Improving gridded snow water equivalent products in British Columbia, Canada: multi-source data fusion by neural network models

Andrew Snauffer[1], William Hsieh[1], Alex Cannon[2], and Markus Schnorbus[3]

[1]Department of Earth, Ocean and Atmospheric Sciences, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[2]Climate Research Division, Environment and Climate Change Canada, PO Box 1700 STN CSC, Victoria, BC V8W 2Y2, Canada
[3]Pacific Climate Impacts Consortium, University House 1, 2489 Sinclair Road, University of Victoria, Victoria, BC V8N 6M2, Canada

Correspondence to: Andrew Snauffer (asnauffer@eoas.ubc.ca)

Abstract. Estimates of surface snow water equivalent (SWE) in alpine regions with seasonal melts are particularly difficult in areas of high vegetation density, topographic relief and snow accumulations. These three confounding factors dominate much of the province of British Columbia (BC), Canada. An artificial neural network (ANN) was created using as predictors six gridded SWE products previously evaluated for BC: ERA-Interim/Land, GLDAS-2, MERRA, MERRA-Land, GlobSnow and ERA-Interim. Relevant spatiotemporal covariates including survey date, year, latitude, longitude, elevation and grid cell elevation differences were also included as predictors, and observations from manual snow surveys at stations located throughout BC were used as target data. Mean absolute errors (MAEs) and correlations for April surveys were found using cross validation. The ANN using the three best performing SWE products (ANN3) had the lowest mean station MAE across the entire province, improving on the performance of individual products by an average of 53%. Mean station MAEs and April survey correlations were also found for each of BC's five physiographic regions. ANN3 outperformed each product as well as product means and multiple linear regression (MLR) models in all regions except for the BC Plains, which has relatively few stations and much lower accumulations than other regions. Subsequent comparisons of the ANN results with predictions generated by the Variable Infiltration Capacity (VIC) hydrologic model found ANN3 to be superior over the entire VIC domain and within most physiographic regions. The superior performance of the ANN over individual products, product means, MLR and VIC was found to be statistically significant across the province.

1 Introduction

In areas of the world with significant alpine regions, many drainage basins exhibit a predominantly nival regime. Under such conditions, which are characteristic of the province of British Columbia (BC), Canada, accurate snow water equivalent (SWE) estimation is particularly crucial for water supply, hydropower generation and flood forecasting and planning purposes. The ability of current methods and products to give accurate SWE estimates is limited by a number of topographic and climate

factors. Forest cover, topography and large SWE accumulations pose challenges for physical models containing multiple sim-
plifications and parameterizations, as well as for satellite-based products, which can experience signal masking and wash-out
(Chang et al., 1996; Derksen, 2008). Likewise reanalysis and assimilation-based products are susceptible to biases arising from
various structural limitations (e.g. elevation biases tied to spatial resolution) and uncertainties in the climate mean state (Dutra

5   et al., 2011; Reichler and Kim, 2008). Mudryk et al. (2015) compared various gridded products across the Northern Hemi-
sphere and observed large spreads in SWE with magnitudes on the order of the SWE interannual variability. These spreads
were seen particularly in alpine regions with complex topography and in the Arctic, a region of large data gaps.

A similar set of gridded products was examined and ranked by Snauffer et al. (2016) over the rugged, forested topography
of British Columbia. Product correlations, biases and MAEs were determined by comparison with in situ manual snow surveys
10   and reported by survey month and physiographic region. While the best products were generally superior performers across
the determined statistics and the majority of regions, all products were found to underestimate SWE and imperfectly represent
interannual fluctuations, particularly in the regions of highest snow accumulation. Against this backdrop, the current work
set out to improve the accuracy of SWE estimates in the same topographically complex region by combining an ensemble of
products with relevant site covariates utilizing a data fusion approach.

15   Numerous models and estimation techniques have attempted to better characterize SWE. Nascent satellite-based efforts
involved the use of multiple linear regression (MLR) to retrieve snow parameters from passive microwave data. Though newer
platforms have become available, many of the early algorithms (Chang et al., 1987; Foster et al., 1997; Tait, 1998) rely on the
Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave/Imager (SSM/I) brightness temperature data,
which is a long passive microwave record covering July 1987 to December 2015 derived from intercalibrated measurements
20   across several spaceborne platforms Wentz (2013). While the early linear approaches performed well in areas of low snow
accumulation, forest cover and topographic relief, mountainous regions with dense vegetation and deep snowpacks were not
as well represented.

Machine learning methods began to become an important tool in the environmental sciences in the 1990s and have since
spread to many application areas (Hsieh, 2009). Early applications of machine learning methods for retrieving SWE were
25   developed by Chen et al. (2001) and Guo et al. (2003) using artificial neural networks (ANNs). Inputs were based on a
snow hydrology model, and outputs derived from a dense media radiative transfer model were iteratively computed until
the ANN results converged on observations. The model improved on previous efforts by improving grain size representation
and applying a spatially distributed snow accumulation and melt model. Tedesco et al. (2004) compared SWE retrieved from
an ANN to values from the spectral polarization difference (SPD) algorithm (Aschbacher, 1989), the Helsinki University of
30   Technology (HUT) snow emission model (Pulliainen et al., 1999), and the Chang algorithm (Chang et al., 1987; Foster et al.,
1997). The ANN used SSM/I 19 GHz and 37 GHz vertical and horizontal brightness temperatures for inputs and the national
operational snow observations of the Finnish Environment Institute or HUT simulated brightness temperatures as the target
data. Under dry snow conditions enforced by only considering days on which the maximum temperature was lower than -5
°C, the ANN outperformed these other methods when trained with observations. Evora et al. (2008) developed a data fusion
35   modeling framework utilizing ANNs, passive microwave data and geostatistics. Predictors for this framework were seven

passive microwave channels and the interpolated minimum temperature. The target SWE field was derived from snow depth and density measured manually at snow stations and interpolated by a kriging with external drift (KED) algorithm using elevation as the secondary variable. More recently Tong et al. (2010) compared SWE predictions of ANNs using microwave brightness temperatures (TBs) to those of the SPD algorithm Aschbacher (1989) and other TB difference algorithms (Chang et al., 1987; Derksen, 2008) in the Quesnel River Basin of BC, finding that ANNs which included the most TB channels outperformed other networks and algorithms. Binaghi et al. (2013) applied radial basis function networks to estimate snow cover thickness in the Italian Central Alps, finding that this approach outperformed inverse distance weight and spline interpolation methods commonly used in similar contexts with limited numbers of homogeneously distributed measurement sites. These and other studies covered in various review papers (Gan, 1996; Evora and Coulibaly, 2009; Shi et al., 2016) demonstrate the promise of more accurate snow estimates via machine learning methods, but they do not incorporate existing SWE products directly.

Manual snow surveys, taken on courses extending tens of meters, cannot be accurately represented by large-scale gridded products with resolutions on the order of tens of kilometers (Mudryk et al., 2015). Inherent SWE differences can be introduced by disparities between mean grid cell elevation and station elevation, as well as other topographic controls. Even higher resolution covariates such as the GLOBE 1-km DEM (Hastings et al., 1999) still fail to adequately reflect important physical contexts at survey sites (e.g. local slope and aspect) which can significantly affect measured values. In spite of the scale mismatch, the combination of large-scale gridded products with covariates such as finer terrain information and physiographic context may capture the spatiotemporal patterns in the manual snow surveys.

The objective of this work is to apply data fusion techniques to readily available gridded SWE products, manual snow surveys and covariates in order to better estimate SWE in BC. Means of six products are evaluated, as are means of the best three of these products based on previous rankings (Snauffer et al., 2016). Relevant covariates are further combined with these sets of products in MLR models as well as nonlinear ANNs.

2 Study Area

The study area, namely the province of BC, can be separated into five physiographic regions, which are shown in Fig. 1. These regions vary considerably in terms of their topography and climate, as well as the amount of snow they receive as reported in Snauffer et al. (2016). Along the entire western edge of BC, the Coast Mountains and Islands is a highly rugged series of mountain ranges and troughs with a mean station SWE of 782 mm in April. The Columbia Mountains and Southern Rockies to the southeast is a very rugged region with high mean elevation and a mean measured April SWE of 565 mm. Possessing a similarly moderate 337 mm SWE measured during April surveys, the Northern and Central Plateaus and Mountains is relatively flat along the plateaus in the north with low mountain ranges scattered throughout. The Interior Plateau in the middle of the province is mostly flat except in the east and has a more moderate April mean station SWE of 283 mm. To the northeast of the province, the Interior Plains of British Columbia, referred to in this work as the "BC Plains", is the region of lowest relief as well as SWE accumulations, having a survey station average of 88 mm in April. Further descriptions of the study area are detailed by British Columbia Ministry of Forests (1995) and Holland (1964).
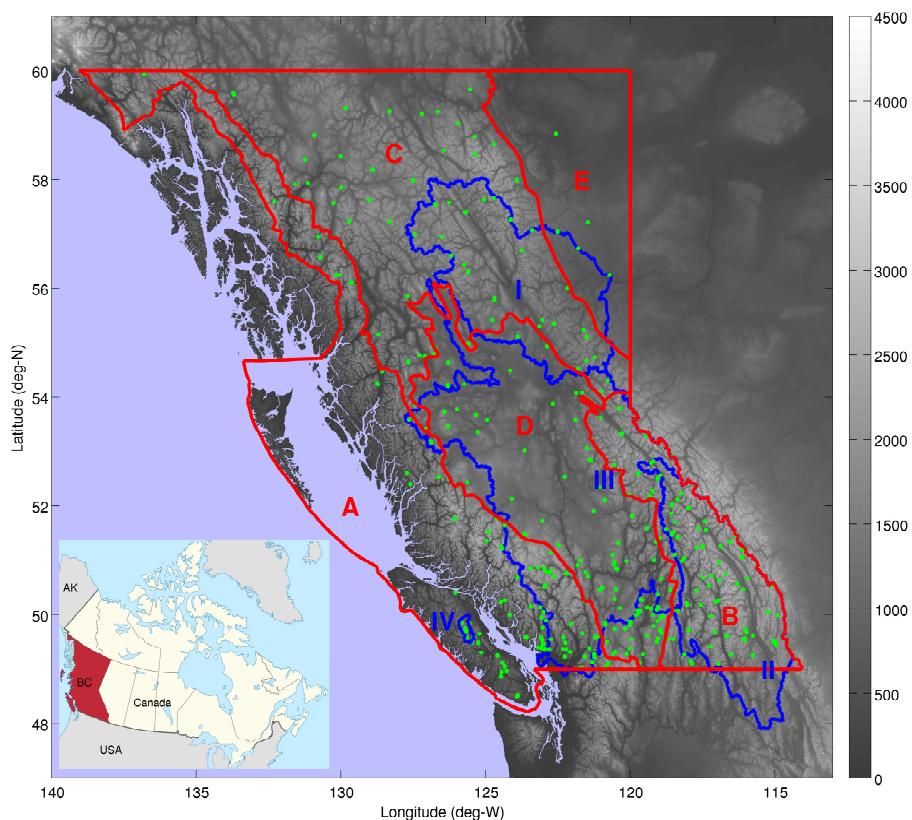
Figure 1. Physiographic Regions of British Columbia, Canada, overlaid on the GLOBE 1-km DEM. Elevations are shown according to the legend with units of meters above sea level. The five physiographic regions are outlined in red: (A) Coast Mountains and Islands, (B) Columbia Mountains and Southern Rockies, (C) Northern and Central Plateaus and Mountains, (D) Interior Plateau, and (E) BC Plains. SWE accumulations generally increase from region A to region E. Outlined in blue is the VIC domain run by the Pacific Climate Impacts Consortium, covering four basins throughout BC: (I) the Peace, (II) upper Columbia, (III) Fraser and (IV) Campbell River watersheds. Manual Snow Survey station locations are indicated by green dots. Inset: A map of Canada (source: *Wikimedia Commons*) shows BC highlighted in red. The province is bordered to the south by the continental United States (USA) and to the west by the US state of Alaska (AK).

3 Methods

Various publicly available SWE data sets have global to hemispheric coverage and can provide information on snow spatiotemporal distribution over BC. This study considered four reanalysis products (ERAInterim, ERALand, MERRA and MERRA-Land), a land data assimilation system (GLDAS2), and an observation-based product (GlobSnow). Key predictors of snow
5 distribution and evolution are these gridded SWE products, and the target data are manual snow surveys conducted throughout the province.

3.1 Input Data

The following gridded SWE products were used as predictors in this study and are described in further detail in Snauffer et al. (2016). ERA-Interim (Dee et al., 2011), a reanalysis product of the European Centre for Medium-Range Weather Forecasts
10 (ECMWF), assimilates land surface, oceanographic, atmospheric and spaceborne measurements from numerous sources using the Integrated Forecast System (IFS). ERA-Interim/Land (Balsamo et al., 2015) is an offline rerun of ERA-Interim using an improved land surface model known as HTESSEL (Hydrology-Tiled ECMWF Scheme for Surface Exchanges over Land) and precipitation adjustments based on the Global Precipitation Climatology Project (GPCP) v2.1. The Modern-Era Retrospective analysis for Research and Applications, or MERRA (Rienecker et al., 2011) ingests a host of land, oceanic and atmospheric
15 data using the Goddard Earth Observing System Data Assimilation System Version 5 (GEOS-5). MERRA-Land (Reichle et al., 2011) uses the land surface model Fortuna-2.5 and makes precipitation adjustments based on the NOAA Climate Prediction Center "Unified" (CPCU) product. The Global Land Data Assimilation System Version 2, known as GLDAS-2 (Rodell et al., 2004) and produced by NASA, runs forcings from the Princeton meteorological data set (Sheffield et al., 2006) on one of four land surface models, including the GLDAS-2 Noah implementation which is used in this study. GlobSnow (Takala et al., 2011),
20 the "mountains unmasked" data set provided by the Finnish Meteorological Institute (FMI), relies on a background field of snow depths created using synoptic observations and a forward model of satellite-based microwave radiometer measurements. Other evaluated products in Snauffer et al. (2016), GLDAS-1 and CMC, do not cover the early part of the study period (1980-2010) and hence are not included in this work.

The target data are manual snow surveys (MSS) conducted by the Snow Survey Network Program for the BC River Forecast
25 Centre (http://bcrfc.env.gov.bc.ca/data/survey/index.htm). At each designated survey site a column of snow is extracted using a Standard Federal Snow Sampler and weighed to determine SWE. The mean of SWE measurements at five to ten points along a snow course is calculated to find the station areal average, which is then reported. Surveys are taken up to eight times a year, once at beginning of months January through June and once mid-month in May and June, and they are currently conducted at 167 stations throughout the province with another 216 stations with historical but no current measurements. Sampling
30 frequencies and temporal ranges vary considerably by station, with some records dating back to 1935.

3.2 Artificial Neural Network Construction

An ensemble ANN (hereafter ANN) was constructed following Cannon and McKendry (2002) using the implementation by Cannon (2015). In total, manual snow survey data from 386 stations across BC was available for training and testing. Of those stations 256 had at least ten years of measurements within the overlapping period of record of the six gridded products

5   (1980-2010). This was considered to be an adequate length to generate representative statistics for testing the ANN by cross-validation. Ten test splits were created by selecting every tenth station, stratified by watershed, such that for each split one-tenth of the stations were excluded from training. The result was an approximately uniform spread of test stations throughout the province in each split, as shown in Fig. 2 (green dots). Inevitably there were large areas with little in situ data, particularly along the coast and in the BC Plains to the north, but the use of all available data outside the test split maximized the spatial

10 coverage of the training set. The MAE and April survey correlation reported at each station were computed using predictions from the ANN which did not include data from that station in its training set.

Training data sets for the individual ANN ensemble members were created using the block bootstrap (Davison and Hinkley, 1997), with each block made up of all training cases for a given station. The optimal number of hidden neurons for each test split was determined by minimizing the out-of-bootstrap RMSE. A hyperbolic activation function was used in the ANN, with

15 initial neuron weights set randomly in the range of $\pm\sqrt{0.8/d_{in}}$, where d is the number of inputs (Thimm and Fiesler, 1997). 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. Finally, predictions from an ensemble of 50 members were averaged for each test split.

In addition to the gridded SWE product(s), the following base set of predictors were included in the model: physiographic

20 region, elevation, elevation biases, latitude, longitude, year, and sine and cosine of the day in the year. These predictors represented major spatial and temporal dimensions to the ANN training, as well as broad location context information (i.e. physiographic region). Efforts to include additional local information were limited by lack of extensive site-specific metadata and appropriate regional scale representations. For instance, while various ground cover products are available, the actual site forest cover is highly dependent on snow course location, time since last maintenance and other factors, which vary considerably

25 across the MSS data set (Tony Litke, personal communication, 2016). As a result, only this base set of covariates has been employed. The main focus of this work, however, was to find which combination(s) of products lead to the best result. While other covariates may incrementally improve the ANN performance, use of this base covariate set should suffice for this purpose.

The gridded SWE product combinations shown in Table 1 were used as predictors in the ANNs. The predictor combinations included each gridded product run individually as well as in combination with better performing products. In addition, all

30 possible two-product combinations of the three best products were run. Only observations for which all six gridded products had SWE values within seven days of the observation date were used, so that all models were based on the same set of observations.

Station elevation biases were calculated as the difference between reported station elevations and the mean grid cell elevations found from the GLOBE 1-km DEM (Hastings et al., 1999), and elevation biases corresponding to each included product
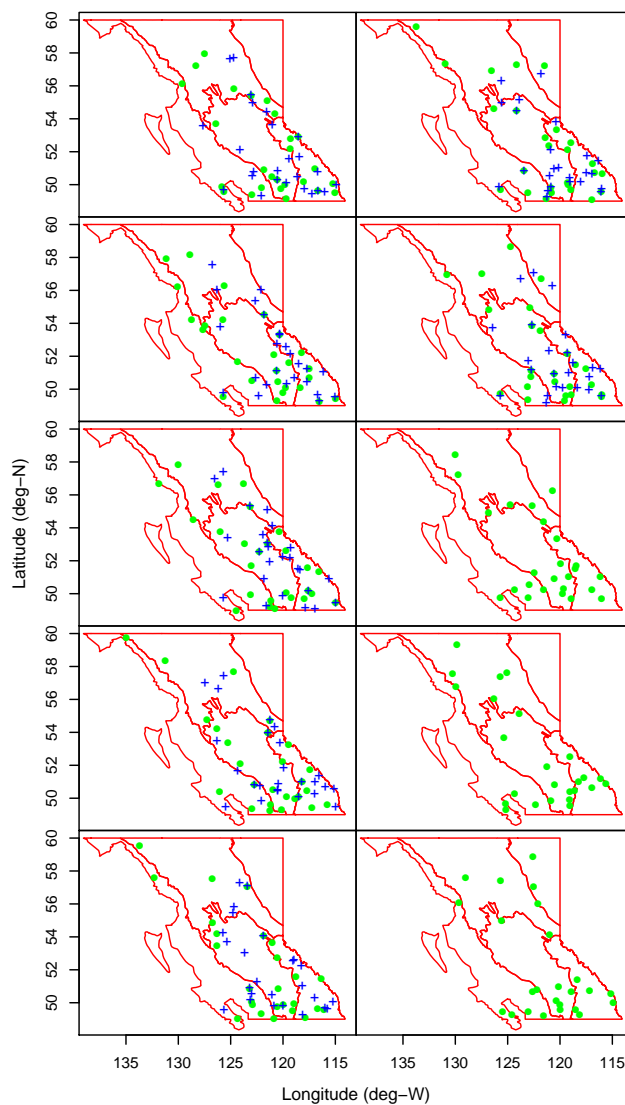
Figure 2. MSS station locations used in evaluating combination products. Green dots show the distribution of ten test splits for comparing means, MLRs and ANNs to gridded SWE products throughout BC by cross validation. Blue crosses show the distribution of seven test splits for comparing ANNs to VIC runs in four BC watersheds. Physiographic regions are outlined in red.

Table 1. Mean station MAEs (in mm SWE) for individual and combinations of products and corresponding MLR and ANN runs, with the mean station MAEs computed using only the VIC stations given in parenthesis.

| Gridded Products | | | | | | | | |
E	GS	ML	M	G2	EL	Mean Product	MLR	ANN
X						411 (400)	214 (217)	184 (184)
	X					407 (396)	220 (218)	196 (194)
		X				362 (349)	211 (209)	178 (180)
			X			315 (303)	196 (191)	172 (168)
				X		308 (294)	206 (205)	177 (176)
					X	305 (290)	202 (201)	173 (176)
				X	X	292 (282)	196 (194)	171 (173)
			X		X	295 (287)	192 (186)	166 (163)
			X	X		295 (287)	190 (186)	170 (165)
			X	X	X	289 (281)	188 (185)	163 (159)
		X	X	X	X	302 (295)	189 (185)	165 (162)
	X	X	X	X	X	316 (311)	189 (186)	167 (167)
X	X	X	X	X	X	328 (324)	189 (186)	168 (166)

were used as predictors. Physiographic regions were represented in the model using discrete logical variables for each region, so-called 1-of-c coding as described in Bishop (1995). Multiple linear regression (MLR) models were run using the same sets of predictors and test stations splits. MLR results are presented alongside those of the ANN models for comparison of linear and nonlinear data fusion techniques.

3.3 Hydrologic Model Comparison

The Variable Infiltration Capacity (VIC) hydrologic model is used by the Pacific Climate Impacts Consortium (PCIC) to conduct hydrologic assessments for the province of British Columbia (Pacific Climate Impacts Consortium, 2014). VIC is a macroscale hydrologic model that solves full surface water and energy balances (Liang et al., 1994), sharing many basic features with land surface models (LSMs) in global climate models (GCMs). The land surface is modeled as grid of large, flat, uniform cells, and sub-grid heterogeneity is handled via statistical distributions. The land-atmosphere fluxes and water and energy balances are simulated at a daily or sub-daily time step. Water can only enter a grid cell via the atmosphere, which means that water in the channel network stays in the channels and subsurface flow between grid cells is not included in the water balance. For the 1/16° grid cells used in the PCIC implementation, this is a reasonable approximation.

The PCIC VIC implementation currently covers four watersheds throughout BC (Pacific Climate Impacts Consortium, 2014), as shown in Fig. 1. Relevant model parameters, specifically snow/rain temperature thresholds and snow albedo decay curves, were explicitly adjusted to reproduce observed SWE at snow pillows, with subsequent automated multi-objective

calibration to reproduce streamflow series and volume errors (Schnorbus et al., 2010). The model is forced with daily observations for minimum and maximum temperature, precipitation, and wind on a 1/16° grid (Schnorbus et al., 2014).

A second set of ANN runs was conducted using a subset of the test stations previously used. Only the MSS stations that were within the VIC domain and that had at least ten years of measurements were used for testing. The 173 stations which

5 met this criterion were divided into seven test splits of 24 or 25 stations each, ordered by watershed, such that for each split one-seventh of the VIC long record stations were excluded from training. This selection results in a relatively uniform spread of test stations for each split over the VIC domain as shown in Fig. 2 (blue crosses). Model training and MAE and correlation computation for each station were performed as in the full-province runs.

4   Results

10 Results of ANN runs for different combinations of products in this study are shown in Table 1. The gridded products used are listed in order of increasing performance across the entire province: ERAInterim (E), GlobSnow (GS), MERRALand (ML), MERRA (M), GLDAS2 (G2) and ERALand (EL). Mean station MAEs are presented for single and average gridded product SWE values plus MLRs and ANNs constructed using test splits of 256 stations across BC. MAEs are also shown for those same combinations based on test splits of 173 stations within the VIC domain. Using the individual products in an ANN with

15 a base set of relevant spatiotemporal covariates improves the mean station MAE by 43% to 55%. ANNs and MLRs based on combinations of the products further improve the mean station MAE, and MAEs for the ANNs are consistently lower than the corresponding values for the MLR models. While ANNs using multiple products all outperform those using those using single products, the best performance was achieved by a combination of the three best products: MERRA, ERALand and GLDAS2. The differences in mean station MAE between the three-product ANN and ANNs that use more products are very

20 small, suggesting that adding the products that do not perform as well does not bring additional useful information into the model. These results are consistent both for ANNs that use test splits of all 256 stations and for those that rely on test splits of the 173 VIC stations.

While comparing snow statistics aggregated across BC is useful for broad comparisons, SWE is a highly inhomogeneous field due to climatic and topographic variability in the province. Figure 3 illustrates the variability at stations across the five

25 physiographic regions of the province. SWE values from manual snow surveys, gridded products and combinations are plotted for representative stations, each of which has a station MAE close to the mean for the encompassing region. Regions are ordered by decreasing mean SWE accumulations in panels (a) through (e). The time interval shown covers snow seasons of 1999, a year of relatively high accumulations across the province, to 2005, a relatively low year. A wide range of accumulations across the regions was observed, with a peak SWE of 50 to 150 mm observed in the BC Plains (panel e) while 500 to over 3000 mm SWE

30 was measured in the Coast Mountains and Islands (panel a). Significant temporal disparities were also seen. Peak SWE in the Coast Mountains and Islands in 1999 was over five times that of 2005, while there is less than a 40% difference for these years in the BC Plains. Underestimates in peak SWE are apparent for most gridded products and snow seasons. The largest of these underestimates were in the Columbia Mountains and Southern Rockies (panel b) and the Coast Mountains and Islands, the
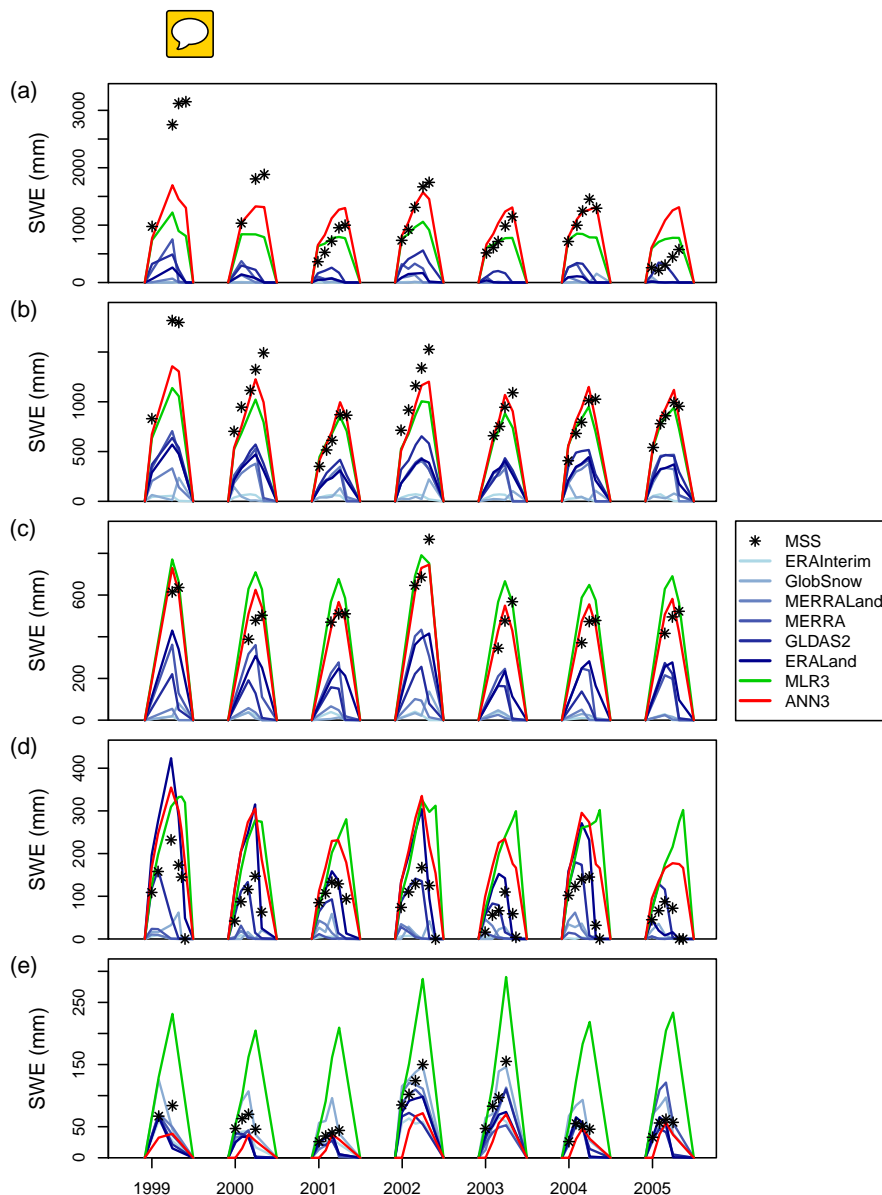
Figure 3. SWE time series for manual snow survey stations, gridded SWE products, MLR3 and ANN3. Gridded products are ordered by increasing overall performance, with the darkest blues for the best performing products. A representative station is shown for each of the 5 physiographic regions of BC: (a) station 1D08 Stave Lake in the Coast Mountains and Islands, (b) station 2A25 Kirbyville Lake in the Columbia Mountains and Southern Rockies, (c) station 4B08 Mount Cronin in the Northern and Central Plateaus and Mountains, (d) station 2F11 Isintok Lake in the Interior Plateau, and (e) station 4C05 Fort Nelson Airport in the BC Plains.

regions of highest accumulations. The MLR model using the combination of the three best products (MLR3) mostly improved the underestimates of SWE apparent at these stations. This model still significantly underpredicted SWE in the two regions of highest accumulations but then also overpredicted in the three other regions with lower accumulations, particularly in the BC Plains (panel a). The ANN using the three best products (ANN3) partially corrected these errors, estimating lower SWE values at the stations with lower accumulations and higher SWE values at those with higher accumulations relative to MLR3.

Further analyzing mean station MAE and April correlation by physiographic region may reveal spatial dependencies and in turn give insight into model performance. Mean station MAEs (Fig. 4) and April correlations (Fig. 5) for the five major physiographic regions of BC in order of decreasing mean SWE accumulations (panels a through e) and for the province as a whole (panel f). Whiskers on the bars represent 95% confidence intervals determined by $n = 5000$ bootstrap samples as described in Jolliffe (2007). Blue bars show mean station values for the six SWE gridded products, with dark blue highlighting the three best performers. Across the four physiographic regions of highest accumulations, ERALand, GLDAS2 AND MERRA are consistently the best performers in both mean station MAE and April correlation values, though in many cases the 95% confidence intervals overlap. In the BC Plains, the region of lowest snow accumulations, GlobSnow has the lowest MAE and a similar April correlation to that of MERRA and ERA-Interim, though the differences in this region are relatively small. The green bars represent different ways of combining all six products: mean, MLR and ANN, while the brown bars have similar crosshatching but use only the three best performing products. In terms of mean station MAEs, the regional story is again similar to that of the entire province for the four regions of highest accumulations. MAEs for the three-product means are slightly lower than those of the six-product means. MLRs show notable performance improvements over multi-product means and are virtually identical for both three- and six-product combinations. ANNs further show improved performance over the MLRs. Three-product ANNs perform very slightly better than the six-product networks and are the best performers of all ANNs. For the BC Plains, both three- and six-product means produce results comparable to those of the individual products, but the MLRs and ANNs have a higher mean station MAE. This result is likely due to the fact that relatively few stations exist in this region, causing the models to be less well trained in the BC Plains. In terms of April correlations, however, the combination products are relatively high across all regions, with the three-product means and ANNs generally being the best performers.

The relative performance of each gridded or combination product relative to a reference was found using the confidence intervals for differences approach of Jolliffe (2007). Briefly, 5000 bootstrap samples of the mean station MAE and April correlation differences between each product and the reference were plotted on box plots (Fig. 6 and Fig. 7), in which the boxes extend between lower and upper quartiles and the whiskers span 95% of the distribution. If the whiskers do not cross the origin, the difference in the product and reference is said to be significant, rejecting the null hypothesis at the 5% level. The mean of all six products (Mean6) was used as the reference and was compared to the remaining gridded and SWE fusion products. Negative mean station MAE differences (Fig. 6) and positive mean station April correlation differences (Fig. 7) indicate that the evaluated product performed better than Mean6. The best three products significantly outperformed Mean6 in MAE across the province and in some of its regions, but Mean6 was often a better performer than the individual products in April correlation. The mean of these best three (Mean3) was consistently better than that of Mean6 under both metrics, and this
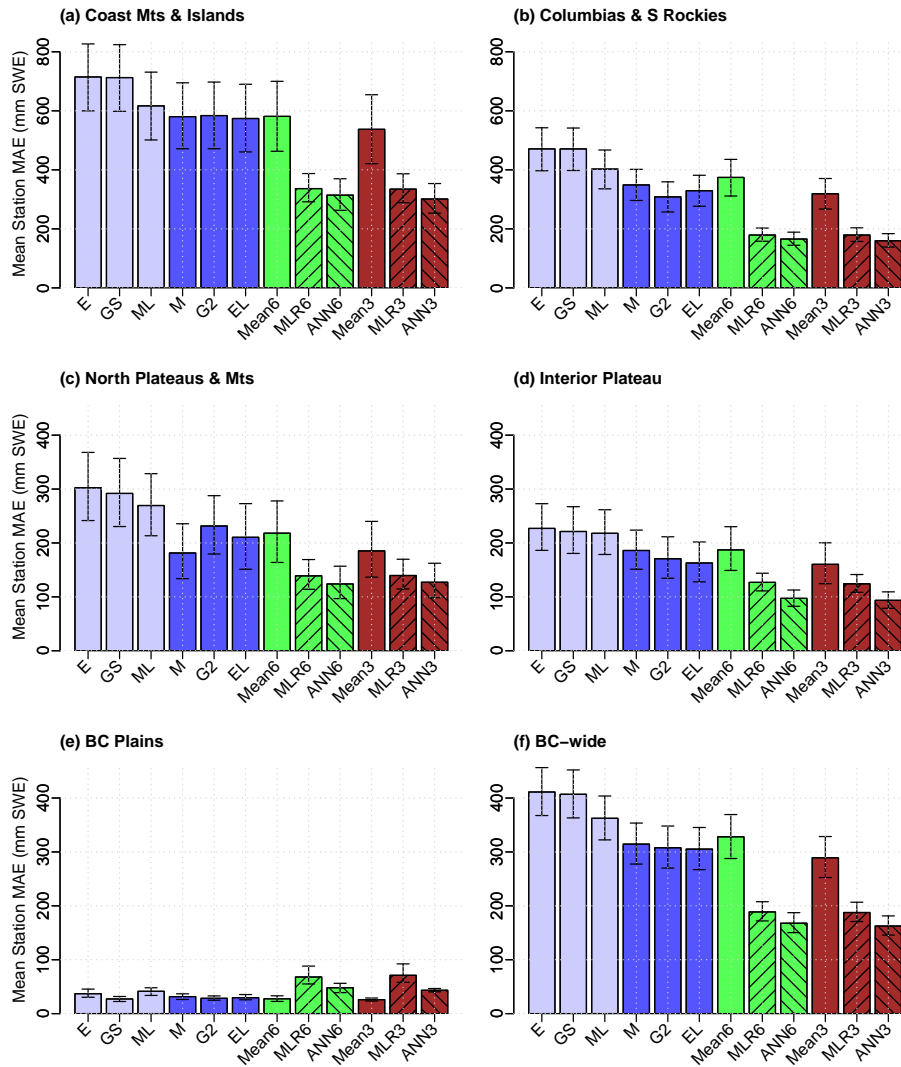
Figure 4. Mean station MAE for several SWE products/combinations for regions of BC in order of descending mean accumulations: (a) Coast Mountains and Islands, (b) Columbia Mountains and Southern Rockies, (c) Northern Plateaus and Mountains, (d) Interior Plateau, (e) BC Plains, and (f) province-wide (combining all data from the five preceding panels). Shown are MAEs with 95% confidence intervals based on $n = 5000$ bootstrap samples for the six gridded products in blue: ERA-Interim (E), GlobSnow (GS), MERRALand (ML), MERRA (M), GLDAS2 (G2) and ERALand (EL). Dark blue indicates three best performing products. Also shown are three fusion techniques (mean, MLR and ANN) using all six products (green) and the three best performing products (brown). Note the vertical scale in panels (a) and (b) is double that of panels (c) through (f).
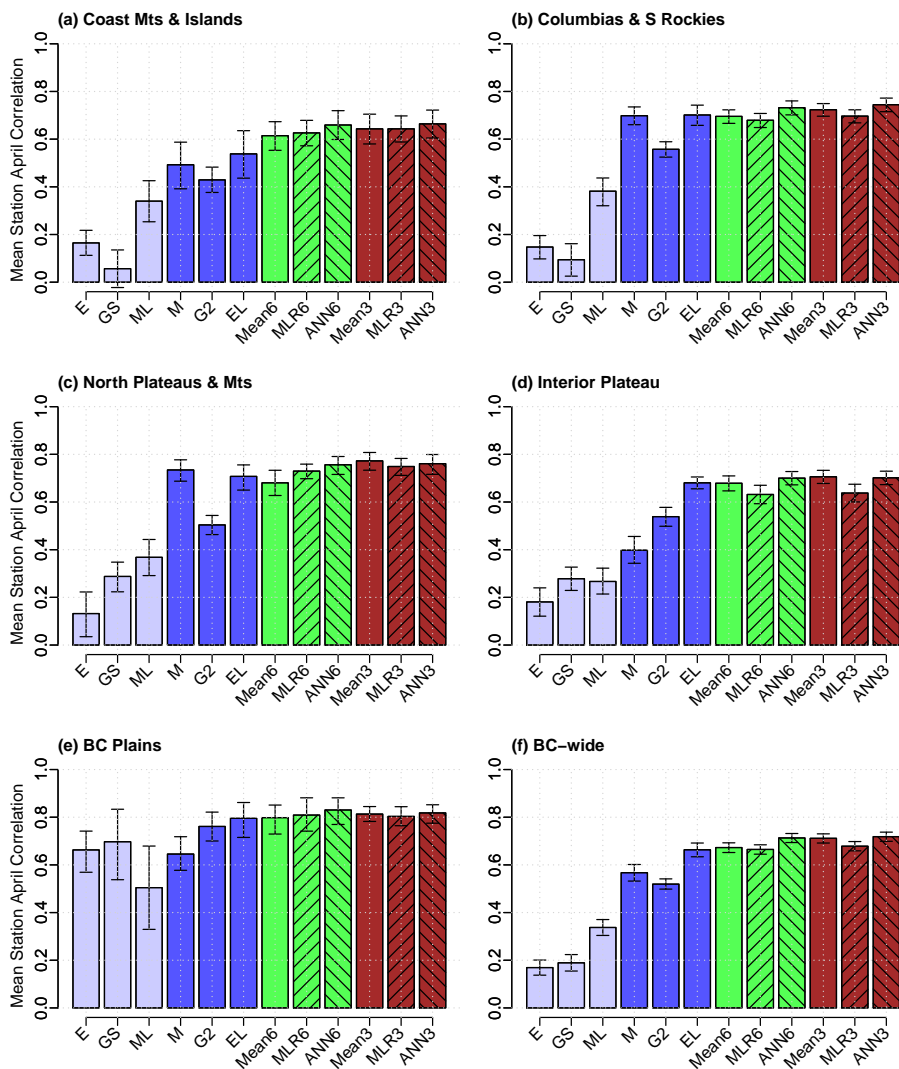
Figure 5. Mean station April Correlation for several SWE products/combinations for regions of BC in order of descending mean accumulations: (a) Coast Mountains and Islands, (b) Columbia Mountains and Southern Rockies, (c) Northern Plateaus and Mountains, (d) Interior Plateau, (e) BC Plains, and (f) province-wide. Product and combination abbreviations and colors are as in Fig. 4.
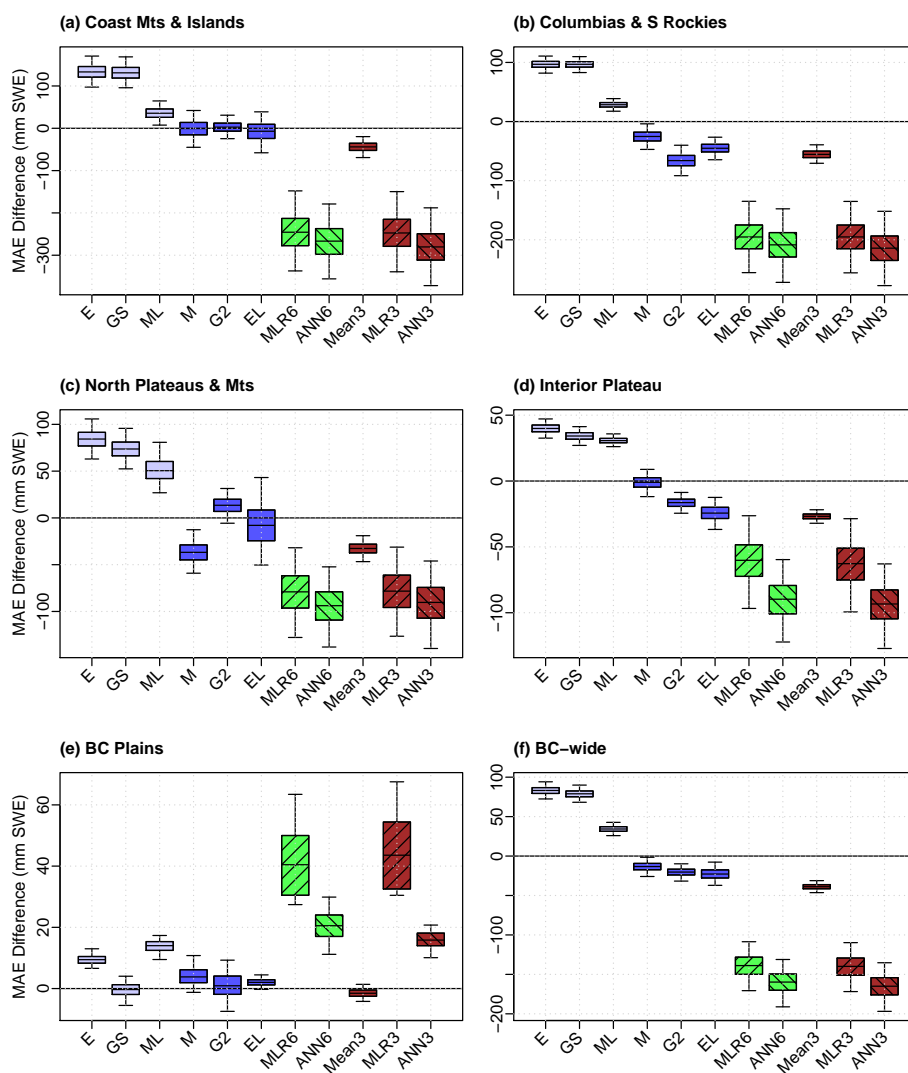
Figure 6. Bootstrap mean station MAE differences between several SWE products/combinations and the reference (the mean of six products). First quartile, median and third quartile, plotted respectively as the bottom, waist and top of each box, as well as 95% confidence intervals indicated by whiskers, are shown for regions of BC in order of descending mean accumulations: (a) Coast Mountains and Islands, (b) Columbia Mountains and Southern Rockies, (c) Northern Plateaus and Mountains, (d) Interior Plateau, (e) BC Plains, and (f) province-wide. Product and combination abbreviations and colors are as in Fig. 4. Note differences in vertical scale.
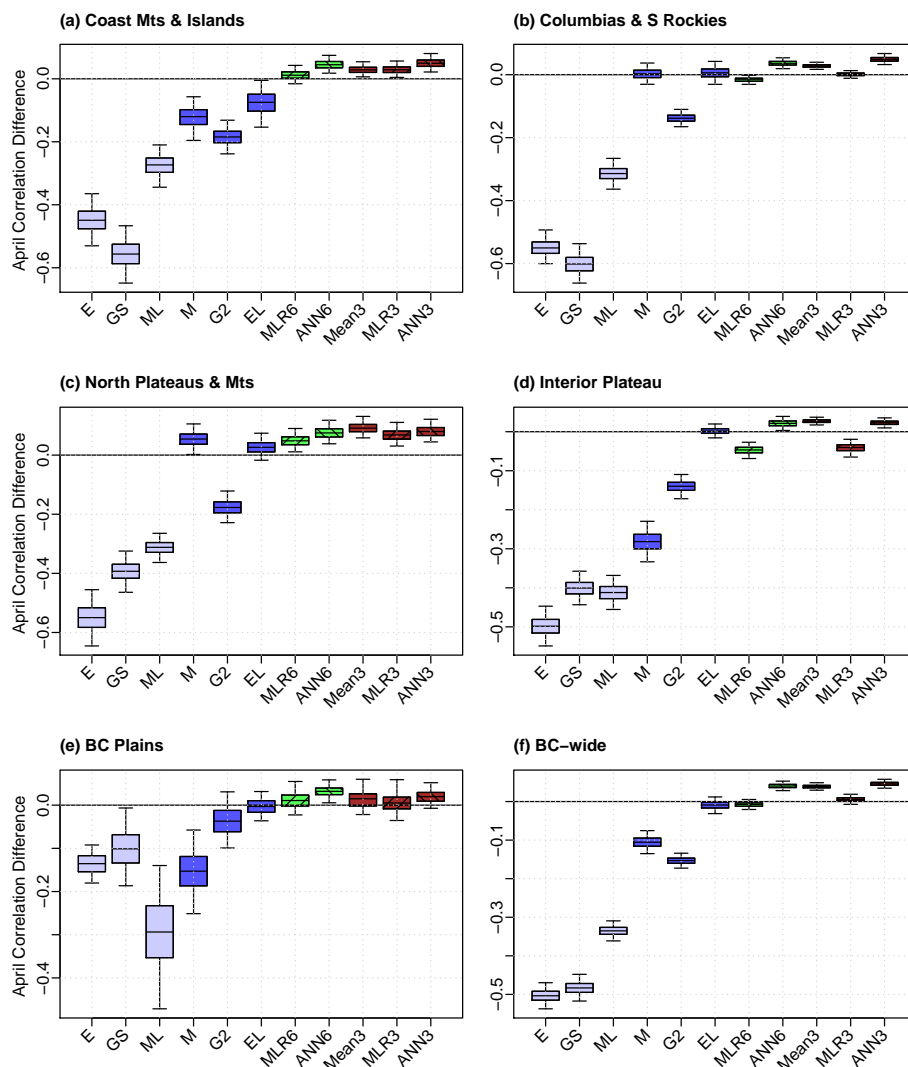
Figure 7. Bootstrap mean station April Correlation differences between several SWE products/combinations and the reference (the mean of six products). Quartiles and 95% confidence intervals are shown for regions of BC in order of descending mean accumulations: (a) Coast Mountains and Islands, (b) Columbia Mountains and Southern Rockies, (c) Northern Plateaus and Mountains, (d) Interior Plateau, (e) BC Plains, and (f) province-wide. Product and combination abbreviations and colors are as in Fig. 4. Note differences in vertical scale.
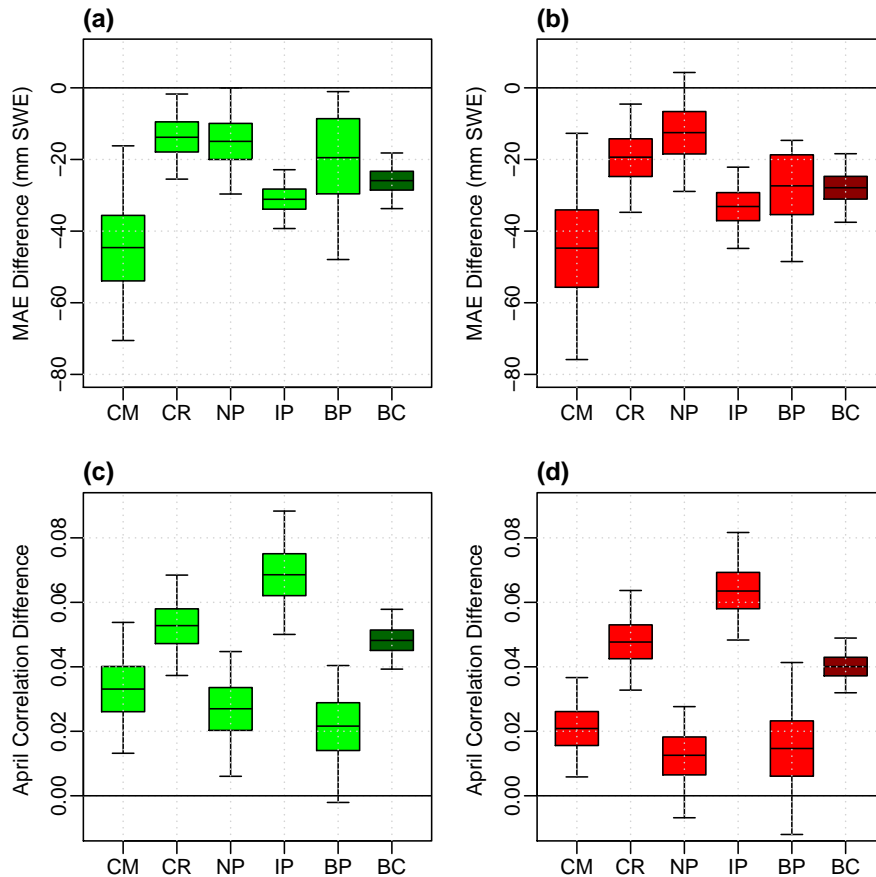
Figure 8. Box plots of bootstrap station differences in MAEs and April correlations showing quartiles and 95% confidence intervals for ANNs of six products, ANN6 (green), and of the best three products, ANN3 (red), with respect to MLR6 and MLR3 respectively. Bootstrap difference distributions are shown for (a) ANN6−MLR6 MAEs, (b) ANN3−MLR3 MAEs, (c) ANN6−MLR6 April correlations and (d) ANN3−MLR3 April correlations. Shown are the differences for each physiographic region: Coast Mountains and Islands (CM), Columbia Mountains and Southern Rockies (CR), Northern and Central Plateaus and Mountains (NP), the Interior Plateau (IP), and the BC Plains (BP), as well as province-wide (BC).

difference is statistically significant across the province and all regions except the BC Plains. The ANN using all six products (ANN6) and one using the best three products (ANN3) also significantly outperformed Mean6 and slightly outperformed corresponding multiple linear regressions (MLR6 and MLR3) province-wide and in all physiographic regions except for the BC Plains. In the BC Plains MLRs were slightly better than Mean6 in April correlation but significantly underperformed in

5   MAE. ANN performance in this region was similar, though the underperformance in MAE was less notable. This failure of the MLRs and ANNs to perform well in MAE was likely due to the lower accumulations of SWE here, accompanied by the scarcity of manual snow survey stations.

A closer examination of ANN performance revealed improvements over comparable MLR. Figure 8 shows the results of $n = 5000$ bootstrap samples of mean station MAE and April correlation differences between the ANNs and MLRs using six products (panels a and c) and the best three products (panels b and d). Negative mean station MAE differences (panels a and b) and positive mean station April correlation differences (panels c and d) indicate the regions in which the ANN performed

5 better than the comparable MLR. ANN improvements over MLR were seen across all regions, and those improvements were statistically significant in most regions and across the province. While non-significant differences between ANNs and MLRs were found for the best three product MAEs and correlations in the Northern and Central Plateaus and Mountains (panels b and d) and for both combinations' correlations in the BC Plains (panels c and d), median ANN improvements over MLR were observed. This result suggests nonlinear modeling of the SWE survey data set may merit further investigation as a tool to better

10 estimate province-wide and regional SWE.

The performance of the six product and best three product ANNs were also compared to VIC runs for the four major watersheds in BC shown in Fig. 1 (Pacific Climate Impacts Consortium, 2014). As these four watersheds do not cover all of the available manual snow stations, station test splits were redrawn and the ANNs runs were conducted separately from those mentioned above. Figure 9 shows the results of $n = 5000$ bootstrap samples of mean station MAE and April correlation

15 differences between the ANNs and VIC using six products (panels a and c) and the best three products (panels b and d). Negative mean station MAE differences (panels a and b) and positive mean station April correlation differences (panels c and d) indicate the regions in which the ANNs performed better than VIC. Across the province, the ANNs outperformed VIC in MAEs and April correlations across most regions. The exceptions included MAEs in the Northern and Central Plateaus and Mountains and BC Plains, where VIC performed slightly better than the ANNs, and April correlations in the Coast Mountains

20 and Islands, where the two are nearly tied. Statistically significant differences were seen only for MAEs in the Columbia and Rocky Mountains and for April correlations in the Northern and Central Plateaus and Mountains. However, across the entire province ANN3 significantly outperformed VIC in both MAEs and April correlations. This result suggests that consideration of a large enough data set and the use of the best gridded products for the province in an ANN can produce better estimates of SWE in BC, both spatially and temporally.

25 **5 Conclusions**

Six gridded products have been used to predict SWE at manual snow survey stations using data fusion techniques. The products estimated lower snow accumulations than were measured, but ANNs combining a single gridded SWE product with a base set of relevant predictors reduced the products' mean station MAEs by 43% to 55% (average 48%) across the province, improvements that are mirrored within most physiographic regions. ANNs which used multiple gridded products as predictors

30 performed better than those using single products. The best performing ANN used the best three products (ERALand, MERRA and GLDAS2), with no improvement seen by including additional products in the ANN. The ANN using the best three products improved mean station MAEs by 47% to 60% (average 53%) over the individual products. Both MLRs and ANNs significantly outperformed the mean of six products in terms of MAEs and April correlations across the province and in four regions. Only
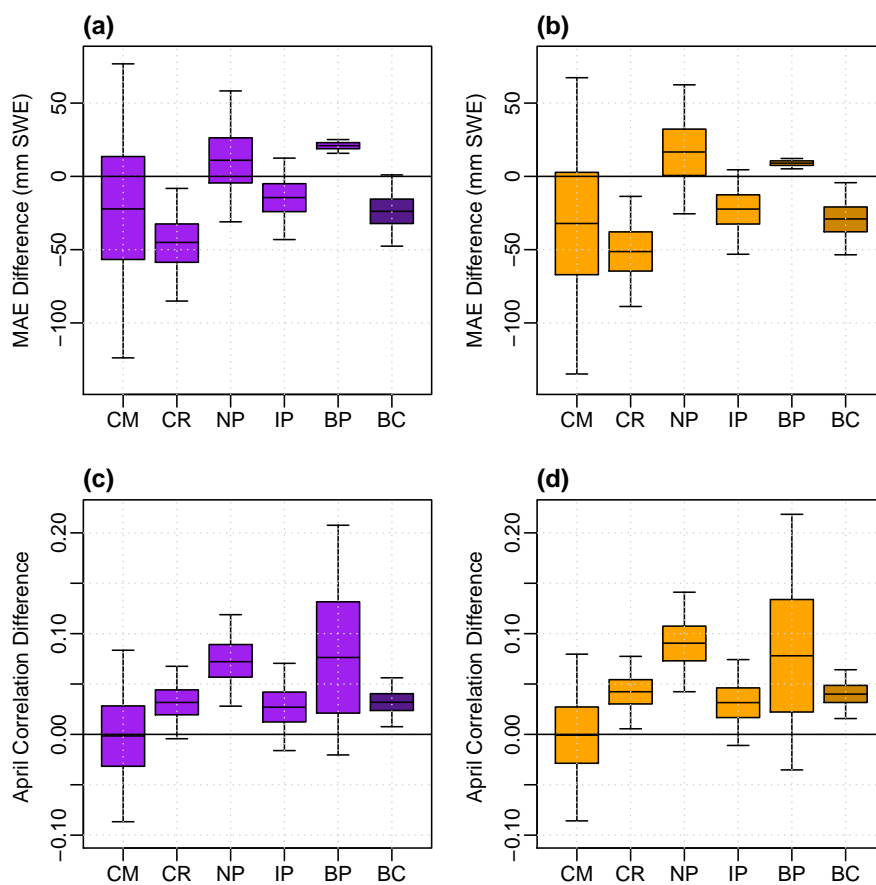
Figure 9. Box plots of bootstrap station differences in MAEs and April correlations showing quartiles and 95% confidence intervals for ANNs of six products, ANN6 (purple), and of the best three products, ANN3 (orange), with respect to VIC. Bootstrap difference distributions are shown for (a) ANN6−VIC MAEs, (b) ANN3−VIC MAEs, (c) ANN6−VIC April correlations and (d) ANN3−VIC April correlations. Notation for the physiographic regions follows Fig. 8.

in the BC Plains, a region of low accumulations and few stations, did the products themselves have lower MAEs, while still underperforming the SWE fusion products in April correlations. The ANNs also outperformed comparable MLRs using the same gridded products; this assessment was statistically significant across the province and in most physiographic regions. Comparing ANNs with runs of the hydrologic model VIC, it was found that the ANNs performed better across the province

5   and in three regions, though statistical significance was found in only one region and across the province for the ANN using the best three products.

While estimating SWE using a fusion of available gridded products and a base set of relevant spatiotemporal covariates poses limits in context and process understanding, from an operational standpoint it can potentially improve the representation of SWE via a far more efficient approach. Further development of the ANNs, including the incorporation of additional covariates

10   and data sources (e.g. terrain information, a simple snow model, etc.), may result in further improvements to the estimation of snow in BC.

6   Data availability

Data sets used in this work have mostly been downloaded from their respective websites. Manual snow survey data is available from the British Columbia River Forecast Centre (http://bcrfc.env.gov.bc.ca/data/survey/index.htm). ERAInterim and

15   ERALand can be obtained from the European Centre for Medium-Range Weather Forecasts Public Datasets web interface (http://apps.ecmwf.int/datasets). MERRA, MERRALand and GLDAS2 can be pulled from the Goddard Earth Sciences Data and Information Services Center (GES DISC) at https://disc.sci.gsfc.nasa.gov/uui/datasets. The GlobSnow "mountains unmasked" data set is not publicly available because it is considered a prototype product. This data set was necessary for this work however, as most of British Columbia is masked in the publicly available data. The GlobSnow "mountains unmasked"

20   data set can be obtained by contacting the GlobSnow team, as described at http://www.globsnow.info/swe/SWE_prototype_ mountains_included/GlobSnow_SWE_prototype_mountain_product_note.txt.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Funding for this work was provided by the Canadian Sea Ice and Snow Evolution (CanSISE) Network awarded through the Climate Change and Atmospheric Research (CCAR) initiative of Canada's Natural Sciences and Engineering Research Council

25   (NSERC). The authors thank Kari Luojus, Juval Cohen and Jaakko Ikonen for making the GlobSnow "mountains unmasked" data set available for inclusion in this study. Tony Litke and Jessica Byers with the Ministry of Environment provided valuable context and saved us much work by packaging and interpreting the manual snow survey data.

References

Aschbacher, J.: Land surface studies and atmospheric effects by satellite microwave radiometry, Ph.D. thesis, University of Innsbruck, Innsbruck, Austria, 1989.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz-Sabater, J., Pappenberger, F., de Rosnay,
5   P., Stockdale, T., and Vitart, F.: ERA-Interim/Land: a global land surface reanalysis data set, Hydrology and Earth System Sciences, 19, 389–407, 2015.

Binaghi, E., Pedoia, V., Guidali, A., and Guglielmin, M.: Snow cover thickness estimation using radial basis function networks, The Cryosphere, 7, 841–854, 2013.

Bishop, C. M.: Neural Networks for Pattern Recognition, Oxford University Press, 1995.

10   British Columbia Ministry of Forests: 1994 Forest, Recreation, and Range Resource Analysis, Crown Publications, 1995.

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.

Chang, A., Foster, J., and Hall, D.: Nimbus-7 SMMR derived global snow cover parameters, Annals of Glaciology, 9, 39–44, 1987.

15   Chang, A., Foster, J., and Hall, D.: Effects of forest on the snow parameters derived from microwave measurements during the BOREAS winter field campaign, Hydrological Processes, 10, 1565–1574, 1996.

Chen, C.-T., Nijssen, B., Guo, J., Tsang, L., Wood, A. W., Hwang, J.-N., and Lettenmaier, D. P.: Passive microwave remote sensing of snow constrained by hydrological simulations, IEEE Transactions on Geoscience and Remote Sensing, 39, 1744–1756, 2001.

Davison, A. C. and Hinkley, D. V.: Bootstrap Methods and Their Application, vol. 1, Cambridge University Press, 1997.

20   Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hersbach, H., Hólm, E., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A., Monge-Sanz, B., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, 2011.

25   Derksen, C.: The contribution of AMSR-E 18.7 and 10.7 GHz measurements to improved boreal forest snow water equivalent retrievals, Remote Sensing of Environment, 112, 2701–2710, 2008.

Dutra, E., Kotlarski, S., Viterbo, P., Balsamo, G., Miranda, P., Schär, C., Bissolli, P., and Jonas, T.: Snow cover sensitivity to horizontal resolution, parameterizations, and atmospheric forcing in a land surface model, Journal of Geophysical Research: Atmospheres, 116, D21 109, doi:10.1029/2011JD016061, 2011.

30   Evora, N. D. and Coulibaly, P.: Recent advances in data-driven modeling of remote sensing applications in hydrology, Journal of Hydroin-formatics, 11, 194–201, 2009.

Evora, N. D., Tapsoba, D., and De Sève, D.: Combining artificial neural network models, geostatistics, and passive microwave data for snow water equivalent retrieval and mapping, IEEE Transactions on Geoscience and Remote Sensing, 46, 1925–1939, 2008.

Foster, J., Chang, A., and Hall, D.: Comparison of snow mass estimates from a prototype passive microwave snow algorithm, a revised
35   algorithm and a snow depth climatology, Remote Sensing of Environment, 62, 132–142, 1997.

Gan, T. Y.: Passive microwave snow research in the Canadian high Arctic, Canadian Journal of Remote Sensing, 22, 36–44, 1996.

Guo, J., Tsang, L., Josberger, E. G., Wood, A. W., Hwang, J.-N., and Lettenmaier, D. P.: Mapping the spatial distribution and time evolution of snow water equivalent with passive microwave measurements, IEEE Transactions on Geoscience and Remote Sensing, 41, 612–621, 2003.

Hastings, D. A., Dunbar, P. K., Elphingstone, G. M., Bootz, M., Murakami, H., Maruyama, H., Masaharu, H., Holland, P., Payne, J., Bryant, N. A., Logan, T. L., Muller, J.-P., Schreier, G., and MacDonald, J. S.: The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, version 1.0, Tech. rep., National Oceanic and Atmospheric Administration, 325 Broadway, Boulder, Colorado 80305-3328, U.S.A., digital data base on the World Wide Web (URL: http://www.ngdc.noaa.gov/mgg/topo/globe.html) and CD-ROMs., 1999.

Holland, S. S.: Landforms of British Columbia: a physiographic outline, 48, Printer to the Queen, 1964.

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

Jolliffe, I. T.: Uncertainty and inference for verification measures, Weather and Forecasting, 22, 637–650, 2007.

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, Journal of Geophysical Research, 99, 14 415–14 428, 1994.

Mudryk, L., Derksen, C., Kushner, P., and Brown, R.: Characterization of northern hemisphere snow water equivalent datasets, 1981–2010, Journal of Climate, 28, 8037–8051, 2015.

Pacific Climate Impacts Consortium: Gridded Hydrologic Model Output, University of Victoria, downloaded from http://tools.pacificclimate.org/dataportal/hydro_model_out/map/ on March 24 2015, 2014.

Pulliainen, J. T., Grandell, J., and Hallikainen, M. T.: HUT snow emission model and its applicability to snow water equivalent retrieval, IEEE Transactions on Geoscience and Remote Sensing, 37, 1378–1390, 1999.

Reichle, R. H., Koster, R. D., De Lannoy, G. J., Forman, B. A., Liu, Q., Mahanama, S. P., and Touré, A.: Assessment and enhancement of MERRA land surface hydrology estimates, Journal of Climate, 24, 6322–6338, 2011.

Reichler, T. and Kim, J.: Uncertainties in the climate mean state of global observations, reanalyses, and the GFDL climate model, Journal of Geophysical Research: Atmospheres, 113, D05 106, doi:10.1029/2007JD009278, 2008.

Rienecker, M., Suarez, M., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M., Schubert, S., Takacs, L., Kim, G., Bloom, S., Chen, J., Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C., Reichle, R., Robertson, F., Ruddick, A., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's modern-era retrospective analysis for research and applications, Journal of Climate, 24, 3624–3648, 2011.

Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J., Walker, J., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, Bulletin of the American Meteorological Society, 85, 381–394, 2004.

Schnorbus, M., Werner, A., and Bennett, K.: Quantifying the water resource impacts of mountain pine beetle and associated salvage harvest operations across a range of watershed scales: Hydrologic modeling of the Fraser River Basin, vol. 423, Pacific Forestry Centre, 2010.

Schnorbus, M., Werner, A., and Bennett, K.: Impacts of climate change in three hydrologic regimes in British Columbia, Canada, Hydrological Processes, 28, 1170–1189, 2014.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling, Journal of Climate, 19, 3088–3111, 2006.

Shi, J., Xiong, C., and Jiang, L.: Review of snow water equivalent microwave remote sensing, Science China Earth Sciences, 59, 731–745, 2016.

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.

Tait, A.: Estimation of snow water equivalent using passive microwave radiation data, Remote Sensing of Environment, 64, 286–291, 1998.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J.-P., Koskinen, J., and Bojkov, B.: Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements, Remote Sensing of Environment, 115, 3517–3529, 2011.

Tedesco, M., Pulliainen, J., Takala, M., Hallikainen, M., and Pampaloni, P.: Artificial neural network-based techniques for the retrieval of SWE and snow depth from SSM/I data, Remote Sensing of Environment, 90, 76–85, 2004.

Thimm, G. and Fiesler, E.: High-order and multilayer perceptron initialization, IEEE Transactions on Neural Networks, 8, 349–359, 1997.

Tong, J., Dery, S. J., Jackson, P. L., and Derksen, C.: Testing snow water equivalent retrieval algorithms for passive microwave remote sensing in an alpine watershed of western Canada, Canadian Journal of Remote Sensing, 36, S74–S86, 2010.

Wentz, F. J.: SSM/I version-7 calibration report, Remote Sensing Systems Tech. Rep, 11012, 46, 2013.