Thank you for your helpful comments and suggestions. Please note that paper has been updated using data obtained through 16 May, 2018 to give a full 2-year time series. This has not caused any major changes in our conclusions.

Reponses are provided in bold font to each comment in italics.

This manuscript presents the results of a nearly 2-year long test at Summit, Greenland of a commercial device (Snowfox) developed to measure snow accumulation in water equivalents (SWE). The Snowfox sensor measures neutrons produced in the earth's atmosphere by cosmic rays, these neutrons are attenuated by water accumulating above the sensor in the form of snow. While this technique has been applied fairly widely in alpine snowpacks, the authors assert that the test described in the manuscript is the first application in the firn on a polar ice sheet.

We note that "fairly widely" is debatable, as the only existing network is operated by the French utility entity in a relatively small portion of the Alps, with little documentation (only the proceeding and white paper cited in the manuscript, neither of which may have been peer reviewed) and no scientific application we're aware of.

The results are very encouraging (perhaps not surprising), but oversold in my opinion. Specifically, I find the claim that this test confirms better than 0.5% accuracy for SWE < 20 cm and better than 0.7% accuracy for SWE up to 140 cm implausible. I note that previous studies with similar devices (cited in this manuscript) conclude that accuracy in the 3 to 5% range could be achieved, and that data with this quality were useful.

We now clarify that these values are estimates of the sensor's precision, rather than accuracy, as based on the daily standard deviation in hourly measurements as function of depth given in Fig. 4. The accuracy can only be determined through validation, although, as explained in the text, little or no bias is found, indicating that the precision is indeed representative of the error.

Note that previous studies were in alpine settings measuring seasonal snow, where soil moisture and higher spatial variability made their measurements, and their validation, much more uncertain. This is explained in the introduction.

My skepticism is based on a combination of 4 factors: 1) the corrections to account for temporal variations in neutron flux at the snow suface and variations in the column of water vapor in the atmosphere above the buried sensor do not seem likely to be accurate to better than 1% (perhaps considerably less precise), 2) the noise in the estimated accumulation (Figures 4 and 5) appears to be at least +/- 2%, 3) comparison to 2 independent manual techniques to measure accumulation do not show consistent agreement at < 1% across the 10 month comparison period, and 4) over the full month long trial the maximum SWE reached just 42 cm, suggesting *strongly that the claimed accuracy at 140 cm is entirely theoretical, and hinting that the same might be true for SWE <20cm.*

Addressing each numbered point:
1) Reasons for neglecting atmospheric pressure are provided where this point is repeated in more detail below.

2) The precision curve is derived from Figure 4 (it's the solid curve). Figure 5. plots the daily difference in accumulation, with 4B giving the percentage change. The mean percentage change is 1.2%, which is close to the 1% error one gets when differencing of two measurements with 0.7% errors.

3) The mean biases between accumulation measured by the cosmic rays sensor and the validation datasets for the ~8 day periods between manual surveys are < 1% of the water thickness. The standard deviations of the differences between cosmic ray and validation dataset estimates are close to or within the expected precision of the cosmic ray sensor and the validation data. This is now explained more clearly in the text.

4) We now make it clear in the abstract and text that these precision estimates are based on the precision curve fit to the daily standard deviations in hourly measurements in up to 56 cm of water shown as the black curve in Figure 4. The loss in precision with thickness is due to the predictable decline in count rate and signal to noise ratio.

Regarding the corrections to raw data, more detail is needed to explain how the neutron monitor at Thule is used to provide No (neutron flux at the surface). It is likely that the Snowfox and the Thule monitors are not measuring neutrons with identical efficiency across the energy range, and I would be surprised if hourly changes in neutron flux were perfectly in sync, given 6 degrees in latitude and nearly 3 km in altitude separation between the 2 sensors.

It is true that our method of correction cannot account for anisotropy of the primary cosmic ray flux. This could lead to some short term error, for example, in accounting for individual Forbush decrease events. However, our method does ensure that on average (i.e. on timescales of multiple days) our solar correction should be unbiased. We employ a factor to account for latitude variation (insignificant here, since both the Thule station and our site are well above the "knee" in neutron intensity, see Hawdon et al. 2014, citation added to paper) and altitude difference (which is significant). Essentially, we estimate that the our site, because of it's higher altitude, is on average 1.19 times as sensitive to solar variations than Thule station. This is based on a regression to monthly neutron monitor data from the global neutron monitor data set, and again, will not work well for individual events. Fortunately, those interested in accumulation and mass balance will be interested in weekly to monthly changes, and short lived solar events will be averaged out.

To describe this is more detail, we have added the following to section 3:

"The unitless scaling parameter *b* accounts for differences in the magnitudes of solar-induced variations between the reference and local sensors due to differences in latitude and elevation, only the latter of which will be significant in this case (Hawdon et al. 2014). We use a value of *b* =1.19 based on regressions to the global neutron monitor dataset. While this single value produces unbiased corrections on weekly or longer timescales, we expect some errors associated with short term variation, such as individual Forbush decrease events."

(Are the Snowfox devices inexpensive enough to use one mounted above the surface to provide No (near a single buried Snowfox, or in the middle of a regional array of buried sensors in a future study)).

A local reference station would certainly reduce uncertainty in corrections, especially for measuring short-term variability as described above. However, it's not obvious to what extent uncertainty would be reduced over longer-time scales, and if this reduction would outweigh the risk of relying on a reference sensor on the ice sheet as opposed to a global-network monitor like THUL.
We have added the following sentence to section 5:
"Finally, additional accuracy on short time scales (< 1 week) may be obtained by deploying a local reference sensor above the surface, removing uncertainties in the incoming cosmic ray corrections due to differences in elevation and latitude."

Atmospheric water vapor is not a simple linear function of atmospheric pressure, varying depending on synoptic conditions in addition to pressure. Assuming solely pressure dependence has to introduce more than 0.5 or 0.7% uncertainty in the derived SWE. Note that there are several data sets on water vapor above Summit that might allow more precise treatment of its impact on neutron flux, or at least provide an estimate of the magnitude of uncertainty introduced by neglecting changes in water vapor that are not just a function of pressure.

The only mechanisms by which atmospheric water vapor would affect our measurements would be (a) attenuating high energy "source" neutrons or (b) moderating fast neutrons produced locally. In regard to (a), we note that neutron monitor stations--which require a much higher precision than our instrument-- do not necessarily correct for water vapor. This is probably because water vapor accounts for only ~1% of the mass of the atmosphere (and less over greenland) and has attenuation properties similar to dry air (attenuation length = 95 g cm^-2 for water, as opposed to 130 g cm^-2 for air).

With regard to (b), the work of Rosolem et al. (2013) do imply some atmospheric vapor effect for a subaerially exposed detector. This effect occurs due to the moderation of fast neutrons as they scatter through the atmosphere. However, we emphasize that our

detector is buried in the snow, and that most of the recorded neutrons are produced in the snow. For this we add the statement at the beginning of section 3:

"We do not apply a correction for atmospheric water variability because nearly all fast neutrons impacting the sensor are produced in the snowpack, rather than the atmosphere."

Rosolem, R., W.J. Shuttleworth, M. Zreda, T.E. Franz, X. Zeng, and S.A. Kurc, 2013: The Effect of Atmospheric Water Vapor on Neutron Count in the Cosmic-Ray Soil Moisture Observing System. J. Hydrometeor., 14, 1659–1671, https://doi.org/10.1175/JHM-D-12-0120.1

I also find that the manuscript is a little sloppy, particularly in describing the corrections applied to convert measured neutron counts to SWE above the sensor. For example, in the discussion of equation 1 used to calculate the relative count rate Nr is defined as the reference sensor count rate while Ns is the reference count rate??? If Nr=Ns then the relative count rate from this equation is always 1. Next page in discussion of equation 3 No is defined as the reference count rate (3rd ref ct rate) at the surface obtained before burial of the sensor (at time = 0). The term N/No in Equation 3 suggests No should be the count rate at SWE = 0 (i.e., flux reaching the snow surface), both N and No should be measured (estimated) at every time (it does not make sense to ratio N at each time to No measured just once, given time variations in both cosmic ray flux and water vapor/pressure).

We agree that Section 3 "Count Rate Correction and Conversion" was poorly presented and we have re-assigned several variables and rewritten the equations to be consistent with review given in Andreasen et al., 2017 (cited in revised paper). The specific points of confusion above have been addressed. We note that only a single "snow free" reference count is needed because the reference is corrected for solar and pressure variations. This is made more clear in the revised text.

Below are listed a variety of additional editorial comments (some are additional examples of sloppiness, a few more substantive), referenced by page/line #.

1/5 "background cosmic ray intensity" is probably not the correct term. What is really needed is variation in the neutron flux reaching the surface above the sensor at Summit, which could vary widely due to solar events (likely to dwarf changes in the flux of "cosmic ray background" impacting the solar system)

Replaced with "incoming cosmic radiation"

1/21 I would be very hesitant to claim that accumulation at Summit is "consistently low in June/July" based on less than 2 year record (not even considering prior results that find different results)

Replaced with “with the lowest accumulation in July and highest accumulation in the autumn of both years.”

1/28 measuring the volume of accumulation (delete “of” before volume)

Corrected

2/14 mˆ2 (superscript)

Corrected

2/29 the statement here that “neutron counts increase with altitude and latitude” (more specifically geomagnetic latitude) demands that more be said later regarding how well a monitor at Thule can constrain neutron flux at Summit

Addressed above

3/4 “calibration data sets” suggest that there will be calibration data presented later. Turns out that all of the (critical) parameters in Table 1 appear to be taken from specs provided by Snowfox vendor.

Replaced with “validation”.

3/12 and 13 Juxtaposing statement attributed to Alley, 1993 that Summit snow has “average surface density of 0.35 gm cmˆ-3” and citation of Dibb and Fahnestock, 2004 is sloppy. Latter paper presents density profiles from 22 “monthly” snowpits sampled at Summit over 2 years and shows that the mean density in the top 99 cm never exceeded 0.336 g cmˆ3 and averaged 0.305 g cmˆ3. This is also relevant to the Snowfox “validation” presented later. (Also note that the “-3” in manuscript should be superscript.)

We now use a mean top 10-cm density of 28 g cm^3 referenced to the recent SUMup datasets described in Montgomery et al. (2018). We also use the SUMup densities for the stake conversion described below. Superscript typo corrected.

3/18 MSF is an acronym for the “Mobile Science Facility”. Until summer 2017 the main science facility at Summit was TAWO.

Corrected

3-4/25-30 and 1-5 (equations 1-3) see comments above. Also, N in Eq 3 is never defined (think this is the actual measured neutron count, at a given time T, from the buried Snowfox)

We now clearly define N as raw counts at the sensor.

4/12 Fig 2 does not show any time series, rather a curve based on assumed performance of the Snowfox sensor.

Corrected

Also confusing to introduce N/No here, since Eq 3 defines N* to be a function of N/No.*

Corrected and clarified

4/16-17 such that the resolution is (delete "that" before resolution)

Corrected

4/30 "42 observations" (snow cores)? ? Earlier in this paragraph it is stated that cores were sample every 10 days. 42 x 10 is 420. 13 Mar 17 to 17 Jan 18 is 310 days. Plot in Fig 7 seems to show 36 cores. These are not all consistent (sloppy)

Corrected.

5/1 Differences in the values of hw derived from any single core by weighing and by measuring the volume of melted snow are not due to "unconstrained errors in the sampling procedure." These are 2 different measurements of the same sample, so the errors have to be in the measurements.

Replaced with "errors in the measurement procedure".

5/3 Given estimate of the mean density in the snowpack from surface down to depth of the Snowfox from 42 (or 36 or 31, whatever may be the actual number) cores over 10 months, what can you say about 1) whether constant value of 0.35 g/cmˆ3 is reasonable, 2) is any variation in the measured density seasonal, 3) does it look like what Dibb and Fahnestock saw, 4)why not use these measured values to convert the stake measurements rather than a constant, loosely defined "surface" value from the literature?

We have revised our approach, as described in section 4.
The reason for the high density value used was that this was supposed to also account for compaction. To make this more clear, we have reformulated the conversion to separately account for the density (using the mean monthly SUMup values) and the compaction rate (as a tuning parameter qualitatively compared to Dibbs and Fahnestocks observations and Zwally and Li 2002's model). The result is very similar to that obtained from the constant density.

5/6 No good justification to use constant value for density, given that you have measured it at fairly frequent intervals, and that Dibb and Fahnestock showed that it is not constant (and was always lower than the assumed constant value used here).

See response to previous point. Also, the surface core densities may not be applicable to the snow stake conversion because 1) the depth of the sample varies with time 2) the base board may impact compaction and 3) this is a point measurement that may not apply over a large area. This is now explained in section 4.

5/19-20 There is an overall decline

Corrected

5/26 0.4 g cmˆ3 is a pretty high value for the density of a wind slab at Summit, also note that it is sloppy to change the units to kg/mˆ3 here

We clarify that these values here are approximate. Units are changed to g cm^{-3}

6/1 should "0.013 cm + 0.007" just be 0.02, or 0.013 +/- 0.007?

Replaced with "The best fit line predicts a standard deviation of 0.005 cm at h_w=0, increasing by 0.007 cm per cm of h_w…"

6/1-9 this paragraph does not support the very high accuracy claims made for Snowfox in the astract.

The abstract quotes exactly these values, which are now clarified to indicate precision, not accuracy.

6/19 "that much of the" (delete extra "the")

Corrected

6/21 seems that the overall rate should refer to 16 May '16 to 18 Jan '18 (not Jan '17)

Corrected to read "2018"

6/28 given that the Snowfox estimated SON accumulation differed by more than a factor of 2 between 2016 and 2017, how confident can one be that JJ are consistently low accumulation months based on the same almost 2 years of record.

Since we do not make a statement about "consistency" here, we assume the reference is to the statement in the abstract, which we have edited as described above.

6/30 “change in water equivalent” (add in)

Corrected

7/1-8 Make it clear what is signified by the “mean difference” (i.e., is Snowfox biased high or low by mean of 0.77 cm vs cores and 0.22 cm vs stakes).

Clarified.

Also consider redoing the stake comparison using a better estimate of density, with seasonal variation (from measured density of the cores in this study and/or values from Dibb and Fahnestock.) (Note that the agreement with Snowfox is likely to be worse using more realistic, lower, density for snow in the top 42 cm of SWE.

Redone as described above and in the text (section 4). The results have not significantly changed.

Also, why not show the comparison to stakes for the entire 20 months? It is unfortunate that the validation cores started almost a year late, but the stakes were measured monthly for a different project since 2003.

The entire time series comparison is now shown in Figure 7B.