**Anonymous Referee #2**

**Which global reanalysis dataset represents better in snow cover on the Tibetan Plateau?**

This paper evaluates snow cover fraction (SCF) of eight reanalysis datasets over the Tibetan Plateau during 2001–2020. The SPIReS EO product is used as the reference dataset and products are evaluated based on their bias, Taylor skill score and its component metrics, and trends by means of a consistency index. The impact of biases in temperature and precipitation on forcing is assessed through comparisons of the reanalysis forcing with a reference dataset (TPMFD). An interesting analysis of SCF parameterization is also provided as are tests to identify the optimal combination of SCF products to include in an ensemble. The manuscript tries to conclude on the impact of snow data assimilation in SCF product performance.

A comparison of SCF products over the TP is an important endeavor. The manuscript is promising and contains useful and interesting analysis. However, I find the authors struggle to present their results in a way that builds logically to their stated conclusions, specifically as it relates to the role of snow data assimilation. The authors have a tendency to make general statements that do not flow logically from the evidence presented in the text. And at times the evidence presented contradicts these general statements.

R: Thank you very much for positive comments, which will encourage us to conduct more in-depth research in the future. Additionally, the comments are highly valuable as they help significantly improve the quality of our manuscript. We have re-summarized the conclusions based on the manuscript content and made extensive revisions to descriptions related to snow data assimilation. Furthermore, we have refined the manuscript presentation to avoid inconsistencies. We will provide detailed responses to the specific modifications in the following.

The methods are not adequately described (see ‘primary concerns - Methods’). The authors struggle to interpret and communicate information contained in the Taylor plots. This may be related to the lack of clarity in how the metrics in the plots were calculated and what they represent. The text, particularly the results, could be better organized and ideas clearly separated. I encourage the authors to make better use of paragraphs. Sometimes, the references don't support the statement as written -i.e. either not the appropriate reference or summary of the study contains inaccuracies. This may be due language issues.

R: We are very inspired by your suggestions for methods. In the revised manuscript, we have organized all methods into Section 2, unlike before where they were scattered across multiple sections. Additionally, we have added more explanations of SCF evaluation methods to make it easier understand for readers. We have corrected references errors in the manuscript, which also guides us to be more rigorous in our future research work. Here, we would like to note that all results in the manuscript have been reprocessed according to the comments of you and other two anonymous referees, with a unified time range of the Water Years 2001–2017 and consistent spatial resolution of 0.5°×0.5°. Descriptions of the results have also been adjusted accordingly. The followings are our point-by-point responses to the comments. Our responses start with "R:".

**Primary concerns**

I am unconvinced by the statement that the reanalysis datasets incorporating snow assimilation always outperform those without snow assimilation [lines 348-351 and elsewhere]. The results, as presented, do not support this firm conclusion but rather suggest a more nuanced finding. In terms of skill scores, the top two products use snow data assimilation. However, JRA55 which also assimilates snow information has a comparable skill score to ERA5, ERA5L and CRAL which do not assimilate snow information. On the other hand, GLDAS which does not assimilate snow information has comparable skill scores to HMASR and CFSR. These findings are presented in the text and seem to contradict the general statement. I suggest the authors provide a more nuanced discussion and revisit how the ideas flow and build towards a specific conclusion.

R: We agree with your viewpoint. There were some inappropriate statements regarding the impact of snow data assimilation in the original manuscript and have reworked the description. Since we cannot quantify the impact of snow data assimilation on the SCF simulation accuracy in reanalysis datasets, we can only attempt to understand its influence through discussion. Your constructive suggestion inspired us to analyze the impact of the snow data assimilation from the perspective of assimilating snow variables in the reanalysis datasets. In L575–588 of change-tracked manuscript, we have discussed in terms of the effect of whether or not the reanalysis dataset directly assimilates SCF variables on its SCF spatial simulation performance. Through a comparison of the SCF annual trends between HMASR and MERRA2, we discuss the influence of data assimilation on the annual trends, as stated in L738 of change-tracked manuscript.

For example, later the authors tie the poor performance of JRA55 to SCF parameterization. Is it also possible that assimilation of poor snow data (JRA55) can degrade performance?

R: We together analyze the JRA55 and CFSR assimilated with SD data in L582 of change-tracked manuscript that "As for JRA55 and CFSR, which although assimilated SD data and have been found good simulations in SD and SWE (Bian et al., 2019; Orsolini et al., 2019), the process of transforming SD to SCF through model parameterization introduced additional errors, thereby leading SD assimilation to only a limited effect on the accuracy of SCF simulations. Compared to JRA55, the SCF parameterization method employed in CFSR is more reasonable, resulting in spatial simulation performance batter than JRA55 by a considerable margin. This indirectly illustrates the impact of parameterization methods on the SCF simulation in JRA55.".

Does the strong performance of GLDAS indicate that performance comparable to the best products that assimilate snow data can be achieved without snow data assimilation?

R: In response to your suggestions in this manuscript regarding the assimilation impact, we have adjusted the perspective for discussing the effect of snow data assimilation on SCF bias in the reanalysis dataset in L575–588 and L738–750 of change-tracked manuscript.

Further, the above results pertain to SCF accuracy and not to the representation of trends where ERA5 and ERA5L (no snow data assimilation) performed best.

R: This conflicting issue you point out was caused by inappropriate generalizations regarding the impact of data assimilation in the original manuscript. We have removed these generalized descriptions. Furthermore, we have provided a more detailed discussion on the optimization effect of snow data assimilation on SCF in reanalysis datasets in terms of both the spatial distribution and the annual trend. This ensures better consistency between the manuscript sections. It is worth noting here that the relatively better annual trend in SCF for ERA5 and ERA5L is attributed to the better simulation of meteorological forcings.

**Methods**

The methods, specifically the statistical approaches, need to be clarified to adequately review the results presented in the manuscript.

R: Thank you for suggestion. We have added many details in the methodology section.

Did you use anomaly fields or raw SCF values in deriving your Taylor plots? Please clarify.

R: After unifying the resolution, all datasets included 1200 pixel points in the TP. These three metrics in the Taylor diagrams of Fig. 2b, including R, RESM, and STDR, were computed for 1200 pixel points within the TP between reanalysis datasets and SPIReS after averaging the SCF climatology from WY 2001 to WY 2017. These details have added in Section 2.3.1.

'Spatial correlation' – was this calculated using all times and locations which would be more of a bulk correlation rather than a spatial or pattern correlation. Please clarify. Using the full time series is fine, you just need to keep this in mind when interpreting the results and remove 'spatial' from talk of correlation.

R: In the original manuscript, "spatial correlation" refers to the correlation coefficient (R value) in the Taylor diagram. Its calculation method, as described above, has been added into the new manuscript. Additionally, we have removed the word "spatial" when describing the results as suggestion.

Standard deviation ratio – unclear what you mean by 'degree of similarity in dispersion patterns'. Note that if you used the full time series then the variability captured by the STDR would be related to the variability in both time and space. Again, it is perfectly fine to use the full 20–year anomaly time series but the terminology and definitions need to be adjusted to reflect what was calculated.

R: To facilitate comparison across reanalysis datasets in one Taylor diagram, we normalized the Standard Deviation (STD) to obtain STDR. The STD for all datasets was computed across the spatial dimension after averaging over the time dimension. STD is an indicator of the dispersion of a set of data. Therefore, STDR represents the consistency in the SCF climatology spatial field values between the reanalysis datasets and SPIReS (Cui et al., 2021). A description related to STDR has been added in L323 of change-tracked manuscript.

Depending how the above metrics were calculated, the wording and interpretations in the results section may need to be revised to accurately reflect the methods.

R: We revised and expanded the method description according to your suggestions, as well as adjusted the wording and interpretation in the results section.

Additionally, approaches to investigate the impact of meteorological forcings and SCF parameterization should be outlined in the methods section. Currently Methods is limited to presenting the metrics used to compare a product with a reference and does not lay out how the authors plan to meet the objectives stated in the introduction. For example, ~L372-375 you mention that you investigate SCF bias by looking at snowfall, temperature, and parameterization thresholds but you do not clearly lay out how you conducted this investigation. Instead, it jumps right into results. This would be ok if the methods were outlined in the methods section but they are not. What methods were used? Some of what is in the results could be moved to the methods.

R: Thank you very much for suggestions. Organizing all the methods used in the manuscript in the Methods section will reduce reader's time and improve their reading experience. We have added a subsection to the Methods section specifically to introduce the analysis methods of SCF bias sources and the methods for generating the optimal combined dataset. The methods for analyzing SCF bias sources include investigating the impact of meteorological forcing and SCF parameterization. Please refer to Section 2.3.2 for the specific additional content.

For precipitation forcing, did you use snowfall or total precipitation? Please clarify.

R: For meteorological forcing, we utilized snowfall rather than total precipitation. We have provided a clarification at L356 in change-tracked manuscript. Additionally, we added information regarding the conversion process from total precipitation to snowfall in the meteorological forcing reference dataset (TPMFD) in Section 2.1.3.

**Additional comments**

Suggest separating analysis of impact of meteoritical forcing and SCF parameterization.

R: Thanks for suggestion. We revised the paragraphs to ensure that each paragraph in the manuscript focuses on a single aspect of the content.

**L286-288:** How did you select these two products as being the 'worst'? If the comparison is relative to SPIRES which has a TP SCF average of 0.13 then MERRA2 at 0.05 is only 0.08 away from SPIRES whereas ERA5 (0.41), ERA5L and JRA55 all appear to have larger absolute biases (with respect to SPIRES reference) compared to MERRA2. I can't judge from the plot where CFSR falls. I think what you are pointing out is the product with the largest positive bias (ERA5) and the product with the largest negative bias (MERRA2).

R: We agree with your viewpoint. We have revised the sentence as suggestion with "In comparison to SPIReS, ERA5 stands out as the dataset showing the highest positive bias, while MERRA2 demonstrates the largest negative bias, with extreme TP average values of 0.41 and 0.05, respectively (Fig. 1b)."

Include mention of the STDR of the various products, not just the correlation on the Taylor plots. The skill score is more than just the correlation, there is some mention of this on Line 300.

R: We have added the STDR values description as suggestion.

The linkages between the maps (Figures 1a and 2a) and absolute bias to the STDR is tricky because it's not clear how it was calculated. i.e. does the STDR reflect the variability in both time and space or just space?

R: According to the description of STDR calculation methodology, we think STDR reflects the SCF spatial variability. Moreover, after supplementing, the relationship between the spatial distribution of SCF climatology and climatological biases of reanalysis datasets (Fig. 1a and 2a) and STDR seems clearer. In Fig. 1a, the ability of reanalysis datasets to capture high and low values of SCF climatological spatial fields reflects their own STD situation. This information can also be obtained from the climatological bias distribution in Fig. 2a. Therefore, the closer the STD of reanalysis datasets is to that of SPIReS, the closer the STDR obtained by the reanalysis datasets are to 1.

**L300:** add a few words about what an STDR close to 1 means.

R: Thanks for suggestion. We added the meaning of STDR close to 1 in L323 of change-tracked manuscript.

**Description of product performance could be a bit more nuanced:**

Although HMSAR is good, and ranks at or near the top in all assessments, it appears to over(under) estimate in the east(west) compared to the SPIRES dataset. These biases probably average out when looking at the TP as a whole.

R: In Section 3.1.1, we have added more descriptions of the good performance of HMASR, GLDAS, and CRAL in terms of TP average of SCF climatology. We have also included the sentence you mentioned: "These biases probably average out when looking at the TP as a whole."

Further, HMSAR has too little spatial variability (low STDR) compared to the reference while GLDAS has a more appropriate amount of 'variability' with STDR of close to 1.

R: Thanks for your suggestion. We have expanded the description of HMASR and GLDAS in L410–417 of change-tracked manuscript.

**L306-307:** unclear what you mean by 'poor Taylor performance'. From Fig2b&c it that E5 captures the amount of variability well enough (STDR) but its large bias results in a high RMSD (and low skill score). I'd try to be more specific about which elements you are talking about.

R: We have revised the manuscript to replace the expression 'poor Taylor performance' with SS values throughout. Additionally, we have incorporated your suggestion to change the sentence to: "Other reanalysis datasets that overestimate SCF climatology, such as ERA5, ERA5L, and JRA55, are able to capture STDR well in some basins, but their large biases result in high RMSE and low SS values in the TP, consistent with Bian et al. (2019)."

**L329-330:** unclear what you mean by 'scattered on the Taylor diagram panels'. Please be more specific. i.e. there is a large spread in STDR, RMSD is higher compared to the westerly basins, correlations are fairly consistent between 0.5 and 0.6 (this is just by eye).

R: We have deleted the expression 'scattered on the Taylor diagram panels' and adopted your suggestion. The associated changes are in L453–460 of change-tracked manuscript.

**L331-332:** Do you mean the regional average SCF?

R: We have replaced 'the regional average SCF' with 'SCF climatological basin-averaged values' for better understanding by readers.

**L333-334:** Greater deviations compared to what? Unclear here and not obvious from the plots.

R: The main aim of this paragraph is to analyze the performance variations of SCF climatology spatial distribution among different basins within the TP. It focuses on the comparison between basins. We have made the description more specific to avoid ambiguity in the result description. Specifically, it now reads: "In particular, the ITP basin shows the poorest SCF spatial performance among basins, with the SS values of the reanalysis datasets <0.15, except for HMASR."

**L367:** start new paragraph with 'Variations'.

R: We have initiated a new paragraph starting from L575 in change-tracked manuscript.

**L415:** MERRA2 and CRAL seem to have strong correlations with temperature and precipitation in the west. I'm not convinced the evidence provided is sufficient to conclude that it is solely the absence of snow assimilation. Maybe use slightly more nuanced language.

R: Thank you for suggestion. We have revised the final attribution conclusions regarding the SCF errors in MERRA2 and CRAL. By comparing MERRA2 with HMASR, besides meteorological forcing, the lack of data assimilation may also contribute to the poor performance of MERRA2 in SCF simulation (L575–582 of change-tracked manuscript). As for CRAL, both snowfall and temperature play equally important roles in influencing the SCF biases. For more details, please refer to Section 3.1.2 of the manuscript (L526–530 of change-tracked manuscript).

**L439:** cut 'leading to a better Taylor performance' and instead just reference the skill scores.

R: We have revised the manuscript to replace 'poor Taylor performance' with 'SS values' throughout.

**L510-512:** You just described how ERA5 and eRA5L have the best agreement in terms of trends and those products don't assimilate snow data. So, the low performance of CRAL might not only be because of lack of data assimilation. At least, not if I follow the logic as presented in your paper. It might be that I am missing something or that the flow of ideas needs to be revised to support your conclusion.

R: Based on your analysis and suggestions, we have revised the conclusions regarding CRAL as described above.

**Section 4.1**– You discuss in terms of good and bad but what do you mean by this? How much can a parameterization alter the SCF and the accompanying SS and CI? Is this the same for all products? When you say it doesn't change the general performance do you mean the rankings of the products or the SS/CI values or both?

R: According to your suggestions for Section 4.1, we have adjusted the structure of this section to avoid confusion between SS and CI values. To maintain consistency with the overall framework of the manuscript, we have separately analyzed the impact of parameterization methods on the spatial distribution and annual trend of SCF. The parameterization process for offline output reanalysis SCF data mainly affects the SCF values rather than the phase variation of SCF in the time series. Therefore, our focus is on the spatial performance in reanalysis datasets caused by different parameterizations, namely SS values. The following is our point-by-point response to the comments on this section.

You discuss in terms of good and bad but what do you mean by this?

R: The SCF SS values obtained using a certain parameterization method are higher among most reanalysis datasets, indicating that this parameterization method has broad applicability and is considered optimal (L784–790 of change-tracked manuscript).

How much can a parameterization alter the SCF and the accompanying SS and CI? Is this the same for all products?

R: The impact of parameterization methods on SCF SS values varies among reanalysis datasets. Compared to the relatively extreme parameterization built in JRA55, significant improvements in SS values are observed for JRA55 SCF when using appropriate parameterization methods. However, for other reanalysis datasets with more appropriate built-in parameterization, the improvements are generally modest.

When you say it doesn’t change the general performance do you mean the rankings of the products or the SS/CI values or both?

R: Parameterization methods does not change the general performance means that there is not much change in the ranking of the reanalysis datasets after changing parameterization methods. We have clarified the previously ambiguous statement in L800–803 of change-tracked manuscript.

**Section 4.2–** not surprising that HMSAR+GLDAS produces the best skill scores because they had the best skill scores already. None of the other products had equivalent but opposite biases of these two products that would even out and improve the overall performance.

R: We agree with your perspective. Mortimer et al. (2020) demonstrated that averaging multiple reanalysis datasets can improve the accuracy of snow water equivalent (SWE) in North America. However, whether this method can enhance the accuracy of SCF in the complex terrain of the TP uncertain. Therefore, after quantifying this process, we have concluded that a two-member combined of HMASR and GLDAS is optimal for the study of SCF spatial scales, whereas a three-member combined of ERA5L, ERA5, and JRA55 is optimal for the study of annual trends.

**L61:** more advanced compared to the standard Modis SCF algorithm?

R: SPIReS employs the spectral unmixing method, which is more precise compared to the band ratio method applied in the standard MODIS SCF. In response to the suggestion from anonymous referee #1, we will provide a supplementary description on the comparison of these two methods in the SPIReS data description, in Section 2.1.1 of the manuscript.

**L73:** Unclear which reanalysis dataset you are referring to. Reanalysis datasets in general?

R: We have added the corresponding reanalysis datasets to avoid confusion (L93 and L96 in change-tracked manuscript).

**L84-86:** unclear. Do you mean reanalysis datasets that assimilate IMS and/or ground data?

R: The SCF data from IMS and ground observations are the reference data for SCF, not assimilated data. We changed the manuscript to “However, only a few studies have assessed the SCF performance of reanalysis datasets over the TP based on SCF data from the Interactive Multisensor Snow and Ice Mapping System (IMS; Helfrich et al., 2007) and ground observations (Li et al., 2022; Orsolini et al., 2019).”

**L139:** specify ‘over mountain areas in the western United States’.

R: We have added as suggestion in L163 in change-tracked manuscript.

**L195:** weekly

R: We have removed as suggestion.

## References


Bian, Q., Xu, Z., Zhao, L., Zhang, Y.-F., Zheng, H., Shi, C., Zhang, S., Xie, C., Yang, Z.-L., 2019. Evaluation and intercomparison of multiple snow water equivalent products over the Tibetan Plateau. J. Hydrometeorol. 20, 2043–2055. https://doi.org/10.1175/JHM-D-19-0011.1

Cui, T., Li, C., Tian, F., 2021. Evaluation of Temperature and Precipitation Simulations in CMIP6 Models Over the Tibetan Plateau. Earth Space Sci. 8, e2020EA001620. https://doi.org/10.1029/2020EA001620

Helfrich, S.R., McNamara, D., Ramsay, B.H., Baldwin, T., Kasheta, T., 2007. Enhancements to, and forthcoming developments in the Interactive Multisensor Snow and Ice Mapping System (IMS). Hydrol. Process. 21, 1576–1586. https://doi.org/10.1002/hyp.6720

Li, Q., Yang, T., Li, L., 2022. Evaluation of snow depth and snow cover represented by multiple datasets over the Tianshan Mountains: Remote sensing, reanalysis, and simulation. Int. J. Climatol. 42, 4223–4239. https://doi.org/10.1002/joc.7459

Mortimer, C., Mudryk, L., Derksen, C., Luojus, K., Brown, R., Kelly, R., Tedesco, M., 2020. Evaluation of long-term Northern Hemisphere snow water equivalent products. The Cryosphere 14, 1579–1594. https://doi.org/10.5194/tc-14-1579-2020

Orsolini, Y., Wegmann, M., Dutra, E., Liu, B., Balsamo, G., Yang, K., De Rosnay, P., Zhu, C., Wang, W., Senan, R., Arduini, G., 2019. Evaluation of snow depth and snow cover over the Tibetan Plateau in global reanalysis using in situ and satellite remote sensing observations. The Cryosphere 13, 2221–2239. https://doi.org/10.5194/tc-13-2221-2019