The manuscript describes the first multi-decadal simulation of a coupled sea ice-ocean model with neXtSIM as the sea ice component. As such it is the first coupled ice-ocean simulation with a brittle rheology. The model system is driven by re-analysis (ERA5) and very well tuned to observed large scale features (ice extent, volume, mean drift speed). The manuscript then describes a mass (volume) balance of Arctic sea in great detail. As a main result, new sea ice formation contributes 25-35% of the total ice growth in winter. This contribution grows over time, mostly consistent with many previous studies.

The manuscript is generally well written (see a few suggestions below) and has an easy to follow structure. The analysis appears to be very thorough and great care has been taken to map model results to observations (and not vice versa, as is often done). The results are interesting and warrant publication in TC.

There is one aspect that I find inconclusive and not supported by the presented results: The authors associate the new ice growth to the brittle sea ice dynamics that set neXtSIM apart from any other large scale sea ice model. While I have no doubt that most of the ice growth takes place in open water or over thin ice (also in this model), I cannot see how the heterogeneity (at the grid scale) of the ice cover that this model features is an essential ingredient to the analysis. For a balanced analysis the authors need to “couch” their work differently.

Any sea ice model that I am aware of uses the sub grid scale parameterisation of ice concentration to simulate unresolved leads. With a simple diagnostic that records new ice...
formation over open water or the thinnest ice class (or great ice), it would be possible to repeat the present analysis with a sea ice model without brittle rheology and I would not expect very different results (although I may be wrong).

To show that the heterogeneity of the ice cover is an essential ingredient that we need to get right, there needs to be a comparison of a model with heterogeneity and without (not clear to me how that can be achieved cleanly, maybe with extra averaging of the concentration fields before they are used for heat flux computations, etc.). The analysis of a single model simulation will only show that there is more ice production in areas of little or thin ice (which could be done with any model. In my opinion, the authors need to rephrase the corresponding parts of their manuscript or present clear evidence that supports their claims of this aspect.

More specifically:

page 1

I3: (second sentence of abstract) “These exchanges strongly depend on openings in the sea ice cover, which are associated with fine-scale sea ice deformations, but the importance of these processes remains poorly understood as most numerical models struggle to represent these deformations without using very costly horizontal resolutions”. This is a strong claim that is unsupported, because even coarse models have sub grid parameterisations (sea ice concentration < 100%) that allow finite exchange. I have not seen any evidence that on average (10-100km to basin scale) the effects of fine-scale sea deformation are important for, e.g. heat exchange in coupled models. For ocean models it is very unclear (I am not aware of any work in that direction, please prove me wrong); regional atmospheric models have been used to illustrate the effects of leads on vertical and horizontal mixing, but in coupled simulations, atmosphere models are too coarse to resolve the forcing by leads. If the authors are aware of evidence that supports their claim, it needs to cited here (or in the introduction).

page 11

I317: “The impact of heterogeneity of the sea ice cover on winter ice production is visible in Figure 11a, and is clearly linked to the growth of young ice (Figure 11b). …” (maybe it is a good idea to have different color scales for 11a and b to stress that one is growth and the other a fraction of total growth).

While I do not question the heterogeneity of the ice cover, it does not become clear from this analysis that the “openings” are important for the net volume changes (growth) in the model. It’s clear that ice can only grow over open water or thin ice. Sea ice models account for this by having at least 2 ice classes (thick ice and “thin ice including open
water”, see Hibler 79), most models have even more (e.g. 3 in neXtSIM or many more in CICE). This is the coarse resolution model’s parameterisation of leads. Heat fluxes and growth rates are computed separately for the individual classes. For this to work, the ice distribution does not need to be heterogenous and it doesn’t matter if the patterns “look similar to maps of observed ice divergence ...” or not.

page 14

l444: “The ability of the sea ice model to simulate fracturing and the subsequent sea ice deformations is used to assess the contributions of leads and polynyas to the mass balance.” This claim is not supported by the presented work. It does not become clear that the fracturing and deformation as simulated by the brittle rheology affect the contribution of leads and polynyas to the mass balance.

page 15

l453: “Our results illustrate the interest of using a brittle rheology framework in ice–ocean coupled modelling. This framework is able to capture the spatial and temporal heterogeneity of the ice cover, opening up the possibility to assess how this heterogeneity affects the ocean surface properties.”

I think that this again is overselling neXtSIM’s rheology. Heterogeneity at the grid scale should not be confused with realism. Further, grid point models always need a few grid points (order 5-10) to represent a feature. Very localised forcing at the grid scale may lead to local effects on the vertical mixing, which will then immediately be smoothed by horizontal processes. It is not clear if heterogeneous heat fluxes have a significant impact on mixed layer properties relative to smoothed heat fluxes. If there is evidence from the literature, please cite it.

Processes that involve thresholds, like biogeochemistry with minimal light requirements, the effect of heterogenous light conditions on net production, etc. are more plausible. Again, I have seen this claim a lot, without any proof or evidence from numerical modeling. Please cite the relevant literature.

L457-460: At higher horizontal resolution (which the authors have deemed too expensive earlier), non-brittle (VP) models also exhibit the heterogeneity (as cited in the introduction), so the advantage of neXtSIM does not become clear.
Data availability: All external (and open) data sources used in the study are listed, but availability of simulation data or code of this study is unclear.

Minor comments and suggestions:

page 4

l99: “The stress values are chosen to match the observed large scale drift and thickness distribution as well as possible, while still maintaining good deformation patterns and statistics.” It is not clear how this is done, what are “good deformation patterns and statistics”?

l106: I stumbled over “twice the ice model time step”, because a model timestep of 450 seconds would be short for VP model, but in the light of Plante et al 2020 or Dansereau et al 2016 (the only other ice models with brittle rheology that I am aware of), who both use timesteps of order 2 seconds to (marginally) resolve fast elastic waves, this seems like a very long time step for a elastic model. It would be useful to state here that the dynamics are solved with a much shorter time step (6s according to table1) to avoid confusion.

page 5

L127 “OPA-neX“ later “OPA-nex” is used (who a lower case “x“).

page 7

l186: “the internal stress is an important term in the Arctic mass balance”, that’s technically not correct. It may have an important effect on the mass balance, but it is a term in the momentum balance. Please rephrase.

l188: “adding a degree of freedom to the simulation”? Unclear, what this is supposed to mean. I would remove it.
“careful”, wording: I hope that everything reported here is based on “careful” analyses, so I would be careful with this adjective. I would replace it by something more descriptive, like “detailed”, “thorough”, if you really need to stress that you are “careful”.

“The evaluation of …” Could be much shorter, e.g.

“The evaluation of small-scale dynamics of sea ice in the coupled neXtSIM/OPA setup provided no qualitative differences in sea ice deformations compared to a standalone setup (Olason et al. 2021a).”

Fig3: no numbers? Mean difference? RMSD?

“look at”, colloquial, rephrase

Figure 4a and associated text (l206-214): Interannual variability is different from PIOMAS, notably the extreme minima in 2007, less so in 2012, are underestimated (not low enough). Instead, the OPA-nex timeseries tends to be more stable than the PIOMAS time series (less mean volume decrease and lower inter annual variability). Is this the ERA5 forcing or some model specifics/parameters?

page 8

“remarkable”: wording. I find this adjective not appropriate. This scientific MS should not be not about selling the results, and as sea ice drift is mostly determined by wind forcing, the agreement may not be as “remarkable” as claimed.

L219 “overestimated”, what is the reason for this overestimation when it did not happen before. Does the ice state change so that the ice becomes more mobile?

L223: “component” here I would use “term”, but that’s a matter of taste

L235: “agree very well” -> “agrees very well”, although I would replace statements like these as much as possible by more quantitative statements (“very well” can mean anything).
l248: “estimations” -> estimates? (also elsewhere)

l257: “significant”? Statistically significant? What is significant about this trend?

l261: “is not well captured by OPA-nex”, but for OPA-nex you can (and have done it) diagnose the individual terms of the model-thermodynamics, whereas Ricker et al could only indirectly estimate the thermodynamic growth. What happens if you use the same method as Ricker et al?

l263: “look at”, colloquial (also elsewhere), replace by “examine” or similar.

l264: “This allows us to explore more deeply the links between the dynamic and thermodynamic contributions to the Arctic mass balance.” How? Without explaining the “how”, this sentence makes little sense and could be dropped.

l304: Liu et al. (2020), please cite the numbers for the trends (or the range) for context.

l313: “To do this, we assume that, …” this assumption is not specific to small scales or brittle rheology. This could be done for any model that has a sub grid scale ice concentration (i.e. virtually all sea ice models) at coarse resolution. (See main comment)

l359: “look into more detail at”, rephrase

L414/415 “was made of” -> consisted of?
L531 Hutter et al has been published in JGR: https://doi.org/10.1029/2021JC017666.

Fig7 caption: “estimations” -> estimates