The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2020-73-RC2, 2020

Interactive comment on "Year-round Impact of Winter Sea Ice Thickness Observations on Seasonal Forecasts" *by* Beena Balan-Sarojini et al.

Anonymous Referee #2

Received and published: 8 July 2020

########## # Summary

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 obserations 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.

########## # 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.

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".

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

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 relexation timescale and why you chose 10 days.

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.

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).

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.

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)?

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.

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?

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

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 iceâĂŘedge forecasts") that you could apply to your ensemble forecasts directly, which would have the advantage that changes in uncertainty/reliability would be captured, too.

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!).

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).

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 differenc, or can it be linked to the continued assimilation of SIC (or ocean variables)?

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 tendeny" 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.

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?

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.

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?

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.

########## # Technical corrections

L25: last -> lasts

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

L91: These -> This

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

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

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

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

L227: "indicates" -> "indicate"

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

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

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

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

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

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

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2020-73, 2020.

The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2020-73-RC1, 2020

Interactive comment on "Year-round Impact of Winter Sea Ice Thickness Observations on Seasonal Forecasts" *by* Beena Balan-Sarojini et al.

Anonymous Referee #1

Received and published: 8 July 2020

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.

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 Cryosat-2, 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.

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.

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.

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? - In regards to July ORA-SIT 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? 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." - 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.

Minor comments & corrections

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

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".

l. 25: change to "lasts into autumn"

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.

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.

l. 91: change to “The Level-3”

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.

l. 168: change to “differ in”

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.

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

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

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.

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

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.

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?

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

l. 369: correct spelling of “Atlantic”

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2020-73, 2020.