

Revision of the Preprint:
“*Evaluation of MARv3.14 over the Greenland Ice Sheet*”
– Response to the Reviewers –

G. Timmermans, B. Noël, C. Kittel, T. Dethinne, N. Ghilain & X. Fettweis

September 2026

Contents

1 Answer to Reviewer 1 (Kristiina Verro)	1
1.1 Specific comments [scientific]	1
1.2 Technical corrections	6
2 Answer to Reviewer 2 (anonymous)	9
3 Bibliography	12

1 Answer to Reviewer 1 (Kristiina Verro)

The authors are presenting a new MAR version, v3.14, and have gone through extensive and thorough model evaluation in the context of simulated surface mass balance (SMB) over the domain of Greenland Ice Sheet. MAR outputs are compared to a range of observations, including in situ measurements, AWS records, satellite-derived melt extent, and albedo products. For example, the discussion on ablation underestimate (Sect. 4.1) describes, in detail, comparison to SMB observations from station to station. They have also conducted a model resolution-dependency analysis.

The paper is generally well structured, but needs some technical editing. I would like to see a structured Conclusions section. There are a lot of comparisons made in this paper, and the Conclusions section is the perfect place to summarize the main points found in the analysis that are of interest for a MARv3.14 (data) user. Please provide concrete, specific outcomes rather than vague, generalized summary. The Results section needs language polishing / stylistic editing. Namely, the over use of parentheses and over-referencing should be reviewed, as it severely disrupts the reading. This is also true in the Discussion section. At times, the Results section paragraphs are hard to read, as there is no clear structure (e.g in 3.4 Albedo representation). I suggest starting with a strict lead sentence, which is the opening sentence of a paragraph that introduces its central point, and then following with supporting sentences with data. As this is a paper using a lot of different datasets to evaluate MAR, the results need to be presented clearly.

I have only some minor scientific comments, mostly asking to expand and elaborate. Overall, the evaluation is done properly and the results make sense. MAR is an important tool in climatological studies, and such a thorough and rigorous improvement and evaluation is appreciated.

Suggestions for improving the content are provided here. I recommend the paper to be published at the TC, after these minor, mostly technical, improvements.

Thank you for your constructive comments on our manuscript. Below you will find our point-by-point response to your comments in blue. Changes applied to the revised manuscript are shown in red.

1.1 Specific comments [scientific]

L31: *Several approaches have been developed to estimate the GrIS SMB including positive degree days, energy balance, regional (RCM) or global circulation models (GCM) (Fettweis et al., 2020).*

An important takeaway of the paper of Fettweis et al. 2020 is not the existence of several approaches of SMB estimation, but that RCMs out-perform other methods when validated against satellite and in situ field observations. I think this is an important motivation of this paper of development of MAR and should be emphasised here.

Indeed, we cited Fettweis et al. (2020) as it discusses and compares different SMB modelling approaches (p. 3937). We now highlight this in the revised manuscript as:

“Several approaches have been developed to estimate the GrIS SMB including positive degree days, energy balance, regional or global circulation models. As highlighted by Fettweis et al. (2020), regional climate models (RCMs) prove to be the most accurate at capturing SMB when evaluated against in situ and satellite observations.”

L90: *Grid-cells with an ice fraction lower than 50 % are discarded to avoid unrealistically high ablation values near the ice sheet–tundra interface.*

I would like a more thorough explanation on what unrealistically high ablation values are, so as not to disregard data semi-arbitrariness from the analysis. Furthermore, is this hard-coded within the model (as in grid-cell fractionation is only from 50% ice fraction and above), or is it that you later only analysed GrIS pixels with grid-cell fractionation > 50% ice?

MAR implements grid-cell fractionation. If a pixel contains, for instance, 80 % of tundra and 20 % of ice, the surface module (SISVAT) will run two times at each MAR iteration, once for each mode. Then the atmospheric module estimates a weighted average of the two modes (tundra/ice) as lower boundary condition (e.g., surface temperature and energy fluxes). For pixels showing an ice fraction lower than 50 %, we consider that model outputs are not representative of the ice sheet conditions, as tundra areas will substantially influence the end result. This methodology is a convention commonly applied in the literature, and to every MAR study over Greenland up to now (e.g. Fettweis et al., 2017; 2020). The same holds for RACMO2.3 in e.g. Noël et al. (2015; 2018). This is now clarified as:

“As commonly adopted in Greenland MAR studies (e.g., Fettweis et al., 2017; 2020), grid-cells with an ice fraction ...”

L147: *We present and discuss the results obtained using a melt rate–based criterion, whereby a grid-cell is classified as melting when the simulated melt exceeds a threshold of 0.2 mm w.e. day⁻¹, following Fettweis et al. (2011) and Puggaard et al. (2025).*

Later, in Sect. 3.5, Sensitivity tests that derived the MARv3.14 threshold of 0.2 ± 0.1 mm w.e. day⁻¹ is referred to Table S8. Puggaard et al. 2025 determined 1 mm w.e. day⁻¹ to be MARv3.12 threshold using ROC and PR curves. Fettweis et al. (2011) tested a range of values, but suggested 8.25 mm WE day⁻¹. I do expect different thresholds in MARv3.14 and previous versions. Your methodology of minimizing the RMSE tests quite low values, 0.1-0.3 mm w.e. day⁻¹ and does not test the values suggested by Fettweis et al. (2011) and Puggaard et al. (2025). Could you expand on the choice of low threshold values in the sensitivity tests? Why such a large difference between 1 (Puggaard et al. 2025), 8.25 (Fettweis et al. 2011) and your 0.2 ± 0.1 mm w.e. day⁻¹? Is it the methodology of determining the threshold, or is it changes between the MAR version (such as implementing meltwater freezing or ecRad?). Could you have hit a local minimum with 0.2 mm w.e. day⁻¹? This is relevant also considering the results you obtained: Figure 6b indicates that the use of a melt threshold of 0.2 mm w.e. day⁻¹ results in an overestimation (>30 days per year) of the mean annual number of melt days over southeastern Greenland.

We tested melt thresholds from Fettweis et al. (2011) (8.25 mm w.e. day⁻¹) and Puggaard et al. (2025) (1 mm w.e. day⁻¹) which led to worse agreement with observations (see Figs. 1 and 2).

The differences in melt thresholds likely arise from the use of different satellite sensors and melt-detection methods: Puggaard et al. (2025) used active microwave backscatter observations from ASCAT, Fettweis et al. (2011) used passive microwave observations from SMMR, SSM/I, and SSMIS, whereas we use passive microwave observations from AMSR-E and AMSR2 following Picard (2025).

We retain the melt-production threshold of 0.2 mm w.e. day⁻¹ because it yields the lowest RMSE among the thresholds tested, including 1 and 8.25 mm w.e. day⁻¹. We therefore retain this threshold for the comparison in the revised manuscript.

While revisiting the melt-extent comparison, we identified an array-allocation error. The observation and MAR arrays were intended to contain 365 daily melt extents for each of the 23 years from 2002 to 2024. However, they were allocated for 29 years, reducing the RMSE by a factor of $\sqrt{23/29} \sim 0.89$, and the mean bias by a factor of $23/29 \sim 0.79$. We have removed these entries and recalculated the affected statistics and figures. The revised calculation also excludes days without complete observational coverage over the comparison mask (approximately 5.7% of the study period). Removing the zero-filled entries rescales the RMSE by the same factor for all thresholds and therefore does not change their ranking. The maps of mean annual melt days and their associated spatial statistics were unaffected by the array-allocation error, as they were calculated separately using the correct 23 years. We have updated the manuscript accordingly.

L215 and L313: *“Note that our results show similar biases and correlation to those obtained from another RCM (i.e., RACMO2.3p2 at 5 km (Noël et al., 2019) using the same near-surface meteorological variables and*

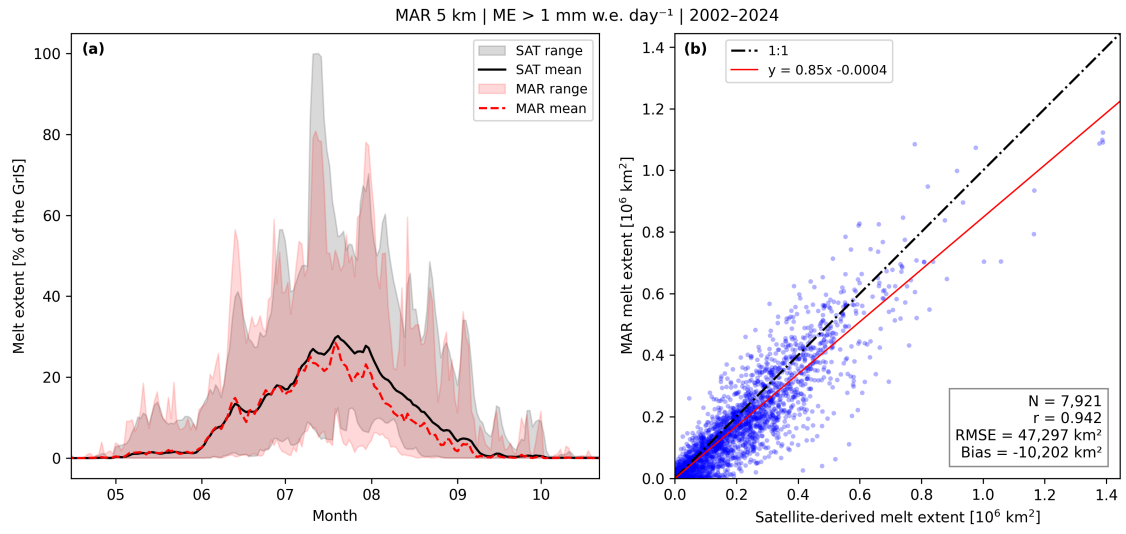


Figure 1: Same as Figure 7 of the preprint using a melt threshold of 1 mm w.e. day⁻¹.

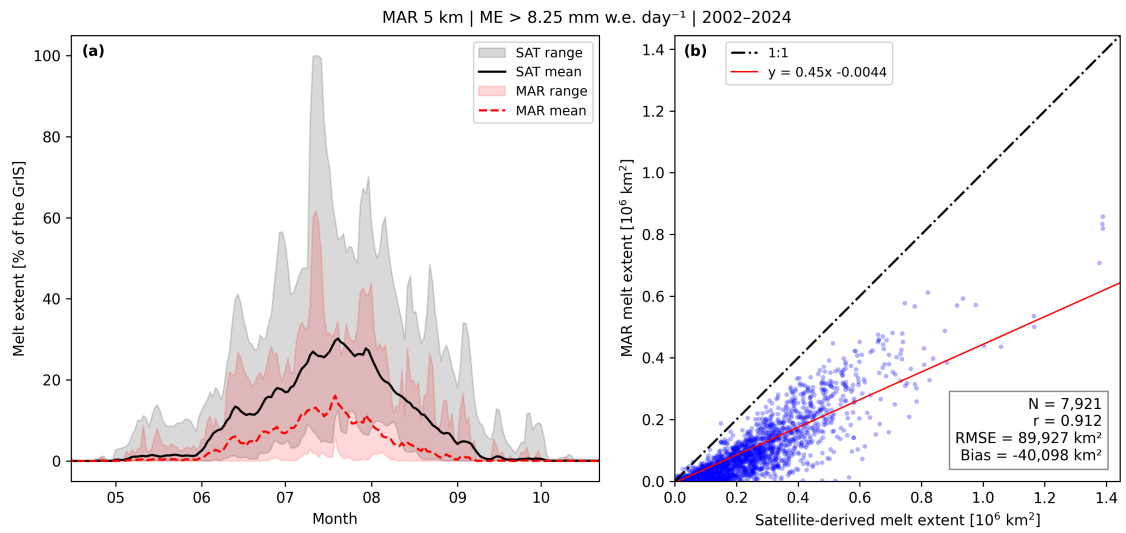


Figure 2: Same as Figure 7 of the preprint using a melt threshold of 8.25 mm w.e. day⁻¹.

radiative fluxes. Both MAR and RACMO2 tend to underestimate high wind speed.”

Also in the Discussion section: that the biases are “of the same order of magnitude as another state-of-the-art RCM at 5.5 km”.

Two points here. Firstly, please support the “similar biases and correlation” and “tend to underestimate” claims by giving numbers, so the reader does not have to go to Noël et al., 2019 paper and compare these results side by side.

Thank you. We have rephrased this statement, to make the comparison more quantitative, now providing the RMSE, bias, coefficient of determination (R^2), and regression slopes for both MAR and RACMO2.3p2 as follows:

“Note that our results align with those obtained from another RCM, RACMO2.3p2 at 5 km (Noël et al., 2019), which was evaluated using the same near-surface meteorological variables and radiative fluxes. For temperature and specific humidity, MAR and RACMO2.3p2 show comparable performance, with RMSEs of 2.49 versus 2.07 °C and 0.36 versus 0.32 g kg⁻¹, biases within ± 0.21 in their respective units, and R^2 values above 0.93. The corresponding regression slopes are close to unity (0.92–0.94 for temperature and 0.99–1.01 for specific humidity). Similar agreement is found for the radiative fluxes, with biases generally within ± 10 W m⁻², R^2 values of 0.75–0.93 for MAR and 0.85–0.96 for RACMO2.3p2, and regression slopes of 0.91–0.95 and 0.94–0.99. Both models tend to underestimate high wind speeds, as indicated by slopes below unity (0.69 and 0.91), although RACMO2.3p2 shows a lower RMSE (0.98 versus 1.98 m s⁻¹) and higher R^2 (0.73) than MAR (0.59).”

The second point is technical: there is one bracket too much: (i.e., RACMO2.3p2 at 5 km, Noël et al., 2019).

Thank you for pointing this out, we removed the extra parenthesis and now refer to “RACMO2.3p2 at 5 km (Noël et al. 2019)”.

L232 and L238, sentences: ... but also reflects a systematic negative bias (~ -0.1), likely originating from the inability of the snow albedo scheme of MAR to capture high fresh snow surface albedo (~ 0.87) in accumulation zones.

and: We find that MAR overestimates low albedo values in ablation zones exposing bare ice in summer. In fact, MAR does not simulate values below 0.35 whereas values as low as 0.1 are measured by AWSs. This discrepancy stems from the MAR albedo scheme, which constrains the prescribed MODIS-derived ice albedo between 0.35 and 0.55 to prevent the development of unrealistically low albedo values over large areas (25 km²).

Please add extra information about the albedo scheme to Sect. 2.1. From the MAR improvements description, the available information there lead me to believe that bare ice albedo (prescribing MODIS observations instead of a surface meltwater dependent bare ice albedo ranging between 0.5 and 0.55 as before) and the possibility for the meltwater to freeze above bare ice, would allow for lower albedo values than 0.35. Furthermore, how are meltpond albedos taken into account? When it comes to fresh snow albedo, is there a maximum value?

In MARv3.14, bare-ice albedo is derived from MODIS observations using the 5th percentile of the observed albedo. The resulting value is then clipped between 0.35 and 0.55. The lower bound is an explicit modelling choice intended to prevent unrealistically low albedo values over the model grid. Thus, although MODIS observations can yield values below 0.35, these values are not allowed in the MAR simulations. This is now mentioned in Sect. 2.1 as:

“In the MAR implementation, however, the resulting ice albedo is constrained to a minimum value of 0.35 and a maximum value of 0.55. The lower bound is imposed to prevent unrealistically low albedo values over the relatively coarse MAR grid, whereas the upper bound discards unrealistically high bare-ice albedos. Consequently, although the MODIS-derived BIA can locally be lower than 0.35, such values are not captured by MAR simulations.”

We also clarify that melt ponds are not explicitly represented in the MAR version evaluated in this study.

“The MARv3.14 snow albedo scheme does not account for melt pond formation and sets fresh snow albedo to 0.875 as the snow MAR albedo scheme (CROCUS) was not updated.”

Figure 5a.

Where are the horizontal line features in the summer albedo on Fig. 5a coming from?

The bare-ice albedo is prescribed from MODIS observations as a single value for each MAR grid cell and is therefore constant in time for a given grid cell. Consequently, when an AWS is located within a grid cell for which MAR exposes bare ice during summer, the prescribed MODIS albedo remains constant over the corresponding period, producing the horizontal features in the scatter plot. In contrast, the observed albedo at the AWS varies in time because the actual bare ice albedo evolves throughout the melt season. This is now clarified as:

“The horizontal features visible in Fig. 5a arise from the fact that MAR prescribes bare ice albedo as a

constant, while the observed albedo varies in time.”

L265. Moreover, MAR suggests the presence of aquifers in this area where melt occurs sometimes 10 m or 20 m below the surface.

Do you mean that meltwater percolates and is stored 10-20m below the surface in MAR’s simulations? Could you expand a little on that? How surface melt water reaches the ocean, and how fast it does so, is poorly understood, but an important process for SMB. These depths, or even deeper, are realistic, e.g. Humphrey et al. 2021 and Forster et al. 2013.

In southeast Greenland, MAR simulates both substantial summer melt and high snowfall accumulation. Meltwater can percolate through the snow/firn column and does not necessarily refreeze or run off immediately, so that liquid water can occasionally persist at depths of 10–20 m as irreducible water content (IWC). However, this retained water should not be interpreted as a simulated firn aquifer. Although observations document firn aquifers in southeast Greenland, MAR does not explicitly represent them. We have therefore removed this sentence from the manuscript to avoid this interpretation.

L340: Additional biases, such as in precipitation, drifting snow erosion and deposition, sublimation, and measurement uncertainties, may also play a role.

I agree. But, is drifting snow erosion to some extent included in MARv3.14 snow scheme or none whatsoever?

Drifting snow erosion/redeposition is not activated in the Greenland MAR configuration. This is an intentional modelling choice, as previous tests showed that activating the drifting-snow scheme did not improve, and in some cases degraded, the agreement with observations for Greenland.

This sentence also raises points not discussed in the paper: drifting snow erosion, deposition and sublimation, while precipitation was evaluated between different MAR resolutions. These topics stay out of the scope of this paper and are difficult to evaluate. But, this sentence raises the question: what about them? I suggest spending a few more sentences explaining why these SMB components were not evaluated in the paper, and referencing MAR papers that have done so (e.g. MAR (v3.11) drifting snow paper by Amory et al. 2021) and what they found.

We agree that these processes nevertheless occur in reality and can locally contribute to SMB variability. Sublimation has a secondary effect on SMB ($\sim 48 \text{ Gt year}^{-1}$) and cannot be directly evaluated using in situ measurements. We added the following clarification:

“Evaluating those processes is beyond the scope of this study. The MAR drifting snow scheme developed by Amory et al. (2021) for Antarctica is not activated for the Greenland configuration.”

L409: Therefore, future developments should prioritize improving the representation of clouds and associated precipitation, which directly affects both radiative fluxes and accumulation.

You are emphasising a conclusion that is not really backed up by this paper. While I am sure that there are biases coming from clouds and precipitation, and while these issues are coupled also to the surface conditions e.g via albedo, this paper does not evaluate clouds and precipitation. You should conclude what you learned from your evaluation, and if there is a suspicion of poor cloud and precipitation processes from the biases that cannot be explained with surface processes (e.g. Sect 3.3), then we should conclude that clouds and precipitation needs further study. I would like to know, however, what future developments should prioritize, considering this evaluation.

We revised the conclusions to first focus on the processes that were directly evaluated, particularly bare-ice albedo and surface topography. Remaining discrepancies that cannot be explained by the evaluated surface processes may point to the representation of cloud processes and associated precipitation that require further investigation. We also refer to Lindsey-Clark et al. (2025) evaluating MAR accumulation using radar measurements as a valuable insight for future model evaluation. These are now clarified in the revised manuscript as:

Remaining SMB biases at 5 km are primarily associated with biases in bare-ice albedo and surface topography, while additional discrepancies, i.e. not explained by the evaluated surface processes, may point to the representation of clouds and associated precipitation, that require further investigation. Previous MAR evaluations based on radar-derived accumulation rates provide insight for future model evaluation in regions where in situ measurements are sparse such as southeast Greenland (Koenig et al., 2016; Lindsey-Clark et al., 2025).

1.2 Technical corrections

L25 (*PR ; the sum of snowfall and rainfall*) - unnecessary space

Done, thank you.

L36: *Since then, numerous developments have been implemented in versions 3.10–3.14, as well as an increase in spatial resolution over the Greenland domain, i.e., from 15 km (Fettweis et al., 2020) to 5 km in this study.*

Increase of the HIGHEST spatial resolution? As one of the next paragraphs states, 30 km MAR simulations exist, so it would be fitting to show the true extent of the model resolution increase or explicitly state that the highest suggested resolution has increased.

The previous benchmark MAR simulations were conducted at 15 km resolution, whereas the simulations evaluated in this study use 5 km resolution. The 30 km simulations mentioned below were used to force ice sheet models with long-term SMB projections. We have clarified this distinction as follows:

“Since then, numerous developments have been implemented in versions 3.10–3.14, as well as an increase in spatial resolution of benchmark MAR simulations over the Greenland domain, i.e., from 15 km (Fettweis et al., 2020) to 5 km in this study.”

Section 2. This is not a requirement, but I would think that a separate Section (2) should be reserved to MAR only. The data you use to evaluate MAR should have a Section of its own (Section 3). This would lead to a better overall structure of the paper.

Thank you, we revised sections accordingly.

L93: *We use SMB measurements from the GrIS SUMup database (Vandecrux et al. (2024) ; ~~purple dots in Figure 1a~~).* L96: *(Ohmura and Reeh (1991) ; ~~green squares in Figure 1a~~).* Leave these for the caption, not useful in the main text.

We modified as *“SUMup database (Vandecrux et al., 2024)”* and *“(Ohmura and Reeh, 1991)”*.

L94: *Here, we compare modelled SMB with measurements derived from stakes, snow pits, firn and ice cores, snowfox records, mass balance profile estimates and $\delta^{18}O$ dating.*

Be more concise with words: either use “SnowFox” records, or better yet, “ground-based SnowFox cosmic ray neutron sensor snow measurement records”.

Thank you, we reformulated as:

“ground-based SnowFox cosmic ray neutron sensor snow measurement records (see e.g., Howat et al., 2018)”

L157: *Applying topographic correction (Eq. (1)), SMB bias and RMSE decrease by 27 % and 20 % respectively, while correlation increases to 0.95 (Fig. 1c).* Double brackets (Eq. (1)) → (Eq. 1), or [Eq. (1)].

Thank you, we now use (Eq. 1) here and where appropriate.

L158: *The linear regression lines, which exhibit slopes lower than 1, indicate that MAR tends to underestimate both high ablation and high accumulation observed values.*

It is noteworthy that elevation correction barely changes the regression.

Indeed, this is now mentioned as:

“Note that the topography correction barely impacts the coefficients.”

L171: *Comparing MARv3.14 to **airborne** radar transects from Karlsson et al. (2016, 2020) leads to biases, RMSEs and correlations of -0.01 m w.e. and 0.02 m w.e. and 0.99, similar to those of MARv3.9.*

→ *Comparing MARv3.14 to **airborne** radar transects from Karlsson et al. (2016, 2020) **reveals** biases, RMSEs and correlations of -0.01 m w.e. and 0.02 m w.e. and 0.99, similar to those of MARv3.9.* (Biases are already in the model and in the radar measurements). Air-bone radar typo is also in Table S1.

Thank you, we revised these accordingly.

L170: *We find that MARv3.14 (5 km) improves on MARv3.9 (15 km), while model differences remain statistically non-significant.* And elsewhere.

A Results section should be readable and understandable on its own, even without looking at tables. Please back the claims of “improving” and “remain statistically non-significant”. No need to write out all Table S1 results, but please back up your claims either with an example or range of values. Similarly to paragraph expanding L174-L178.

For conciseness, we added the following sentence:

“For instance, the RMSE between modelled and observed SMB decreases by 8.7% in MARv3.14.”

Figure 1: At least inside one Figure, please use the same units, either mm w.e or m w.e

SMB maps are commonly shown in mm w.e. yr^{-1} (equivalent to $\text{kg m}^{-2} \text{yr}^{-1}$). For consistency with Fig. 8 (of the manuscript) showing SMB changes in mm w.e. yr^{-1} , we prefer keeping Fig. 1a (of the manuscript) as is. We deem that reader can easily convert m w.e. in mm w.e. in scatterplots.

L193: *At coarser resolutions (Table S3), the agreement with observations decreases slightly in terms of RMSE, correlation and bias. However, the differences are small, and the RMSEs remain systematically lower than the corresponding standard deviations.*

A Results section should be readable and understandable on its own, even without looking at tables. Please back the claims of “decreases slightly” and “differences are small”.

We added the following:

“For instance, the RMSE increases by 8-93% at 30 km relative to 5 km resolution, depending on the variable.”

Sect 3.3 *Radiative fluxes modelling* but also everywhere else.

The frequent use of brackets to give information and side notes makes it very hard to read. If every sentence has a parenthetical aside, it disrupts reading flow. In some cases, one sentence has several. Rewording some sentences would make the language better. For example: “*Except for the RMSEs which are slightly larger for shortwave fluxes (12 W m^{-2}), no significant differences are found between the annual and the summer (JJA) statistics (Table S4).*” JJA is already defined as summer months previously in the text. Instead of referencing tables in brackets, you can use full sentences: “Table S4 presents ... etc.” In cases of sentences such as these: “*The annual (resp. JJA) standard deviations (in W m^{-2}) are 129.95 (resp. 93.60) for SWD, 50.08 (resp. 35.98) for LWD, 99.01 (resp. 91.84) for SWU, and 38.98 (resp. 18.85) for LWU.*” You could write two sentences, one for annual, one for JJA.

Another example is from Sect. 3.5: *Figure 7a not only shows that MAR accurately reproduces the onset of the melt season (around early June) as well as its ending (mid September), but also that the timing of the peak melt extent is well captured, on average in mid-July.* → *Figure 7a shows that MAR accurately reproduces the onset of the melt season around early June, its ending in mid September, as well as the timing of the peak melt extent in mid-July.*

Please cut down the use of parentheses. If the information is essential to the argument, put it in the main sentence.

We corrected all the examples you mentioned and tried to reduce the amount of information given in parentheses in the rest of the text.

We also tried to refer to Tables and Figures without brackets, following your advice. For instance, in Sect. 3.2 (4.2 in the revised version) (*Near-surface meteorology*), we changed:

At coarser resolutions (Table S3), the agreement with observations ...

To:

“The results for coarser resolutions are listed in Table S3. In general, the agreement with observations ...”

We hope this improves readability.

Sect 3.3 *Radiative fluxes modelling* but also everywhere else

This relates to the previous point. The referencing Table S4 is done in multiple places, and in brackets. The whole section is really a description of Table S4 and Figure 3, so you can lead with “Table S4 and Figure 3 present...”, and then you don’t need to refer to Table S4 separately any more. Same true everywhere else in the Results section.

In the revised version, the section begins with the following sentence: *Figure 3 and Table S4 present a similar comparison, performed for SEB radiative fluxes, i.e., downward/upward shortwave (SWD/SWU) and longwave radiation (LWD/LWU).*

We modified other parts of the texts. For instance in the revised manuscript, Sect. 3.4 (4.4 in the revised version) starts with the following: *“To evaluate modelled albedo representation, we compare MAR to observations from OLCI GEUS and PROMICE & GC-NET. For a meaningful grid-cell by grid-cell comparison, we*

averaged daily OLCI GEUS albedo values onto the MAR grid. As listed in Table S6, this yields a mean bias of -2×10^{-3} , RMSE of 0.08, and a correlation of 0.70.”

L221: *Daily OLCI GEUS albedo values (Vandecrux et al., 2021)* ... also in the other parts of Sect 3.

Be mindful of over-referencing. You have introduced the OLCI GEUS (Vandecrux et al., 2021) dataset in Sect 2. Model and observational data. Later in the text, you can just call it OLCI GEUS without referencing. Same true for the SMB dataset of Vandecrux et al. (2024) in Sect. 3.1.. After establishing reference in Sect. 2, call it GrIS SUMup database. Same true for “AWS observations from the PROMICE and GC-NET networks (Fausto et al., 2021; How et al., 2022; Vandecrux et al., 2023)...” in Sect. 3.2.

In Sect 3.1 (4.1 in the revised version), we replaced “*in situ SMB measurements from Vandecrux et al. (2024), and Ohmura and Reeh (1991)*” by “*in situ SMB measurements from combined GrIS SUMup / Ohmura and Reeh (1991) data sets.*”

In Sect 3.2 (4.2 in the revised version), we replaced ... *observations from the PROMICE and GC-NET networks (Fausto et al., 2021; How et al., 2022; Vandecrux et al., 2023).* by “... *observations from the PROMICE and GC-NET networks.*”

Idem for Sect. 2.5 (3.3 in the revised version).

L269: “*Focussing on the melt extent, it appears that it is slightly ($\sim 2\text{--}5\%$ of the GrIS area) overestimated during the first half of the melt season and underestimated after the peak.*”

The extent of the overestimation is given, but not underestimation. Also, there is no space between “ $\sim$ ” and the number.

Good catch, thanks. We also added the underestimation range as: “... *overestimated during the first half of the melt season and underestimated after the peak ($\sim 3\%$ of the GrIS area).*”

Sect. 3.4 Albedo representation

The Section needs better structure. Use a lead sentence, followed by the supporting sentences for each paragraph. Start with an introductory sentence for the whole Sect. 3.4, as you did in Sect 3.1-3.

As aforementioned, the section now begins with “*To evaluate modelled albedo representation, we compare MAR to observations from OLCI GEUS and PROMICE & GC-NET.*”

L299: *Figure 4c shows that the RMSE exceeds the OLCI-GEUS standard deviation over a large fraction of the GrIS (hatched)...* No need to add “(hatched)”, as the colour of the points, it belongs to caption, not text.

We removed “hatched” accordingly.

Figure 6 and Figure 7:

This information belongs to Sect 2.6, not into the Captions: using a melt-based criterion of 0.2 mm w.e. day⁻¹ (although fine if you want to really emphasise this), satellite data are derived from the ascending and descending passes (combined with “OR” criterion) of the AMRS-derived product Picard (2025) from Picard and Fily (2006). The comparison is made over the satellite data grid using an IDW interpolation. Furthermore, I cannot see Picard and Fily (2006) referenced in Sect 2.6 and the “the AMRS-derived product Picard (2025) from Picard and Fily (2006)” is either typo or confusing sentence structure.

Indeed, we deem it better to emphasize the criterion we used and kept the sentence as is. We also added the reference to Picard and Fily (2006) in Sect. 2.6 (3.4 in the revised version).

L311: “*In the previous sections, we evaluated the latest MAR version (MARv3.14) at 5 km resolution against a set of observational data including in situ SMB (Fig. 1), meteorological (Fig. 2) and SEB (Fig. 3) records as well as albedo (Figs.4 and 5) and remotely sensed melt extent (Figs. 6 and 7).*”

No need to reference the Figures here again. A simpler sentence without all the parentheses would make it easier to read: In the previous sections, we evaluated MARv3.14 at 5 km resolution against a set of observational data including in situ SMB, meteorological and SEB records as well as albedo and remotely sensed melt extent.

Indeed, we applied your suggestion in the revised manuscript.

L321: *To investigate the origin of the ablation biases at 5 km, we compare MAR to summer (JJA) SMB ...* Summer is already established as JJA, so please pick one to limit usage of brackets.

Thank you, we now refer to “*JJA SMB*”.

L326: “Interestingly, the comparison with pressure transducer data (Figure S3a) yields similar results to those shown in Figure 1b, despite the smaller sample size (137 vs 3012 points in Fig. 1).”

Please write out what comparison was shown in Figure 1b.

We modified as follows: “Interestingly, the comparison of MAR-simulated SMB with pressure transducer data (Figure S3a) yields similar results to those obtained using SUMup data, shown in Figure 1b, despite the smaller sample size (137 vs 3012 points in Fig. 1).”

L338: “However, while this radiative bias should result in an ablation overestimation of only ~ 0.7 mm w.e., we observe a bias of ~ 2 mm w.e..”

Double punctuation.

Good catch, thank you.

Sect. 4.1 Ablation underestimate

Here, the first paragraphs with station comparisons have metrics to back up claims of biases. But, from KAN_L and KAN_M comparisons, not. Only words like “underestimation”, “overestimation”, “small” are used. You need to use key findings in numbers consistently to prove physical points, explain biases, or compare with literature.

We added numbers in the text accordingly:

- “MAR slightly underestimates albedo at UPE_L and KAN_L, by 0.03 on average.”
- “...MAR overestimates SMB, by 3.2 and 3.6 mm w.e. day⁻¹”

Sect 5. Conclusions

Please provide concrete, specific outcomes rather than vague, generalized summary.

There were a lot of comparisons made, and a Conclusions section is the perfect place to summarize the main points found in the analysis that are of interest for a MARv3.14 (data) user.

We added the following: “Regarding other variables, MARv3.14 well represents SEB radiative fluxes (RMSE of 11–35 W m⁻²) as well as climate fields, including temperature, specific humidity, wind speed and surface pressure, with RMSEs lower than the natural variability. The melt extent evolution is overall well captured ($= 0.95$), although spatial differences remain between MAR and AMSR.”

2 Answer to Reviewer 2 (anonymous)

This work evaluates a new version of MAR against a range of ground and satellite observations across Greenland. There is thorough statistical analysis spatially from both satellite and station measurements. Overall, this manuscript is well-written and provides a clear benchmark of the MAR model. It is particularly useful to compare benefits from higher resolution versus computation time. The evaluation is fairly comprehensive for mean values, and it does acknowledge areas for model improvement, e.g., cloud representation.

Thank you for your review and constructive comments. Below you will find our point-by-point response to your comments in blue. Changes applied to the revised manuscript are shown in red.

I believe a couple relatively minor improvements are needed. First, in lines 113-4 the authors state that they only compare MAR to observed radiative fluxes. This is true through the results, in which case I think the language in the abstract and elsewhere would be more precise to say that comparisons are to surface radiative fluxes rather than mentioning “SEB fluxes”, e.g., line 9. However, Table 2 and section 4.1 actually include turbulent heat fluxes. This inclusion means that the turbulent fluxes should be described in the methods, including if they’re bulk or eddy covariance, uncertainty, and which direction is positive, and the language that only radiative fluxes are included should be removed.

Thank you. We now describe the turbulent heat flux data in Section 3.2 (“Automatic weather stations data”) as:

“Sensible and latent heat fluxes are not directly measured (e.g., via eddy covariance) but are instead calculated using a bulk aerodynamic method based on Monin–Obukhov similarity theory (Fausto et al., 2026). These derived turbulent fluxes show large uncertainties due to their sensitivity to surface roughness parameterization (z_0), stability functions, and surface temperature estimates, and should therefore be interpreted with caution (Fausto et al., 2026). For both sensible and latent heat fluxes, positive values denote energy flux directed towards the surface.”

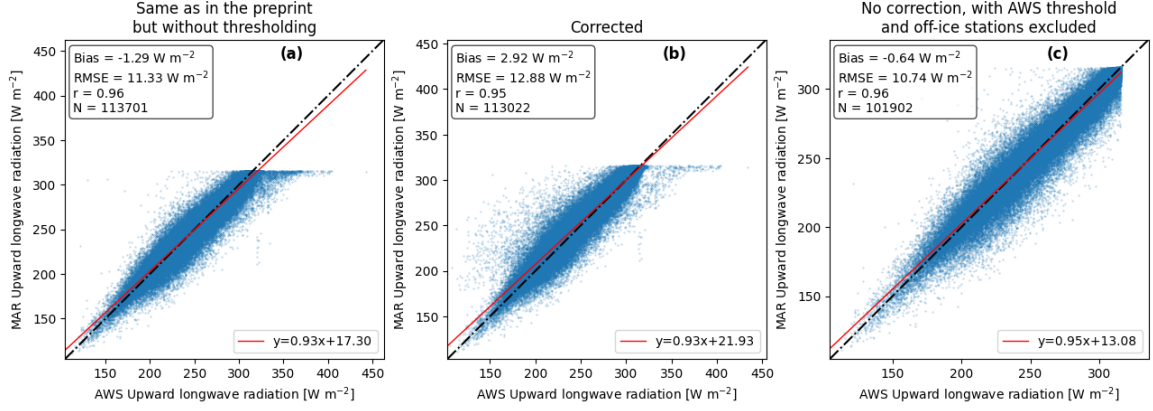


Figure 3: Comparison of MAR at 5 km resolution with AWS LW_U measurements. Panel (a) shows the comparison without applying the 315.7 W m^{-2} threshold to the AWS observations. Panel (b) shows the comparison after applying the LW_U correction proposed by Sledd et al. (2026); the number of observations is slightly lower because the correction requires SW_U measurements. Panel (c) shows the comparison retained in the revised analysis, without applying the correction, but excluding AWS observations exceeding the 315.7 W m^{-2} threshold as well as the off-ice stations KAN_B, NUK_B, WEG_B, and SER_B.

Second, the authors note unphysical longwave fluxes on lines 118-120. There are unphysical upward longwave fluxes because the downward looking radiometer absorbed shortwave radiation reflected from the surface (e.g., Kuipers Munneke et al. 2009). While the biases are only obviously wrong when upwelling longwave is greater than $\sim 315.7 \text{ W m}^{-2}$, the values are biased for all times with incoming shortwave radiation. I doubt this will significantly impact the results, but the upwelling longwave fluxes should be corrected (e.g., Sledd et al. 2026) instead of using biased values.

We thank the reviewer for this relevant remark. As suggested, we investigated the implementation of the empirical solar-heating correction proposed by Sledd et al. (2026), which accounts for the potential low-level daylight bias in upward longwave radiation measurements (see Figure 3).

For some stations, the correction produced encouraging results (see Figure 4 for the KPC_L station). However, the empirical relationship underlying the correction, between ΔLW_U and SW_U , is weak at several stations (with R^2 values of only 0.01–0.1). The correction is therefore poorly constrained at these sites, and its application may introduce additional, station-dependent uncertainty rather than providing a robust bias correction.

For completeness, we also explored the correction proposed by van Tiggelen et al. (2025), which additionally accounts for wind speed. This approach likewise did not lead to clear improvements. We therefore decided to keep AWS observations uncorrected.

Nevertheless, we excluded AWS observations exceeding the theoretical upper limit (i.e., $(1-\epsilon)LW_D + \epsilon\sigma T_m^4 \sim 315.7 \text{ W m}^{-2}$, with $T_m = 273.15 \text{ K}$ and $\epsilon \simeq 0.95$), as well as off-ice stations KAN_B, NUK_B, and WEG_B, located on tundra areas, and SER_B, which is located on bedrock. At these stations upward longwave radiation threshold for melting ice is not applicable. We acknowledge that excluding observations above this threshold does not remove potential solar-heating biases below the threshold. These residual observational biases remain a source of uncertainty in the LW_U evaluation.

This is clarified in the revised manuscript as:

“Because this threshold [315.7 W m^{-2}] is invalid for surfaces exceeding 0°C , off-ice stations (i.e., KAN_B, NUK_B, and WEG_B on tundra, and SER_B on bedrock) were excluded specifically from the upward longwave flux evaluation. Although downward-looking radiometers can absorb reflected shortwave radiation—introducing a positive daylight bias in upward longwave measurements (e.g., Kuipers Munneke et al., 2009)—testing empirical solar-heating corrections (e.g., van Tiggelen et al., 2025; Sledd et al., 2026) yielded inconsistent performance across stations. We therefore retained the uncorrected upward longwave measurements throughout the analysis and acknowledge potential residual solar-heating biases.”

Some additional small language changes would improve readability, also noted below.

L 94: What are the snowfox records?

The SnowFox instrument is a cosmic-ray neutron sensor that estimates snow water equivalent/accumulation in situ (see e.g. Howat et al., 2018). We fixed the capitalization issue and added a reference to Howat et al.

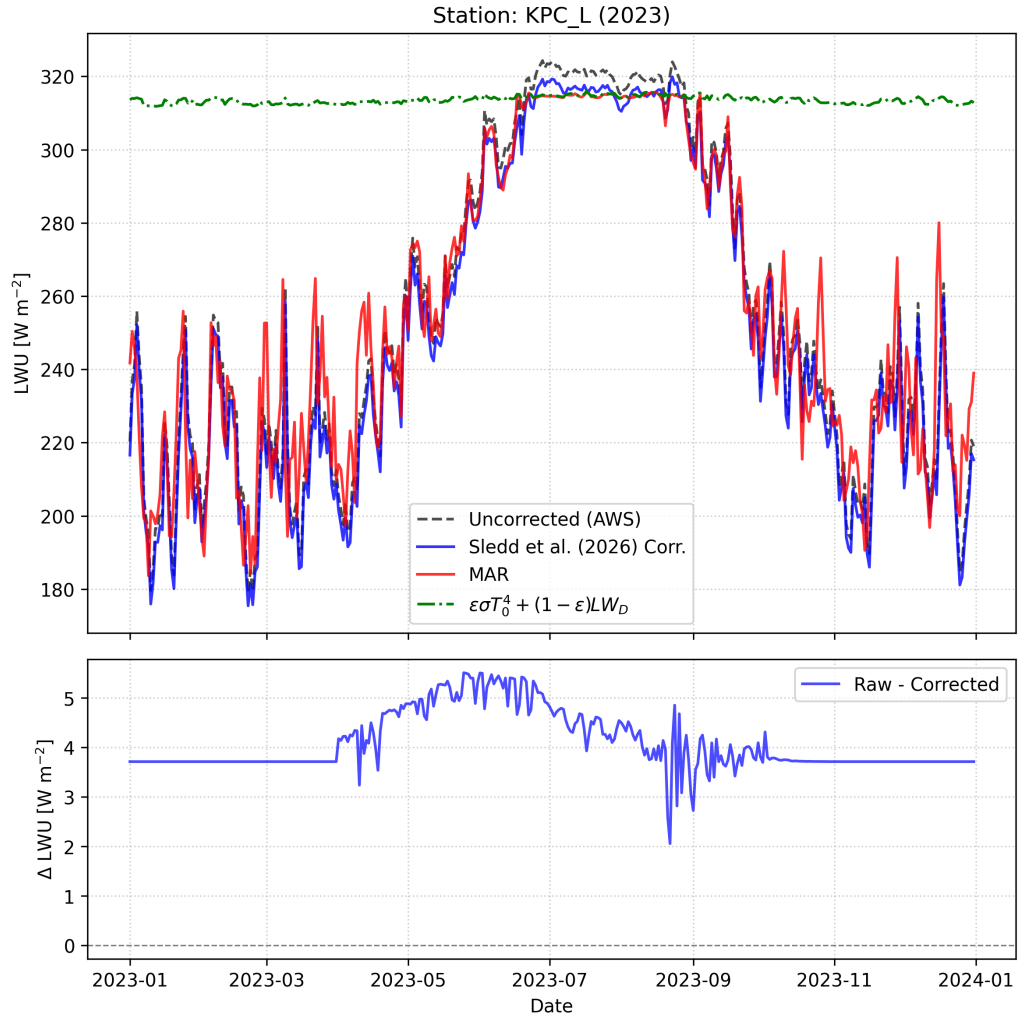


Figure 4: Comparison of raw (grey dashed line), corrected (blue), and MAR (at 5km resolution) (red) LW_U fluxes for summer 2023 at KPC_L station.

(2018) in the text to avoid confusion: “... ground-based SnowFox cosmic ray neutron sensor snow measurement records (see e.g., Howat et al., 2018) ...”

L 170-171: How can a version both be an improvement and statistically the same?

This was confusing, we reformulated as: “*We find that differences between MARv3.14 (5km) and MARv3.9 (15km) are statistically non-significant.*”

L 201: A nearly 10 W m^{-2} LWD bias seems notable compared to the magnitudes of upwelling shortwave and longwave fluxes, no?

We agree that the LWD bias of -9.78 W m^{-2} is notably larger than that found for the other radiative fluxes. However, this bias remains relatively small compared with the magnitude of LWD variability (standard deviation). In particular, the annual (resp. JJA) standard deviation of LWD is 50.08 (resp. 35.98) W m^{-2} , such that the mean bias represents approximately 20% (resp. 27%) of the observed variability. We have clarified this point in the revised manuscript as:

“*Despite being larger than the biases of the other fluxes, this bias remains substantially lower than the observed LWD variability (standard deviation of 50.08 W m^{-2} annually and 35.98 W m^{-2} in JJA)*”

L 249-50: I found this sentence difficult to understand the first time I read it. Maybe revise for clarity?

Indeed, there was a typo claiming albedo “underestimation” instead of “overestimation” in the ablation zone. This is probably where the confusion originated. We corrected this as:

“*This underestimation of absorbed shortwave energy in the ablation zone is consistent with the modelled ablation underestimation, that is especially pronounced for SMB values below $\sim -4 \text{ m w.e.}$, as shown in Fig. 1b-c.*”

L 262: Given this concern, is there are better threshold/metric to use for melt to compare the model and satellite observations? Is the threshold for melt from MAR only considered at the surface or does it include melt in the full column?

The melt threshold holds for the melt in the full column.

It is also possible to use a criterion on the amount of liquid water at the surface of the snow mantle to compare the model to satellite observations. But it involves opting for a depth threshold and a liquid water content threshold. Also, as MAR considers bare ice as impermeable, it is difficult to use a liquid water content based criterion.

L 265-6: Would this still be included as melt from MAR? Is there a way to define surface melt in MAR for comparison with the observations?

This is currently not implemented, i.e., the model outputs do not mention at what depth the melt occurs.

L329: At the end of this sentence, it would have been helpful to mention that the following paragraphs are referencing Table 2.

We added a reference to Table 2.

L 433: You’re using the density of ice, but if you’re considering surface melt wouldn’t that include some fresh snow with lower density for some locations (e.g., Fausto et al. 2018)?

The pressure-transducer ablation measurement technique used here measures ice ablation rather than snow ablation, which motivates our use of ice density. We now also provide the value in millimetres of water equivalent per day (mm w.e. day^{-1}), which is independent of the initial snow or ice density. We modified the text as follows:

“*Using the density of pure ice, $\rho = 917 \text{ kg m}^{-3}$, a radiative bias of $F = 1 \text{ W m}^{-2}$ corresponds to a melt rate of approximately $\Delta z \approx 0.28 \text{ mm of ice day}^{-1}$, i.e., $\sim 0.26 \text{ mm w.e. day}^{-1}$.*”

3 Bibliography

Amory, C., Kittel, C., Le Toumelin, L., Agosta, C., Delhasse, A., Favier, V., and Fettweis, X.: Performance of MAR (v3.11) in simulating the drifting-snow climate and surface mass balance of Adélie Land, East Antarctica, Geoscientific Model Development, 14, 3487–3510, <https://doi.org/10.5194/gmd-14-3487-2021>, 2021.

- Fausto, R. S., How, P., Vandecrux, B., Lund, M. C., Box, J. E., Mankoff, K. D., Andersen, S. B., van As, D., Bahbah, R., Citterio, M., Colgan, W., Jakobsgaard, H. T., Karlsson, N. B., Kjeldsen, K. K., Larsen, S. H., Olsen, C., Oraschewski, F. M., Rutishauser, A., Shields, C. L., Solgaard, A. M., Stevens, I. T., Svendsen, S. H., Langley, K., Messerli, A., Bjørk, A. A., Andersen, J. K., Abermann, J., Steiner, J., Prinz, R., Hynek, B., Lea, J. M., Brough, S., and Ahlstrøm, A. P.: PROMICE | GC-NET automatic weather station data, *Earth System Science Data*, 18, 2829–2873, <https://doi.org/10.5194/essd-18-2829-2026>, 2026.
- Fettweis, X., Tedesco, M., van den Broeke, M., and Ettema, J.: Melting trends over the Greenland ice sheet (1958–2009) from spaceborne microwave data and regional climate models, *The Cryosphere*, 5, 359–375, <https://doi.org/10.5194/tc-5-359-2011>, 2011.
- Fettweis, X., Box, J. E., Agosta, C., Amory, C., Kittel, C., Lang, C., van As, D., Machguth, H., and Gallée, H.: Reconstructions of the 1900–2015 Greenland ice sheet surface mass balance using the regional climate MAR model, *The Cryosphere*, 11, 1015–1033, <https://doi.org/10.5194/tc-11-1015-2017>, 2017.
- Fettweis, X., Hofer, S., Krebs-Kanzow, U., Amory, C., Aoki, T., Berends, C. J., Born, A., Box, J. E., Delhasse, A., Fujita, K., Gierz, P., Goelzer, H., Hanna, E., Hashimoto, A., Huybrechts, P., Kapsch, M.-L., King, M. D., Kittel, C., Lang, C., Langen, P. L., Lenaerts, J. T. M., Liston, G. E., Lohmann, G., Mernild, S. H., Mikolajewicz, U., Modali, K., Mottram, R. H., Niwano, M., Noël, B., Ryan, J. C., Smith, A., Streffing, J., Tedesco, M., van de Berg, W. J., van den Broeke, M., van de Wal, R. S. W., van Kampenhout, L., Wilton, D., Wouters, B., Ziemen, F., and Zolles, T.: GrSMBMIP: intercomparison of the modelled 1980–2012 surface mass balance over the Greenland Ice Sheet, *The Cryosphere*, 14, 3935–3958, <https://doi.org/10.5194/tc-14-3935-2020>, 2020.
- Forster, R. R., Box, J. E., Van Den Broeke, M. R., Miège, C., Burgess, E. W., Van Angelen, J. H., Lenaerts, J. T., Koenig, L. S., Paden, J., Lewis, C., et al.: Extensive liquid meltwater storage in firn within the Greenland ice sheet, *Nature Geoscience*, 7, 95–98, 2014.
- Howat, I. M., de la Peña, S., Desilets, D., and Womack, G.: Autonomous ice sheet surface mass balance measurements from cosmic rays, *The Cryosphere*, 12, 2099–2108, <https://doi.org/10.5194/tc-12-2099-2018>, 2018.
- Humphrey, N. F., Harper, J. T., and Meierbachtol, T. W.: Physical limits to meltwater penetration in firn, *Journal of Glaciology*, 67, 952–960, <https://doi.org/10.1017/jog.2021.44>, 2021.
- Karlsson, N. B., Eisen, O., Dahl-Jensen, D., Freitag, J., Kipfstuhl, S., Lewis, C., Nielsen, L. T., Paden, J. D., Winter, A., and Wilhelms, F.: Accumulation Rates during 1311–2011 CE in North-Central Greenland Derived from Air-Borne Radar Data, *Frontiers in Earth Science*, Volume 4 - 2016, <https://doi.org/10.3389/feart.2016.00097>, 2016.
- Karlsson, N. B., Razik, S., Hörhold, M., Winter, A., Steinhage, D., Binder, T., and Eisen, O.: Surface accumulation in Northern Central Greenland during the last 300 years, *Annals of Glaciology*, 61, 214–224, <https://doi.org/10.1017/aog.2020.30>, 2020.
- Koenig, L. S., Ivanoff, A., Alexander, P. M., MacGregor, J. A., Fettweis, X., Panzer, B., Paden, J. D., Forster, R. R., Das, I., McConnell, J. R., Tedesco, M., Leuschen, C., and Gogineni, P.: Annual Greenland accumulation rates (2009–2012) from airborne snow radar, *The Cryosphere*, 10, 1739–1752, <https://doi.org/10.5194/tc-10-1739-2016>, 2016.
- Kuipers Munneke, P., van den Broeke, M. R., Reijmer, C. H., Helsen, M. M., Boot, W., Schneebeli, M., and Steffen, K.: The role of radiation penetration in the energy budget of the snowpack at Summit, Greenland, *The Cryosphere*, 3, 155–165, <https://doi.org/10.5194/tc-3-155-2009>, 2009.
- Lindsey-Clark, J. Y., Grinsted, A., Vandecrux, B., and Schøtt Hvidberg, C.: Greenland Monthly Accumulation Maps (1960–2022): A Statistical Semi-Empirical Bias-Adjustment Model, *EGU sphere*, 2025, 1–45, <https://doi.org/10.5194/egusphere-2025-2516>, 2025.
- Noël, B., van de Berg, W. J., van Meijgaard, E., Kuipers Munneke, P., van de Wal, R. S. W., and van den Broeke, M. R.: Evaluation of the updated regional climate model RACMO2.3: summer snowfall impact on the Greenland Ice Sheet, *The Cryosphere*, 9, 1831–1844, <https://doi.org/10.5194/tc-9-1831-2015>, 2015.
- Noël, B., Van De Berg, W. J., Van Wessem, J. M., Van Meijgaard, E., Van As, D., Lenaerts, J. T., Lhermitte, S., Kuipers Munneke, P., Smeets, C. P., Van Ulft, L. H., et al.: Modelling the climate and surface mass balance of polar ice sheets using RACMO2–Part 1: Greenland (1958–2016), *The Cryosphere*, 12, 811–831, 2018.

- Noël, B., van de Berg, W. J., Lhermitte, S., and van den Broeke, M. R.: Rapid ablation zone expansion amplifies north Greenland mass loss, *Science Advances*, 5, eaaw0123, <https://doi.org/10.1126/sciadv.aaw0123>, 2019.
- Ohmura, A. and Reeh, N.: New precipitation and accumulation maps for Greenland, *Journal of Glaciology*, 37, 140–148, <https://doi.org/10.3189/S0022143000042891>, 1991.
- Picard, G.: Near real-time Greenland melt extent from AMSR2 passive microwave observations at 10 km resolution (v2025), <https://snow.univ-grenoble-alpes.fr/opendata/melt-AMSRU-greenland-10km.zarr>, based on Picard and Fily (2006), *Remote Sensing of Environment*, 104(3), 325–336. Data from AMSR2 (JAXA). Accessed 2025-10-22., 2025.
- Puggaard, A., Hansen, N., Mottram, R., Nagler, T., Scheiblauer, S., Simonsen, S. B., Sørensen, L. S., Wuite, J., and Solgaard, A. M.: Bias in modeled Greenland Ice Sheet melt revealed by ASCAT, *The Cryosphere*, 19, 2963–2981, <https://doi.org/10.5194/tc-19-2963-2025>, 2025.
- Sledd, A., Gallagher, M. R., Shupe, M. D., Cox, C. J., Hawley, R., Town, M. S., Guy, H., Marshall, H.-P., Neely III, R. R., Pettersen, C., Walden, V. P., Hebson, C., Martin, A., Olson, E., and Pickell, D.: Summer subsurface temperature variability in the percolation zone of southwest Greenland: high resolution observations of the top meter of firn, *EGU sphere*, 2026, 1–40, <https://doi.org/10.5194/egusphere-2026-1842>, 2026.
- van Tiggelen, M., Smeets, P. C. J. P., Reijmer, C. H., Kuipers Munneke, P., and van den Broeke, M. R.: Institute for Marine and Atmospheric Research Utrecht (IMAU) Antarctic automatic weather station data, including surface radiation balance (1995–2022), *Earth System Science Data*, 17, 4933–4955, <https://doi.org/10.5194/essd-17-4933-2025>, 2025.
- Vandecrux, B., Amory, C., Ahlstrøm, A. P., Akers, P. D., Albert, M., Alley, R. B., de Castro, M. A., Arnaud, L., Baker, I., Bales, R., Benson, C., Box, J. E., Brucker, L., Buizert, C., Chandler, D., Charalampidis, C., Cherblanc, C., Clerx, N., Colgan, W., Covi, F., Dattler, M., Denis, G., Derksen, C., Dibb, J. E., Ding, M., Dixon, D., Eisen, O., Fahrner, D., Fausto, R., Favier, V., Fernandoy, F., Freitag, J., Gerland, S., Harper, J., Hawley, R. L., Heuer, J., Hock, R., Hou, S., How, P., Humphrey, N., Hubbard, B., Iizuka, Y., Isaksson, E., Kameda, T., Karlsson, N. B., Kawakami, K., Kjær, H. A., Kreutz, K., Munneke, P. K., Lazzara, M., Lemeur, E., Lenaerts, J. T. M., Lewis, G., Lindau, F. G. L., Lindsey-Clark, J., MacFerrin, M., Machguth, H., Magand, O., Mankoff, K. D., Marquette, L., Martinerie, P., McConnell, J. R., Medley, B., Miège, C., Miles, K. E., Miller, O., Miller, H., Montgomery, L., Morris, E., Mosley-Thompson, E., Mulvaney, R., Niwano, M., Oerter, H., Osterberg, E., Otosaka, I., Picard, G., Polashenski, C., Reijmer, C., Rennermalm, A., Rutishauser, A., Scanlan, K., Simoes, J. C., Simonsen, S. B., Smeets, P. C. J. P., Smith, A., Solgaard, A., Spencer, M., Steen-Larsen, H. C., Stevens, C. M., Sugiyama, S., Svensson, J., Tedesco, M., Thomas, E., Thompson-Munson, M., Tsutaki, S., van As, D., den Broeke, M. R. V., van Tiggelen, M., Wang, Y., Wilhelms, F., Winstrup, M., Xiao, J., and Xiao, C.: The SUMup collaborative database: Surface mass balance, subsurface temperature and density measurements from the Greenland and Antarctic ice sheets (2024 release), *Arctic Data Center*, <https://doi.org/10.18739/A2M61BR5M>, version: urn:uuid:35eb18e0-8107-4458-bc62-3cb194bb395f, 2024.