Interactive comment on "Improving gridded snow water equivalent products in British Columbia, Canada: multi-source data fusion by neural network models" *by* Andrew Snauffer et al.

Andrew Snauffer et al.

asnauffer@eoas.ubc.ca

The authors greatly appreciate the comments in RC2: 'Review of "Improving gridded snow water equivalent products in British Columbia, Canada: multi-source data fusion by neural network models" by Snauffer et al.', by Anonymous Referee #2, 14 Aug 2017 (https://doi.org/10.5194/tc-2017-56-RC2). The following are responses to those comments.

Reviewer Comments (C#:), followed by Author Responses

C1: Though I understand ANNs to be very powerful machine learning algorithms, I

would like the authors to clarify why they have chosen an ANN instead of other machine learning algorithms such as Support Vector Machines (SVM)? Some other methods are more computationally efficient and provide very similar results to ANNs as shown by Forman and Reichle (2014).

Response: A number of alternate machine learning methods were investigated in the course of this work. Bayesian Neural Networks (BNN) and Support Vector Machines (SVM) runs were executed but did not result in improvement upon the ANN in spite of additional computational cost. Mentions of these investigations will be added to the manuscript. While SVM has been documented to improve brightness temperature modeled over North America (Forman and Reichle 2014), the authors of the submitted manuscript have previously found that in an evaluation of four different non-linear methods across nine different environmental data sets, no single non-linear method consistently outperformed the others (Lima et al., 2015). Though an exhaustive exploration of machine learning algorithms was not the objective of this manuscript, the trials performed indicated that of the investigated algorithms the ANN was the best method for this particular application and data set.

C2: In Table 1, please provide the names of the SWE products and there acronyms in the legend to make it easier for the reader to identify quickly the different products.

Response: The reviewer comment will be incorporated as suggested.

C3: In this study, only the Mean Absolute Error (MAE) is used to determine the performance of each product/method. Please provide reasoning for this or add other metrics such as bias. The bias potentially gives more information on the performance of the method by indicating if it over/underestimates the measurements. This could actually help understand why ANN3 outperforms ANN6.

Response: This work attempted to better estimate SWE magnitudes and temporal variations on a regional scale. MAE and April correlation were chosen as metrics to examine these quantities. In the mountainous province of BC, the typically high SWE

accumulations are not well represented by any of the gridded products, leading to overwhelmingly negative biases. Hence in this region, bias generally follows MAE, which was noted by Snauffer et al. (2016). Mean station biases by product and physiographic region are shown here in Fig. 1 (same as Fig. 2 in author response AC1). This figure shows similar patterns to those of mean station MAEs in Fig. 4 in the submitted manuscript. Mean station biases for ANN6 and ANN3 are found to be comparable across the province. Biases are also significantly reduced by MLR models, though the MAE and April correlation box plots in Figure 8 of the submitted manuscript confirm the ANN models statistically outperform the MLR models.

C4: I understand that the authors have analysed in depth the six SWE products in Snauffer et al. (2016). Nonetheless, why only try different combinations of the 3 best SWE products with the ANN? Some other combination might actually prove better since the different SWE products don't all use the same inputs and modelling schemes to estimate SWE. Though this might be out of the scope of this current paper, I suggest the author provide a reasoning why they only tested combinations of the 3 best SWE products and I also suggest they test other combinations in a future study.

Response: Table 1 details the gridded products used in various run. Each product was run in the ANN individually, and combinations of products were run based on their previously evaluated performance. All combinations of the best three were run, and the best four through six products were also run. It is not anticipated, for instance, that adding the poorest performing product to an ANN of the best three will outperform an ANN of the best four. If the reviewer has suggestions for additional combinations or better ways to explain this combination approach, please indicate this.

C5: I would also like more clarifications on the selection of the ANN parameterization. Since there are many parameters in an ANN, this would help understand the results. Even if the parameters are the default ones from a given algorithm, please provide them.

Response: The following additional ANN parameterization details will be incorporated.

An ensemble ANN (hereafter ANN) made up of 50 members was constructed following Cannon and McKendry (2002) using the implementation by Cannon (2015). ANN topology consisted of an input layer, one hidden layer and an output later with a single node for SWE. A single hidden layer is adequate to model any nonlinear continuous function (Hsieh, 2009). The hyperbolic tangent sigmoid function was used as the hidden layer transfer function, the identity function was used for the output layer, and predictor and predictand variables were standardized (zero mean and unit standard deviation) prior to model training. Initial model weights were set randomly in the range of $\pm\sqrt{0.8/d_{in}}$, where d_{in} is the number of inputs (Thimm and Fiesler, 1997). The optimization function employed was the Broyden-Fletcher-Goldfarb-Shanno or "BFGS" method (Fletcher, 1970; Nash, 1979). To prevent overfitting, early stopping was used to regularize individual ensemble members; ANN training was stopped when validation error, as monitored on the out-of-bootstrap cases, reached a minimum. The optimal number of hidden nodes for each test split was determined by minimizing the ensemble's out-of-bootstrap RMSE over the training data. Following model selection and training, predictions from the 50 ensemble members were averaged for each test split, and statistics for each test split station were calculated accordingly.

C6: The authors need to do a more thorough analysis of the predictor variables. Which predictor is more statistically significant? Are there correlated variables? Etc.

Response: In statistics, regression coefficient estimation becomes very uncertain when there are correlated inputs. Unlike statistics, machine learning does not attempt to estimate "regression coefficients", and correlated inputs do not affect prediction accuracy. Predictor variables used in the ANNs, especially the SWE gridded products, are certainly correlated given that they attempt to estimate the same quantity in the same time period. These correlations will introduce large uncertainties in regression coefficients as well as in finding the most significant predictors. The objective of this work was not to use the results of a machine learning model to evaluate the statisti-

cal significance of or correlations in the constituent input data, but rather to produce a result which has demonstrably better performance.

References:

Cannon, A. J.: monmlp: Monotone Multi-Layer Perceptron Neural Network. R package version 1.1.3, 2015.

Cannon, A. J. and McKendry, I. G.: A graphical sensitivity analysis for statistical climate models: application to Indian monsoon rainfall prediction by artificial neural networks and multiple linear regression models, International Journal of Climatology, 22, 1687-1708, 2002.

Fletcher, R.: A New Approach to Variable Metric Algorithms. Computer Journal, 13(3), 317-322, 1970.

Forman, B. and Reichle, R. H.: Using a Support Vector Machine and a Land Surface Model to estimate Large-Scale Passive Microwave Brightness Temperatures Over Snow-Covered Land in North America. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8 (9), pp.4431-4441, 2014.

Hsieh, W.W.: Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels. Cambridge University Press, 2009.

Lima, A.R., Cannon, A.J., and Hsieh, W.W.: Nonlinear regression in environmental sciences using extreme learning machines: a comparative evaluation. Environmental Modelling Software, 73, pp.175-188, doi.org/10.1016/j.envsoft.2015.08.002, 2015.

Nash, J.C.: Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation. Adam Hilger, Bristol. 2nd Edition, 1990, Institute of Physics Publications, Bristol, 1979.

Snauffer, A. M., Hsieh, W. W., and Cannon, A. J.: Comparison of gridded snow water equivalent products with in situ measurements in British Columbia, Canada, Journal of

Hydrology, 541, 714-726, doi:10.1016/j.jhydrol.2016.07.027, 2016.

Thimm, G. and Fiesler, E.: High-order and multilayer perceptron initialization, IEEE Transactions on Neural Networks, 8, 349–359, 1997.
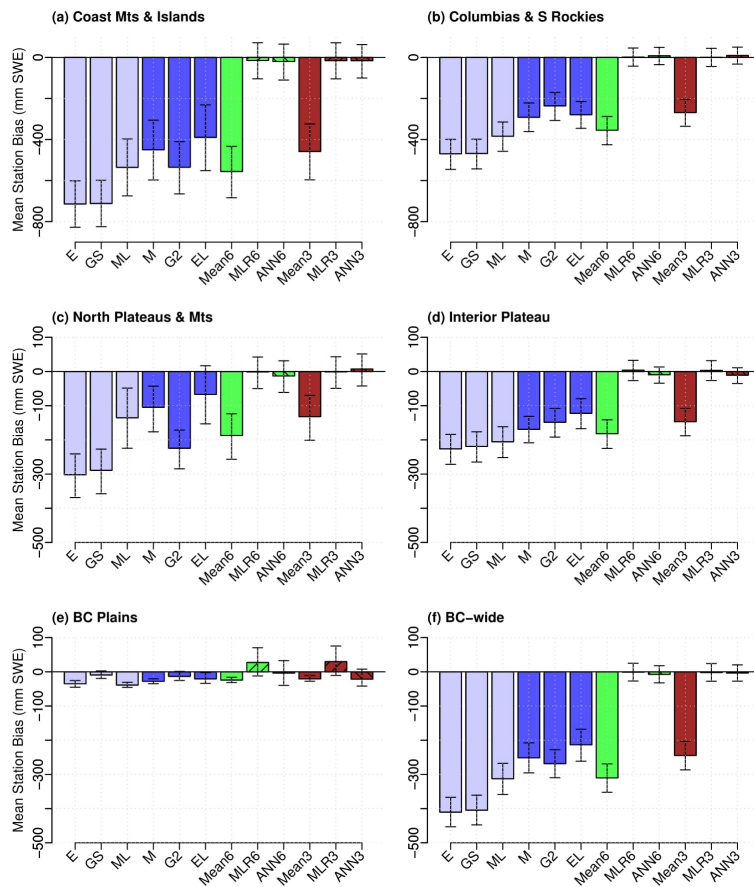
Fig. 1. Mean station bias for several SWE products/combinations for regions of BC in order of descending mean accumulations. Regions, products and combinations are as in Figs. 4 and 5 in submitted manuscript.