Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-593-AC2, 2020

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Application of machine learning techniques for regional bias correction of SWE estimates in Ontario, Canada" *by* Fraser King et al.

Fraser King et al.

fdmking@uwaterloo.ca

Received and published: 3 April 2020

Reviewer 2

General Comment:

This very interesting paper of King, et al. compares different methods for reducing the bias between in situ measurements of SWE and the gridded SNODAS estimates for the region of Ontario. The correction methods include simple mean bias subtraction, linear regression and machine learning methods. The paper is very well written and it is worth to be published after some minor changes. Some comments and recommendations:

General Comment Response:

We thank the reviewer for their comments, and we will work to incorporate their suggested changes to improve our currently submitted manuscript. Our responses to each of the reviewer's questions/comments is included below.

Specific Comment 1:

First of all and most important the applied machine learning methods are not described at all and references are missing. I don't think that all readers of this journal are familiar with Decision Trees (DT) and Random Forest (RF) methods. Therefore a short description should be included, especially explaining the RF model in more detail, which shows the best results, and what's the difference to the DT models. Related to that comment, it doesn't make too much sense to mention on page 5 (line 30) that you run the model with a forest size of 100 trees and tree depth of 15, when you don't explain what that parameter mean.

Specific Response 1:

We thank the reviewer for this comment, and agree that additional details should be included in the text which further describe the methodology behind the decision tree (DT) and random forest (RF) techniques we employ in this work. We have updated the document to include further references/details regarding what these techniques are and how they operate, along with further descriptions of what parameters like forest size and tree depth mean with respect to the RF model in section 2.3 of the manuscript.

Specific Comment 2:

Additionally, there are some points which are not clear to me and which should be comment clarified before publishing the paper: You didn't explain how you handled the scaling issue when you compare point data and gridded data (up- or downscaling?). Since you could identify a change in the bias between the first and the second half of the period, it would be reasonable to split the analysis into these two periods and fit

different models and take 2 different means separately for each period.

Specific Response 2:

In our analysis, we compare gridded estimates of SWE from SNODAS (1 km resolution) to snow survey estimates (which is essentially point data taken over 10 m). Due to the relatively high spatial resolution of SNODAS, along with the fact that the in situ measurement sites are taken at distances > 1 km from each other, we do not compare multiple in situ points to a single grid cell. This allows us to complete a simple point to grid cell comparison where we assume the snow survey SWE estimate is representative of the wider, containing grid cell. The snow survey sites are selected to generally be representative of the area around it and are not just random point measurements which would contain higher variability in their estimates. However, this assumption of representativeness across the grid cell introduces additional uncertainty, as SWE is highly variable at even small spatial scales, and we have therefore included additional details in the paper to make these uncertainties clearer to the reader. Furthermore, we agree with the reviewer that due to the change in the intensity of the bias post-2014, a description of how the bias correction models perform over these separate two periods would be interesting and complimentary to our analysis. Therefore, we have also updated the results section 3.3 of the paper with the results of this test.

Specific Comment 3:

On page 3 you specify the 383 locations with in situ measurements. In line 14-15 you write that an average SWE is estimated taken from 10 fixed sampling stations. What does this mean? Is this the average SWE for Ontario estimated from 10 stations, or is this the average for each of the 383 stations taken from the 10 surrounding stations??

Specific Response 3:

What we are referring to on page 3 is the method by which in situ measurements are retrieved (snow survey), where a sampling location is selected and then 10 point

measurements are taken using a snow coring device over approximately 10 meters at that location. These 10 SWE measurements along the snow survey are then averaged together to provide a single SWE estimate for that location. This is the technique used at all 383 in situ measurement sites. We now include additional details on how these measurements are retrieved to add further clarity to the reader in section 2.1 of the manuscript.

Specific Comment 4:

Page 5: You should mention that the period of 1981-2010 is used for calculating the climatology, which is not clear.

Specific Response 4:

We thank the reviewer for noticing this detail, and we now make the temporal period used for the calculation clear in the paper on page 5 (section 2.2.2).

Specific Comment 5:

Also, you should explain why you have used the difference between the precipitation estimates from NRCAN and the SNODAS! It would be interesting to see the results if you would include actual meteorological observations as predictors (for example available at: https://data.noaa.gov/dataset/dataset/globalsurface-summary-of-the-day-gsod, provided by the National Centers for Environment Information). I could imagine that in that case the importance of these variables would not be neglectable and could further improve the bias correction.

Specific Response 5:

We thank the reviewer for the comment; we have also considered this option; however, we have chosen to limit meteorological data to basic monthly climate normals. Analogous to the choice of model complexity, there is always a trade-off between accuracy, complexity and the risk of over-fitting. Using a large set of predictors requires a more complex model, which increases the risk of over-fitting. Therefore we have chosen to

only include monthly normal surface temperature and precipitation, as these two variables are usually readily available and characterize the type of climate reasonably well. The rational behind including climate variables was that, on average, snow characteristics (like density, albedo, ice content) vary between different climates. It is true that these characteristics would be predictable (to some extend) from the actual evolution of these meteorological forcings; however, the processes that govern such characteristics are very complex and involve long-term memory effects, which would require a much more complex model (like an LSTM), which would approach the complexity of physical snow models. Considering the data requirements and complexity of this approach, we believe that the use of monthly normals represents the best compromise. As for the reason, the difference between SNODAS average precipitation and NRCan normals was used, rather than total precipitation from NRCan (or SNODAS): this choice was made because notable biases in the precipitation fields used by SNODAS over Canada were found early on in the analysis, and it appears obvious that the size of these biases would have a first-order effect on the resulting SWE bias in SNODAS. At the same time, in order to reduce the number of input variables, we did not want to include multiple, possibly redundant, precipitation variables.

Specific Comment 6:

Page 7: When you write in 3.2.1 about mean bias, I suppose that this mean bias is calculated as the average of the mean bias of all stations? Similar to that I'm a bit confused about what you write on page 8 regarding SLR. I was assuming that you fit a regression model for each station individually. But that seems to be not the case, otherwise I could not understand why there should be a bias overcorrection. It would be nice if you could clarify this, whether you fitted separate models for each station or not.

Specific Response 6:

The reviewer is correct in that the mean bias is calculated as the difference between

the average SWE across the full temporal period for SNODAS minus the average SWE for all 383 survey sites (ie. the two dashed lines in the timeseries Figure 1.b). The reviewer is also correct in that we did not fit a SLR model to each station, but instead trained a single model across all survey sites for our full temporal period (as well as the partitioned upper and lower regions in Figure 2 to see if multiple models showed improvement; and we found they did not). The bias overcorrection in the linear techniques like SLR stems from the fact that the SLR is attempting to model a linear relationship across all years which is problematic due to the nonlinearity in the bias introduced post-2014. This results in an overcorrection in some periods and an undercorrection in others.

Specific Comment 7:

Although you wrote in the beginning that you took 75% for training, you didn't mention if all the calculated verification measures refer to the remaining 25% testing period.

Specific Response 7:

The reviewer is correct that the calculated verification measurements on model performance when performing validation testing refer to the remaining 25% of the dataset, we now make this clearer in the text in section 3.3.

Specific Comment 8:

In the legend of Figure 2 you write Lower and Upper. Shouldn't it be southern and northern?

Specific Response 8:

We thank the reviewer for noticing this naming discrepancy and we have updated the Figure 2 legend to show Southern and Northern instead of Lower and Upper.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-593, 2020.