

Review of revised manuscript for “Using an ensemble of artificial neural networks to convert snow depth to snow water equivalent over Canada” – Konstantin Franz Fotios Ntokas, Jean Odry, Marie-Amelie Boucher and Camille Garnaud.

Now: “Investigating ANN architectures and training to estimate SWE directly from snow depth”

General Comments

Dear authors and editor,

I am pleased to see the authors have taken my comments into consideration, and I am happy with their reply for all comments that I do not re-discuss below. The manuscript has clearly improved its structure and readability, and I (again) want to state that the method presented here to estimate SWE is a good contribution to scientific progress that deserves to be published. In fact, most of my comments do not relate to the validity or robustness of the methods, but to the contextualisation of the approach, its applicability, and some statements made by the authors. In my opinion there are still some issues that need to be seriously addressed before publication. Most of them relate to the answers you have given to my previous report, but a couple new minor issues have arisen from re-reviewing the text. Line numbers refer to the **track changes version** of the manuscript.

- 1) I am still not convinced about how you contextualise your study and your aims. You have to convince the reader that your manuscript/method provides a relevant improvement with respect to Odry et al 2020. It is not just about writing “these are the knowledge gaps”, as you do in line 84, but also why those knowledge gaps must be tackled. The knowledge gaps must logically arise from and be linked to the introduction (especially the paragraph before describing the results of Odry et al 2020). Something like: “This is a follow-up study [..]. While they did XXX, they did not consider PPP, so here we further test YYY and ZZZ, which is important because NNN. We hypothesize that (hypothesis 1) and (hypothesis 2). Furthermore, we aim to...”. Also, line 92 is confusing when you write “We also take the opportunity to...”. It reads as if that is a new aim and another third dataset, but to me it sounds like a repetition from the previous sentence. Please think about this carefully from a reader’s perspective.
- 2) Thank you for your explanation about why not to use snow depth time series. However, your argument that you want to use data that is only available in near-real-time has made me realise that there is something inconsistent in your aims and method, and the following issues are linked to each other. You don’t have and don’t need real-time snow depth data, but you do need real-time precipitation data. Therefore, in what circumstances is the model going to be really applicable for operational use? I am guessing that you will only be able to use it at sites where there is real-time meteorological data available, and then a single snow depth measurement is provided at some point in time. From my ignorance about operational use over Canada, is this a realistic application? If so, this must be stated more clearly somewhere. This links to my comments on the aims of the study and the method. If not, you should reconsider what the aim of this approach to estimate SWE is (I think it can be very valuable for several applications, but this should be clearer in the text).
- 3) Related to the previous comment, the ANN is trained, validated and tested with ERA5 meteorological data, but these are not available in real time (you discard snow density from ERA5 because it is not available in real time). Therefore, I am assuming in real-time operational use, only in-situ meteorological station data will be used. Therefore, we don’t really know if the ANN will perform well for the real-time operational application. You apply a lapse rate for temperature, but precipitation can also vary a lot between ERA5 and point

locations. The model will be trained from dynamics and features of reanalysis data, which can differ from station data. Again, if you want to keep the real-time operational use as one main aim for your method, then an independent validation should be provided with station data. Are there meteorological data available for some of your snow survey locations? If so, you should provide an additional validation for real-time use.

- 4) I have realised that it is not right to say that your method estimates SWE directly from snow depth (which is now even in the title, so I do not think it is accurate). You also need meteorological data. Given that, your method might be more comparable to a temperature index model, than to the simple regression models that you compare it with, which need only snow depth and simple geographical data (elevation, region, day of the year...). For instance, a recently published paper estimates SWE directly from snow heights (<https://doi.org/10.5194/hess-25-1165-2021>), but they really only need snow height and its temporal change. I think it should be stated, especially in the introduction and conclusions, why you decide to compare your ANN with simple regression models (Jonas, Sturm), given that your method requires more data, and it is then not surprising that it performs better. This should also be a limitation of the method, but even if I suggested that you write more about limitations, you only added that the ANN does not perform well for the very high and very low values of snow density. The amount of data required does not only mean “how large the data is” but also the type of data. In that sense, your method requires more data than other simple regression models.
- 5) Regarding Table 1, and your new statement in line 706-707. I agree that information on the short term time scales is relevant, especially due to fresh snow density effects. However, I think the justification to include short term accumulated precipitation comes rather from the effect on snow density. Even if it is not the target variable, the effects might be still “hidden” in SWE. Given that correlation between SWE and “n days precip.” increases, I find the choice of 10 days arbitrary, because it is not justified by the data. What is the effect on the score of explanatory variables if you choose (or add) n=3 or n=5 instead?
- 6) The new structure of section 3 and 4 is great, as well as Table 2 and 3! Much clearer and logical now.
- 7) After more thoughts on Figure 10, I think the histograms on (a) and (b) provide little information. It is hard to compare simulations vs observation, but also (a) vs. (b). I think a scatter plot would be a lot more informative. Since you do not mention outliers here (or very high values) you could cut the x and y axis to 2000mm. I know the scatter is already shown in Figure 15, but here it would provide zoomed in information. Similarly, the scatter plots in Figure 15 should be cut to 4000mm, or even 3000, as long as it is stated in the text that some outliers (probably 0.0001%) are outside the figure limits. Further, why is the origin of Figure 15 not at zero-zero? Also, include “colour shows scatter density” in the caption. Similar applies to Figure 3a,b,c,e, the x axis should be cut to where the bins are not visible anymore.
- 8) Line 676-678. I agree, but then it might be worth adding the Odry 2020 configuration in Table 6, for comparison.
- 9) Finally, I suggest that the github repository to reproduce the study is a little clearer. It took me a long while to understand the logical order of the codes, and what the folders are and where they come from. A clearer README file explaining the workflow (in addition to the figure in “OO.Overview.pdf” would be highly appreciated.

Other technical corrections:

- Table 5 is not referenced in the text anymore. You “lost” it when crossing it in line 555.

- I like your clarification (to me) about how the MLP ensemble works: “For clarification, when simulating the test data set for each record, the snow class is determined and the associated MLP ensemble is taken in the multiple MLP ensembles model. This returns one ensemble for one record, as in the single MLP ensemble model.” It should be included in the text.
- Line 367: I think it is better “From snow depth, snow density, total precipitation, and temperature, we obtain the following explanatory variables.”
- Line 425: Rephrase, it is hard to read. Maybe “the characteristics shown in Table 2 (first six rows in Options column), are tested...” ?
- Line 445: Should be “Sec. 4.1.3 and 4.1.4.”
- Line 566: Swap order of MMLP and SSMLP, for consistency with the rest of the manuscript where SMLP is shown first.
- Line 567: Take (not takes).
- Line 575: “All performance metrics are smaller, except MBE.”
- Line 599: Should be “ephemeral snow class in Fig. 13”.
- Line 689: Perhaps reiterate here what the large gain in reliability is based on (what figure or metric).
- Line 703. A dot instead of comma after “analysed”.
- Line 709: “such as” instead of “e.g.”.