Author response to Comment RC1 on gi-2021-16

We thank the reviewer for her/his time to assess the manuscript and for the valuable feedback suggestions. Below, we respond to each suggestion and comment one by one. The reviewer comments are highlighted in blue while our responses are kept in black.

The manuscript "Evaluating methods for reconstructing large gaps in historic snow depth time series" by Aschauer and Marty is an interesting technical study focusing on interpolation of snow depth time series. The topic is relevant for the community focusing on the snow modeling and time series analysis. The paper summarizes in an effective way concepts which are known by the community, by applying six different interpolation methods to a snow dataset from Switzerland. I think the manuscript represents a good contribution to the current scientific discussion on the topic. I have a few suggestions that the authors could consider in the revised version of the manuscript:

- Lines 43-45: I only partially agree with this statement. The fact that temperature index models require only precipitation and temperature data is the only reason why someone should apply them. In fact, they provide only a very rough physical representation of snow processes. Moreover, it seems that the authors do not consider in the modeling approach processes such as snow redistribution driven by wind or gravitation. These processes are generally not dominant if the choice of the location of the monitoring station is performed properly, but the authors may comment on it.

[Answer]: We clarified the sentence and added two additional sentences in which we take up the points from your comment and highlight the simplistic nature of temperature index models and the associated problems.

- Line 70: I think time series available for this dataset are generally longer than the period selected by the authors (1999-2020). Could they provide shortly a justification for this choice?

[Answer]: Yes, many stations from our station set cover longer time ranges than 1999-2020. We wanted to have as many gap free station records in our set of evaluation and predictor stations as possible and therefore decided to take the last 21 years because the most recent period has the highest station density in the Swiss station network. Additionally, the 21 year time period was chosen because we wanted to have a long enough data set on the one hand (containing a few well known snow abundant and snow scarce years) and a common (realistic) length of available snow depth time series for the training period on the other hand.

We think that 21 years per station is enough for the cross validation. If the period would be longer, the “leave one winter out” approach would probably become problematic due to potential shifts in snow climatological dependencies since many of our models assume temporal stationarity. We added two sentences to the manuscript to clarify our choice of the selected period for our station dataset.

- Line 87: should “station or both” be “station of both”?

[Answer]: Yes, has been corrected to “station of both”.

- Interpolation methods: I fully understand that it is not possible to test all available interpolation methods, but I was surprised that a standard interpolation method like Kriging with external drift

or let's say Kriging based methods were not considered. Could the author please motivate this choice?

[Answer]: As stated in the discussion [line 265 ff] we also tested the GIDS method during early phases of the manuscript preparation. In the reference paper of GIDS [https://doi.org/10.1016/S0168-1923(98)00102-6], the authors also compare this technique to Kriging based methods. The findings there are pointing towards the direction that Kriging is not superior to GIDS and also has problems in sparse station networks.

However, when trying to do spatially continuous interpolation of snow depth data and not only for a single target (i.e. a single station), Kriging might be a suitable approach. We added a part to the discussion section where we explain our reasoning on why we did not include Kriging based methods in our set of compared methods.

- Line 178: Is a threshold of 1 cm really relevant for tourism?

[Answer]: In fact, for winter tourism higher HS thresholds such as 30 cm are more often used [e.g. https://doi.org/10.5194/tc-13-1325-2019]. However, as the winter tourist expects a nice wintery scenery, a threshold of 1 cm and the question whether a snow cover is present has also some importance.

- The Matiu/WNR method should be consistently named in the manuscript

[Answer]: We named the WNR method consistently throughout the manuscript

- Line 230 should "derived form" be "derived from"?

[Answer]: Yes, has been corrected accordingly.

- Line 237 I think that beside the value of HSmax it would be interesting to compare when HSmax occurs using the different methods.

[Answer]: We calculated the difference in days of the date of HSmax measured - date of HSmax reconstructed. We added Figure A3 with the distributions of these differences to the appendix and inserted a sentence in the results section.

- Figure 4: it is very hard to read the values of r2, RMS and BIAS, please move them to a Table

[Answer]: We removed the numbers from the figure and moved them to the newly added Table 2.

- Although I acknowledge the difficulty of summarizing the results of this work in a graphical form, I find the quality of Figures 4, 5 and A1 low. The black lines in the blue box plots are hardly readable and in general the size of the subplots it is too small allow the reader a quantitative interpretation of the results.

[Answer]: We changed the color of the median lines in the boxplots in order to make them appear more visible. Additionally, we made the subplots a little larger by getting rid of in-between labels.

- Line 298: It would be nice to have a ranking of the methods to be applied in these circumstance.

[Answer]: We added an overview table (Table 3) to the discussion which summarizes the suitability of different methods in different situations (dense and sparse station networks, gaps in neighboring stations) and for different applications.

- The authors correctly point out that data interpolation is an important step to have longer time series for climatological analysis. However, after data interpolation the next step would be to homogenize stations. Homogenization of snow depth time series is an actual research topic (e.g., Marcolini et al., 2019) and it would be nice to see in the discussion part some comments given by the authors about the effect of the different interpolation method on the quality of the resulting time series. I do not expect a quantitative assessment, since it would probably result in another paper due to the required amount of work, however some qualitative insights would be interesting to stimulate further research in this direction.

[Answer]: You are correct; often gap filling is a precursory step of homogenizing a climatological measurement series. We added a paragraph in which we qualitatively assess the effect of filling gaps on homogenization.

Marcolini, G., Koch, R., Chimani, B., Schöner, W., Bellin, A., Disse, M., & Chiogna, G. (2019). Evaluation of homogenization methods for seasonal snow depth data in the Austrian Alps, 1930–2010. *International Journal of Climatology*, *39*(11), 4514-4530.