# Author responses to reviewer 1 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.

Marty et al. present an evaluation of spatially gridded datasets of snow cover over Switzerland, with high spatial resolution (1km) and long duration (60+ years). They evaluate different datasets, with and without assimilation using ground observations using different metrics across elevation, and also compare long-term trends. The manuscript falls well within the topic of TC. It is well written and results are discussed critically, for which I compliment the authors (especially sec 3.4 on limitations is great). However, there are a few issues that remain unclear.

**Major points:**

The novel contribution of this study with respect to previous studies is not completely clear – from the literature review in the intro it seems that a comparison of the "new" dataset has already been performed (Scherrer et al 2024) and the dataset itself has been created and validated in Michel et al. 2024. Please highlight the differences between the past studies and the current one more clearly, as well as the novelty of this particular study.

We agree to have been biased in this regard (this point was also mentioned by reviewer 2): 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 have changed the manuscript text in the Introduction accordingly.

I assume one novelty is the elevational analysis and temporal aggregation unit analysis. While the first one is evident and well explained, the second seems very minor to me after reading the paper. First, all of the figures are in the supplement, and often the results are presented as similar across temporal units. Moreover, the motivation behind the different temporal units is not evident. And also why daily was not a temporal unit.

You are right, there is less emphasis on temporal aggregation than on elevational analysis. The main reason is the fact that elevation dependent differences are larger than temporal aggregation dependent differences and that the main message is always the same: "lower performance for smaller time aggregations". Moreover, when describing Fig: S2, S3, we also emphasized the differences due the shorter time aggregations in absolute and relative terms. We agree that the corresponding description of Fig. S4 so far was only brief. We therefore extended the description in the revised manuscript.
The motivation behind the used temporal units was given by the following facts: Climatological analyses are often provided by yearly or monthly reports and we wanted to assess the uncertainty of the new snow products with the goal to include them in future such reports. Moreover, knowing about the need for timely (near real time) public information about possible current extraordinary situations, we also assessed the weekly aggregation. Daily aggregations were by purpose not assessed as quantile mapping is not expected to do more than a climatological bias correction, meaning biases at short time scales, like on a single weeks or days are not necessarily corrected. This is demonstrated by the fact that already the analysis weekly values revealed relative errors of more than 100 % at low elevations. Additionally, model results of the current day are only available after 15:00 local time, which is too late for timely public information. We have added this information to 2.3 in the revised manuscript.

On the other hand, another important factor is seasonality. Did you consider how error metrics vary across the snow season? E.g, if they are constant or increase/decrease towards the end of the season?

Yes, we did specifically analyze how the error metrics vary across season. The beginning and end of the snow-covered season has generally a lower performance than mid-winter also at higher elevations because the situation is similar as at low elevations during the entire winter. This implies the transition seasons between no-snow and snow also at higher elevations have the same potential problems as at low elevations during the entire winter. These problems involve among others high spatial variability and no information on the soil temperature, which is decisive for the survival of potential snow fall. But since our focus was between November and April this seasonality issue does only affect the 1000 and 1500 m elevation band. We have added this explanations to 3.2 in the revised manuscript.

While the authors have a great choice of evaluation metrics, including MAAPE, which seems very interesting, there needs to be some consideration of whether the metrics (and the associated figures and statistics) refer to spatial, temporal, or their combined spatiotemporal variability. For example, L176-180 is unclear (and also not really relevant to know computational details like how your array looks like). It would be great if you could identify what the metrics and variability refer to, i.e., where you average over space, time (years or other), or where you show variability across space or time or both. Also the order of calculation matters, so if you first do metrics, then average (e.g., over weeks, or over gridcells), or first average and then do metrics. Less for bias, but significantly for all other metrics. The ordering of calculations is not completely clear from the manuscript.

This aspect indeed has potential for improvement. We always first averaged over time for each elevation band. This means that the boxplot shows the variability across space for each temporal aggregation in each elevation band. This means in the case of the model-to-model intercomparison (Fig.4) the boxplots were created based on the number of grid points per elevation band (as listed in Table S2). In the case of model-to-station intercomparison (Fig. 5), the boxplots were created based on the number of stations per elevation band (as listed in Table S2). We have added this missing information to 2.3 in the revised manuscript.

**Minor points:**

Abstract is sometimes confusing. It mentions two datasets, named old and new, which remains unclear. My suggestion is to try to make the abstract as self-contained as possible. Also, some numbers would make it less vague.

We have reworded the Abstract in the revised manuscript to make it more coherent and also added some numbers to make it more concrete.

L40 “many applications”, please provide some examples.

We have extended the sentence with: “…limit their usefulness for climatological applications beyond station-based analysis, i.e. the provision of altitude-dependent region- or country-wide snow information.”

Besides the general intro to snow, the introduction focuses exclusively on the history of gridded snow datasets in Switzerland. Since the topic of spatially gridded snow cover datasets is not trivial and can be tackled (in theory) in multiple ways (stations, remote sensing, modelling), maybe a broader introduction into gridded climate datasets and gridded snow in particular might be useful for readers.

We added some sentences regarding a broader introduction into gridded climate datasets in the Introduction in the revised manuscript.

Introduction and Methods are somewhat mixed, since the used models/datasets are presented in the introduction, but then in the method the models/datasets are not described further. I guess there are other studies presenting this in detail, but for completeness, I suggest including a brief summary of the key model characteristics and meteorological input for the different datasets.

We believe that the used models/datasets are also shown in Fig. 1 in the Methods section. Nevertheless, we streamlined the Introduction and added a brief summary of the meteorological input and the models used in the Method section of the revised manuscript.

L114 some reference would be useful

We assume that the reviewer refers to "The data of these stations have been carefully quality-controlled and gap-filled in separate steps." 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 (we will added this information to the text) among others involving the relationship between depth of snowfall and snow depth. According to MeteoSwiss and SLF experts, there is unfortunately no publication available documenting the various QC methods, but we added a reference for the gap-filling.

L135 unclear if for the climatological analysis the reference period was 1991-2020 or 1999-2023.

We agree that this paragraph is confusing and changed the last sentence to: "When investigating performance differences between OSHD-CLQM and OSHD-EKF, as well as to have enough in-situ data (Fig. 2, Table S2) available in the different elevation bands (mean per elevation band is 20 stations, minimum 14 stations, maximum 34 stations), the evaluation is based on the period 1999-2023."

L151 so relative trend is based on the Theil-Sen slope, but relative to what? Theil-Sen intercept, mean over the whole period, something else?

This sentence has been deleted in the revised manuscript as we do not show any relative trend results.

Sec 2.4 did you compare the difference in trends between CLQM and Comb?

Yes, see section 3.3.1: There we write: “The OSHD-Comb trend magnitude is marginally weaker than the OSHD-CLQM trend magnitude and thus closer to the station-based trend magnitude for all investigated elevations with the exception of the 2000 m band.”

Related, has the meteo input (the temp and precip grids) been tested for homogeneity? Otherwise, I guess the snow trends could reflect input dishomogeneities as well…(ok this comes around L405...)

No, we did not check the meteo-inputs for homogeneity, as they are potentially inhomogeneous due to their creation method, as described in section 3.4.

Sec. 2.5 Since you use relative errors, I guess relative bias would also be interesting? While absolute bias increases with elevation, relative one should decrease, no?

Yes, this is true for SWE, but not necessarily for HS. We did not specifically analyze relative bias, but when combining the results of Fig. 6 with Fig. 8, we can calculate that the relative bias for HS is decreasing with elevation only until about 1500 m, i.e. if the absolute bias is small (see Fig. 5).

L190 “because HS has been derived from SWE” but this is true for all elevations.

Yes, this is true for all elevations, but the used SWE2HS algorithm sometimes seems to underestimate HS at highest elevations (Third sentences in 3.1).

L210 “boxplots consisting of the 25 yearly values” but there more points than this in the boxplots?

Thanks for this tip, this information was wrong. Since the boxplots show the spatial variability (see our answer to your last major point), they were created based on the number of grid points per elevation band (as listed in Table S2) in the case of the model-to-model intercomparison

(Fig.4). In the case of model-to-station intercomparison (Fig. 5), they were created based on the number of stations per elevation band (as listed in Table S2). We have corrected this information in the revised manuscript.

Fig 4: very unusual choice for the whiskers to go from 5$^{th}$ to 95$^{th}$ percentile. Why not the standard boxplot variant with 1.5*IQR from the box edges (up to the largest value, if within range)? Also because your choice highlights a lot of "outliers", which are not really outliers, but continuous variability, in my opinion. One could also do the other standard whiskers that go to min and max.

Thanks for this statistical help. We used the 5$^{th}$ to 95$^{th}$ percentiles as we thought that this is easier to communicate to practitioners. However, we agree with your argument about the outliers and now recalculated all boxplots with standard 1.5 IQR whiskers. This did reduce the number of outliers. We have changed the corresponding text and exchanged the figures in the revised manuscript.

Fig4 and 5: for bias a line at y=0 would be useful. Or some light background grid lines in all panels.

Good idea, which is now included in both figures in the revised manuscript.

Fig5: if the focus is on the comparison between CLQM and EKF, it would be useful to show the boxes side-by-side (e.g., with different fill or line colours) and not in separate panels. If the focus is on comparing by elevation, it's fine like this.

Yes, we chose this representation because our focus is the comparison by elevation band.

Fig. 6: a polynomial of first degree should be a straight line…

Thanks a lot for the hint, this indeed wrong. It is a polynomial fit of second degree. We have changed this in the revised manuscript.

Besides Fig6, is it possible to produce the same figure as Fig5 also for the non-assimilated stations, or put them side-by-side to compare the performance metrics between assimilated and non-assimilated stations?

Technically this would be possible, but it would not be comparable to Fig. 5 as the non-assimilated stations cover any time period between 1999-2023 (median 12 years, minimum 5 years) instead of 25 years as in Fig. 5.

L278-284 this paragraph feels a bit off in the current section. Or what does it refer to? To all stations assimilated and not? It also contains something on climatology and trends…

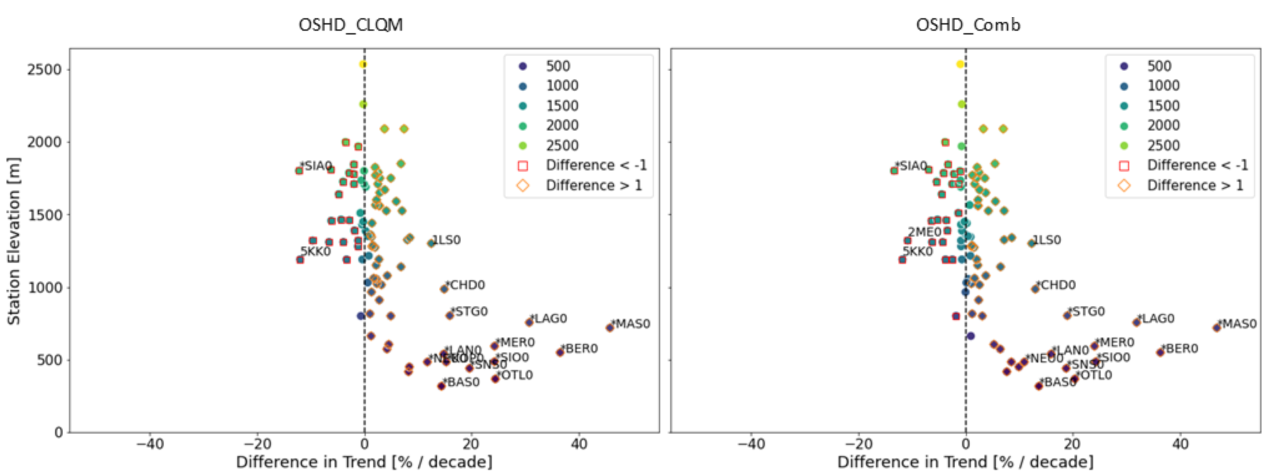
We agree. This paragraph rather belongs to Fig. S4. We added this information and rephrased it in the revised manuscript.

Fig7 could you please increase resolution or use vector graphics? It's not possible to zoom in easily.

You are right: We have increased the resolution of the original Fig.7 (now Fig.8) in the revised manuscript.

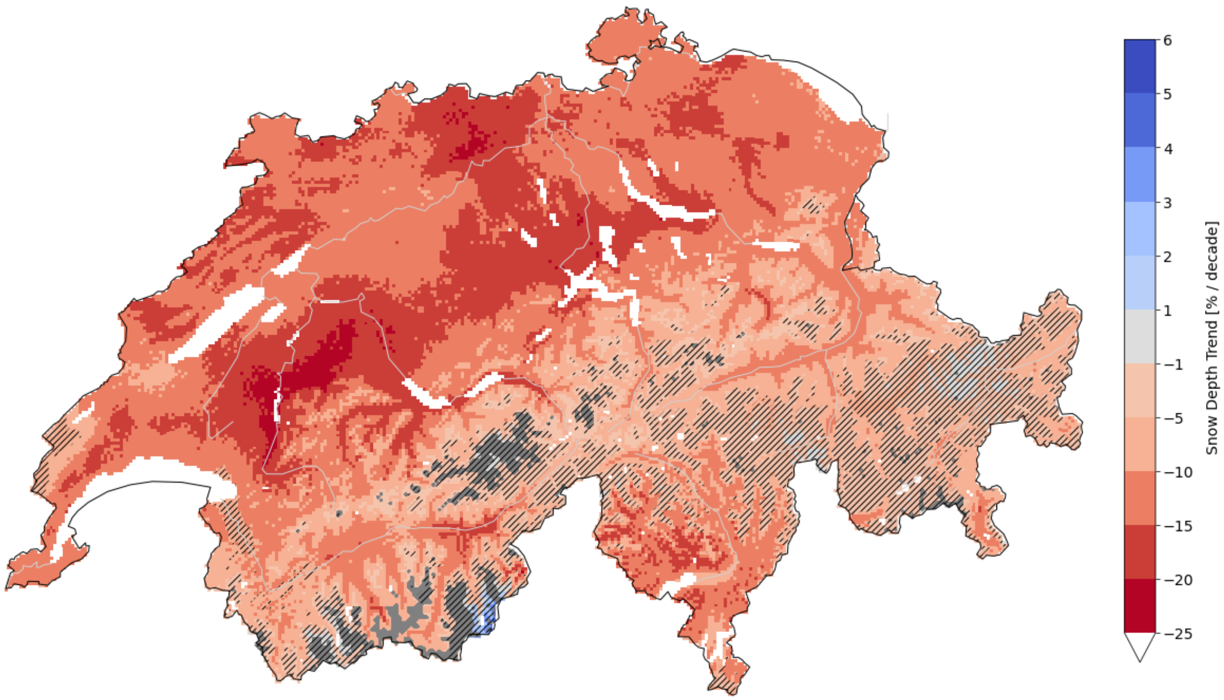
Fig9, by chance, do you also have a relative trend figure of this?

We are not sure what you exactly mean. In our opinion the most meaningful way to show Fig. 9 in relative terms is the difference with respect to the mean snow depth at each station (see the below figure, where stations that show a difference greater than ± 10%/decade are labeled). As expected from Fig. 5, the relative difference shows large values at low elevation. On the other hand, above 1000 m only 4 stations show differences larger 10 %.




Similarly, I guess a relative trend map (Fig10) could also be useful?

As shown below and added as supplementary Figure S8, we additionally produced a relative trend map based on Fig. 10. Similar as above, the largest relative trends are shown at low elevation (mainly Swiss plateau), which reach almost -25 % per decade. In the Alps typical relative trends are between -5 and -10 % per decade.




L427 unclear, might be resolved with a more detailed method section description

This refers to the fact that the daily degree factor of the temperature index model is seasonally dependent, but the same for all pixels, i.e. derived from the mean of all station measurements in the last 25 years but applied to the entire 60 years. We agree that this was unclear and we rephrased this section in the revised manuscript.

L429 Why not use tmin and tmax instead of tmean? It's also much more stable over time, considering the 60year period.

We agree, but gridded data of tmin and tmax is only available back to 1971 in Switzerland.

L461 please repeat the reasoning from Michel et al. 2024 here, shortly.

Generally, quantile mapping is known to not necessarily correctly preserve extreme events. The applied quantile mapping method can by definition not really capture extreme values, as they are corrected according to the correction of the 99th quantile. We have added this information and rephrased the corresponding sentence in the revised manuscript.

Conclusion could be a bit more general. E.g., what are the implications of the results? Can the fairly simple degree-day model be trusted or is the station data maybe more accurate? What are use cases of such a dataset in climatology, hydrology, ...?

The question if station or SPASS data can be more trusted depends on the application and region in focus. Generally, as shown in our study, at low elevation station data is more thrustful. At higher elevations SPASS data from larger regions and longer time aggregations have the advantage of being less locally dependent and available in pre- and after season periods (early autumn and late spring). Anyway, we added some more implication related sentences to Conclusion in the revised manuscript.

Not necessarily for this study but have you considered evaluating the grids with remote sensing? Like with optical-derived snow presence?

The SWE dataset has already been compared with remote sensing data (doi.org/10.3929/ethz-b-000652501).

PS: Thanks for the review invitation, I had SPASS reading the paper, or Spaß, as we spell it here :)

# 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 based on the work of Aschauer et al. (2023) 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 have changed the Introduction text in the revised manuscript accordingly.
We agree with your suggestion regarding the content of the Introduction and Methods section. We have changed the introduction text accordingly and have moved 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 comment below.

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 the answers to your minor comments.

**Minor comments**

L13: Do you mean climatology?

Thanks for this hint. There was a word missing. The sentence has been rephrased due to other comments.

L47-50: Suggest referencing Figure 1 here.

Yes, this makes sense after mentioning OSHD_EKF. We have added a reference to 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 followed in the revised manuscript.

L77: Do you mean Section instead of chapter?

Yes, that’s wrong wording, we have exchanged “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 have added 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 have made up for it in the revised manuscript by moving a part of the Introduction into Methods and by rephrasing 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.

This was so far only in Fig. 1, we realized that we forgot to provide a description in the Method section, which you will now find it there 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 clarified 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 therefore added 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 (2024), which we now have added.

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 have broken this down into these two points in the revised manuscript.

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

We have removed “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 (which was added in the revised manuscript) 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 have removed “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 have added 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 have added 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 have added 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 have added 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 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 have removed 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 have now added this information and a corresponding reference of the zero QM bias (doi.org/10.1175/JCLI-D-14-00754.1) in the revised manuscript.

L190: revise to 'and therefore has not been'

We have corrected 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. We have now improved the structure and rephrased this paragraph 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. The main reason for this better performance at high elevation is the fact that the error indices in this analysis reflect the performance of the quantile mapping step - and that the signal-to-noise ratio of the quantile mapping is increasing with elevation due to the larger absolute mount of snow. We have added 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 have now specified 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 (dx.doi.org/10.3354/cr032187), 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 have added 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 corrected this.

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

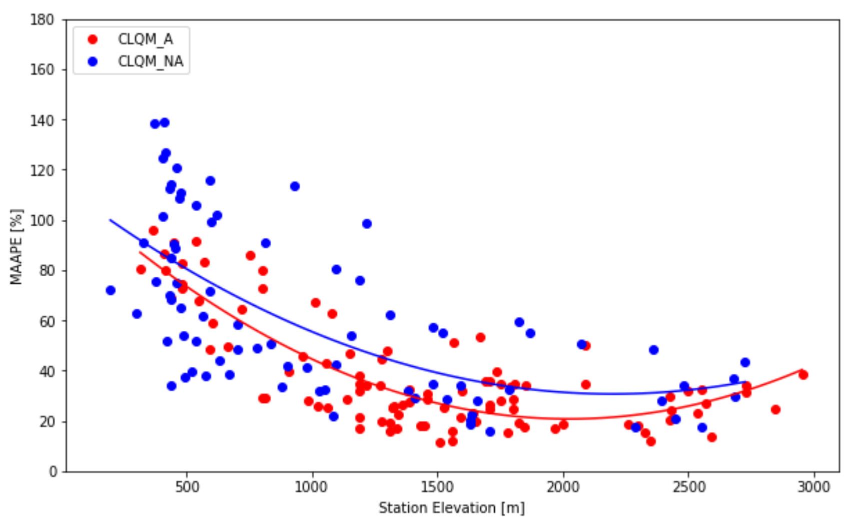
We mean that there is hardly any difference between the found BIAS for the assimilated and non-assimilated stations. We have clarified this in the revised manuscript and moved the sentence after next to the front to emphasize what we mean.

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?

Thanks for this hint. Yes, you are right it is a second order polynomial. We have corrected this. Regarding MAAPE instead of bias, we produced the corresponding MAAPE plot below to demonstrate 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 now added this information to the revised manuscript text.

L279: 'slightly higher' instead of 'slightly increasing'

We have changed 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 hint, you are perfectly right. We have rephrased the sentence in the revised manuscript accordingly.

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

We have changed 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 have added 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 have restructured this section accordingly in the revised manuscript.

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 have therefore restructured this section (i.e. moved the anomaly map to the end of this section) and now 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. We have added this reference and a link to the figure in this sentence in the revised manuscript.

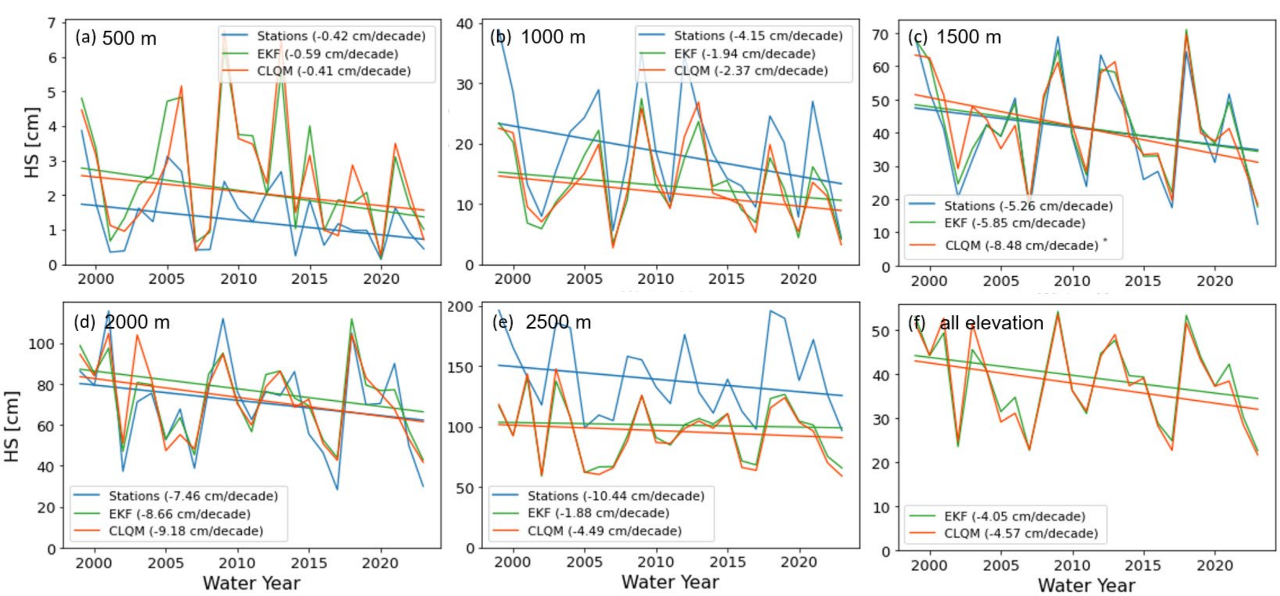
L208: albeit only during the last 25 years

We guess you refer to L308. Thank you very much, this is correct indeed and important to mention. We have added this information to this sentence in the revised manuscript.

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 was unclear, we have now added 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 that 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 the further analysis and to 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 have therefore added this information to the Methods section.

L383-384: also important for energy balance

We agree and have added this information to the revised manuscript.

L388: please reference the corresponding figure.

Right, we added a reference to Fig. 7.

L392: perhaps ‘see’ rather than ‘detect’

We have replaced “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 have changed this in the revised manuscript.

L393: revise to ‘demonstrate a significant decreasing’

We agree and have changed 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 added 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 we already discuss other issues in the former sections. For this reason we explicitly named Chapter 3 ‘Results and Discussion’. Moreover, reviewer 1 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. Therefore, we have moved the rationale for not analyzing daily values to the Method section and have restructured the remaining part as suggested.

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 regions 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 have clarified this in the revised manuscript by providing a link to the Method section.

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

We agree and have corrected this 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 have clarified this in the revised manuscript by adding sentences.

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 now rephrased this part to make it clearer and added some corresponding sentences to the description of Fig. S4.