Author response to reviews of “Lagrangian reconstruction of snow accumulation and loss on Antarctic sea ice”

To the editor and referees assigned to this submission:

We thank the reviewers for the constructive input and thoroughness in their reports, which will greatly improve the quality of our study.

Please note that we are providing this same document to both referees, as some of our comments pertain to feedback from both referees.

In accordance with *The Cryosphere*’s review process and guidelines, we have not prepared a revised manuscript at this time. This Author Comment outlines our planned revisions regarding the issues raised by referees. Here we describe our intentions to amend or maintain existing text, update analyses where specified, and conduct additional model simulations. We intend to incorporate any new results into the manuscript.

We have divided this document into major issues and detailed responses. We distilled the primary concerns raised by both referees into five major issues, which we address individually with **key proposed changes highlighted using bold text**. We then respond in detail to the original text of the two referees’ comments (including those listed as “Minor comments” by Referee #2), with their text in blue and ours in black. Where they pertain to major issues, we refer back to our earlier comments. References are included at the bottom of this document.

We observe five major concerns raised by one or both reviewers:

- **Major Concern 1:** Snow-ice formation is: **(a)** not represented in the model, and **(b)** instead estimated with overstated confidence using a comparison with satellite snow depth retrievals. *(Referees #1 and #2)*
- **Major Concern 2:** Precipitation forcing for the model comes from ERA5 only, and other reanalysis products and the ERA5 ensemble are not tested. *(Referee #1 only)*
- **Major Concern 3:** The model is re-initialized annually to avoid model drift, but this prevents the model from retaining interannual memory in areas of multiyear ice. *(Referee #1 only; note minor comment from Referee #2)*
- **Major Concern 4:** The model calibration routine may not be robust to different subsets of snow buoys. *(Referees #1 and #2)*
- **Major Concern 5:** Treatment of superimposed ice in the modeled snow budget. *(Referee #1 only)*

Major Concern 1a: Lack of explicit snow-ice formation

We acknowledge that explicit snow-ice formation is not represented in our snow process model. Indeed, we are aware that adding snow-ice formation could have enabled the estimation of snow depth, a variable of key interest for the polar science community. Yet we have opted not to do so for several reasons expanded on below.

Overall, we believe that accurately representing snow-ice formation is not feasible without introducing a coupled ocean–ice–snow model configuration, which is beyond the scope of this study. This is because less complex approaches are insufficient for yielding realistic estimates of snow-ice conversion:

Possible approach #1: As adopted in the CASSIS snow model (Lawrence et al., 2024), one can automatically assign a fraction of new snowfall to snow-ice; the authors specify a value of 55%. While straightforward to implement, we do not believe this simplistic approach reflects the wide spatiotemporal variability in snow-ice fractions observed from in situ measurements (e.g., 7–58%; Lange et al., 1990; Jeffries et al., 2001). Therefore, we do not implement it.

Possible approach #2: Alternatively, one could try estimating snow-ice formation from hydrostatic balance using sea ice thickness (SIT) from a satellite-based product. However, this approach would be inappropriate because snow-ice is already "baked in" to SIT retrievals. For example, for retrievals using radar altimetry, the main radar scattering horizon occurs at or above the interface between snow and frozen snow-ice (Willatt et al., 2010), so the instantaneous (or monthly) retrieved SIT already includes an unknown proportion of snow-ice. New snow-ice formation therefore cannot be independently constrained from hydrostatic balance using the retrieved SIT and simulated snow accumulation. While a zero-ice-freeboard assumption could be used, this would yield inaccurate results for snow depth in most regions—especially those featuring thicker sea ice, such as the western Weddell and Amundsen-Bellingshausen sectors (Kacimi & Kwok, 2020; see their Fig. 10). Several other methodological challenges would arise: the monthly resolution of existing SIT products would require careful temporal interpolation to enable daily snow-ice formation estimates; the leading SIT estimates of Bocquet et al. (2024) rely on climatological snow depths, which would introduce an inconsistency with our model; and this overall approach would reproduce the existing subsumption of WASSAIL model errors into the snow-ice estimates.

Possible approach #3: Lastly, one could explicitly simulate sea ice growth and thus estimate snow-ice formation along our Lagrangian parcel trajectories using hydrostatic balance. This would be similar to the approach of Arndt et al. (2024), who coupled SNOWPACK to a simple thermodynamic ice growth model for the western/northern Weddell Sea. This is a worthy effort and, we believe, the only rigorous solution to this problem. However, it would require incorporating (at minimum) a 1-D sea ice model, which introduces several issues.

First, this is a complex endeavor due to the strong and poorly constrained influence of spatially variable ocean–ice heat fluxes across the Southern Ocean, which exhibit at least a five-fold range, up to 50 W m^{-2} in some regions (Wilson et al., 2019). The heat fluxes (and hence ice growth rates) closely depend on mixing-induced ocean–ice feedbacks that cannot be reliably simulated unless a 1-D ocean model is additionally incorporated. The snow-sea ice coupled model approach was largely feasible in Arndt et al. (2024) because the study focused on a limited region of the Southern Ocean and thus adopted a constant ocean heat flux of 3 W m^{-2}, although even this is a significant simplifying assumption. In their study, when the ocean heat flux was doubled from 3 W m^{-2} to 6 W m^{-2}, a 6-cm change occurred in simulated snow depth for one buoy (note, no other buoys were tested and, as Referee #1 has pointed out, there is substantial variability in buoy point measurements). While coupled ocean–ice–snow modeling could potentially address this challenge, we regard it as beyond the scope of this study.

Additional uncertainties with a coupled model approach include whether snow-ice formation can be adequately parameterized using the hydrostatic assumption (e.g., flooding and snow-ice formation occur in the same time-step when the sea ice freeboard is negative), as observations suggest this process occurs on much longer timescales (e.g., Provost et al., 2017), and, sometimes, snow-ice does not even form after flooding. As such, the hydrostatic assumption may overestimate flooding and snow-ice formation (please see Wever et al., 2021).

More critically, adding coupled ice–ocean modeling to the snow process model would merit concomitant validation, requiring, ideally, coincident in situ observations for sea ice and ocean properties. Considering the limited number of coincident snow, sea ice, and ocean observations in the Southern Ocean, carrying out this approach would introduce an even larger number of uncertainties and biases, going against the spirit of Referee #1's review. Accordingly, we believe the extensive effort described here is better served by a separate future study, which we would warmly welcome in a collaborative effort with Referee #1. We are hopeful that future observational efforts, such as Antarctic InSync, will be particularly valuable in this regard.

For reasons above, we will not parameterize snow-ice formation in the model for the revised manuscript. While our results yield a budget for meteoric freshwater intercepted by Antarctic sea ice rather than a complete budget of snow depth, we believe this focus is novel, appropriate for constraining the represented snow mass budget terms, and relevant to the readership of *The Cryosphere*.

Although this is mentioned in the current manuscript (Line 186, Figure 2, Line 587, Lines 608-610, Figures 6 and 7, Lines 711-712, Lines 745-747, Line 763, and Lines 784-788) and reflected in the title (which refers to a reconstruction of "snow accumulation and loss" rather than "snow depth"), **we will add text at the following locations that further clarifies the omission of snow-ice formation in our methodology and the significance of estimated freshwater fluxes** (in addition to the updates to Section 3.5 described in Major Concern 1b):

- Abstract, Line 18: "*While our model does not explicitly represent snow-ice*, comparison with satellite snow depth retrievals suggests that…"
- Section 1 (Introduction), Lines 78-79: "... which integrates snow accumulation, erosion, and transformation mechanisms (Sect. 2.3)*, albeit without representing snow-ice formation*."
- Section 3.2 (Climatological snow mass budget), Line 594: "*This model budget term includes snow incorporated into snow-ice, due to the lack of simulated snow-ice formation in the model, and is therefore best interpreted as a combined meteoric freshwater flux. It excludes seawater incorporated into snow-ice. Interannual variability in snowfall and freshwater release…*" The similar text on Lines 608-610 will be deleted, as this phrasing and location is clearer.

- Figure 6 caption: "... lead trapping and *meteoric freshwater* release upon sea ice melt, which includes snow *incorporated into* snow-ice."
- Figure 6: The legend for the red line and the label for the red shaded region will be amended to "*Meteoric freshwater* release to ocean *upon ice melt* (includes *snow in* snow-ice)".
- Figure 7 caption: "... red represents *release of meteoric freshwater* to the ocean, with panel [j] including snow *incorporated into* snow-ice."
- Figure 7j: The panel title will be amended to "*Meteoric freshwater* released to ocean upon sea ice melt".
- Section 4 (Conclusions), Lines 784-788: A summary of the rationale above regarding SIT retrievals and sea ice modeling will be added.

Major Concern 1b: Snow-ice formation estimates

We appreciate both reviewers' perspective that our inference of snow-ice formation from the model–satellite difference is described with overstated confidence and may not be robustly supported by the methodology. **We will both de-emphasize and further contextualize this result in the revised manuscript**. We intended this to be a back-of-the-envelope estimate, and it will remain a small portion of this overall work, currently limited to just three paragraphs in Section 3.5 and described using cautious language in the Abstract ("comparison… suggests…"). We are admittedly confused by Referee #1's suggestion that it was elevated in the original manuscript to "a demonstrated retrieval".

Following both reviewers' recommendations, however, **we will moderate our existing claim and add text in Section 3.5** emphasizing that the snow-ice estimates are not necessarily direct evidence of the phenomenon but rather an approximate amount of potential snow-ice formation consistent with the best evidence to date.

We appreciate Referee #1's enumeration of additional contributions to the model–satellite residual: satellite retrieval error (Major Concern 1b), annual re-initialization without model memory (Major Concern 3), buoy representativeness (minor) and sensitivity during calibration (Major Concern 4), omission of liquid-water processes (Major Concern 5), saline-snow processes (minor), and forcing uncertainty (Major Concern 2). In Section 3.5, **we will add further text underscoring these sources of uncertainty**, noting that these terms may contribute biases of compensating sign to the inferred snow-ice formation:

- *Satellite retrieval errors*: anticipated source of positive bias in inferred snow-ice, as the AMSR-E/2 and CryoSat-2 retrievals both likely underestimate snow depth, as described in Sections 2.1.2 and 2.2.2 (both to be updated; see responses to Referee #2 under "Minor comments")
- *Annual re-initialization without model memory:* anticipated source of negative bias in inferred snow-ice within areas of multiyear ice, where AMSR-E/2 is used to initialize model snow depth
- *Calibration uncertainties:* anticipated source of negative bias in inferred snow-ice, if the negative bias in model validation (Section 3.1) remains after calibration updates described in Major Concern 4

- *Precipitation forcing:* contributes error of unknown sign, though this should be mitigated by the scaling of ERA5 precipitation in calibration
- *Other model and input data errors:* contributes error of unknown sign

Note that our **proposed changes to Fig. 8 to unify the time periods used for the CryoSat-2 and AMSR comparisons with the model** (see our response to Referee #2 under "Minor comments") will remove one notable source of error in the snow-ice formation inference.

Major Concern 2: Reanalysis precipitation forcing

We thank Referee #1 for their encouragement to estimate the uncertainty in precipitation forcing and propagate it through to the overall freshwater flux estimate. In response to their recommendation, we have carefully considered options for doing so. However, we have determined that it is not possible to quantitatively constrain the uncertainty in reanalysis snowfall at present. **We intend to expand the discussion on precipitation uncertainty in Section 2.1.4 with additional details**, as described below.

Uncertainty estimates from ERA5 ensemble: As noted by Referee #1, ERA5 provides a 10-member perturbed ensemble (Ensemble of Data Assimilations, EDA) at a reduced time and space resolution. The ensemble incorporates random errors in the global observing system, sea surface temperatures, sea ice concentration, and model parameterizations (Hersbach et al., 2020). In response to Referee #1's suggestion, we conducted a preliminary analysis of the spread in cumulative snowfall over one year in the ERA5 EDA (**Figure A**, below).

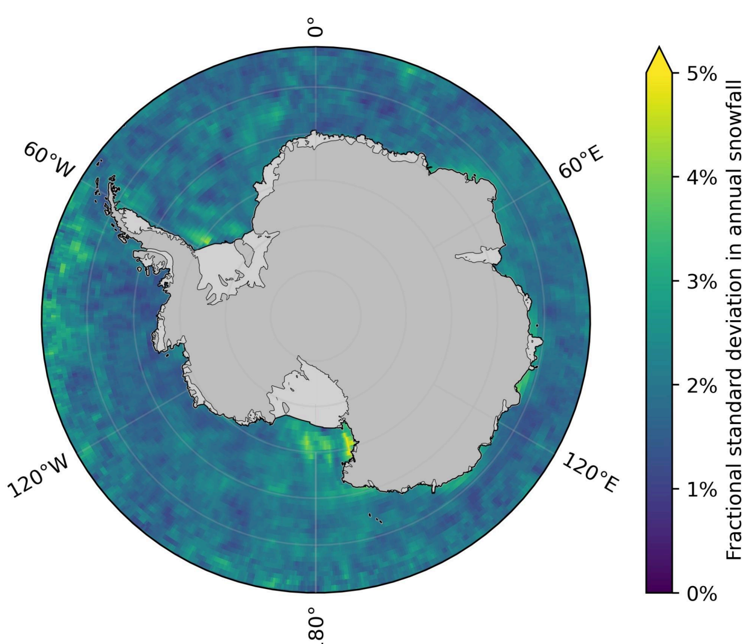

Figure A: *Example showing the standard deviation in annual cumulative snowfall in 2020 across the 10 members of the ERA5 Ensemble of Data Assimilations (EDA), expressed as a percentage of the average annual snowfall.*

We obtain a fractional standard deviation across the 10 ensemble members of approximately 1–3% in annual snowfall over the ice-covered Southern Ocean. We believe this is an unrealistically low and potentially misleading error estimate for the following reasons: (1) the ERA5 EDA is known for being underdispersive; (2) ECMWF notes that the "EDA system was not actually designed for uncertainty estimation" but instead offers "a *relative* estimate of uncertainties in terms of spatial and temporal distribution", because it "mostly account[s] for random errors and *not for systematic ones*" (emphasis added); and (3) a spread of 1–3% is significantly less than the differences between available reanalysis products (see below). For these reasons, we do not use the ERA5 EDA to generate propagated uncertainty estimates. Please see ECMWF's ERA5 EDA documentation for more details: https://confluence.ecmwf.int/plugins/viewsource/viewpagesrc.action?pageId=111158117.

Different reanalysis products: While we mention in Section 2.1.4 that Boisvert et al. (2020) find "a spread of ~200 mm per year in snowfall magnitude across five [reanalysis] products", we now believe our use of this statistic was an overestimate of the relevant uncertainty. **We will add to Section 2.1.4 a summary of the following clarifying details from Boisvert et al.:**

- Boisvert et al. find that ERA5, MERRA, and JRA-55 exhibit the highest fidelity to CloudSat retrievals of snowfall from 2007–2010 over the Southern Ocean, while MERRA-2 and ERA-Interim produce amounts of cumulative snowfall that deviate markedly from CloudSat. For this reason, we consider ERA5, MERRA, and JRA-55 potentially usable but believe that MERRA-2 or ERA-Interim should not be included in a meaningful uncertainty estimate.
- Of the three best-performing reanalysis products, their differences from CloudSat within the seasonal ice zone (SIZ) are generally <100 mm SWE (ignoring areas north of the sea ice maximum), which is approximately 25% of the Southern Ocean-wide annual average snowfall of 370 mm SWE for ERA5, MERRA, and JRA-55 (Fig. 7 and Table 3 in Boisvert et al., 2020). We believe this approximate estimate of ±25% is the most relevant uncertainty statistic that can be provided for ERA5. That said, Boisvert et al. note that even CloudSat is uncertain and cannot be regarded as a perfect source of ground truth.
- However, differences *between* ERA5, MERRA, and JRA-55 appear notably smaller (<50 mm SWE within the SIZ) and exhibit a similar spatial pattern of positive and negative anomalies from CloudSat (Fig. 5 in Boisvert et al., 2020). This suggests that structural biases are shared across the best-performing reanalysis estimates. Annual snowfall over the Southern Ocean from ERA5 and MERRA differs by only 27 mm (7%, from their Table 3; note this includes areas without ice cover), while the difference between ERA5 and JRA-55 is much larger, 107 mm (29%). MERRA production ended in 2016, and therefore only ERA5 and JRA-55 would be available for forcing WASSAIL. Using a two-member reanalysis "ensemble" of ERA5 and JRA-55 is unlikely to generate a more meaningful estimate of uncertainty in accumulated snow on sea ice than the analysis already conducted by Boisvert et al. (2020).
- Critically, we calibrate WASSAIL to the snow buoy observations and apply the scaling coefficient γ_{new} to reanalysis snowfall. After this calibration, uncertainty estimates reflecting differences across reanalyses would no longer be an accurate description of the uncertainty in model-estimated freshwater fluxes.

Lastly, we regard our current approach as reasonable given that past work has similarly used a single reanalysis product for snow modeling purposes: CASSIS (Lawrence et al., 2024) uses only ERA5, Maksym and Markus (2008) used only ERA-40, and Kwok and Cunningham (2008) used a single ECMWF product, for example, as did a 1-D snow-ice modeling study forced by ERA5 atmospheric fluxes (Arndt et al., 2024). In contrast, NESOSIM (Petty et al., 2018) and Blanchard-Wrigglesworth et al. (2018) tested multiple reanalysis products, but their models are less computationally expensive to calibrate and run for additional realizations. We note that the studies introducing SnowModel-LG for snow on Arctic sea ice (Liston et al., 2020; Stroeve et al., 2020) required the separate, second paper by Stroeve et al. to assess sensitivity to the choice of reanalysis product; they used only two reanalyses (MERRA-2 and ERA5).

Major Concern 3: Interannual memory

We appreciate Referee #1's recommendation to run multiyear simulations without annual re-initialization in order to retain interannual memory of simulated net snow accumulation, snow density, and superimposed ice thickness. We agree that this would be ideal. In response, we have implemented this capability into the WASSAIL model code and are exploring the feasibility of adopting this approach. **If it yields acceptable results, we will switch from annual re-initialization to using exclusively multiyear simulations**.

A key challenge is the 15% annual budget imbalance (Figure 6), which we believe originates from the initialization of multiyear sea ice (MYI) parcels using AMSR-E/2 snow depth on 15 February of each year. It is possible that a continuous, multiyear simulation could produce significant model drift in snow properties in areas of multiyear ice, particularly simulated bulk density. **If adopting multiyear simulations ultimately degrades the reconstruction quality, we will maintain the current approach but detail this finding in Section 2.3** (Lines 225-229). We will also mention the fractional sea ice area that is affected by loss of interannual memory in the model; we estimate this at ~23%, based on the average MYI area at the annual sea ice minimum over 1979–2024 (**Figure B**) divided by the annual mean sea ice area.

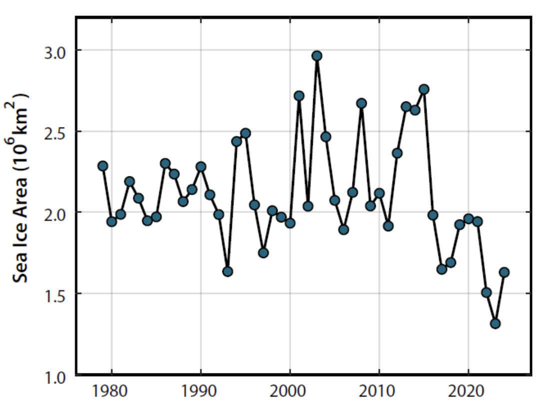

Figure B: *Area of Antarctic sea ice that is 1+ year old over 1979–2024, based on the annual sea ice minimum in Bootstrap sea ice concentration data. Reproduced from Webster, Arndt, Bliss et al., "Changes in Arctic and Antarctic sea-ice properties and processes", Nature Reviews Earth and Environment, in press.*

That said, we dispute Referee #1's assertion that the model "does not represent… interannual memory on [MYI]", which implies a total loss of interannual memory, i.e., resetting snow depth to zero on 15 February of each year. We emphasize that our model does not do this. MYI parcels are reset to the AMSR-E/2 snow depth, which means they retain a nonzero snow depth—though not the model-generated net accumulation from the previous year.

Lastly, we note that most past work has taken similar approaches to our study in regards to multiyear simulations:

- CASSIS (Lawrence et al., 2024) similarly uses annual simulations of snow on Antarctic sea ice that re-initialize each year on 20 February, except using climatological snow depths based on ASPeCt observations rather than interannually varying snow depths, as in our work.
- NESOSIM (Petty et al., 2018) uses only partial annual simulations (summer to spring) of snow on Arctic sea ice that re-initialize each year on 15 August using snow depths from an observational climatology from 1954–1991, scaled linearly using the summer melt duration of the given year.
- Lagrangian snow reconstructions for Arctic sea ice by Blanchard-Wrigglesworth et al. (2018) and Kwok and Cunningham (2008) as well as the Eulerian reconstruction for Antarctic sea ice of Maksym and Markus (2008) re-initialize simulations annually.

The exception is SnowModel-LG (Liston et al., 2020), which uses multiyear runs initialized in 1980, albeit with an assumption of zero snow depth in summer.

Major Concern 4: Calibration robustness

We agree with Referees #1 and #2 that the calibration routine may be sensitive to the particular 75%–25% split of snow buoys used as the observational reference. In response, we have conducted exploratory testing using different random splits of buoys. The calibrated model parameters indeed differ enough that we concur a more robust approach is warranted. We appreciate Referee #2's comment that successive halving remains an appropriate choice for the optimization algorithm.

We intend to update the calibration methodology by running successive halving repeatedly using different random buoy partitioning between training and validation sets, a technique known as Monte Carlo cross-validation (MCCV), then averaging the resulting parameter sets. We anticipate that MCCV will not only make the final parameters more resilient to overfitting but will also likely improve the model bias compared to the validation set of buoys. This is because the random split used in the paper was a particularly imbalanced one, as noted in Section 3.1, Lines 540-541 ("... suggesting that unexpected differences within the randomly split buoy data can explain much of the calibrated model's difficulty reproducing the validation buoy set"). **We will update Figure 3 to show the optimization evolution across the ensemble of random splits.** An added benefit of this updated approach will be new uncertainty bounds on the eight model free parameters—including γ_{new}, the scaling coefficient for ERA5 snowfall.

Major Concern 5: Superimposed ice

We are confounded by the concerns raised by Referee #1 that superimposed ice is not represented in the model. This is incorrect. Superimposed ice formation from snow melt processes is already explicitly represented in the model, and the amount released to the ocean annually is shown in Figure 7h. Please see Lines 211-212, 217-218, 234-236, 242-251, 309-313, 448-469, and 471-473 and Table 1 of the original manuscript for descriptions of how superimposed ice is treated in the model.

Responses to Referee #1

This manuscript addresses an important and timely problem: how snow on Antarctic sea ice modulates the Southern Ocean freshwater budget and the sea-ice mass balance. The topic is clearly worth publication-quality study. However, in its present form, the paper overstates what its modeling framework can support. The central claim, that the reconstruction implies widespread snow-ice formation consuming roughly 49–60% of remaining snow and yielding an annual meteoric freshwater input of 237 mSv, more than half the circumpolar freshwater flux from sea-ice melt, is not robustly demonstrated given the model design, forcing uncertainty, observational representativeness issues, and the authors' own stated omissions. The manuscript itself explicitly acknowledges that it does not represent snow-ice conversion, meltwater/liquid-water effects, or interannual memory on multiyear ice, yet those omissions directly target the same processes the paper later elevates into headline conclusions.

Please see our responses above to Major Concerns 1a and 1b regarding snow-ice formation and Major Concern 3 regarding interannual memory.

While snow melt processes and superimposed ice formation are included in the model, as mentioned in our response to Major Concern 5, we do acknowledge on Lines 341-344, 585-586, and 782-784 that other liquid-water effects on snow metamorphism are not explicitly represented. This is an important limitation. Yet some of these effects may be incorporated through the tuning parameters, and the reconstruction seems to reproduce the contour of snow ablation events well over multi-day timescales (e.g., **Figure C** below) despite not being a full hydrological model of the snowpack. We note the omission of liquid-water processes should not significantly impact our "headline conclusion" regarding meteoric freshwater input to the ocean.

The most serious problem is a mismatch between the simulated quantity and the inferred quantity. The model is calibrated against snow-buoy surface-height-change records, which the manuscript treats as a closer analogue to snow accumulation because basal snow-ice conversion is not represented. But the paper then compares this reconstructed accumulation to satellite snow depth products and interprets the residual largely as snow-ice formation. That residual also contains biases from satellite retrieval error, annual re-initialization, multiyear-ice memory loss, buoy representativeness bias, the omission of liquid-water and saline-snow processes, and forcing uncertainty. In other words, the paper turns an underconstrained residual into a physical estimate of Antarctic-wide snow-ice conversion. That is not convincing as presently shown.

We are taking steps to explicitly address the issues the reviewer raises here: we are placing more care in our interpretation of the residual as snow-ice formation and expanding our discussion of satellite retrieval biases (Major Concern 1b), carrying out new multiyear simulations to preserve interannual memory that is otherwise lost during annual re-initialization (Major Concern 3), revising our calibration routine to increase robustness to the buoy partitioning (Major Concern 4), and expanding our discussion of precipitation forcing uncertainty (Major Concern 2). As described in the response to Major Concern 5 above, the model already represents superimposed ice formation.

As with liquid-water metamorphism, we do not include explicit saline-snow processes in the model. While we acknowledge these processes could be potentially impactful for snowpack characteristics, we have opted to avoid adding additional complexity to the model since there is an insufficient number of observations to develop a well-constrained parameterization of saline snow processes. It is an interesting idea though.

A second major weakness is that the regions most central to the manuscript's calibration and strongest claims, especially the Weddell Sea, are precisely where perennial and multiyear processes matter most. The draft states that parcels are initialized every 15 February and then reset "without memory of the prior year," and that this "limits the model's influence on snow over multiyear sea ice." It also initializes pre-existing snow density to a spatially uniform 320 kg m^{-3}. Yet independent Weddell literature shows that older/perennial ice in the western Weddell carries distinctive snow and meteoric-ice signatures, that Snow Buoy records often represent second-year and even 2-to-4-year snow regimes, and that regionally adjusted densities differ materially between perennial and seasonal ice. This means the initialization strategy likely imprints Figure 8f, weakens autumn conditions, and undermines the western/northern Weddell comparisons.

We agree that multiyear free-running simulations would be ideal, as mentioned in our response to Major Concern 3. We are carrying out these simulations and will evaluate their effects on snow accumulation, snow density, and superimposed ice thickness in areas of multiyear ice.

We emphasize that the calibration simulations along snow buoy trajectories in the Weddell Sea are already multiyear by design, as some buoys survive for 2+ years and no annual re-initialization (or loss of memory) occurs in these simulations.

Third, the forcing and validation strategy is too fragile for the paper's quantitative claims. The manuscript uses ERA5 snowfall as the core depositional forcing, while also noting a spread of about 200 mm yr^{-1} among Southern Ocean reanalyses and then tuning snowfall upward by 32% during calibration. ERA5 itself provides uncertainty information through an ensemble, but the draft does not propagate forcing uncertainty through to the headline freshwater-flux estimate. Similarly, the passive-microwave product used for both annual initialization and later evaluation is subject to dry-snow-only applicability and a 50 cm retrieval limit, while comparisons to Antarctic observations show substantial biases for older ice regimes. CryoSat-2 products are likewise affected by saline basal snow, radar-penetration uncertainties, and sparse validation. These are not minor caveats. They are central to the interpretation of Figure 8 and to the inferred snow-ice conversion.

The reviewer makes a good point in that atmospheric forcing is one of the key factors contributing to uncertainties in snow process model simulations. We address uncertainties in ERA5 snowfall forcing in our response to Major Concern 2. We appreciate their suggestion to explore the use of ERA5's perturbed ensemble for an uncertainty analysis, which we have assessed in **Figure A** above. We have proposed additions to the text to provide readers with greater clarity on the accuracy of ERA5 precipitation.

In regards to the inferred snow-ice conversion, please see our response to Major Concern 1b. While we appreciate the reviewer takes issue with these comparisons to AMSR-E/2 and CryoSat-2 snow depth retrievals, it is important we retain them, with an expanded discussion of their caveats, given that they are among the best pan-Antarctic set of snow depth retrievals currently available. We note that the limitations mentioned by the reviewer concerning the satellite retrievals are highlighted directly in the original text (Lines 110-116, 195-199; also see added text in response to Referee #2, under their "Minor comments").

My recommendation is **reject in present form**, with encouragement to resubmit only after substantial redevelopment. The work is promising, but the current paper makes quantitative claims that outstrip the demonstrated reliability of the method.

Major methodological weaknesses

The most important scientific omission is the absence of explicit meteoric-ice formation pathways from the prognostic snow budget: snow-ice formation from flooding/seawater infiltration, superimposed ice formation from meltwater refreezing, and the accompanying salinity effects. The manuscript acknowledges that snow-ice conversion is not represented explicitly and later estimates it from model-minus-satellite differences. That is a methodological shortcut, not a demonstrated retrieval. Independent Weddell Sea studies published in Arndt (2022), show that these omitted processes are not peripheral. In the northwestern Weddell, superimposed ice and snow ice averaged about 0.11 ± 0.11 m and 0.22 ± 0.22 m, respectively, together amounting to 3%–54% of total ice thickness. A broader Weddell Snow Buoy/SNOWPACK analysis found mean maximum snow-ice thicknesses of 16 cm across all analyzed buoys, 35 cm on snow-ice-bearing floes, and about 27% of snow accumulation converted to snow ice, with much larger values possible regionally. Those are order-of-magnitude effects for a snow-depth study, not second-order corrections.

We agree with the reviewer that snow-ice formation is an important process in the snow mass budget. As we detail in our response to Major Concern 1a above, the exclusion of snow-ice formation in the model is a result of careful thought on what process-oriented information is available to construct robust parameterizations. To explicitly represent snow-ice formation, information on sea ice thickness is required and it is that very information that is lacking for Antarctic sea ice.

With regard to superimposed ice formation, which is included in the model, please see our response to Major Concern 5 above.

This omission also matters specifically for the paper's freshwater interpretation. Snow ice forms from seawater-soaked snow and is therefore saline; superimposed ice forms from internal melt/refreezing and changes both snow depth and snow thermal properties. The manuscript argues that the "mass budget of total meteoric freshwater" should remain robust despite uncertainties in conversion processes. I do not think that is shown. The partitioning, timing, vertical location, salinity signature, and ocean-coupling consequences of the freshwater pathway depend directly on whether snow remains snow, becomes snow ice, becomes superimposed ice, or is redistributed into ridges or leads. In Antarctic conditions, those pathways are central to the physical meaning of the freshwater flux, not a cosmetic post-processing issue.

Our estimation of meteoric freshwater fluxes to the ocean is a combination of two reservoirs: snow and snow that has been incorporated into snow-ice. We detail new clarifications to the text and figures above in our response to Major Concern 1a that emphasize this accounting, which we firmly believe is robust for meteoric freshwater. Note that seawater incorporated into snow-ice and brine wicking into the snowpack are not part of the meteoric freshwater budget.

We reiterate that superimposed ice formation is included in the model and accounted for separately in the analysis, as shown in Fig. 7h; please see our response to Major Concern 5 above.

Redistribution of snow into leads and the associated freshwater fluxes are also explicitly represented in the model; see Section 2.4.6, Fig. 2, Fig. 5b, Fig. 6b, and Fig. 7i.

For the reviewer's suggestion that snow deposition in ridges should be represented in the model, we respectfully disagree as the snow model is not coupled with a sea ice model and satellite-derived ridge fractions are unavailable for the full period of interest. However, we have noted the importance of ridges and small-scale redistribution processes in multiple sections in the original manuscript, including on Lines 408-410, 415-418, 450-453, 571-573, 657-659, and 805.

A closely related weakness is the neglect of liquid-water effects, including melt metamorphism, infiltration, enhanced destructive metamorphism, reduced compactive viscosity in wet snow, and melt-freeze evolution. The manuscript explicitly states that these effects are not represented and that rainfall and snowmelt are assumed to refreeze rapidly. Yet Antarctic observations and process studies show that internal melt-freeze cycles, superimposed ice formation, saline basal layers, and wet-snow metamorphism materially affect snow depth, density, radar penetration, and surface energy balance. In the Weddell Sea, internal melt-freeze cycles and superimposed ice are latitude dependent, and recent observations show superimposed-ice thicknesses up to and above 0.1 m regionally, with associated changes in conductive properties and internal ice melt. Under those circumstances, using a dry, single-layer framework to infer continent-scale snow-ice conversion from residuals is too optimistic.

While we agree with the reviewer that all of the small-scale processes noted above are important for snowpack characteristics and melt processes, adding more complex and often poorly-constrained physics to a model does not necessarily lead to improved model simulations. There is a practical limit on what physics should be represented in a model, which is determined by the purpose of the model. For example, in the Los Alamos Sea Ice Model (CICE), snow-ice formation is readily simulated and used to inform the scientific community of the potential large-scale changes that occur under a warmer climate (e.g., Holland et al., 2021). Only recently has CICE begun to incorporate even some of the small-scale liquid-water processes described above due to computational constraints and a lack of well-validated parameterizations, yet the model fits its purpose in that it is a necessary tool for understanding snow in the sea-ice system. For the WASSAIL model, its purpose is to quantify the large-scale fate of meteoric freshwater on Antarctic sea ice and, in particular, to constrain the partitioning of snowfall between accumulation, sublimation, lead trapping, and melt. Its purpose is not to monitor the micro-evolution of individual deposition layers nor to track percolation channels and blockage within the snowpack. Other models are better suited for those purposes and explicitly represent many of the processes listed by the reviewer.

The annual re-initialization strategy is another major liability. The manuscript states that parcels are initialized every 15 February from AMSR-E/2 snow depth, with pre-existing snow density set to 320 kg m^{-3}, and then reset the next year "without memory of the prior year." The same section concedes that this limits the model's influence on snow over multiyear sea ice, and later the paper acknowledges a residual budget drift reflecting net depletion of snow originating from multiyear sea ice. This is especially problematic because western Weddell multiyear/perennial ice is one of the few Antarctic sectors where older ice is common and thick, and because Snow Buoy analyses in the Weddell indicate that many records represent second-year ice and some represent 2-to-4-year snow regimes. A model that forgets prior-year snow and meteoric-ice history cannot be treated as fully credible in exactly the sector where its validation story is strongest.

We agree that multiyear free-running simulations that retain interannual memory would be ideal and are exploring their feasibility; please see our response to Major Concern 3.

That logic also likely explains the suspiciously homogeneous autumn snow-density field in Figure 8f. If snow density is initialized everywhere to 320 kg m^{-3} in mid-February and only later evolves through simplified wind and overburden parameterizations, then the relatively uniform autumn density structure is at least partly an initialization artifact. That would be particularly important in the Weddell Sea, where regionally adjusted perennial-versus-seasonal snow densities of 340 and 264 kg m^{-3} have already been used in the literature following Arndt (2022a), precisely because older/perennial and seasonal ice regimes differ materially. The paper should not present autumn density heterogeneity as an unconstrained emergent result when the initial condition is spatially uniform.

We acknowledge that resetting snow density to 320 kg m^{-3} in mid-February prevents a realistic density evolution in areas of multiyear ice. While this only impacts ~23% of the Antarctic sea ice area, the reviewer's concern will be directly addressed through multiyear simulations (see Major Concern 3).

Compared to point-scale density measurements, the modeled results may seem homogeneous. When accounting for the model grid scale and domain, however, we respectfully disagree with the reviewer that the autumn snow density field is homogeneous. The average density fields in all seasons exhibit clear and often substantial circumpolar variations (Fig. 8f), reflecting the fact that there is no spatially uniform model initial condition for density within areas of first-year ice—approximately three-quarters of the Antarctic sea ice area. Furthermore, in all areas, density evolves dynamically based on the parameterizations of wind-enhanced compaction, which uniquely determines snow density at the time of deposition based on the first 100 hours of wind speeds experienced (Section 2.4.2), and mechanical compaction, which occurs gradually over time (Section 2.4.1).

Data sources and validation

The ERA5 snowfall forcing is a critical uncertainty that the manuscript underplays. The draft itself notes that Southern Ocean reanalyses agree in broad spatial patterns but differ by roughly 200 mm yr^{-1} in snowfall magnitude. By simple conversion, that corresponds to roughly 0.6–0.8 m of annual snowfall depth at densities of 340–264 kg m^{-3}, before compaction and losses, a very large perturbation relative to the climatological snow depths discussed in the paper. The authors then tune snowfall deposition upward by 32%, which strongly suggests compensation between forcing bias and process tuning. ERA5 is an excellent product, but it is still a reanalysis with uncertainty information that should be exploited, not treated as essentially fixed. Studies comparing reanalysis snowfall forcings have shown that comparatively modest snowfall differences can materially alter modeled snowpack and sea-ice evolution.

We agree that precipitation forcing is a key source of uncertainty. Please see the details in Major Concern 2 clarifying its magnitude (which we now believe is substantially less than the 200 mm yr^{-1} estimate cited above), our assessment of precipitation error within the ERA5 perturbed ensemble (the "uncertainty information" that the reviewer is referring to; see **Figure A**), and our planned clarifications in the text.

We are confused by the reviewer's concerns about "compensation between forcing bias and process tuning" and treating ERA5 precipitation as "essentially fixed", neither of which is reflected in the manuscript. The model calibration procedure is explicitly designed to correct for biases in precipitation forcing by applying a scaling coefficient to snowfall, $\gamma_{new} = 1.32$ in the current manuscript (see Section 2.1.4, Section 2.4.2, and Lines 553-559 in Section 3.1).

The snow-buoy treatment is also not persuasive enough for a circumpolar calibration framework. The manuscript is aware that Snow Buoys measure changes in the surface interface relative to a deployment baseline, and it explicitly uses that fact to argue that buoy data are a better analogue for modeled snow accumulation than for true snow depth. But the paper does not adequately grapple with the representativeness cost of that choice. The AWI Snow Buoy methods paper warns that time series from level ice may not represent the overall snow-mass distribution, notes that snow in ridges can be substantially deeper than on level ice, and shows that snow-ice and superimposed-ice formation can change the pure-snow layer without much change in measured surface elevation. A later Weddell analysis goes further and states directly that Snow Buoys are usually deployed on level ice and may therefore underestimate snow accumulation and snow-ice formation relative to the broader floe and ridged pack. So the calibration data are neither true snow depth nor clearly representative of grid-cell snow mass. That is a serious observation-operator problem, not just random noise.

We recognize that snow buoys are not wholly representative of all snow conditions on sea ice. To address the reviewer's concern, we will expand the discussion in Section 2.2.1 to highlight the caveats of using buoy data due to these sampling limitations, citing the AWI snow buoy methods paper by Nicolaus et al. (2021). As that study notes, it is not yet possible to accurately correct for this potential underestimation bias from snow buoys, given the lack of spatiotemporally-resolved information on the surface fraction of ridged Antarctic sea ice. There are no alternative methods for calibrating a snow process model on a scale as large as the snow buoy observational network.

In situ snow surveys are equally fraught with biases for model validation due to their tendency to avoid heavily deformed ice, and airborne surveys are fraught with retrieval biases and uncertainties. As a consequence, we respectfully retain the methodology but follow the reviewer's thoughts in being more direct about the buoy data caveats.

Nonetheless, the fact that "snow-ice… can change the pure-snow layer without much change in measured surface elevation" is actually advantageous for our purposes, because this allows the buoy surface height records and simulated net snow accumulation to remain intercomparable despite our model not resolving snow-ice formation (as noted in Section 2.2.1).

The satellite comparisons in Figure 8 are therefore meaningful only as qualitative consistency checks, not as quantitative proof of Antarctic-wide snow-ice conversion. For AMSR-E/2, the product documentation states that the snow-depth algorithm is applicable only under dry-snow conditions and has an upper retrieval limit of 50 cm because of limited penetration depth, while Antarctic passive-microwave evaluations note difficulties over multiyear ice and show substantial underestimation for older algorithms or comparable products. The manuscript also uses AMSR-E/2 snow depth to initialize the model every year, which makes later comparison with AMSR-E/2 only partly independent, especially in autumn and in multiyear-ice regions most sensitive to the initial condition. That circularity needs to be stated plainly.

This information about the limitations of passive microwave snow depth retrievals was already provided in the original manuscript (Lines 110-116), so we have taken no action to restate the information.

After continuous multiyear simulations are implemented and results are updated (see our response to Major Concern 3), the circularity described by the reviewer will no longer exist. In the current results, this partial interdependence only impacts the ~23% of the Antarctic sea ice area covered by multiyear ice and only for the inference of snow-ice from AMSR-E/2 data (not the estimate based on CryoSat-2).

CryoSat-2 is not a clean arbiter either. Kacimi and Kwok show that high-salinity basal snow can bias CryoSat-2 freeboards by at least a few centimeters, that the resulting Antarctic ice thickness/volume can shift materially under plausible corrections, and that validation remains difficult because seasonally and regionally diverse field datasets are lacking. Fons et al. also stress that Antarctic radar-altimetry snow-depth retrievals carry substantial uncertainty due to radar penetration, external snow assumptions, and sampling biases in ship-based comparisons. In a saline Antarctic snowpack, CryoSat-2 residuals cannot be straightforwardly mapped to snow-ice formation without stronger independent constraints.

As mentioned in our response to Referee #2, we will add further discussion of potential CryoSat-2 biases to Section 2.2.2: "Nonetheless, the retrieval is still vulnerable to biases arising from interface misclassification, such as anomalously thin snow depth estimates during periods of surface melt." We will add additional details on the effects of saline basal snow and possible retrieval biases to Section 2.2.2.

Interpretation of figures and conclusions

Figure 3 does not by itself demonstrate robust generalization. It shows that the calibration procedure improves the fit to the chosen buoy split, but the validation set still carries a notable negative bias, and the split is only one random partition of 39 buoys from a single basin. That is not enough to establish that the tuned parameterization is transferable to the circumpolar Antarctic. A blocked cross-validation by sector, ice regime, and buoy type would be far more convincing than a single random 75/25 split. As shown, Figure 3 mainly demonstrates that the model can be tuned to Weddell buoy behavior, not that the inferred Antarctic-wide process partition is correct.

We agree with the reviewer's concern about possible sensitivity to the partitioning of the snow buoys used for calibration vs. validation. As detailed in our response to Major Concern 4, we will implement a more robust technique, known as Monte Carlo cross-validation (MCCV), which will reduce sensitivity to any particular random partitioning. That said, we are confused by the referee's comment suggesting a "cross-validation… by buoy type", as there is only one type of snow buoy used.

The current manuscript acknowledges that the calibration focuses on the Weddell Sea, but notes that the region features a diversity of snow and meteorological conditions albeit generally lower snowfall rates (Lines 483-485). We are not aware of another body of observational records of snow evolution on Antarctic sea ice outside of the Weddell Sea, so we are limited to these buoys for model calibration. We look forward to future opportunities to validate the model using observational time series of snow evolution collected in other regions, such as during upcoming Antarctica InSync campaigns.

Figure 5b is one of the least convincing parts of the paper. The plotted cumulative process contributions imply very large compaction and very rapid melt-related losses over short intervals, yet those processes are not independently validated at the event scale. The text itself admits that the model likely produces an unusually large superimposed-ice amount for buoy 2014S9, about 32 cm , and that liquid-water effects are not explicitly represented. Observational work in the Weddell records mean superimposed-ice thicknesses closer to 0.11 ± 0.11 m, with regional increases but still much smaller typical values. I therefore do not find the partitioning in Figure 5b physically trustworthy enough to support detailed causal claims about what fraction was compaction, what fraction was warm-air melt, and what fraction was rain. This panel is presently more illustrative than evidentiary.

While we believe the model results presented in Fig. 5 are robust and accurate, we appreciate the reviewer's comment because it spurred us to look into the simulation evolution along the trajectory of snow buoy 2014S9 in greater detail and identify an error in our description of the amount of superimposed ice formation.

To respond to the reviewer's concerns, we present below in **Figure C** an expanded view of Fig. 5 focused on the events the reviewer refers to, namely the two episodes of observed rapid snow surface height loss that occurred over one-week periods in January 2025 (each marked with red stars). The model attributes these snow ablation episodes to rapid melt from above-freezing air temperatures (gold line in **Figure C**), as we describe in Section 3.1, Lines 574-589. Nowhere in the manuscript text is it suggested that "very large compaction… over short intervals" occurred; this is not possible in the model's parameterization of gradual mechanical compaction (Section 2.4.1), nor do we present any evidence for it in Fig. 5b. As wind speeds were low (< 8 m s^{-1}) during the two ablation events (blue line in **Figure C**) and simulated sublimation losses were minimal (Fig. 5b), snow redistribution or sublimation are also not satisfactory explanations. The available evidence therefore remains consistent with rapid melt, and the model's melt parameterization produces a close correspondence with the observed surface-height decreases.

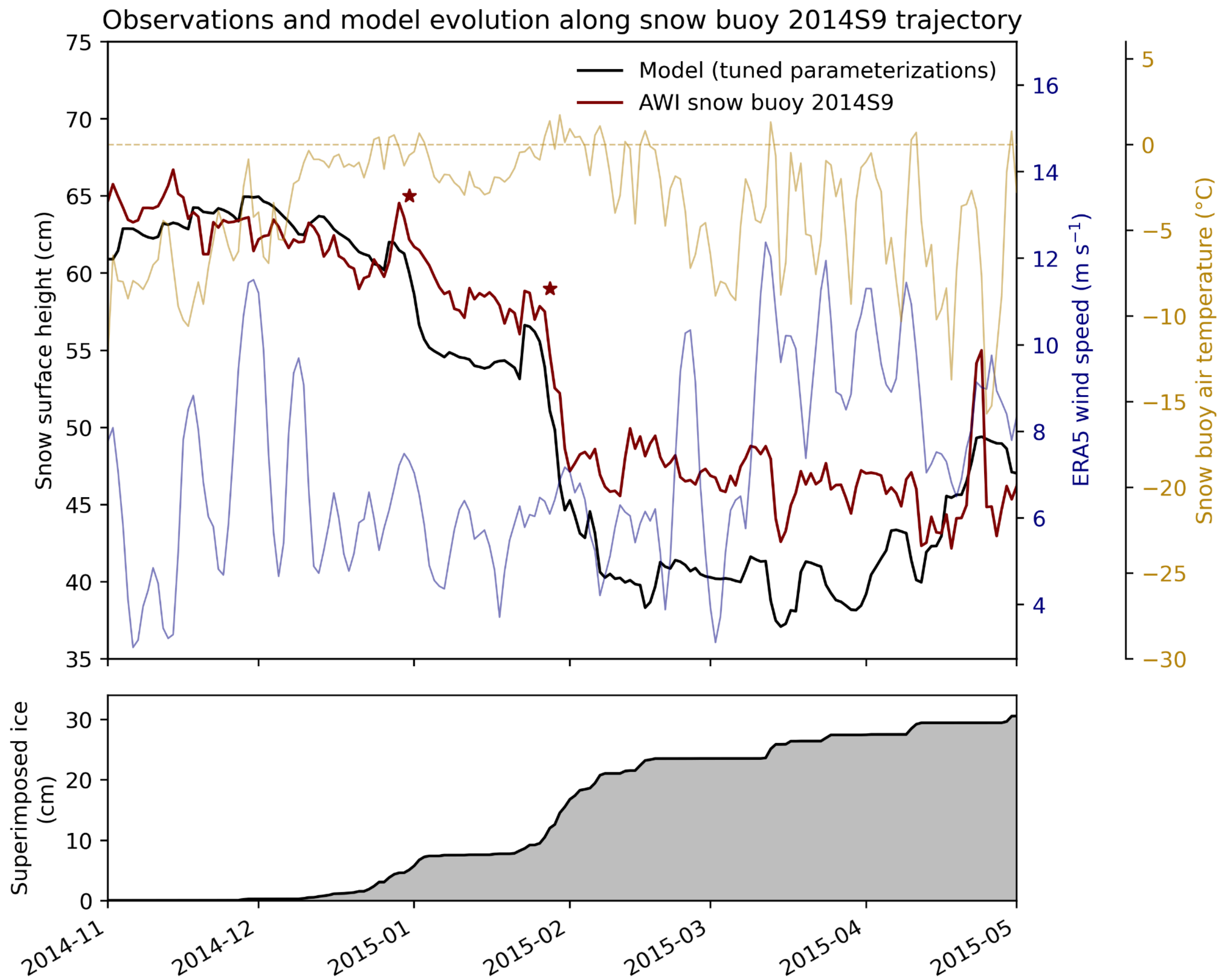

Figure C: *Observations and WASSAIL model evolution along the trajectory of snow buoy 2014S9, as shown in Fig. 5 in the manuscript, expanded to focus on the time period of summer snow ablation events.* ***At top:*** *Black and red lines represent the simulated and observed snow surface height changes, respectively, while the blue line shows ERA5 10-m wind speed along the buoy track and the gold line shows the buoy-measured air temperature (with a horizontal line indicating 0°C). Two red stars mark the rapid snow ablation events discussed.* ***At bottom:*** *Superimposed ice thickness simulated by the WASSAIL model.*

However, we realized that we made an error in Line 582 in stating that the model simulated 32 cm of superimposed ice formation in early 2015. As shown in **Figure C**, only 16 cm of superimposed ice formed from snow melt through the end of January 2015. This is more consistent with the superimposed ice thicknesses of 11 ± 11 cm observed in the northwestern Weddell Sea by Arndt et al. (2021), where the buoy was located in January. The originally quoted value of 32 cm is the amount of superimposed ice formed over the entirety of the buoy 2014S9 record, through September 2015, by which point the buoy had drifted to the northeastern Weddell Sea (Fig. 5b). We will revise Section 3.1 to correct this error and add the details above.

While the final amount of 32 cm of superimposed ice formation is indeed large, we note it is only modestly higher than superimposed ice thicknesses of 25 cm observed in an unspecified Antarctic location (Haas et al., 2001), up to 20 cm measured in the Bellingshausen and Amundsen Seas (Jeffries et al., 1997), and 24 cm simulated along the trajectory of snow buoy 2014S12 in the northern Weddell Sea (Arndt et al., 2024). Furthermore, recent observations presented at the SCAR Open Science Conference in August 2026 reported between 70–80 cm of superimposed ice at two locations in the northwestern Weddell Sea (S. Arndt, unpublished data), an intriguing finding suggesting that substantially greater superimposed ice formation than 32 cm can occur in the region.

Incidentally, after submitting our original manuscript, we realized that this buoy 2014S9 record was not only examined by Boisvert et al. (2020) but also Arndt et al. (2016), who determined that the melt event in early January 2025 occurred while the buoy was located in an area of diurnal freeze-thaw cycles followed by continuous strong snow melt, as detected by their passive microwave algorithm. We will add this context to Section 3.1.

Figure 8 is the paper’s inferential centerpiece, but it is not strong enough for the claims placed on it. The manuscript subtracts satellite snow depth from reconstructed net snow accumulation and interprets the resulting 20–30 cm positive residual over much of Antarctica in spring as evidence of widespread snow-ice formation. That residual, however, includes at least six components: true snow-ice conversion, superimposed-ice effects, liquid-water metamorphism and infiltration, AMSR/CryoSat retrieval error, annual initialization artifacts, and representativeness mismatch between level-ice point observations and gridded satellite fields. The authors themselves acknowledge that the CryoSat-2 and AMSR-derived inferences are not perfectly consistent. Under those circumstances, “widespread snow-ice formation of that magnitude” should be stated as a hypothesis to be tested, not as a strong result.

We agree with the reviewer that the snow-ice formation estimate should be stated more cautiously. We are placing more care in our interpretation of the residual as snow-ice formation and adding discussion of other contributors to the residual, as detailed in Major Concern 1b.

The final quantitative conclusion, 237 mSv of annual meteoric freshwater input, “more than half” the sea-ice-melt flux, is therefore too definite. Even if the broad message that snow-on-sea-ice freshwater matters is almost certainly correct, the exact number depends on a chain of assumptions that has not been sufficiently stress-tested. The manuscript should at most present this as a provisional estimate with large uncertainty bounds, after propagating forcing, process, initialization, and observation uncertainty. In its current form, the headline number reads as far more secure than the evidence warrants.

We concur that quantifying the uncertainty in this freshwater flux estimate is a worthy goal and that the 237 mSv estimate is not a precisely constrained quantity. That said, conventional error propagation is not feasible: the model calibration routine (Section 2.5) is designed to mitigate the effects of forcing biases and imperfect snow process parameterizations, and their uncertainties cannot be separated and quantified in a straightforward manner—especially after model calibration has altered their magnitude.

Instead, we provide a validation of the model's snow evolution using independent snow buoy measurements in Section 3.1. This cannot be directly translated to uncertainty bounds on the circumpolar freshwater flux, but nonetheless offers an empirical measure of model performance. We consider 237 mSv to be our best estimate of the meteoric freshwater flux, and we will revise the text in Section 3.3 to emphasize that this estimate is subject to the same sources of model and input data uncertainties reflected in the validation statistics.

Required revisions and additional analyses

To become publishable, the manuscript would need a stronger separation between what is directly simulated, what is empirically constrained, and what is only inferred by residual. In practice, that means the authors should either incorporate explicit snow-ice and superimposed-ice physics, including liquid-water and saline-snow effects, or substantially soften the inferential claims and reframe the snow-ice discussion as a bounded hypothesis. The present hybrid approach is too internally inconsistent for such strong conclusions.

The paper also needs explicit multi-year-memory experiments. At minimum, a second configuration should carry snow state, density, and meteoric-ice proxies across years, especially for perennial/multiyear sectors in the western and northern Weddell. Without that, the reader cannot assess how much of Figure 8 and the freshwater budget are artifacts of the annual reset. The forcing should be tested as an ensemble, not a single tuned realization. The obvious first step is to bracket ERA5 with at least one or two alternative reanalyses and to use ERA5's own uncertainty information where possible. The freshwater-flux headline should then be reported as an uncertainty range propagated from forcing, calibration, and retrieval uncertainty. A single deterministic number is not defensible at present.

Finally, the validation framework needs an explicit observation operator. Snow Buoys measure one thing, passive microwave another, radar altimetry another, and ship or ice-core observations another. Those cannot be interchanged casually. The authors should distinguish among surface-interface change, pure-snow depth, snow mass, meteoric-ice thickness, and total freshwater content, and only compare model output to an observation after transforming it into the observed quantity. That would immediately improve the logic of the paper.

We thank the referee for their constructive feedback and this summary of key concerns. As detailed above, we are incorporating many of these suggestions into the revised manuscript and are confident it will be a stronger study for it.

Responses to Referee #2

This article presents a new model for snow on Antarctic sea ice: the University of Washington Snow on Antarctic Ice Lagrangian (WASSAIL) model. The model uses a combination of reanalysis data (for atmospheric forcing) and observation-derived ice drift to estimate snow mass and bulk density over Southern Ocean sea ice. The model includes parameterized representations of various snow processes including snow compaction, sublimation, blowing snow loss to leads, changes due to ice motion, superimposed ice, and melt, but currently excludes snow-ice formation. Model processes are parameterized and the parameters are calibrated against a set of buoy snow observations. Snow budget terms from the model output are examined. Seasonal averages are compared to passive microwave and altimetry-derived estimates of snow depth.

I think the model is promising overall and represents a good step forward in Antarctic snow-on-sea-ice modelling, and I also commend the authors for the easy accessibility of the model source code and clear documentation. The article is generally clearly written and well-referenced, with good-quality figures. However, in my view, some of the model development methodology and scientific reasoning require expansion and/or clarification, particularly pertaining to the omission of snow-ice and some details of the model calibration process. I think some additional revisions/testing are required before the manuscript would be suitable for publication. I include my comments with suggestions below.

We thank the referee for their thoughtful and constructive assessment of our study, including their detailed suggestions below. We will make revisions throughout the manuscript to address these concerns.

Major comments:

The authors take great care to make careful choices of parameterizations of several snow processes, which makes the exclusion of snow-ice formation somewhat puzzling to me. I understand that there are non-trivial difficulties with this (lack of knowledge of ice thickness being a key limitation), and that validation with buoys remains possible despite the exclusion of this process because of their tracking of snow surface height relative to their initialization baseline, but nevertheless it seems like a large omission given the prevalence of snow-ice conversion on Antarctic sea ice. It is not clear to me why the authors chose to omit this process rather than attempt to e.g. parameterize and calibrate it as has been done with other snow processes. The frequent attribution of budget inconsistencies and validation discrepancies to snow-ice formation seems hasty to me also. I think the authors need to take more caution in how this discussion is framed. I also urge caution at the interpretation of model-observation differences as evidence of snow-ice formation, since other processes/biases may factor into the differences. Snow-ice itself may also introduce further uncertainty into remote sensing measurements, so I think these inferences need to be stated with more caution than is currently employed.

We appreciate the reviewer’s recommendations regarding snow-ice formation. We discuss our revisions arising from these suggestions in our response to Major Concerns 1a and 1b at the top of this document. We will be revising the manuscript to better explain the omission of explicit snow-ice formation, clarify the significance of estimated freshwater fluxes, moderate our claim about snow-ice formation inferred from the model–satellite residual, and discuss other sources of uncertainty that contribute to the residual.

I am sympathetic to the challenges of calibrating models with limited observational dataset availability, and the successive halving method is in my view a reasonable choice for this particular calibration problem, but I wonder about the reliance of the model parameters on a specific choice of buoys. Calibration is currently performed once for a single selection of calibration buoys; how much do the optimal parameters vary when different sets of buoys are (randomly) selected? I think more testing may be necessary to examine the robustness of this parameterization, and to guard against potential overfitting.

We agree with the reviewer's concern about possible sensitivity to the partitioning of the snow buoys used for calibration vs. validation. As detailed above in Major Concern 4, we will implement the specific technique that they have suggested, a framework known as Monte Carlo cross-validation (MCCV), to increase the robustness of the calibration routine and make the final model parameters more resilient to overfitting.

Minor comments:

Some of the figure text is small, please increase the size for legibility (e.g. Fig 7, Fig 8)

We will increase font sizes in the Fig. 3 legends, Fig. 7 titles and colorbars, and Fig. 8 vertical labels and colorbars.

You mention some other modelling efforts (e.g. CASSIS); can you provide a sense of how well WASSAIL performs relative to other snow-on-sea-ice simulations? This could also help contextualize the impact of the missing snow-ice process.

We appreciate this suggestion and will add a summary of the following to Section 3.1:

- CASSIS (Antarctic) shows a bias of –7.3 cm and RMSE of 13.5 cm compared to Operation IceBridge data and a bias of –9.4 cm and RMSE of 17.2 cm compared to AMSR snow depth (Lawrence et al., 2024). We discount the comparison to ASPeCt observations because the model was calibrated to the same data set.
- Comparable summary statistics for SnowModel-LG (Arctic) are not available, but comparison with Operation IceBridge data shows mean differences along coincident tracks varying by year from 0 cm to 8 cm while comparison with passive microwave snow depths shows differences on the order of 10–30 cm (Stroeve et al., 2020).

Thus, while the validation data sets and methodologies vary, WASSAIL appears reproduce snow accumulation with similar fidelity to other reconstructions of snow depth, based on the validation snow buoy statistics (Section 3.1). This will be highlighted.

If I'm interpreting things correctly, model output is interpolated to a regular lat/lon grid; would this not introduce artefacts at the pole? I'm curious of the motivation for using that grid rather than a polar grid (e.g. EASE).

The use of a uniform lat/lon grid is necessary for computational tractability. This is relevant both when regridding input forcing data and when binning model output from Lagrangian parcel drift trajectories onto the grid for visualization and analysis purposes. 2-D interpolation or binning would be significantly more challenging on an EASE grid. That said, the current binning and regridding procedures does not introduce artifacts; the binning uses a simple 2-D histogram routine and the remapping uses a geospatial Python toolbox designed for this purpose (*xesmf*). Note that Antarctic sea ice does not exist at the pole (90°S). Additionally, as mentioned in Section 2.4.8, area-weighted (latitude-dependent) binning and averaging are used for accuracy in grid-based calculations.

Also, is the final output resolution 0.75x0.75 degrees? I think this needs to be stated more explicitly earlier in the text for clarity; I was initially under the impression that the resolution was 0.25.

Yes. This will be clarified in Section 2.4.8: "… coarsened by a factor of 3 to 0.75° × 0.75° *(the final model output resolution)*…". A note on the output resolution will additionally be added to Section 2.1.

Line-by-line comments:

115: bias of -19 cm; is this for AMSR2 – OIB or OIB – AMSR2? Please specify

This is AMSR2 – OIB, so it represents a low bias for AMSR2. This will be clarified in the text.

155-156: Snowfall includes a scaling parameter; is a scaling approach also used for the total precipitation?

Rainfall is not scaled in a similar way to snowfall, which is reflected in the existing language and Eq. 8. We can think of good reasons to adopt either approach. Ultimately, however, the calibration of γ_{new} is based on snow accumulation changes, so we think it makes sense to restrict that scaling to new snowfall. The impacts of rainfall on snow melt are separately subject to calibration using γ_{rain}.

Figure 1 looks good overall, but at 100% zoom it is very difficult to distinguish the calibration and validation sets in panel b. If possible, please consider making this panel larger or otherwise adjust how the lines are formatted for clarity.

We will increase the legibility of the two different line formats in Fig. 1b by: (1) cropping the map more tightly around the data extent, allowing the lines to be wider; and (2) experimenting with a more visible inner color in the validation buoy lines.

Section 2.2.2: I think it would be worth mentioning how the [CryoSat-2] retrieval can still have biases e.g. from melt; although it can often identify the snow-ice interface, it can still have difficulties

Good point. We will add mention of this in Section 2.2.2: "*Nonetheless, the retrieval is still vulnerable to biases arising from interface misclassification, such as anomalously thin snow depth estimates during periods of surface melt*."

233: How is SIC treated in calibration mode? Are there times where remote sensing SIC disagrees with in situ? (or e.g. grid edge effects)

Thanks for noticing the need for clarification here. SIC is not used for parcel creation or destruction in calibration mode, since the parcel being simulated is a single point (the buoy location) rather than a grid-scale "floe". Nonetheless, nearest-neighbor SIC is still used in the parameterization of blowing snow lost to leads, as this is assumed to be a non-local redistribution process affected by nearby leads. This will be clarified in Line 255: "…, *and SIC is referenced only in the parameterization of blowing snow loss*."

Due to the grid-scale/point-scale mismatch, remotely sensed SIC for a grid cell encompassing a buoy location can plausibly range from anywhere between 1–100% without there being disagreement. We haven't checked for instances of satellite SIC being exactly 0% at a buoy location, but in such a case (which would likely reflect rapid melt conditions within the MIZ), we would infer that the buoy remains drifting on sea ice.

269: To clarify, does this mean that the inverse scaling factor (gamma_dens) is what is calibrated in the model, whereas k_n remains fixed? I.e. is the scaling factor essentially a way to calibrate k_n?

Yes, the inverse scaling factor γ_{dens} serves to calibrate k_n, as reflected in Table 1 and Eq. 2. We will clarify this in the text: "Here we adopt a *fixed value* of 4000 K…".

316: Consider specifying here that the "available snow depth divided by the time step" option is so that you do not remove more snow than the available snow depth

We will clarify this: "… or the available snow depth divided by the time step, *thus limiting melt to the available snow:…*".

325: Is the factor of 12.5 prescribed from literature?

This value represents the ratio of the specific heat of water (1.0 cal g^{-1} $°C^{-1}$) to the latent heat of fusion (79.7 cal g^{-1} at 0°C), as described in Anderson (1976), after appropriate unit conversions. We will note this in the text: "*The 12.5 $°C^{-1}$ prefactor represents the ratio of the specific heat of water to its latent heat of fusion (Anderson, 1976).* ROS melt applies…".

515-516: As phrased, this reads like the goodness-of-fit metrics are also used for the calibration when only RMSE is used, please clarify

This will be clarified by adding the following sentence: "*Only RMSE influences the calibration*."

557-558: Although the studies cited do have divergent conclusions on reanalysis biases, I think it's worth mentioning that they examine different time periods [2] (e.g. CloudSat is ~2006-2016, drifting station records are from ~1950s-1991) which may suggest different biases in different time periods (particularly also when there may be discontinuities in products assimilated in ERA5 across decades). Though regardless uncertainties are indeed high.

We appreciate this insightful point. We will add a note that the North Pole drifting station records used by Barrett et al. (2020) are from 1979–1991. Additionally, the subsequent sentence will see the following addition: "… from diverse validation data sets, techniques, *and time periods*, as well as possible hemispheric differences."

Figure 4: It's difficult to see which lines are dashed; consider distinguishing buoys vs model in other ways (e.g. use a lighter shade of the line colour, different line weights, etc).

We will make the buoy lines more distinct while still capturing fluctuations on short time scales.

604: If the model were run for multiple years during calibration, could you constrain the model drift further and close the budget? Or are you concerned that this would cause the free parameters to take on unphysical values?

Note that the calibration routine is run by evolving the model forward along the snow buoy trajectories, both in time and space. The existing calibration runs are therefore inherently multiyear, as some buoys survive for 2–3 years. Please see our response regarding multiyear spin-up runs in free-running mode (after calibration) in Major Concern 3 above.

608-609: It's not clear to me why snow-ice formation is grouped with snow release to the ocean if snow-ice is not simulated, please elaborate. Is it implicitly assumed to be part of the budget term?

We agree that the wording of this text was confusing. Only the meteoric fraction of snow-ice (the fraction derived from snow) is included within the pink slice of Fig. 6b. As detailed in our response to Major Concern 1a above, we will make changes to this section's text, the caption of Figure 6, labels in Figure 6, and Figure 7 to clarify that this budget term represents the combined meteoric freshwater release to the ocean—specifically, snow plus snow incorporated into snow-ice.

Figure 7: How do you account for year-to-year differences in marginal ice cover for these averages? I.e. since these are multiyear averages, it's possible that grid squares near ice margins may only include a few years of data in the average. I suggest checking this and either discussing possible impacts or applying a minimum threshold for the number of available years.

This is an excellent point, though we believe it is more relevant to Fig. 8 than Fig. 7. We will adopt changes to Fig. 8 to more accurately represent the margins of the SIZ.

In Fig. 7, the climatological budget terms are calculated for each grid cell as the sum of all years' contributions divided by length of the total record (22 years). Therefore, years with anomalously large sea ice extent do not have outsized influence in the result. A grid cell at the SIZ margin with only one year represented will have its budget terms divided by 22. We feel this is an appropriate representation.

In Fig. 8, however, the average snow depths were computed after zero-filling grid cells with no snow depth in a given year. Thanks to the reviewer's comment, we realized this yields the visual effect of an excessively broad SIZ, the result of a few years with anomalously large sea ice extent. Despite average snow depths around the margins (accurately) being close to zero, imposing a minimum threshold for the number of available years would produce a more representative visualization, as the reviewer has suggested. We will modify Fig. 8 accordingly.

736: Are you comparing the 2003-2025 time period from the model with CryoSat-2 from 2010-2021? If so, some of the disagreement may be spurious; I suggest comparing 2010-2021 from the model and CS2.

Yes, the comparison is over different time periods, as noted in the text. We agree that a more representative comparison is warranted. We will restrict the model (Fig. 8a) and AMSR (Fig. 8d) average fields to the CryoSat-2 period of 2010–2021. As Fig. 8 is probably too large to accommodate an additional row, we will also add a figure to the Appendix reproducing the current Fig. 8a, over the full period of 2003–2025.

750: The correspondence being less robust for CS2 makes me wonder how much of this is due to snow-ice formation vs how much of this is from other discrepancies (particularly if different time periods are being compared here).

It is indeed notable that the comparison in Fig. 8 is more robust for CryoSat-2 than for AMSR, given the model initialization from AMSR snow depth. We will reassess these results after updating the time periods, as described above.

References

Anderson, E. A. (1976). *A point energy and mass balance model of a snow cover*. *NOAA Technical Report NWS 19*. National Oceanic and Atmospheric Administration. Retrieved from https://repository.library.noaa.gov/view/noaa/6392

Arndt, S., Willmes, S., Dierking, W., & Nicolaus, M. (2016). Timing and regional patterns of snowmelt on Antarctic sea ice from passive microwave satellite observations. *Journal of Geophysical Research: Oceans*, *121*(8), 5916–5930. https://doi.org/10.1002/2015JC011504

Arndt, S., Haas, C., Meyer, H., Peeken, I., & Krumpen, T. (2021). Recent observations of superimposed ice and snow ice on sea ice in the northwestern Weddell Sea. *The Cryosphere*, *15*(9), 4165–4178. https://doi.org/10.5194/tc-15-4165-2021

Arndt, S., Maaß, N., Rossmann, L., & Nicolaus, M. (2024). From snow accumulation to snow depth distributions by quantifying meteoric ice fractions in the Weddell Sea. *The Cryosphere*, *18*(4), 2001–2015. https://doi.org/10.5194/tc-18-2001-2024

Barrett, A. P., Stroeve, J. C., & Serreze, M. C. (2020). Arctic Ocean precipitation from atmospheric reanalyses and comparisons with North Pole drifting station records. *Journal of Geophysical Research: Oceans*, *125*(1), e2019JC015415. https://doi.org/10.1029/2019JC015415

Blanchard-Wrigglesworth, E., Webster, M. A., Farrell, S. L., & Bitz, C. M. (2018). Reconstruction of snow on Arctic sea ice. *Journal of Geophysical Research: Oceans*, *123*(5), 3588–3602. https://doi.org/10.1002/2017JC013364

Bocquet, M., Fleury, S., Rémy, F., & Piras, F. (2024). Arctic and Antarctic sea ice thickness and volume changes from observations between 1994 and 2023. *Journal of Geophysical Research: Oceans*, *129*(11), e2023JC020848. https://doi.org/10.1029/2023JC020848

Boisvert, L. N., Webster, M. A., Petty, A. A., Markus, T., Cullather, R. I., & Bromwich, D. H. (2020). Intercomparison of precipitation estimates over the Southern Ocean from atmospheric reanalyses. *Journal of Climate*, *33*(24), 10627–10651. https://doi.org/10.1175/JCLI-D-20-0044.1

Haas, C., Thomas, D. N., & Bareiss, J. (2001). Surface properties and processes of perennial Antarctic sea ice in summer. *Journal of Glaciology*, *47*(159), 613–625. https://doi.org/10.3189/172756501781831864

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz‐Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Holland, M. M., Clemens-Sewall, D., Landrum, L., Light, B., Perovich, D., Polashenski, C., et al. (2021). The influence of snow on sea ice as assessed from simulations of CESM2. *The Cryosphere*, *15*(10), 4981–4998. https://doi.org/10.5194/tc-15-4981-2021

Jeffries, M. O., Worby, A. P., Morris, K., & Weeks, W. F. (1997). Seasonal variations in the properties and structural composition of sea ice and snow cover in the Bellingshausen and Amundsen Seas, Antarctica. *Journal of Glaciology*, *43*(143), 138–151. https://doi.org/10.3189/S0022143000002902

Jeffries, M. O., Krouse, H. R., Hurst-Cushing, B., & Maksym, T. (2001). Snow-ice accretion and snow-cover depletion on Antarctic first-year sea-ice floes. *Annals of Glaciology*, *33*, 51–60. https://doi.org/10.3189/172756401781818266

Kacimi, S., & Kwok, R. (2020). The Antarctic sea ice cover from ICESat-2 and CryoSat-2: freeboard, snow depth, and ice thickness. *The Cryosphere*, *14*(12), 4453–4474. https://doi.org/10.5194/tc-14-4453-2020

Kwok, R., & Cunningham, G. F. (2008). ICESat over Arctic sea ice: Estimation of snow depth and ice thickness. *Journal of Geophysical Research*, *113*, C08010. https://doi.org/10.1029/2008JC004753

Lange, M. A., Schlosser, P., Ackley, S. F., Wadhams, P., & Dieckmann, G. S. (1990). 18O concentrations in sea ice of the Weddell Sea, Antarctica. *Journal of Glaciology*, *36*(124), 315–323. https://doi.org/10.3189/002214390793701291

Lawrence, I. R., Ridout, A. L., Shepherd, A., & Tilling, R. (2024). A simulation of snow on Antarctic sea ice based on satellite data and climate reanalyses. *Journal of Geophysical Research: Oceans*, *129*(1), e2022JC019002. https://doi.org/10.1029/2022JC019002

Liston, G. E., Itkin, P., Stroeve, J., Tschudi, M., Stewart, J. S., Pedersen, S. H., et al. (2020). A Lagrangian snow-evolution system for sea‐ice applications (SnowModel‐LG): Part I—Model description. *Journal of Geophysical Research: Oceans*, *125*(10), e2019JC015913. https://doi.org/10.1029/2019JC015913

Maksym, T., & Markus, T. (2008). Antarctic sea ice thickness and snow-to-ice conversion from atmospheric reanalysis and passive microwave snow depth. *Journal of Geophysical Research: Oceans*, *113*(2), 1–18. https://doi.org/10.1029/2006JC004085

Nicolaus, M., Hoppmann, M., Arndt, S., Hendricks, S., Katlein, C., Nicolaus, A., et al. (2021). Snow depth and air temperature seasonality on sea ice derived from snow buoy measurements. *Frontiers in Marine Science*, *8*, 655446. https://doi.org/10.3389/fmars.2021.655446

Petty, A. A., Webster, M., Boisvert, L., & Markus, T. (2018). The NASA Eulerian Snow on Sea Ice Model (NESOSIM) v1.0: initial model development and analysis. *Geoscientific Model Development*, *11*(11), 4577–4602. https://doi.org/10.5194/gmd-11-4577-2018

Provost, C., Sennéchael, N., Miguet, J., Itkin, P., Rösel, A., Koenig, Z., et al. (2017). Observations of flooding and snow‐ice formation in a thinner Arctic sea‐ice regime during the N‐ICE2015 campaign: Influence of basal ice melt and storms. *Journal of Geophysical Research: Oceans*, *122*(9), 7115–7134. https://doi.org/10.1002/2016JC012011

Stroeve, J., Liston, G. E., Buzzard, S., Zhou, L., Mallett, R., Barrett, A., et al. (2020). A Lagrangian snow evolution system for sea ice applications (SnowModel‐LG): Part II—Analyses. *Journal of Geophysical Research: Oceans*, *125*(10), e2019JC015900. https://doi.org/10.1029/2019JC015900

Wever, N., Leonard, K., Maksym, T., White, S., Proksch, M., & Lenaerts, J. T. M. (2021). Spatially distributed simulations of the effect of snow on mass balance and flooding of Antarctic sea ice. *Journal of Glaciology*, *67*(266), 1055–1073. https://doi.org/10.1017/jog.2021.54

Willatt, R. C., Giles, K. A., Laxon, S. W., Stone-Drake, L., & Worby, A. P. (2010). Field investigations of Ku-band radar penetration into snow cover on Antarctic sea ice. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(1), 365–372. https://doi.org/10.1109/TGRS.2009.2028237

Wilson, E. A., Riser, S. C., Campbell, E. C., & Wong, A. P. S. (2019). Winter upper-ocean stability and ice-ocean feedbacks in the sea ice-covered Southern Ocean. *Journal of Physical Oceanography*, *49*(4), 1099–1117. https://doi.org/10.1175/JPO-D-18-0184.1