Referee #1 – Authors reply

In the following, we provide a point-to-point answer to the referee's comments. The referee's comments start with RC:*, authors replies begin with AR: and are formatted italic with light-blue color.*

A markup of manuscript changes are shown in boxes with new text in dark blue and ~~removed text in red.~~

RC: The authors present a prototype sensor to measure SWE based on muonic cosmic ray. They derive SWE by fitting the count rate with a few manual measurements which were performed during one season on a Swiss glacier. Reliable methods to measure SWE temporally continuously in alpine environments are urgently needed. Studies as presented in this paper are therefore highly welcome and fit well into TC. I liked reading the manuscript, which has a clear structure and illustrative figures. The language is, with a few exceptions, easy to read. I suggest to accept the manuscript as soon as the following points have been addressed.

AR: We thank the referee for his/her dedicated time to our manuscript and the constructive feedback, which will significantly improve our manuscript.

RC: The described potential for reliable daily data seems much too optimistic when looking at Fig. 3a. The daily fluctuation shown therein make physically no sense. Even neglecting the large fluctuation in February, the daily signal (for the µ-CRSG and n-CRSG) often demonstrates strong negative changes during the accumulation season (e.g. April), which make only sense if you have large snow erosion. Please show a comparison with the daily snow depth change and discuss this problem. In light of the problem with daily values, a sentence like "the µ-CRSG promises to infer sub-daily SWE estimates with a higher precision than the n-CRSG" is quite bold!

AR: We added snow depth to Fig. 3, where accumulation periods agree fairly well with increases in both sensor's measurements. However, ablation periods cannot be compared directly to snow depth measurements because of the process of densification. While the snowpack densifies, SWE remains constant. Nonetheless, the additional comparison to snow depth further builds trust in our measurements.

Our conclusion that the µ-CRSG provides a higher precision than the n-CRSG is based on a theoretical approach. Following a Poisson distribution, the uncertainty within the count rates corresponds to the square root of the count rate divided by the total count rate. With lower count rates, this precision is lower. Since the muons are so abundant on the Earth's surface, the count rate is significantly higher, and thus, the theoretical precision too. We adapted the sentence to emphasize our reasoning behind it. To better understand the daily fluctuations, however, more measurements and different field experiments would be necessary.

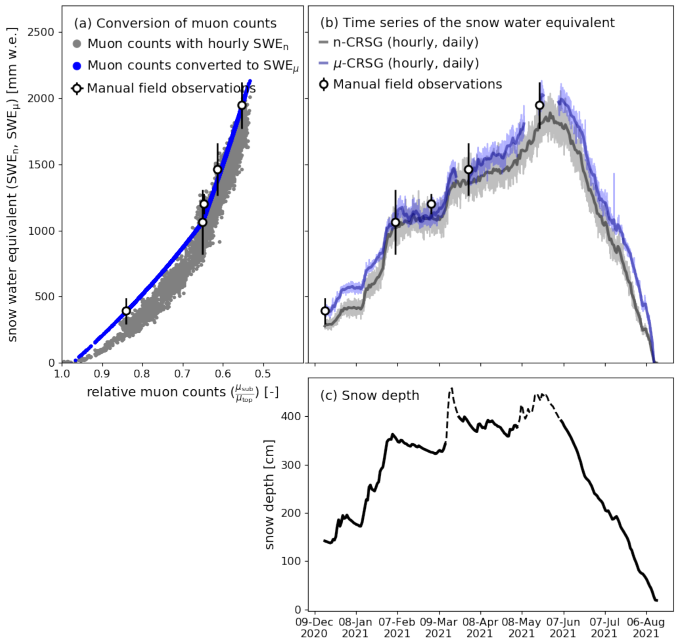

Fig. 3 From muon count rates to SWE. Panel (a) shows the relative muon count rates plotted against SWE that is measured by the n-CRSG (grey dots) and measured manually (white dots). Blue dots represent the SWE that is directly inferred from the relative muon count rate with the conversion function given in Eq.6. Panel (b) shows the time series of SWE inferred from neutron counts (grey) and muon counts (blue) at a daily resolution. Light grey and light blue represent hourly observations of the n-CRSG and μ-CRSG, respectively. Panel (c) shows snow depth measured at the site (solid line) with data gaps that were complemented from a nearby high-altitude station (dashed line).

[Section 4.2]

The independent snow depth measurements provide a further comparison to the n-CRSG and μ-CRSG estimates. Accumulation periods identified by the n-CRSG and μ-CRSG agree well with increases in snow depth measurements (Fig.3b and c). However, periods with snow erosion cannot be directly compared to the snow depth measurements because a decrease in snow depth can have two causes; densification of the snowpack, snow erosion or ablation. Therefore, a comparison of such time periods is not straightforward.

[Section 4.3]

Based on the theoretical precision estimation, the μ-CRSG promises to infer sub-daily SWE estimates with a higher precision than the n-CRSG. In addition, the hourly observations vary less around the daily mean for the μ-CRSG than for the n-CRSG (Fig.3b). Nonetheless, the μ-CRSG contain some inter-daily fluctuations that are larger in the μ-CRSG estimates than the n-CRSG. To understand these, further investigations are needed.

RC: You defined the uncertainty of the manual observations as the standard deviation of several observations during the same field day. These measurements usually have different SWE because neither the glacier surface nor the snow depth is perfectly homogenous. The account for these differences all same day measurements are usually referred to common snow depth (assuming that the bulk snow density is constant), which is often the one from the snow depth sensor, because this location should be undisturbed throughout the whole season. The same should be done when comparing the CRSG measurements with the manual measurements as both measurements usually have a different snow depth. Could please elaborate your procedure to deal with this issue?

AR: Thank you for this comment. We agree that it is an important point. In most field campaigns, especially those where we had a deep snowpack, all measurements are taken within the same snow pit. Often, we did not have enough time nor human power to excavate several deep snow pits on the same day. Snow cores are drilled where the snow pit is excavated afterwards. Hence, the spatial variability between the different measurements should be reduced to a minimum.

The spatial variability between the manual and automatic measurements is not straightforward to evaluate. The n-CRSG is placed about 7m away from the ultra sonic ranger to reduce influences of the metallic mast parts on the n-CRS (see newly added figure Fig. S1). In turn, we reduce the influence of the n-CRSG on the ultra sonic ranger footprint.

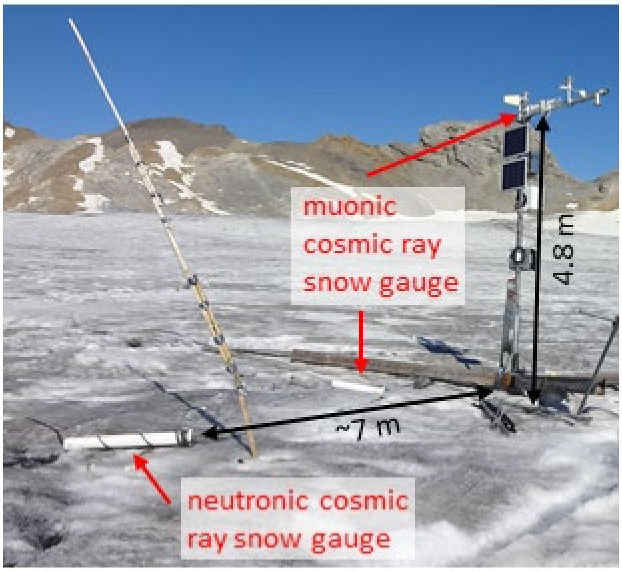

Fig. S1: Measurement setup on the Glacier de la Plaine Morte with the sub-snow neutronic cosmic ray snow gauge and the top and sub muonic cosmic ray snow gauge. [Photo courtesy: M. Huss]

Gugerli (2020) provides an example to demonstrate the low spatial variability at the site on Plaine Morte. For example, on 4 March 2020, 18 random snow depth measurements within 30m radius of the station were obtained. On average, the snow depth is 377±11 cm, which shows a spatial variability of ±3%. Simultaneously, the average snow depth differs by 3% from the ultra sonic measurement (389cm).

While the spatial variability of snow depth is low, the standard deviation of the SWE observations varies between 4% and 25% for measurements between October 2016 and May 2021. On the day with a 25% variation between the SWE observations (16 Dec 2020), the observations are based on three snow core samples. In conclusion, the variability in snow depth is low compared to the variability in samples within the same snow pit and/or over several snow pits.

Following the comment of referee #1 and #2, we now base our validation of the n-CRSG on re-calculated field data. Instead of calculating SWE for each sample, we derive an average manually-

obtained bulk snow density and multiply it with the undisturbed and autonomous snow depth observation by the ultra sonic ranger. The validation of the n-CRSG does not change significantly with this new approach. Generally, the temporal variability in the agreement between the field observations and the n-CRSG estimates improves.

We modified the manuscript as follows.

[2.2. Data]

The uncertainty of these manual observations is defined as the standard deviation of several observations during the same field day. Typically, the observations are taken within the same snow pit due to time restrictions. These snow pits and snow cores are located within a 30m radius of the station's mast. For each field campaign another location was used to avoid sampling a disturbed snowpack. SWE is calculated by multiplying the manually-obtained average bulk snow density with the autonomous and undisturbed daily snow depth observations by an ultra sonic ranger.

[3.1.2. Inferring SWE from neutron counts]

The 22 manual measurements are significantly and highly correlated with a coefficient of determination of 0.969 (Fig. 1a). On average, then-CRSG agrees with the manually obtained SWE with an underestimation of ~~-1~~ -2% and an uncertainty of ~~±12~~ ±10% (one standard deviation, Fig. 1b). The root mean square error amounts to 112mm w.e. Please also note that 50% of the manual field observations are obtained from snowpacks that are deeper than 1130 mm w.e.

[4.2. Evaluation of SWE inferred by muon counts]

Moreover, the validation with manual field observations may include a spatial and methodological uncertainty. While there is a spatial distance between the n-CRSG, µ-CRSG and the manual snow measurements, which are obtained at a different location on each field day, the site is rather flat and independent snow depth measurements have shown little variability within a 30m radius around the station (Gugerli, 2020). The manual snow observations are based on two main approaches; short tube samplings within snow pits and snow core samplings. To avoid an influence of the measurement approach, the estimated bulk snow density is based on the average over several samples. Hence, we limit uncertainties related to the measurement approaches, and integrate them within the standard deviation over all samples obtained at the same day.

RC: I cannot get the SWE numbers shown in Fig 3a with the given Equation 6 using the relative muon count rate as input? For a muon count rate of 0.7 for example, I get 89 and not ca. 1000 mm SWE as given by Figure 3a?

AR: Thank you for checking our calculations! We agree that it is crucial to have them correct. Indeed, there is a conversion mistake of a factor of 10 (from cm to mm). We apologize for this mistake and corrected the equation so that we get mm w.e. and not cm w.e.

We now calculate a SWE of 875 mm w.e. with a differential muon count rate of 0.70 (Eq. 6). This is also the value presented in Fig.3a.

We also adapted the figure in the supplement, where the fit corresponds to cm water depth and not to mm water depth.

RC: Is there a physical explanation behind the two-part conversion function, i.e. why there is a change in slope between 1000 and 1500 mm water depth in Fig. S1?

AR: This is a very interesting and important point, which requires more research. We are convinced that we also measure a softer component in addition to what is fairly certainly a pure muon component at depth. To distinguish between these two components, however, further field deployments and experiments would be necessary.

We added the following to the manuscript.

[Section 4.2]

With the fit between relative muon count rates and manually obtained SWE, the condition of having 0mm w.e. for a relative muon count rate of 1.0 is not fulfilled. Either a third part of the conversion needs to be introduced, or the fit needs to be repeated with more manual measurements. The transition within the conversion function could be caused by a softer component of the ionizing radiation from secondary cosmic rays. The ratio of the soft and hard component could also be location and especially elevation dependent. Hence, a site-specific calibration could be necessary. Nonetheless, this remains highly speculative and further measurement experiments would be needed to investigate it in more depth. A robust statistical evaluation of the presented conversion function is not possible nor representative because only two manual field measurements remain independent.

RC: Looking at Fig. 3a it may rather be a three-part conversion function?

AR: Yes, as mentioned in the manuscript, it may even be a three-part function. A function of the type used for the n-CRSG (e.g., Howat et al., 2018; Gugerli et al., 2019), which was derived for a similar purpose, might also be appropriate here.

RC: Please better discuss if the application of a μ-CRSG in other locations will always need some site-specific calibration or not.

AR: We believe a site specific calibration function could be needed given that the ratio of the two components (softer and harder component) could be location dependent. We believe this ratio should particularly depend on elevation. It should be straightforward to answer this question definitively through shielding experiments at different elevations. Another approach would be to try to eliminate or minimize one of the components, leaving a simpler and more universal calibration function.

[Section 4.2]

The transition within the conversion function could be caused by a softer component of the ionizing radiation from secondary cosmic rays. The ratio of the soft and hard component could also be location and especially elevation dependent. Hence, a site-specific calibration could be necessary. Nonetheless, this remains highly speculative and further measurement experiments would be needed to investigate it in more depth. A robust statistical evaluation of the presented conversion function is not possible nor representative because only two manual field measurements remain independent.

[Section 5]

In future studies, more manual measurements, further measurement experiments and simulations can improve our understanding of this measurement approach in addition to validate the presented conversion function.

Minor points:

L17: … measures the attenuation of incoming secondary neutrons on the ground below the snowpack to infer SWE.

AR: Done

L21: with an n-CRSG above the snowpack

AR: Done

L43: Would be nice to have a picture to demonstrate the setup of the instruments.

AR: We fully agree. We did not include a photo because of the figure limitation given for the manuscript type "Brief communication" in The Cryosphere. We added a photo of the setup in the supplement as Fig. S1 (see above) instead.

L45: …parameterizations have previously been investigated (please reference).

AR: We changed the sentence as follows:

> For the n-CRSG, parameterizations have previously been ~~investigated~~ applied and discussed to correct for changes in atmospheric pressure and incoming cosmic ray fluxes (e.g., Howat et al. 2018; Gugerli et al. 2019).