Snow depth retrieval over Pan-Arctic sea ice (2012-2021) using multi-source data and machine learning models

Mengmeng Li[1] , Jianwei Ma[2,*], Yang Li[3], Juha Karvonen[4], Bin Cheng[4], Yingfei Wang[1], Haili Li[5], Yafei Nie[6] and Zheng Duan[7]

This manuscript uses AMSR2 passive-microwave brightness temperatures, satellite-altimetry-derived snow depth, SnowModel-LG snow density, and OIB and MOSAiC observations to retrieve pan-Arctic sea-ice snow depth. It compares four machine-learning approaches (multiple linear regression (MLR), random forest (RF), LightGBM, and long short-term memory (LSTM)) and further evaluates the effect of the retrieved snow depth on sea-ice-thickness (SIT) estimation. The topic is within the scope of The Cryosphere. The use of multiple data sources, the comparison of different model architectures, and the feature-importance analysis are potentially valuable. I recommend accepting this article after moderate revisions.

Major comments：

1. There appear to be inconsistencies between the table headings, the numerical values, and the descriptions in the text. For example, Table 3 contains an additional “ALL” column, whereas the numerical rows appear to contain seven values corresponding to LGBM, MLR, RF, LSTM, SM, Rost, and MW99.

2. The manuscript interprets the difference between the internal testing results and the independent OIB/MOSAiC results as a “classic indicator of overfitting.” This is possible, but the difference may also result from several other factors, including differences between the ASD and OIB/MOSAiC error structures, spatial and temporal sampling, scale mismatch, and differences in snow-depth distributions. The current experiments do not fully isolate overfitting from dataset shift or reference-product

uncertainty.

3. The Results repeatedly interpret high importance of PR(89), PR(06), and other features as direct evidence of specific snow microstructural processes, such as depth-hoar development and peak metamorphism. Permutation importance indicates the predictive contribution of a feature within a particular model and dataset. It does not, by itself, establish a unique physical mechanism, especially in the presence of highly correlated GR and PR variables and possible atmospheric effects at 89 GHz.

4. Please replace causal or overly definitive statements such as “directly tracking,” “demonstrates,” “signifies,” and “providing direct evidence” with more cautious language. The authors should also mention the effect of feature collinearity on permutation-importance rankings.

5. The correlation coefficients against MOSAiC are very low for all products, and the LSTM correlation is reported as 0.13 with a p-value of 0.08. The manuscript correctly notes the point-to-grid scale mismatch, but it should not state that LSTM “most accurately captures the mean snow depth state” solely on the basis of a low bias and MAE, particularly when the differences in MAE among the products are small.

6. The SIT results show a much smaller bias for the machine-learning-based estimates than for FT4 using MW99. This is an important result, but the conclusion that this “confirms” MW99 as the major source of operational SIT error may be too strong. Differences may also result from sea-ice density, snow density, freeboard processing, ice-type classification, matchup selection, and correlations between the snow depth and SIT reference products.

Minor comments:

1. The title should consistently use “pan-Arctic.” such as CryoSat-2, AMSR2, and sea

ice thickness should be used consistently throughout the manuscript.

2. The manuscript requires comprehensive language and copy editing. Several language and formatting errors are present, including:“root-means-square-error,” which should be “root-mean-square error”; “horizonal,” which should be “horizontal”; “addictive linear interactions,” which should be “additive linear interactions”; “oceanatmosphere,” which should be “ocean–atmosphere”; “year-around,” which should be “year-round”; and “Cryosat2,” which should be consistently written as “CryoSat-2.”

3. Suggested revisions to key statements in the abstract: Original statement: The 89 GHz polarization ratio has a significantly greater influence on snow-depth retrieval over multi-year ice. Suggested revision: The 89 GHz polarization ratio showed greater predictive importance over multi-year ice than over first-year ice.Original statement: Our long-term snow products offer users a reliable, high-accuracy dataset. Suggested revision: The resulting long-term snow-depth products provide an observationally evaluated dataset for Arctic sea-ice and snow studies, with uncertainties that vary across ice types, seasons, and the extrapolation period.

4. Use consistent model naming and capitalization. Please use one form consistently: LightGBM rather than alternating between LGBM and LightGBM, unless LGBM is explicitly defined as the abbreviation. Similarly, use consistent forms for MLR, RF, LSTM, OIB, MOSAiC, FYI, MYI, and SIT.

5. The manuscript contains several typographical and language problems, including:“generalisability” and “generalization” should be standardized according to the journal’s preferred English style;

6. In Figure 5 caption, “Artic” should be “Arctic”.

7. Line 676, “retrievals..” contains an extra period;

8. Line 490, “The Rost product” should use lower-case “the” when it appears in the middle of a sentence.

9. Line 514, “The value in parentheses represent” should be “The value in parentheses represents”.

10. The captions of Figures 2-7 should provide sufficient methodological information. The figure captions should indicate whether the plotted values are daily or monthly, whether the distributions are based on all collocated samples, and whether the comparisons use the same spatial and temporal mask for all products. Figure 4 should also clarify the meaning of the “point of flawless agreement.”