Review of: "Modelling the Point Mass Balance for the Glaciers of Central European Alps using Machine Learning Techniques" by Anilkumar et al.

Summary

In this paper, the capabilities of different machine learning (ML) models in predicting point glacier mass balance are explored. The used data is composed of monthly meteorological data from ERA5-Land together with direct mass balance measurements in Central Europe from the Fluctuations of Glaciers database. The study is an important next step to explore which ML models are most suitable for applications of mass balance estimates. Further, they assess the data required for the different models and the importance of each meteorological variable. Both are very interesting and important questions for the potential future use of ML models in this field, also in light of increasing data availability in the future.
The study is well designed, but I think parts could be improved to make the results more solid and the manuscript easier to follow by the readers. I have divided my proposed changes into General Comments and specific/Line by Line comments.

General Comments

- **GC1**: I think it would be good to give more information on the values of the used mass balance observations (e.g. How is the distribution? Are they located in the yearly ablation regions of glaciers or also some in the accumulation regions?). Also, why are only annual mass balance observations used and no seasonal ones? This probably could improve the analysis of Feature Importance performed separately for accumulation and ablation months.

- **GC2**: If just the raw ERA5-Land data is used as input it is probably hard to asses feature importance due to the very complex topography which is poorly represented. In general, how did you deal with the downscaling of the meteorological data to the glacier location? In particular, how do you deal with the height difference between the ERA5-Land grid point and the glacier elevation or how do you deal with poorly resolved precipitation? (Could this be an explanation for why you could not find the expected importance of precipitation during the accumulation months?)

- **GC3**: Results sections 3.2, 3.3, 3.4, 3.5, 3.6: The last sentences of the first paragraph are not needed and could be incorporated at the end of the sentences where relevant things are discussed, e.g. '(Fig. 3).' at the end of the sentence, like is done in L307. This makes it easier for the reader to check your described findings by themselves in the plots. In the second paragraph, you can point to that this information is available in the supplementary in more detail.

- **GC4**: To make it easier for the reader to interpret the Figures you could include subfigure tags (e.g. (a), (b), (c), …) and describe in the Caption more precisely what is shown in each subfigure. Also, increase the font size where needed.

- **GC5**: You should use the same units in the text and figures, e.g. in the text L241 it says RMSE value of 1.071 mwe, but in Figure 3 the y-axis shows 1071 (with no unit given).

Specific comments

- **L190**: define which months are accumulation months and which months are ablation months, should be done earlier in the manuscript (is only defined in L335)

- **L204**: How is stabilizing the training metrics defined? We can not see this from Figure 2, maybe include a similar subplot as the right one for training performance.

- **L207**: Also here, how is stabilizing defined? 'This suggests that all models have successfully fit the data.': Doesn't it only shows that the results do not get better if we give the models more data than 50%, and it tells us nothing about how successful the fit is?

- **L209**: also here a plot suggested under L204 would be helpful to see the explained increase in training MAE

- **L217**: How do you see this? (Smaller box in Figure 2 left?)

- **L240**: Instead of 'This is depicted in Fig 5.' just right '(Fig. 5).' at the end of the sentence

- **L247**: define somewhere in the manuscript what are 'ablation meteorological variables'

- **L261**: is 'cost' the same as 'penalty'? If so you should be consistent and use one or the other throughout the manuscript.

- **L304**: How do you conclude this ranking? From Figure 3 and Figure 4, it looks like RF and SVM are closer than SVM and NN.

- **L326**: To which graphs are you linking here? Maybe include the figure number.

- **L348**: Probably you could not find the expected importance of precipitation because it is poorly resolved in the climate input data (see GC2).

- **Table 1**: Why is 'Number of trees' listed two times?

- **Figure 2**: See GC4. In the caption also explain which quantiles are shown in the box plot on the left. And explain how the two plots are connected (are the yellow boxes on the left representing the quantiles of the lines in the right plot?) Add the unit to the y-axis. Currently wrong caption: 'Training and testing RMSE (in mm we) and r values for varying the size of the training dataset for each of the models:' but only shown is MAE.

- **Figure 3**: See GC4. In caption: e.g. how are training and testing data split in this plot, 70%/30% or different, include (a), (b), (c) and (d) and explain also in the caption which performance measure is shown in which subplot. Add units to the y-axis where needed.

- **Figure 4**: See GC4. Maybe you can include the information of Figure 3 into this figure and delete Figure 3 (e.g. "RMSE: 0.95/1.08 mwe" and include a legend at the empty subplot space lower right with "RMSE: Training/Testing"). For the y-equations don't write y=0.744x + (-338.433) instead write y = 0.744x – 338.433. Is the high precision of numbers with three decimals meaningful for the RMA regression?

- **Figure 5**: hard to distinguish in the legend what are the solid lines and what are the dash-dotted lines. In the caption mention which test score is shown and explain briefly what the negative scaled RMSE is.

- **Figure 6**: In the caption mention which test score is shown and explain briefly what the negative scaled RMSE is.

- **Figure 7**: increase the font size, In Caption mention which test score is shown. Also include the test score name in the y-axis (currently only 'Test score').

- Maybe you could combine Figures 5, 6 and 7 into one Figure.

- **Figure 8**: increase the font size. Because the x-axis is limited to 13 maybe add the numbers in the plot for features which go beyond this limit. Maybe include the abbreviations of meteorological variables in the caption or the text, so you can understand the plot without having a look in the supplementary. And you can also use the abbreviations in the result sections.

- **Supplementary S1**:
  - general: give more meaningful names to the individual sheets
  - sheet3: no explanation of what is shown on this sheet, include references in the text or delete this sheet