Interactive comment on "A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967–2015" *by* Jordi Bolibar et al.

Ben Marzeion (Referee)

ben.marzeion@uni-bremen.de

Bolibar et al. present the results of a new approach to reconstruct glacier mass balances at times and/or locations where meteorological conditions (and some topographical information) are known, but no observations of glacier mass balance exist. Their approach, based on a neural network algorithm, adds considerable diversity to the existing group of reconstruction methods. The thorough validation of the results leads to great confidence in the robustness of the method.

Except for some minor issue listed below, the manuscript is very clear and easy to follow. The data set produced and presented here will be of great use for the community.

C1

I particularly appreciate the great care that has been taken in documenting the test for overfitting in the supplementary material.

I recommend publication once the authors have gone through the list of questions/suggestions below.

Specific/minor comments:

- P1 L9: please specify "1 \sigma" instead of "\sigma" for clarity

- P1 L10: the "moderately" should only apply to the 1980s, I think

- P1 L10: avoid line break within negative number

- P1 L12: unclear, what "this period" refers to

- abstract: why are no uncertainties given for the values of the different massifs? (also concerns the conclusions)

- P2 L8: "these points" refers to the points of MB measurements, but this reference is not very clear here; also, it's not the points that show nonlinear variability, but the measurements at the points; suggest to rephrase

- P2 L23: there more four global parameters in the Marzeion et al. (2012) model, and I wouldn't necessarily say they were "optimized", because that "optimization" was very subjective...

- Fig. 1: the figure certainly works well for presentations etc., but I'm not sure it is necessary here, since the text describes very well what is done, and there is little to be gained from the figure.

- P3 L14-15: it would be great if you can add a sentence or two here, specifying how any difference in the altitude of the glaciers' centroids and the reanalysis grid points were treated (lapse rates or similar?)

- P4 L22 or lower: It might be worth pointing out/discussing that the density of observa-

C2

tions used in the LOGO cross validation is denser towards the end of the reconstruction interval, when presumably, also the quality of the meteorological data are higher, such that the uncertainty of the methods might be underestimated for the (roughly) first half of the period. I also wonder if/how this interferes with your assessment of the model's ability to reconstruct the more neutral MB values during 1967-1984?

- Fig. 2: since there are so many lines, it is somewhat hard to see the distribution. Particularly in the lower panel, a histogram for showing the distribution of the accumulated values (vertically, to the right of the panel) would be quite interesting. It would be possible to see, e.g., how/if the area weighted mean differs from the "ensemble" mean and/or median, if the distribution is (a)symmetric, etc. Just a suggestion to consider.

- Fig. 3: why are no uncertainties included for the decadal averages?

- Fig. 4: great figure! But a bit busy (just visually); would it be possible to mute the background image a bit (and then perhaps change the text color to black) so that the colors of the glaciers stand out more?

- P8 L16: it's more than three parameters: one local (the temperature sensitivity) and four global ones (precipitation correction factor, precipitation lapse rate, temperature threshold for solid precipitation, and melt temperature threshold); see Figs. 4-7 in Marzeion et al. (2012)

- P8 L22: perhaps clarify that the 38 glaciers are not the global sample used for calibration

- P8 L31: I believe that the CV results in the Marzeion et al. (2012) study are also influenced by the global "optimization" (see above) of the four parameters; probably, a focus on the Alps would have led to a different parameter choice, and hence different CV results.

- P10 L1 and following: another reason for the different behavior around the 2003 "break point" might be that the Marzeion et al. (2012) model, by construction, cannot

capture the lasting effect that the extreme 2003 year may have had on albedo; while your model may be able to capture this (I guess – I'm not sure) by essentially taking the time as an additional predictor?

- Fig. S2: would it be possible to re-arrange the legend such that it is easier to compare the "B" to the "M" lines (e.g., shift the lowest line in the legend to the right)?

_____