Reviewer 2:

This paper describes the use of a professional UAV for atmospheric sampling, in this example for water vapour isotopes above central Greenland. Such a platform is a welcome addition to ground-based measurements, especially under hars conditions such as in this paper.

The data presented are still only a few, not enough for new scientific insights, but enough for the proof of principle, and as such this paper fits the AMT journal.

I have several (smaller) comments and remarks, and I invite the authors to use them for slight changes in the paper.

2.4 I'd like a bit more details on the used UAV: power, total stored energy, the actual payload in kgs, and the way of launching (only later in the paper it becomes apparent it is launched by some 12G launching system). Of course, the list is in the appendix, but some key numbers in the main text would be welcome.

We have added important descriptors for the platform to the text in Section 2.4, line 161.

Revised text: "The S2 is capable of conducting fully autonomous flights in unimproved areas such as an ice sheet in part due to its pneumatic launch system. The aircraft can adjust to changing wind conditions in real-time, ensuring a high degree of stability for predefined mapping or atmospheric sampling applications (Elston et al. 2015b). The aircraft can carry up to a 3.5 kg payload for up to 90 mins. At arctic temperatures with the payload used in this study, we found 45 mins of flight time typical and apt for climbing 1600m and including needed sampling time."

What is the mass of the flasks used?

We have added the mass of the flasks (181 grams for each flask) and volume is ~ 500 cc on line 188 of manuscript.

Revised text: "The glass flasks are approximately 181 grams each, 500cc in volume, and include a supported dip tube to ensure the sample is adequately flushed during fill."

Table 1: the numbers within parentheses seem too large, especially for SPGSW, which is so close to SLAP2 that virtually only that uncertainty would add up to the primary uncertainty: (0.2^2+0.3^)≈0.36‰

We do not calibrate our secondary references to the closest primary standard. Instead, we calibrate to an interpolated line of the correction amount determined from all the primary standards. As such, we include all primary standards in the error propagation. For SPGSW that would be (1.2^2+0.3^2+0.3^2+0.2^2)^(1/2) = ~1.288

We have included a reference for this procedure to the caption of Table 1.

Revised text: "Additional details describing the calibration scheme can be found in *Jones et al.*, 2017."

line 290 the ‰ sign (or the word per mil) is lacking twice

Fixed.

Figure 5 and text. Sure d-excess is a powerful comparison; nevertheless, the individual isotopes D and 18O themselves would be equally interesting. As d-excess from the pods is somewhat higher than that of the tower, it would be interesting to see if this is caused by D or 18O, or an interplay of both.

In an earlier draft of the paper, we did show both but the graphs were so identical, with both D and 18O and their relative difference from the tower setup, that we decided just to report dxs. We believe it is an interplay

of both. A line to mention this has been added to line 310. As well, too additional figures have been added to Appendix A separating the two different isotopes.

Revised text: “The positive relation is seen in both δD and $\delta^{18}O$ implying that the positive bias is due to an interplay of both measurements. Figures of separated δD and $\delta^{18}O$ can be found in Appendix A.”

Line 308: The "Euclidean distance in the measurement domain" is also explained in appandix D, please mention that in the text.

We have added a second reference to appendix D.

And what is the difference between that and 'just' taking for example the average of the height of the max gradient in water vapor content and similar in other parameters?

In a way, we are already doing that though measuring water vapor content is something we can measure for flasks on the ground and not airborne. Euclidean distance between observation points is a gradient, just the normalized aggregate of all observable parameters. The reason we use the clustering index is because gradients alone returns multiple candidates for PBL location where the index considers the difference in ‘likeness’ between two regions and is less ambiguous. Discussion for this difference is in Appendix D.

Figure 6 also indicate the PBL that you actually took, based on your in-flight method at the time.

Added.

lines 345-350. While I of course see the advantage of taking duplicate samples, the alternative would have been 6 (or 8) altitudes, with also advantages! Why has the duplicate choice been made? (or one duplicate and 4 single ones, or any other combination). Would you choose differently for the next campaigns? May be something to discuss in the conclusions and outlook chapter?

This was certainly a difficult decision! We decided on pairs because of the difficulty of creating a benchtop equivalent test of both the low temperatures and mid-flight vibration exerted on the flasks in the field. We believe that, after the apparent reproducibility of pairs from this field campaign, pairs won’t be needed. As well, by alternating sample pods used on the drone vs taking the 2 meter sample, we can determine if any of the pods are experiencing leaks leading to a previous flight with such a pod being flagged for bad data. We have added a line to mention this.

Revised text at line 364: “Paired sampling was motivated primarily by the inability to test the low temperatures, the 12G forces exerted on the flasks during launch, and inflight vibration forces in a “benchtop” setting.”

Line 373 what is AGL ? Caption figure 7 gives meters ABL ? The plots all simply state "height"

The typo of ABL has been changed to AGL in Figure 7. AGL is “above ground level” and the acronym is defined at line 71. We also define it now in the Figure 7 caption.

Fig 8 only shows the d2H (18O is in the appendix). Apparently the authors do not think the isotope measurements are worth a 2H-18O relation plot (or d-excess for that matter)? Perhaps the July 12 points would be worth a plot or table for the 2H-18O relation?

Or alternatively still a third horizontal axis in fig 8 and show the 18O's as well (different marks color, and slightly displaced in height).

We choose not to originally include a 2H-18O relation plot as we intend for this paper to focus on the process of producing the isotope values for water vapor in air. That said, we have included it as an appendix figure and referenced it in the caption of Figure 8 where the 18O data is referenced as well. We have also now included dxs as a column of the data product.

The conclusions are a bit long compared to the rest of the text, so it can be shortened, and renamed "conclusions and outlook", as the last paragraph is about the future perspective (and need not be shortened).

We have significantly reduced the length of the conclusion without affecting the future perspective portion of it.

Revised text: "We have presented a UAV-isotope sampling platform and methodology capable of measuring atmospheric water vapor and its stable isotopes within the planetary boundary layer (PBL) and lower troposphere in a polar environment. We utilize a fixed-wing UAV (Black Swift Technologies) with flight times in excess of 45 minutes with the capability to reach 1,600m AGL. Multiple nose cones allow for collection of air in 8 glass flasks, enclosed within a 3D printed support structure that critically withstands 12Gs of force during takeoff. In this study, the total system is used to sample above and below an algorithmically-detected PBL, resulting in the first measurements of atmospheric water isotopes above and below the PBL on the high-altitude Greenland Ice Sheet.

Across four sample-taking missions at the EGRIP ice core site in 2019, we observed significant variation in water isotopes on either side of the PBL; the variability exceeded our conservative precision estimates of 2.8‰ in δD and 0.45‰ in δ18O. These results form the basis for future campaigns to collect high-temporal density measurements (flights every 4-6 hours across many weeks) at key missing scales that will improve ice-to-atmosphere modeling and mixing processes, ice sheet mass balance, satellite detection algorithms, moisture tracking, ice core science, and modeling the hydrologic cycle in general.

A field campaign for return to EastGRIP is scheduled for summer 2022. Future improvements to the UAV-isotope system will be primarily focused on logistical improvements that increase the number and frequency of flights. Additional flight crew will be available for nighttime flight missions. To ensure a balanced diurnal flight schedule over weeks of time, with the goal of one flight every 4-6 hours, a precessing schedule of calibration times will be used. Each calibration will be done every 2-4 days, lasting 12 hours, starting at different times of day. This ensures that we do not consistently lose the ability for UAV sampling at the same time for every calibration, e.g. from 12pm-12am. The combination of these improvements will allow the potential maximum number of flights per day to increase from two to as many as six, while balancing the timing of calibration. In flight, we will carefully regulate the rate of ascent and include better performing temperature and humidity sensors with minimal time constants, all of which will reduce hysteresis for PBL detection. We plan to leverage an existing anemometer used by the autopilot in order to assist in the correction as well as produce an additional 2D wind speed for the flight. Additional improvements will include a lighter pump and manifold system that should allow greater flight time. Beyond Greenland, this platform is readily adaptable to other scientific disciplines, and will be used in an upcoming permafrost project to measure atmospheric methane emissions and soil moisture content in Alaska."

Additional comment:

In the time since submission, we have determined that the first two flights of our six total do not have sufficiently useful calibrations. This was discovered from referencing lab notebooks from which it was found that calibration protocols were not correctly followed at the beginning of the field campaign. The midair isotope values for both the June 3rd and June 6th flights are precise and the midair isotope gradients real, but the values are not accurate to the standard necessary for reporting in this paper. To remain conservative in this pilot study, we are choosing to omit the data from the text and data product. Please note, this does not change any conclusions in the paper and only necessitates minor explanatory changes in the discussion, which has already focused on the June 12th flight onwards.