Authors' specific response regarding revision #2 to comments by Anonymous Referee #1

April 22, 2021

Black text: Reviewer's comment

Blue text: Authors' response; The identification of lines, figures and tables refer to the version with track changes.

1 General Comments

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 rereviewing the text. Line numbers refer to the track changes version of the manuscript

The authors would like to thank the referee #1. The review will surely improve our study further. In the following, we address each specific comment and explain how we incorporated them into the manuscript.

2 Specific Comments

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

gap 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.

We thank the referee for the critical thinking from a reader’s perspective. We rewrote the paragraph in line 80-100, focused on the gaps in Odry et al. [2020] and tried to logically derive the aims and goals of the current study.

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).

   We address comment 2 and 3 together after comment 3.

3. ln 54: 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 point4) 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.

We would like to clarify the concept behind the study and address comments 2 and 3 together.

The ANN ensemble is trained on manual snow surveys where snow depth and SWE are both available. Furthermore, the meteorological data for training comes from a reanalysis, in this study from ERA 5. The main goal of the study is to test the applicability of ANN ensembles to the conversion problem. With this respect, it appeared coherent to test it using the longest historical record possible, using meteorological data that is consistent in time (i.e. reanalysis) so that the differences in performances can by mainly attributed to the ANN model itself. The actual application in real time will be a second step that is more specific to the used atmospheric and hydrological forecasting system. In operational use, the recently available Regional Deterministic Reforecast System (RDRS; Gasset et al. [2021, in revision] in revision) would be used. Note that this product was just released and was not available in time for this study. RDRS has similar dynamics and physics as the operational Global Environmental Multiscale Model (GEM) used operationally at Environment and Climate Change Canada. The historical records are much shorter for this kind of data (snow depth and SWE) and also it is not consistent for the whole period. Some testing would be required before RDRS can actually be used, but it will be more specific to the forecasting system.

When simulating SWE, in-situ snow depth measurements could be taken (e.g. the sonic sensors provided by Meteorological Service of Canada and ECCC [2020]). As meteorological data, one could use an operational "nowcast" from an atmospheric model that include a land data assimilation system such as the Canadian Land Data Assimilation System (CaLDAS; Carrera et al. [01 Jun. 2015]). CalDAS is forced by real-time precipitation analyses from the Canadian Precipitation Analysis (CaPA; Fortin et al. [2015]), which combines simulated background precipitation fields with observed data (in situ and radars). Furthermore, the proposed method can be applied onto assimilated snow depth data in CaLDAS. Currently in CaLDAS, only snow depth data is assimilated and subsequently converted to SWE using the simulated density to initialize the land surface scheme. The proposed method would allow for two important upgrades: First, it would allow to assimilate snow depth data (converted to SWE) as well as SWE data, thus increasing the quantity of assimilated observations, and second, it would avoid using the simulated density, which is very hard to simulate accurately. In the introduction (line 80-81), we now explain that we built a model to estimate SWE from in-situ snow depth measurements and several indicators derived from gridded meteorological time series. The information about operational use as stated in this response is given in the conclusion, line 689-701.

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.

We wanted to emphasis in the title the new finding of the current study that SWE is the direct output of the ANN ensemble model. We think that our model is probably an intermediate between snow modeling and regression model. Neural networks are in fact multivariate regressions. So they are comparable to regression models, they are just a more complex regression. The model uses meteorological time series, but derives certain indicators from them and cannot be used in the same way as a degree-day model because it does not dynamically accumulates nor melts snow. The only purpose of this model is to convert snow depth to SWE, and it was intended like that (not as a replacement of a snow accumulation and melt model). Actually, the ANN only sees the indicators, not the time series. Theoretically, it would be possible to build a regression using the exact same inputs used by the proposed ANN, but it would probably be more complicated to deal with the data (inter-correlation, normality, non-linearity,...). ANN is known to be more powerful to identify non-linear relationships among variables than regression. However, for less confusion, we deleted 'directly' in the title and reformulated it in line 84. Further, we added some information in the conclusion about model limitation regarding data requirements in line 681-683.

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?

   We agree with the referee that the choice of 10 days is taken arbitrarily and further investigation of the explanatory variables needs to be done to cover short term effects with respect to SWE as the new target variable. This information has been added to the manuscript in revision #1 in line 667-674. This study took the same input variables as the previous study Odry et al. [2020] (except snow density from ERA5) and changed the output variable to identify the effect of that change. We do not believe that a random calculation of n=3 or n=5 gives us a lot of information at this point, and thus will not perform them.

6. The new structure of section 3 and 4 is great, as well as Table 2 and 3! Much clearer and logical now.

   We would like to thank the referee for his helpful suggestion and are happy that it made the manuscript clearer.

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.

   We agree that the histograms in Fig. 10a and b provide little information and we would like to take up the suggestion of the referee. We show a zoom in of the scatter plot with adjusted axes in Fig. 10a and b and give some information in the caption. Similarly, we adjusted the axes in Fig. 15 and state the portion of outliers not shown in the caption and line 619-620. Furthermore, we updated the x-axis of Figure 3a,b,c and e and give the information about non-shown outlier in the caption and line 356-358.

8. Line 676-678. I agree, but then it might be worth adding the Odry 2020 configuration in Table 6, for comparison.

   In Table 6, we added the configuration of the ANN ensemble of the previous study, proposed by Odry et al. [2020]: The table is linked to the conclusion statement in line 644.

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.

   We refined the README file in the github repository to improve the guidance through the code structure.

3 Technical Corrections

All the technical corrections were made, and we indicate the lines where they appear in the revised version with track change.

- Table 5 is not referenced in the text anymore. You "lost" it when crossing it in line 555. line 533-534
- 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 426-428
- Line 367: I think it is better "From snow depth, snow density, total precipitation, and temperature, we obtain the following explanatory variables." line 371-372
- Line 425: Rephrase, it is hard to read. Maybe "the characteristics shown in Table 2 (first six rows in Options column), are tested..." ? line 417-418
- Line 445: Should be "Sec. 4.1.3 and 4.1.4." line 434; note that this was a problem with latexdiff in the previous revision. Thus, it does not appear in blue in the current revision.
- Line 566: Swap order of MMLP and SSMLP, for consistency with the rest of the manuscript where SMLP is shown first. line 539-540
- Line 567: Take (not takes). line 540
- Line 575: "All performance metrics are smaller, except MBE." line 547
- Line 599: Should be "ephemeral snow class in Fig. 13". line 569
- Line 689: Perhaps reiterate here what the large gain in reliability is based on (what figure or metric). line 656
- Line 703. A dot instead of comma after "analysed". line 668
- Line 709: "such as" instead of "e.g.". line 674

References

M. L. Carrera, S. Bélair, and B. Bilodeau. The canadian land data assimilation system (caldas): Description and synthetic evaluation study. *Journal of Hydrometeorology*, 16(3):1293 – 1314, 01 Jun. 2015. doi: 10.1175/JHM-D-14-0089.1. URL `https://journals.ametsoc.org/view/journals/hydr/16/3/jhm-d-14-0089`$_1.xml$.

V. Fortin, G. Roy, N. Donaldson, and A. Mahidjiba. Assimilation of radar quantitative precipitation estimations in the canadian precipitation analysis (capa). *Journal of Hydrology*, 531:296–307, 2015. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2015.08.003. URL https://www.sciencedirect.com/science/article/pii/S0022169415005624. Hydrologic Applications of Weather Radar.

N. Gasset, V. Fortin, M. Dimitrijevic, M. Carrera, B. Bilodeau, R. Muncaster, E. Gaborit, G. Roy, N. Pentcheva, M. Bulat, X. Wang, R. Pavlovic, F. Lespinas, and D. Khedhaouiria. A 10 km north american precipitation and land surface reanalysis based on the gem atmospheric model. *Hydrology and Earth System Sciences Discussions*, 2021:1–50, 2021, in revision. doi: 10.5194/hess-2021-41. URL https://hess.copernicus.org/preprints/hess-2021-41/.

Meteorological Service of Canada and ECCC. Historical climate data, 2020. date of access: Oct 2019, https://dd.weather.gc.ca/climate/observations/.

J. Odry, M. A. Boucher, P. Cantet, S. Lachance-Cloutier, R. Turcotte, and P. Y. St-Louis. Using artificial neural networks to estimate snow water equivalent from snow depth. *Canadian Water Resources Journal / Revue canadienne des ressources hydriques*, 0(0):1–17, 2020. doi: 10.1080/07011784.2020.1796817. URL https://doi.org/10.1080/07011784.2020.1796817.