This study examines the sensitivity of sea ice prediction skill in the Barents and Kara Seas in two seasonal forecasting systems, SINTEX-F2 and SINTEX-F3, to different initialization strategies for the ocean and sea ice components. The SINTEX-F3 system is initialized with SIC and SIT from the CMCC C-GLORS reanalysis and is shown to have better skill than when initializing SST alone. The assessment is limited to a May initialization and focuses on summer sea ice extent. The forecasts are compared with those from other systems hosted on Copernicus (C3S) and are shown to have similar skill in this case. The authors also investigate different sources of predictive skill in the region of interest through a regression analysis.

However, most of the results are not new. It is already known that initializing SIC improves sea ice prediction skill beyond initializing SST alone, particularly in summer. Initializing SINTEX-F3 with SIC and SIT from the C-GLORS reanalysis appears to bring the system more in line with existing systems, but does not demonstrate anything novel with respect to initialization. While I think it is reasonable to report on the system and that the regression analysis is of interest, the authors should be more transparent about the scope of the study and what it does and does not demonstrate relative to what has already been established in the literature. The title, abstract, and introduction give the impression that the study will test the impact of assimilating SIT observations, when this is not the case. I have detailed my concerns below, including comments on how the regression analysis could be clarified, missing information, incorrect characterization of how the models on C3S were initialized, and a general misunderstanding of the spring predictability barrier that has implications for the study's conclusions.

Scope and novelty:

- I see some issues regarding the characterization of this study in terms of scope and novelty. From the title, abstract, and throughout the study, the authors are not very transparent about what is actually being assimilated in the best-performing experiment configuration (that being, the one abbreviate F3-SI) in which sea ice concentration and sea ice thickness are set to those from a reanalysis product (C-GLORS from the CMCC). It is never stated what sea ice observations are assimilated into C-GLORS, which I think is only limited to SIC. The emphasis in the study though is on SIT initialization, which is based on the unconstrained (at least directly) reanalysis---i.e., no SIT observations are assimilated in that reanalysis. This information is central to the methodology and scope of the study and needs to be included.

- The authors compare the experiments done with SINTEX-F3 with similar hindcast records from several operational centres hosted on Copernicus C3S. However, they claim incorrectly (e.g. Table 1) that sea ice is only included in the initialization of SINTEX-F3 and the model from DWD. To the best of my knowledge, all the other models included in that table are initialized over most of their historical records from in-house reanalyses or reanalyses from other centres that do in fact assimilate SIC. This switches to an operational analysis in some cases for more recent years (e.g. for ECCC), but which still includes assimilation of satellite-based SIC observations. I'm surprised by this oversight, given that SINTEX-F3 is initialized from a reanalysis that is also used to initialize CMCC's hindcasts on C3S. The authors need to look through the documentation for those models on C3S, correct this table, and put the scope of the study into better context with respect to what is already done in the other systems on C3S. I also don't think it's accurate to include SIT and Snow in the "Init" column of SINTEX-F3, since no SIT or snow observations are assimilated in C-GLORS (as far as I know). Finally, I think it's also important to include Figure S1 in the main manuscript because it is directly relevant to the claimed achievements of SINTEX-F3; but, in fact, the CMCC hindcasts show better and more widespread skill being initialized with a sea ice state from the same reanalysis used to initialize SINTEX-F3. It would be more accurate to frame what's being done in the study as bringing the SINTEX system more up to speed with existing seasonal forecasting systems; even so, comparison with the other models should indicate that there is still room for improvement there.

- The impact of SIT initialization is also not thoroughly investigated in my view, since there is no experiment performed that only constrains SIC to the C-GLORS reanalysis for comparison. The benefit of F3-SI relative to the different F2 experiments and F3-SST experiment can come from constraining both the SIC and SIT to the C-GLORS reanalysis. The contribution from SIT on its own is only indirectly inferred from the regression analysis performed. While that analysis suggests SIT plays a role, the magnitudes of the regressions are hard to compare because the units are different depending on what variable is being regressed (see my more detailed point on this below). It is therefore not clear to me from this analysis if the memory from SIE or SIT is dominant (or other predictors for that matter). The authors should also better contextualize these results with other studies. The role of SIT as a source of predictability in the Barents-Kara sea was covered in Bushuk et al 2017 (see their Fig. 4). There are also relevant results pertaining to the sensitivity of constraining SIC vs both SIC and SIT (from reanalysis) for the Barents-Kara Seas in Dirkson et al 2017 (see their Fig. 6).

Bushuk, M., Msadek, R., Winton, M., Vecchi, G. A., Gudgel, R., Rosati, A., & Yang, X. (2017). Skillful regional prediction of Arctic sea ice on seasonal timescales. *Geophysical Research Letters*, *44*(10), 4953-4964.

Dirkson, A., Merryfield, W. J., & Monahan, A. (2017). Impacts of sea ice thickness initialization on seasonal Arctic sea ice predictions. Journal of Climate, 30(3), 1001-1017.

Regression Analysis & Interpretation

- There is not enough information provided on the regression analysis performed to adequately understand what was done. Is the "predictand" in the regression always the July-August-Sep SIE index from NSIDC data set, or is it sometimes also the same quantity from the various forecast experiments? This needs to be clarified.

- It is also not clear to me why the regression was done onto a normalized value, making the "predictand" unitless but the "predictor" unit-retaining. While this puts the regression value plotted in the units of the predictor, it is hard to compare magnitudes between different sources of predictability (like SIE and SIT). Even if the predictand is not normalized, it would be more straightforward to just compare results between different predictors using the correlation coefficient (as in the similar analysis performed in Bushuk et al 2017). This would allow you to make a more robust claim about the relative roles of SIE, SIT, and other predictors at the time of initialization to SIE skill at the various lead times. Still though, the actual impact is not knowable without doing an experiment where only SIC is constrained to the reanalysis.

- All of the figure captions also currently have the mistake of stating that the units of the predictors are [units / sigma] when they are actually just [units]. Since the predictand is normalized by \sigma, the units of the regression coefficient should just be the units of the predictor. I.e., the regression coefficient is:
  corr(X,Z) * (sigma_x / sigma_z), where X is the predictor anomaly and Z=Y/\sigma_y is the normalized SIE anomaly. The correlation, corr(X,Z), is unitless, sigma_x has units of "X", and sigma_z is unitless, making the regression coefficient have units of "X". Note that this error would be resolved if the regression coefficient was replaced by the correlation coefficient.

- Finally, the authors seem to misinterpret the sign of the regression coefficient, giving preference for a certain "direction" of the physical process being described. For

example, a negative value for the heat flux regression doesn't mean that the heat flux itself is negative, it means that the heat flux and SIE are negatively correlated, such that when one is larger than normal, the other is smaller than normal and vice versa. Also remember that what's being regressed here are anomalies, so it's just a description of the relationship of the two variables when they differ from normal; the original sign of the predictor (e.g. heat flux) is not retained in the regression.

Misunderstanding of the "spring predictability barrier"

- The authors seem to have misunderstood the spring predictability barrier discussed in previous literature. It in fact refers to an inherent predictability barrier that exists in spring, where forecast skill for the summer dramatically increases for forecasts initialized after June compared to forecasts initialized before. This "barrier" does not mean that sea ice is not predictable for initializations earlier in the spring, it just means that skill is notably lower than later start dates. When saying it's an "inherent" barrier, what I mean is that it exists even in perfect-model forecasts, where initial conditions (including SIT) are perfectly prescribed. That is in fact where it was first identified. What has been found, however, is that better initialization of SIT can make this barrier less extreme than without initialization. The authors should revise all mentions of this phenomenon and modify the phrases in the introduction and conclusions regarding the "overcoming" of this "barrier". The sentence in the conclusions that the "spring predictability barrier is not an intrinsic limit of the climate system" is factually incorrect and needs to be removed. There is a similar incorrect sentence in the intro. The points I have made here are all discussed and investigated thoroughly in Bonan et al 2019 and references therein:

Bonan, D. B., Bushuk, M., & Winton, M. (2019). A spring barrier for regional predictions of summer Arctic sea ice. *Geophysical Research Letters*, *46*(11), 5937-5947.

Specific comments:

Fig. 1b and throughout: the Barents-Kara seas region should be extended to the coastline, given that there is relevant skill outside the boundary considered in the Kara Sea.

Fig. 3a: I don't see the off-green curve for SINTEX-F3 in this plot. I only count 9 curves and there are 10 labels; it should be the same as the yellow curve in Fig. 3b right? Maybe also restate that the forecasts were initialized in May and that the curve shows the skill for the target month indicated (and/or add "target month" to the horizontal axis).

Figs. 4-6. Why is the 3-month period ASO included in these figures if the predictand is JAS sea ice extent and you're looking at what possible predictors lead to its skill? ASO includes the month October, which is after the 3-month period used to define the predictand, so why is it relevant? Please also state in each of these figures what you're regressing onto. Is it the observed normalized SIE from NSIDC or sometimes the forecast SIE? I assume it's the former but this should be stated explicitly.

Line 79-80: Give a hint to the reader here as to why you focus on the Barents-Kara Sea.

Section 2.1: Please add more information on how each of these data sources will be used (i.e., initialization, verification, predictability attribution, etc.). Also, much more information is needed on what is assimilated into C-GLORS related to sea ice and the ocean in the Arctic.

L149-150: Are the initial conditions truly just set directly to those from C-GLORS? Is there no nudging performed? I assume that there are strong shocks because of this. Also, it's confusing to say, "this leads to six additional ensemble members", since it sounds like you're creating different groups of members within the same experiment/configuration with different initialization strategies. Really, you're just describing different experiments/configurations and how many members are in each one, correct?

L233-235: I don't see how the subsurface ocean initialization disturbs the sea ice initialization if the sea ice initial conditions are simply set to C-GLORS; i.e. their values at initialization time don't depend on the subsurface ocean, do they?

L255-256: Should quickly restate here why you're doing the regression analysis.

Fig. 3d: Why is the regression larger for F3-SST and F3-3DVAR than that for F3-SI?

L321: In the caption, replace the period by a color ":" before "Regression values of the total meridional advection…" since you're listing the different components.

Section 3.3 I don't understand why you have started a new subsection here; how is this distinct from section 3.2?

L384-386: Please relax the certainty of this claim. E.g. could say that the initial SIT "appears to support" a more realistic ice-albedo feedback and … Also, should state what findings in the results lead to this conclusion. I.e. In section XX, we showed that … implying that SIT appears to support …

L391-393 and Fig. 9: I don't think Fig. 9 and the reference to it are appropriate here in the conclusions. It's relevant to state that the observation coverage is sparse, but this should be included in the methods section. I also don't think it's necessary to include Fig. 9 in the

main manuscript given the focus is on the Barents-Kara sea and there is such little coverage there. It should be just as easy to state as much without the figure.

L393: “constrains” to “limits”. The word “constrains” in this context usually means that the initial conditions are more tethered to observations, and you’re saying the opposite.

L400-403: Should avoid claims like “there is no denying the fact that …” in scientific writing. Also the Zeng et al 2023 study didn’t show that the atmospheric initialization improved sea ice initial conditions, they showed that including the surface heat budget in the initialization alongside SIC and SST initialization improves forecast skill in the summer. However, this seemed to be the case more for other regions, but not the Kara Sea in summer.

L410: Please soften this claim that "this study has large implications for the necessity of long-term sea ice observations data". The key results in this study are already established in the literature.