Author responses to reviewer 2 comments on:

"SPASS – new gridded climatological snow datasets for Switzerland: Potential and limitations"

by Marty et al. in *the Cryosphere*

We thank the reviewer for the time to assess our work and for the valuable feedback and suggestions. We respond to each point of the reviews below. The reviewer comments are highlighted in blue while our responses and comments are kept in black.

General summary

The manuscript presents gridded snow datasets for Switzerland. The gridded SWE datasets analysed are described in Micheal 2024 but in this study the authors apply a conversion from SWE to HS (SWE2HS) to obtain HS values which are then evaluated instead of SWE. The authors perform comparisons with in-situ data to understand the biases in the gridded data with respect to elevation and various aggregation methods. Trend analysis is also presented.
The manuscript focuses on four versions of gridded snow datasets produced and used operationally by SLF's snow hydrological service (OSHD). One without snow data assimilation spanning 1962-present and one with snow data assimilation spanning 1999-present (described in Mott et al. 2023). A derived dataset that applies a quantile mapping procedure (described in Michel et al. 2024) to extend the characteristics of the shorter time series dataset with data assimilation across the full period (1960-present). A combined dataset comprised of the original output with snow data assimilation for 1999-present and the quantile mapped data for 1962-1998.
Although the topic is consistent with the journal and the analysis seems reasonable the article needs to be revised to clearly differentiate this manuscript from existing work. I recommend the authors restructure the introduction and methods and provide a clear storyline that is woven throughout the paper.

Major revisions

Novelty - I had to refer to Scherrer et al. 2024 and Michel et al. 2024 to deduce the knowledge gap this paper aims to fill and to understand how it differs from previous work. My impression from those works is that this paper fills the gap specified in Michel et al. 2024 pp.8970 "The snow climatology produced for Switzerland is used here as an example, and will be studied and validated in more detail in a future work." The text should be revised to bring out this point. My reading is you are doing a more thorough evaluation of Michel et al. 2024 and you are also producing a snow depth dataset by applying SWE2HS and evaluating HS and not SWE.
I suggest a revised introduction where the authors succinctly present the companion works and

clearly outline the gap that is being filled by this analysis. Focus your introduction on what you are doing and why. Many of the details of the models currently in the introduction could be moved to an expanded methods/data section which would also include additional details about the models (e.g. forcing data) not currently provided.

Your reading/understanding regarding novelty is right. We agree to have been biased in this regard (the same issue was also mentioned by reviewer 1). Michel et al. (2024) introduced the quantile mapping method, which was used to create the new SWE datasets. They also provided some first validation of these datasets. Moreover, Scherrer et al. (2024) compared preliminary versions of these SWE datasets to other gridded SWE datasets. The main novelty of our study is the creation of the corresponding gridded snow depth datasets and thus possible comparisons to station-based measurements based on elevation and time aggregation dependent analysis as well as elevation dependent trends. This procedure allows to provide quantitative information on time aggregation- and elevation-dependent uncertainties, which was missing so far. We will change the manuscript text accordingly.

We agree with your suggestion regarding the content of the Introduction and Methods section. We will change the introduction text accordingly and move some of the details to the Method section. Please see also our corresponding answers under “Minor comments”.

I’m not fully convinced by the conclusion to use the combined dataset for trends. The time series and trends for entire elevation bands for CLQM versus the combined dataset do not appear to be sufficiently different to support for the use of the combined time series for trend analysis which has a known inhomogeneity. Indeed, the EKF dataset seems to provide different information at a local scale (e.g. Fig. 7) and the comparison with in-situ suggests it is of ‘higher quality’ but this does not necessarily make it better for trends.

We agree, that for the trends the combined dataset is not necessarily better. In our opinion we just compare the trends of the combined datasets with the trends of the other datasets. Please see our answer to your corresponding Minor comment below.

There is insufficient information presented in the methods to support the interpretation of results, especially as it pertains to Section 3.4. See specific minor comments.

Please see our answer to the mentioned specific minor comments.

Text needs to be revised to improve flow of logic and overall readability. Multiple concepts are often lumped together and not properly explained, or are conclusions drawn without corresponding support in the text or figures.

After reading all the minor comments below, we understand what is meant with this remark. You find our answer to each comment below. We believe that the revised manuscript will have a better logic and readability due to the changes made.

Improved and consistent references to literature throughout.

We are sorry this general comment is difficult to fulfill without more specific information. However, we hope it helps that the revised manuscript will also contain several new references, which among others are added due to some of your minor comments.

Minor comments

L13: Do you mean climatology?

Thanks for this hint. There was a word missing: The sentence now reads: "The comparison against in-situ station data for yearly, monthly and weekly aggregated values at different elevation bands demonstrates only slightly better performance scores for the higher quality dataset, which demonstrates the good performance of the quantile-mapping method which was used to produce the long-term climatological dataset from the higher quality dataset."

L47-50: Suggest referencing Figure 1 here.

Yes, this makes sense after mentioning OSHD_EKF. We will reference Fig. 1 in the revised manuscript.

L50-54: suggest revising sentence to make it more clear that the QM mapping method is presented in Michel et al. 2024. E.g., "This method, presented in Michel et al. (2024), allows..."

Good suggestion, which we follow in the revised manuscript.

L77: Do you mean Section instead of chapter?

Yes, that's wrong wording, we will exchange "chapter" with "section".

Section 2.1

- what is the temperature and precipitation forcing? Uncertainties related to forcing data are discussed in later parts of the paper (Sect. 4.3 in particular) but you did not inform the reader of the source of the forcing data in the methods.

That's right, thanks for this hint. We will add this important information to the Method section.

- general framework is outlined in the Figure 1 but this also needs to be detailed in text form. The schematic is not a sufficient replacement.

We agree and we will make up for this in the revised manuscript by moving a part of the Introduction into Methods and by rewriting another part.

- briefly outline the OSHD model. i.e. a temperature index model run at 1km spatial resolution, plus any other relevant details. It is described in other papers so a full description shouldn't be needed but you still need to provide enough information for the reader to interpret the results and to aid them through the discussion.

Similar as above, we now realize that forgot to write this in the Method section, but we will do it in the revised manuscript.

L97: Do you mean uncertainty or error rather than bias. Also, please expand. Are there certain conditions under which you expect the conversion to be more or less accurate? Maybe reference Aschauer et al. (2023) here.

We mean uncertainty. Yes, there are certain conditions, like rain-on-snow events, which contribute to the overall uncertainty, and which are mentioned in Aschauer et al. (2023). We will change this in the revised manuscript and reference Aschauer et al. (2023).

L94-95: In the introduction you present that Switzerland has a strong SWE network (L35-36) but then only evaluate HS. In the introduction you state that you evaluate HS because it is more accessible to the non-scientific public and is needed for 'other applications' (L65). Here (L94-95) you state that the comparison is limited to HS because you are using daily in situ data. The statements are not fully consistent. Both can be true, but the flow of logic needs to be clear.

We agree that this has potential for misunderstandings. We will therefore add to L35-36 that the SWE measurements are performed only bi-weekly and at only about 10% of the stations, which measure daily HS.

L103-104: Are there any references to support the statement about in situ sites being in flat sheltered fields? Maybe Grünewald and Lehning 2015?

Yes, that would be one option. But the best reference is The WMO Guide to meteorological instruments and methods of observation (2021), which we will reference in the revised manuscript.

L103-104: Here you are combining two issues into one. The problem you are describing is twofold - one is point to grid representativity; the other is that the points are biased to a certain landcover type and condition (the flat field statement).

We agree and will break this down into these two points in the revised manuscript.

L106: Suggest simply 'relative to station data are expected' (remove the intentional part).

We will remove "and intentional" in the revised manuscript

L114: Unclear if the data are quality controlled by the data provider or by you. By you, please detail these 'separate steps'. If by the data provider, provide a reference if possible and information on gap-filling as appropriate.

The data have been cured by the data providers. The details of the applied methods have changed over time. In general, these are physical threshold checks, as well as temporal und spatial consistency checks among others involving the relationship between depth of snowfall and snow depth. Unfortunately, there is no publication available documenting the various QC methods.

L115: Remove 'Technically'. Start sentence with 'Each station'

We will remove "Technically" in the revised manuscript.

L115-117: Did you also look at the results if you used the intersecting grid cell and compare with the smallest elevation difference method?

Good question. We did not do this exact comparison as in steep topography with 1km resolution the intersecting grid cell can have a large elevation difference to the corresponding station due to sub-grid topographic variability. As elevation is crucial for snow, we decided to compare our approach with mean value of the intersecting grid cell and its 8 neighboring grid cells. As the differences between the two methods were small and ambiguous, we decided to stick with our method.

L115-117: Again, you need to work on more clearly building your argument. Here you need to tell the reader why you compared it with the eight nearest grid cells and took the one with the smallest elevation difference. i.e. grid-point mismatch and changes in HS with elevation.

We agree and will add the above information that we choose this method because snow depth is heavily controlled by elevation.

L123: Can you provide a reference here, please?

Good point. We will add the following references:
doi.org/10.5194/tc-18-6005-2024
doi.org/10.1002/joc.5902

L127-129: What do you mean by ‘hardly any’? It might be helpful to have a figure showing the hypsometry and number of stations in the various elevations bands. I found myself looking for such a figure many times while reading the paper.

We are sorry, the information you were looking for is in Table S2, but the reference to this table was missing here. We will add it. There are only 0.2 % of all grid points below 250 m.

L154: What do you mean by ‘quite well’. Please be more specific. There are a number of these vague statements throughout. A figure showing the number of sites by elevation band versus the land area elevation distribution would support this claim.

Again sorry, the information you were looking for is also in Table S2, but the reference to this table was also missing here. We will add it to the revised manuscript.

Figure 2: what about a hypsometry vs sites per elevation band or something as a sub-figure? This would support your 'quite-well' statement as well as the lines where you say there 'are hardly any' sites. Is there a way to show the sites that were not included?

We agree that such information is important. This is why Table S2 was created. We believe the figure is already now quite loaded with all the legends, direction symbols and color scales.
We are sorry, what do you mean with “sites that were not included”?

L176: remove extra '.' after 'band' and before 'elevation'.

Thanks for this hint. We will remove it in the revised manuscript.

L188-190: Is this supported in the literature somewhere or is it conjecture? If SWE bias is close to 0 and HS bias not close to 0 then it is likely due to the conversion but this does not come across clearly in the text as written. Also, be careful to go one step at a time and clearly lay out each piece for the reader – i.e. why was this expected from the QM procedure (L187).

Quantile mapping forces the distribution of the corrected data to match that of the observed data in the reference period. Since the mean (bias) is a property of the distribution, matching the distribution means the mean (and thus the bias) is also corrected. This is the reason why a successful application of QM should produce a bias close to zero by design. And right, the HS bias is caused by the conversion. We will add this information and a corresponding reference of the QM zero bias (doi.org/10.1175/JCLI-D-14-00754.1) in the revised manuscript.

L190: revise to 'and therefore has not been'

We will correct this as suggested in the revised manuscript

L194: 'clear improvement in score performance when going from low to high elevations' text is a bit strange in the context of the text that follows. Not sure the phrasing here captures what is presented next. Maybe in the subsequent text it would help if you compared low to high for the same temporal period(s) instead of mixing everything together.

We fully agree. This will be better formulated/structured in the revised manuscript

L195: Does the lower MAAPE at annual and monthly at the 500m elevation band (compared to weekly) have to do with zeroes and all that to which the MAAPE method isn't well suited?

No, this caused by the fact the difference for yearly means in relative or absolute terms is smaller because possible spatio-temporal differences (one region has snow in week X/month X and the other region has snow in week Y/ month Y)) get more smoothed over longer time aggregations. This is not only the case for the 500 m band, but also for all other elevations.

L200: Any suggestions as to why the performance increases with elevation?

Good question. Temperature-index models are known to have the largest problems at temperature around 0°C (typical winter temperature at 500m in Switzerland), because of several reasons: 1) Snow events are more intermittent, with frequent rain–snow transitions. 2.) Rain-on-snow and short warm spells can cause melt not well captured by temperature-only models. 3) Snowpack may be shallow or short-lived, and energy inputs from rain, radiation, or turbulent heat fluxes can play a larger role. These problems get smaller with increasing elevation as the snowpack is continuous, temperature dominates melt timing and the influence of rain and warm spells is decreasing. We will add this missing information to the revised manuscript.

L214-215: Even with the MAAPE metrics which is supposed to reduce impact of small values do you still expect relative errors to be larger for small values (i.e. low elevations) compared to larger values (i.e. higher elevations)?

Yes, why not? MAAPE is just limiting the absolute size of large relative error. As illustrated in Fig. 7 mean HS at 500 m is about 3 cm, whereas mean HS at 2000 m is about 80 cm. This huge absolute difference is the main reason that MAAPE is decreasing with elevation.

L225: The same analysis as what? Same as Figure 4? Please specify

Yes, the same as Fig. 4. We will specify this in the revised manuscript.

L230: Do these poor scores correspond with certain locations or certain months or time periods or is it random?

we checked and found that the lowest R-scores generally come from the few lowest pixels, i.e. from two separate regions north and south of the Alps, which are often characterized by different snow conditions (Scherrer et al. 2008), regardless of the time in the winter season, if there is snow at all. This possible divergence is smaller for yearly values as there is a higher chance for compensation than for monthly or weekly values. We will add this information to the manuscript.

L236-238: Also, is the fact that you need another model to go from SWE to HS not another confounding factor?

Right, to a minor part sure, but this was already the case in the former analysis (Fig. 4)

L264: 'In a separate step'

Right, we will correct this.

L266: What do you mean by 'very similar'?

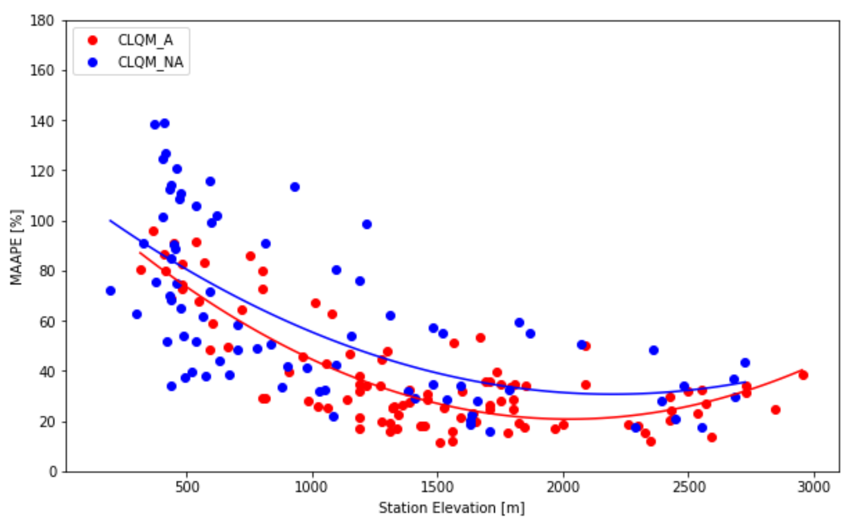
We mean that there is hardly any difference between the found BIAS for the assimilated and non-assimilated stations. We will clarify this and move the sentence after next to the front to emphasize what we mean.

L2658-269: Is it completely and only due to the flat field observations or is this only one of multiple confounding factors? What about elevation differences at point vs grid cell, or are these fully mitigated by the 8-nearest grid cell approach?

Most of the bias must be due to the flat field observation. A small part is also due to the SWE2HS conversion (Fig. 5). The 8 nearest grid cell approach can not be responsible as the median elevation difference (grid minus station) is +3 m and largest difference +78 m.

Figure 6 – Do you mean second order polynomial? Did you consider making a similar plot but with MAAPE instead of bias?

Yes, you are right it as second order polynomial. We will correct this.
We produced a corresponding MAAPE plot to show you how it looks like. Non-assimilated show a larger variability (probably also due to the smaller time period) and slightly higher MAAPE (as expected from the BIAS plot) but the same elevation dependence as the assimilated stations.



L278: Do you mean all grid points with coincident in situ stations or all grid points in Switzerland? Unclear.

It’s for all grid points corresponding to the stations as written in the figure caption. We will also add this information in the revised manuscript text.

L279: ‘slightly higher’ instead of ‘slightly increasing’

We will change this to “slightly better” in the revised manuscript

L283: Is this analyzed in your paper or elsewhere? If elsewhere, please add appropriate reference.

In the next sentence we write that this has been already analyzed in other studies and provide the corresponding references.

L283-284: Lack of language precision: your comparison showed that HS is lower with the gridded data but did not demonstrate why it is lower.

Thanks for this clarification, you are perfectly right. We will rephrase the sentence in the manuscript accordingly.

L285: revise Section 3.3 subheading to ‘Evaluation of trends’

We will change this in the revised manuscript

L289-290: Is this also in Michel et al. 2024 paper or is this a new finding? Not clear what is new to this paper and what is in a previous study.

This is a new finding as Michel et al. (2024) did not analyze elevation dependent differences and their station-based analysis was limited to a qualitative visual comparison of the annual snow evolution at a few selected stations.

L291: Recommend highlighting and reiterating the fact that it would only change or improve things from 1999 onward.

We agree and will add this information in the revised manuscript.

L292-293: 'where the benefit of using OSHD-Comb can be nicely demonstrated' again, you are jumping ahead here.

We agree and will rephrase this paragraph accordingly

L293: How does an anomaly map (although interesting) supports your discussion of trends? Also, the anomalies for this year look fairly different to me. There are some interesting differences that could be expanded upon.

We agree that such an anomaly map is not directly related to trends. We will therefore restructure this section (i.e. move the anomaly map to the end of this section) and provide additional information on the shown differences.

L307: was this clearly demonstrated in your analysis? or was this in the other paper? Unclear. If in your analysis maybe add a sentence to summarize those results in the relevant results section.

This is described in detail in Michel et al. (2024) and additionally shown in Fig. 1. Will add this references and link the figure to this sentence in the revise manuscript.

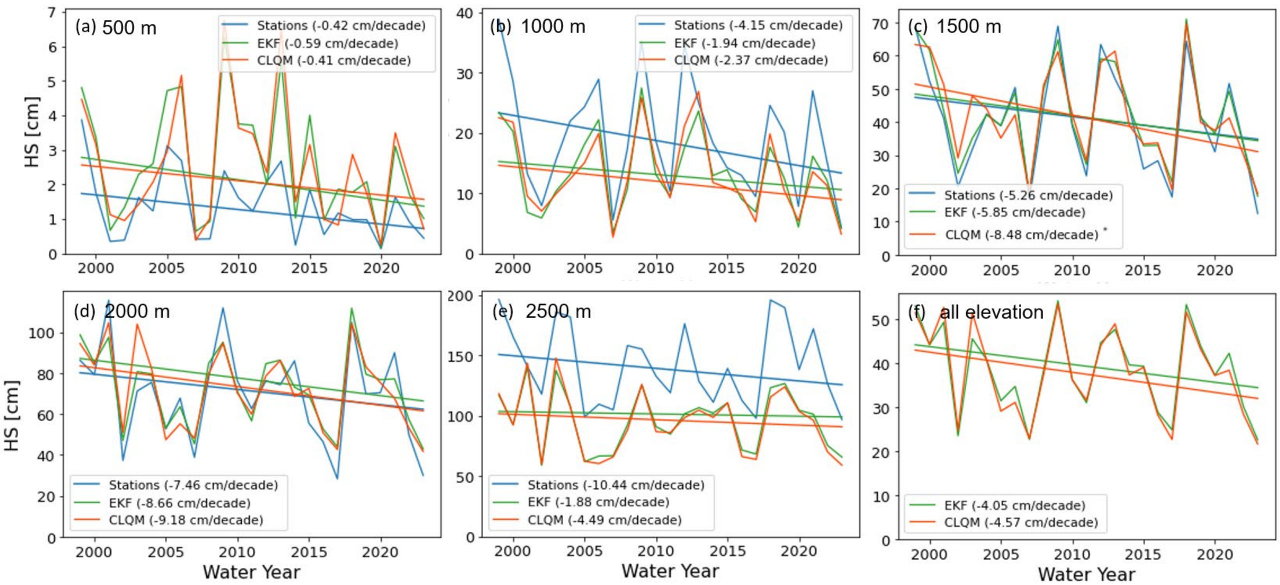
L208: albeit only during the last 25 years

We guess you refer to L308. Thank you very much, this correct indeed and important to mention. We will add this information to the sentence.

Figure 8 and associated text –

- By extending both back to 1960 you are muting the differences. There are 40years where the time series is the same and only 25 years when they might differ. I know it's only 25 years, which is considered short for climatology, but I'd like to also see the trends calculated over that period to be able to better compare the two models.

Please find below a similar plot as Fig. 8, but only for the last 25 years. The results demonstrate no significant difference between CLQM and EKF, i.e. at some elevation CLQM trends agree slightly better with station trends, at other elevation EKF trends agree slightly better.



- Second, if the trends are similar then why concatenate two different time series which might introduce an artificial break in the time series? Why not simply recommend using CLQM for trend analysis?

Thank you for this note. We actually do not recommend not to use OSHD-CLQM (or recommend to use OSHS-Comb) for trend analysis in the current text. We only show that OSHD-Comb has advantages for the application case of anomaly maps. If a user does not want to use two different datasets for trends and anomalies (with its possible misunderstandings due to inconsistences) he/she may also just use OSHD-Comb.

L340: what do you mean by ‘it’s tempting to fill in missing snow information with that from a grid’? Did you fill in missing data with gridded data?

We mean that ‘it’s tempting to fill in missing snow information with that from a grid point in case you have specific problem at a location where there is no station nearby. Since this unclear yet, we will add this additional information to the revised manuscript.

L358: Given that you identified sites with erroneous observations or inhomogeneities why not exclude these sites from the analysis entirely?

We agree the few sites, which were known to be inhomogeneous, could have been excluded. However, we intentionally kept them to check if they will stick out during analysis and see if other stations demonstrated a similar behavior to find potentially inhomogeneous stations or grid points.

L360-361: This is the first mention of meteorological input data. You haven’t told the reader what the input forcing is, so they have no frame of reference for the sites listed. Need to add this type of information to the methods/data.

We fully agree (this point was also mentioned by reviewer 1) and will therefore add this information to the methods section.

L383-384: also important for energy balance

We agree and will add this information to the revised manuscript.

L388: please reference the corresponding figure.

Right, we will add a reference to Fig. 8.

L392: perhaps ‘see’ rather than ‘detect’

We will replace “detect” with “see”.

L392-293: better agreement with station fluctuations compared to what?

Better agreement compared to the year-to-year fluctuations at stations. We will change this in the revised manuscript.

L393: revise to ‘demonstrate a significant decreasing’

We agree and we will change this in the revised manuscript

L395-398: Perhaps mention why does the fact that it coincides with the snow line mean it is expected to have the largest change in number of snow days? You need to explicitly lay out to the reader of these connections that are implied.

We will add the information that is caused by the fact that below 1000 m there is hardly any snow days left and above 1000 m the absolute decrease is smaller as the mean winter temperature is still below melting.

Section 3.4

- This section reads more like discussion than results. Consider renaming it.

We would like to keep it as is, because part of the discussion is already in the former sections and because reviewer explicitly liked this section.

- There are some interesting pieces in this section but could be improved by synthesizing and presented the information in a more coherent manner. I suggest reorganizing the section to put all the arguments about forcing data together and all the arguments about OSHD-Comb together. Also, moving some of the details about forcing data and precipitation partitioning to the methods may also help the reader anticipate and understand what is being presented. You can remind the reader by repeating key pieces but this should not be the first place you bring for example, where the precipitation data came from. Finally, the rationale for not analyzing daily values (L429-431) should be in the methods.

Some of these points were also mentioned by reviewer 1. We agree that the structure of this section is not coherent. We will move the rationale for not analyzing daily values to the mnethod section and will restructure remaining part as suggest.

L403: Is it limited to inhomogeneities or are there also systematic biases in one or both forcings? Are they regional? by elevation?

Yes, it is limited to inhomogeneities. Since the products are based on interpolation of station values it can have an impact on region or/and elevations.

L405-406: Here, are you talking about the impact of combining two datasets into 1 time series or are you talking about the forcing data? Unclear.

Here we write about the potential inhomogeneity of the forcing data. We will clarify this in the revised manuscript.

L425: Suggest 'input data are a reason' rather than input data are 'the' reason.

We agree and will correct this as suggested in the revised manuscript.

L432: Other reasons for what? Other reasons for not analyzing at the daily time scale? Please clarify.

Yes, another reason for not analyzing at the daily scale. We will clarify this in the revised manuscript.

L469-470: I don't feel that you really presented enough to really support this claim/argument. I believe you have probably done the analysis but how it is presented here and what you have chosen to present doesn't seem to really support this piece (at least not sufficiently)

In our opinion, we demonstrated this with several figures throughout the manuscript, especially when comparing Fig. 4 with Fig S2 and Fig. 5 with Fig. S3. However, we will rephrase this part to make it clearer and add some corresponding sentences to the description of Fig. S4.