Reply to Reviewer R. S. Fassnacht

We thank you for your comments. We believe that your informed opinion and our effort to address your observations have resulted in sensible improvements of the quality of our manuscript.

For an easier interpretation of the contents of this reply we decided to adopt the following colour scheme

Reviewer comments

Authors reply

Proposed changes (lines refer to the updated manuscript)

General

The authors use 35 SWE measurement location-dates to calibrate six sites with cosmic ray neutron sensing (CRNS) probes during one winter of accumulation and then evaluate at 13 sites with 111 SWE location-dates. This is a very relevant "brief communication" that could be a decent paper. However, more information is needed to explain how the calibration (ground-truth) and evaluation SWE data were collected – see below. It is unclear how the data were collected and more importantly whether the data are representative of the area around the CRNS probe.

We thank you for this comment and we acknowledge the limited description of the adopted manual SWE sampling methodology. We would like to stress that this was done to preserve the compact structure proper of this kind of publication. In fact, we opted to give a quick description of the methodology while citing the appropriate existing literature which gives an extensive description of the sampling techniques. Nevertheless, we are glad to expand on each of your useful comments in the following replies, hoping to fully address every nuance for the sake of clarity and scientific rigor.

L91-92: The measurement footprint for this setup has a radius of approximately 20 m, a feature that allows its SWE estimates to be directly compared with point-scale manual samples.

L111-116: Each vertical core was taken with an Enel Valtecne EV2 corer which has an internal diameter of 6 cm and is composed by a modular setup of stainless-steel sections with a length of 50 cm each (Lopez-Moreno, 2020b). The vertical coring process is repeated for at least three times for each sampling campaign, resulting in multiple SWE and bulk density values. Subsequently, those measures are averaged to obtain an estimate of the SWE. For our analysis we used these averaged values assuming they represent the (point-scale) amount of SWE at each site.

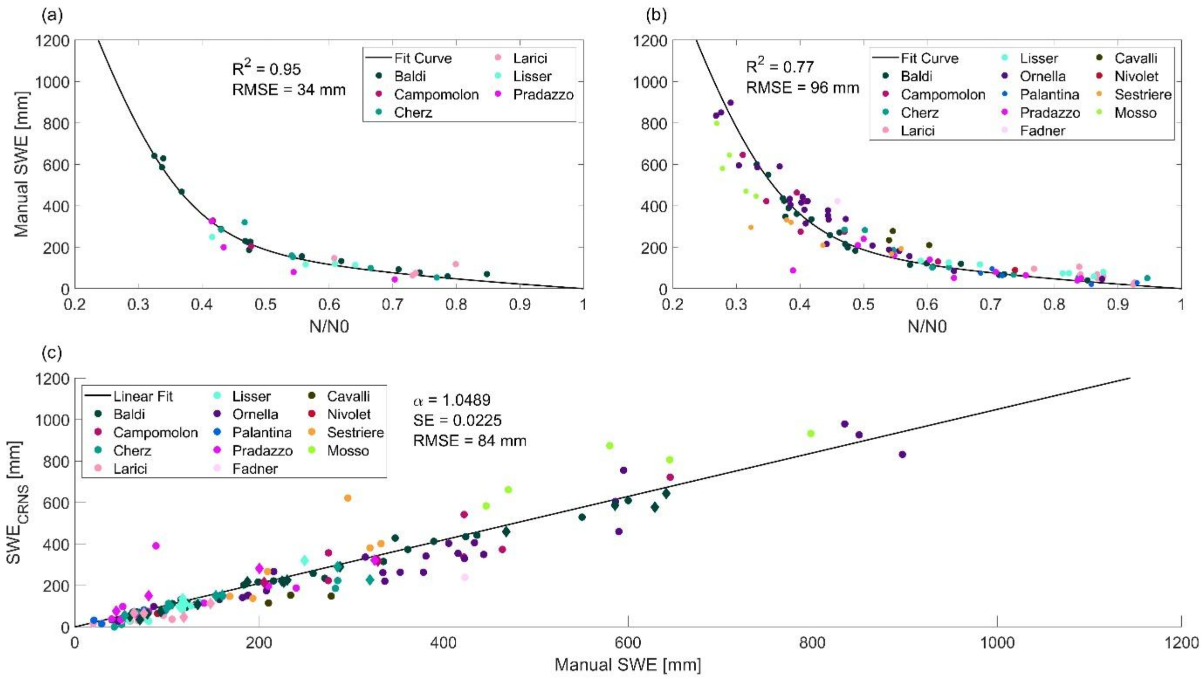
L121-122: The measurements and, crucially, the layers identification are carried out by expert researchers and technicians who have been extensively trained by the Italian Association for Snow and Avalanches (AINEVA).

L124-125: as it has been reported in existing literature focused on snow resources of our area of study (e.g. Colombo et al., 2022; Guyennon et al., 2019)

The first results are shown in Figure 2. These are informative. However, it would be useful to consider a different way to show the points, at least the calibration points – I cannot tell some of the colours apart. I suggest using different symbols, and perhaps the elevation colour ramp from Figure 1.

We thank you for this comment. To address a comment by the other reviewer we propose a division of Figure 2 in 6 panels that represent the regression curve and linear fit for the subsets of: (i) calibration data, (ii) validation data from calibration sites, and (iii) validation data from other sites. We believe that this modification may help in distinguishing which data belongs to each site addressing the concerns you expressed.

OLD FIGURE 2:



NEW FIGURE 2:

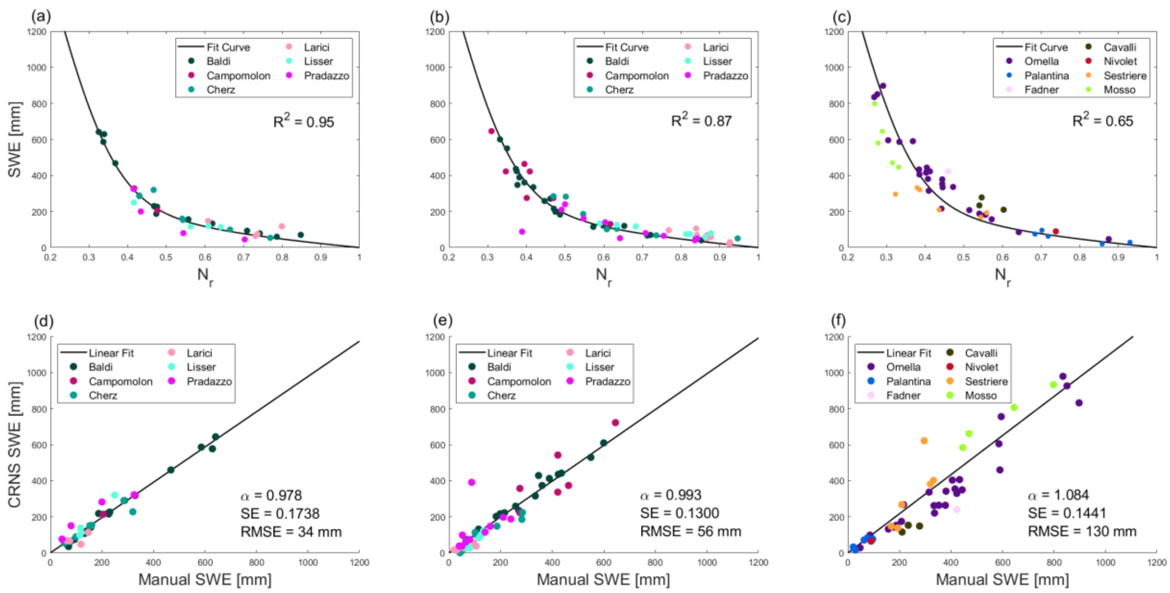


Figure 2: Normalised daily neutron count rate (x-axis) against SWE direct measurements (y-axis) for each site (coloured dots), and regression curve (black line) for the calibration dataset (a) and validation dataset divided between the six sites chosen for calibration (b) and the remaining seven sites(c). Manual SWE measurements plotted against SWE_{CRNS}; coloured dots represent the calibration (d), validation in the same sites as calibration (e), and validation at different sites (f) dataset respectively; the black lines illustrate the linear regression fit performed for each dataset.

For the Data (and Figure 2), can you provide more information about the 35 location-day samples? A table in an Appendix or Supplementary Materials would be fine. There appear to be 13 samples from Baldi, 3 from Larici, etc. An indication on when these samples were taken would be informative. The calibration station (a) are also used in the evaluation (b). Better explain how and why the data were split into groups.

Thank you for this comment. We propose to add an additional column in Table S1 containing the number of samples taken at each site. Furthermore, we propose to add two figures in the Supplementary Material showing the monthly distribution of the manual SWE samples of the calibration and validation datasets respectively.

Name	Elevation [m a.s.l.]	Latitude [° N]	Longitude [° E]	# Manual SWE samples (# used for calibration)	Time Series Length
Lisser*	1422	45.9543	11.6510	11 (4)	11/08/2023 – 01/09/2025
Campogrosso	1462	45.7285	11.1684	/	12/11/2023 – 01/09/2025
Palantina	1492	46.1189	12.4759	5	26/09/2023 – 01/09/2025
Sant'Antonio	1495	46.5720	12.4858	/	21/12/2023 – 01/09/2025
Faverghera	1603	46.0751	12.3013	/	07/08/2023 – 01/09/2025
Larici*	1606	45.9636	11.4193	12 (4)	10/08/2023 – 01/09/2025
Campomolon *	1732	45.8732	11.2729	9 (2)	10/10/2023 – 01/09/2025
Monte Baldo	1756	45.7731	10.8657	/	13/09/2023 – 01/09/2025
Malga Losch	1757	46.2624	11.9595	/	06/10/2023 – 01/09/2025
Casera Doana	1887	46.4812	12.5410	/	19/09/2023 – 01/09/2025
Baldi*	1913	46.4152	12.0731	35 (15)	21/09/2023 – 01/09/2025
Coltrondo	1929	46.6652	12.4435	/	20/09/2023 – 01/09/2025
Cima Dodici	1965	45.9790	11.4868	/	18/07/2024 – 01/09/2025
Val Visdende	2001	46.6538	12.6174	/	09/11/2023 – 01/09/2025
Cherz*	2010	46.5153	11.8782	15 (6)	01/11/2023 – 01/09/2025
Monte Rite	2013	46.3823	12.2484	/	09/07/2024 – 01/09/2025
Fadner	2155	46.9256	11.8614	2	25/10/2024 – 01/09/2025
Pradazzo*	2195	46.3559	11.8227	17 (5)	25/11/2023 – 01/09/2025
Ornella	2227	46.4757	11.8863	25	05/10/2023 – 01/09/2025
Cavalli	2255	46.4694	10.8194	3	30/10/2024 – 01/09/2025

Monte Piana	2262	46.6134	12.2503	/	05/10/2023 – 01/09/2025
Tre Cime	2375	46.6176	12.3178	/	10/07/2024 – 01/09/2025
Sestriere	2490	44.9540	6.9111	7	15/10/2023 – 01/09/2025
Nivolet	2555	45.5192	7.1714	1	02/01/2024 – 01/09/2025
San Martino	2580	46.2793	11.8864	/	11/07/2024 – 01/09/2025
Mosso	2901	45.8754	7.8717	5	28/12/2023 – 01/09/2025

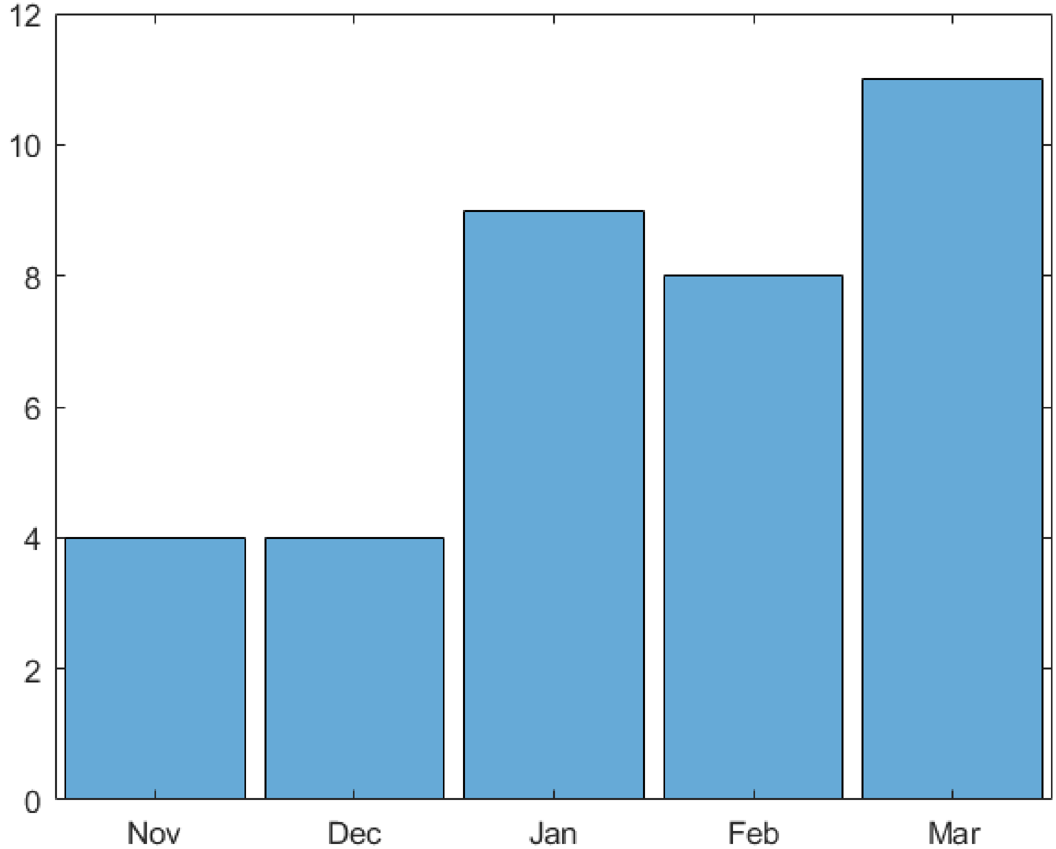


Figure S1: Distribution showing in which month the manual SWE samplings used for calibration (35 data points in total) were taken.

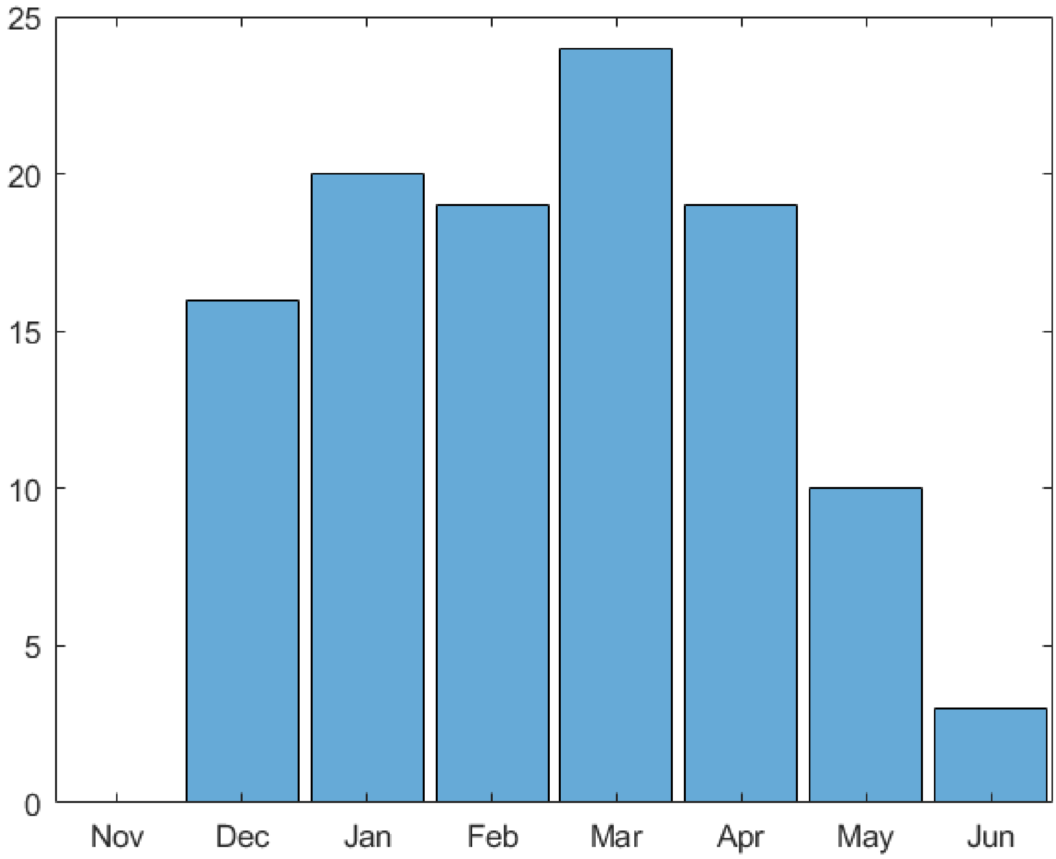


Figure S2: Distribution showing in which month the manual SWE samplings used for validation (111 data points in total) were taken.

More information is necessary in explaining how SWE was measured. It appears that samples were either taken as a single core or as a series of horizontal cores. Which method was used for which site and which date?

- For the single vertical core, provide the details about the sampler, including the inside diameter/cross-sectional area (see López-Moreno et al., 2020; https://doi.org/10.1002/hyp.13785). Further, how many samples were taken? For the single corer, at least three samples are usually measured (see López-Moreno et al., 2013; https://doi.org/10.1016/j.advwatres.2012.08.010). No uncertainty can be assigned to a single sample.

  We thank you for this comment and for providing useful references. The vertical corings were performed with the Enel-Valtecne EV2. It is widely adopted in Italy both by environmental protection agencies and research labs (as it is also stated in Lopez-Moreno et al., 2020). As you mentioned, multiple corings are necessary to ascertain the uncertainties of SWE samplings. We apologise for not reporting this in detail, however, for each manual SWE measurement taken with the EV2 at least three corings were taken to assess SWE. Further details on the methodology adopted for the vertical corings can be found in the cited Berni and Giancanelli (1966) which has been providing the guidelines for this kind of measurements in the Italian Alps for more than half a century.

L113-118: Each vertical core was taken with an Enel Valtecne EV2 corer which has an internal diameter of 6 cm and is composed by a modular setup of stainless-steel sections with a length of 50 cm each (Lopez-Moreno, 2020b). The vertical coring process is repeated for at least three times for each sampling campaign, resulting in multiple SWE and bulk density values. Subsequently, those measures are averaged to obtain an estimate of the SWE. For our analysis we used these averaged values assuming they represent the (point-scale) amount of SWE at each site.

- For the horizontal coring, it is unclear how representative horizontal cores are (see the previous references and citations therein, including Elder et al., 2009). Although Valt et al. (2012) and Valt (2019) seem to indicate that the method used here (especially the vertical coring) seems sufficient, these citations are from conference proceedings? [The latter should be 2018.] The method of measuring SWE is crucial for this paper, yet the validity and error are poorly explained. Sampling horizontally with a tube oversamples the middle and undersamples the sides. Further, measurements were only taken within layers. This does not provide a representative sampling of the entire snowpack. How were the layers identified? How do we know that they are homogeneous?

  We thank you for pointing out the typo in reporting Valt 2018 reference, we propose to correct it with Valt (2019). As we mention in our manuscript, most of the SWE measurements we leveraged were not originally meant for CRNS calibration and validation purposes. In fact, horizontal coring of homogenous layers is a quite common methodology in the Italian Alps, as it is carried out regularly to assess avalanching risks. Even though the correct identification of homogenous layers within the snowpack may be subjected to human error (as it is the case for every instance of manual measurement), these identifications and measurements are carried out by researchers and technicians that undergo extensive training provided by the Italian Interregional Agency for Snow and Avalanches (AINEVA), therefore we are confident in the reliability of our dataset. Although we concur on the fact that vertical corings, when available, are preferable to horizontal ones, we would like to stress that horizontal coring still gives a reliable estimate of the entire snowpack SWE. This has been the consensus for our Area of Interest (i.e. the Italian Alps) and has been reported and accepted in peer-reviewed papers (e.g. Colombo et al., 2022 and Guyennon et al., 2019). As we underline both in Section 2 and in the Conclusions, a standardised and shared methodology to retrieve SWE samples aimed specifically for CRNS validation has yet to be defined. We hope that our work may provide a reasonable starting point for advancing the community effort in this direction.

  L124-125: (Valt, 201~~9~~8) as it has been reported in existing literature focused on snow resources of our area of study (e.g. Colombo et al., 2022; Guyennon et al., 2019)

- Finally, were samples taken over an area? The snowpack is not homogeneous and the CRNS collects data in an area with a 200 to 300 meter radius around the CRNS probe. With only measuring SWE near the CRNS probe, we don't know what SWE (and soil moisture) that the CRNS probe is actually sampling/measuring. This heterogeneity is crucial and explored in many papers. Without having ground-truth, it is unclear what to make of any of the results. Figure 2a presents the calibration and is dependent on four measurements at Baldi (SWE > 400 mm). Without those four measurements, there is limited correlation. Fortunately, Figure 2b seems to show that the calibration works for the evaluation.

  Samples were not taken over an area since the network's sensors have a limited footprint with a radius of ~20 m as declared by the probe's manufacturer. We propose to explicitly report the footprint. We suppose that your comment on a much wider footprint is based on a different typology of CRNS probes which consists only of a detector on a mast and give an "areal" SWE estimate with the tradeoff of a much lower saturation threshold (few hundreds mm water equivalent). This characteristic makes the aforementioned setup unapplicable in most Italian high mountain sites. As we stated in the methodology, most of the SWE sampling campaigns we leveraged for our study were not originally meant for validating the CRNS probes. To avoid misrepresentation, we discarded data with reported snow depth too distant to what was measured by the AWSs. Therefore, we reckon that snowpack heterogeneity between the probe's location and each manual sample collection site plays only a minor role, making the manual and CRNS datasets comparable.

L92-93: The measurement footprint for this setup has a radius of approximately 20 m, a feature that allows its SWE estimates to be directly compared with point-scale manual samples.

Specifics

- Line 23: you are not "validating" the parameterization, but rather "evaluating," since you don't know the SWE sensed by the CRNS, nor the actual SWE unless you measure all the snow, i.e., you cannot truly scale from measurements to an area.

We thank you for this comment, however we disagree with this statement. As we reported in the reply to the last general comment, the CRNS probes used in this study measure on a footprint which has a ~20 m radius. At this scale, the most relevant source of heterogeneity in the snowpack water equivalent is given by snow depth variations. In our data filtering process, to address this issue, we discarded manual samples taken in points where HS did not match what had been measured at the AWS. Therefore, we believe that the process of comparison between CRNS data and point-scale SWE samples is indeed an instance of validation.

- Line 25: why is elevation relevant?

  It is stated in the results regarding the correlation between SWE time series from monitored and unmonitored sites, which sensibly decreases with elevation differences while it seems to be less affected by horizontal distance (see Fig. 3). However, we propose to modify the abstract to clarify this point.

L24-28: Further results show that correlations between SWE series from coupled monitored and unmonitored sites are strongly affected by elevation differences. ~~which can be extended to unmonitored sites if they have monitored counterparts at similar elevation. This finding overcomes the need for year-round accessibility and increases the number of potential sites for continuous SWE retrieval~~.

- Line 33: what does "snowmelt dates are projected to be anticipated by as much as" mean? This is unclear.

  Thank you for pointing out this lack of clarity. We used the term "snowmelt dates" as it is proposed by the cited Vorkauf et al. 2021. It refers to the calendar day when the snowpack completely melts resulting in an absence of snow cover. We propose to change the manuscript to clarify this point.

L35-36: ~~snowmelt dates are~~ the dates when complete snowpack waning happens in the water year are

- Figure 1: show/distinguish the calibration versus evaluation sites. It is unclear what "unmonitored" means

Unmonitored sites are those where manual SWE samplings are not performed during the snow season. We propose to clarify this in the Methods Section and in the description of Figure 1.

L85-90: **Figure 1: Elevation map of the Alps and northern Italy. ~~Monitored and unmonitored s~~Sites of the network are indicated by circles and diamonds ~~respectively~~; circles correspond to locations where manual SWE samples are available (i.e. monitored sites), while diamonds are used for location without available manual SWE samples (i.e. unmonitored sites); the symbols filling follows a scale of blue to represent the elevation of each site, where darker shades correspond to higher elevations. Produced using Copernicus WorldDEM-90 © DLR e.V. 2010-2014 and © Airbus Defence and Space GmbH 2014-2018 provided under COPERNICUS by the European Union and ESA; all rights reserved.**

L197-198: (i.e. the remaining 13 sites of the network where manual SWE samplings were not available)

- Line 91: why does N0 vary as a function of elevation?

  The number of neutrons reaching the ground (in absence of snowpack or other obstacles), strongly depends on atmospheric thickness which diminishes with elevation.

- Data and Methods: use sub-sections or at least new paragraphs within this section to keep the piece separate.

  Thank you for this comment. To obtain a tidier structure of the manuscript we divided Section 2 in 4 sub-sections (see reply to the other reviewer). The proposed sub-sections are: 2.1 Network and station setup, 2.2 Manual SWE sampling, 2.3 CRNS data processing and SWE computation, and 2.4 Calibration and validation against manual samples.

- Line 107: it is a force gauge or a scale. Or is it a "dynamometer?"

  It is a scale. We apologise for the lack of clarity caused by the fact that scales in the native language of all the authors are also referred to as dynamometers. We propose to change the term "dynamometer" with "scale" to avoid any confusion.

  L120: ~~dynamometer~~ scale.

- Figure 2: it should be clearly stated that the curve in (b) is derived from the data in (a).

  We thank you for this. We propose to address this lack of clarity in the modified Fig. 2. (See above)

- Figure 2c: are these the points from (a) and (b) replotted? This is not clear. What are the pink diamonds?

  For the Manual SWE they indeed are points from (a) and (b) replotted. However, (c) showed the linear regression of SWE obtained from the regression curve depicted in the other panels where diamonds represented calibration data (maintaining the colours assigned to the stations in the other panels). In any case, we believe that lack of clarity pointed out in this comment and in the previous one have been addressed by the proposed changes in Fig. 2 which now clearly separates the data showed in three subsets (i.e. calibration, validation at calibration sites, and validation at different sites) both for regression curve and linear fit.

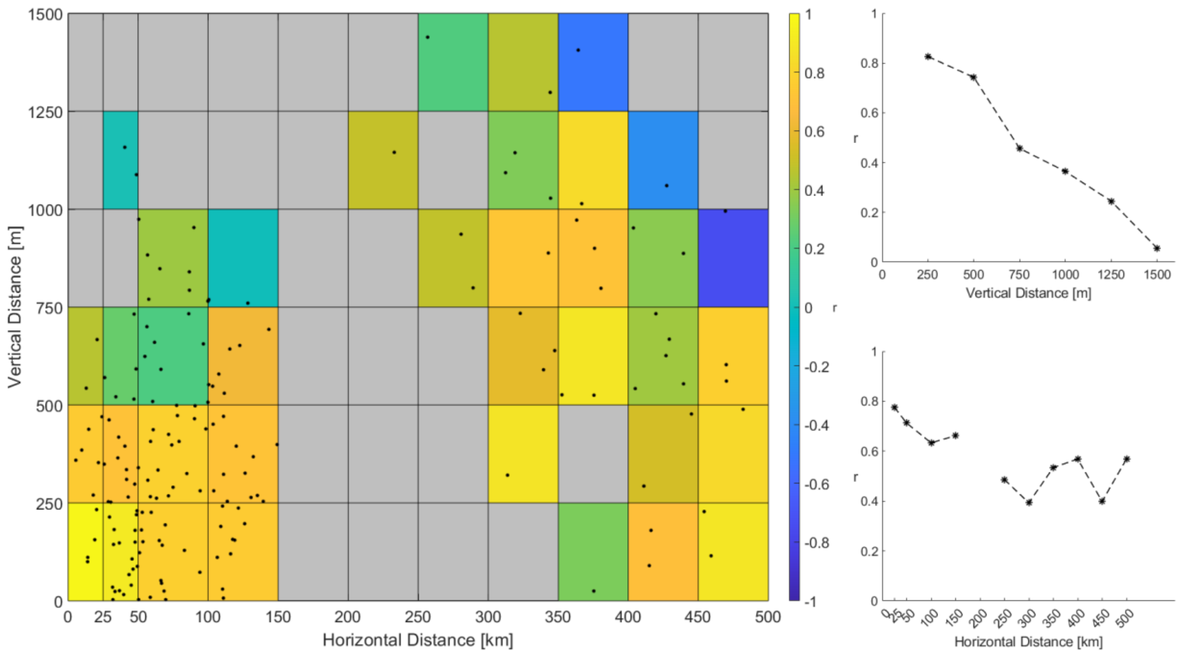
- How do the unmonitored sites fit in? That is not clear. What does unmonitored mean? Does this mean that there are no field SWE measurements? Is SWE computed from the fitted equation?

  Thank you for this comment that helps us to further clarify this detail. Unmonitored sites are the sites of the CRNS network that are unreachable during the snow season and, therefore, lack any manual SWE measurement. SWE time series are computed from the regression curve presented in the earlier part of the manuscript. We believe that an important asset of the network lays in the fact that these probes allow to get SWE data even at sites unreachable during the snow season. Although we recognise that further investigation into the probes performances at these sites are desirable, our intention for the part showing correlations between monitored and unmonitored sites is to provide preliminary steps to guide and inform on additional research on this crucial matter.

- Is Figure 3 only between monitored and unmonitored? How were the “distance classes” (I think those are the different colour blocks) decided? These seem random? What about showing the R value for each point in Figure 3a? This plot needs to be better explained. Is this a common plot? Provide a citation.

  Thank you for this comment. Each black dot in Fig. 3 shows a pair of unmonitored-monitored sites, for a total of 78 points given by the possible unique combinations of 13 monitored vs 13 unmonitored. Following your comments as well as the ones from the other reviewer, we propose to modify Fig. 3 description to improve its clarity.

  Distance classes were decided based on the feature of the network, that is composed mainly by sites that have a horizontal distance of less than 50 km among themselves with a lack of sites with similar (i.e. less than 750 m) elevations in the 200-400 km distance range. We include below a figure analogous to the current Fig. 3 but relative to preliminary analysis that was carried out with evenly divided distance classes of 50 km (except for the 0-25 km and 25-50 km) and elevation classes of 250 m.



As you would notice, there are several grey tiles (i.e. blocks without any couple of stations meeting the distance criteria) in the aforementioned sections and, moreover, many coloured tiles contain only one or two points. Given that the colour of each tile represents the average correlation computed from the points it contains, from our perspective, having less than three points in a tile may hinder its overall representativity. That said, we propose to add some explanation in sub-section 2.4 to show that the classes choice was indeed an informed decision.

L198-205: For each of the 78 possible pairings of unmonitored and monitored stations, we computed the correlation coefficient between their N_r and SWE_{CRNS} time series, considering only periods where the reference AWS data showed presence of snow on the ground (i.e. HS > 0 cm). We assigned the obtained correlations to different classes based on the vertical and horizontal distances between each pair of stations and computed the average correlation for each class. The chosen divisions for the vertical distances are 250 m, 500 m, 750 m, 1000 m, and 1500 m; while for the horizontal distances we chose 25 km, 50 km, 100 km, 200 km, and 500 km. We defined these divisions based on the spatial distribution of the network's sites and pursuing two goals: (i) minimize the number of void classes, and (ii) avoid classes containing only one couple of stations to make its average correlation informative.

L279-282: **Figure 3: (a) Heatmap of the average correlations of the SWE_{CRNS} time series classified by the horizontal (x-axis) and vertical (y-axis) distances between unmonitored and monitored stations. The black dots represent each pair of monitored-unmonitored stations for which r was computed for a total of 78 pairings. Average r values computed for each ~~depending on~~ vertical (b) and horizontal (c) distance~~s~~ classes. The x coordinates of each point represent the upper limit of its distance class.**

- There is no real discussion of the results.

We acknowledge this comment. However, we note that a network of CRNS probes used for SWE measurement constitutes an unprecedented novelty in the existing literature both for its scale and the number of measures leveraged. Therefore, we were only able to partially discuss our results comparing them with what had been presented in studies where CRNS applications for SWE are presented only at single sites.

Addition to the references

- Colombo, N., Valt, M., Romano, E., Salerno, F., Godone, D., Cianfarra, P., Freppaz, M., Maugeri, M., Guyennon, N.: Long-term trend of snow water equivalent in the Italian Alps, Journal of Hydrology, 614(A), 128532, doi:10.1016/j.jhydrol.2022.128532, 2022.
- Gottardi, F., Carrier, P., Paquet, E., and Laval, M. T.: Le NRC: une décennie de mesures de l'équivalent en eau du manteau neigeux dans les massifs montagneux français, International Snow Science Workshop 2013, 33, 926–930, https://arc.lib.montana.edu/snow-science/objects/ISSW13_paper_O2-08.pdf, 2013.
- Jitnikovitch, A., Marsh, P., Walker, B., and Desilets, D.: Snow water equivalent measurement in the Arctic based on cosmic ray neutron attenuation, The Cryosphere, 15, 5227–5239, doi:10.5194/tc-15-5227-2021, 2021.
- Lopez-Moreno, J. I., Leppänen, L., Luks, B., Holko, L., Picard, G., Sanmiguel-Vallelado, A., Alonso-González, E., Finger, D. C., Arslan, A. N., Gillemot, K., Sensoy, A., Sorman, A., Ertaş, M. C., Fassnacht, S. R., Fierz, C., and Marty, C.: Intercomparison of measurements of bulk snow density and water equivalent of snow cover with snow core samplers: Instrumental bias and variability induced by observers. Hydrological Processes. 2020; 34: 3120–3133. doi:10.1002/hyp.13785, 2020a.

Sincerely in name of all the co-authors,

Mario Gallarate (mario.gallarate@unito.it)