Response to the review of tc-2020-73 titled "Year-round Impact of Winter Sea Ice Thickness Observations on Seasonal Forecasts"

We are grateful to the two anonymous reviewers for the thorough review, insightful comments and generally positive response to our article titled "Year-round Impact of Winter Sea Ice Thickness Observations on Seasonal Forecasts". All of their remarks are addressed. Please find below our responses to the reviewers' comments and suggestions (*in blue italics*).

Response to Anonymous Referee #1

Balan-Sarojini and co-authors present a study examining the impact of sea ice thickness (SIT) assimilation on seasonal forecasts of the northern hemisphere sea ice cover. In its approach and scope the study covers new ground; several of the key findings are substantive and represent a significant advance in our understanding of sea ice predictability and performance of seasonal-scale forecasts. The authors make good use of newly available, state-of-the-art ice thickness fields and strike a nice balance between more fundamental questions of prediction system performance, and applied questions related to improving forecast skills of Arctic sea ice models. The paper is well structured and makes good use of figures to illustrate key points. The scientific approach is well described and appropriate for the problem at hand. The first half of the paper (up to and including Section 3.2, Fig. 7) is particularly compelling and self-contained. The latter part of the manuscript, while interesting, is less compelling with some of the writing lacking clarity and the paper losing focus with respect to the goals laid out in the introduction and implicit in the title. If this part of the paper is retained in full, tightening the text and improving readability of sections 3.3-3.5 in particular would make the paper more accessible.

We thank the reviewer for the constructive comments on the article. Regarding sections 3.3-3.4 we propose to follow the reviewer's suggestion on improving the readability of these sections in the revised manuscript. We also justify the presence of these sections in the introduction by pointing out that this work takes a forecasting system end-to-end perspective, from observations, modelling to forecast products. Thus, in the revised version we clarify that the paper has three main foci sections, as the reviewer has noticed: i) assessing the seasonal forecasting system performance using new sea ice thickness (SIT) observational information (sections 3.1 and 3.2) , ii) improving Arctic sea-ice forecast skill by understanding the errors in the coupled forecast model and the data assimilation system through targeted diagnostics (sections 3.3), and iii) quantifying the impact of sea-ice improvements on seasonal forecasts of atmospheric variables (section 3.4). We agree with the reviewer that sections 3.3 and 3.4 were not properly motivated in the introduction, and we intend to do so in the revised manuscript. Section 3.3 describes the inconsistency between the errors in the coupled forecast system and the analysis, a key result that points out that the dominant source of error lies in the atmospheric forcing rather than in the sea-ice model formulation or data assimilation, and indicates that improvement of atmospheric fluxes from atmospheric reanalyses is urgently needed to improve the Arctic sea-ice related forecasts. Section 3.4 explores the impact on forecast skill of atmospheric variables. Although seasonal predictions of sea ice can be an end by itself, a prime objective of ECMWF is the forecast of atmospheric variables. Therefore, a key part of the evaluation methodology for system developments includes the verification of impact of atmospheric variables. In a first instance, this acts as a sanity check to make sure there are no obvious degradations, which adds robustness to the developments. In a second instance, it helps to quantify the impact on forecasts arising from small incremental improvements, which helps to put the SIT impact in the context of other model/data assimilation improvements. Hence, we would prefer to keep sections 3.3-3.4 in the same article for future reference. Please note that there is no section 3.5 in the article, and we believe it is a typo. Please also see our response to Reviewer 2's specific comment on L12-L13. [L115-L116, L120-L127]

At the same time, a few aspects of the paper can be improved or require further thought, as outlined below. First, given the central role the SMOS/Cryosat-2 data set plays in this study, one would like to see some discussion of how errors and uncertainties in the ice thickness data set may have impacted forecast skill and in particular some of the regional patterns observed in the thickness-assimilation runs. As shown in Ricker et al. (2017) uncertainties due to the fundamentally different retrieval approaches for SMOS and Cryosat2, and to a lesser extent the optimal interpolation and data merging schemes, vary significantly by region. For example, over the Canadian Basin with mostly thick, multiyear ice the data product is dominated by bias/errors in Cryosat-2 data whereas in the Bering or Okhotsk Sea thinner ice weights errors towards those associated with SMOS thickness retrievals. It would be important to establish whether differences in thickness-field uncertainties have any impact on model performance and regional or temporal contrasts in forecast bias. This is also relevant for the integrated analyses of parameters such as the Integrated Ice Edge Error or ice volume at the pan-Arctic scale which may gloss over important regional contrasts in model performance.

We completely agree with the reviewer that the thickness uncertainties should be considered in sophisticated assimilation of SIT. However, in this work we have only taken the very first step in SIT assimilation by using a simple nudging method to constrain SIT from the merged Cryosat2-SMOS product without considering the observational uncertainties. In a merged product like CS2SMOS, it is difficult to represent the sensor-specific errors properly. As the reviewer commented, sensor-specific errors could affect regional SITs, i.e., over multiyear thick ice over the Canadian Basin, errors associated with Cryosat-2 retrievals dominate whereas in the Bering or Okhotsk Sea with mostly seasonal thin ice, errors associated with SMOS retrievals dominate. As reported in Ricker et al. 2017, the relative error is maximum in the thickness range of 0.5-1.0 m in the merged product where both CS2 thick ice and SMOS thin ice retrieval errors are maximum. We add this point in Section 2.1.2 and in the concluding section mentioning it as an area which needs to be explored in future studies of SIT assimilation. For instance, the uncertainty in the retrievals could be taken into account by perturbing the observations in the ensemble of data assimilations. We add a sentence on this aspect in the revised manuscript. Equally, the verification will benefit from having records of SIT with easy-to-use uncertainty estimates. The practice followed by HadISST2, which provides an ensemble of SST records to cover different sources of uncertainty in SST, is proving very convenient. [L200-L208, L790-L795, L800-L803]

Second, the paper lacks detail on the representation of ice thickness and key ice growth, melt and deformation processes in the LIM2 prognostic thermodynamic-dynamic sea ice model used in this study. It would be important to provide more detail, in particular as to whether any of the parameterizations that are part of the Fichefet & Morales Maqueda (1997) – FMM97 – model have been updated or *changed. Of potential concern in FMM97 – based on description in their original paper – would be the limited representation of surface melt processes and their impact on ice albedo as well as physically unrealistic representation of internal ice melt (with internal "storage" of solar heat up to a 50% threshold). These shortcomings, which may have been addressed in updates but if so the paper needs to make this explicit, do not necessarily limit applicability of the model in the context of seasonal ice forecasts. However, they are problematic in diagnosing some of the linkages between surface forcing, energy storage and the seasonal ice cycle explored in Section 3.3, since FMM97 in its original form may be ill suited to examine in particular the spring-summer-fall transitions in terms of the surface radiation balance or rates of ice thinning and decay. Given these potential concerns, it would be instructive in Section 3.3 to examine the proportion of up/downwelling shortwave fluxes (or albedo, for that matter) to get a better perspective on the sensitivity of sea ice as represented in FMM97 to variations in downwelling shortwave energy. Such a detailed analysis may well be beyond the scope of the present paper. If so, this may be an argument to remove the latter parts of the paper as the basis for a separate, more detailed study. The first part of the paper (up to Section 3.3) is substantive enough and fully in line with the title of the paper.*

We acknowledge that LIM2 is a simple sea ice model, which we have in our current operational system since 2017. The original version of FMM97 is used. As the reviewer points out, it has several limitations of surface melt processes such as for instance, representation of melt ponds which could affect the accurate representation of surface albedo over sea-ice. However, we note that only the ocean reanalysis ORAS5 actually makes use of the albedo computed by LIM2 (which is too high in summer), while the atmospheric reanalyses and the forecasting system use the same climatological albedo (based on SHEBA campaign observations; Beesley et al. 2000). This means that the differences found in Figure 10 cannot be attributed to different albedo biases in the atmospheric reanalyses and forecasts. We would also like to point out that a recent comparison study (Pohl et al. 2020) shows that, overall, the broadband albedo over Arctic sea-ice derived from MERIS observations is comparable to that in the ERA5 atmospheric reanalysis in terms of the seasonal cycle on the large spatial scales. We find that the forecast albedo over ice is comparable to that in ERA-5 and ERA-Interim atmospheric reanalyses. Moreover, it has been shown that the downwelling short wave radiation has a negative bias over the central Arctic in both the atmospheric reanalyses and the reforecasts (Hogan et al. 2017, Balan-Sarojini et al. 2019). We add the specific points in the model description part. We also emphasize that, although the manuscript does not cover sea-ice model improvements, recent developments in modelling sea-ice processes affecting the sea-ice melt/growth (https://forge.ipsl.jussieu.fr/nemo/chrome/site/doc/SI3/manual/pdf/SI3_manual.pdf) are being considered for inclusion in upcoming versions of the ECMWF forecasting systems. [L245-L246, L255-L266, L795-L800]

Third, starting with the discussion of sea ice volume at the pan-Arctic or northern hemisphere scale the paper began to veer off-course a bit in terms of the goals laid out in the introduction. While total ice volume is a great integrator and a relevant variable in a global context, I was not able to tell whether the authors were assuming that it can also serve as an integrated measure of model performance in terms of ice concentration/extent and ice thickness. Given the regional contrasts in model performance apparent *in the early figures of the paper the wholesale discussion of ice volume is somewhat problematic. For example, the interpretation of the seasonal ice volume cycle in terms of a single "freezing rate" (p. 17, top paragraph) is too simplistic since increases in ice volume during fall and winter occur through a combination of ice deformation and ice growth inside the ice pack as well as advance of the ice edge in marginal seas. Without an in-depth analysis some of the earlier figures and a solid understanding of how well the sea ice model is capturing the relevant processes, Figures such as Fig. 8, don't add that much to the paper and could be relegated to supplemental materials or cut completely.*

We agree to the reviewer that an integrated quantity like Arctic sea ice volume is a result of many dynamic and thermodynamic sea-ice processes and lacks regional details. However, integrated SIV is a key indicator for understanding of the Arctic energy cycle, an important process that needs to be represented in reanalyses and seasonal forecast. It is useful to compare the contrasting SIV seasonal cycles in coupled and uncoupled mode, and with/without SIT observational constraint in the initialization, since this helps to identify the origin of errors in the systems in the specific operational set up. As noted in L304-305, SIC increments in the ORAs do affect analyzed SIV. We add a few sentences to discuss the benefits and caveats of using pan-Arctic sea-ice volume as a diagnostic for model performance in the revised manuscript. [L506-L513]

Finally, just a few minor points: - Comparing bias in ice thickness (Fig. 1) with bias in ice concentration (Fig. 3) it's striking that regions with near-zero bias in thickness (e.g. East Siberian Sea, Chukchi Sea in November) show up as having significant bias in ice concentration; moreover despite substantial contrasts in thickness biases between reference and ice thickness runs (Fig. 1c&d) the biases in ice concentration are near indistinguishable (Fig. 3 g&h). How can this be explained?

We thank the reviewer for raising this important question, which we failed to comment in the original version. The sea-ice thickness bias in ORA-SIT in these areas is very small, about 0.1 to 0.05 m (below the contour interval). The presence of concentration bias (Figure 3, similar pattern in ORA-SIT and ORA-REF) in regions with negligible thickness bias in ORA-SIT is suggestive of fast growth processes in the forward model, faster than the timescales intrinsic to the assimilation of sea-ice concentration. We add this explanation in the revised manuscript. [L359-L361, L383-L391]

- In regards to July ORASIT biases in ice concentration, it was striking to see much larger bias in the ORA-SIT than in the reference runs. Why would the simulations that performed (understandably) so much better in replicating ice thickness in March through assimilation of ice thickness data perform much worse in replicating ice concentration in July?

A very pertinent question, we asked ourselves as well. As noted in L182-L185 of original text, sea-ice thickness observations are only available and assimilated for November-March. The ORA-SIT sea ice thickness from April-October are not constrained by observations. The fact that ORA-SIT has larger errors than ORA-REF in SIC for July is attributed to the overestimation of the melt in the forced model, as discussed in the last two paragraphs of section 3.1. The negative summer SIC bias gets worse in ORA-SIT than that in ORA-REF due to the fact that the ORA-SIT starts from a thinner ice state compared to ORA-REF without CS2SMOS thinning. Please also read our response to L208-209 in the Reviewer2's response part. Indeed, the assimilation of sea-ice concentration is trying hard to compensate for this excess of sea-ice melt (Figure 4). The reasons for this excess sea-ice melt during this season is further investigated in section 3.3 and attributed to errors in forcing fluxes. This is one of the main outcomes of this work. We revise the manuscript to bring out the answer to the reviewer's question more clearly by adding the above discussion in Section 4. [L774-L789]

Note that this finding also seems to contradict your statement in l. 185 that "The non-availability of the observations for the melt season in a way provides an opportunity to test the predictability of winter SIT from summer initial conditions."

Thanks for this point. We agree that the sentence is confusing and we remove it in the revised manuscript. [L328-L330]

You discuss your findings in terms of Arctic ice concentration and thickness but your figures include regions outside of the Arctic proper (such as the Okhotsk Sea). Please clarify whether both model output and assimilated data cover the entire northern hemisphere sea ice or a subset of that data. This is relevant in particular for figures like Fig. 5 which references "nh" in the figure label (for northern hemisphere?) but refers to Arctic sea ice area in the caption.

Thanks for pointing out the confusion on the definition of Arctic domain in the article. We would like to clarify that we have used a global model, so its output includes the entire northern hemisphere. And we exploit the full spatial coverage of CS2SMOS data set, which covers all the regions where sea ice has ever been observed in recent decades, so it can safely be treated as representing the entire northern hemisphere. This information is clearly stated in S2.1.2 and S2.2.1 in the revised manuscript.

We also add a definition of pan-Arctic (as the sea ice area included is of the whole of NH which is essentially the Arctic sea ice, the Okhotsk sea ice and the Baltic sea ice) in the beginning of the Results section and use the term 'pan-Arctic' wherever appropriate except for Figures where we have explicitly mentioned the Arctic domain area in the caption. In this work, we are interested in the large-scale impact and not in grid-point scale impact. [L194-L195, L340-L342]

Response to Minor comments & corrections

l. 2/3: change to "in its early stage"

thanks, it is done. [L14]

l. 20 "near-surface temperature forecasts of early freezing season initialized in May": This phrase is confusing and not entirely clear, please revise to clarify what specifically is forecast with respect to "freezing season".

thanks, it is rephrased as "near-surface temperature forecasts of early freezing season (Sept-Oct-Nov) initialized in May". [L43-L44]

l. 25: change to "lasts into autumn"

thanks, it is done. [L52]

l. 80: "it is relevant as cross-check variables evaluation" – not entirely clear what's referenced here – should it be "they are relevant because they allow for cross-checking between variables"? Please clarify.

thanks for pointing it, the reviewer is right. It is rephrased as "Although the datasets used for verification are not strictly independent, evaluation using those datasets is relevant as it allows cross-checking between variables, for instance between SIC and SIT assimilation.". [L144-L147]

l. 81: "SIT verification is also conducted as a sanity check of the nudging approach" – You lost me at "sanity check" – what exactly are you doing here? Please explain.

By 'sanity check' we meant that SIT verification using CS2SMOS dataset (Figure 1) is a basic test to check whether the nudging works in the first place. L81 of the original text is rephrased as "SIT verification using CS2SMOS dataset is also conducted as a sanity check of the nudging approach: the approach works, the differences with respect to CS2SMOS should be smaller in ORA-SIT than in ORA-REF.". [L147-L150]

l. 91: change to "The Level-3"

thanks, it is done. [L162]

l. 145: "LIM2 has a single sea ice category to represent sub-grid scale ice thickness distribution" – this needs further clarification. To calculate an effective conductive heat flux through the ice Fichefet and Morales Maqueda (1997) assumed a uniform thickness distribution bounded by zero and twice the average thickness. This parameterization was only applied in calculating heat fluxes through ice and lateral melt rate but did not enter into any of the ice dynamics components of the model. Given that ice thickness initialization is central to this manuscript, a clearer description of what exactly was implemented is needed.

The reviewer is right. As we responded earlier in the main comments section, the original version of Fichefet and Morales Maqueda (1997) LIM2 version is implemented in our operational system. We explicitly mention it in the revised version. [L245-L246]

l. 168: change to "differ in"

thanks, it is corrected. [L302]

l. 233: "These results clearly show. . ." – Some clarification is needed here, since I interpret Fig. 4 as indicating that through May (but not the entire melt season), the effects of SIT assimilation are evident, beyond that the reference run performs better through the end of melt. In linking SIC increments to SIT assimilation please also consider the points raised in the general comments above.

Thanks for pointing it out. We agree with the reviewer that a positive impact on SIT is seen till May and a negative impact is seen till September. As we have already described the nature of impact in the preceding paragraph, this is a general summary sentence at the end of the subsection. To avoid confusion we rephrase the sentence as "These results clearly show the long-lasting effect of the SIT information: the SIT constraint was only applied during the growth season from November to March (grey shading), but its impact, whether positive or negative, is evident in sea ice concentration throughout the melting season even in the presence of SIC assimilation”. [L416-L420]

l.238: “(units are. . .” – This should be part of the figure legend or caption, and not be buried in the main text.

The caption has it already. We remove it here, thanks. [L424-L425]

l. 245: change to “melt season forecasts are considerably reduced”

thanks, it is done. [L437]

l. 251: The top labels of the figure panels are cut off and it’s not clear that they’re actually needed (“bias for sia in area nh” – would need to be explained; also: is nh Northern Hemisphere? If so, what is the difference between this data for northern hemisphere and the Arctic sea ice area as indicated in the figure caption?); the color scale needs better labeling.

Thanks, we confirm that the Arctic domain we have considered everywhere, unless it is specifically mentioned in the Figure captions, are pan Arctic which is defined in the revised Results section too. We remove the confusing term in the figure panel in the revised paper. [L194-L195, L340-L342, P25]

l. 265: insert “are” in “that are to be expected”

thanks, it is done. [L472]

l. 268: Fig. 6 - This figure should be cleaned up a bit as well; there’s no need for two top labels (the upper one is more descriptive anyways, but even that’s not needed given the explanation in the caption); the color bar needs proper units. Fig 7: Same comments apply – the 1e12 and 1e11 squeezed right next to the figure panel label and disjunct from the axis label (with units of square meters) are less than ideal and need to be cleaned up.

thanks for the suggestion, it is done. Please see our response to Reviewer 2’s comment on L250-262 also. [P28, P29]

l. 287: Fig. 8: It’s not clear to me how an axis label of 10ˆ1 3 mˆ3 translates into 10ˆ12 mˆ3 as the figure caption claims. Why not put an axis label in kmˆ3?

thanks for spotting the typo, it is done. [P32]

l. 361, Figure 11: same comments as for Fig 6 apply

thanks, it is done. [P41]

l. 369: correct spelling of “Atlantic”

the typo is corrected. [L683]

Response to Anonymous Referee #2

Balan-Sarojini et al. study the impact of Cryosat2/SMOS winter ice thickness (SIT) observation nudging on a lower-resolution version of the ECMWF ocean/sea-ice reanalysis (ORA) system and on associated coupled seasonal forecasts initialized from that reanalysis system. The SIT constraint suppresses an otherwise too strong annual SIT/SIV cycle in the ORA and provides overall thinner SIT conditions toward the end of the northern winter (except in the perennial ice regions north of Greenland and the CAA), which turn into decreased sea-ice extent in the ORA in summer (despite sea-ice concentration assimilation). The thinner/less extensive initial ice is benefitial for seasonal forecasts initialized before July, but forecasts initialised in late summer tend to be deteriorated. The authors show that this is linked to too-strong spring/summer melt in the ORAs (when no SIT constraint is available), leading to low-biased ice and warm biased sea-surface initial conditions in summer, in combination with a too-late/too-weak refreeze in the coupled forecast system. Balan-Sarojini et al. show evidence that the latter points can be explained at least partly with the surface radiation budget in the atmosphere-forced ORAs and in the coupled forecast model. The study is scientifically sound, well-written, contains appropriate graphics and references, and provides interesting insights into the impact of ice thickness observations on forecasts in the specific system used which might be helpful to understand other systems, too. I do have quite a number of remarks, most of which are however minor. The maybe most demanding recommendation is to compare against simple climatological benchmark forecasts where appropriate. In summary, I recommend publication of this work in The Cryosphere subject to minor(-to-major) revisions as detailed in the following.

We thank the reviewer for the positive remarks on our article. The main suggestion to verify the reforecasts against a climatological benchmark forecast is appreciated. Comparison against a climatological benchmark forecast is a very useful background information for evaluating the predictive skill of multi-model-ensemble forecasting systems (for example as in Zampieri et al. 2018), and we add it in the revised manuscript, even if benchmarking dynamical seasonal forecast against climatology is not the main objective of the paper.[L478-L483, L499-L504, P29]

Response to Specific comments

L12-13: "we find significant improvement of up to 28% in the September sea ice edge forecast started from April" - From the abstract it does not become clear that the paper is almost completely focussed on biases (and how these affected by constraining SIT) and not on interannual anomalies. In the summary section you state very clearly that this is the case (L441-442), but I think it should be mentioned in the abstract, too. Without that information, the sentence in L12-13 leaves one wondering how such a significant forecast improvement can be reconciled with the "May predictability barrier". In this context, see also my recommendation below to consider comparing with a climatological benchmark forecast where appropriate.

Thanks for pointing it out. We mention in the revised abstract that change in biases is the main impact. [L29]

L57: Zampieri et al. 2018 - There's also a follow-on paper demonstrating reasonable skill of ECMWF S2S sea-ice forecasts in the Antarctic: Zampieri et al. 2019 "Predictability of Antarctic Sea Ice Edge on Subseasonal Time Scales".

We add here the reference of the suggested paper on the Antarctic sea ice skill too. [L102]

Eq. 1: It probably doesn't make a big difference, but can you specify whether this is "floe-thickness" or "effective thickness" (thickness when evenly distributed over grid cell)?

Thanks, we mention that it is the "floe thickness". [L283-L284]

L162-164: "We have also tested the sensitivity to different nudging strengths by running variants of ORA-SIT with a relaxation time scale of 20, 30 and 60 days" - If you mention this, I would expect that you also say something about the impact of the relaxation timescale and why you chose 10 days.

As the relaxation time scale increases from 10 days to 60 days, lesser constraint on SIT is found. We chose the time scale of 10 days for 2 reasons: 1) it fits to the length of the assimilation window, and 2) we first wanted to look at the forecast impact of the initial conditions with the maximum observational constraint. We add a line on this point. [L293-L298]

L201-205: "slight underestimation over the central Arctic and overestimation over the Canadian Archipelago still remain in November. This is probably caused by the lack of SIT observations during the months preceeding November" - Given the relaxation timescale of 10 days, I assume that this difference goes back almost completely to the first half of November? That would confirm that you could omit the word "probably"; that's a rather obvious link.

We agree with the reviewer. Please also see our response to the next remark. [L357]

L208-209: "The largest impact occurs in March, probably because at this month the SIT observations have been assimilated during the preceeding 5 months" - similar to what I say in the previous point, I assume that the SIT state responds according to the relaxation timescale. This implies that, on a monthly scale, the largest impact should occur in the month with the largest bias, no matter for how many months relaxation has been active before that month (as long as it's at least one month).

Thanks for raising this point. We agree that the relaxation timescale sets the degree of observational constraint as expected and that the largest impact occurs in the month with the largest bias. The reviewer could be right on the last point. But we can only confirm that statement after conducting assimilation experiments with each month observationally constraint as if the observations were only available for that particular month. Indeed, this is something we want to experiment in the future. So we would prefer to keep the word "probably". [L380, L367-L369]

L210: "with a slight clockwise displacement" - you could mention that this is consistent with the mean climatological Arctic drift pattern (transpolar drift, Beaufort gyre) and thus likely a consequence of the mean advection.

Thanks, we add this point. [L372-L374]

L217-218: "In November [...] the SIT constraint has very little impact on SIC biases" - Could the reason be that (in addition to the fact that no SIT corrections are applied in the previous months) the thickness corrections made in Nov need more time to influence the sea-ice concentration, because that requires a "cross-impact" through other processes (dynamics and thermodynamics)?

As explained in our response to Reviewer 1's related comment starting with "Finally, a few minor comments", we now explain the seasonal cycle of the differences in SIC bias better in the revised manuscript. Firstly, there is no SIT nudging happening in the preceding months. Secondly, the negligible changes in SIC bias between ORA-REF and ORA-SIT is suggestive of fast growth processes in the forward model which is faster than the timescales intrinsic to the SIC assimilation. We provide this explanation in the revised manuscript. [L359-L361, L383-L391]

L225: "large positive increments from May to October, indicative of strong underestimation of SIC in the ORAs" - To be precise, should "in the ORAs" rather be "in the (hypothetical) forced model without SIC assimilation"? After all, the SIC assimilation makes sure that the SIC underestimation doesn't get too strong.

Thanks for the suggestion. Indeed, the assimilation of SIC reduces the errors in concentration, that would be otherwise larger. We modify the sentence as " …indicative of the strong underestimation of SIC in the forward model…". [L403-L404]

L232-235: Isn't the even bigger difference in the SIC increments after May (even though these are for the worse) even more strongly showing the long-lasting impact of the SIT corrections on the SIC assimilation?

The reviewer is right.

L243: "low bias" could be mistaken for "negative bias", maybe better say "weak bias" or "small bias" or similar

That is true, thanks, it is changed to "small bias". [L433]

L250-262: To compute the IIEE, do you use the ensemble-median ice edge (50%- contour of sea-ice probability where SIC=15% is used to determine "presence" or "absence" of sea-ice in each ensemble member) or do you compute it for each member individually and average the IIEEs afterwards? That would make a difference, so this should be specified. Related, note that there's a probabilistic version of the IIEE ("Spatial Probability Score", Goessling and Jung 2018 "A probabilistic verification score for contours: Methodology and application to Arctic iceâĂ˘ Redge forecasts") that you ˘ could apply to your ensemble forecasts directly, which would have the advantage that changes in uncertainty/reliability would be captured, too.

For simplicity, we compute the IIEE for the ice edge of the ensemble mean. Thanks for the suggestion on the SPS metric. We appreciate that different possibilities of computing IIEE give different results, and the SPS again can give a different result. However, in light of the large differences between the forecasts in the present study, the differences are probably small. We test these other approaches and document in the revised paper whether they would lead to noticeable differences in the figures. We regret to note an error in the markings of significant changes in the original Figure 6 in the submitted version and we replace it with the corrected Figure 6 which doesn't change the results qualitatively. [L452-L455]

Fig. 6 caption and throughout the paper: DelSole and Tippett (2016) just apply the sign test (a special case of the binomial test with p=0.5), only that they visualize how the outcome develops from forecast case to case like a random walk. I would recommend to refer to the test simply as the sign test (which in fact dates back to 1710!).

We thank the reviewer for pointing us to the historical roots of this test. We would like to keep citing DelSole and Tippet (2016) as the most recent and most relevant piece of work in applying and refining this long-known test for the field of climate and weather forecasts. We follow the reviewer's suggestion to refer to it as the sign test (also in Figures 7, 11 and 12 captions) and modify the text as "...sign test as recommended by DelSole and Tippet (2016)". [P28, P29, P41, P44]

Sect. 3.2 and Fig. 7: 1) Can you please explain how the bias correction is performed? Is this simply done for each gridcell individually? Do you just subtract the mean concentration bias (difference as a function of time of the year and lead time, averaged over 2011-2016/17), possibly with a correction that makes sure concentration values remain bound between 0 and 1? Or is quantile normalization involved? 2) Related to the bias correction, I would find it very useful if the forecast errors could be compared against a climatological benchmark forecast. The latter could be based simply on the same period (2011-2016/17), or on the preceding decade (to make it more "operational"). I would expect that the uncalibrated forecasts are worse than climatology for most lead times (except the first one or two months?), but the calibrated might beat the climatology for a few months? In the summary section you say very clearly that you are "yet to demonstrate the benefit of interannual varying data on bias-corrected forecast scores", but I think it would be rather easy and revealing to add a climatological reference (even if it reveals clear limitations of current sea-ice forecast skills).

1) We perform a simple bias correction like so: for each combination of grid cell, start date and forecast lead time, we compute the mean forecast error over all forecasts, which is then subtracted from the "raw" forecast value in order to obtain the bias-corrected forecast value. We do not clip the bias-corrected forecast values to make sure they are between 0 and 1. Although this should be done when issuing forecasts, Johnson et al. (2018) have shown that it makes negligible difference for forecast skill assessment. [L475-L478]

2) We appreciate the comparison to a climatological reference forecast is an interesting point, and we include the climatological reference in Figure 7. However, we do not plan to dwell on this point, since this is not the main point of the work, and including further discussion on the performance of climatology would distract the readers from the main point: to determine whether initialization with CS2SMOS improves or deteriorates the forecast. [L478-L483, L499-L504, P29]

Fig. 8, top: Can you provide an explanation why the SIV in the ORAs converge from May to September, so that the large SIT difference in spring is completely "forgotten", whereas the coupled forecasts maintain *much of the initial offset? Is there some fundamental reason why the forced (vs. coupled) atmosphere would cause such a difference, or can it be linked to the continued assimilation of SIC (or ocean variables)?*

As the reviewer suspects, the explanation for the different behavior of the ORAs and the forecasts is that the ORAs are constrained by the same SIC (but no SIT) information in summer, which leads to the convergence of the sea ice state in the ORAs during that time of the year. This effect can also be appreciated from Figure 4, which shows that the SIC assimilation increments of ORA-SIT in summer are much more positive than those of ORA-REF, suggesting that SIC assimilation needs to work harder to keep ORA-SIT on track (compared to ORA-REF), but overall acts to bring the ORAs closer together in the absence of ORA-SIT information.

In contrast, the coupled forecasts do not have a similar constraint and thus tend to keep the offset in the initial conditions throughout the forecast. However, Figure 8a shows that FC-SIT and FC-REF tend to converge as well, although much more slowly than the ORAs. This is mentioned in the revised text. [L528-L533]

Eq. 2: The way the melt energy tendency is defined, is seems to be really just the derivative of (area-averaged) SIT (times a constant factor), right? Also, maybe it's better to use partial d's to make clear that these are not material (Lagrangian) derivatives. Related, you could also mention that changes in SIT through divergence as well as advection are included, implying that the "melt energy tendency" can in principle also change through dynamics. I understand that, by averaging over a large area (almost the whole Arctic), most of any dynamical effects would be compensating each other, but being clear about this would be good.

Yes, MET is simply proportional to temporal changes in effective sea ice thickness. As the reviewer points out, MET can also change dynamically through lateral ice transports, but here we average over all ocean north of 70N, which should be a sufficiently large area to average out this effect and should mainly leave thermodynamic effects as the drivers of MET. We clarify this point and also use the more appropriate partial d's in the revised manuscript as the reviewer suggested. [L555, L559-L562]

L314-316 & Fig. 9: The plot caption reveals that for the forecasts you look only at the first-month MET, but you do not mention/explain/motivate this in the text. Further, do I understand correctly that, by considering just the first month of the respective forecasts instead of a "closed" seasonal cycle, the annual integral of MET is not expected to be zero (while it should be zero for the ORAs)? In fact it looks a bit like it's rather negative (average build-up of sea-ice volume), can you confirm this?

In Figure 9 we want to isolate the changes in MET when switching from forced (=analysis) to coupled (= forecast) mode. To avoid seeing mainly the effect of feedbacks arising from the model sea ice state drifting away from the analyzed state (most notably the ice-albedo feedback), we decided to compile the annual cycle of forecasted MET from lead-month 1 data. This motivation is clarified in the revised manuscript.

The reviewer is also right that the average of the MET annual cycle in the ORAs is close to zero (in fact about ~+0.3 Wm-2, in agreement with the long-term sea ice melt), while it is ~-4.8Wm-2 in FC-REF. The negative value suggests that lead-month 1 forecasts on average produce too much ice in winter and melt too little ice in summer. This point is noted in the revised manuscript. [L565-L570, L572-L576]

Fig. 10 and corresponding text: I am wondering to what extent turbulent fluxes (in particular sensible) could also play a role, for example, with stronger downward spring/summer sensible heat fluxes in the forced ORAs compared to the coupled forecasts (acknowledging that there might not be a corresponding observational dataset to compare against). Too high near-surface temperatures that could generate too strong downward sensible heat flux would be consistent with a positive downwelling longwave bias in ERA-I, even if clouds also seem to play a role there. If differences in turbulent fluxes are too small to be important, please mention that.

Sea ice and near-surface air temperature are close to 0°C during the melting season in both our reanalyses and forecasts. Because of this weak vertical temperature gradient sensible heat fluxes will be generally small over sea ice in summer. Nevertheless, to be sure, we checked mean difference in sensible heat fluxes in ERA-Interim and the forecasts and differences over sea ice were uniformly small (<2Wm-2 for May and July averages), confirming that differences in this field cannot explain the found differences in MET. A short note on this is added to the revised manuscript. [L616-L619]

L351-352: "Significant cold biases are present in forecasts for most of the start months and lead months" - Is this also true over Arctic sea ice in winter? If so, how can it be reconciled with Batrak and Müller (2019) "On the warm bias in atmospheric reanalyses induced by the missing snow over Arctic sea-ice"? I thought that the surface coupling is similar in the system studied here?

Yes, cold biases in near-surface temperature (T2m) are present in the forecasts over Arctic sea ice in winter when considering ERA5 reanalysis as the truth. Batrak and Mueller's findings on warm biases in sea-ice temperature in a group of atmospheric reanalyses (including ERA5) without realistic representations of snow over sea ice, and sea ice thickness, is based on verification against observations, and reanalysis products. The reviewer is right that using atmospheric or ocean reanalysis for verification of Arctic surface temperature can be misleading, since there is large uncertainty in them, as Batrak and Mueller 2019 show in their Figure 3. Verifying against observations is not easy, since due to the scarcity of observational campaigns over sea ice, the verification will have large representativeness error, and definitively not suitable for seasonal forecasts. So, while it is clear that assimilation of SIT has a sizeable and significant impact on T2m forecasts via SIC forecasts, we do not have enough information to assess if this contributes to the reduction of the mean error in T2m. We modify the manuscript along these lines. [L648-L657]

Fig. 12: I was a few times slightly confused when looking at this figure, intuitively thinking that the lower panels show differences between FC-SIT and FC-REF that could be directly combined with the biases shown in the upper panels. But the lower panels show the differences in mean absolute error, which is alright but easily misleading. I suggest to use a different colourbar for the lower panels so that the different flavours of "temperature" (signed vs. unsigned) is more intuitively reflected.

We have made bold figure labels (‘t2m bias’, ‘t2m diff in MAE’) next to the top and bottom colour panels respectively and also made the figure caption clearer. [P44]

Response to Minor comments & corrections

L25: last -> lasts

thanks, it is done. [L52]

L80: "as cross-check variables evaluation" - I recommend to reformulate.

Thanks. As both the reviewers pointed out, it is rephrased as “Although the datasets used for verification are not strictly independent, evaluation using those datasets is relevant as it allows cross-checking between variables, for instance between SIC and SIT assimilation. ” [L144-L147]

L91: These -> This

thanks, the grammar is corrected. [L162]

L168: "differ on" -> "differ in" / "differ regarding"

thanks, it is replaced with “differ in”. [L302]

L208: "gradients on" -> "gradient in the" or "gradients of the"

thanks, it is replaced with “gradients in the”. [L367]

L212: "end of melt season" -> add "the"

thanks, it is done. [L376]

L217: "reduced up to" -> "reduced by up to"

thanks, it is done. [L383]

L227: "indicates" -> "indicate"

thanks, it is done. [L407]

L228: "at marginal seas" -> "in the marginal seas"

thanks, it is done. [L409]

L232: "on an average" -> "on average"

sorry, “on an average” is correct. [L414]

L232-233: "in ORA-SIT analysis" -> add "the"

thanks, it is done. [L415]

L265: "that to be" -> add "are"

thanks, it is done. [L472]

L288: "is smaller" -> "are smaller"

thanks, it is done. [L535]

There are a few more such tiny things, please check carefully!

thanks, we check the revised manuscript and correct where necessary. [L14, L437, L683, L720]

References

Pohl, C., Istomina, L., Tietsche, S., Jaekel, E., Stapf, J., Spreen, G., and Heygster, G.: Broad Band Albedo of Arctic Sea Ice from MERIS Optical Data. The Cryosphere, 14, 165-182, 2020.

Hogan, R., Ahlgrimm, M., Balsamo, G., Beljaars, A., Berrisford, P., Bozzo, A., pe, F. D. G., Forbes, R., Haiden, T., Lang, S., Mayer, M., Polichtchouk, I., Sandu, I., Vitart, F., and Wedi, N.: Radiation in numerical weather prediction, ECMWF Technical Memorandum, https://doi.org/10.21957/2bd5dkj8x, https://www.ecmwf.int/node/17771, 2017.

Balan-Sarojini, B., Tietsche, S., Mayer, M., 465 Balmaseda, M., and Zuo, H.: Towards improved sea ice initialization and forecasting with the IFS, https://doi.org/10.21957/mt6m6rpwt, https://www.ecmwf.int/node/18918, 2019.

Zampieri, L., Goessling, H. F., and Jung, T.: Bright prospects for Arctic sea ice prediction on subseasonal time scales, Geophysical Research Letters, 45, 9731–9738, 2018.

Goessling, H. F. and Jung, T.: A probabilistic verification score for contours: Methodology and application to Arctic ice-edge forecasts, Quarterly Journal of the Royal Meteorological Society, 144, 735–743, https://doi.org/10.1002/qj.3242, URL http://doi.wiley.com/10.1002/qj.3242, 2018.

Year-round Impact of Winter Sea Ice Thickness Observations on Seasonal Forecasts

Beena Balan-Sarojini[1], Steffen Tietsche[1], Michael Mayer[1,2], Magdalena Balmaseda[1], Hao Zuo[1], Patricia de Rosnay[1], Tim Stockdale[1], and Frederic Vitart[1]

[1] *The European Centre for Medium-Range Weather Forecasts, , Shinfield Rd, Reading RG2 9AX, UK*

[2] *Department of Meteorology and Geophysics, , University of Vienna, Vienna, Austria*

Monday 2nd November, 2020

Abstract

Nowadays many seasonal forecasting centres provide dynamical predictions of sea ice. While initializing sea ice by assimilating sea ice concentration (SIC) is common, constraining initial conditions of sea ice thickness (SIT) is only ~~at~~ in its early stages. Here, we make use of the availability of Arctic-wide winter SIT observations covering 2011-2016 to constrain SIT in the ECMWF (European Centre for Medium-Range Weather Forecasts) ocean–sea-ice analysis system with the aim of improving the initial conditions of the coupled forecasts. The impact of the improved initialization on the predictive skill of ~~Arctic~~ pan-Arctic sea ice for lead times of up to 7 months is investigated in a low-resolution analogue of the currently operational ECMWF seasonal forecasting system SEAS5.

By using winter SIT information merged from CS2 and SMOS (CS2SMOS: CryoSat2 Soil Moisture and Ocean Salinity), substantial changes of sea ice volume and thickness are found in the ocean–sea-ice analysis, including damping of the overly strong seasonal cycle of sea ice volume. Compared with the reference experiment, which does not use SIT information, forecasts initialized using SIT data show a reduction of the excess sea ice bias and an overall reduction of seasonal sea ice area forecast errors of up to 5% at lead months 2 to 5. Change in biases is the main forecast impact. Using the Integrated Ice Edge Error (IIEE) metric, we find significant improvement of up to 28% in the September sea ice edge forecast started from April. However, sea ice forecasts for September started in spring still exhibit a positive sea ice bias, which points to too slow melting in the forecast model. A slight degradation in skill is found in the early freezing

season sea ice forecasts initialized in July and August, which is related to degraded initial conditions during these months. Both the ocean reanalyses, with and without SIT constraint, show strong melting in the middle of the melt season compared to the forecasts. This excessive melting related to positive net surface radiation biases in the atmospheric flux forcing of the ocean reanalyses remains and consequently degrades analysed summer SIC. The impact of thickness initialization is also visible in the sea surface and near-surface temperature forecasts. While positive forecast impact is seen in near-surface temperature forecasts of early freezing season (Sep-Oct-Nov) initialized in May (when the sea ice initial conditions have been observationally constrained in the preceding winter months), negative impact is seen for the same season when initialised in August month when the sea ice initial conditions are degraded. We conclude that the strong thinning by CS2SMOS initialization mitigates or enhances seasonally dependent forecast model errors in sea ice and near-surface temperatures in all seasons.

The results indicate that the memory of SIT in the spring initial conditions ~~last~~ lasts into autumn, influencing forecasts of the peak summer melt and early freezing seasons. Our results demonstrate the usefulness of new sea ice observational products in both data assimilation and forecasting systems, and strongly suggest that better initialization of SIT is crucial for improving seasonal sea ice forecasts.

1 Introduction

Sea ice is an integral part of the Earth system as it regulates the heat, moisture and momentum flux exchange between the polar oceans and the atmosphere. Decline in Arctic sea ice is a visible indicator of the changing climate. Forecasting Arctic sea ice has advanced significantly in the last decade, with most forecasting centres using prognostic sea ice models operationally, allowing us to explore the sea ice forecast skill on long lead times from weeks to months to seasons. Possibilities of economically viable shorter shipping routes across the Arctic in the summer are constantly being explored. Monthly and seasonal outlooks of sea ice products are therefore in great demand especially by the Arctic communities, maritime and resource extraction industries.

Moreover, there is increasing scientific evidence that warming and sea ice loss in the Arctic due to climate change affect the European weather and climate (Balmaseda et al. (2010), Mori et al. (2014), Overland et al. (2016), Ruggieri et al. (2016)). Unlike sea ice concentration and extent, long records of satellite observations of sea ice thickness are sorely lacking (Laxon et al. (2003), Kwok and Rothrock (2009), Haas et al. (2010), Meier et al. (2014), Sallila et al. (2019), Scarlat et al. (2020)).

Since reliable estimates of long-term, basin-wide sea ice extent and volume are needed for understanding climate change and for initializing numerical weather forecasts, there is growing interest in using improved and new types of sea ice observations in data assimilation systems (Lindsay et al. (2008), Blanchard-Wrigglesworth et al. (2011), Tietsche et al. (2013), Sigmond et al.

(2013), Balmaseda et al. (2015)). Earlier studies propose that long-term memory in the winter sea ice thickness can potentially improve summer sea ice extent forecasts (Guemas et al. (2016), Tietsche et al. (2014), Day et al. (2014)). They concluded that potential predictability mainly originates from the persistence or advection of sea ice thickness anomalies, interaction with ocean and atmosphere and changes in the radiative forcing.

While assimilation of sea ice concentration (SIC) is routinely done in operational sea ice forecasting, assimilation of sea ice thickness (SIT) is at its early stages (Allard et al. (2018), Xie et al. (2018), Mu et al. (2018), Fritzner et al. (2019)). These studies have found that SIT initialization improves sea ice forecasts in forced ocean–sea-ice forecasting systems which were run for short time periods spanning from 3 months up to 3 years. Blockley and Peterson (2018) reported for the first time the positive impact of winter SIT initialization on the skill of seasonal forecasts for summer sea ice forecasts using a fully-coupled atmosphere–ocean–sea-ice model. All of these studies used either European Space Agency's Cryosat-2 (CS2) radar altimeter freeboard SIT measurements alone (Laxon et al. (2013), Hendricks et al. (2016)) or merged with SMOS radiometric measurements (Kaleschke et al. (2012),Tian-Kunze et al. (2014)) in a dataset called CS2SMOS (Ricker et al. (2017)).

Currently SIC is the only sea ice variable assimilated in the ECMWF ocean-sea–ice data assimilation system. Although the ECMWF sea ice reanalysis and reforecasts compare well with other systems (Chevallier et al. (2017), ~~Uotila et al. (2018) , Zampieri et al. (2018)~~ Uotila et al. (2019) , Zampieri et al. (2018) , Zampieri et al. (2019)), they are affected by noticeable errors (Tietsche et al. (2018)). There are large biases in sea ice forecasts from months to seasons, pointing to uncertainties in both the models and observations used in the assimilation and forecasting systems. Here we explore the pathway to improve the initialization using observations of sea ice thickness which covers both the thick and thin ice regions of the Arctic. We then assess the impact of the changed sea ice initial condition on the forecast skill on long lead times of months to seasons. Compared to Blockley and Peterson (2018), who looked only at summer forecast skills, our study for the first time assesses the forecast impact of SIT initialization on all seasons using a fully-coupled seasonal forecasting system. We use the ECMWF coupled ensemble seasonal forecasting system SEAS5 and CS2SMOS thickness observations.

Our study takes a forecasting system end-to-end perspective, from observations, modelling to forecast products. The rest of the article is organised as follows. Section 2 describes the methodology of sea ice thickness initialization and forecasting, including a brief description of ocean–sea-ice models, the assimilation system, the atmosphere-ocean–sea-ice coupled forecasting system, observations used and the experimental set-up. Section 3 presents the main results ~~:~~ and has three main foci: i) assessing the impact of new SIT observations on the analysed sea ice state and the impact of the changed sea ice initialization on seasonal range ~~forecasts~~sea-ice forecasts (sections 3.1 and 3.2), ii) improving Arctic sea-ice forecast skill by understanding the errors in the coupled forecast model and the data assimilation system through targeted diagnostics (sections 3.3), and iii) quantifying the impact of sea-ice improvements on seasonal forecasts of atmospheric variables (section 3.4). Finally, Section 4 provides the summary of the findings with concluding remarks.

2 Observations and Methods

The procedure followed here to assess the impact of SIT information follows a twin experiment approach. Each of the experiments consists of two distinctive steps: 1) the production of a set of ocean and sea ice initial conditions by conducting twin ocean–sea-ice assimilation experiments (ocean–sea-ice reanalyses; abbreviated as ORA), which only differ in the use of SIT information ; and 2) the production of a set of twin retrospective seasonal forecast (reforecast) experiments, initialized from the respective ORA. The ORA twin reanalyses are a low resolution variant of the currently operational ORAS5 (Zuo et al. (2019)). The seasonal forecast experiments are also low resolution versions of the operational ECMWF seasonal forecasting system SEAS5 (Stockdale et al. (2018), Johnson et al. (2018)). The impact of SIT in the ocean initial conditions and seasonal forecast is then evaluated, using verification against observational datasets and other more specific diagnostics. The verification will also use fields from ORAS5 and ERA-5 (ECMWF atmospheric Re-Analysis-5); Hersbach et al. 2019) reanalyses. Although the datasets used for verification are not strictly independent, ~~it~~ evaluation using those datasets is relevant as ~~cross-check variablesevaluation~~it allows cross-checking between variables, for instance between SIC and SIT assimilation. SIT verification using CS2SMOS dataset is also conducted as a sanity check of the nudging approach~~compared to the reference experiment~~: if the approach works, the difference with respect to CS2SMOS should be smaller in ORA-SIT than in ORA-REF. In this section we first describe the sea ice information used for both initialization and verification, and then offer a brief description of the experimental set-up.

In addition to the sea ice data sets described below, the initialization step uses ocean observations: sea surface temperature, sea-level anomalies from altimeter and in-situ temperature and salinity, which are the same as those used in ORAS5, as described in Zuo et al. (2019).

2.1 Sea Ice Observational ~~information~~Information

2.1.1 Sea Ice Concentration Product: OSI-401-b

The two ocean–sea-ice reanalysis experiments presented here assimilate the sea ice concentration product of the EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSI SAF, www.osi-saf.org; product identifier OSI-401-b (Tonboe et al. (2017))). ~~These~~ The Level-3 OSI SAF SIC product (OSI-401-b) is produced as daily-mean fields with only a few hours latency. In contrast to the operational ORAS5 system, which uses Level-4 SIC data, experiments presented in this study use Level-3 SIC data. The main difference is that Level-4 products rely on gap-filling, whereas Level-3 products have missing data, for instance if the satellite has a temporary malfunction, or if certain areas like the North Pole are not observed. The OSI-401-b SIC observational estimate is based on SSMIS

(Special Sensor Microwave Imager / Sounder) measurements. SIC is provided as the percentage of an ocean grid point covered by sea ice. The product comes in a polar stereographic grid of 10km horizontal resolution with varying pole hole size.

The impact of Level-3 SIC observations in the initialization is reported to have neutral forecast impact on seasonal sea ice forecasts and positive impact on sub-seasonal range (Balan-Sarojini et al. (2019)). The OSISAF OSI-401-b SIC data set is also used for verification of SIC and sea ice edge.

2.1.2 Sea Ice Thickness ~~product~~Product: CS2SMOS

A recent initiative led by the Alfred Wegener Institute provides a merged product of Arctic-wide winter ice thickness that combines thick-ice retrievals by CryoSat2 (CS2) satellite and thin-ice retrievals by the Soil Moisture and Ocean Salinity (SMOS) satellite. This merged sea ice thickness observational product, CS2SMOS (https://spaces.awi.de/display/CS2SMOS, Ricker et al. (2017)), is the first ever multi-sensor ice thickness product for the Arctic. CS2 (Hendricks et al. (2016)) measures freeboard (the height of the ice or snow surface above the water level) using altimetry, whereas SMOS (Tian-Kunze et al. (2014)) measures brightness temperatures in the L-band microwave frequencies. Both measurements are converted to ice thickness in metres. Due to their different measurement principles, SMOS retrievals should be reliable for ice thinner than about 1 m and CS2 retrievals for ice thicker than 1 m. The merged product can hence represent the entire thickness range covering the whole Arctic with reasonable accuracy (Ricker et al. (2017)). CS2 and SMOS are merged using an optimal interpolation scheme to produce the CS2SMOS product, which is available on a weekly basis on an Equal-Area Scalable Earth Grid version 2 (EASE2) grid with ~~25km horizontal resolution~~ 25 km horizontal resolution covering all regions in the Northern Hemisphere where sea ice can be expected. Both the CS2 and SMOS retrievals are not possible in the melt season due to signal contamination owing to the presence of melt ponds, and wet and warm snow and ice surfaces~~.~~ ~~It~~ , therefore it is only available for 5 full months from November to March of the ice growth season every year.

In a merged product like CS2SMOS it is difficult to appropriately represent observational uncertainties. For instance, sensor-specific errors could affect regional sea ice thickness: over multi-year thick ice in the Canadian Basin, errors associated with Cryosat-2 retrievals dominate, whereas in the Bering or Okhotsk Sea with mostly seasonal thin ice, errors associated with SMOS retrievals dominate. As reported in Ricker et al. (2017) , the relative error is maximum in the thickness range of 0.5-1.0 m in the merged product, where relative uncertainty is high for both CS2 and SMOS.

The CS2SMOS SIT information without observational uncertainties has been assimilated in one of the twin ORA experiments, during the November-March period. It has also been used for verification of initialization in those months. We emphasize that this dataset does not provide SIT information during the period April–October. Nevertheless, there is still substantial impact in the April–October period from constraining sea ice thickness during the

November–March period, as we will see in Section 3 – a truly year-round impact.

2.2 Methods

2.2.1 ~~Ocean–sea-ice~~ Ocean–Sea-Ice Reanalysis Experiments

In order to assess the impact of new sea ice thickness observations on the assimilation, we carry out two ORAs as shown in Table 1. They are 1) a reference experiment with SIC assimilation (ORA-REF), and 2) an experiment with SIC assimilation and sea ice thickness constraint (ORA-SIT). Experiments ORA-REF and ORA-SIT are run for the time period January 2011 to December 2016, because these are the full years for which CS2SMOS observations were available at the time of experimentation. Note that ORA-REF is a continuation of a longer experiment which started in 2005 and ORA-SIT starts from ORA-REF on the 1st of January, 2011.

Experiment name	SIC constraint	SIT constraint	Time period	Description
ORA-REF	Yes	No	2011-2016	SIC assimilation
ORA-SIT	Yes	Yes	2011-2016	SIC assimilation and SIT nudging

Table 1: Specifications of the ocean–sea-ice assimilation experiments.

Our reanalysis experiments are forced by near-surface air temperature, humidity and winds as well as surface radiative fluxes from the atmospheric reanalysis ERA-Interim (ERA-I) (Dee et al. (2011)) until 2015 and from the ECMWF operational analysis from 2015 to 2016. We use the same set-up of NEMOVAR (Variational data assimilation system for NEMO (Nucleus for European Modelling of the Ocean) ocean model) used in ORAS5 (Zuo et al. (2019)) – in particular, almost the same observations are assimilated. The only differences are the following: a) a coarser model resolution as described below, b) different assimilated SIC observations compared to the current operational one and, c) a longer assimilation window of 10 days instead of 5 days.

The ocean general circulation model used in these experiments is NEMO version 3.4 (Madec (2008)) with a horizontal resolution of approximately 1° and 42 vertical layers. The grid is tripolar, with the poles over Northern Canada, Central Asia and Antarctica enabling higher resolution across the Arctic than at the equator. The first model layer is 10 m thick, and the upper 25 levels represent approximately the top 880 m. Both the horizontal and vertical resolution in our setup is lower than that of the current operational system, which has a horizontal resolution of approximately 0.25° and 75 vertical levels. The time step is one hour.

The prognostic thermodynamic-dynamic sea ice model used is LIM2 (Louvain-la-Neuve Sea Ice Model) in its original version (Fichefet and Maqueda (1997)). The vertical growth and decay of ice due to thermodynamic processes is modelled according to the three-layer (one layer for snow and two layers for ice) Semtner scheme (Semtner (1976)). The ice velocity is calculated from a momentum balance considering sea ice as a two-dimensional continuum in dynamical interaction with the atmosphere and ocean. Internal stress within the ice for different states of deformation is computed following the viscous-plastic (VP) rheology proposed by Hibler III (1979). LIM2 has a single sea ice category to represent sub-grid scale ice thickness distribution, and open water areas like leads and polynyas are represented using ice concentration. Melt ponds are not modelled ~~.~~ which could affect the accurate representation of surface albedo over sea-ice. However, we note that only the ocean reanalysis ORAS5 actually makes use of the albedo computed by LIM2 (which is too high in summer), while the atmospheric reanalyses used for verification and the forecasting system use the same climatological albedo (based on SHEBA campaign observations; Beesley et al. (2000)). Moreover, a recent comparison study (Pohl et al. (2020)) shows that, overall, the broadband albedo over Arctic sea-ice derived from MERIS observations is comparable to that in the ERA5 atmospheric reanalysis in terms of the seasonal cycle on large spatial scales. The forecast albedo over ice is comparable to that in ERA-5 and ERA-Interim atmospheric reanalyses. LIM2 has a time step of 1 hour and is coupled to the ocean at every time step.

As for ORAS5, both experiments here use the variational data assimilation using NEMOVAR in a 3D-Var FGAT (First Guess at Appropriate Time) configuration as described in Mogensen et al. (2012). The length of the assimilation window is 10 days in our experiments. Assimilated observations comprise temperature and salinity profiles, altimeter-derived sea level anomalies and sea ice concentration. Sea-surface temperature is constrained to observations by a strong relaxation. A global freshwater correction is added to reproduce the observed global-mean sea-level change. The assimilation of the SIC is done separately from the ocean variables, and is described in Tietsche et al. (2015) and Zuo et al. (2017).

In addition to the observations assimilated via NEMOVAR, the SIT in experiment ORA-SIT is constrained to the CS2SMOS via a linear nudging technique (Tietsche et al. (2013), Tang et al. (2013)). The relationship between the modelled and observed sea ice thickness in a grid point is described by the following equation:

$$SIT^n = SIT^m - [\frac{\Delta t}{\tau}(SIT^m - SIT^o)] \quad (1)$$

where SIT^n is the nudged thickness, SIT^m is the modelled floe thickness, SIT^o is the observed floe thickness, Δt is the sea ice model time step of 1 hour, and τ is the nudging coefficient corresponding to a relaxation time scale of 10 days. The choice of a 10-day relaxation time scale makes sense as a first trial, since it is consistent with the length of the assimilation window. To facilitate the nudging, the CS2SMOS weekly observations in EASE2 grid have been interpolated to daily gridded fields in ORCA 1° grid. The weekly to daily interpolation is done by appropriately weighting two adjacent weekly-mean fields. We have also tested the sensitivity to different nudging strengths by running variants of ORA-SIT with a relaxation time scale of 20, 30 and 60 days. By construction, as the relaxation time scale increases from 10 days to 60 days, SIT is less constrained to CS2SMOS. In this study, ~~but in this study~~we only use the experiment with the strongest constraint (10-day relaxation time) for initializing the ensemble reforecasts, because this time scale fits with the length of the assimilation window, and we aimed for a strong observational constraint in order to obtain a strong forecast impact.

2.2.2 Coupled Reforecast Experiments

In order to assess the impact of CS2SMOS sea ice thickness initialization on sea ice forecasts, we performed 2 sets of twin coupled ocean–sea-ice-atmosphere reforecast experiments as shown in Table 2, which only differ ~~on~~ in the ocean–sea-ice initial conditions, provided by the data assimilation experiments shown in Table 1. The reference reforecast (FC-REF) is initialized by ORA-REF, and reforecast experiment FC-SIT is initiailized by ORA-SIT. Comparison of results from these two sets of reforecasts allows quantifying the impact of SIT information on the seasonal forecasts.

Experiment name	Start years	Lead mon	Ens. size	Initial condition	Description
FC-REF	2011–2016	7	25	ORA-REF	SIC initialization
FC-SIT	2011–2016	7	25	ORA-SIT	SIC and SIT initialization

Table 2: Overview of the reforecast experiments. For each of the start years, forecasts are started on the 1st of every calendar month.

The reforecast experiments are carried out using a version of the ECMWF coupled seasonal forecasting system. The coupled model consists of the same ocean and sea ice model (NEMO3.4/LIM2) used for our reanalysis experiments, and is coupled to the ECMWF atmospheric model, Integrated Forecast System (IFS) version 43r3. It is run with a horizontal resolution of 36 km, corresponding to a cubic octahedral reduced Gaussian grid at truncation TCo319 and 91 vertical levels (SEAS5 is run with IFS cycle 43r1 at the same atmospheric resolution, but with 0.25° horizontal resolution and 75 vertical levels in the ocean). The coupled model also includes the land surface model HTESSEL (Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land) and the ocean surface wave model WAM. The coupling of the atmosphere and ocean is done using a Gaussian interpolation method, and the coupling frequency is 1 hour. For more details on SEAS5 see (Stockdale et al. (2018), Johnson et al. (2018)).

Both reforecasts are started from the 1st of each month of each year 2011–2016, resulting in 72 forecast start dates overall. Note that out of all months of each year in the 2011-2016 period only winter (December-April) months are directly constrained by November-March observations as the CS2SMOS data is only available for those 5 full months. The initial conditions for the remaining 7 start months (May-November) of each year are indirectly affected by the thickness constraint applied earlier in the ice growth season in the reanalysis. The ~~non-availability of the observations for the melt season in a way provides an opportunity to test the predictability of winter SIT from summer initial conditions. The~~ forecast initialized from each start date has 25 ensemble members for both sets of reforecasts.

3 Results

Here we first assess the impact of sea ice thickness observations on the estimation of sea ice properties in the ORA initial conditions, and then we evaluate the impact on the skill of seasonal forecast of sea ice area, sea ice edge, sea ice volume and 2m temperature. When possible, we use the observational datasets for verification. However, as mentioned above, sea ice thickness and volume (SIV) can not be verified properly for the months April-October, due to the lack of sea ice thickness observations. In those cases, we will describe the impact in terms of differences between experiments. We use the term pan-Arctic to refer to all regions of the Northern Hemisphere that are potentially covered by sea ice.

3.1 Impact of Sea Ice Thickness Initialization on the Sea Ice Reanalysis

Figure 1 shows the SIT bias with respect to the CS2SMOS observations for ORA-REF (Figure 1a, c) and ORA-SIT (Figure 1b, d), for March (Figure 1a, b) and November (Figure 1c, d). The ORA-REF suffers from large ice thickness bias of up to 1.4 m. The predominant bias pattern is an underestimation of ice thickness by more than 1 m in the central Arctic, and an overestimation in the Beaufort Gyre and the Canadian Archipelago of the order of 1 m. This pattern is present for all the months when CS2SMOS is available. In March, widespread overestimation in the coastal Arctic seas is also present. These biases are much reduced or absent in ORA-SIT. Most of the large-scale pattern of underestimation and overestimation of sea ice in ORA-REF is not present in ORA-SIT in March. However, slight underestimation over the central Arctic and overestimation over the Canadian Archipelago still remain in November. This is ~~probably~~ caused by the lack of SIT observations during the months preceeding November. In contrast, the estimation of the March conditions benefit from the availability of SIT information in the preceeding winter. We note that the bias in ORA-SIT over the Laptev, East Siberian and Chukchi Seas is very small, about 0.1 to 0.05 m of magnitude (below the contour interval).

Figure 2 shows the difference in SIT between ORA-SIT and ORA-REF for March, July, September and November. The difference patterns between ORA-SIT and ORA-REF are quite consistent for all the months, characterized by a thickening of the thick ice over the Central Arctic and North of Greenland, and a thinning of the thin ice area over the Beaufort and Siberian Seas, thus enhancing the spatial gradients ~~on~~ in the sea ice thickness distribution. The largest impact occurs in March, probably because at this month the SIT observations have been assimilated during the preceeding 5 months. The impact of SIT winter information lasts well into the summer months, with a slight clockwise displacement of the thinning, and a reduction of the thickening, which by September has roughly halved. The shift in the thinning pattern is consistent with the mean climatological transpolar Arctic drift pattern and is thus likely a consequence of the mean advection. The impact during March and November is consistent with a reduction of the bias in ORA-REF (Figure 1a and c). Since

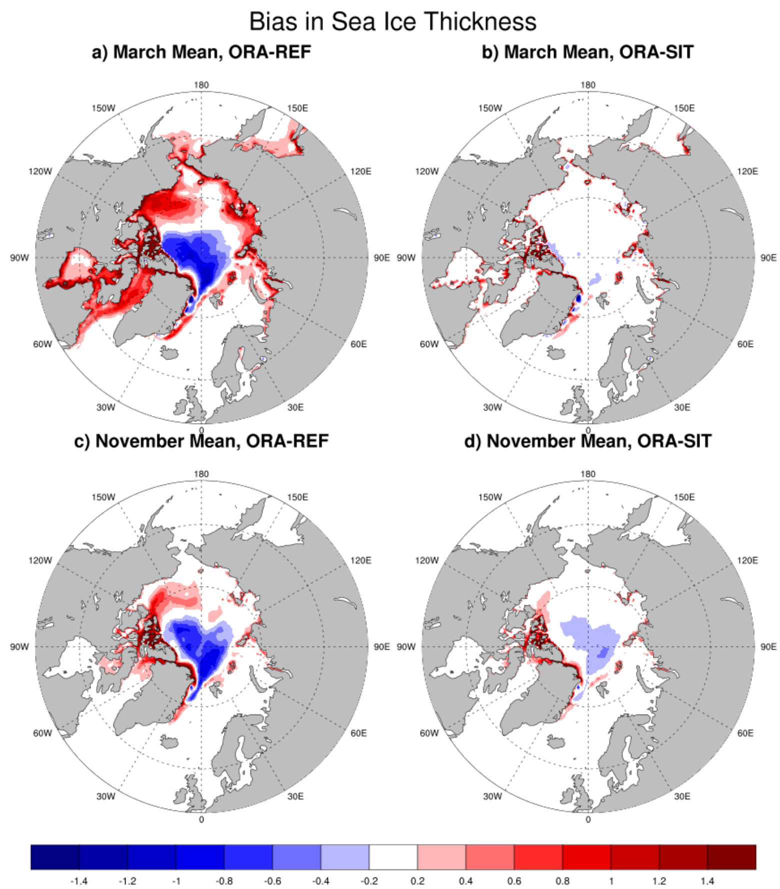

Figure 1: Bias in monthly-mean (2011-2016) sea ice thickness (m) in experiment a) ORA-REF and b) ORA-SIT, for March (a, b) and November (c, d). The reference is CS2SMOS observations. ORA-REF is the ocean–sea-ice assimilation experiment with no sea ice thickness constraint. ORA-SIT is the assimilation experiment with a thickness relaxation time scale of 10 days.

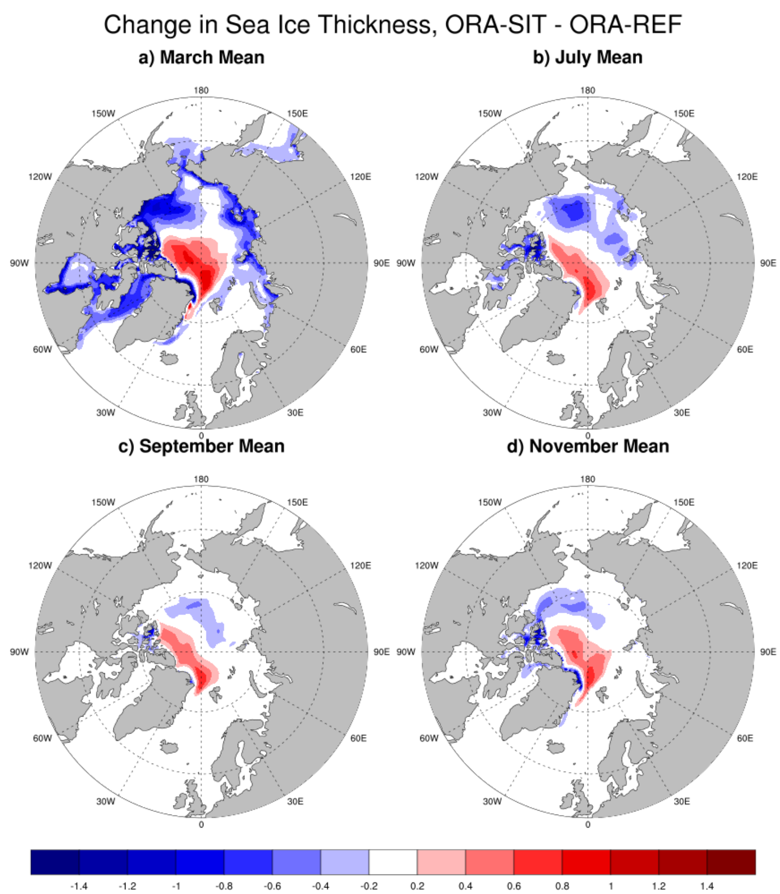

Figure 2: Difference in monthly-mean (2011-2016) sea ice thickness (m) between experiments ORA-SIT and ORA-REF for a) March and b) July and for c) September and d) November months.

basin-scale SIT observations are not available for the end of the melt season, biases are unknown.

The thickness constraint also affects the biases in SIC. Figure 3 shows the SIC bias w.r.t. OSI-401-b SIC as well as the SIC difference between ORA-REF and ORA-SIT. In March, the month of sea ice maximum, ORA-REF shows mostly an overestimation of SIC all around the sea ice edge, over the Davis Strait, northeast of Greenland, Bering Sea and Okhotsk Sea. In ORA-SIT this bias is uniformly reduced by up to 10% . In November (Figure 3g, h and i), when the sea ice edge is expanding with newly frozen ice, ORA-REF has similar SIC overestimation biases over the ~~sea~~ ice edge, but this time the SIT constraint has very little impact on SIC biases. This is because of no SIT nudging happening in the preceding months. Also, the very small changes in SIC bias between ORA-REF and ORA-SIT over Chukchi and East Siberian Sea regions of negligible ice thickness bias in ORA-SIT (Figure 1d) is suggestive of fast growth processes in the forward model which is faster than the timescales intrinsic to the SIC assimilation. The ORA-REF biases in July are characterized by a weak underestimation of SIC. Notably, in ORA-SIT there is an increase of the negative SIC bias of more than 10% over the Pacific and Siberian Arctic sectors towards the end of melt season, with July and August (not shown) months being the most affected.

To gain some insight into the degradation of the July SIC bias in ORA-SIT we look at the mean annual cycle of the SIC assimilation increments. The assimilation increments are indicative of the corrections that the assimilation of SIC

Bias and ~~change~~ Change in Sea Ice Concentration

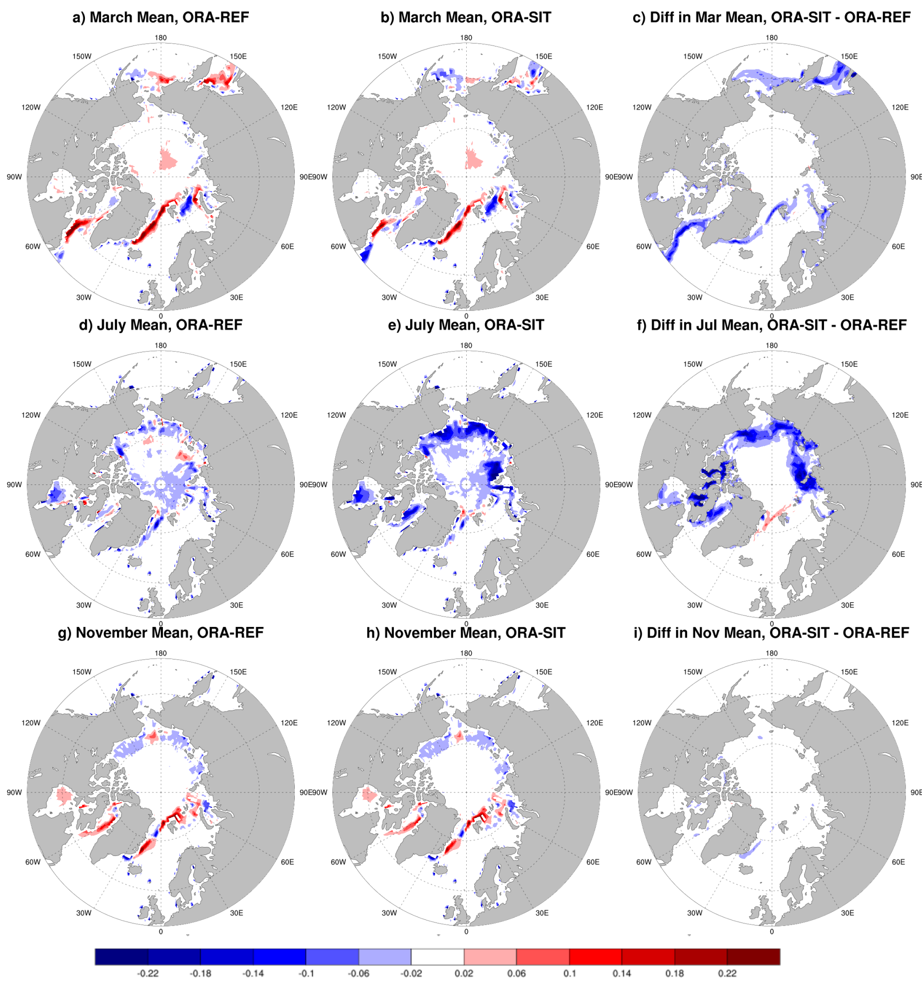

Figure 3: Bias in monthly-mean (2011-2016) sea ice concentration w.r.t. OSI-401-b observations for ORA-REF (a, d, g), ORA-SIT (b, e, h), and the difference between ORA-SIT and ORA-REF for (c, f, i). Panels (a, b, c) are for March, (d, e, f) for July, and (g, h, i) for November.

observations exerts to compensate for errors in the sea ice state. Figure 4 shows the mean annual cycle of the area-averaged assimilation increments in ORA-REF (blue) and ORA-SIT (green). In both experiments, the assimilation increments exhibit a clear seasonal cycle, with large positive increments from May to October, indicative of strong underestimation of SIC in the ~~ORAs~~forward model, and weak negative increments from December to March. The differences in SIC increments over the Arctic between the two experiments peaks during July, with ORA-SIT increments about 9% per month higher than in ORA-REF. The results in this figure ~~indicates~~ indicate that 1) both ORAs melt sea ice too fast during the summer months, as shown by negative SIC biases ~~at~~ in the marginal seas of the Arctic Ocean where thin sea ice resides during the summer months (Figure 3d and e) ; and 2) the SIT assimilation exacerbates the summer SIC biases in ORA-SIT (as seen in eg: Figure 3e) due to corrected but thinner sea ice at the begining of the melt season in almost all marginal seas of the Arctic Ocean (Figure 2a).

From January to May and from November to December, on an average less ice is being taken away by the increments in the ORA-SIT (green) analysis than that in ORA-REF (Figure 4). These results clearly show the long-lasting effect of the SIT information: the SIT constraint was only applied during the growth season from November to March (grey shading), but its impact, whether positive or negative, is evident in sea ice concentration throughout the melting season even in the presence of SIC assimilation.

Annual ~~cycle~~ Cycle of the ~~mean~~ Mean of Sea Ice Concentration

~~increments~~Increments

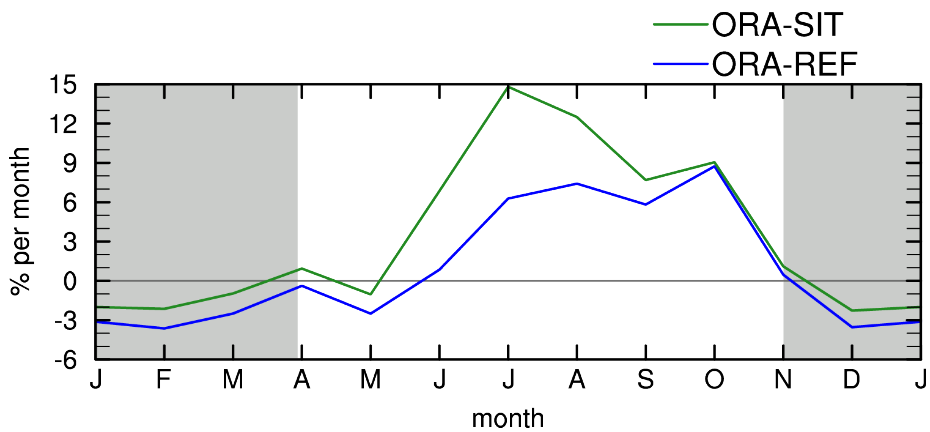

Figure 4: Annual cycle of the mean of the SIC increments in ORA-SIT (green), and ORA-REF (blue), averaged over north of $70°N$ during 2011-2016. The grey shading shows months (January to March, and November to December) with CS2SMOS SIT nudging.

Bias in the Sea Ice Area ~~forecasts~~Forecasts

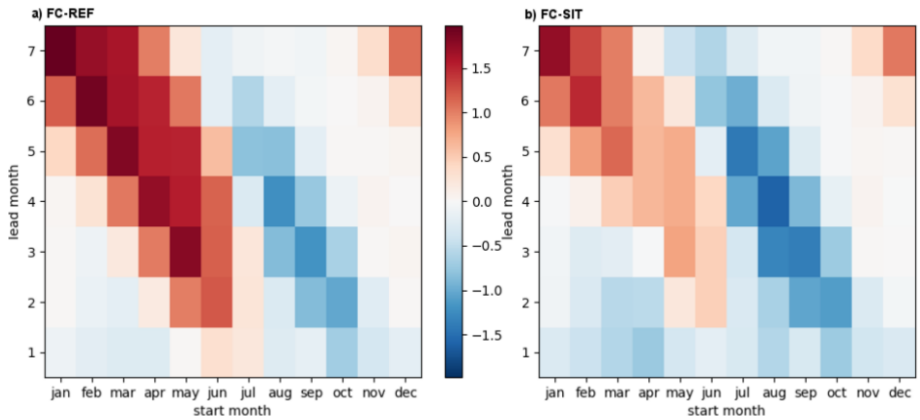

Figure 5: Bias in the forecast of ~~Arctic~~ pan-Arctic sea ice area (~~$\times 10^{12}$m^2~~$\times 10^6$km^2) w.r.t. ORAS5 as a function of start and lead month for 2011–2016, a) in the reference reforecast FC-REF and b) in the SIT-initialised reforecast FC-SIT. Red colour denotes over-prediction of sea ice area, and blue colour denotes under-prediction.

3.2 Impact of ~~ice thickness initialization~~ Ice Thickness Initialization on ~~sea ice forecasts~~Sea Ice Forecasts

Figure 5a gives an overview of bias in sea ice area in the FC-REF reforecast w.r.t. ORAS5 reanalysis as a function of forecast start and lead months~~(units are 10^6 km^2)~~. ORAS5 is preferred to OSISAF for the verification of integrated sea ice area because of its complete spatial coverage. The figure shows the forecast bias for different forecast lead times (y-axis) as a function of forecast starting month (x-axis). Errors at lead month 1 are generally small throughout the year. However, for longer lead times, there is a strong over-prediction of sea ice area in summer months, and a moderate under-prediction of autumn sea ice conditions, consistent with too slow melting and refreeze respectively. The forecast biases are generally small in winter months.

These three bias regimes, in general – ~~low~~ small bias in winter, positive bias in summer and negative bias in autumn – seem to be mostly independent of start months. These biases shown in FC-REF are quite similar to those in SEAS5 (not shown) which are discussed in more detail in Stockdale et al. (2018). The positive biases in the melt season forecasts ~~is~~ are considerably reduced with the SIT initialisation in FC-SIT started in January to June and the negative biases in the forecasts is worsened in FC-SIT started in July to October (Figure 5b). The forecasts for winter months remain unbiased in FC-SIT. Note that the bias regimes in the forecasts are very different from the bias regimes in the reanalysis (Section 3.1), which tends to have too much ice in winter and too little ice in summer.

Impact of thickness initialization has not only improved the biases in summer SIC forecasts, but it has also improved the sea ice ~~extent~~ edge forecasts as measured by the Integrated Ice Edge Error (IIEE) (Figure 6). Seasonal forecasts of ice edge are in great demand for exploring economically viable Arctic shipping routes. IIEE is one of the recent user-relevant sea ice metrics on ice extent or ice edge (Goessling et al. (2016), Bunzel et al. (2017)). Since it can be decomposed into over- and under-prediction, it is more useful than the traditional basin-wide sea ice extent error. ~~It is calculated using uncalibrated~~

For simplicity, we assess ice edge forecasts by using the deterministic IIEE metric calculated from the ice edge of the ensemble mean SIC forecasts. We have also tested probabilistic metrics like the Spatial Probability Score suggested by Goessling and Jung (2018) and found that they give very similar results.

IIEE for all lead months and start months verified against OSI-401-b suggests reduced error in sea ice edge (blue colours) in FC-SIT overall. The most striking feature is the significant improvement in summer forecast error for lead months 2–7 in FC-SIT compared to FC-REF. The main contribution to the error reduction of up to 30% in summer forecasts comes from the reduction of the model bias leading to consistent over-prediction (not shown). For the traditional September sea ice extent forecast starting in April, an improvement of 28% is found. Forecast verification in October and November from July and August starts show a slight degradation, caused by under-prediction (not shown). This could again be due to the indirect effect of a thinner starting point in FC-SIT (Figure 2b) and a lower, degraded SIC in the ORA-SIT reanalysis (Figure 3e), combined with the already existing slow refreeze nature of the model.

Figures 5 and 6 point out that the impact of ice thickness initialization on the forecast bias and errors is strongly dependent on season and lead time. Most seasons and lead times are improved but some are, perhaps inevitably, deteriorated. To measure the overall impact on forecast error and make a statement about potential skill improvements that are to be expected for operational forecasts, we aggregate FC-SIT and FC-REF for all start months from January 2011 to December 2016 and compute the area-integrated mean absolute forecast error (MAE) of sea ice concentration for each lead month. In order to obtain

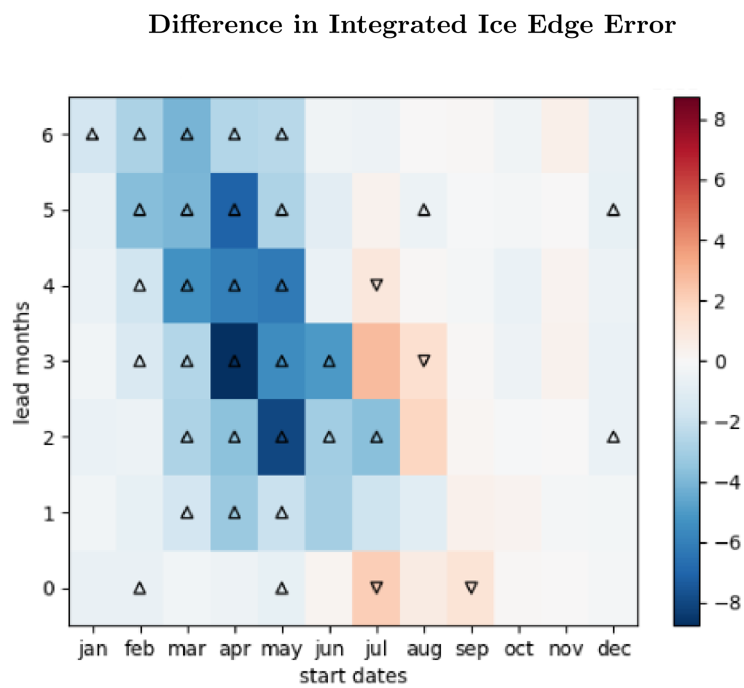

Figure 6: Difference in Integrated Ice Edge Error in ~~10^{11} m^2~~ 10^5 km^2 between FC-SIT and FC-REF reforecasts 2011–2016 w.r.t. OSI-401-b observations. Blue colour denotes reduced error in sea ice edge in FC-SIT and red colour denotes increased error in FC-SIT. Black triangles represent ~~statistically significant (DelSole and Tippett (2016)) changes~~ statistical significance at the 5% level ~~.~~according to the sign test (DelSole and Tippett, 2016) .

Mean Absolute Error in ~~Integrated~~ SIC ~~forecasts~~Forecasts

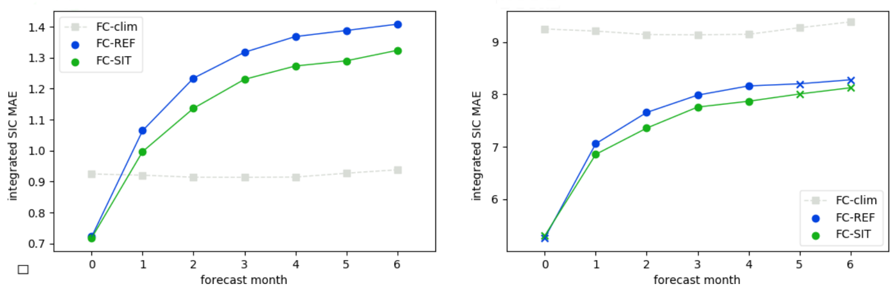

Figure 7: Spatially integrated SIC mean absolute error over lead month for all FC-REF and FC-SIT forecasts (~~no.~~ 72 forecasts each first of ~~forecast start months, $n = 72$;~~ the month from January 2011 to December 2016) w.r.t OSI-401-b observations. Panel a) shows the error in $10^6\,\mathrm{km}^2$ without bias correction, panel b) the error in $10^5\,\mathrm{km}^2$ after bias correction. Lead months for which the reduction of forecast error in FC-SIT passes ~~the DelSole and Tippett (2016)~~ a statistical significance test at the 5% level ((DelSole and Tippett, 2016)) are marked by filled circles, insignificant changes are marked by crosses. The error of a simple climatological reference forecast is also shown as FC-clim.

the bias-corrected forecast value, for each combination of grid cell, start date and forecast lead time, we compute the mean forecast error over all forecasts, and then subtract it from the "raw" forecast value. Comparison against a climatological benchmark forecast is a very useful background information for evaluating the predictive skill of ensemble forecasting systems (e.g. Zampieri et al. (2018)). The climatological reference forecast for a given target month and year is constructed by using the verification data valid for the same calendar month but different years from the range of target months considered (January 2011 to June 2017).

Averaged over all start dates and grid points, Figure 7 shows that the MAE of sea ice area is substantially improved in FC-SIT. When no bias correction is applied prior to computing the MAE (Figure 7a), FC-SIT forecasts are significantly better in each lead month, with maximum error reduction of about 10%.

However, skill assessments of seasonal forecasts are conventionally made after a forecast calibration where the impact of the forecast bias is removed. By this measure, a reduction of forecast bias does not by itself count as an improvement. As Figure 7b shows, removing the respective bias of FC-SIT and FC-REF prior to computing the MAE results in a smaller error reduction: errors in FC-SIT are significantly lower only in lead months 2–5, by up to 5%. Figure 7 demonstrates that, although the thickness initialization predominantly reduces the bias, it also leads to an improvement in the skill of sea ice area forecasts that is relevant for operational forecasting systems.

The importance of forecast biases is illustrated by benchmarking the errors of the dynamical forecasting system against a simple statistical reference forecast: Figure 7 also shows the errors of a climatological reference forecast (FC-clim). Without bias correction, errors of both FC-REF and FC-SIT are larger than those from FC-clim already after one month, while after bias correction, both FC-REF and FC-SIT have lower errors than FC-clim for all lead months.

Finally, we analyse the impact of SIT initialization on forecasts of pan-Arctic sea ice volume. Though an integrated quantity like pan-Arctic sea ice volume is a result of many dynamic and thermodynamic sea-ice processes and lacks regional details, it is a key indicator for understanding of the Arctic energy cycle, an important process that needs to be realistically represented in reanalyses and seasonal forecasts. It is useful to compare the contrasting SIV seasonal cycles in coupled and uncoupled mode, and with/without SIT observational constraint in the initialization, since this helps to identify the origin of errors in the systems in the specific operational set up. Figure 8 shows the sea ice volume forecast climate at different lead month for the forecasts starting in May (top) and August (bottom). The forecast climate is computed by averaging the reforecast starting at a given calendar month for the years 2011-2015. Seven months forecasts started in August lead to February of the following year. Since the ORAs are not available in January and February, 2017, the year 2016 is not accounted for in this figure. For reference, the sea ice volume estimates of ORA-REF and ORA-SIT reanalyses are also shown. It is remarkable that the shape of the seasonal cycle is largely preserved between FC-REF and FC-SIT,

Time ~~evolution~~ Evolution of ~~mean~~ Mean Sea Ice Volume ~~forecasts~~Forecasts

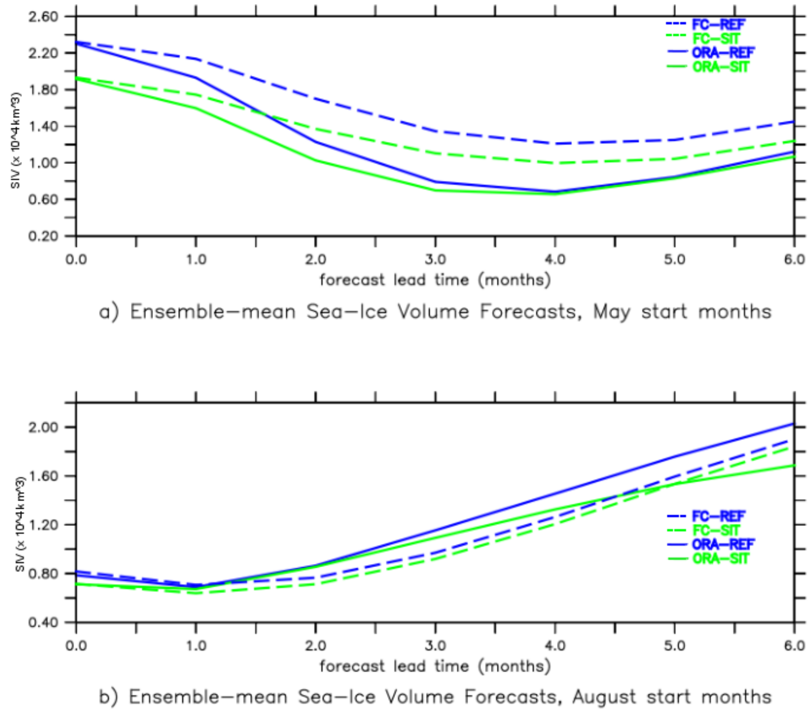

Figure 8: Time series of ensemble-mean sea ice volume (units are ~~10^{12} m^3~~10^4 km^3) forecasts averaged over 2011–2015, for May start date (a) and August start date (b) in reference reforecast (FC-REF, dashed blue line) and reforecast with thickness initialization (FC-SIT, dashed green line) compared to their own reanalyses, ORA-REF (solid blue line), and ORA-SIT (solid green line).

maintaining the initial offset during the whole forecast range. The figure clearly shows that FC-SIT starts from a thinner ice state than FC-REF in both initial months.

The May starts show large differences between the forecasts and the ORAs: Both FC-SIT and FC-REF show a slower SIV decrease (lower melt rate) than the ORAs from June to September, and also a slower refreeze during October and November. The explanation for the different behavior of the ORAs and the forecasts is that the ORAs are constrained by the same SIC (but not the same SIT) information in summer, which leads to the convergence of the sea ice state in the ORAs during that time of the year (also seen in Figure 4). In the coupled forecasts, there is no similar constraint and they tend to converge slower than the ORAs. The melt rate of the ORAs here are consistent with those in ORAS5 (see ~~Uotila et al. (2018)~~ Uotila et al. (2019) or Mayer et al. (2019)). Compared to the May starts, differences between FC-SIT and FC-REF ~~is~~ are smaller in the August ~~start~~starts, and so is their agreement with the ORAs. Again, the FC-SIT shows smaller values than FC-REF from the begining, and both forecast sets exhibit a parallel SIV evolution. The shape of the seasonal cycle in the forecasts is different from the ORAs; the forecasts initialized in August show a slower refreeze during October than the ORAs. However, after October, the SIV increases faster in the forecasts than in ORA-SIT, and it continues increasing more or less at the same rate until the end of January in the forecasts, while in ORA-SIT (solid green line) the freezing rate slows down after November. As a result by the end of January the forecast SIV is higher than in ORA-SIT. ORA-REF without the thickness constraint has the highest SIV in the ice growth season. In the next section we examine the discrepancies in SIV changes between ORAs and FCs in more detail.

3.3 Linking ~~sea ice analysis~~ Sea Ice Analysis and ~~forecast errors~~ Forecast Errors to the Arctic ~~surface energy budget~~Surface Energy Budget

In order to investigate the physical causes of sea ice errors in the ORAs and forecasts, the Arctic surface energy budget is considered. We estimate melt energy tendency (MET), which is the energy used to melt sea ice and energy released in the process of freezing, and is proportional to SIV changes. It is defined as in Mayer et al. (2019):

$$MET = L_f \rho (\cancel{\frac{dSIT}{dt}} \frac{\partial SIT}{\partial t}) \tag{2}$$

where L_f denotes latent heat of fusion (-0.3337x10^6 J kg^{-1}), ρ represents sea ice density (assumed constant at 928 kg m^{-3}), and SIT, the grid-point averaged sea ice thickness. Thickness changes are computed as exact monthly differences. MET can also change dynamically through lateral ice transports, but here we average over the ocean area north of $70°N$, which should be a sufficiently large area to average out any dynamical effects and should mainly leave thermodynamic effects as the drivers of MET. Figure 9 shows the MET mean annual cycle (2011-2015) north of $70°N$ for ORA-REF, ORA-SIT, FC-REF, and FC-SIT. ~~The values for the forecasts are compiled from one-month~~

~~forecasts from every calendar month~~In order to isolate the changes in MET when switching from forced ORA mode to coupled forecast mode and to avoid seeing mainly the effect of feedbacks arising from the model sea ice state drifting away from the analyzed state (most notably the ice-albedo feedback), we compile the annual cycle of forecasted MET from lead-month 1 data from each start date. Assimilation increments of SIC proportionally affect SIV in the ORAs (Tietsche et al. (2013), Tietsche et al. (2015)). The resulting MET increments are shown for both ORA-REF and ORA-SIT as well. We note that the MET annual cycle of ORA-REF is very similar to that of ORAS5 (not shown) and that the average of the MET annual cycle in the ORAs is close to zero (in fact about +0.3 W/m^2 (Mayer et al. (2016) , Mayer et al. (2019)), in agreement with the long-term sea ice melt), while it is -4.8 W/m^2 in FC-REF.

Figure 9 clearly shows that ORA-REF exhibits the most pronounced annual cycle of MET, with strongest melting in summer and strongest freezing in winter. Earlier studies have shown that the MET annual cycle is exaggerated in ORAS5 (Uotila et al. 2019; Mayer et al. 2019) and hence also in ORA-REF. ORA-SIT has a damped MET annual cycle, as the thickness constraint during winter prevents overly strong SIV accumulation. Lower SIV at the end of winter consequently leads to weaker melting in summer. However, summer melt in ORA-SIT is likely still too strong, as this experiment features a negative SIC bias in summer despite realistic SIT and SIC earlier in the year, when CS2SMOS data is available (see Figure 3e).

Both FC-REF and FC-SIT agree very well with each other and exhibit a

Mean ~~annual cycle~~ Annual Cycle of ~~melting energy~~Melting Energy

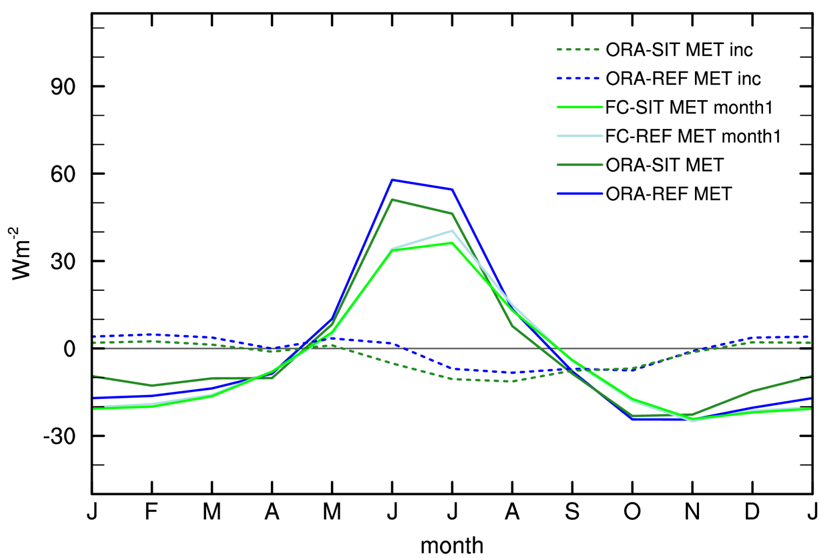

Figure 9: Mean annual cycle of MET over ocean area north of $70°N$ in ORA-REF, ORA-SIT, FC-REF (lead month 1), FC-SIT (lead month 1). MET increments for ORA-REF and ORA-SIT are shown as well.

Mean ~~annual cycle~~ Annual Cycle of ~~surface net radiation~~Surface Net Radiation

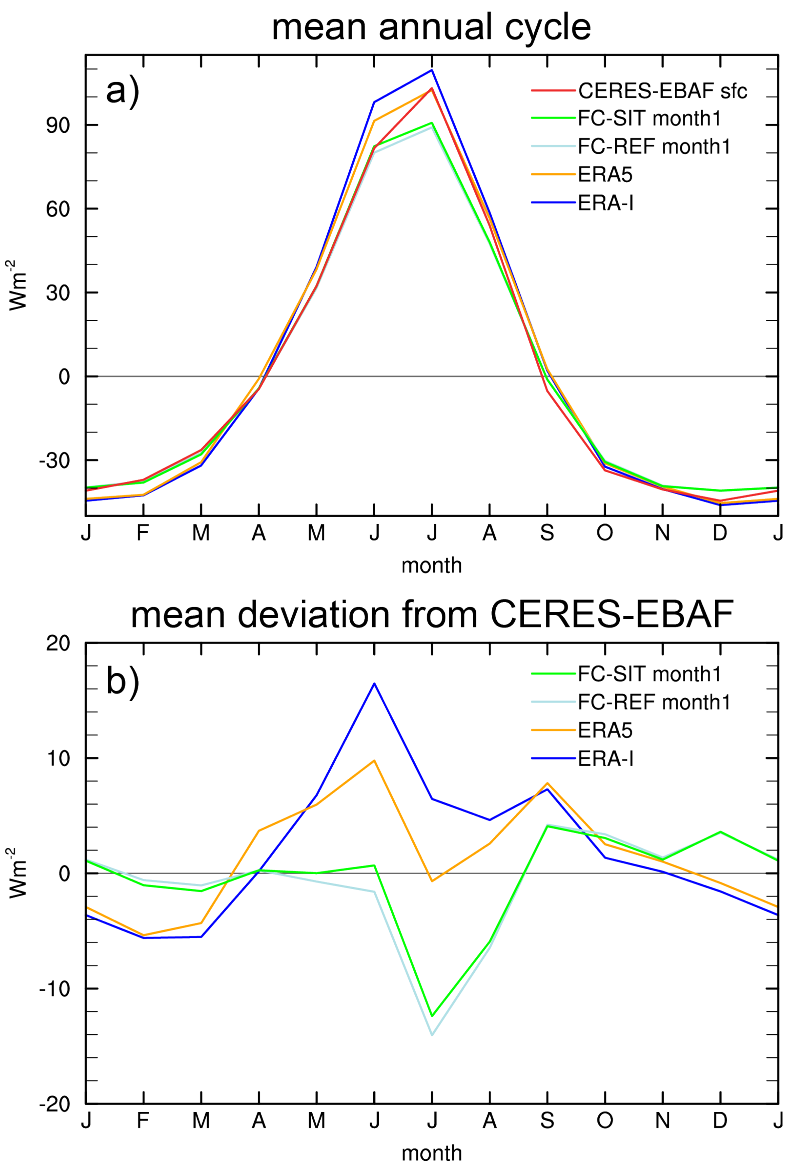

Figure 10: a) Mean annual cycle of surface net radiation, Rad_S (W/m^2) over ocean area north of $70°N$ from ERA-I, ERA5, FC-REF (lead month 1), FC-SIT (lead month 1), and CERES-EBAF, and b) Mean deviation of Rad_S from CERES-EBAF for FC-REF, FC-SIT, ERA-I and ERA5.

much weaker MET annual cycle than the ORAs (Figure 9). The difference between the forecasts and the ORAs in May and June melting cannot be explained by the MET increments (neutral impact at that time), which points to differences in energy fluxes into the sea ice as a cause.

We therefore compare the mean annual cycle of surface net radiation (Rad_S) over ocean north of $70°N$. Figure 10a shows the 2011-2015 annual cycle of Rad_S from FC-REF, FC-SIT, ERA-I, ERA5, and the satellite-based product Clouds and Earth's Radiant System – Energy-Balanced and Filled Surface edition 4.0 (CERES-EBAF; Kato et al. (2018)), which we use as reference.

We consider Rad_S from ERA-I as a good proxy for Rad_S seen by the ORAs, due to two reasons: 1) ORAs use ERA-I forcing during most of the study period, and 2) ORAs does not output Rad_S term; although it is not exactly identical e.g. due to different albedo in the ORAs. ERA-I exhibits a positive Rad_S bias in summer, peaking in June at 15 W/m^2, while FC-REF and FC-SIT agree quite well with CERES-EBAF, especially in May and June, when MET discrepancies with the ORAs are large (Figure 9). Thus the Rad_S bias of ERA-I can explain a large fraction of the overly strong MET in the ORAs during May and June, and the discrepancy between the ORAs and the forecasts.

The mean deviation of Rad_S from CERES-EBAF (Figure 10b) clearly indicates that forecasts are closer to the observational product than the atmospheric reanalyses in May and June. This positive Rad_S bias of ERA-I should be considered alongside the results by Hogan et al. (2017), who found a negative bias in downwelling shortwave radiation in ERA-I due to excessive low-level clouds.

Our results can be explained by the positive bias in downwelling longwave radiation in ERA-I outweighing the shortwave flux bias. Figure 10 also shows results for ERA5, which is closer to CERES-EBAF than ERA-I, which indicates a reduced cloud bias in this more recent atmospheric reanalysis and gives rise to the expectation of improved MET in future ocean reanalyses forced by this product. We also note that the mean difference in sensible heat fluxes in ERA-Interim and the forecasts and differences over sea ice were uniformly small (generally <2 W/m^2 in summer; not shown), confirming that differences in this field cannot explain the found differences in MET.

Additional information on the realism of summer MET in the forecasts can be obtained from the sea ice area forecast bias of FC-SIT, as displayed in Figure 5b. It shows that FC-SIT May starts exhibit a strongly reduced positive bias compared to FC-REF. The bias reduction can be attributed to the improved initial conditions in FC-SIT, but the fact that the sea ice area bias remains positive from July onward indicates that MET in the forecasts is too low in summer. Figure 10b suggests that Rad_S is too low in the forecasts in July and August, which probably contributes to the positive SIA bias remaining in FC-SIT (Figure 5b).

The October MET (Figure 9) indicates stronger refreeze in the ORAs (lower MET values) compared to the forecasts. This is consistent with negative MET increments present in the ORAs, which act to speed up refreeze in the reanalyses (see Figure 9). The less negative MET values of the forecasts in October are consistent with the too weak freezing and consequent underestimation of sea ice in autumn in the August starts.

Area-averaged net radiation of all considered products agrees well with CERES-EBAF in October (see Figure 10), and also difference maps show only a weakly positive Rad_S bias of the reanalyses and forecasts compared to CERES-EBAF (not shown). Hence, errors in other physical terms such as ocean-ice fluxes must play an important role in fall, but more detailed investigations are beyond the scope of this paper.

3.4 Impact of ~~ice thickness initialization~~ Ice Thickness Initialization on ~~forecasts~~ Forecasts of ~~atmospheric variables~~Atmospheric Variables

To discuss the impact of the sea ice thickness constraint on the atmosphere, we first assess the differences in the forecast means (or biases) between FC-SIT and FC-REF. Figure 11a shows the bias in 2m temperature (t2m) (averaged over $50 - 90°N$) in FC-REF as a function of start dates and lead months. ~~Significant~~ When verified against ERA5, significant cold biases are present in forecasts for most of the start months and lead months except for non-significant warm biases in November forecasts started in August, September and October months. We note that using atmospheric or ocean reanalysis without realistic representation of snow over sea ice, and sea ice thickness, for the verification of pan-Arctic surface temperature can be misleading, since there is large uncertainty associated with these products (Batrak and Müller (2019)). Verifying against observations is not easy, since due to the scarcity of observational

Difference in ~~mean t2m~~ Mean T2m and ~~mean sea level pressure forecasts~~Mean Sea Level Pressure Forecasts

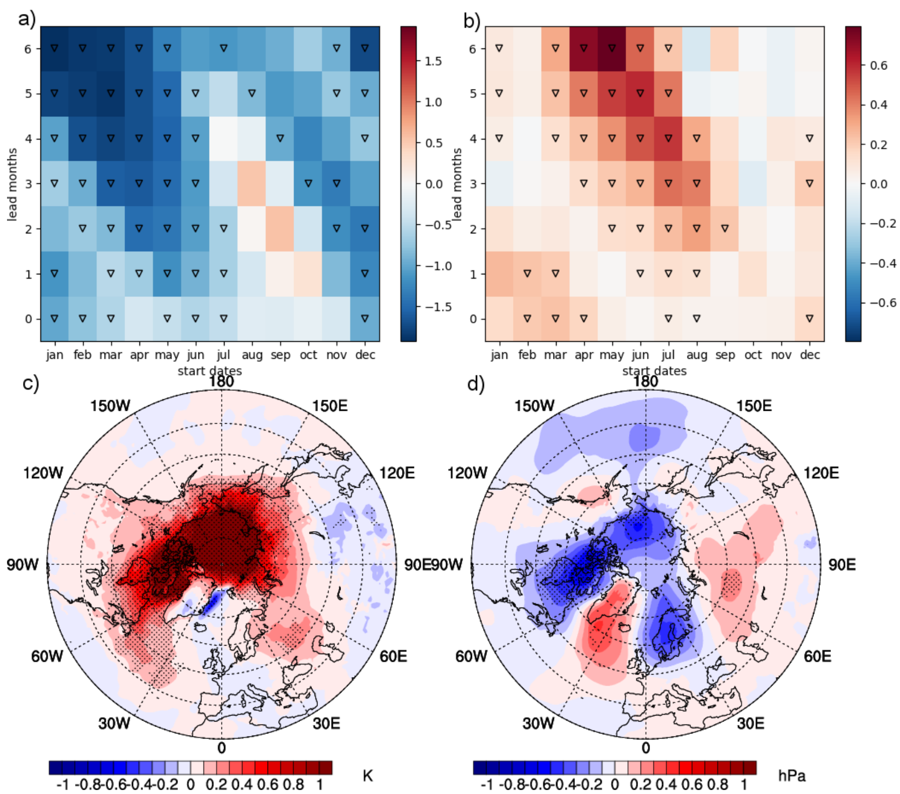

Figure 11: Mean forecast differences between FC-SIT and FC-REF 2011-2016: a) bias in mean 2m temperature (in K) north of 50°N w.r.t. ERA5, as a function of start dates and lead months, in FC-REF, b) similar to a), but difference in mean 2m temperature (in K) between FC-SIT and FC-REF. Triangles denote significant changes according to ~~DelSole and Tippett (2016)~~ the sign test as recommended by DelSole and Tippett (2016) at the 5% level. Mean forecast difference (FC-SIT - FC-REF) for SON aggregated from May, June, July, August start dates of c) 2m temperature and d) mean sea level pressure. Dots indicate areas of significant changes on the 95% level according to Komolgorov-Smirnov test.

campaigns over sea ice, the verification will have large representativeness error, and hence is not suitable for seasonal forecasts verification. Mean differences in t2m (Figure 11b) are generally positive with very few and non-significant exceptions, which is expected from the generally reduced sea ice cover in FC-SIT. Strongest warming with area averages of 0.5K can be found during fall for forecasts started between March and September. February and March start dates show a moderate but significant warming at short lead times, but otherwise changes are relatively small for October to February start dates. Also, changes in summer temperatures are small compared to those in fall. Inspection of temperature difference patterns between FC-SIT and FC-REF indicates that differences in summer are confined to areas around the sea ice edge (not shown), while changes in fall are more widespread (see Figure 11c). The warming pattern in fall appears as a diagonal feature in Figure 11b, which suggests that changes depend more on season than on forecast lead time. Therefore, to gain more insight into the spatial structure of the changes, Figure 11c and d show forecast differences in 2m temperature and mean sea level pressure in SON, respectively. To find robust changes, the differences are aggregated from forecasts started between May and August, yielding samples of 600 forecasts. Moreover, aggregation along the diagonal maximizes the signal (compare to Figure 11b).

Widespread temperature differences >1K can be seen over the Arctic Ocean and the Canadian Achipelago in SON (Figure 11c), but significant warming spreads also south to North America and Eurasia. Solar radiation in the Arctic is very weak for SON. Hence, the warming in FC-SIT must stem from enhanced fluxes of heat from the ocean to the atmosphere, with a possible positive feedback from increased atmospheric water vapour. The fluxes are enhanced in FC-SIT due to larger areas of open waters and increased SSTs, both a result of reduced sea ice concentration. Furthermore, we find warming over the Northwest ~~Atantic~~Atlantic, which is related to the warmer SSTs present already in the initial conditions from ORA-SIT (not shown). Another area of significant warming in FC-SIT relative to FC-REF can be seen over Eastern Europe and Western Russia. This warming seems consistent with patterns of mean sea level pressure differences shown in Figure 11d. They show lower pressure in FC-SIT over Scandinavia and higher pressure over central Russia, which together suggest more southerly winds in the region of warmer temperatures. Furthermore, mean sea level pressure changes indicate lower pressure over the Arctic Ocean and the Canadian Archipelago, i.e. in areas of maximum warming. In addition, there are positive pressure differences southeast of Greenland. Altogether, the patterns in sea level pressure difference resemble a wave-like response, but it should be kept in mind that only some parts of these changes are statistically significant. Nevertheless, we note that qualitatively similar and significant changes are also found in 500hPa geopotential forecasts for SON (not shown), suggesting that the features seen in Figure 11d are indeed robust.

We now turn to the question whether changes in the forecast mean constitute a forecast improvement or a forecast deterioration in the sense that they lead to an overall reduction of model biases. Since forecast bias is strongly dependent on region, season and lead time, aggregating over many seasons and lead months

Bias and ~~difference~~ Difference in MAE in ~~t2m forecasts~~T2m Forecasts

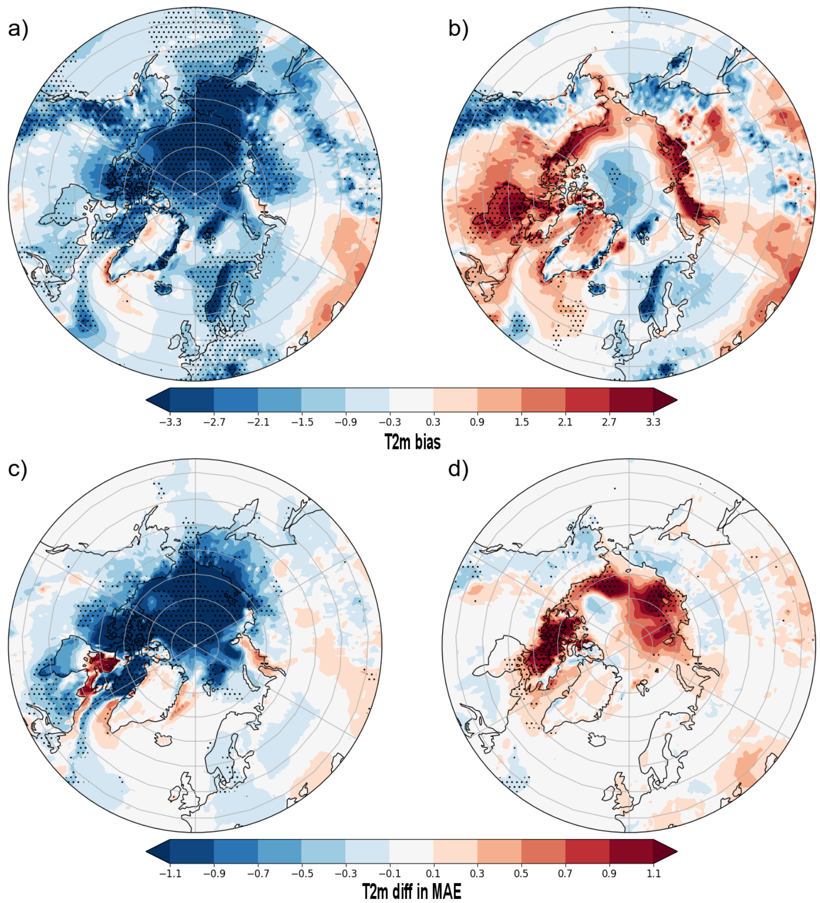

Figure 12: ~~Verification~~ Bias and difference in MAE of 2m temperature against ERA5 for SON forecasts started in May (a,c) and August (b,d) respectively: Bias (in K) of FC-REF is shown on the top (a,b), and MAE difference (in K) between FC-SIT and FC-REF at the bottom (c,d). Differences significant at the 5% level according to the sign test as recommended by DelSole and Tippett (2016) are stippled. The homogeneous warming of FC-SIT w.r.t. FC-REF shown in Figure 11c results in MAE for SON t2m being reduced for May start dates c) and increased for August start dates d).

hampers physical understanding of the impact of thickness initialization. We therefore focus only on forecasts for the September–November (SON) season, where the impact on 2m temperature is strongest.

As Figure 12a and b show, the 2m-temperature forecast bias for the SON season have a strong dependence on the start and lead month. Cold biases are clearly dominating the entire hemisphere in May forecasts, whereas a mixture of warm and cold biases is visible in August forecasts, with predominantly warm biases over the ice edge. As discussed previously, the thickness initialization leads to a homogeneous warming of 2m temperature (Figure 11c), which is not very sensitive to the time of initialization.

To determine whether the mean change leads to an increase or a reduction in the bias, we compute changes in mean absolute error (MAE) of 2m-temperature forecasts without the usual calibration. This is shown in Figure 12c and d. Mean absolute forecast errors are substantially reduced in SON (by more than 1K) over the entire ice cover and some adjacent regions (Figure 12c). In this case, the thickness initialization helps to mitigate the model bias. Conversely, when initializing forecasts in August, mean absolute forecast errors are increased over the marginal Seas of the Arctic Ocean and the Canadian Archipelago (Figure 12d). This points to an ~~exeacerbation~~ exacerbation of the model biases by the thickness initialization. However, the negative impact for August start dates is not as significant as the positive impact for May start dates.

Forecast skill changes on other atmospheric fields have been explored as well. The picture for circulation-related fields such as mean sea-level pressure and 500 hPa, geopotential height (not shown) is less clear compared to 2m-temperature, indicating that much of the statistically significant changes found at the near-surface temperature in the Arctic are due to local thermodynamic effects.

4 Summary and Concluding Remarks

In this paper we use 6 years of Arctic-wide sea ice thickness observations of January, February, March, November and December months during 2011 to 2016 to constrain the modelled sea ice thickness, and study the impact on the ocean–sea-ice reanalysis. Coupled forecasts of the ocean–sea-ice-wave-land-atmosphere are initialized using these data assimilation experiments, and the forecast skill of ~~Arctic~~ pan-Arctic sea ice for lead times up to 7 months is investigated. To our knowledge this study provides the first comprehensive assessment of coupled seasonal sea ice forecasts with thickness initialization for all the seasons. It complements to the study by Blockley and Peterson (2018), who reported the positive forecast impact on summer season only. This paper does not delve into the technical implementation of sea ice observational information in the ECMWF systems as reported in Balan-Sarojini et al. (2019), but instead it focuses on 1) collating the relevant scientific results on the impact of sea ice thickness information alone on seasonal forecasts, 2) conducting targeted diagnostics to gain understanding of the results, and 3) providing a more thorough discussion on the impact.

Constraining initial conditions by nudging to CS2SMOS ice thickness results in a substantial reduction of sea ice volume and thickness in the ocean–sea-ice analysis. This reduces some of the existing forecast biases in SEAS5 and improves forecast skill in the melt season, but in turn increases the errors during autumn, when the existing sea ice forecast bias is negative.

The impact of sea ice thickness initialization on seasonal forecast skill for Arctic sea ice variables, namely sea ice cover, sea ice thickness, sea ice volume and sea ice edge, is mostly positive for seasonal forecasts started from January to June start dates. We find significant improvement of up to 28% in the traditional September sea ice edge forecasts started from April start dates as shown by Integrated Ice Edge Error. However, sea ice forecasts for September started in spring still exhibit a positive sea ice bias, which points to too slow melting in the forecast model. Neutral forecast impact for November and December start dates is found. However, a slight degradation is seen in autumn forecasts started from July and August start dates, which is shown to be due to errors in the sea ice initial conditions. Both the ocean reanalyses, with and without SIT constraint, show strong melting in the middle of the melt season compared to the forecasts. This excessive melting is shown to be due to positive net surface radiation biases in the atmospheric flux forcings of the ocean reanalyses. Compared to the forecasts, strong freezing is seen throughout the freeze season in the ocean reanalysis without SIT constraint. With SIT constraint applied from November to March, the existing strong freezing is somewhat damped in the late freeze season. The exact causes of the differences in freezing between the reanalyses and forecasts require further investigation. Aggregating all the forecasts started in January to December, positive forecast impact of up to 5% skill improvement for integrated SIC is found at 2-5 lead months. Thinning of sea ice by CS2SMOS mitigates or enhances seasonally dependent forecast model error.

We reiterate that the sea-ice thickness observations are only available and assimilated for November-March. The ORA-SIT sea ice thickness from April-October is not constrained by observations. The fact that ORA-SIT has larger errors than ORA-REF in SIC for July is attributed to the overestimation of the melt in the forced model. The negative summer SIC bias gets worse in ORA-SIT than that in ORA-REF due to the fact that the ORA-SIT starts from a thinner ice state compared to ORA-REF without CS2SMOS thinning. Indeed, the assimilation of sea-ice concentration is trying hard to compensate for this excess of sea-ice melt as seen in the annual cycle of the pan-Arctic sea ice increments and melting energy tendencies. The reasons for this excess sea-ice melt during the summer season is investigated and is attributed to errors in forcing fluxes in the ORAs as just summarised. This key result points out that the dominant source of error lies in the atmospheric forcing rather than in the sea-ice model formulation or data assimilation in our experiments, and indicates that improved atmospheric fluxes from atmospheric reanalyses is urgently needed to improve the Arctic sea-ice related forecasts.

In this work we have only taken the very first step in SIT assimilation by using a simple nudging method to constrain SIT without considering the observational uncertainties. An area which needs to be explored in future studies of SIT assimilation is the use of thickness uncertainities. For instance, the uncertainty in the retrievals could be taken into account by perturbing the observations in the ensemble of data assimilations. We also note that this study does not cover recent sea-ice model improvements such as modelling sea-ice processes affecting the sea-ice melt/growth, which are being considered for inclusion in upcoming versions of the ECMWF forecasting systems. The use of multi-category sea ice models in coupled forecasting systems is another step forward in this direction. Since uncertainty of Arctic seasonal sea ice forecasts is reported to grow at a higher rate over thin ice regions than over the central Arctic (e.g. Blanchard-Wrigglesworth et al. (2017)), we recommend observational constraint of SIT for both the thick and thin ice regions in ORAs.

The impact of sea ice thickness initialization on atmospheric variables has also been investigated. Changes in ensemble mean 2m-temperature over the ~~Arctic~~ pan-Arctic region are significant for SON forecasts initialized from May to August start dates. The impact is also seen in mean sea level pressure and to certain extent in 500hPa geopotential height and is mostly local and thermodynamically driven, except for some remote impact over the north west Atlantic ocean. Similar to the sea ice edge forecasts, positive forecast impact is seen for 2m-temperature forecasts for the early freezing season, SON, started in May and negative impact for the same season is seen when started in August when the initial conditions are degraded. Statistically significant changes in 2m-temperature mean absolute error are predominantly due to corresponding local changes in errors in the sea surface temperature and sea ice variables. There is no clear change in forecast skill of upper atmospheric circulation in our experiments. Our results illustrate that information on sea ice thickness is relevant for identifying model errors and for exploiting the long-term memory present in ice thickness for seasonal forecasts of sea ice and near-surface temperatures. Constraining SIT in the initialisation alters biases arising due to both errors in the forcing and the sea-ice model. Though the SIT assimilation is not expected to solve these underlying problems per se, by moving the model state closer to reality, it helps us to better understand the errors in our system, as well as improving forecast skill scores in the meantime. As atmospheric forecast errors are dominated by biases, we are yet to demonstrate the benefit of interannual varying data on bias-corrected forecast scores. Robustness of impact on upper atmospheric variables and possible teleconnections need to be further assessed which would require a longer study period and larger sample size.

These findings demonstrate that making use of recently-available, spatially and temporally rich sea ice thickness observations from satellites for the ice growth season has the potential to significantly improve 1) the sea ice state in currently operational ocean–sea-ice reanalyses and, 2) the seasonal forecasts in operational forecasting systems. Our study also emphasizes the potential of future sea ice satellite missions for Earth System reanalysis and forecasts.

Acknowledgements

We acknowledge the European Union Horizon 2020 SPICES (Space-borne observations for detecting and forecasting sea ice cover extremes) project (640161) for funding this work. MM's work was partially supported by the Austrian Science Fund project 33177. The production of the merged CryoSat2-SMOS sea ice thickness data was funded by the ESA project SMOS and CryoSat-2 Sea Ice Data Product Processing and Dissemination Service, and data from 01/11/2011 to 31/12/2016 were obtained from Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research (AWI).

References

Allard, R. A., Farrell, S. L., Hebert, D. A., Johnston, W. F., Li, L., Kurtz, N. T., Phelps, M. W., Posey, P. G., Tilling, R., Ridout, A., et al.: Utilizing CryoSat-2 sea ice thickness to initialize a coupled ice-ocean modeling system, Advances in Space Research, 62, 1265–1280, 2018.

Balan-Sarojini, B., Tietsche, S., Mayer, M., Balmaseda, M., and Zuo, H.: Towards improved sea ice initialization and forecasting with the IFS, https://doi.org/10.21957/mt6m6rpwt, URL `https://www.ecmwf.int/node/18918`, 2019.

Balmaseda, M., Hernandez, F., Storto, A., Palmer, M., Alves, O., Shi, L., Smith, G., Toyoda, T., Valdivieso, M., Barnier, B., Behringer, D., Boyer, T., Chang, Y.-S., Chepurin, G., 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., Tang, Y., Yin, Y., Vernieres, G., Wang, X., Waters, J., Wedd, R., Wang, O., Xue, Y., Chevallier, M., Lemieux, J.-F., Dupont, F., Kuragano, T., Kamachi, M., Awaji, T., Caltabiano, A., Wilmer-Becker, K., and Gaillard, F.: The Ocean Reanalyses Intercomparison Project (ORA-IP), Journal of Operational Oceanography, 8, s80–s97, https://doi.org/10.1080/1755876X.2015.1022329, URL `http://www.tandfonline.com/doi/full/10.1080/1755876X.2015.1022329`, 2015.

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, Quarterly Journal of the Royal Meteorological Society, 136, 1655–1664, https://doi.org/10.1002/qj.661, URL `http://doi.wiley.com/10.1002/qj.661`, 2010.

Batrak, Y. and Müller, M.: On the warm bias in atmospheric reanalyses induced by the missing snow over Arctic sea-ice, Nature communications, 10, 1–8, 2019.

Beesley, J., Bretherton, C., Jakob, C., Andreas, E., Intrieri, J., and Uttal, T.: A comparison of cloud and boundary layer variables in the ECMWF forecast model with observations at Surface Heat Budget of the Arctic Ocean

(SHEBA) ice camp, Journal of Geophysical Research: Atmospheres, 105, 12 337–12 349, 2000.

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. Climate, 24, 231–250, https://doi.org/10.1175/2010JCLI3775.1, URL `http://journals.ametsoc.org/doi/abs/10.1175/2010JCLI3775.1`, 2011.

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, Climate Dynamics, 49, 1399–1410, https://doi.org/10.1007/s00382-016-3388-9, URL `https://doi.org/10.1007/s00382-016-3388-9`, 2017.

Blockley, E. W. and Peterson, K. A.: Improving Met Office seasonal predictions of Arctic sea ice using assimilation of CryoSat-2 thickness, The Cryosphere, 12, 3419–3438, 2018.

Bunzel, F., Notz, D., and Tietsche, S.: Definition of a new set of observation based metrics relevant for regional applications. SPICES Deliverable, D8.2 Report, 2017.

Chevallier, M., Smith, G. C., Dupont, F., Lemieux, J.-F., Forget, G., Fujii, Y., Hernandez, F., Msadek, R., Peterson, K. A., Storto, A., et al.: Intercomparison of the Arctic sea ice cover in global ocean–sea ice reanalyses from the ORA-IP project, Climate Dynamics, 49, 1107–1136, 2017.

Day, J. J., Hawkins, E., and Tietsche, S.: Will Arctic sea ice thickness initialization improve seasonal forecast skill?, Geophysical Research Letters, 41, 7566–7575, https://doi.org/10.1002/2014GL061694, URL http://doi.wiley.com/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, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, URL http://doi.wiley.com/10.1002/qj.828, 2011.

DelSole, T. and Tippett, M. K.: Forecast Comparison Based on Random Walks, Monthly Weather Review, 144, 615–626, https://doi.org/10.1175/MWR-D-15-0218.1, URL http://journals.ametsoc.org/doi/10.1175/MWR-D-15-0218.1, 2016.

Fichefet, T. and Maqueda, M. A. M.: Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics, Journal of Geophysical Research, 102, 12 609–12 646, https://doi.org/10.1029/97JC00480, URL http://doi.wiley.com/10.1029/97JC00480, 1997.

Fritzner, S. M., Graversen, R., Christensen, K. H., Rostosky, P., and Wang, K.: Impact of assimilating sea ice concentration, sea ice thickness and snow depth in a coupled ocean-sea ice modelling system, 2019.

Goessling, H. F. and Jung, T.: A probabilistic verification score for contours: Methodology and application to Arctic ice-edge forecasts, Quarterly Journal of the Royal Meteorological Society, 144, 735–743, https://doi.org/10.1002/qj.3242, URL http://doi.wiley.com/10.1002/qj.3242, 2018.

Goessling, H. F., Tietsche, S., Day, J. J., Hawkins, E., and Jung, T.: Predictability of the Arctic sea ice edge, Geophysical Research Letters, 43, 1642–1650, https://doi.org/10.1002/2015GL067232, URL http://doi.wiley.com/10.1002/2015GL067232, 2016.

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., et al.: A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales, Quarterly Journal of the Royal Meteorological Society, 142, 546–561, 2016.

Haas, C., Hendricks, S., Eicken, H., and Herber, A.: Synoptic airborne thickness surveys reveal state of Arctic sea ice cover, Geophysical Research Letters, 37, 2010.

Hendricks, S., Ricker, R., and Helm, V.: User Guide-AWI CryoSat-2 Sea Ice Thickness Data Product (v1. 2), 2016.

Hibler III, W. D.: A Dynamic Thermodynamic Sea Ice Model, J. Phys. Oceanogr., 9, 815–846, 1979.

Hogan, R., Ahlgrimm, M., Balsamo, G., Beljaars, A., Berrisford, P., Bozzo, A., pe, F. D. G., Forbes, R., Haiden, T., Lang, S., Mayer, M., Polichtchouk, I., Sandu, I., Vitart, F., and Wedi, N.: Radiation in numerical weather prediction, ECMWF Technical Memorandum, https://doi.org/10.21957/2bd5dkj8x, URL https://www.ecmwf.int/node/17771, 2017.

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S., Mogensen, K., Zuo, H., and Monge-Sanz, B.: SEAS5: The new ECMWF seasonal forecast system, Geoscientific Model Development Discussions, 2018, 1–44, https://doi.org/10.5194/gmd-2018-228, URL https://www.geosci-model-dev-discuss.net/gmd-2018-228/, 2018.

Kaleschke, L., Tian-Kunze, X., Maaß, N., Mäkynen, M., and Drusch, M.: Sea ice thickness retrieval from SMOS brightness temperatures during the Arctic freeze-up period, Geophysical Research Letters, 39, L05 501, https://doi.org/10.1029/2012GL050916, URL http://doi.wiley.com/10.1029/2012GL050916, 2012.

Kato, S., Rose, F. G., Rutan, D. A., Thorsen, T. J., Loeb, N. G., Doelling, D. R., Huang, X., Smith, W. L., Su, W., and Ham, S.-H.: Surface irradiances of edition 4.0 clouds and the Earth's radiant energy system (CERES) energy balanced and filled (EBAF) data product, Journal of Climate, 31, 4501–4527, 2018.

Kwok, R. and Rothrock, D. A.: Decline in Arctic sea ice thickness from submarine and ICESat records: 1958–2008, Geophys. Res. Lett., 36, L15 501, https://doi.org/10.1029/2009GL039035, 2009.

Laxon, S., Peacock, N., and Smith, D.: High interannual variability of sea ice thickness in the Arctic region, Nature, 425, 947, 2003.

Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J., Haas, C., Hendricks, S., Krishfield, R., Kurtz, N., Farrell, S., and Davidson, M.: CryoSat-2 estimates of Arctic sea ice thickness and volume, Geophysical Research Letters, 40, 732–737, https://doi.org/10.1002/grl.50193, URL http://doi.wiley.com/10.1002/grl.50193, 2013.

Lindsay, R. W., Zhang, J., Schweiger, A. J., and Steele, M. A.: Seasonal predictions of ice extent in the Arctic Ocean, J. Geophys. Res., 113, C02 023, https://doi.org/10.1029/2007JC004259, URL http://onlinelibrary.wiley.com/doi/10.1029/2007JC004259/abstract, 2008.

Madec, G.: NEMO ocean engine, Tech. rep., Institut Pierre-Simon Laplace (IPSL), URL http://www.nemo-ocean.eu/About-NEMO/Reference-manuals, 2008.

Mayer, M., Haimberger, L., Pietschnig, M., and Storto, A.: Facets of Arctic energy accumulation based on observations and reanalyses 2000–2015, Geophysical research letters, 43, 10–420, 2016.

Mayer, M., Tietsche, S., Haimberger, L., Tsubouchi, T., Mayer, J., and Zuo, H.: An improved estimate of the coupled Arctic energy budget, Journal of Climate, 32, 7915–7934, 2019.

Meier, W. N., Hovelsrud, G. K., Van Oort, B. E., Key, J. R., Kovacs, K. M., Michel, C., Haas, C., Granskog, M. A., Gerland, S., Perovich, D. K., et al.: Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity, Reviews of Geophysics, 52, 185–217, 2014.

Mogensen, K., Balmaseda, M. A., and Weaver, A.: The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4, Tech. Rep. 668, European Centre for Medium-Range Weather Forecasts, 2012.

Mori, M., Watanabe, M., Shiogama, H., Inoue, J., and Kimoto, M.: Robust Arctic sea-ice influence on the frequent Eurasian cold winters in past decades, Nature Geoscience, 7, 869–873, https://doi.org/10.1038/ngeo2277, URL `http://dx.doi.org/10.1038/ngeo2277`, 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, Quarterly Journal of the Royal

Meteorological Society, 144, 529–538, https://doi.org/10.1002/qj.3225, URL `http://doi.wiley.com/10.1002/qj.3225`, 2018.

Overland, J. E., Dethloff, K., Francis, J. A., Hall, R. J., Hanna, E., Kim, S.-J., Screen, J. A., Shepherd, T. G., and Vihma, T.: Nonlinear response of mid-latitude weather to the changing Arctic, Nature Climate Change, 6, 992, 2016.

Pohl, C., Istomina, L., Tietsche, S., Jäkel, E., Stapf, J., Spreen, G., and Heygster, G.: Broadband albedo of Arctic sea ice from MERIS optical data., Cryosphere, 14, 2020.

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, https://doi.org/10.5194/tc-11-1607-2017, URL `https://www.the-cryosphere.net/11/1607/2017/`, 2017.

Ruggieri, P., Buizza, R., and Visconti, G.: On the link between Barents-Kara sea ice variability and European blocking, Journal of Geophysical Research: Atmospheres, 121, 5664–5679, 2016.

Sallila, H., Farrell, S. L., McCurry, J., and Rinne, E.: Assessment of contemporary satellite sea ice thickness products for Arctic sea ice, The Cryosphere, 13, 1187–1213, 2019.

Scarlat, R. C., Spreen, G., Heygster, G., Huntemann, M., Paţilea, C., Toudal Pedersen, L., and Saldo, R.: Sea Ice and Atmospheric Parameter

Retrieval From Satellite Microwave Radiometers: Synergy of AMSR2 and SMOS Compared With the CIMR Candidate Mission, Journal of Geophysical Research: Oceans, p. e2019JC015749, 2020.

Semtner, A. J.: A Model for the Thermodynamic Growth of Sea Ice in Numerical Investigations of Climate, J. Phys. Oceanogr., 6, 379–389, 1976.

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, Geophysical Research Letters, 40, 529–534, https://doi.org/10.1002/grl.50129, URL `http://onlinelibrary.wiley.com/doi/10.1002/grl.50129/abstract`, 2013.

Stockdale, T., Alonso-Balmaseda, M., Johnson, S., Ferranti, L., Molteni, F., Magnusson, L., Tietsche, S., Vitart, F., Decremer, D., Weisheimer, A., Roberts, C. D., Balsamo, G., Keeley, S., Mogensen, K., Zuo, H., Mayer, M., and Monge-Sanz, B.: SEAS5 and the future evolution of the long-range forecast system, ECMWF Technical Memorandum, https://doi.org/10.21957/z3e92di7y, URL `https://www.ecmwf.int/node/18750`, 2018.

Tang, Y. M., Balmaseda, M. A., Mogensen, K. S., Keeley, S. P. E., and Janssen, P. E. A. M.: Sensitivity of sea ice thickness to observational constraints on sea ice concentration, Tech. Rep. 707, European Centre for Medium-Range Weather Forecasts, Reading, 2013.

Tian-Kunze, X., Kaleschke, L., Maaß, N., Mäkynen, M., Serra, N., Drusch, M., and Krumpen, T.: SMOS-derived thin sea ice thickness: algorithm baseline, product specifications and initial verification, The Cryosphere, 8, 997–1018, https://doi.org/10.5194/tc-8-997-2014, URL http://www.the-cryosphere.net/8/997/2014/ http://www.the-cryosphere.net/8/997/2014/tc-8-997-2014.html, 2014.

Tietsche, S., Notz, D., Jungclaus, J. H., and Marotzke, J.: Assimilation of sea-ice concentration in a global climate model – physical and statistical aspects, Ocean Science, 9, 19–36, https://doi.org/10.5194/os-9-19-2013, URL http://www.ocean-sci.net/9/19/2013/os-9-19-2013.html, 2013.

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, Geophysical Research Letters, 41, 1035–1043, https://doi.org/10.1002/2013GL058755, URL http://onlinelibrary.wiley.com/doi/10.1002/2013GL058755/abstract, 2014.

Tietsche, S., Balmaseda, M. a., Zuo, H., and Mogensen, K.: Arctic sea ice in the global eddy-permitting ocean reanalysis ORAP5, Climate Dynamics, https://doi.org/10.1007/s00382-015-2673-3, URL http://link.springer.com/10.1007/s00382-015-2673-3, 2015.

Tietsche, S., Balmaseda, M., Rosnay, P., Zuo, H., Tian-Kunze, X., and Kaleschke, L.: Thin Arctic sea ice in L-band observations and an ocean reanalysis, The Cryosphere, 12, 2051–2072, https://doi.org/10.5194/tc-12-2051-2018, URL `https://www.the-cryosphere.net/12/2051/2018/`, 2018.

Tonboe, R., Lavelle, J., Pfeiffer, R., and Howe, E.: Product User Manual for OSI SAF Global Sea Ice Concentration (Product OSI-401-b), 2017.

Uotila, P., Goosse, H., Haines, K., Chevallier, M., Barthelemy, A., Bricaud, C., Carton, J., Fuckar, N., Garric, G., Iovino, D., Kauker, F., Korhonen, M., Lien, V. S., Marnela, M., Massonnet, F., Mignac, D., Peterson, K. A., Sadikni, R., Shi, L., Tietsche, S., Toyoda, T., Xie, J., and Zhang, Z.: An assessment of ten ocean reanalyses in the polar regions, Climate Dynamics, (in press), 1–64, 2018.

Uotila, P., Goosse, H., Haines, K., Chevallier, M., Barthélemy, A., Bricaud, C., Carton, J., Fučkar, N., Garric, G., Iovino, D., Kauker, F., Korhonen, M., Lien, V. S., Marnela, M., Massonnet, F., Mignac, D., Peterson, K. A., Sadikni, R., Shi, L., Tietsche, S., Toyoda, T., Xie, J., and Zhang, Z.: An assessment of ten ocean reanalyses in the polar regions, Climate Dynamics, 52, 1613–1650, 2019.

Xie, J., Counillon, F., and Bertino, L.: Impact of assimilating a merged sea-ice thickness from CryoSat-2 and SMOS in the Arctic reanalysis, The Cryosphere, 12, 3671–3691, 2018.

Zampieri, L., Goessling, H. F., and Jung, T.: Bright prospects for Arctic sea ice prediction on subseasonal time scales, Geophysical Research Letters, 45, 9731–9738, 2018.

Zampieri, L., Goessling, H. F., and Jung, T.: Predictability of Antarctic sea ice edge on subseasonal time scales, Geophysical Research Letters, 46, 9719–9727, 2019.

Zuo, H., Balmaseda, M. A., and Mogensen, K.: The new eddy-permitting ORAP5 ocean reanalysis: description, evaluation and uncertainties in climate signals, Climate Dynamics, 49, 791–811, https://doi.org/10.1007/s00382-015-2675-1, URL `http://link.springer.com/article/10.1007%2Fs00382-015-2675-1`, 2017.

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., and Mayer, M.: The ECMWF operational ensemble reanalysis-analysis system for ocean and sea-ice: a description of the system and assessment, Ocean Science, 2019, 779–808, https://doi.org/10.5194/os–15-779-2019, URL `https://doi.org/10.5194/os-15-779-2019/`, 2019.