

Author response to Comment CC1 on gi-2021-16

Thank you Michael for your time to give the manuscript a read and for your 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.

I've been following this work since over 1 year and am happy it's being published. Congratulation to the authors for this evaluation, which is immensely helpful for future climatological assessments of station observations of snow depth. Also, they are to be thanked for making the code publicly accessible.

I suggest using consistent naming of the Matiu/WNR method in figures, tables, and text. Personally, I prefer WNR, since it's not "my" method...

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

The comparison of more traditional approaches to ML tools is very useful.

The only thing that I found odd is the large bias and errors in dHS1. A higher one for HSmax is to be expected, because one value (maximum) can behave very sensitive. But if daily HS is reconstructed well, as well as HSavg, why not dHS1? However, I reproduced the results of dHS1 based on Matiu et al. 2021 and found basically the same order of bias in dHS1 for the WNR method. Even though, it appears to be highest in middle elevations (1000-1500m) and lower otherwise, and the bias decreases with a higher SCD threshold, e.g it's halved for 2cm, and becomes negligible (almost 1/6) for 5cm. There seems to be a minimal positive bias for low HS in daily reconstructions, negligible for HSavg, but enough to introduce errors in dHS1.

[Answer]: We thank for this input and just want to remind that HSmax, despite being one single value, usually is the result of snowfall accumulation over several days or months. Regarding dHS1, we calculated the number of snow days for thresholds of 1, 2, 5, 10 and 30 cm (dHS1, dHS2, dHS5, dHS10, dHS30). As you write above, the BIAS decreases with increasing threshold for snow covered days. We added some bits to the discussion and included Table A1 in the appendix.

Initially, we also had the suspicion that the positive BIAS for dHS1 could arise from the fact that our methods are rounding predictions to the nearest integer and that accordingly values between 0.5 and 1 cm are still contributing to dHS1. However, after having a look at dHS2 which is still positively biased for all methods except of BCS, we believe it is unlikely that rounding to integers is the source of the large positive biases and did not do further investigations towards that direction.

Maybe the authors could provide a table in the applicability and limitations section to summarize the reconstruction methods evaluated? Showing e.g. "Best/Good", "OKish", and "Not recommended" methods, depending on parameter (daily, seasonal, ...) and network density? I know, this might involve some arbitrary choices, but could be useful nonetheless. And, this could highlight the dHS issue and warn against using IDW for snow - in addition to what the authors already write in the text...

[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.