Author response to Comment RC2 on gi-2021-16

Thank you Mr. Lopez-Moreno for your time to assess the manuscript 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.

The manuscript presents the comparison of different methods to fill gaps in snow series. This is a task that has generated many doubts to snow reseachers and this paper provides very useful information for readers. The paper has a clear structure, is well written and conclusions are sound and clear. Therefore, I recommend the publication of the article, with just a few comments that authors may consider to prepare a revised version of the manuscript.

1- in my opinion, it would be interesting to present some analysis to show how differenet methods are suitable to fill gaps of different length, as probably there will be important differences among accuracy scores and methods.

[Answer]: We originally planned to include different gap lengths in our analysis but decided to use only a unique gap length (one winter) for our method comparison. Additional gap lengths require decisions regarding the training period and how the gaps are created. Furthermore, the amount of results blows up and it will become more difficult to compare methods easily. This is why we decided to keep the study simple and only consider long gaps of a whole winter of missing data.

2- As you can include other categorical variables in the Random Forest, authors can test or at least discuss other possible predictors that might refine the results. In example clasiffy if gaps occur in low/average/high snow years; or it existed different dominant weather types or atmospheric patterns in a given year when gaps must be filled.

[Answer]: During early phases of manuscript preparation, we tested different versions of Random Forest interpolation where we also included other categorical variables such as binned quantiles of the mean of all used predictor stations. These versions did not add improvement to the simpler RF version we present in the paper. However, it is worth noting that this kind of predictors are also possible to be used in Random Forest interpolation and might be useful in certain circumstances. We added a corresponding sentence to the discussion.

3- Authors may discuss to which extent the use of more physically based (when possible) may improve the error estimators compared to the degree day model. Researchers from CEN uses adjusted crocus/safran simulation to fill gaps in snow series (see https://doi.org/10.1002/joc.6571 as example). In a similar way bias corredted ERA-land series could be used for some areas, or used as "virtual" best correlated stations.

[Answer]: We added a paragraph to the discussion where we cover the potentially beneficial use of more physics-based snow models such as SNOWPACK or CROCUS if the necessary data is available.

Additionally, we added a paragraph in which we discuss the possibility of using reanalysis data instead of neighboring station data for the spatial interpolation methods or as input for the snow model.

Despite of the fact that we believe it is out of scope of the paper to assess estimations of snow depth data from reanalysis products, we did a quick assessment of the potential using ERA5-land reanalysis data. We tested three different schemes for gap reconstruction:

1. ERA5-land snow depth data without any bias correction from the closest grid point (ERA5nobc)
2. ERA5-land snow depth data from the closest grid point with the same mean ratio bias correction applied to the BCS method (ERA5mrbc)
3. ERA5-land snow depth data from the 9 surrounding grid points as input to the RF method (ERA5rf)

We applied the same leave-one-winter-out cross validation at the evaluation stations as for the other methods. The scores for HSavg, HSmax and dHS1 are listed in the following table:

		ERA5nobc	ERA5mrbc	ERA5rf
HSavg	**r2**	-10.51	0.84	0.86
	RMSE	73.69	8.68	8.14
	BIAS	53.13	-0.02	0.25
HSmax	**r2**	-3.99	0.5	0.78
	RMSE	100.16	31.69	21.08
	BIAS	72.21	-21.62	-6.84
dHS1	**r2**	-1.74	-0.6	0.52
	RMSE	91.51	69.74	38.05
	BIAS	78.51	56.46	27.82

These single grid point approaches perform clearly inferior to all tested methods in the paper. The Random Forest downscaling approach (ERA5rf) can compete with IDW for HSavg and HSmax but is not able to reach the performance of the other methods. dHS1 is more biased than any other method with BIAS of +27.8 days. For the rare case, that meteorological data is also missing when snow data is missing, it would be interesting to first downscale temperature and precipitation and then use that as input for a snow model. However, as stated in the discussion, the probability that there are gaps in temperature and precipitation is lower than for snow. In case a local measurement is available as input for a snow model, we believe this will always be superior to the reanalysis driven approach.

Moreover, reanalysis products often suffer of an elevation dependent precipitation or temperature bias, which is crucial in regard to a highly temperature sensitive variable as snow cover. Additionally, higher resolution reanalysis products like ERA5 are not available for the historic gaps in the first half of the 20th century. Nevertheless, it would be interesting to assess the potential of different reanalysis products for snow depth reconstruction in a follow up study.

Looking forward to see your revised manuscript,

Ignacio López-Moreno