A review of « Impact of updated radiative transfer scheme in RACMO2.3p3 on the surface mass and energy budget of the Greenland ice sheet » by van Dalum, et al., submitted to *The Cryosphere*.

Overview

This manuscript presents the updated radiative transfer scheme of RACMO2.3p3, its impact on surface mass balance (SMB), and surface energy budget (SEB). The updated scheme enables the representation of subsurface warming causing by radiation penetration.
Considering these new developments, the results of SMB, SEB, and their respective components are compared with various in-situ observations (automatic weather station of K-transect, PROMICE and temperature profile at Summit) and the former version of the model (RACMO2.3p2).
The SMB representation is improved, compared to the previous version of the model, in the percolation zone and more generally around the ice margin. Comparison of subsurface temperatures in different snow layers at Summit is good. This updated radiative transfer scheme enables therefore to improve the representation of the subsurface melt extension.

The authors correctly evaluate an original radiative transfer scheme in RACMOp. The manuscript is very interesting and I'm curious to read what this brings to the RACMOp projections. However, I raised many points that should be considered to improve the manuscript. Since they mainly concern the presentation and form, I recommend a minor revision as it doesn't request further developments or experiments.

General comments

This study first describe and evalue the new and original representation of a physical process trough an updated radiative scheme in RACMOp. Without discussing anything on the scientific background of this manuscript, in its current state, the way of expressing ideas sometimes does not enable an immediate understanding. Indeed, some passages require several readings for a complete understanding. For instance, in section 4, the links between figures and explanations or justifications of processes present are not very clear and could be improved and better explained. A series of comments were made in an attempt to fill this lack of immediate understanding throughout the manuscript (see specific comments).

Moreover the structure of the manuscript is rather surprising. The updated version of the radiative scheme is partially evaluated in van Dalum et. al. (2020, published in TC), but the part concerning radiation penetration was not evaluated there. After reading the title and the abstract, I mainly expected in the present to see an evaluation of RACMO with this new radiative scheme, and in a second step a comparison of the two budgets to assess their respective influence. However, the manuscript is constructed the opposite way which might be not very intuitive in my opinion. I would therefore suggest to first bring the sensitivity experiment which represents an evaluation of the radiation penetration part of the radiative scheme alone and describe the internal energy absorption (what its presented in the title), then to show its impact on the subsurface temperature profile, on the SEB and finally on the SMB and its components compared to previous RACMO versions.
Indeed, discussing and describing the internal energy abortion regionally (as for example in paragraph 5.1) firstly would help to provide a basis for the subject discussed throughout the results, and would undoubtedly lead to a better understanding on first reading.
Several times the authors insist on the evaluation of the new albedo and radiative transfer scheme, but the part concerning albedo is mainly evaluated in a previous paper (van Dalum et al., 2020, published in TC). Although closely related, it is obvious that links have be made with this first paper. However, one would not expect to see an evaluation of the new albedo scheme in this

manuscript (despite physical connections between the albedo and radiative scheme). Yet the distinction between the two is weak in the abstract, in the introduction and even in the methodology. Furthermore, comparing RACMO2.3p2 with RACMO2.3p3 (main parts of the manuscript) amounts to compare both improvement in the radiative scheme and albedo preventing the reader to assess “the impact of (the) updated radiative transfer scheme in RACMO2.3p3 on the surface mass and energy budged of the Greenland ice sheet”.

In the introduction, the author could relate the new radiation scheme more to other works and references: Has this already been done with another climate model? What is the relevance of using this new scheme and not another one? (...?) In general, the manuscript could benefit from adding more broad-scale context and impacts. While exhaustive evaluation of climate models are required and welcome to know the biases of the models (to be put in perspective with the projected changes), the broader scientific interest/question could be more detailed to move away from the papers that can be found in Geoscientific Model Development (ie, only evaluation without any/very few scientific discussion).

In order to bring more scientific credibility to the manuscript, I highly advise authors to reformulate each comparison described in the text, and to evaluate its significance using a simple statistical test (student test for example) or at least a comparison of statistical variability (RMSE) to the observed biases. Each time the authors qualify an increase/decrease/... terms with poor scientific value ("considerable", "important",...) are used that do not reflect scientific rigour. This comment is valid for all the variables studied, whether they concern the SMB, the SEB, the temperature. I would suggest the author to define at least a threshold for stating about “important” changes.

Specific comments

- P. 1, L. 1: The abstract requires an introduction sentence to situate this work in its broader scientific context (also true for its conclusion). The authors should also try to introduce in the abstract what the study brings in a wider scientific context than the internal and technical improvement of the model used.

- P. 1, L. 3: « […] as subsurface heating by radiation penetration now occurs ». I suggest to reformulate for instance like this: « […] as the representation of radiation penetration enables to simulate subsurface heating. »

- P. 1, L. 18: « Snow and ice melt typically dominate the SMB [...] », in your SMB equation (2) on P. 4, melt is not a component of the SMB and can’t then dominate SMB? Since it rather runoff that dominates the SMB over these areas, this sentence should be rephrased.

- P. 2, L. 34: Concerning melt, don't the downscaling techniques, commented above (statistical downscalling), enable us to avoid problems linked to poorly represented topography with a coarse resolution? Doesn't this lack of precise representation of the topography affect the SMB precipitation component more than the runoff component (and associated melting)?

- P. 2, L. 36-42: Throughout this paragraph, the theoretical description of radiation penetration and its influences is well written. However, the link with modeling is poorly introduced. For instance, I suggest to move this sentence: « Parametrizations of radiative … (Fettweis et al., 2017). » to place it at the end of the paragraph, and rephrase it by adding information about how these processes are now represented (or not) in RCMs.

- P. 4, eq (2): Authors detailed runoff components, but not erosion (ER) ones. Following the SMB equation over Greenland in Lenaerts et al. (2012, 2014), drifting snow (DS) erosion is also associated with DS sublimation. Could you clarify the different components in your equation (2)?

- P. 4, L. 99-100: Radiative scheme is called each hour. Have some sensitivity experiments been carried out on the call frequency of the radiative scheme? Does this have an influence on SEB and subsurface heating results?

- P. 4, L. 109: Could you specify the order of magnitude or the average height of the snow/ice layers considered for internal energy absorption and SEB?

- P. 5, L. 137: Has an evaluation of the model performance forced by the ERA5 reanalysis already been carried out? Could this influence/improve the results of the new radiation scheme evaluation?

- P. 8, L. 167: What elevation difference threshold did you choose to not use an AWS?

- P. 8, L. 170: For a homogeneity of the method, I suggest to chose a single RACMO grid point, the closest one, for comparison with SMB and SEB observations, as for subsurface comparisons.

- P. 8, L. 176: Please specify that you are talking about annual SMB, in the main text and in captions of your figures even if the units give some clues on this (also for SMB components).

- P. 8, L. 178: Which statistical parameter corresponds to 20 mm w.e. yr^{-1}? It would make more sense to use a local statistical comparison given the large variation in intensity of the components of SMB and SMB (20 mm w.e. yr^{-1} seems to be high over the centre of the ice sheet but particularly low over margins where SMB has more viariabily). As suggested in general comments, a t-test or a comparison to the interannual variability of each grid point (RMSE) could be performed. This could be plotted on the spatial representations of the results as hatched if the results are significant or not. This will also strengthen the conclusions of the manuscript.

- P. 8, L. 181: Integrated SMB is usually given in Gt yr^{-1}.

- P. 8, L.186: « a strong SMB increase » Please specify that you are referring to Figure 2b.

- P. 8, L188: « The outer rim of the ice sheet, except in the southeast, is characterized by a strong SMB increase. In Rp2 at 11 km, the bare ice albedo of the majority of the outermost glaciated points is 0.30 due to contamination with tundra albedo, causing too much melt and runoff. This artifact is mitigated for higher resolutions and is solved in Rp3 (Van Dalum et al., 2020), lowering the snow melt and runoff and increasing the SMB. »
When reading this paragraphs, I understand that this bias is mostly caused by the spatial resolution (mix between tundra and ice albedo). The lower also leads to a surestimation of associated melt and runoff. For me, this is not corrected by the improvements in Rp3. If the runoff biases really comes from the resolution, and since both Rp2 and Rp3 simulations are performed at the same resolution, it would be better to state that the biases is compensated by the improvements of the radiative scheme and not mitagtaed (since you have’nt corrected the resolution that leads to the contamination)..

- P. 9, L. 194: Why subsurface melting of ice creates pore space? Could you more explain this affirmation and not just refer to van Dalum et al (2020). It deserves to be clearly explained, especially since it is repeated in the conclusion. This result seems counterintuitive since melt leads to an increase in density and thus to a reduction in available space.

- P. 10-11, subsection 3.2 and Figure 4: If you consider melt events (and not per year), why do you chose a threshold per year (250 mm yr^{-1} in the caption of the Figure 4)?

- P. 10, L. 210: Results averaged over ice sheet should be given in Gt yr^{-1}, same comment than in L. 181.

- P. 13, L. 240: Note that R^2 is the abbreviation for determination coefficient of a regression line (that expresses the part of a variable that can be expressed by another) while the coefficient correlation (that associates variations of two variables) is noted by R. Annotation (R^2 to R) or the use of a non-adapted variable (correlation to determination) should also be corrected in the Table 1, Figure 5, 6 and B1 captions depending on what you meant.

- P.13, L. 241-243: « There is a tendency for Rp3 to underestimate LWd during cold and dry, cloud-free winter (small SWd) conditions, resulting in underestimated LWu for the same cold days. » Please could the authors clarify these explanations and what they are basing this reasoning on?

- P. 14, Figure 7: Do you have an idea of why Rp3 improves SHF of Rp2 during Jul-Aug-Sep at S6? (while Rp3 has poorer results for this variable in the rest of the comparison)

- P. 15, L. 259-261: « The introduction of a new radiation penetration and snow albedo scheme reduces the differences with observations for Rp3 compared to Rp2. This is especially noticeable at the onset of the accumulation season [...] »: Is this amelioration statically significant? At what point is improvement considered? Do you have hypotheses to explain these discrepancy (especially in summer for SWn)?
More generally, comparison between Rp2 and Rp3 (mainly Table 1 and Figure 7) suggest a relatively limited amelioration (and even deterioration of the results) poorly discussed in the text.

- P. 16, L. 296-297: « For bands 7 to 12, almost all incoming radiation is absorbed at S6 due to the large grain radius and density. For these bands a larger fraction of energy contributes to the SEB at S6 compared to Summit. » : Figure 9 for Bands 7 to 12 at S6 suggests the opposite, the authors would probably write « a lower fraction » instead of « a larger fraction »?

- P.17, L. 308 and 309: Please specify what do you mean by « typical winter » (cloud, precipitaitons, …?) and « extrodinary warm summer day ».

- P. 21, L. 377: Please justify negligible.

- p. 22: Several times in the results sections, Rp3 reveals a weakness in the representation of turbulent fluxes (LHF and SHF). This deserves a point of improvement in the model in your conclusion.

<u>*Tables and Figures*</u>

- Figure 2: « Subsurface temperature profiles are available for Summit » Please specify that is the green dot.

- Table 1: Please align the numbers to the right for better visibility ; Add unit for bias in the legend.

- Figure 5: Please add bias and its unit in the legend.

- Figure 6: Please add unit for bias in the legend.

- Figure 7: For more visibility, could you enlarge or thicken the coloured lines in the legend (especially for LWn, SHF and LHF)?

- Figure 7: At certain places in the figure, the values of the three time series are similar so that it's complicated to distinguish them. This is not a problem when they are all three grouped together, but when two are grouped together and not the third one, it is difficult to distinguish which one is hidden behind the other. This can lead to a misunderstanding of the figure.

- Figure 10c: Please justify why this specific grid point is choosen (SE GRL, star on Figure 2).

Additional references

P. 2, L. 32: « The SMB components can be statistically downscaled to an even higher resolution of up to 1 km (Noël et al., 2019) ». The authors can add these two references about dwonscalling :

Franco, B., Fettweis, X., Lang, C., and Erpicum, M.: Impact of spatial resolution on the modelling of the Greenland ice sheet surface mass balance between 1990–2010, using the regional climate model MAR, The Cryosphere, 6, 695–711, doi:10.5194/tc6-695-2012, 2012

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 Discuss., https://doi.org/10.5194/tc-2019-321, accepted, 2020.

- P. 2, L. 33: I suggest the authors to refer to Fettweis et al. (2020) for a more detailed RCM evaluation from different models.