Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-593-AC3, 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: 28 April 2020

Reviewer 3

General Comment:

This work evaluates several bias-correction methods (simple subtraction, single and multiple linear regression, decision trees, and random forests) to SNODAS, resulting in a new data product that shows improved fidelity to in situ observations. The authors further develop a simple water balance analysis that exhibits the improved consistency of the inferred melt of the corrected model to streamflow observations. This work rep-

resents important progress to advancing the application of machine learning to water resources management in regions of snowmelt-dominated streamflow regimes.

General Comment Response:

We thank the reviewer for their comments and suggestions for improving the manuscript, and we will work to incorporate these changes into the article. Our responses to each of the reviewer's questions/comments is included below.

Specific Comment 1:

The potential strengths of machine learning are highlighted but a justification for the selection of random forests (RF) is not particularly apparent. The authors mention applications of support vector machines and neural networks in geosciences detailed in Lary et al., 2009, a study of aerosol optical depth, but neglect to review specific literature around machine learning applications in SWE estimation (e.g. Wrzesien et al., 2017, Snauffer et al., 2018, Xue et al., 2018). A review of such advances is warranted.

Specific Response 1:

We thank the reviewer for their suggestion to include additional motivation behind our selection of the random forest technique for bias correction. As mentioned by the reviewer, this choice primarily stems from the strengths this technique has shown in previous literature for bias correcting data in the geosciences (Reichstein et al., 2019; Shen, 2018; Lary et al., 2009). However, we agree that additional motivation with respect to bias correcting SWE would be beneficial, and we have now included additional literature focusing on the application of random forest bias correction towards SWE datasets from Wrzesien et al., 2017, Snauffer et al., 2018, Xue et al., 2018, Zahmatkesh et al., 2019 and Lv et al., 2019, in section 1 and section 4.2 of the manuscript.

Specific Comment 2:

RF model structure and hyperparameter descriptions should be moved to the methods section. The authors mention RF is run with a forest size of 100 and maximum tree

depth of 15, but it is unclear how these hyperparameters were selected beyond a mention of "sensitivity tuning experiments". Generally hyperparameters should be tuned using a standard method (e.g. grid search, particle swarm optimization, evolutionary strategy, etc.) on each test split and reported accordingly. Is the maximum number of terminal nodes for a given tree specified or are the trees allowed to grow to full extent?

Specific Response 2:

We thank the reviewer for this comment and question. During the model training phase of our analysis, we experimented with a variety of values for forest size and maximum tree depth to find a balance between model accuracy and run time efficiency. This sensitivity experiment was performed through a brute-force grid search approach of nudging each parameter value to find a set of parameters which exhibit both high general performance (low RMSE and bias), and an efficient RF model runtime. This test resulted in the selection of a forest size of 100, along with a max tree depth of 15. As per the maximum number of leaf nodes for each tree, this was left to allow each tree to grow to its full extent. We have moved some of the general model structure details (along with the hyperparameter descriptions) into the methods section and have also included further details on how the hyperparamaterization was performed in the same section (2.3) to add further clarity.

Specific Comment 3:

RF and DT are stated to be trained on 75% of the data and evaluated on the remaining 25% test set, but are also evaluated using a 10-fold cross-validation, resulting in an average RMSE reduction of 4.7 mm. The change to bias is unclear, as is the motivation for using both a 75-25 and 10-fold split structure. Since you've appropriately gone to the effort to run a full 10-fold cross-validation, why aren't you just using these results?

Specific Response 3:

When training and running our RF model, we used a 75/25 split (75% training and

25% testing) of our dataset to help mitigate against potential model overfitting while maintaining good model performance (low bias and RMSE). We experimented with a variety of values for the training and testing set and found the 75/25 ratio provided a balance between strong model performance, and a large test set of data to compare against. This train/test ratio also aligns with standard RF test sizes as mentioned in the SciKit-learn documentation (Pedregosa et al. 2011). After calculating our results, in order to further mitigate against potential model overfitting and to evaluate model performance on unseen data, we then went ahead and employed an additional 10-fold cross validation which resulted in an average RMSE reduction which was complimentary to our 75/25 structured model. Our 75/25 model was therefore used as the primary structure for our results since it was the original model developed and employed for bias correction, reported similar results (< 1.5 mm SWE difference) to our followup CV experiments, and was overall much more efficient to run.

Specific Comment 4:

The manuscript would be strengthened with a description of the efforts you've undertaken to mitigate temporal and spatial auto-correlation in your training and test sets. The manuscript would be strengthened with further descriptions of the efforts you've undertaken to mitigate overfitting. A comparison of training and validation errors would be an appropriate way to do this.

Specific Response 4:

In order to mitigate against spatial auto-correlation, we broke the training and testing datasets spatially as seen in Fig. 2 of the manuscript into northern and southern regions, to evaluate model performance in areas with differing magnitudes of bias and station densities. With respect to mitigating against temporal autocorrelation, we use monthly averaging of the biweekly station data which does help to some extent, however in order to fully avoid issues with auto-correlation, we would need to employ a strategy of removing stations/periods which are consistently correlated, and this would

introduce new biases in the training dataset for our model. Overall, stations are usually selected in a representative manner by the Conservation Authorities who collect measurements throughout the region, and we trust in the integrity of the station network to help mitigate this issue.

Specific Comment 5:

In Table 2, what are Year Id and Month Id? Are you using straight numerical values, cyclical temporal sin-cos pairs, 1-of-c indicators (Bishop, 1995)?

Specific Response 5:

The Year Id and Month Id predictors are 1-of-c indicators (numerical values of 0 or 1) with 0 representing the absence of either a month/year and 1 representing the presence of a month/year.

Specific Comment 6:

The water balance analysis averages melt over a watershed associated with a given stream gauge, asserting the stream gauge provides a reasonable estimate of snowmelt while at the same time neglecting evapotransportation and rainfall (actually any precipitation). Such an assertion requires that evapotransportation and subsequent precipitation are not as significant a signal as snowmelt to runoff. This may be true, but it should be backed up by analysis and references, or minimally one of these. Baseflow should also be at a minimum mentioned.

Specific Response 6:

These are fair comments and we agree that the argument can be strengthened by quantitative data. We have conducted an analysis of the dominant hydrological components across all three catchment areas, based on climate normals obtained from NRCan/CFS for the period of 1980-2010. The figure is included in this response (Fig.1 below) and could be included in supplementary material if required. It shows that in all cases average liquid precipitation (rain) during the spring freshet season exceeds po-

tential evapotranspiration, so that it can be argued that snowmelt places a lower bound on the spring freshet volume. A nuance here is that the snowmelt peak estimated following the method of Erler et al. (2019) can (and does) exceed the streamflow peak due to routing delay within the catchment area. The peak of negative SWE differences (which is shown in Fig. 7 of the manuscript) is shown in the Figure for comparison: it is evident that the value is significantly lower than the former snowmelt estimate and does not exceed the streamflow peak. The reason is that negative SWE differences do not include water from additional snowfall during the melt period. Comparing SnoDAS SWE differences with those estimated from NRCan climate normals and streamflow, it is clear that the uncorrected SnoDAS values are unphysical, while the bias-corrected values appear reasonable. For a detailed discussion of the variables shown in the Figure and how they were processed, see section 3.2 and S2 of Erler et al. (2019); the Figure is analogous to their Fig. 2 and the datasets and methods employed are the same. The reason that this figure was not included initially is that it is based on climate normals for a period before our analysis period. Unfortunately the PET and snow depth data used in the figure are not available past 2010, so that it was not possible to update the figure. Curation of a new PET dataset (for just this figure) would be beyond the scope of this study.

Specific Comment 7:

You conclude that MBS and SLR exhibit an inability to capture year-to-year variability present in the bias, but interannual correlations are not present in the analysis. The ability of bias-correction methods particularly of the non-linear flavor to capture changes over time is arguably one of their greatest strengths, as simple offsets are more easily calculated, as you have done. A simple correlation calculation may serve as further evidence of the utility of the nonlinear method.

Specific Response 7:

We thank the reviewer for this comment and agree that the inclusion of interannual

correlations between in situ the bias corrected SWE datasets would further highlight the utility of nonlinear techniques. These results have been included in section 3.3 of the manuscript.

Specific Comment 8:

Fig 5 is hard to read with the scales and lines used, especially the in situ values, which are key to the plot. No description of shading used is given in the figure caption. Suggest changing line thicknesses/colors and/or adjusting scales, orientation, or paneling to make better use of available space.

Specific Response 8:

The Fig. 5 caption has been updated to include a description of the shaded regions (95% sampling confidence intervals). We have also updated line thickness for the in situ data to improve visibility for the reader.

References:

Bishop, C. M. (1995). Neural Networks for Pattern Recognition, Oxford University Press. Snauffer AM, Hsieh WW, Cannon, AJ, Schnorbus, MA, 2018. Improving gridded snow water equivalent products in British Columbia, Canada: multi-source data fusion by neural network models. The Cryosphere, 12(3), 891-905.

Erler, A. R., Frey, S. K., Khader, O., d'Orgeville, M., Park, Y.-J., Hwang, H.-T., Lapen, D. R., Peltier, W. R., & Sudicky, E. A. (2019). Simulating Climate Change Impacts on Surface Water Resources Within a Lake-Affected Region Using Regional Climate Projections. Water Resources Research, 55(1), 130–155. https://doi.org/10.1029/2018WR024381

Lary, D. J., Remer, L. A., MacNeill, D., Roscoe, B., & Paradise, S. (2009). Machine Learning and Bias Correction of MODIS Aerosol Optical Depth. IEEE Geoscience and Remote Sensing Letters, 6(4), 694–698. https://doi.org/10.1109/LGRS.2009.2023605

Lv, Z., Pomeroy, J. W., & Fang, X. (2019). Evaluation of SNODAS Snow Water Equivalent in Western Canada and Assimilation Into a Cold Region Hydrological Model. Water Resources Research, 55(12), 11166–11187. https://doi.org/10.1029/2019WR025333

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, Journalof Machine Learning Research, 12, 2825–2830, 2011.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. Nature, 566(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Shen, C. (2018). A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. Water Resources Research, 54(11), 8558–8593. https://doi.org/10.1029/2018WR022643

Wrzesien, M. L., Durand, M. T., Pavelsky, T. M., Howat, I. M., Margulis, S. A., & Huning, L. S. (2017). Comparison of methods to estimate snow water equivalent at the mountainrange scale: A case study of the California Sierra Nevada. Journal of Hydrometeorology, 18(4), 1101-1119.

Xue, Y., Forman, B. A., & Reichle, R. H. (2018). Estimating snow mass in North America through assimilation of AMSR-E brightness temperature observations using the Catchment land surface model and support vector machines. Water Resources Research, 54(9), p.6488.

Zahmatkesh, Z., Tapsoba, D., Leach, J., & Coulibaly, P. (2019). Evaluation and bias correction of SNODAS snow water equivalent (SWE) for streamflow simulation in eastern Canadian basins. Hydrological Sciences Journal, 64(13), 1541–1555. https://doi.org/10.1080/02626667.2019.1660780

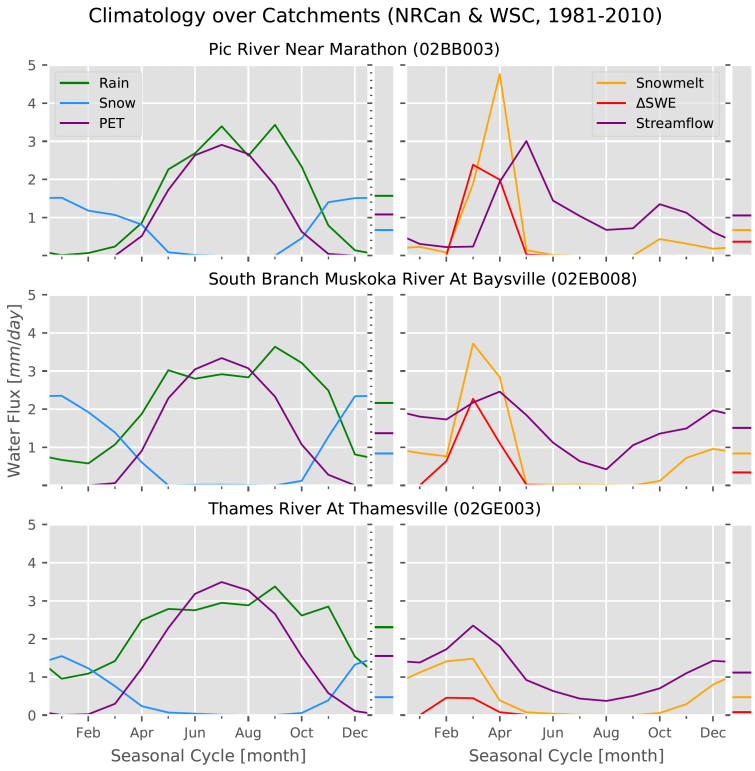

Fig. 1. Catchment water flux climatology (1981-2010) for NRCan data and stream gauge data from the Water Survey of Canada.