-HC

OBS = 5.20
MOD = 2.07
IIEE = 3.48

OBS = 5.20
MOD = 4.07
IIEE = 1.96

2015 - CTRL-HC

2015 - ThkDA-HC

OBS = 4.63
MOD = 2.51
IIEE = 3.51

OBS = 4.63
MOD = 3.91
IIEE = 2.18
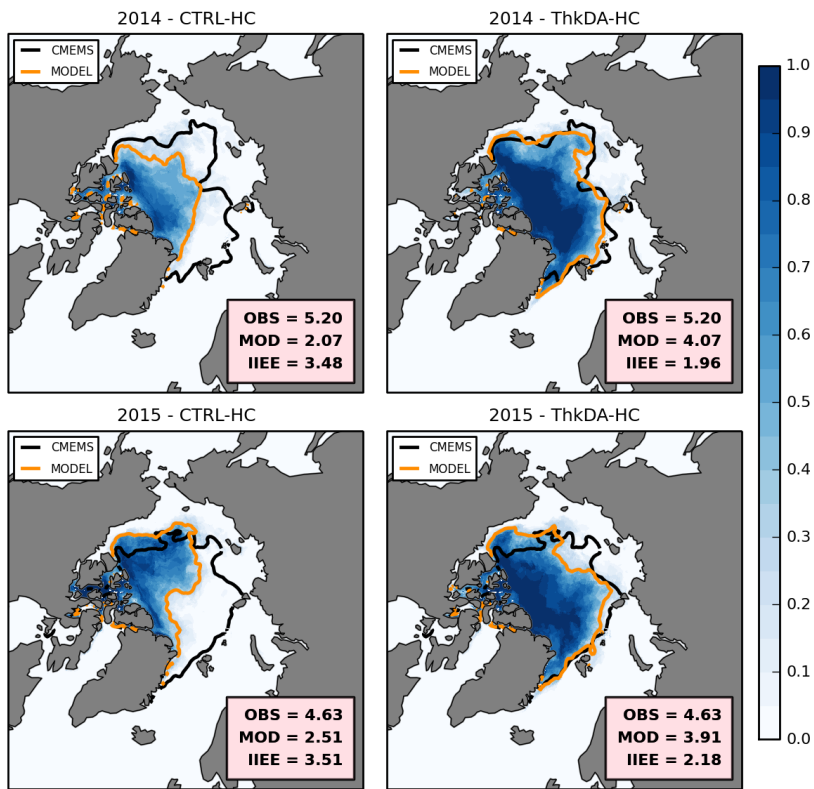
Figure 5: September mean probability of sea ice for the CTRL-HC (left) and ThkDA-HC (right) seasonal predictions for all years from 2011 (top) to 2015 (bottom). Contours of 15% concentration are overlain to represent the sea ice edge for the ensemble mean (orange) and CMEMS reanalysis product (black). Probability is defined at each point as the proportion of ensemble members that have at least 15% ice concentration. The CMEMS extent, modelled extent and corresponding Integrated Ice Edge Error (IIEE) are included, for each plot, in the lower-right corner (units: 10^6 km^2).
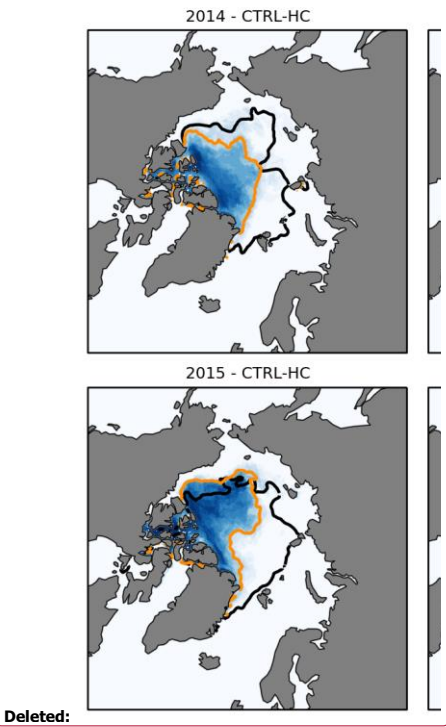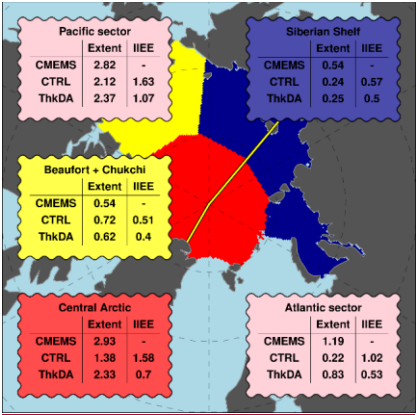
Figure 6: Mean September Arctic sea ice extent for 2011-2015 from the CMEMS reanalysis (using OSI-SAF) compared with modelled extent and ice edge error (IIEE) from the CTRL and ThkDA seasonal predictions (units x10^6 km^2). Data are shown for 3 regions distinguished by the underlying shading and corresponding box colours: combined Beaufort + Chukchi Seas (yellow), combined Kara, Laptev and East Siberian Seas (dark blue) and the central Arctic (red). Also shown (pink boxes) are corresponding statistics for the Atlantic and Pacific sectors of the Arctic Ocean, defined by splitting the Arctic Ocean (i.e., red + yellow + dark blue) along 30°W and 140°E longitude (yellow lines) – which roughly follows the Lomonosov Ridge.
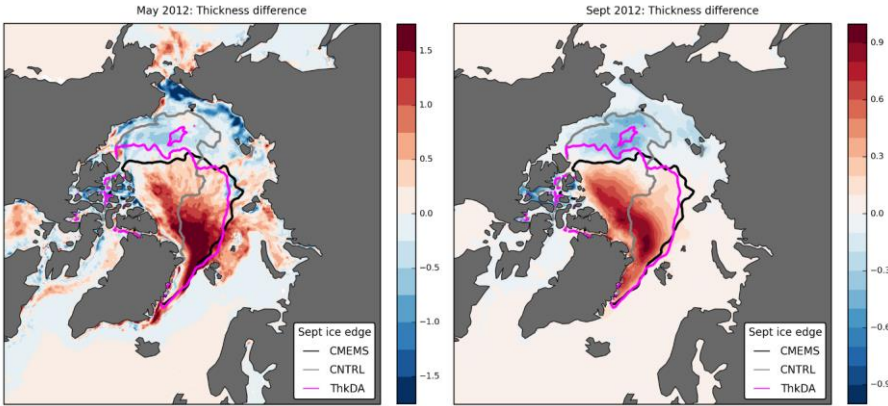
Figure 7: Thickness difference (m) between ThkDA and CTRL experiments in May and September 2012 (contour shading) with differences calculated as ThkDA – CTRL in each case. (left) shows the difference between the reanalyses fields used to initialise the seasonal predictions on 1st May 2012. (right) shows 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).
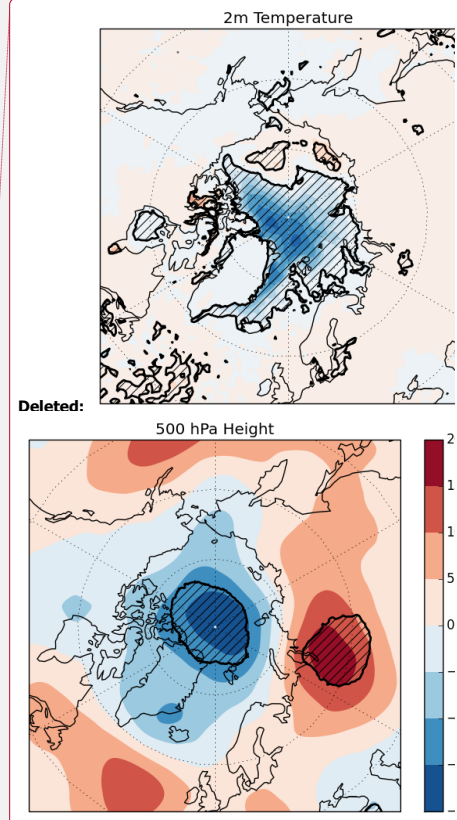
Deleted: 6

Deleted: around the start of

Deleted: hindcasts
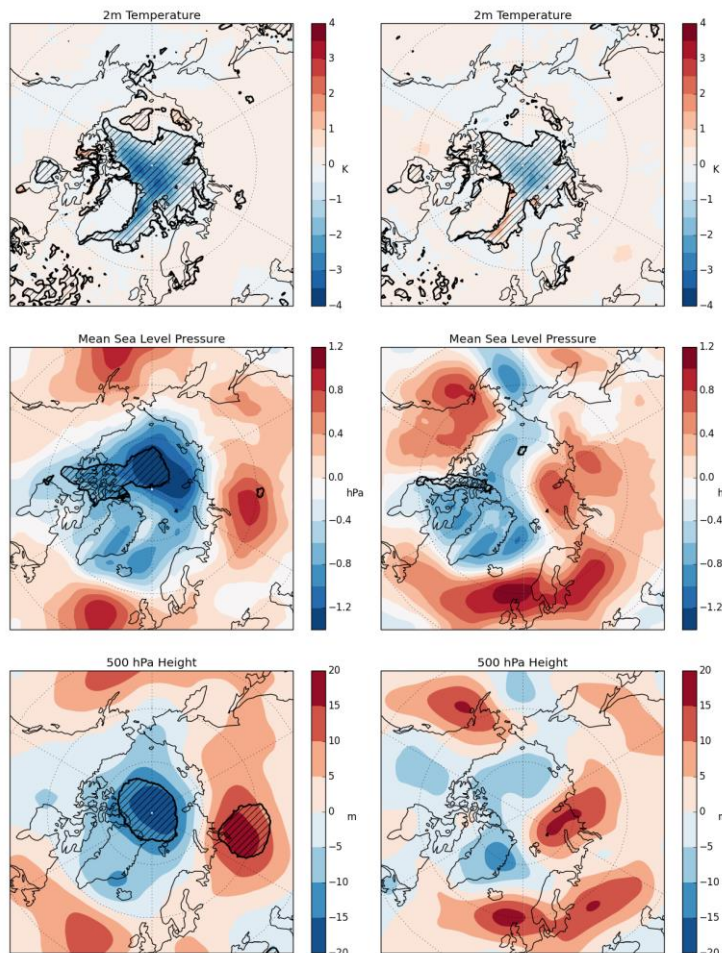
Figure 8: Differences in mean fields (left) and root-mean-square-error (RMSE; right) between the ThkDA and CTRL September predictions averaged over all ensemble members for the 5-year period 2011-2015. Fields shown are: (top) near-surface 2m air temperature (T2M; K); (middle) mean sea-level-pressure (MSLP: hPa); (bottom) 500 hPa geopotential height (z500; m). Differences are calculated as ThkDA – CTRL meaning that areas of blue (red) denote that ThkDA predictions/RMSEs 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. Further details can be found at the beginning of Section 4.2.

5