We thank Prof. Margulis for the encouraging evaluation of our work and for the constructive suggestions, which will help us improve the manuscript. We respond to each specific comment below and describe the changes we intend to make in the revised manuscript.

Specific comments (SC):

SC1: Line 59-67. In discussing other observationally-constrained snow datasets, you should consider the mention of Cortés and Margulis (2017), Liu et al. (2021), and Fang et al. (2022), which demonstrate multi-decadal SWE estimation over High Mountain Asia, the Andes, and the Western U.S. respectively.

Reply to SC1: We are grateful to Prof. Margulis for having suggested these three references, which are directly relevant to the paragraph in question, and we will cite them in the revised version of the manuscript. Cortés and Margulis (2017), Liu et al. (2021), and Fang et al. (2022) present multi-decadal, observation-constrained snow reanalyses for the Andes, High Mountain Asia, and the western United States respectively, in which fractional snow-covered area retrieved from the Landsat record is assimilated to produce sub-kilometre SWE estimates. Since the Landsat archive extends back to the mid-1980s, these reanalyses span periods of up to nearly four decades. Their example shows that the temporal coverage of observation-constrained SWE products is determined by the length of the assimilated observational record and by the design of each individual product, rather than being an inherent limitation of the product class, and we will revise the paragraph to make this distinction explicit.

SC2: Figure 2, panel (b). It may make it easier for the reader to interpret the SWE if you centered the mean annual temperature colorbar on zero degrees Celsius. It would provide visual evidence of regions that are below freezing for a significant fraction of the year and thus more likely to have snowfall than rainfall.

Reply to SC2: We fully agree that centring the temperature colour scale on zero degrees Celsius would make the figure easier to interpret, and we will revise the colour scale accordingly in Fig. 2, panel (b), and in Fig. A2, panels (a) to (d), where the same scale is used. We thank Prof. Margulis for the suggestion.

SC3: Section 3.1. Given that the IT-SNOW is the reference product, a bit more detail on how it works would help the reader. For example, how is the assimilation implemented? Is it in near-real-time? Retrospective? What is the assimilation method used? While these details are in the provided references, having the basic summary would be helpful.

Reply to SC3: We agree that, given the central role of IT-SNOW as the reference product of this study, a basic overview of how it is generated would benefit the reader without requiring consultation of the original publications.

To answer the Reviewer's questions directly, IT-SNOW is a retrospective product. It was generated by rerunning the S3M Italy operational chain over the whole historical period with the station data and satellite products available for that period, and it is therefore not an archive of the outputs produced in near real time by the operational system. Within this rerun, assimilation is performed once per day. The measurements from the snow depth monitoring network are spatialised into a gridded snow depth map by means of linear regressions between the measured values and physiographic predictors, and this map is then assimilated into the model. Assimilation is restricted to the period from December to April, since outside these months the ultrasonic sensors may detect the vegetation growing beneath them after melt-out, which would introduce

spurious signals into the spatialised maps. The scheme itself follows a Newtonian relaxation, or nudging, approach: at each grid cell the modelled SWE is corrected by a fraction of its departure from the observation-based estimate, and this fraction decreases with distance from the snow depth sensors, on the assumption that the spatialised maps are more reliable in densely instrumented areas (Avanzi et al., 2023). The conversion from measured snow depth to SWE relies on the modelled snow density, on the assumption that the spatial variability of density is considerably lower than that of snow depth (Mizukami and Perica, 2008). A blended snow-covered area product combining Sentinel-2, MODIS/VIIRS and EUMETSAT H SAF H10 retrievals is used in addition, not as a further assimilated quantity but as a mask applied to the spatialised snow depth maps prior to assimilation, as detailed in our reply to SC4.
In the revised manuscript, we will complement the description currently provided in Sect. 3.1 with a condensed version of this summary.

SC4: Line 151: How informative is the snow covered area in updating SWE? It typically has limited instantaneous relationship with SWE (i.e. often saturates near 100% for much of the accumulation season). Does the assimilation more heavily rely on the snow depth observations as a constraint?

Reply to SC4: We thank Prof. Margulis for pointing this out. The instantaneous relationship between snow-covered area and SWE is indeed weak over a consolidated snowpack, where SCA saturates near 100 % for much of the accumulation season, and the design of the IT-SNOW assimilation is consistent with this limitation. In S3M Italy, the SCA maps are in fact not assimilated directly: the quantitative constraint on the modelled snowpack comes entirely from the snow depth maps, which are converted to SWE using modelled snow density and assimilated through a spatially distributed Newtonian relaxation scheme (Avanzi et al., 2023). The blended SCA maps are used to remove snow-free pixels from the snow depth maps before assimilation. Their information content therefore enters the system as a binary, snow / no-snow constraint on the extent of the assimilated fields, correcting the regression-based spatialisation where it would otherwise extend snow into areas observed as snow-free, rather than as a constraint on snow mass. The main role of assimilating SCA is, then, to get the snow line right, which is crucial especially during rain-on-snow events (Parajka et al., 2019). We also recognise that the current wording at lines 150 to 155, which introduces the SCA maps as the first of two assimilated data sources, may suggest a direct assimilation of SCA. In the revised manuscript, we will rephrase this passage to state explicitly that the SCA maps act as a mask applied to the snow depth fields prior to assimilation, and that the quantitative update of the modelled SWE relies on the snow depth observations.

SC5: Line 158: It would be helpful to the reader to provide some of the bulk verification statistics from previous work (i.e. define what is meant by "consistent" and "low mean bias" more quantitatively).

Reply to SC5: We agree that the terms "consistent" and "low mean bias" might remain vague without the corresponding figures, and that reporting the main verification statistics gives the readers a more concrete measure of the reliability of the reference dataset. These statistics are available from the evaluation reported in Avanzi et al. (2023), which relies on two complementary sources of independent data. The first is the Sentinel-1-based C-SNOW snow depth product, against which the bias and the root mean square error were computed at each grid cell between the time series of the two products over the period September 2016 to April 2020, and then averaged across the evaluated mountain areas: the resulting mean bias is close to zero (−0.01 m), with a mean RMSE of 0.22 m. The second consists of in situ snow depth and SWE measurements from three focus regions spanning different snow climates, from the deep and seasonally consistent Alpine snowpack of Aosta Valley and Lombardy to the more ephemeral and maritime snow cover of Molise. Across these regions and measurement types, the root mean square errors are typically of the order of 30 to 60 cm for snow depth and 90 to 300 mm for SWE, with the width of these ranges partly reflecting the regional differences in verification skill, which Avanzi et al. (2023) relate to the density of the assimilated sensor network. In the revised manuscript, we will integrate these statistics into the text at line 158, so that the qualitative wording is replaced by the quantitative results it refers to.

SC6: Figures 3 and 4. I would suggest adding a third row at the bottom showing the bias as a percent of the reference. Given the large difference in mean annual SWE or SCD values, it would be complementary to show the percentage bias in addition to the absolute bias.

Reply to SC6: We agree that a relative measure of the discrepancies can complement the absolute one. We have therefore computed the percentage bias for both mean annual SWE and SCD. Figures R1 and R2 below reproduce the current Figs. 3 and 4, respectively, with percentage bias (dividing the bias by the corresponding IT-SNOW reference value) added as a third row of panels.

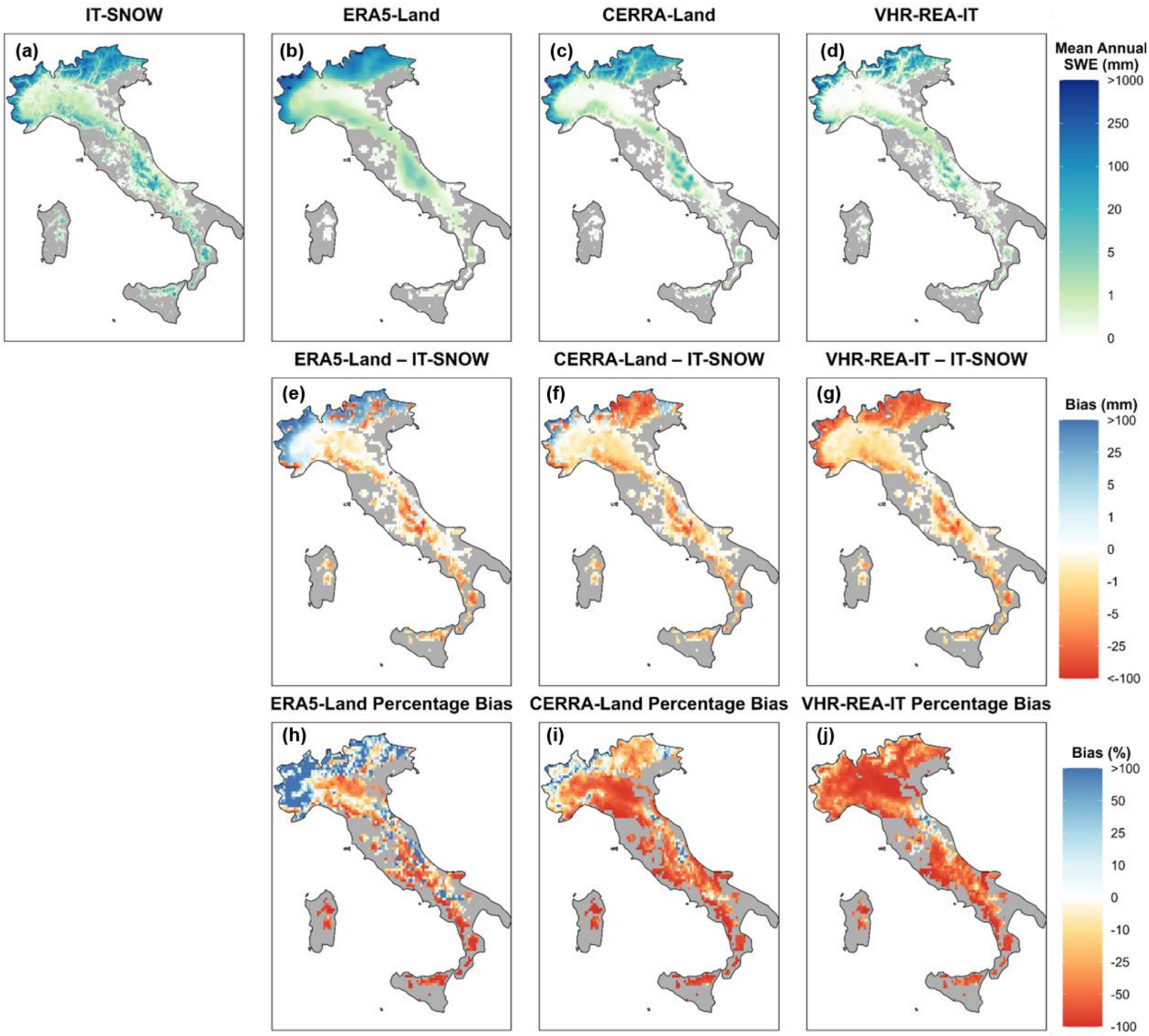

Figure R1. Mean annual snow water equivalent (SWE), September 2010–August 2024. (a–d) Mean annual SWE for IT-SNOW and the three land-surface products at native resolution. (e–g) Bias (Product − IT-SNOW) on the common 9 km grid. (h–j) Corresponding percentage bias.

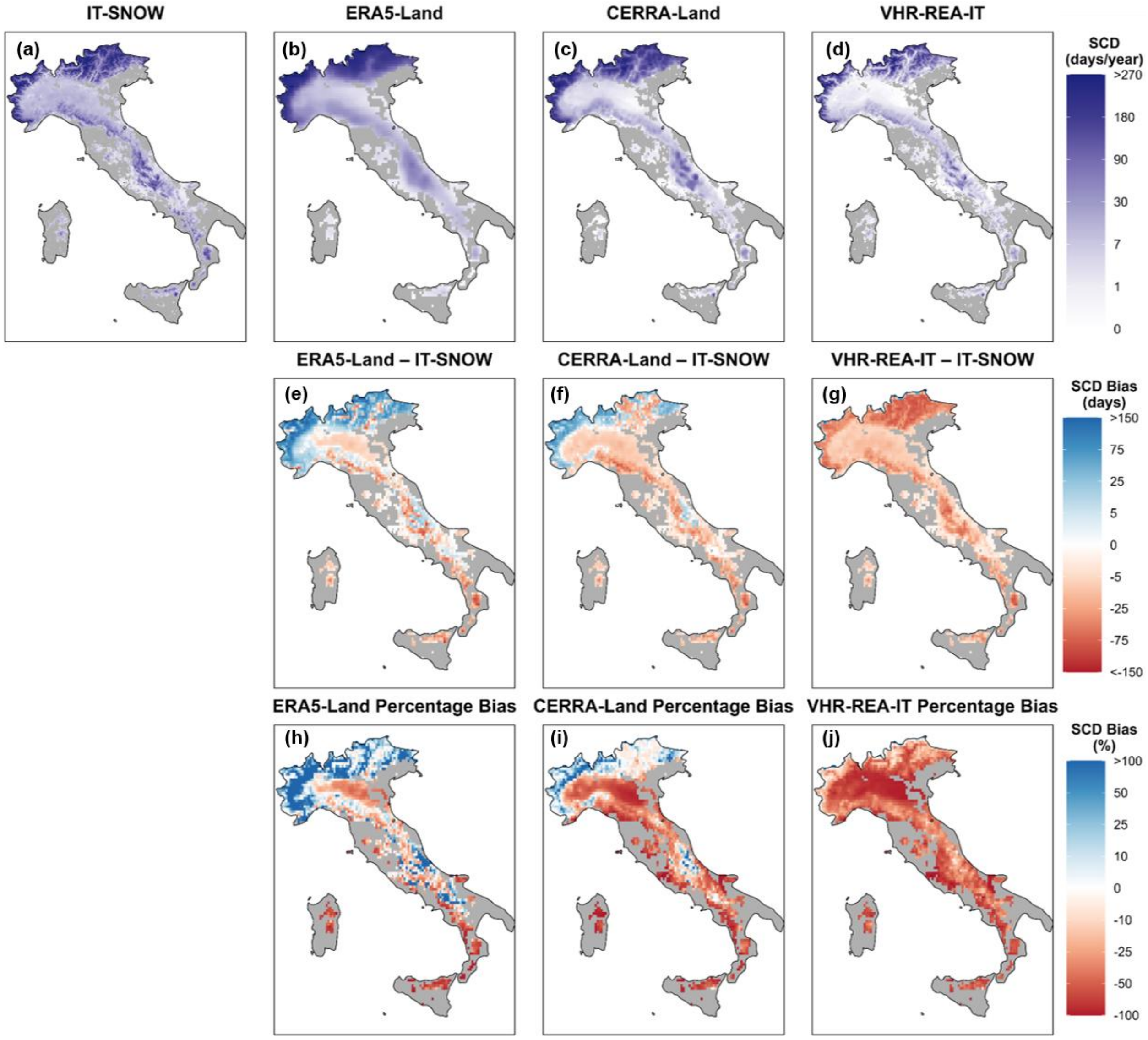

Figure R2. Mean annual snow cover duration (SCD), September 2010–August 2024. (a–d) Mean annual SCD for IT-SNOW and the three products at native resolution. (e–g) Bias (Product − IT-SNOW) on the common 9 km grid. (h–j) Corresponding percentage bias.

Having examined these figures, we would nonetheless like to comment on what they add and on where they can be best placed.

The analysis domain comprises all grid cells in which the IT-SNOW mean annual SWE exceeded 1 mm in at least one year of the study period (thus excluding the cells in grey), and therefore includes cells where the reference mean annual SWE amounts to a few millimetres and the reference mean annual SCD to a few days. Since the reference values in the denominator are positive, the percentage bias has the same sign as the absolute bias everywhere, and the two rows differ only in the relative weight they give to each cell. Over the Alps and the higher Apennines, where the reference values are large enough for the normalisation to be stable, the patterns of the percentage bias follow those of the absolute bias, and the largest absolute discrepancies correspond to comparatively moderate percentage values. Where the reference values are small, in contrast, differences of little practical consequence in absolute terms correspond to percentage values of several hundred per cent, and the percentage bias saturates the colour scale over much of the lowland part of the domain; the saturation there reflects the magnitude of the reference value rather than that of the discrepancy.

Since the percentage bias adds no information on the sign or the spatial pattern of the discrepancies and is difficult to interpret over the low-elevation part of the domain, and in order to keep two already composite figures readable within the concise format of a Technical Note, we propose to include Figs. R1 and R2 in a new Supplement, referring to them explicitly in Sects. 5.1 and 5.2 together with a short note on the interpretation of the percentage bias where the reference values are small. We would of course be glad to reconsider this placement should the Reviewer and the Editor prefer these panels in the main text.

SC7: Line 359: Follow-on studies to those cited above (i.e., Liu et al. (2022) and Fang et al. (2023) looked at relationships between SWE errors and meteorological forcing (including ERA5 and ERA5-Land). These might be useful for context and supportive of the findings in this manuscript.

Reply to SC7: We thank Prof. Margulis for these further references. Both works address the same question that motivates our analysis, namely how far discrepancies in SWE estimates from large-scale products can be traced back to their meteorological forcing and how far other factors are involved, and they do so for products that partly overlap with those evaluated here, including ERA5-Land, across mountain regions with contrasting snow climates. Their conclusions are consistent with ours in attributing a substantial part of the differences in simulated SWE to the meteorological forcing, while showing that the forcing alone does not account for all of them, which is the same distinction we draw in Sect. 5. They therefore provide a valuable wider context for the interpretation developed there and for the behaviour of ERA5-Land in particular. We will accordingly add these references in Sect. 5.1 and in the concluding discussion of the transferability of the evaluation framework.

SC8: I would suggest the authors consider whether the Alps vs. Apennines results should be discussed in terms of seasonal vs. ephemeral snow. Liu et al. (2022) partitioned their analysis of HMA snow into seasonal and ephemeral and found varying errors for the different regimes. From the results shown, it appears the Apennines may be more ephemeral and that may explain some of the errors seen.

Reply to SC8: This is a valuable suggestion, which we will adopt in the revised manuscript, as the contrast between the Alpine and mid- and lower-elevation Apennine results indeed can be framed well in terms of snow regime.

We have classified the study domain into seasonal and ephemeral snow. Rather than the peak SWE criterion of Liu et al. (2022), we adopted a snow persistence criterion, which requires no variable or threshold beyond those already used in the manuscript. For each grid cell and each of the 14 snow years, we computed the longest sequence of consecutive days with IT-SNOW SWE at or above the 5 mm threshold already adopted for the SCD indicator. Cells for which this quantity, averaged over the study years, is at least 60 days are classified as seasonal, and the remaining cells as ephemeral, following the threshold used by Petersky and Harpold (2018) and originating from Sturm et al. (1995). The resulting classification is shown in Fig. R3 below.

Figure R3. Seasonal and ephemeral snow regimes over the study domain, classified from IT-SNOW mean annual SCD using a 60-day threshold.

The resulting classification broadly supports the Reviewer's interpretation. Seasonal snow is largely confined to the higher Alpine sectors, whereas ephemeral snow dominates most of the Apennines. In the Apennines, seasonal snow is restricted to the uppermost altitudes, with the most notable examples being the highest massifs of the central sector, which feature substantial areas characterized by a seasonal snow regime. This is consistent with previous descriptions of seasonal but highly variable snow presence in these high-elevation areas (De Bellis et al., 2010; AINEVA, 2023). This classification therefore provides a useful framework for interpreting the contrasting product performance between the Alps and the Apennines. In the revised manuscript, we will introduce the classification and the supporting references, briefly describe the snow conditions of the higher Apennines in Sect. 2, present the regime map and related material in the Supplement, and use this distinction to support the discussion of regional performance differences in Sect. 5.

SC9: Line 393: The fact that the highest resolution (and coupled) model is presumably most computational expensive and yet worst in some error metrics is perhaps worth emphasizing more. Given that meteorological forcings are so important in SWE evolution, but typically relatively coarse, does not guarantee that results will be more accurate at higher resolution. This emphasizes the importance of observations in high-resolution applications.

Reply to SC9: We agree with Prof. Margulis that this point deserves greater emphasis than it currently receives, and we will bring it out more clearly in the revised manuscript.
The results are indeed instructive in this respect. VHR-REA_IT combines the finest grid spacing with a fully coupled atmosphere and land configuration, and is presumably also the most computationally demanding among the analysed products, yet it shows the most widespread underestimation of mean annual SWE and SCD and the weakest daily correlations. Higher resolution and a more elaborate representation of land and atmosphere interactions therefore do not in themselves guarantee more accurate SWE estimates. This is probably due to the fact that, among the characteristics distinguishing the three products, VHR-REA_IT is the only one in which no observational constraint on the meteorological variables is applied within the

regional domain, with observations entering only through the ERA5 lateral boundary conditions. As discussed in Sect. 5.1, we consider this the most plausible explanation for its performance, although the products differ in several respects at once. We will therefore expand the discussion at line 393 and in the concluding remarks accordingly, highlighting that the assimilation of meteorological observations remains essential in high-resolution applications.

SC10: Section 5.2: Can you provide a bit more detail on how the models parameterize snow covered area? Is it a prognostic variable or is it parameterized as a function of snow depth? Is it binary or fractional?

Reply to SC10: We thank Prof. Margulis for this question, which concerns a model feature that the current version of the manuscript does not describe.

In all three products, snow cover fraction is a diagnostic rather than a prognostic variable. It is represented as a fraction of the grid cell rather than as a binary indication of snow presence or absence. The prognostic snow variables include SWE and snow density, from which snow depth is diagnosed through the mass–volume relationship (snow density × snow depth = water density × SWE). Snow cover fraction is then computed as function of the simulated snow amount, expressed either as SWE or as diagnosed snow depth depending on the product, and reaches complete coverage when a product-specific snow amount threshold is reached. The three schemes therefore share the same general structure, although the specific formulation of this diagnostic differs among the products.

To ensure a consistent comparison across the four datasets, we did not use their native snow-cover-fraction fields. Instead, snow cover duration was derived directly from the daily SWE fields by applying the same 5 mm threshold to IT-SNOW and each of the three evaluated products and counting the number of snow-covered days within each snow year. The SCD metric is thus defined consistently across all four datasets and does not depend directly on their native snow-cover-fraction parameterisations.

In the revised manuscript, we will add a brief description of these parameterisations to Sect. 3.2 and clarify in Sects. 4.1 and 5.2 that the analysed snow cover duration is derived exclusively from daily SWE.

SC11: Figure 5: This is a more general comment relevant to some of the other figures too, but Figure 5 provides a useful example. The authors should consider whether using a more discretized colorbar will help communicate the message more clearly. In this case, having a continuous colorbar to demonstrate correlations is not as visually useful as perhaps discretizing the colorbar into a few bins where "poor", "moderate", and "high" correlation would be easier to infer.

Reply to SC11: We appreciate this suggestion and will revise Fig. 5 accordingly, since a discretised colour scale indeed conveys the differences in temporal agreement more clearly than a continuous one.

SC12: Figures A1-B3. Would using consistent colorbars for bias in these plots be easier to interpret or do the authors think using different colorbars for different variables are helpful?

Reply to SC12: Our intention in using a distinct palette for each variable was to make the quantity displayed immediately identifiable across the large number of similarly structured appendix panels. For a given variable, the same palette and the same limits are applied to all products and periods, so the panels remain directly comparable, and all palettes are diverging and centred on zero, so the sign and magnitude of the bias are read in the same way throughout. We would therefore prefer to keep the current colouring, since we believe it helps the reader to recognise immediately that the panels show different variables, while remaining consistent across the multiple figures referring to the different products.

Minor suggested edits/corrections (MS):

MS1: Line 112: Is "SCIA" an acronym? If so, define in first instance.

Reply to MS1: Thank you for catching this. SCIA is indeed an acronym, standing for "Sistema nazionale per la raccolta, elaborazione e diffusione di dati Climatologici di Interesse Ambientale", and we will define it at its first occurrence.

MS2: Line 352: Use "SWE" instead of "snow water equivalent".

Reply to MS2: Corrected. We will also check for any further instances elsewhere in the manuscript.

References

AINEVA: Appennini, resoconto stagione 2022-2023, Associazione Interregionale Neve e Valanghe, available at: https://aineva.it/wp-content/uploads/Pubblicazioni/resoconto2022-23/appennini2022-23.pdf, 2023.

Avanzi, F., Gabellani, S., Delogu, F., Silvestro, F., Pignone, F., Bruno, G., Pulvirenti, L., Squicciarino, G., Fiori, E., Rossi, L., Puca, S., Toniazzo, A., Giordano, P., Falzacappa, M., Ratto, S., Stevenin, H., Cardillo, A., Fioletti, M., Cazzuli, O., Cremonese, E., Morra di Cella, U., and Ferraris, L.: IT-SNOW: a snow reanalysis for Italy blending modeling, in situ data, and satellite observations (2010–2021), Earth Syst. Sci. Data, 15, 639–660, https://doi.org/10.5194/essd-15-639-2023, 2023.

Cortés, G. and Margulis, S.: Impacts of El Niño and La Niña on interannual snow accumulation in the Andes: results from a high-resolution 31 year reanalysis, Geophys. Res. Lett., 44, 6859–6867, https://doi.org/10.1002/2017GL073826, 2017.

De Bellis, A., Pavan, V., and Levizzani, V.: Climatologia e variabilità interannuale della neve sull'Appennino Emiliano-Romagnolo, Quaderno Tecnico ARPA-SIMC n. 19/2010, ARPA Emilia-Romagna, Bologna, 2010.

Fang, Y., Liu, Y., and Margulis, S. A.: A western United States snow reanalysis dataset over the Landsat era from water years 1985 to 2021, Sci. Data, 9, 677, https://doi.org/10.1038/s41597-022-01768-7, 2022.

Fang, Y., Liu, Y., Li, D., Sun, H., and Margulis, S. A.: Spatiotemporal snow water storage uncertainty in the midlatitude American Cordillera, The Cryosphere, 17, 5175–5195, https://doi.org/10.5194/tc-17-5175-2023, 2023.

Liu, Y., Fang, Y., and Margulis, S. A.: Spatiotemporal distribution of seasonal snow water equivalent in High Mountain Asia from an 18-year Landsat–MODIS era snow reanalysis dataset, The Cryosphere, 15, 5261–5280, https://doi.org/10.5194/tc-15-5261-2021, 2021.

Liu, Y., Fang, Y., Li, D., and Margulis, S. A.: How well do global snow products characterize snow storage in High Mountain Asia?, Geophys. Res. Lett., 49, e2022GL100082, https://doi.org/10.1029/2022GL100082, 2022.

Mizukami, N. and Perica, S.: Spatiotemporal characteristics of snowpack density in the mountainous regions of the western United States, J. Hydrometeorol., 9, 1416–1426, https://doi.org/10.1175/2008JHM981.1, 2008.

Parajka, J., Bezak, N., Burkhart, J., Hauksson, B., Holko, L., Hundecha, Y., Jenicek, M., Krajčí, P., Mangini, W., Molnar, P., Riboust, P., Rizzi, J., Sensoy, A., Thirel, G., and Viglione, A.: MODIS snowline elevation changes during snowmelt runoff events in Europe, J. Hydrol. Hydromech., 67, 101–109, https://doi.org/10.2478/johh-2018-0011, 2019.

Petersky, R. and Harpold, A.: Now you see it, now you don't: a case study of ephemeral snowpacks and soil moisture response in the Great Basin, USA, Hydrol. Earth Syst. Sci., 22, 4891–4906, https://doi.org/10.5194/hess-22-4891-2018, 2018.

Sturm, M., Holmgren, J., and Liston, G. E.: A seasonal snow cover classification system for local to global applications, J. Climate, 8, 1261–1283, https://doi.org/10.1175/1520-0442(1995)008<1261:ASSCCS>2.0.CO;2, 1995.