We thank the reviewer for their careful reading of the manuscript and constructive remarks, which helped improve the quality of the manuscript. We have addressed all the recommendations, see details below (reviewer's comments in black, our replies in blue).

Manuscript Synopsis

This manuscript evaluates the usage of the CICE sea ice model for seasonal forecasting in NOAA's Unified Forecast System by examining 12 month simulations of CICE driven by reanalysis atmospheric forcing and an ocean 1-D mixed layer model. Using root mean square errors in integrated quantities, the authors evaluate hemispheric wide errors in sea ice extent and sea ice volume - but not spatial errors.

My evaluation of this manuscript is to reconsider after major revisions. These major revisions (major comments 1, 2 and 4 below) can mostly be dealt with by the supply of additional information either in text or in response. Major comment 3 is a recommendation which could enhance the manuscripts applicability, but is not critical to its publishability.

The most pressing issue, which may stem solely from my lack of understanding of running CICE in stand-alone mode and comes down to a lack of details with respect to the oceanic boundary condition (Section 2.2). While I understand the authors usage of the mixed layer ocean model for the sea ice thermodynamics, it does not explain what was used for bottom boundary dynamical conditions – and leads me to believe there are none, or more precisely, it assumes an unmoving ocean. Perhaps this could be resolved with some rather simple re-working of the model introduction (i.e. more to Section 2.2 then just the thermodynamic lower B.C.). As the location of the thickest ice is predominantly set by dynamic processes (driven against the Canadian Archipelago and northern Greenland) – and the location of the thickest ice in the model setup is not modelled very well (in the Beaufort Sea and central Canada Basin), this suggests a possible deficit in dynamical tendencies that requires a better articulation of the dynamical lower boundary conditions. [Scientific Quality: Methods not adequately explained.]

Thank you for bringing this to our attention. Indeed, we assumed the ocean to be stationary, and there was no specified dynamical condition for the bottom boundary. We have now included this clarification in the manuscript:

It's important to note that the mixed layer ocean model used in this study is a simplified one-dimensional stationary model that doesn't include horizontal advection in the ocean. This limitation does impact the model results, as will be demonstrated later.

A 2nd major omission of the manuscript is the mistaken impression implied in the introduction (ll. 37-41; and granted it is not stated explicitly) that CICE has mainly been used for climate simulations. This is far from the truth, as CICE is already in use in many operational seasonal, sub-seasonal and NWP systems, which I detail further below. [Originality: Manuscript does not adequately present current scientific Understanding.]

Thank you for pointing this out. We have now updated this in the manuscript:

CICE, originally developed for long-term climate research, has been used in seasonal prediction applications with success and challenges, for example, in the Global Seasonal forecast systems at the UK Met Office and the Canadian Seasonal to Interannual Prediction System (CanSIPSv2) (Arribas et al., 2011; MacLachlan et al., 2015; Lin et al., 2020). In addition, Martin et al (2023) attributed the enhanced skill in CanSIPSv2, compared to the previous version, to an improved sea ice initialization procedure.

Lastly, while the skill metrics used in the manuscript are acceptable – but very climate oriented – they can be enhanced quite easily to assess integrated errors as opposed to errors in integrated quantities (IIEE), simply by changing the order in which the operations are performed. This point I would be willing to give the authors some leeway with. The regional results (Section 3.4) do address this issue somewhat – although Figure 11 is not correct, but Figure 12 does not suggest any cancellation of error (only deficits in sea ice). Nevertheless it could highlight further errors not elicited by Figures 2 and 3. [Scientific Quality: Validation methods could be improved. Significance: Application of results could be improved through enhanced spatial Information.]

See below.

I appreciate that this manuscript may already have been through at least one round of review and revision, so I hope my suggestions do not pose too onerous a task.

My recommendation is Reconsider after Major Revision.

Major Comments

1. Section 2.2: No explanation is given for lower (ocean) dynamic boundary condition. While the authors dedicate a sub-section (Section 2.2) to the ocean boundary conditions, this only explains the lower boundary condition for Sea Surface Temperature (SST) – which is chosen to be a mixed layer 1-D ocean – and would only effect the model thermodynamics. No explanation for specification of the dynamical boundary conditions, or more precisely, the ocean surface currents is given. Presumably, this would have a large effect on the sea ice dynamics, which it is impossible to make informed decisions regarding without further information.

The one-dimensional mixed layer ocean model has a thickness of 20 m and is stationary, solely designed for sea ice thermodynamics. The omission of horizontal advection in the ocean from this setup has notable implications for the results, as highlighted in the manuscript.

2. ll. 37-41: The introductory paragraph gives the false impression that the CICE sea ice model is primarily used for climate simulations (Note: I do not deny that the model was initial constructed for this purpose), implicitly implying that its introduction as the sea ice component for NOAA's Unified Forecasting System is a novel usage. The exact phrasing used (ll 39-40) "its suitability for seasonal forecasting needs to be assessed." I do agree, or at least do not disagree with that statement, however, some credit through citation is deserving to the multitude of operational (and quasi-operational) systems currently in use for seasonal forecasting use throughout the world for over a decade (Note: many of these, by necessity, are self-serving):
(1) UK Met Office GloSea4/5 system (https://doi.org/10.1175/2010MWR3615.1, https://doi.org/10.1007/s00382-014-2190-9, https://doi.org/10.1002/qj.2396)
(2) Korea Met Agency version of GloSea5 (https://doi.org/10.5194/tc-2018-217)
(3) ASSESS-S1/2, Australia Bureau of Meteorology version of GloSea5 (https://www.publish.csiro.au/es/ES17009, https://www.publish.csiro.au/ES/ES22026).
(4) CanSIPSv2 GEM-NEMO-CICE component (https://doi.org/10.1175/WAF-D-19-0259.1, https://doi.org/10.1175/WAF-D-22-0193.1, https://www.tandfonline.com/doi/pdf/10.1080/07055900.2023.2252387

6 of 11 dynamical contributors to the ARCUS July 2023 sea ice outlook (https://www.arcus.org/sipn/sea-ice-outlook/2023/july (5 listed below as UK Met Office is one of 6)

(5) RASM/NPS (https://doi.org/10.5194/gmd-11-4817-2018)
(6) ArcIOAM, National Marine Environmental Forecasting Center, China (https://doi.org/10.5194/gmd-14-1101-2021)
(7) FIO-ESMv1.0, Qingdao, China (https://doi.org/10.3389/fmars.2020.00504
(8) FGOALS-f2 V1.3, Institute of Atmospheric Physics, China (https://doi.org/10.1029/2019MS002012)
(9) Unified Forecast System, NOAA (https://doi.org/10.1029/2022GL102392). Note: This is the prototype system being evaluated in this manuscript. and a final system via a google search for CICE seasonal forecasts
(10) SLAV/INMIO/CICE, Marchuk Institute of Numerical Mathematics / Shirshov Institute of Oceanology / Hydrometeorological Centre of Russia (https://doi.org/10.1515/rnam-2018-0028)
to which I will also add shorter range S2S, monthly and short range (< 10day) systems:
(11) Global Ensemble Prediction System, Environment and Climate Change Canada (https://doi.org/10.1002/qj.4340) (S2S/monthly/extended)
(12) Global Ice Ocean Prediction System / Regional Ice Ocean Prediction System, Environment and Climate Change Canada (short range) (https://doi.org/10.1002/qj.2555, https://doi.org/10.1175/MWR-D-17-0157.1, https://doi.org/10.5194/gmd-14-1445-2021)
2
(13) Forecasting Ocean Assimilation Model (FOAM), UK Met O ce (https://doi.org/10.5194/gmd-7-2613-2014) (short range)
(14) Prototype UK Met Office Coupled System for NWP (https://doi.org/10.1175/WAF-D-20-0035.1) (short range)

While I would not expect (the list was longer than even I had initially assumed!) the authors to cite each and every one of these, it would still be appropriate to underline the usage of the CICE model in existing operational systems – and highlight and reference their results against the sea ice predictability of some of these earlier systems, particularly when an assessment of sea ice performance has been undertaken. See point 4 for one such possible connection.

Thank you so much for bringing this to our attention and for providing a comprehensive list of references. We have revised the manuscript accordingly, as mentioned earlier.

3. RMSE skill measure: While the RMSE quantification of error for the hemispheric domain is adequate, it is also relatively non-standard. More usual to be found in seasonal papers is anomaly correlation (https://doi.org/10.1002/grl.50129, https://doi.org/10.1007/s00382-014-2190-9), which eliminates bias. However, more modern skill estimates account for both the area of the sea ice extent along with its position, through skill measures like Integrated Ice Edge Error (https://doi.org/10.1002/2015GL067232). Implementation of this would be simple enough. All you need do is commute the order in which the area and rmse error operations are performed. [I.e. calculate the square error of ice existence $(M - O)^2$, where M is modelled ice > 0.15 concentration and O is observed ice > 0.15 concentration in any grid cell, and then perform your summation. Note: Since (M - 0) is 1/0 it does not matter whether you square or take absolute value. Taking the square easily allows you to generalize for an ensemble, where M is replaced with P, the fraction of ensemble members with ice (https://doi.org/10.1002/qj.3242). You could also consider replacing the RMSE of total ice volume with the integrated square error of grid cell ice volume (it is preferable to add square error, not root mean square error). The latter will then give you a double penalty for having ice volume in the Beaufort Sea, but little over the Canadian Archipelago and Greenland. I will not insist on the authors doing this, but it could enhance the applicability of their results.

We appreciate your suggestion regarding IIEE and the accompanying reference. Indeed, the computation of IIEE and RMSE diverges in the order of summation operations. Unfortunately, the existing NSIDC data available to us provides only hemispheric totals. The calculation of IIEE requires data at each grid point, which is currently impractical for us due to time constraints.

While the current circumstances hinder our ability to calculate IIEE, we acknowledge its significance as a valuable metric. We are committed to exploring the inclusion of IIEE in future studies when the availability of more data permits.

4. Although it is also a characteristic of the IEEE as well (so that will not solve this problem), the RMSE of an area integrated quantity will be inherently larger when that integrated quantity is larger. Thus it may be natural for the RMSE in February and March to be large solely because the ice extent is large during those periods. It would be a more accurate assessment of whether the ice area predictability is better or worse by comparing the RMSE with the interannual variability for that time of year. That being said, deterioration of predictability seems also to occur for mid to late winter in many other assessments of sea ice predictability through correlation skill assessments (https://doi.org/10.1002/grl.50129, https://doi.org/10.1007/s00382-014-2190-9, https://doi.org/10.1175/WAF-D-22-0193.1). Perhaps the authors can comment on this – and cite previous seasonal sea ice skill Assessments.

It is a good point that SIE has a strong interannual variability in late winter. We have added a measure of RMSE in terms of percentage of the observations, and a ratio of annual maximum and minimum SIE. We also added:

The decline in forecast skills for late winter and the preference for fall start dates seen here align with findings in Peterson et al. (2014), Martin et al. (2023).

5. Figure 11: Figure 11 is an error/omission. Figure 11 is identical, save for 6 month offset x-axis to the correctly attributed Figure 12.

Could you say more to the error in Fig. 11? Figs. 11 and 12 are identical, differing only in the choice of initial months. Fig. 11 is for experiments starting on April 1, close to the annual maximum SIE, and Fig. 12 is for experiments starting on Oct. 1, close to the annual minimum SIE. Both sets of experiments, regardless of the initial month, reveal a positive bias in both SIE and SIV in the BKG Seas and Baffin Bay. We attribute this bias to the model's lack of northward oceanic heat transport.

Minor Comments

1. Section 2.1: You should emphasize that you are forcing the sea ice integration with "0-hour lead" reanalysis forcing. In other words, you are not performing a true seasonal forecast, where the forcing is also of long lead time.

Since this is in standalone mode, atmospheric forcing is prescribed and time-varying. We have added that "The *prescribed time-varying* atmospheric boundary forcings used in this study are derived from the 6-hourly archives obtained from CFSR".

2. Section 2.2: The above point then begs the question as to why "0-hour lead" SST forcing does not lead to a better ice concentration (not necessary thickness) integration as described in Guemas et al (2014; http://dx.doi.org/10.1007/s00382-014-2095-7). Perhaps it did – the explained reason for abandoning due to it "result(ing) in an unrealistic increase in basal melt," was for reasons of unrealistic thermodynamics, it may still have resulted in a more accurate sea ice concentration integration – likely

at the cost of a more unrealistic sea ice thickness integration. Perhaps the authors can expand their Explanation.

Appreciate the reference provided. The performance degradation of CICE was unfortunately influenced by a positive bias in SST from the CFSR. It's reassuring to note that the ocean temperature in ORAS4 exhibited strong performance, as highlighted in Guemas et al. (2014), surpassing that of CFSR. We have incorporated the reference to Guemas et al. (2014) into the manuscript, explicitly acknowledging that a minor yet persistent positive bias in CFSR SST data led to the unrealistic basal melt seen in the initial experiment. Consequently, we had to abandon the original approach.

3. Section 2.3: It is very important you explicitly specify you initialized to the Cryosat-2/SMOS dataset (Ricker et al, 2014, https://doi.org/10.5194/tc-11-1607-2017). Otherwise readers will be confused on how you initialized sea ice with thickness less than 1m.

Unfortunately, we utilized the raw CryoSat-2 data without knowledge of CryoSat-2/SMOS as mentioned in Ricker et al. 2017. Consequently, there are large uncertainties in sea ice thickness over thin ice regimes. We have acknowledged this limitation in the manuscript, and it is an aspect we intend to enhance in future experiments.

Minor Presentation Comments

1. Perhaps this is pedantic, but the units should really be on the colour bar (Figures 2, 3, 4, 5, 6, 8, 9), or on the y-axis (Figures 7, 11, 12) if possible, and not just (could be additionally) in the figure title.
Note: Figure 9, lacks units completely – although fairly obviously ºC/K. The latter at least needs to be corrected.

Thanks for your suggestion. We are able to add units to all colorbars successfully.