Response to Topical Editor

We greatly appreciate your valuable suggestions and comments. We studied your comments carefully and made substantial revisions accordingly. We hope that our corrections meet the requirements. We have also provided a clean version of the revised manuscript in which all revisions have been incorporated. The revised contents are marked in red in the manuscript. Our modifications and responses to the reviewers' comments are listed as follows. The reviewers' comments are in blue text, while the authors' responses are in black text. All revisions in the revised manuscript are in red text.

Comments to the author:
Dear Donghang Shao,

Thank you for a thorough revision of the manuscript and for addressing the reviewers' comments.
I have one main question that I will also ask separately to reviewer #1 (without starting a new review round):
I have quickly checked the GHCN and it seems that it contains direct SWE observations (such as from snow pillows at the SNOWTEL sites). So I do not understand why they are not suitable for the training of the model. You switched to the hemispheric scale snow course dataset but I do not understand why it is different, or better suited to your use, than the GHCN? Could you add maybe a sentence in the manuscript to clarify your choice of the HSSC?

Reply: This suggestion has been accepted, and the text has been clarified and modified.
GHCN applies pointwise observations on SWE (Menne et al., 2016), not distributed observations from snow courses. Therefore, GHCN SWE data cannot adequately reflect the spatial distribution characteristics of SWE or represent SWE at the spatial scale. The snow courses of hemispheric-scale snow course (HSSC) observational datasets are transects in which SWE is sampled manually at multiple locations with typical conditions to eliminate uncertainty in the regional-scale spatial variability of SWE due to the influence of snowpack characteristics and land cover type (Pulliainen et al., 2020).
Pulliainen et al.'s (Pulliainen et al., 2020) study of HSSC from Eurasia and North America showed that the typical exponential autocorrelation length of HSSC observations is 150~250 km. The HSSC dataset is representative of regional-scale SWE, and it is extremely suitable for comparison and validation with SWE satellite data or SWE reanalysis data.

References
Menne, M., Durre, I., Korzeniewski, B., McNeal, S., Thomas, K., Yin, X., Anthony, S., Ray, R., Vose, R., and Gleason, B.: Global Historical Climatology Network–Daily (GHCN-Daily), Version, 3, V5D21VHZ, 2016.
Pulliainen, J., Luojus, K., Derksen, C., Mudryk, L., Lemmetyinen, J., Salminen, M., Ikonen, J., Takala, M., Cohen, J., Smolander, T., and Norberg, J.: Patterns and trends of Northern Hemisphere snow mass from 1980 to 2018 (vol 41, pg 861, 2020), Nature, 582, E18-E18, 10.1038/s41586-020-2416-4, 2020.

We added a detailed description of the reason for the selection of the HSSC data in Section 2.4.1.
The snow courses of the HSSC dataset are transects in which SWE is sampled manually at multiple locations with typical conditions to eliminate uncertainty in the regional-scale spatial variability of SWE due to the influence of snowpack characteristics and land cover type (Pulliainen et al., 2020).

Additionally, your responses generally answer very well the reviewers' demands, but the quality of the English in the added material could be improved. I list some points below and I encourage a full proof-read of the manuscript before the next submission. The manuscript will be accepted once these minor comments are fixed.

Reply: Thank you for your suggestion; the text has been modified.

We have thoroughly checked and revised the English version of the article, including grammar problems and some sentences with unclear meanings, which makes the article more understandable. In addition, this manuscript has been revised by a native English speaker from American Journal Experts (https://www.aje.com/).

Sincerely,
Baptiste Vandecrux

(lines refer to the updated manuscript)

l. 23: "are 0.22 and 19.92 mm at different altitude intervals" This sentence is not clear. is it the average for different elevations? Is it the maximum MAE and RMSE among all elevation bins (if yes then which one?)?

Reply: The values of 0.22 and 19.92 mm are the average values of MAE and RMSE for the RRM SWE product at different altitude gradients.

"0.21 and 27.00 mm at different regions" Same thing here, are the two numbers MAE and RMSE or are they the min max values (of MAE or RMSE?) for different areas (then which regions?)?

Reply: The values of 0.21 and 27.00 mm are the average values of MAE and RMSE for RRM products in Russia, Canada, and Finland.

Consider using a structure like:
"The MAE (resp. RMSE) ranges from x1 mm (resp. y1 mm) for areas within E1-E2 m elevation to x2 (resp. y2) mm within the E3-E4 m elevation range. They are best in this region (spell out which one) and worst in that region (spell out)."

Reply: This suggestion has been accepted, and the text has been modified.

We have modified the original sentence to:

"The MAE ranges from 0.16 for areas within <100 m elevation to 0.29 within the 800-900 m elevation range. The MAE is best in the Russian region and worst in the Canadian region. The RMSE ranges from 4.71 mm for areas within <100 m elevation to 31.14 mm within the >1000 m elevation range. The RMSE is best in the Finland region and worst in the Canadian region."

l.175-176 "spatially uniformly distributed" The comment of the reviewer here is still not addressed: how do you select these points? how do you ensure that the selected samples are uniformly distributed? Please be thorough in your description so that someone that would like to reproduce your study would end up selecting (almost) the same samples for training.

Reply: This suggestion has been accepted, and the text has been modified.

We added a detailed description of the training sample selection method in Section 2.3.

"During the RRM model training process, we reconstructed the training data to try to extract training samples that are uniformly distributed in space as much as possible. First, a scan window of 250 km × 250 km (10 × 10 pixels) was created. Then, each gridded SWE data point participating in training is scanned, and the sample numbers in each scan window are counted. Finally, the mean value n of the sample numbers in all scan windows is taken as the number of training samples to be selected in each scan window. For the scan window with sample numbers higher than n, n samples are randomly selected from the scan window. For the scan window with sample numbers lower than n, all samples in the scan window are selected as training samples."

l.261-265 Please refer to Table 2 in these two statements.

Reply: This suggestion has been accepted, and the text has been modified.

"The RRM SWE product has a significant advantage over the multisource data average method, and its accuracy is much higher than that of the simple multisource data average method (Table 2). Based on the above verification results, the accuracy of the RRM SWE is significantly improved; the RRM SWE dataset has higher accuracy than that of any single grid SWE dataset, and it also fills the gap in the original SWE data in terms of spatial and temporal resolutions."

l.278: I don't understand what you mean by "relative to itself"

Why is the RRM not so good below 400m? Do you mean that another product is better there? Then which one? Can you point at a graph or table that highlights a better product below 400 m? You can also merge this discussion point with your description of Figure 5.

Reply: This suggestion has been accepted, and the text has been modified.

We are very sorry for the misunderstanding of this sentence due to our lack of clarity. In fact, the meaning we intended to convey is as follows. The estimation accuracy of the RRM SWE product for the high value range of SWE (SWE > 400 mm) is lower than that for the low value range of SWE (SWE < 400 mm) (Fig. 4). The main reason for this is that the training accuracy of the RRM model for the high-value range of SWE is affected by the small number of stations that observe the high-value range of SWE.

We have modified the original sentence to:

"The estimation accuracy of the RRM SWE product for the high-value range of SWE (SWE > 400 mm) is lower than that for the low-value range of SWE (SWE < 400 mm) (Fig. 4). The main reason for this is that the training accuracy of the RRM model for the high-value range of SWE is affected by the small number of stations that observe the high-value range of SWE."

l. 284 "training model" change to "the RRM presented here". "in detail" change to "as predictor" or remove.

Reply: This suggestion has been accepted, and the text has been modified.

"Unfortunately, our current RRM presented here does not consider these factors as predictors, which is a limitation of the current RRM SWE product."

l.285 "remain challenging" Can you refer to a plot or a section in the study that would illustrate the poorer performance of the SWE products in rugged terrain? Maybe add a reference that deals with SWE estimation in rugged terrain.

Reply: This suggestion has been accepted, and the text has been modified.

"Finally, in complex terrain with an elevation interval >1000 m, the RRM SWE product performed poorly, with an RMSE of 31.14 mm (Fig. 5), and the integration of SWE products remains challenging (Mortimer et al., 2020)."

References
Mortimer C, Mudryk L, Derksen C, Luojus K, Brown R, Kelly R, Tedesco M. 2020. Evaluation of long-term Northern Hemisphere snow water equivalent products. Cryosphere,14(5), 1579-1594. doi: 10.5194/tc-14-1579-2020

l.291-296: This >5 line long sentence is redundant with Figure 5. Consider shortening by only presenting the metrics or the elevation bins that are discussed in the text. The rest can be found in the Figure.

Reply: This suggestion has been accepted, and the text has been modified.

We have reduced the original sentence to:

"The above verification results show that the MAE, RMSE, R and R^2 between the RRM SWE product and measured SWE perform well at altitude gradients of <100 m, 100-200 m, 200-300 m, 300-400 m, 400-500 m, 500-600 m, 600-700 m, 700-800 m, 800-900 m, 900-1000 m and >1000 m (Fig. 5)."