Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-593-RC3, 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.

Anonymous Referee #3

Received and published: 14 April 2020

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

Comments:

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.

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?

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?

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.

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)?

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.

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.

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.

References:

Bishop CM, 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.

Wrzesien ML, Durand MT, Pavelsky TM, Howat IM, Margulis SA, Huning LS, 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 BA, Reichle RH, 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.

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