Answers to reviewers: TC-2021-281

Temporal stability of long-term satellite and reanalysis products to monitor snow cover trends

Ruben Urraca and Nadine Gobron

REFEREE #1 – Chris Derksen

This study uses a reference dataset of point snow depth measurements to assess the performance and stability of snow extent and snow cover duration from reanalysis and satellite-derived products. This is important to quantify because changes to the quality and quantity of satellite data and the data sources assimilated into reanalysis can introduce spurious trends and temporal discontinuities into multi-decadal time series. The analysis is focused on ERA5 and the NOAA snow chart climate data record (NOAA-CDR), which are two widely used datasets that provide snow information back to the 1960s. Overall, I found the analysis to be comprehensive in scope, sound in the overall approach, and clearly explained.

I have a number of both major and minor comments, mostly in an effort to further clarify the methods and tighten the messaging. This was a really enjoyable paper to review, thanks to the authors for their efforts.

MAJOR COMMENTS

Lines 61-68: Some additional context/examples could be provided in this paragraph.

First: "The transition between different sensors (e.g., JAXA GHRM5) or increasing the number of satellite sources used (e.g., IMS, NOAA CDR)…" It may not be clear to some readers that the IMS product is actually manually derived by analysts from multiple sources of satellite imagery (as opposed to an objective retrieval like the JAXA product). This is noted later on line 88, but this could be mentioned in this introductory paragraph.

Second: The ESA GlobSnow and Snow CCI products are derived from the passive microwave satellite record, which is composed of SMMR + SSM/I + SSMIS data, which is another example of how discontinuities can be introduced through changing instruments during the satellite era. (Incidentally, we have found there are differences in the validation statistics for Snow CCI SWE performance related to the different passive microwave sensors. This work is under review, but it would be interesting to also include the Snow CCI dataset in the analysis you present in this work.)

Answer: Thanks for the comments. We have rephrased the paragraph including the reviewer's suggestions:

"The temporal coverage of satellite products is limited by the satellite/sensor used, so different satellite instruments are combined to produce Climate Data Records (CDRs). For instance, JAXA GHRM5 combines optical data from NOAA's AVHRR and MODIS sensors, whereas both ESA GlobSnow and ESA snow CCI SWE combine passive microwave data from SMMR, SSM/I and SSMIS sensors. The transition periods between different sensors are the main source of instability in these products, but stability issues can also arise due to sensor degradation and orbital drifts (e.g., AVHRR data). The increasing number of satellite sources can also alter the stability of products derived manually by analysts from multiple sources of satellite imagery (e.g., IMS and NOAA CDR)."

Figure 1: It's unfortunate no data from Canada were used in this study (particularly in the context of the trend analysis in Figure 9, which gives the impress of negative trends in the Eurasian sector and no trends over Canada, which is not the case). There is an updated snow depth dataset for Canada described here: Brown, R., C. Smith, C. Derksen, and L. Mudryk. 2021. Canadian in situ snow cover trends 1955-2017 including an assessment of the impact of automation. Atmosphere-Ocean. DOI: 10.1080/07055900.2021.1911781. For future reference, the Canadian Historical Daily Snow Depth Database should soon be available here (or contact the authors of the above paper): https://catalogue.ec.gc.ca/geonetwork/srv/eng/catalog.search#/metadata/63dca4bb-a29a-43b0-828b-7eccb03de456

Answer: Many thanks for providing us the link to the Canadian snow cover dataset. We agree that Canada was the main spatial gap in our study, particularly for trend analysis. We have processed all the stations available in the Canadian Historical Daily Snow Depth Database. Out of them, 57 passed our selection criteria for trend/stability analysis, and 34 were classified as spatially representative for the stability assessment.

As mentioned by Brown et al 2021, the number of Canadian stations significantly decreases before 1955 and after 2010. If we kept our original study period for the trend analysis (1950-2020), the number of Canadian stations available drops below 10. Therefore, to cover most Canadian regions, we have reduced the study period for the trend analysis from 1950-2020 to 1955-2015.

Section 2.2: How did you ensure that the snow depth observations retained for analysis were not assimilated into ERA5? This issue must be addressed specifically in the text to ensure independence between the reanalysis and validation datasets.

Answer: C3S/ECMWF currently does not provide the list of snow stations assimilated by ERA5. This information could be included in future updates of ERA5. ECMWF specifies that ERA5 assimilates SYNOP stations, and some RHIMI, GHCN and ECCC stations may report to the SYNOP network as well. Thus, some stations used for the validation could be currently assimilated by ERA5.

We acknowledged this issue in our original submission, but we have extended the discussion on the potential limitations of our study in the revised version.

"ERA5 assimilates 3507 snow depth observation from SYNOP, but the exact list of stations assimilated is not yet available (ECMWF, personal communication). It is likely that some RHIMI, GHCN, and ECCC stations were assimilated by ERA5, since some stations of these networks also report to SYNOP. This could compromise the independence of our validation set, particularly for the spatial accuracy analysis of ERA5 between 2005-2015 (Section 3.3). Particularly, the magnitude of the bias and RMSE will be artificially reduced at those stations assimilated by ERA5. However, the impact of this issue in stability analysis, which is the main goal of the study, should be smaller. The step changes and trends observed in the ERA5 bias do exist and are due to changes in the ERA5 model, independently of whether some stations are assimilated or not by ERA5. It is true that the assimilation of the stations could explain some of the step changes observed in ERA5 (e.g., RHIMI 1991). In this case, discontinuities will be higher at those stations assimilated by ERA5, but they will also appear at stations not being assimilated."

Section 2.4: Very interesting comparison with the analysis of Hori et al (2017). I'm not fully clear on how the SCF surrounding the station was determined: "In this study, we used the SCF in the surroundings of the station measured at RIHMI stations to analyze the correlation between SD at the station and the surrounding SCF (Fig. 2)." Was IMS data used to determine SCF? What distance was used around the station (IMS pixels contained by coarser ERA5/ERA5-Land pixels as described in Section 2.3?)?

Answer: The snow cover fraction around the station is visually assessed at RIHMI stations. The exact definition of these measurements is given by Bulygina et al 2011:

"The snow cover extent over the near station territory and the snow cover characteristics are visually determined at morning observations. The amount of snow covering the visible area around a meteorological station is estimated on a scale of one to ten (10–100%; or zero in the absence of snow)."

We have extended the description of the SCF around the station in the methods section, to clarify that it is a visual measurement made at the stations:

"The SCF in the surrounding of the stations is visually assessed at RIHMI stations (Bulygina et al 2011). We used these measurements to analyze the correlation between SD at the station and the surrounding SCF (Fig.2)."

Section 2.4.1: It is noted that "Stability was evaluated by analyzing how the annual bias in both SD and SCD changed temporally." and that stability was analyzed separately for the RIHMI and GHCN networks. But how were step changes statistically determined (the vertical lines in Figures 4 and 5)? Line 198: Why was the interval of four years selected to compare the bias difference before and after a step discontinuity? Was there any testing performed to confirm that this was some sort of ideal number?

Answer: Some vertical lines (years of step changes) are determined by significant changes in the product algorithm: 1979 corresponds to the transition between ERA5 and ERA5 backward extension, 2004 corresponds to the addition of IMS snow data to ERA5 model. The exact year of other discontinuities (e.g., ERA5 1991) was determined with a window function that calculated how the magnitude of the step changes when varying the step year from 1950 to 2020 by intervals of 1 (Fig A1). The relative maximum or minimum was selected as the discontinuity year.

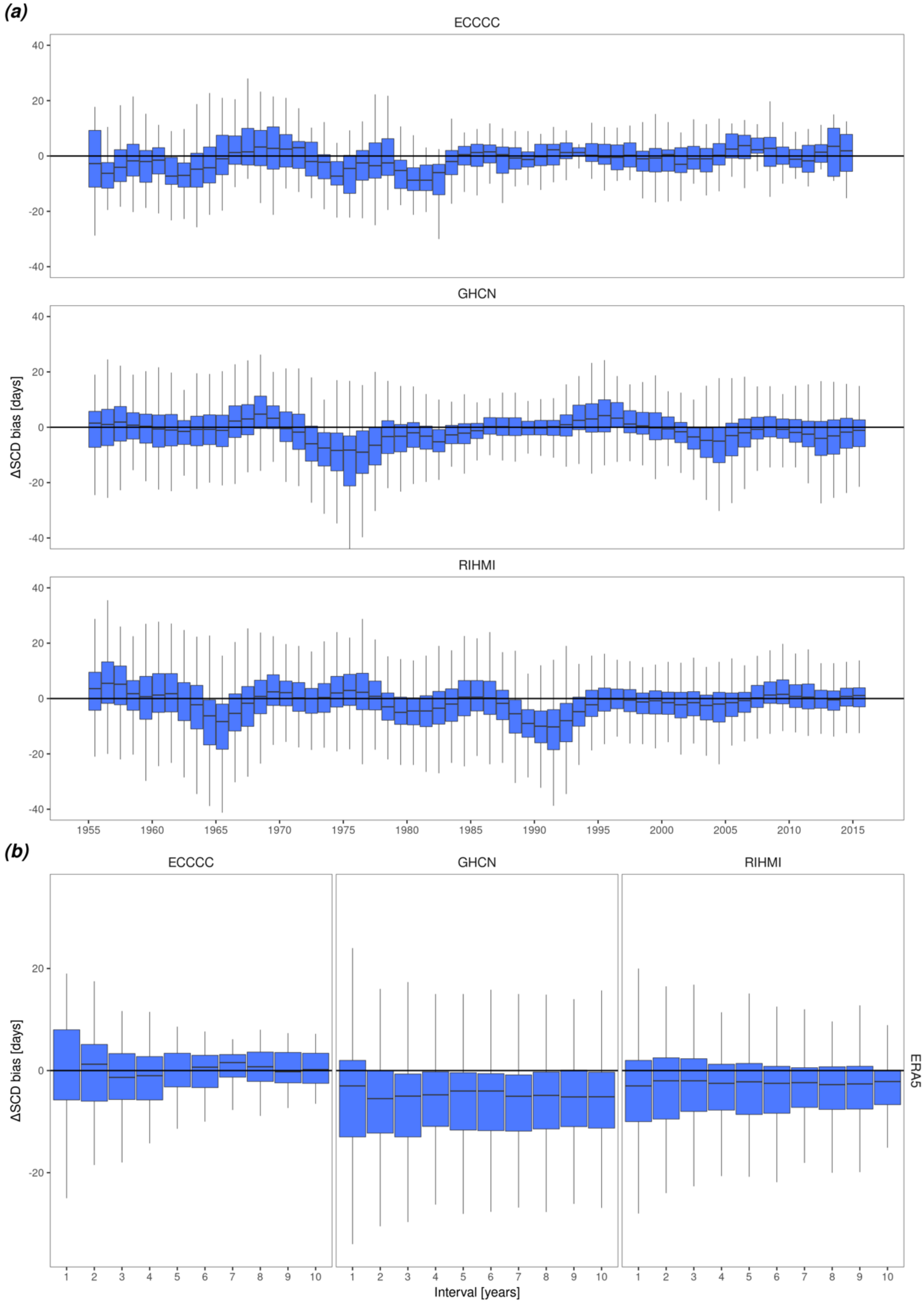

Fig A1 Sensitivity analysis to determine the exact year of step discontinuities **(a)** and the interval used to estimate the magnitude of the discontinuity **(b). (a)** Change in ERA5 SCD Δbias (before – after) when the step year varies from 1955 to 2015. **(b)** Change in ERA5 SCD Δbias (before – after) during the 2004 discontinuity when the number of years used for its calculation (interval) is changed from 1 to 10.

The interval of 4 years used to compare the bias after and before the step was a compromise between two effects.

- The interval should be long enough to remove the effects of inter-annual snow cover variability on Δbias.
- The interval should be as short as possible to remove the effect of underlying trends in the bias on Δbias.

We evaluated the magnitude of both effects with a sensitivity analysis, measuring how the magnitude of Δbias varied with an increasing interval, from 1 to 10 years by 1-year intervals. The results show that Δbias variability stabilizes after 4-5 years. Therefore, we used an interval of 4 years

We have added this figure as supplementary material, and we have included the following explanation in the methods section:

"The interval of four years was chosen based on a sensitivity analysis (Fig A1). This interval needs to be long enough to remove the effects on inter-annual variations of the snow cover, but too long intervals may be affected by the underlying trends in the bias. Therefore, the shortest interval after ∆bias has stabilized was chosen."

Section 2.5: What is the justification for including the stations which failed the spatial representative test in the trend analysis?

Answer: The main goal of the spatial representativeness test is to reduce the uncertainty of the point-to-pixel comparison, discarding stations in which point measurements are not representative of the larger surrounding region covered by the reanalysis pixel. However, stations discarded for the point-to-pixel validation are still valid for the trend analysis.

For instance, most of the stations classified as low representative are in coastal regions (Fig 3). However, snow cover trends at these coastal stations are still meaningful for the trend analysis. Indeed, some of these coastal regions such as Eastern USA and Eastern Canada are those experiencing a larger snow cover retreat. Besides, trends observed at coastal stations are coherent with those observed at inland locations. Thus, we believe that keeping these stations provides additional valuable information to trend analysis, without interfering with the representatives of this analysis.

Section 3.1: I appreciate the effort taken to quantify the spatial representativeness of the point measurements. This is a long standing problem in the validation of gridded snow products at variable resolutions, which is usually acknowledged but not addressed analytically. So these results are very interesting…

Answer: Many thanks for the comment.

Line 291: "This suggests that the H-TESSEL land model used in both ERA5 and ERA5-Land tends to systematically overestimate SD, most likely due to an excessive snowfall, when no data is assimilated (ERA5 before 1979, ERA5 above 1500 m, ERA5-Land)." I find the messaging in this sentence to be confusing. If the overestimation is related to H-TESSEL, this implies that uncertainty

in snow parameterizations in the model lead to overestimation of snow depth, but then the problem of excessive snowfall is mentioned. Does this not imply that precipitation bias is the source of the positive SD bias as opposed to the land model?

Answer: We have rephrased this sentence to clarify that the most likely cause of the SD bias is a precipitation bias, as suggested by Orsolini et al 2019:

"As suggested by Orsolini et al 2019, the most likely cause of the snow depth overestimation in both ERA5 and ERA5-Land could be a precipitation bias, which is only corrected by the assimilation of snow depth observations in ERA5 (after 1979 and below 1500 m)."

Section 3.2.2: The bias trend in the NOAA CDR in fall is an important finding, and corroborates previous work which found similar issues with this product in this season. This is important because numerous studies continue to cite a positive trend in October snow extent over Eurasia, despite increasing multiple lines of evidence (this study provides a new line of evidence) which outline inhomogeneity in the NOAA CDR. I found lines 335-340 to be somewhat confusing, and suggest this text be edited for clarity. The study of Mudryk et al (2017) could also be considered, which showed (1) the NOAA CDR trends in October and November are non-physical and not consistent with other datasets, and (2) NOAA CDR trends in spring are stronger than other datasets. (Mudryk, L., P. Kushner, C. Derksen, and C. Thackeray. 2017. Snow cover response to temperature in observational and climate model ensembles. Geophysical Research Letters. 44, doi:10.1002/2016GL071789.)

Answer: Tanks, we have rephrased the paragraph adding the new references:

"Brown and Derksen (2013) suggested that the opposite effect during the spring season could be expected but was not observed. Theoretically, an improved detection of snow melting could lead to a stronger spring trend, introducing an artificial negative trend in the CDR. In this line, Derksen (2014) reported a tendency of NOAA CDR to map less snow in spring since 2007 than the multi-dataset composed by NOAA CDR, MERRA and ERA-Interim. Mudryk et al. (2017) also found that NOAA CDR has a stronger spring trend than that of other datasets. We analyzed this issue by evaluating the snow cover duration trends in spring. Negative trends in spring bias only appear at some Russian stations (Fig.7a). However, the number of stations showing significant trends in spring is smaller, and the magnitude of these trends is much lower than those in fall and winter. Despite this issue could exist in some specific regions, the impact at global scale is negligible (Fig.A8)."

Section 3.3: "Both ERA5 and ERA5-land use a threshold (5 cm) larger than the one applied to the stations (2.5 cm)…" In reading Section 2.4, I was wondering about the impact of these different thresholds on the validation analysis. I understand the decision to apply 2.5 cm to the snow depth measurements because this is supported by Figure 2 and is consistent with Hori et al (2017). But calculating bias with slightly different thresholds to convert SD to SCD seems problematic. Can you report on any sensitivity analysis which determines how the bias calculations are related to the choice of threshold as applied to the snow depth measurements?

Answer: Thanks for the suggestion. We have included the sensitivity analysis requested by the reviewer in the results section. We evaluate the changes in snow cover duration bias when changing the station threshold from 0 to 10 by intervals of 2.5 cm.

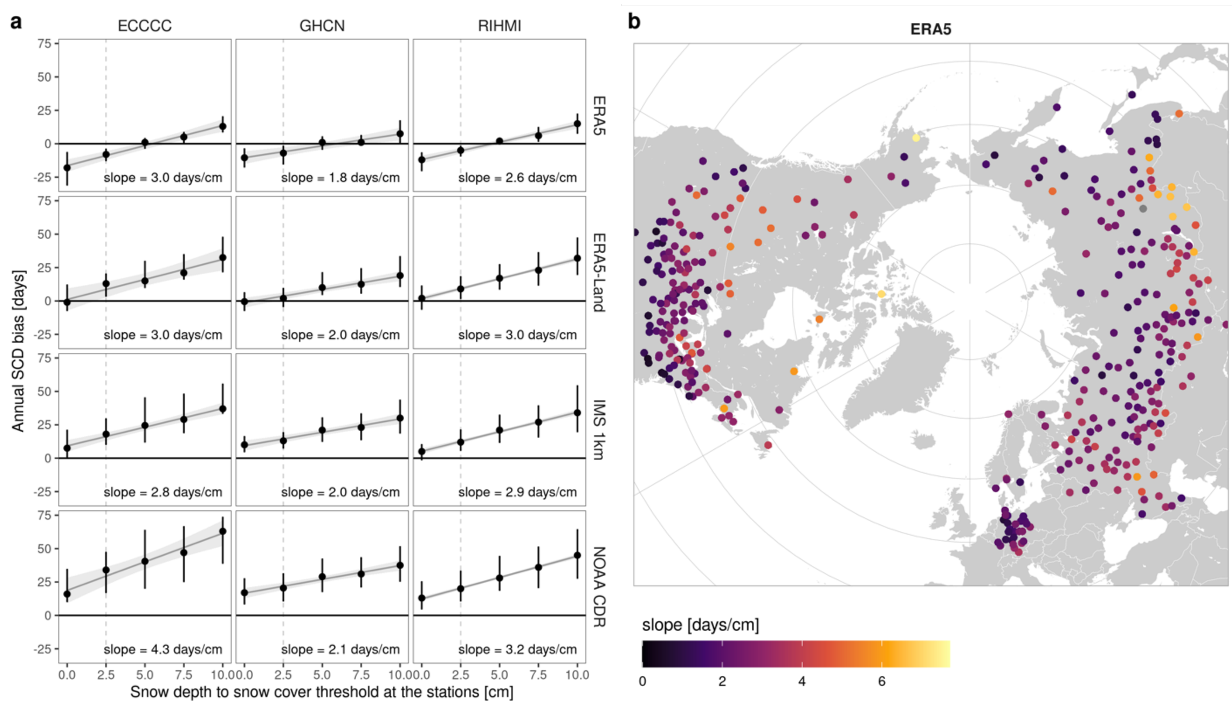

Figure A2. Sensitivity of the snow cover duration (SCD) bias on the snow depth to snow cover threshold used at the stations. (a) Variation of the SCD bias (median ± interquartile range) per product and network when changing the threshold from 0 to 10 cm, by intervals of 2.5cm. (b) Spatial analysis of the rate of change [days/cm] for ERA5. Both figures are derived with data from 2015.

We have discussed the results of the sensitivity analysis in the Results section:

"*In SCD, ERA5 presents a constant underestimation (IQR) of around [-9.4, -5.5 days] while ERA5-Land keeps overestimating [2.4, 11.2 days]. As above mentioned, the SCD bias strongly depends on the threshold used to convert SD to SC. Both ERA5 and ERA5-land use a threshold (5 cm) larger than the one applied to the stations (2.5 cm). This could explain why ERA5 has a negative SCD bias despite having an unbiased snow depth. Indeed, when the ERA5 threshold is applied to the stations (Fig. A2), ERA5 SCD bias is close to zero in the three networks. We could be tempted to use the same threshold in stations and product. However, the thresholds applied by products need to be validated as well, and we can only do it deriving independent thresholds for the station measurements. In this study, we have used visual snow cover measurements made at RHIMI stations, but other data sources such as high-resolution satellite imagery could also be useful.*

We investigated further this issue with a sensitivity analysis that evaluates how the SCD bias changes with different values of snow depth to snow cover threshold during 2015 (Fig.A2). The magnitude of SCD bias is similar between networks, indicating a good consistency between their measuring methods. However, the magnitude of SCD bias strongly varies between products. When a threshold of 2.5 cm is used, the mean SCD bias varies as follows: 24.8 days (NOAA CDR), 14.3 days (IMS), 8.0 days (ERA5-Land) and -6.7 days (ERA5). These differences are the result of the different thresholds applied by the products, as well as their different snow depth biases (in case of reanalysis). Orsolini et al. (2019) already pointed out that the different thresholds applied by reanalysis datasets was one of the main limitations for inter-comparing them.

The sensitivity analysis also shows that changing 1 cm the station threshold leads to changes in the annual SCD bias of around 2-3 days. These changes are constant between products but vary between networks (ECCC = 2.8-4.3 days/cm, GHCN = 1.8-2.1 days/cm, RIHMI = 2.6-3.2 days/cm), due to the different snow *conditions in each station. Stations with more daily SD values close to the threshold are more affected by changes in the threshold."*

Section 3.5: I suggest moving this into Section 4, because it is largely discussion material and does not present new analysis.

Answer: We have moved it into a new Section 4, but we have kept the Conclusions in a separate section (new Section 5).

Conclusions: The key result with respect to ERA5 is clearly stated on line 460: "In the reanalysis, data assimilation creates a trade-off between accuracy and stability." For applications like NWP, the instantaneous best estimate is the highest priority, but this of course does not ensure the temporal consistency required for climate monitoring. The key result for the NOAA CDR is communicated less clearly: "Overall, most of the trends/discontinuities observed are larger than the actual snow trends and the GCOS stability requirements, making these products inappropriate for climate applications without correction, particularly ERA5." I suggest re-phrasing this to provide an assessment more clearly focused on the NOAA CDR. This study provides a new line of evidence that autumn trends are very problematic in this dataset, but there are seasons and regions in which the product is suitable for climate analysis (e.g. MAM as shown in Figure 10b).

Answer: We have rephrased NOAA CDR conclusions as follows:

"NOAA CDR presents a positive artificial trend in SON and DJF. These results provide another line of evidence supporting the problematic fall trends in NOAA CDR and reveal that a similar trend appears in Europe of eastern North America during winter. Despite the numerous studies highlighting the inconsistency of NOAA CDR fall trends with in-situ measurements and with other datasets, some studies keep claiming a positive snow cover trend in fall based solely on NOAA CDR data (Cohen et al., 2021). Using NOAA CDR without correction in SON and DJF should be avoided. NOAA CDR could still be valid after correction, or in other regions and seasons (e.g., MAM) not affected by artificial trends"

Section 4 could also highlight that studies continue to claim there is a positive trend in autumn snow extent based solely on the NOAA CDR (https://doi.org/10.1126/science.abi9167) and do not acknowledge the literature which has identified problems with this dataset, so your study once again points out that this dataset is problematic in the autumn.

Answer: We have rephrased NOAA CDR conclusions as shown in the previous comment.

MINOR COMMENTS

Line 18: change 'snow cover decrease is aggravated' to 'snow cover decrease is coincident to decreasing snow depth…'

Answer: Changed.

Line 30: not clear what is meant by 'global circulation'.

Answer: Global atmospheric circulation. We have clarified in the text.

Paragraph 1 of the Introduction: The Stocker et al (2013) reference for snow trends and snow-albedo feedback is a little out of date. Updated SAF estimates are in the IPCC SROCC Chapter 3, and the Thackeray et al (2019) paper provides a fairly current review. (Meredith, M., M. Sommerkorn, S. Cassotta, C. Derksen, A. Ekaykin, A. Hollowed, G. Kofinas, A. Mackintosh, J. Melbourne-Thomas, M.M.C. Muelbert, G. Ottersen, H. Pritchard, and E.A.G. Schuur, 2019: Polar Regions. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate [H.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer (eds.)].)

Answer: We have included the updated values provided in IPCC SROCC Chapter 3.

Line 33: suggest changing to '…such as the Arctic and high elevations.'

Answer: Changed.

Line 33: "Notably, only 11 long-term stations are available in the Southern Hemisphere (SH)." Very interesting! Is there a reference for this statement?

Answer: The statement was extracted from IPCC AR5:

"Measurement challenges are particularly acute in the Southern Hemisphere (SH), where only about 11 long-duration in situ records continue to recent times: seven in the central Andes and four in southeast Australia."

We have included the corresponding reference. Nevertheless, we have relaxed the sentence as follows:

"Long-term snow measurements are particularly limited in the Southern Hemisphere (SH) (Stocker et al., 2013)."

Line 46: Is there a reference for the S-NPP VIIRS dataset, as is provided for the others in this list?

Answer: We have added the reference to the product user manual.

Line 50: This is a very minor point, but the most recent citation for the GlobSnow dataset (v3) is: Luojus, K., J. Pulliainen, M. Takala, J. Lemmetyinen, C. Mortimer, C. Derksen, L. Mudryk, M. Moisander, P. Venäläinen, M. Hiltunen, J. Ikonen, T. Smolander, J. Cohen, M. Salminen, K. Veijola, and J. Norberg. 2021. GlobSnow v3.0 Northern Hemisphere snow water equivalent dataset. Scientific Data. doi: 10.1038/s41597-021-00939-2.

Answer: Updated.

Line 87: "Since 2004, ERA5 also assimilates the IMS product but only over altitudes below 1500 m." Could add a reference to the Orsolini et al (2017) paper here.

Answer: We have added the reference.

Line 101: "…but snow observations are not directly assimilated." This is a small point but make clear that both the in situ snow depth and the IMS data are not assimilated into ERA5-land.

Answer: We have clarified it as follows:

"Neither in-situ snow depth measurements nor IMS data are directly assimilated by ERA5-Land."

Line 114: Some older citations could be added to provide readers with more background on the CDR and IMS: Robinson, D., K. Dewey, and R. Heim. 1993. Global snow cover monitoring: an update. Bulletin of the American Meteorological Society. 74(9): 1689-1696. Helfrich, S., D. McNamara, B. Ramsay, T. Baldwin, and T. Kasheta. 2007. Enhancements to, and forthcoming developments in the Interactive Multisensor Snow and Ice Mapping System (IMS). Hydrological Processes. 21: 1576-1586.

Answer: We have added both references.

Line 120: remove 'around'

Answer: Removed.

Line 141: typo 'sires'

Answer: Corrected.

Line 223: "…stations are located either on peaks (Fig. 3b) or in the valley…" This wording is quite specific. Perhaps just emphasize that elevation gradients around the stations create uncertainty?

Answer: We have rephrased it as follows:

"On mountainous regions, the spatial representativeness of the stations decreases due to the large elevation gradients (Fig.3b)"

Line 267: I would not refer to the NOAA CDR as having a "retrieval algorithm" since it is analyst-derived. How about "The positive bias is explained by changes in the analysis approach to produce the snow charts, which since 1999…"

Answer: We have rephrased it as follows:

"The positive bias could be explained by changes in the analysis approach to produce the snow charts, which since 1999 considers a pixel snow-covered when only a 42% of the IMS pixels within the pixel were snow-covered."

Line 270: can you provide a reference to the NOAA CDR product manual?

Answer: Yes, we have added the corresponding reference.

Line 273: "…but a positive trend is observed since 1990 in fall and winter." Add a reference to Figure A2 here.

Answer: Done.

Line 285: The study of Mortimer et al (2020) focuses on the ERA5 discontinuity in 2004, not 1980. (please double check the other citations)

Answer: The reviewer is right. Indeed, Mortimer et al (2020) reference was included in the discussion of the 2004 discontinuity.

Here we just wanted to state that ERA5 tends to have a positive bias in regions and periods when it does not assimilate station data (as reported by Mortimer et al (2020) above 1500 m). However, we have removed the reference from this section to avoid any misunderstanding.

Line 324:Instead of Derksen, 2014, could cite Brown and Derksen (2013) here.

Answer: We have changed the reference.

Line 327: change 'algorithm' to 'analysts'

Answer: Done.

Line 392: "In regions such as Europe, spring SCD reductions add up to the decreasing SD, increasing, even more, the annual SCD trends." Awkward wording. I think the point is that in Europe both SD and SCD are decreasing, with the trend towards shallow snow depth amplifying the shorter SCD. In Russia, the snow cover season is shortening, despite positive SD trends in some areas, which means the spring melt signal driven by warming temperatures overrides any increase in snow accumulation during the winter.

Answer: We have rephrased as follows:

"In regions such as Europe, both SD and SCD are decreasing, with the trend towards shallow snow depth amplifying the shorter snow season. In Russia, spring SCD is also decreasing despite the positive trends in SD. This means that the spring melt driven by warming temperatures overrides any increase in snow accumulation during winter."

Line 453: I very much appreciate the comment that while multi-product ensembles are preferred for historical trend analysis, it is still important to quantify the performance of individual products over time.

Answer: Thanks for your comment.