Response to comments from Reviewer 1

General comments:

This paper demonstrates that the estimations of the Antarctic ice sheet surface melt volumes by the regional climate model RACMO2 can be improved significantly if a deep multilayer perceptron (MLP) model is employed to correct the RACMO2 simulation results. In recent years, various types of machine learning techniques, which include the MLP model, also known as artificial intelligence (AI) are widely used in many directions in accordance with improved performance of computers as well as rapid increases in relevant data size. Therefore, the topic addressed by the authors is opportune. As far as I know, such an attempt is very new in the glaciological research community. In this respect, this study is of high value. In addition, this informative paper is well written and organized. Therefore, this reviewer recommends its publication in the journal TC although several points in the paper can be improved prior to its publication. Please find below my specific comments and suggestions on the paper.

We thank the reviewer for taking valuable time to review this manuscript, and for providing very thoughtful and constructive comments for improving its quality. We carefully considered all of the reviewer's suggestions and modified the manuscript accordingly. For a better overview and clarity we numbered and colored the reviewer's comments, as follows:

(0) The comments from the reviewer are colored in black.
Our answer to each comment is colored below in blue.
The proposed changes to the original manuscript are then highlighted in red.

We hope that the reviewer agrees that we took her or his suggestions into serious consideration, and that we adapted the manuscript in a satisfying way.

Specific comments:

(1) Title: Suggest rephrasing "Antarctica" → "Antarctica Ice Sheet". I think it is better to state explicitly that "ice sheet" surface melt is investigated in this study.

We agree that it would be better to narrow down the tittle to 'Antarctica Ice Sheet'.

Improving Surface Melt Estimation over ~~Antarctica~~ ***Antarctic Ice Sheet*** Using Deep Learning: A Proof-of-Concept over the Larsen Ice Shelf

(2) L. 3: "simulations": simulations with a climate model? Please clarify.

Yes, the term 'simulations' in Line 3 is confusing, we have clarified it as suggested.

This study aims to demonstrate the potential of improving surface melt simulations ***with a regional climate model*** by deploying a deep learning model.

(3) L. 11: “XGBoost”, Its definition is needed here. In my opinion, most readers of the journal TC are not so familiar with machine learning methods.

Thanks for pointing out this issue. We agree that such a terminology can confuse readers who are not familiar with machine learning, and it might be too detailed for abstract. Thus, we decide to remove it from the abstract, but in the main text, we have added a short definition in addition to its reference in Line 180-181.

Moreover, the deep MLP model outperforms conventional machine learning models ~~(e.g., random forest regression, XGBoost)~~ and a shallow MLP model.

L. 180 ... we have also built a multivariate linear regression model, a boosting model (XGBoost ~~,~~ ***: a highly effective and widely-used tree boosting system;*** Chen and Guestrin, 2016), ...

(4) L. 14: “the heterogeneous drivers of melt”: Its intention is unclear to me. Can the authors describe it more concretely?

The term 'heterogeneous drivers' refers to air temperature, topography, katabatic wind, to name a few; they have been added to the text.

However, for one location (AWS 18) the deep MLP model does not show improved agreement with AWS observations, likely ~~due to the heterogeneous drivers of melt within the corresponding coarse resolution model pixels~~ ***because surface melt is driven to a large extent by other factors (e.g., air temperature, topography, katabatic wind) than albedo alone***.

(5) L. 45: “remote sensing albedo observations” It is worth to mention MODIS here.

We agree with the reviewer. It has been added in the text, as suggested.

..., and remote sensing albedo observations ***from the moderate resolution imaging spectroradiometer (MODIS)***.

(6) L. 60: “scatterometry”: Can the authors specify the type of the sensor used for the comparison?

Yes, we agree with the reviewer. We have included the full name of the scattermetry, i.e., QuikSCAT, in the text.

... and (3) a previous comparison between RACMO2 albedo and ~~scatterometry~~ ***radar backscatter from the Quick Scatterometer (QuikSCAT)*** revealed that ...

(7) L. 60 ∼ 61: “RACMO2 melt can be both higher and lower than observations”: depending on what conditions? More detailed explanations are useful for readers.

Yes we agree with the reviewer. A comprehensive comparison between the QuikSCAT scatterometry and RACMO2 is in the paper by Trusel et al. (2013). We merely want to highlight that RACMO2 is not not only biased in one direction. In this area, we see RACMO2 both overestimating and underestimating surface melt. Therefore, it is a good test for the application of our model that should be able to both enhance and reduce melt estimates from RACMO2.

... revealed that ***both positive and negative values of the difference between RACMO2 and observed surface melt exist in this area*** (Trusel et al., 2013).

(8) Figure 1 caption: "The elevation information ": "The elevation information for tundra areas "?

The digital elevation model shown in Fig. 1 does not only provides information about mountainous areas but also the lowlands, however, these ice surfaces are very homogeneous, resulting in mono-colored (cyan) areas. We have changed the figure caption as below:

~~The elevation information are~~ ***Surface elevation is*** derived from the ETOPO1 1 arc-minute global relief model ...

(9) L. 85: Section 2.2.2: I would like to confirm whether the authors considered possible factors that can disturb albedo measurements (e.g., tilt of the AWS mast and/or sensors, surface hoar, etc) and corrected the measured albedos.

The solar radiation observations are corrected for sensor tilt and riming. In Line 91, we changed the text to:

More details of the experimental setup ***and data corrections*** are found in Kuipers Munneke et al. (2018b), Smeets et al. (2018), and Jakobs et al. (2020).

(10) L. 85: Please explain whether the ice surface appears or not during the austral summer ablation period at each AWS.

To our knowledge, no bare ice areas appear at the surface at these AWS locations.

(11) L. 89 $\sim$ 90: "Air temperature (T2m), air pressure (p), and relative humidity (RH) ": It is better to indicate that they are measured at the surface.

Yes, we agree. We have clarified it in our text.

The same sensor also measures down- and upwelling longwave radiation (R_l). Air temperature (T_{2m}), air pressure (p), and 90 relative humidity (RH) ***measured at 1-4 m above the surface*** are corrected for heating of the shielded housing by solar radiation, especially in situations with low wind speed.

(12) L. 98: "evolution of surface albedo": The following description on "surface albedo" (The albedo scheme is based on the metamorphism of snow grains determining the amount of incoming radiation that is absorbed in the snowpack) is only on "snow albedo". Therefore, indicating explicitly like "evolution of snow albedo" is better.

The reviewer is correct. We have changed the text as suggested.

RACMO2 has a scheme for calculating the evolution of ~~surface~~ ***snow*** albedo, a key parameter for the surface energy balance in summer, and an important factor in determining surface melt.

(13) L. 111 ∼ 112: "The MODIS albedo itself is corrected for variations in solar zenith angle, and cloud cover (Figure 2 Block II-3) to allow comparison with the RACMO2 albedo.": Its intention is a bit unclear. In the previous part, the authors state that white-sky albedo from MODIS is used in this study, so, can I assume that the authors converted white-sky albedo to (blue-sky) albedo in this process? Please clarify.

The reviewer's assumption is correct, and we agree that it is unclear here. Previously, clarification was given at a later stage in Section 3.5. Now, it has been also briefly included here.

The MODIS ***white-sky*** albedo is ***converted to blue-sky albedo by*** ~~corrected~~ ***correcting*** ~~for~~ variations in solar zenith angle, and cloud cover (Figure 2 Block II-3) to allow comparison with the RACMO2 ***blue-sky*** albedo which is also a blue-sky product.

(14) Figure 2 caption: What is the difference between solid and dashed lines?

The dashed line is the line that intersects more than three other lines. To avoid misunderstanding, we have now changed the dashed line into a solid one, for consistency.

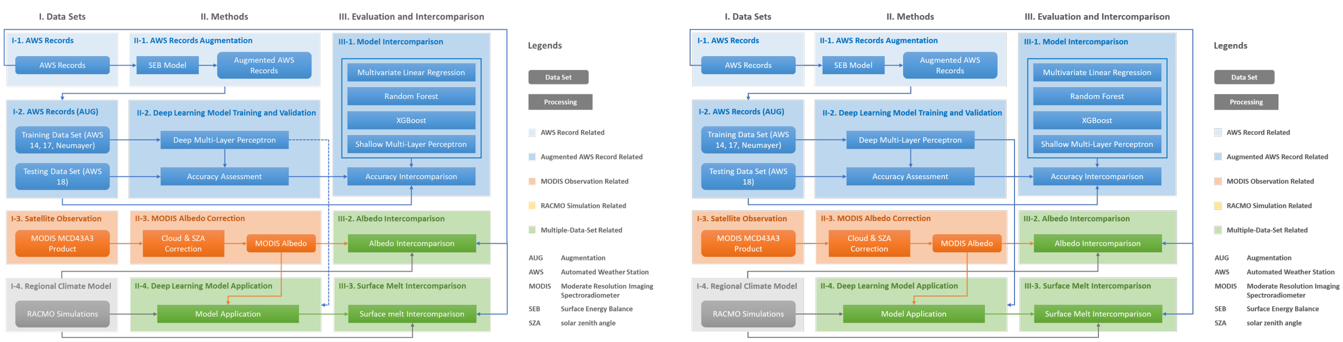

Figure 1: Old Figure 2 on the left, and new Figure 2 on the right.

(15) L. 141 ∼ 143: It is useful for readers if the authors briefly describe how the calibration was performed.

The least constrained settings in the surface energy model is the roughness length for turbulent exchange of energy. It is varied until the bias between observed and modelled temperatures is minimized.

The model ~~has been previously calibrated~~ ***settings (mainly for turbulent exchange of energy) were calibrated*** by minimizing the difference between observed and modelled surface temperature and subsurface temperatures.

(16) L. 161: As seen later in Fig. 7, discrepancies between in-situ measured and MODIS albedos are often found at all the AWSs. I would like to know the authors' thoughts regarding whether the deep MLP model performance can be improved if the accuracy of MODIS albedo (e.g., RMSE) is somehow considered here (e.g., input monthly RMSE values of MODIS albedo). If the authors can add some comments on this issue in Sect. 5, it will be useful for readers.

The reviewer is absolutely correct. That is also the reason why we have performed correction in two different modes (i.e., MLP: MODIS-observed albedo + RACMO2 and MLP: AWS-observed albedo + RACMO2) in Section 4.3. In Fig. 9b, Fig. 10b, and Fig. 11b, they show that once we replace the MODIS albedo by AWS albedo, the corrected surface melt is much closer to the observed ones. These findings are summarized in Section 4.3.

(17) L. 192: "the MODIS white-sky albedo product (MCD43A3) is a clear sky product": In my humble opinion, this description is incorrect: White-sky albedo is defined as albedo in the absence of a direct component when the diffuse component is isotropic.

We agree on the definition of the white-sky albedo. Here, we are referring to the fact that the MODIS MCD43A3 product is based on the Level 2 Surface Reflectance Product (MOD09, MYD09), which is a cloud-cleared (i.e. clear sky) and atmospherically corrected product. In this way the MODIS (white-sky) albedo product is a clear sky product that cannot be measured in cloudy conditions. We have clarified in the text.

~~However,~~ ***Therefore, we use*** the MODIS ~~white-sky~~ albedo product (MCD43A3)***, which*** is a clear sky product ***based on observed clear sky surface reflectances from the Level 2 Surface Reflectance Product (MOD09, MYD09).*** ~~and~~ ***The MODIS white-sky albedo hence*** needs to be converted to ~~total~~ ***blue***-sky albedo by correcting for the influence of changing solar zenith angle and cloud cover, both of which have a significant impact on surface albedo over snow (Kuipers Munneke et al., 2008).

(18) L. 258: "0.21 lower in no-melt conditions": compared to what? It seems to me that the authors compare melt and no-melt periods here. If so, the value must be 0.24.

Yes, the reviewer is correct. We were comparing the 'no-melt' (R^2=0.74) to the overall condition (R^2=0.95), but it would make more sense to compare no-melt to melt conditions. We have changed the corresponding text in the manuscript.

The R^2 is ~~0.21~~ ***0.24*** lower in no-melt conditions than ***melt conditions*** because of a number of notable outliers during no-melt periods (Figure 4c).

(19) L. 261: "'handle' of the 'sword'": I could understand the intention of "sword"; however, the intention of "handle" is unclear to me.

We admit that the term 'handle' can lead to confusion. Thus, we have replaced it by a more detailed explanation, i.e., points along/close to x and y axes, in the text.

The ~~'handle'~~ ***points along/close to x and y axes*** of the 'sword' originates mainly from no-melt periods (Figure 4c).

(20) L. 281: "neural network": This technical name appears for the first time in this paper here. I think it should be introduced much earlier in this paper.

Yes, we agree. We have add a definition of deep learning in the introduction section.

L. 38 ... Therefore, we propose a deep learning method that uses the albedo observations from remote sensing to correct for the shortcomings of climate models. ***Deep learning is a machine learning technique extracting multiple levels***

of abstraction/representation of data based on multiple processing layers, i.e., artificial neural networks (LeCun et al., 2015). To date, deep learning has been widely applied in Earth system science to analyze and correct mismatches between model simulations and observations (Reichstein et al., 2019) as they execute much faster than physics-based models.

(21) L. 307: "clear-sky": How did the authors confirm the "clear-sky conditions"? Also, do the authors mean this is in-situ measured information? Or model simulation results? Please clarify.

The clear-sky conditions are evaluated based on the daily cloud mask provided by MODIS MOD09GA product. We have clarified it in the text.

In L. 79 ... to obtain observational information about cloud coverage at AWS locations, cloud classifications are taken from the MOD09GA daily surface reflectance product (i.e. 'MODIS/Terra Surface Reflectance Daily L2G Global 1 km and 500 m SIN Grid', Vermote and Wolfe, 2015), also archived in GEE.

For clear-sky conditions ***indicated by MOD09GA*** (Figure 6), AWS 14 and AWS 17 show higher correlations with MODIS ($R^2 = 0.28$ and 0.20, respectively) than RACMO2, whereas for AWS 18 this is reversed with better correlations between AWS and RACMO2 ($R^2 = 0.40$).

(22) L. 333: clear-sky" → "white-sky"; "all-sky" → "blue-sky". Also see my comment on L. 192.

We agree to use "white-sky" and "blue-sky" here.

To translate ~~clear-sky~~ ***the white-sky*** MODIS albedo to ~~an all-sky~~ ***the blue-sky*** albedo, we used optical depth simulated by RACMO2 at its horizontal resolution of 27 km.

(23) L. 381 382: "It is plausibly due to the overestimations in 2m air temperature by RACMO2 simulations (Figure 11a).": If the deep MLP model is implemented in RACMO2, and a fully two-way coupled configuration between RACMO2 and the deep MLP model becomes possible, do the authors think that the overestimation can be solved? Please discuss.

It would be very unlikely that the MLP model would be two-way coupled to RACMO2. Theoretically, the best solution would be to nudge the simulation of surface albedo in RACMO2 to a time series of MODIS albedo observations. This would be the best two-way coupling between model and observations. However, in a technical sense, this is a very problematic setup. And, from another perspective, surface albedo is an important parameter that should be allowed to evolve due to the snow physics in the model, rather than due to an observational constraint.

(24) L. 452: "the state-of-the-art deep learning architectures and models": Please specify them more in detail.

We have explained "the state-of-the-art deep learning architectures and models", as below:

Refinement of the deep-learning-based framework includes developing a module to switch the deep MLP correction, and testing the state-of-the-art deep learning architectures and models***, e.g., applying recurrent neural network architectures such as long short-term memory (LSTM; Hochreiter and Schmidhube, 1997) and transformer (Vaswani***

et al., 2017) to help the deep learning model take in more temporal information, and/or using a convolutional neural network to learn a better representation of MODIS albedo information within a RACMO2 grid.

Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. Neural computation, 9(8), pp.1735-1780.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).

Technical corrections:

When refereeing a figure in running text, the abbreviation "Fig." should be used when it appears in running text and should be followed by a number unless it comes at the beginning of a sentence. See the following link for more in detail: https://www.the-cryosphere.net/submission.html#figurestables Please check throughout the manuscript carefully again.

Thanks very much for this detailed information about formatting. We will systematically change the 'Figure' to 'Fig.' when citing a figure in the running text.

L. 12: "the resulting,": The comma is unnecessary?

Yes, the comma can be deleted. Now, it has been removed.

When applying the trained deep MLP model over the entire Larsen Ice Shelf, the resulting~~,~~ corrected RACMO2 surface melt shows ...

L. 63: "The Antarctic Peninsula is the mildest region of Antarctica " → "The Antarctic Peninsula (Fig. 1) is the mildest region of Antarctica "

Changed as suggested.

The Antarctic Peninsula ***(Fig. 1)*** is the mildest region of Antarctica, ...

L. 81: "demonstrate": I think using "confirm" instead of "demonstrate" is better here.

Changed as suggested.

To ~~demonstrate~~ ***confirm*** the spatiotemporal melt pattern in the study area, ...

L. 82: "as an indicator surface melt" → "as an indicator of surface melt"

Sorry for the grammar error, we have added the missing 'of' as suggested.

..., we derived the backscattering coefficient drops as an indicator ***of*** surface melt ...

L. 86: "AWS 14, AWS 17, and AWS 18 are automatic weather stations " → "AWS 14, AWS 17, and AWS 18 (Fig. 1) are automatic weather stations "

Changed as suggested.

AWS 14, AWS 17, and AWS 18 ***(Fig. 1)*** are automatic weather stations installed and operated by ...

L. 87: "Shortwave radiation" Add "down- and upwelling" before this.

We agree it is necessary to explain the type of shortwave radiation. We think it might be better to term it as: incoming/reflected shortwave radiation.

British Antarctic survey (BAS). ~~Shortwave~~ ***Incoming/reflected shortwave*** radiation (~~R_s~~ $S_\downarrow$ ***and*** $S_\uparrow$), and ...

L. 107: "done" → "performed"; Using a formal word is better in a scientific paper.

Changed as suggested.

The deep learning model needs to be trained (~~Figure 2~~ ***Fig. 2*** Block II-2), which is ~~done~~ ***performed*** on a reference data set derived from AWS observations.

L. 138: "the ground heat flux" → "the subsurface conductive heat flux"; I assume the authors do not consider interaction between the ice sheet and the ground (bedrock).

We think it might be better to term it as 'the ground heat flux at the snow surface'.

..., and G is the ground heat flux ***at the snow surface***, ...

L. 252: "Figure 4" → "Fig. 4a"

Changed as suggested.

The cross-validation of the additional daily surface melt (Ma) predicted by the deep MLP model (~~Figure 4~~ ***Fig. 4a***) shows ...

L. 254: "0.95 and 0.42" Indicate these values with units in the text.

The missing units have been added.

The overall RMSE and MAE are 0.95 ***mm w.e. per day*** and 0.42 ***mm w.e. per day***, respectively.

Figure 5 caption: In my humble opinion, "Dynamics of" should be rephrased to something like "Temporal changes in"

Changed as suggested.

Figure 5. ~~Dynamics of~~ ***Temporal changes in*** the corrected regional atmospheric climate model version 2.3p2 (RACMO2) surface melt ...

Figure 7 caption: Same as the comment on the Figure 5 caption.

Changed as suggested.

Figure 7. ***Temporal changes in*** ~~A~~***a***lbedo ~~dynamics~~ from the regional atmospheric climate model version 2.3p2 (RACMO2) simulations, ...

Figures 10 and 11: Most of the captions can be omitted by referring to the caption of Fig. 9.

The captions of Fig. 10-11 has been modified as suggested.

Figure 9. Application of the deep multilayer perceptron (MLP) model to RACMO2 and MODIS data at automatic weather station (AWS) 14, and the results of: (a) discrepancies of input meteorological parameters among the AWS observations, regional atmospheric climate model version 2.3p2 (RACMO2) simulations, and moderate resolution imaging spectroradiomete***r*** (MODIS) observations; (b) daily surface melt time-series from the original RACMO2 simulations, AWS observations, and MLP estimations using different input data set for albedo (from either MODIS or AWS observations) and the other meteorological parameters from RACMO2 during the austral summer 2013/2014; (c) a scatter plot of monthly surface melt from the deep MLP model estimations and the original RACMO2 simulations compared to AWS observations; and (d) annual melt fluxes (July–June) at AWS 14 from observations, RACMO2, the deep MLP predictions, and QSCAT (QuikSCAT) estimations (Trusel et al., 2013).

Figure 10. As Figure 9, but for AWS 17.

Figure 11. As Figure 9, but for AWS 18