Review of a manuscript for *The Cryosphere*

Retrieval and parametrisation of sea-ice bulk density from airborne multi-sensor measurements

by A. Jutila et al.

Overall:

In this manuscript the authors present the approach to bulk sea ice density retrievals from parallel airborne measurements of total thickness, snow thickness and surface freeboard. For their study the authors use the data retrieved during airborne IceBird campaigns over the Beaufort/Chukchi Sea and Canadian Arctic in the spring seasons of 2017 and 2019.

The authors provide new, generally higher than was used before, estimates for the bulk densities for different types of sea ice. They further propose a new nonlinear parameterization linking observable ice freeboard with sea ice density to be potentially used in satellite based retrievals of sea ice thickness/volume.

The paper is clearly written and results, including figures, are well presented. I therefore consider the manuscript deserves to be published after some moderate modifications according to the comments provided below.

Major comment:

My only major comment concerns a new freeboard to density model proposed by the authors. The model is based on exponential fit to the data collected by the authors and offers at the moment RMSE values for the fit itself (model calibration error). However, since the model has a potentially high applicability in the algorithms for ice thickness/ice volume retrievals from satellite-based sensors, it makes sense to have its predictive skills to be tested properly.

Generally, a good agreement with data can be achieved via applying a data model complex enough and hence overfitting; it will not guarantee nevertheless any decent predictive skills for such model.

Since the authors have aggregated a significant volume of measurements for this study, a bootstrapping aproach (or block bootstrapping in case if autocorrelation in the series is substantial) can be used to test the model predicted vs measured values. This routine will provide a more realistic value for the RMSE to be used in future potential uncertainty estimates – RMSE for prediction.

Other (minor) comments:

Sec 2.5: "…a sporadically observed by the ALS at fractures (leads) of the seaice cover and we manually selected the corresponding elevations". Is the ALS used onboard receives returns from open water areas too, or the authors refer to refrozen leads only? Would it be possible to use the measured surface temperatures to support the detection of leads? Or this is actually already a part of the procedure for these z-control points identification?

Line 167: Please consider adding a most recent reference to Rosel et al., https://doi.org/10.5194/tc-15-2819-2021; where this effect is also considered.

Line 180: Please clarify the formulation/ application of the level ice criterion. I find it to be not too informative; it is nearly a copypaste from Rabenstein et al which suffice from the same issue.

Lines 185-188: Discussion on age assignment to sea ice along the flight track is somewhat unclear: do the authors refer to an average thickness estimated for level ice only, or for all (level+deformed) ice along a specified transect/transect segment?? If this is the latter, how long the transect segment length used for the age assignment?

Line 238: It can be useful to mention directly (though this is also apparent from eq. 4) that uncertainty in \sigma_rho_i includes spatially variable uncertainty in measured \sigma_rho_s, and hence both uncertainties vary along the track.

Table 3:

Table 3 shows numerous numbers with redundant precision in FYI density/density uncertainty estimates. Decimals can be eliminated throughout the table (and the text too in many places) by rounding to the nearest integer to leave significant figures only. E.G. 929.7\pm17.9 -> 930\pm18.

Line 263: typo? "…combined they results…"

Line 289: "… ice due to air incorporated in the pore spaces and to an increasing degree in MYI." Please consider rewriting the sentence. The meaning is clear it only appears awkward.

Line 291: "Despite the indirect measurement method, we are able to detect a difference in FYI bulk density between 2017 and 2019 that can be linked to the high sea-ice deformation in 2017". Please consider referring to Figure 8 here, provided that my comment to figure 8 below is justified."

Figure 8: From the figure it appears that there is a tendency towards higher uncertainties for lower values of density. This is especially clear for FYI where the data may form two groups clusters one below and one smaller group above the fit line. I wonder if these two groups of data points originate from different campaigns? Or could this be only the artefact of the data visualization? This is not to be ruled out (at least for me) as this figure is quite busy. I see no similar tendency for the MYI densities.