Review of "Seasonal Sea Ice Prediction with the CICE Model and Positive Impact of CryoSat-2 Ice Thickness Initialization" by Shan Sun and Amy Solomon.

This manuscript investigates the forecast skill of the sea-ice model CICE in stand-alone simulations. The analysis includes both the Arctic and Antarctic. The model CICE is shown to adequately reproduce the seasonality of the sea ice extent and volume, but with a larger difficulty in forecasting the winter and spring conditions. A particular interest is placed on the influence of the initial ice thickness on the forecast skill. Specifically, the use of the CFSR ice thicknesses, which over-estimates the ice thickness, is shown to decrease the forecast skills regardless of the initiation month. This is largely improved when ice thickness from Cryo-Sat2 is used instead. The simulations are presented as a baseline for future studies on the forecast skill of coupled models using CICE as the sea-ice component.

This manuscript is relatively well written, although many sentences are too long and confusingly constructed. The results are interesting and will interest many in sea-ice modelling community. Nonetheless, the manuscript suffers from an unclear problem statement in the introduction and from a tendency to describe figures without much in depth interpretations. This is especially important given that other studies have looked into the impact of ice thickness on forecast skills.

I believe this manuscript have potential for publication, but require major revisions.

Major points:

There is a tendency (mostly in the introduction) in lumping too many ideas in complicated sentences. The text should be revised in that regard.
The introduction lacks a problem statement and should be revised to clearly identify the scientific questions that the analysis aims to answer. A problem statement is vaguely formulated at L37-45 but mixed with the broader context. I believe that adding a paragraph devoted to the problem statement (mainly on the influence of the initial ice thickness on the forecast skills) would largely clarify the scope of the manuscript. In particular, it should clarify what information the stand-alone CICE simulations can bring that has not yet been documented.
There is very little mention of the sea-ice dynamics (in the experiment setup, results and discussion) although it should largely affect the sea-ice extent, especially in a year-long simulation. The thermodynamics and dynamics contribution to the SIE is briefly investigated in Figure 7, although this analysis not clear (the methods are not described) and needs to be clarified.
Much of the results are very descriptive and not thoroughly discussed. I believe that a comprehensive assessment needs to be include before publication. For instance, many statements are vague and general (e.g. the forecast skill is reduced in the alt-init simulations), despite the figures presenting much information. More in depth analysis of the results could include, for instance, explaining the differences in the SIC and SIV forecast skill patterns, and how it relates to the initial ice thickness.

We thank the reviewer for their careful reading of the manuscript and many constructive suggestions, which helped improve the quality of the manuscript. We have incorporated them in the revision, see details below (reviewer's comments in black, our replies in blue).

Specific points:

L17-19: This sentence is confusingly constructed. Perhaps dividing it in smaller sentences would be clearer.

We have simplified this sentence to:

"A relatively thin material layer between the atmosphere and ocean, sea ice amplifies radiative climate feedback with a higher surface albedo than open water."

L20-22: This sentence is currently confusing as it covers too much while being too vague. For instance, what "forecast" and "predictability" are we talking about (weather conditions? Ocean? Climate?). The vague reference to the impact of sea ice conditions on teleconnections also needs expanding.

This sentence is rewritten as:

"Reliable sea ice prediction is important not only for the polar regions but also is expected to improve predictability at mid-latitudes at subseasonal to seasonal time scales due to teleconnections."

L23: It is not clear what we are talking about here. Weather?

This sentence is rewritten as:

"Finding sources of weather predictability at S2S time scales is a challenging research topic."

L23: I am not sure that "leading" is the right verb... Perhaps "Driving"?

This sentence is rewritten as

"Finding sources of predictability for weather at S2S time scales is a challenging research topic."

L29: The use of "In particular" is confusing here, as we were discussing the influence of sea ice on weather predictability, but now jump to sea ice forecasting.

This sentence is rewritten as:

"Weather predictions from a numerical model incorporating a sea ice model have higher skill than those based on persistence of Arctic sea ice (Grumbine, 2003; Hebert et al., 2015; Intrieri et al., 2022). In addition, sea ice forecasts at seasonal to interannual timescales relies on the accuracy of the sea ice initial conditions (Holland et al., 2011; Blanchard-Wrigglesworth et al., 2011b; Wang et al., 2013) and in particular, a more realistic sea ice thickness initialization (Krinner et al., 2010; Chevallier and Salas-Mélia, 2012; Day et al., 2014a; Collow et al., 2015; Allard et al., 2018; Blockley and Peterson, 2018; Schröder et al., 2019, to name a few)."

L37-39: Do I understand that here, you validate the sea ice model component of a fully coupled ice-ocean-atm model, in a first step towards investigating how it feedbacks with the other components?

The standalone sea ice model is evaluated here, while a fully coupled ice-ocean-atm model is under development. We have modified this paragraph to:

"In order to separate various feedbacks among the components of a fully coupled model, in this study we aim to evaluate the seasonal prediction skill of CICE in standalone mode with prescribed atmospheric forcings. It is essential to assess the skills of each module separately in an uncoupled mode before assembling them in a fully coupled model, as it is easier to reveal the strength and weakness of a module in such a controlled environment. This work is even more relevant as CICE, originally developed for climate research, is to be used in the S2S prediction in operation."

L37-45: The structure of this paragraph makes it difficult to understand the scope of the paper. It first indicates that the aim is to isolate feedback processes between coupled model components, then that it is to validate the sea-ice model used in NOAA UFS, in stand-alone simulations. However, I believe that the real goal here is to assess the influence of initial thicknesses on the ice predictability within CICE. This needs to be clarified.

The initial goal of the project was to evaluate the model skill with a given set of initial conditions and atmospheric forcing. Half way through the project, we noticed an obvious bias in the ice thickness in the 'given' initial conditions. We decided to reduce this bias by using satellite observations and evaluate how much skill can be gained from this alone. We have made this clearer in the text.

"We also investigate the sensitivity of prediction skill in the standalone CICE when a bias in the ice thickness initialization is removed."

L51-53: This sentence needs revisions.

This sentence is rewritten as:

"The linear function of salinity (ktherm=1) was used for the freezing temperature. The elastic-viscous-plastic rheology (kdyn=1) were specified for the sea ice dynamics. The ice strength was set to be closely related to the ridging scheme (kstrength=1). In addition, CICE needs the information of sea surface temperature (SST) which can be either prescribed or generated by its own built-in mixed layer model. We chose the latter in this study for the sake of consistency between the SST and the ice state."

L55: I suggest starting a new sentence after "experiment".

Done.

Section 2: Some information on the dynamical component should be provided (e.g., I assume it is the standard EVP rheology and strength parameters in CICE ?).

Done. See above under L51-53.

L66-69: This long sentence could be improved.

This sentence is rewritten as:

"Furthermore, monthly mean estimates of the Arctic ice thickness are available from satellite observations of CryoSat-2 during boreal winter (October to April) since 2011 (Grosfeld et al., 2016). Additional CICE experiments were carried out by initializing ice thickness from this dataset instead of CFSR."

Section 3: What is defined as a "reliable forecast"? This statement is made at various places throughout the result section, but sounds rather subjective.

This sentence is rewritten as:

"The model is able to make SIE forecasts in good agreement with observations at all lead times for the austral winter, spring and especially fall."

L82: It should be specified here (not later) that Figure 1 only shows simulations from 2014.

We have modified Fig.1 with all-year mean and standard deviation to show the interannual variability.

L84: There is very significant inter-annual variability in Arctic (and Antarctic) sea ice extent, yet here you say that the inter-annual differences are small? This needs to be clarified. For instance, you show 2014, a year where the winter maximum was relatively small (~14.9 million km2) and the summer minimum remarkably large (5.0 million km2). It is possible that conclusions drawns from Figure 1 are not representative of different years, such as 2012.

We meant to say the conclusion from year 2014 in Fig.1 is valid for all other years, despite year to year variations. We redid Fig.1 with an all-year mean and standard deviation. This sentence is rewritten as:

"The Arctic SIE forecast matches the NSIDC observations better in the warm season than in the cold season both at 0.5-month and 5.5-month lead times, and the positive bias is biggest in winter. The biggest interannual variabilities in the predicted Arctic SIE occurs in September as shown by the standard deviation."

L94: This result is very surprising to me, as the summer minimum is usually described as being more difficult to forecast and more dependent on the early summer meteorological conditions.

This is indeed the case in many models. Here we see a spring barrier with a lower skill for the summer prediction in Fig.2a. We have added this sentence in the manuscript:

"The control experiments have the lowest RMSE in the Arctic SIE in fall at all lead times up to 12 months. There is an obvious skill barrier in spring for the summer forecast."

L91-102: Together, these results are confusing and should come with some analysis and explanation. The ice extent results seem to indicate that the model overestimates the ice growth, but that does not show in the SIV results. On the contrary, the SIV results seem to indicate biases in the spring and summer melt, but this does not show in the ice extent results. Why? Is this expected? Also, changing the initial thickness improves the overall skill but does not seem to change the temporal patterns. Does this implies that the thickness influences the error magnitude but not the predictability patterns?

We redid Fig.1 to show the biases in SIE and SIV in the control experiment at 0.5-month and 5.5-month lead time to show
(1) there is a positive bias in both SIE and SIV at 0.5-month lead time in the Arctic during the cold season;
(2) the delay of max/min peak in SIV with longer lead time indicates the positive bias in SIV has a long-lasting effect on SIV and SIE.

Then in Fig.2, we plotted the bias in SIE and SIV at all lead times in both control and alt-init experiments to show
(1) biases in SIE and SIV are in general smaller in the magnitude in the alt-init compared to the control experiment, while the bias pattern is still the same;
(2) the difference in SIV from "alt-init" and "control" is bigger with longer lead times in summer and fall forecast;
(3) the phase shift between the biggest bias in SIE and SIV is obvious.

L108: spatial distribution

Figure 2: A couple things that are concerning in these figures:

Why is there a low concentration spots at the North Pole, but only for the top-middle and bottom-right panels, and of different size?

We didn't realize that there was no measurement at the North Pole in the original CryoSat2 data set. We have filled up this hole.

What is that line of low concentration North of Franz Josef Land, running from the Laptev Sea to Svalbard? It is a very strange location and orientation for such an LKF, and it is seen both in April and October. More confusing, it is seen in the observations but only in October. Is it real or is it an artefact?

We investigated the origin of the line of low concentration north of Franz Josef Land, and found a bug in the namelist used to run CICE, which produced wrong wind and a divergence along the model grid line of i=1 where a bipolar grid was used. We have rerun all the experiments and this line is gone.

The AMSR2 observation data used in Fig. 4 was taken from day 1 of an additional model run initialized with AMSR2 (not documented here). However, the day 1 output in October was already impaired by the bug mentioned above when the ice was very thin. We have now replaced the figure by the "raw" AMSR2 data IC and the line is gone.

L117: On the initial thickness being the dominant source of error: how did you determine this dominance? There are no results on the contribution from other sources (dynamics mass balance, thermodynamics mass balances). This statement is also contradictory with the fact that the errors are large in the winter sea ice at all lead time: they are thus are likely dominated by other factors than initial ice thickness. The use of CryoSat also does not seem to change this pattern.

We verified the ice thickness against the CryoSat2 data set. We agree that there are other sources of errors since this model still has biases when initializing with CryoSat-2, as shown in Fig.2. We see the initial ice thickness is an "obvious" error at time=0. We have added a discussion in Sec 3.1. This section is rewritten as

"As seen in Figs. 4 and 5, a positive bias in SIT in the initial conditions appears to be one source of the error in SIT as well as SIA in the seasonal forecast. Both SIT and SIA forecasts improved when this bias was removed in the alt-init experiment."

L122-123: This sentence is confusing and needs to be re-organised.

This sentence is rewritten as

"Fig.6 presents SIV initially in January 2016, its forecast for December 2016 in both experiments and the comparison to CryoSat-2 observations."

L129-130: I would edit to: "the SIE and SIV in the alt-init run in the fall ended up closer to NSIDC and PIOMAS than in the control run."

Done.

L130-137: How are the dynamics and thermodynamics contribution measured? I find it somewhat surprising that the dynamics tendency is exclusively negative. Also unclear to me is how do you define the thermodynamics tendency in sea ice extent? The thermodynamics is usually defined by column physics, and not directly related to changes in area.

We made an error in the unit in Fig.7(c): the tendencies are measured in "%/day" for ice area and "cm/day" for ice volume per unit area. The area tendency from dynamics and thermo-dynamics (daidtd and daidtt) and the volume tendency from dynamics and thermo-dynamics (dvidtd and dvidtt) are achieved in the standard model output, which are shown in the attached Figs. s1 and s2, valid for January 2016 and July 2016, with IC of January 1, 2016. The dynamics tendency for sea ice area (SIA) in January is a mixture of positive tendency near the ice edge and negative near coastlines, overall dominated by the negative. The dynamics tendency for SIA in July is smaller than its thermo-dynamic component.

The physical mechanism of sea ice dynamics can also be regarded as creating leads. Therefore, its tendency on SIA tends to be negative in all seasons.

L138-144: How can the top melt be influenced by the ice thickness? It is more intuitive for the bottom melt, as the reduced thickness would also reduce the insulation, but it could also be mentioned.

Unlike the basal melt, the top melt is not directly related to the ice thickness. Some surface properties like long wave emissivity and aerodynamic roughness are a function of the ice thickness. We have revised the text to…

"The biggest difference is a larger basal melt in the Arctic region between 120°E and 120°W in the alt-init experiments, where the ice thickness is mostly below 1.5 m compared to above 1.5 m in the control experiment. The basal melt pattern is consistent with the fact that the conductive heat flux at the ice bottom is inversely proportional to the ice thickness (Hunke, et al. 2015)."

L140: remove "apparently".

Done.

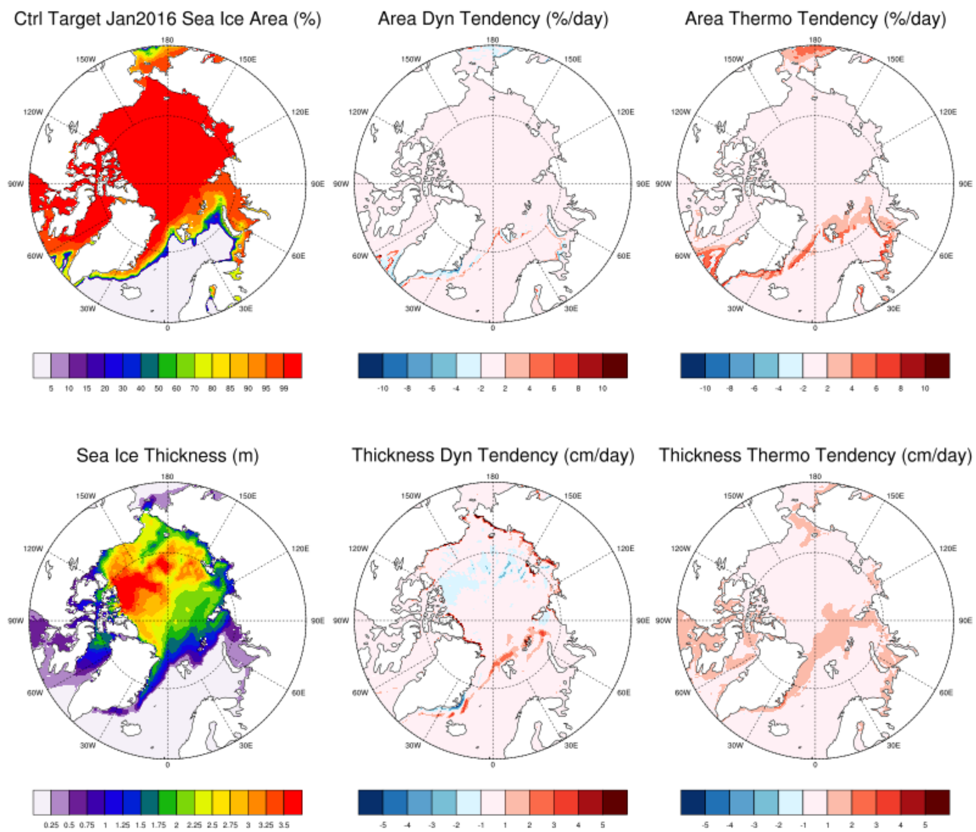

Fig. s1: top, from left to right: ice area, its dynamics and thermodynamics tendency, all monthly mean for January 2016; Bottom: same as top, except for the ice thickness. Initialized on January 1, 2016.

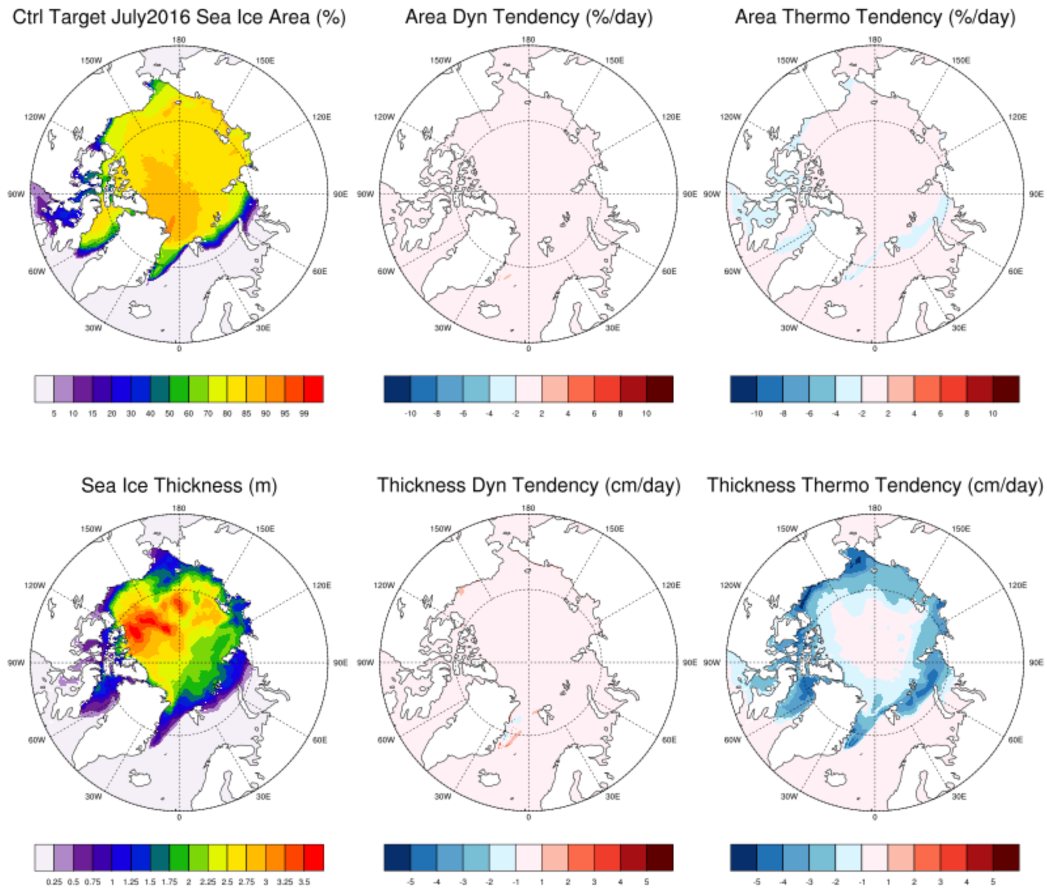

Fig.s2: same as in Fig.s1, except valid for July 2016.