Response to Editor (TC-2019-1)

> Thank you for the handling of our manuscript. Below are private comments to you. We also include a response to A. Winstral. In addition, we include a 'track changes' and also a 'clean' version of our revision. We believe that we have handled all necessary revisions and that we have clearly demonstrated the improvement of our method. We hope to hear from you soon.

Comments to the Author:
Dear authors
Thanks for submitting the revised version of your manuscript. You have addressed all points in your replies raised by the reviewers. However, I believe you have failed to incorporate the arguments you bring forward to actually support your conclusions. I find it not sufficient to mention in the reply that you could have done this or that, or you would have even tried it, but it would not have changed the results.

> In our first revision, we made numerous additions. Also, we tested several ways of building the regression model, and we found the results to be the same. So, perhaps there was a miscommunication. It was not that we could have tried a new approach. We actually did a new approach, but it did not make a difference. We state this explicitly in the revised paper.

I expect you to take the arguments by reviewer #2 more seriously and make the appropriate changes to the manuscript. I think it is not too difficult and you are actually close to a valuable contribution that includes a comprehensive, objective analysis with statistically supportable claims. If you are prepared to revise the manuscript accordingly I am happy to reconsider it for publication.

> In the latest version you will find a side-by-side comparison of our model to three other widely used models. We apply all four models to three datasets:
> 1. The western North America snow pillow dataset (close to 2 million data points).
> 2. The complete western North America snow course / aerial marker data set (100,000+ data points).
> 3. The northeast USA collection of data sets (100,000+ measurements).
>
> Our model has lower bias and RMSE in all cases, except for two very small northeast USA datasets (totaling about 1500 data points). We believe that this is a very convincing case that our approach is an improvement over existing methods.

Jürg Schweizer

Response to A. Winstral Review of TC-2019-1

> Thank you for the recent review of our manuscript. We appreciate the investment of time that you have made in improving our analysis. Below, we provide point-by-point responses to the May 2019 review of our second draft. Reviewer comments are in black. Our responses are indented and in blue. All references to table/figure/line numbers refer to the 'clean' version of revision #2. One notable change is that we have revised our regression analysis to use 'winter precipitation' and 'winter temperature' rather than mean annual precipitation and mean February temperature. This reduced our RMSE values.

Though I firmly stated that the authors had not presented convincing evidence to support their claims, they chose to not implement and include any of my major suggestions for revisions to make the work more convincing. Instead they have largely chosen to reason away my suggestions. I don't wish to repeat my review nor get into a lengthy debate on the relevancy of my comments. Instead I'll just go with what the presented data say to me. The authors have derived a method for estimating snow densities that includes climatological variables, which provide a means of capturing spatial heterogeneity. They developed and tested this method using data from snow pillows and snow courses. By the authors own admissions (lines 46-50) these data all come from "relatively simple topography". Generally these are all located in flat, wind-sheltered locations. On the other hand the Sturm model, a well-regarded and oft-cited research piece, is based on similar data as well as data collected on manual traverses representing a range of topographic positions and snow deposition zones. The presented comparison with the Sturm model is conducted only at SNOTEL sites (i.e. flat, wind-sheltered). These same exact sites were used in both the calibration of the new model and for comparisons with the Sturm model. Results show that overall the new model performs better. However, if one were to eliminate taiga sites, which were not well represented in the Sturm data, the overall results are very similar (see table below).

> We do not believe this to be an entirely correct statement. In Table 4 of the paper, we provide comparative results for 'all data' and also for the data broken out by snow class. In every row of this table, the results show that our model has a smaller bias and RMSE than Sturm's model.
>
> Note also that, in the revision, we now compare our model to Sturm at all of the northeastern USA data (nearly all snow course, not snow pillow). Our model has lower bias and RMSE than Sturm for all of these data sources. We also now compare our model to Sturm at all (100,000+ measurements) NRCS snow course sites in western North America. These sites are independent from the model training data and use a different method (snow coring instead of snow pillow). Our model has lower bias and RMSE than Sturm for this independent data set.

Though the authors acknowledged that I raised an important point regarding their splitting of data in which all stations were included in both training and validation sets they state only that a station-dependent splitting method produced "extremely close" results to the original without actually presenting those results.

Yes, this was a good point, and we tested it out after your first review. The results were the same. It felt redundant to us to present numeric data (identical) for both of these cases in the revised manuscript. In our new revision, we now explicitly state (beginning of section 2.2) that we did our model building / validation both ways, and that the results were the same.

Given how similar the provided results are for non-taiga performance of the two models, the only conclusion I can draw from the presented data is that at sheltered, taiga sites the new model performs better. At other sheltered locations, the new model might be better or worse.

Table 4 shows that the current model has smaller biases and RMSEs for all snow classes, not just Taiga. For the non-taiga sites, our RMSE is usually ~20% smaller than Sturm. And, our biases are much smaller than the biases of Sturm.

Additionally, the new application of our model, side-by-side with Sturm for two independent data sets shows that our model consistently has lower RMSE and bias.

The new method is trained in the same conditions and sites as the validation was conducted whereas the Sturm method is based on a greater diversity of data from independent sites. Given this unbalanced methodology and the closeness of the results, I find the comparative assessments of model performance at sheltered, non-taiga sites to be uncertain.

See our above remarks. We believe the modifications we have made alleviate this point, which is similar to several above points.

Results averaged over all snow classes except taiga (taken from snow class percentages provided in Section 3.1 and results in Table 4)

	Sturm et al.	Multi-variable two-equation
R^2	0.97	0.97
rmse (mm)	72.1	67.8
bias (mm)	1.76	-2.26

Regarding the usefulness and accuracy of the new methodology at sites other than typical SNOTEL stations for estimating densities nothing has been shown (e.g. the manuscript-referenced crowd-sourced and Lidar data that are gathered in a variety of topographic settings). No data is presented for anything other than flat, wind-sheltered locations therefore conclusions on model performance in any other conditions are not possible.

We showed in our original paper our model applied a great variety of data (snow course, mostly) from the northeastern USA. The northeastern US snow courses (NY and ME Snow Survey sites) are not typical SNOTEL stations and include both forested and open sites in both flat and topographically complex settings. At Hubbard Brook, the snow courses are forested areas about ¼ hectare in size. Thompson Farm in Durham, NH, USA includes a forested site and open site. Sleeper's River in Vermont also includes forested and open field sites across a range of topographic classes.

In addition to this, we have now included all western North America snow course data (100,000+ measurements). These are completely different data from the SNOTEL snow pillow data.

So, we have great confidence in the ability of our model to perform in a wide variety of environments.

Yes, the maps look pretty and capture greater heterogeneity but how accurate are they? A theoretical argument can easily be made that the Sturm method – based on a more diverse set of data – would actually be better.

I would be remiss if I didn't further comment on one of my major suggested revisions that wasn't followed up on: direct comparisons to the Jonas et al. method. Rather than including direct comparisons to the Jonas method (based on over 11,000 observations and cited over 140 times) as suggested, the authors have chosen to make comparisons to the simple Pistocchi method that is solely a function of day-of-year (based on 206 observations and cited 5 times). According to the authors this was necessary because the Jonas method is dependent on month of year, elevation, and a geographic "offset" term. Certainly the former two variables are available to the authors. The presence of the offset term however leads the authors to conclude that the Jonas model cannot be applied to other regions while implying that Jonas et al. did not "construct" their model for such applications. Yet in referring to the importance of the offset term Jonas et al. state, "However, the minor importance of regional effects suggests that the model may also be applicable in other regions with similar snow climatologic conditions." In fact, if one averaged the regional offset term over all the data records in the Jonas application it comes to a mere 3 kg/m3. One could in fact, easily optimize the Jonas "offset" term to the data presented here in the calibration set. (It is my personal belief that any comprehensive analysis should include this). Yet even a straight-out-of-the-box Jonas application using the average offset or none at all would be insightful. If, in fact, the presented model performs better than the Jonas model at the tested sites then this would show that there are indeed regional constraints present in the Jonas model that must be accounted for. On its own, these insights on regionality for one of the most widely cited density parameterizations would be relevant. Of course, this would also lend greater credence to the model presented in this work as well. This entire analysis including optimization could probably be done in less than a day. Without optimizing the offset term, this is about 5 lines of code that could even be handled in a spreadsheet. I truly don't understand why this revision wasn't undertaken.

We now present results for the Jonas model applied to (1) the SNOTEL data, (2) the northeastern USA data, and (3) all NRCS snow course data. Please see sections 3.3, 3.4, and 3.5. The present model has lower bias and RMSE than Jonas for (1), for four out of 6 datasets in (2) and for (3). The two datasets for (2) where Jonas has lower RMSE and bias are the smallest datasets by 1-2 orders of magnitude.

Unfortunately, save for the inclusion of the Pistocchi model comparison, the authors have presented no new data. It is my opinion that the presented work still does not compellingly support the stated conclusions, particularly this one (lines 432-33), "The results presented in this study show that the regression equation described by equations (5, 7-8) is an improvement (lower bias and RMSE) over other widely used bulk density equations."

The present model is now compared against three others (Sturm, Jonas, and Pistocchi) for many different datasets from different regions and different methodologies. The results objectively show that the present model has lower values for bias and RMSE.

The authors have the data available to make this a much more extensive and scientifically supportable work (e.g. significance tests would be a nice touch).

We are not 100% sure which specific tests you are asking for. We have provided metrics such as RMSE, R2, and bias. If you are referring to significance tests related to the linear regression analysis, the p values for each variable were 0, and we have now stated that in the manuscript. We also (this was discussed in the original paper and the discussion remains) indicate the use of adjusted R2 values in the construction of the model. This is important since one can always add more predictors and improve the R2. The adjusted R2 ensures that we are only adding in more predictors that produce enough of an improvement in R2 to justify the inclusion.

I think I have provided several constructive ideas on how to go about this. At the very least, if they weren't to follow up on these suggestions they need to objectively evaluate their results and in my opinion, substantially scale back their claims.

In summary, our model is compared to three others for three different datasets, representing a wide variety of methodologies (for measuring SWE) and physical locations.

1. Western North America snow pillow data
2. Western North America snow course data
3. Northeast USA data (mostly snow course, some snow pillow)

For all three of these datasets, our model had the smallest errors and biases.

Converting Snow Depth to Snow Water Equivalent Using Climatological Variables

David F. Hill[1], Elizabeth A. Burakowski[2], Ryan L. Crumley[3], Julia Keon[4], J. Michelle Hu[5], Anthony A. Arendt[6], Katreen Wikstrom Jones[7], Gabriel J. Wolken[8]

[1]Civil and Construction Engineering, Oregon State University, OR, USA
[2]Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, NH, USA
[3]Water Resources Graduate Program, Oregon State University, OR, USA
[4]Civil and Construction Engineering, Oregon State University, OR, USA
[5]Civil and Environmental Engineering, University of Washington
[6]Applied Physics Laboratory, University of Washington
[7]Alaska Division of Geological & Geophysical Surveys, Fairbanks, AK, USA
[8]Alaska Division of Geological & Geophysical Surveys, Fairbanks, AK, USA; International Arctic Research Center, University of Alaska Fairbanks, Fairbanks, AK, USA

Correspondence to: David F. Hill (david.hill@oregonstate.edu)

Abstract. We present a simple method that allows snow depth measurements to be converted to snow water equivalent (SWE) estimates. These estimates are useful to individuals interested in water resources, ecological function, and avalanche forecasting. They can also be assimilated into models to help improve predictions of total water volumes over large regions. The conversion of depth to SWE is particularly valuable since snow depth measurements are far more numerous than costlier and more complex SWE measurements. Our model regresses SWE against snow depth (*h*), day of water year (*DOY*) and climatological (30-year normal) values for winter (December, January, February) precipitation (*PPTWT*) and the difference (*TD*) between mean temperature of the warmest month and mean temperature of the coldest month, producing a power-law relationship. Relying on climatological normals rather than weather data for a given year allows our model to be applied at measurement sites lacking a weather station. Separate equations are obtained for the accumulation and the ablation phases of the snowpack. The model is validated against a large database of snow pillow measurements and yields a bias in SWE of less than 2 mm and a root-mean-squared-error (RMSE) in SWE of less than 60 mm. The model is additionally validated against two completely independent sets of data; one from western North America, and one from the northeast United States. Finally, the results are compared with three other models for bulk density that have varying degrees of complexity and that were built in multiple geographic regions. The results show that the model described in this paper has the best performance for the validation data sets.

1 Introduction

In many parts of the world, snow plays a leading-order role in the hydrological cycle (USACE, 1956; Mote et al., 2018). Accurate information about the spatial and temporal distribution of snow water equivalent (SWE) is useful to many stakeholders (water resource planners, avalanche forecasters, aquatic ecologists, etc.), but can be time consuming and expensive to obtain.

Snow pillows (Beaumont, 1965) are a well-established tool for measuring SWE at fixed locations. Figure 1 provides a conceptual sketch of the variation of SWE with time over a typical water year. A comparatively long accumulation phase is followed by a short ablation phase. While simple in operation, snow pillows are relatively large in size and they need to be installed prior to the onset of the season's snowfall. This limits their ability to be rapidly or opportunistically deployed. Additionally, snow pillow installations tend to require vehicular access, limiting their locations to relatively simple topography. Finally, snow pillow sites are not representative of the lowest or highest elevation bands within mountainous regions (Molotch and Bales, 2005). In the western United States (USA), the Natural Resources Conservation Service (NRCS) operates a large network of Snow Telemetry (SNOTEL) sites, featuring snow pillows. The NRCS also operates the smaller Soil Climate Analysis Network (SCAN) which provides the only, and very limited, snow pillow SWE measurements in the eastern USA.

SWE can also be measured manually, using a snow coring device that measures the weight of a known volume of snow to determine snow density (Church, 1933). These measurements are often one-off measurements, or in the case of 'snow courses' they are repeated weekly or monthly as a transect of measurements at a given location. The simplicity and portability of coring devices expand the range over which measurements can be collected, but it can be challenging to apply these methods to deep snowpacks due to the limited length of standard coring devices. Note that there are numerous different styles of coring devices, including the Adirondack sampler and the Mt. Rose / Federal sampler (Church and Marr, 1937). The NRCS operates a large network of snow course sites (USDA, 2011) in the western United States.

There are a number of issues that affect the accuracy of both snow pillow and snow coring measurements. With coring measurements, if the coring device is not carefully extracted, a portion of the core may fall out of the device. Or, snow may become compressed in the coring device during insertion. These effects have led to varying conclusions, with some studies (e.g., Sturm et al., 2010) showing a low SWE bias and other studies (e.g., Goodison, 1978) showing a high SWE bias. As noted by Johnson et al. (2015) a good rule of thumb is that coring devices are accurate to around ± 10%. Also, studies comparing different styles of snow samplers report statistically different results, suggesting that SWE measurements are sensitive to the design of the specific coring device, such as the presence of holes or slots, the device material, etc. (Beaumont and Work, 1963; Dixon and Boon, 2012). With snow pillows, some studies (e.g., Goodison et al., 1981) note that ice bridging can lead to low biases in measured SWE, with the snow surrounding the pillow partly supporting the snow over the pillow. Other studies (Johnson and Marks, 2004; Dressler et al., 2006; Johnson et al., 2015) note a more complex situation with SWE under-reported at times, but over-reported at other times. Note that when snow pillow data are evaluated, they are most commonly compared to coring measurements at the same location.

All methods of measuring SWE are challenged by the fact that SWE is a depth-integrated property of a snowpack. This is why the snowpack must be weighed, in the case of a snow pillow, or a core must be extracted from the surface to the ground. This measurement complexity makes it difficult to obtain SWE information with the spatial and temporal resolution desired for watershed-scale studies. Other snowpack properties, such as the depth h, are much easier to measure. For example, using a graduated device such as a meterstick or an avalanche probe to measure the depth takes only seconds. Automating depth measurements at a fixed location can easily be done using low-cost ultrasonic devices (Goodison et al., 1984; Ryan et al., 2008). High-spatial-resolution measurements of snowpack depth are commonly made with Light Detection and Ranging (LIDAR). One example of this is the Airborne Snow Observatory program (ASO; Painter et al., 2016). The comparatively high expense of airborne LIDAR surveys typical limits measurements geographically (to a few basins) and temporally (weekly to monthly interval).

Given the relative ease in obtaining depth measurements, it is common to use h as a proxy for SWE. Figure 1 shows a conceptual sketch of the variation of SWE with h over a typical water year. Noting the arrows on the curve, we see that SWE is multi-valued for each h. This is due to the fact that the snowpack increases in density throughout the water year, producing a hysteresis loop in the curve. A large body of literature exists on the topic of how to convert h to SWE. It is beyond the scope of this paper to provide a full review of these 'bulk density equations,' where the density is given by $\rho_b = \mathrm{SWE}/h$. Instead, we refer readers to the useful comparative review by Avanzi et al. (2015). Here, we prefer to discuss a limited number of previous studies that illustrate the spectrum of methodologies and complexities that can be used to determine ρ_b or SWE.

Many studies express ρ_b as an increasing function (often linear) of h. In some cases (e.g., Lundberg et al., 2006) a second equation is added where ρ_b attains a constant value when a threshold h is exceeded. A single linear equation captures the process of densification of the snowpack during the accumulation phase, but performs poorly during the ablation phase, where depths are decreasing but densities continue to increase or approach a constant value. Other approaches choose to parameterize ρ_b in terms of time, rather than h. Pistocchi (2016) provides a single equation while Mizukami and Perica (2008) provide two sets of equations, one set each for early and late season. Each set contains four equations, each of which is applicable to a particular 'cluster' of stations. This clustering was driven by observed densification characteristics and the resulting clusters are relatively spatially discontinuous. Jonas et al. (2009) take the idea of region- (or cluster-) specific equations and extend it further to provide coefficients that depend on time and elevation as well. They use a simple linear equation for ρ_b in terms of h and the slope and intercept of the equation are given as monthly values, with three elevation bins for each month (36 pairs of coefficients). There is an additional contribution to the intercept (or 'offset') which is region-specific (one of 7 regions).

These classifications, whether based on region, elevation, or season, are valuable since they acknowledge that all snow is not equal. McKay and Findlay (1971) discuss the controls that climate and vegetation exert on snow density, and Sturm et al. (2010) address this directly by developing a snow density equation where the coefficients depend upon the 'snow class' (5 classes). Sturm et al. (1995) explain the decision tree, based on temperature, precipitation, and wind speed, that leads to the classification. The temperature metric is the 'cooling degree month' calculated during winter months only. Similarly, only precipitation falling during winter months was used in the classification. Finally, given the challenges in obtaining high quality, high-spatial-resolution wind information, vegetation classification was used as a proxy. Using climatological values (rather than values for a given year), Sturm et al. (1995) were able to develop a global map of snow classification.

There are many other formulations for snow density that increase in complexity and data requirements. Meloysund et al. (2007) express ρ_b in terms of sub-daily measurements of relative humidity, wind characteristics, air pressure, and rainfall, as well as h and estimates of solar exposure ('sun hours'). McCreight and Small (2014) use daily snow depth measurements to develop their regression equation. They demonstrate improved performance over both Sturm et al. (2010) and Jonas et al. (2009). However, a key difference between the McCreight and Small (2014) model and the others listed above is that the former cannot be applied to a single snow depth measurement. Instead, it requires a continuous time series of depth measurements at a fixed location. Further increases in complexity are found in energy-balance snowpack models (SnowModel, Liston and Elder, 2006; VIC, Liang et al., 1994, DHSVM, Wigmosta et al., 1994, others), many of which use multi-layer models to capture the vertical structure of the snowpack. While the particular details vary, these models generally require high temporal-resolution time series of many meteorological variables as input.

Despite the development of multi-layer energy-balance snow models, there is still a demonstrated need for bulk density formulations and for vertically integrated data products like SWE. Pagano et al. (2009) review the advantages and disadvantages of energy-balance models and statistical models and describe how the NRCS uses SWE (from SNOTEL stations) and accumulated precipitation in their statistical models to make daily water supply forecasts. If SWE information is desired at a location that does not have a SNOTEL station, and is not part of a modeling effort, then bulk density equations and depth measurements are an excellent choice.

The present paper seeks to generalize the ideas of Mizukami and Perica (2008), Jonas et al. (2009), and Sturm et al., (2010). Specifically, our goal is to regress physical and environmental variables directly into the equations. In this way, environmental variability is handled in a continuous fashion rather than in a discrete way (model coefficients based on classes). The main motivation for this comes from evidence (e.g., Fig. 3 of Alford, 1967) that density can vary significantly over short distances on a given day. Bulk density equations that rely solely on time completely miss this variability and equations that have coarse (model coefficients varying over either vertical bins or horizontal grids) spatial resolution may not fully capture it either.

Our approach is most similar to Mizukami and Perica (2008), Jonas et al. (2009), and Sturm et al., (2010) in that a minimum of information is needed for the calculations; we intentionally avoid approaches like Meloysund et al. (2007) and McCreight and Small (2014). This is because our interests are in converting *h* measurements to SWE estimates in areas lacking weather instrumentation. The following sections introduce the numerous data sets that were used in this study, outline the regression model adopted, and assess the performance of the model.

2 Methods

2.1 Data

2.1.1 Snow Depth and Snow Water Equivalent

In this section, we list sources of 1970-present snow data utilized for this study (Table 1). With regards to snow coring devices, we refer to them using the terminology preferred in the references describing the datasets.

2.1.1.1 USA NRCS Snow Telemetry and Soil Climate Analysis Networks

SNOTEL (Serreze et al., 1999; Dressler et al., 2006) and SCAN (Schaefer et al. 2007) stations in the contiguous United States (CONUS) and Alaska typically record sub-daily observations of *h*, SWE, and a variety of weather variables (Figure 2a). The periods of record are variable, but the vast majority of stations have a period of record in excess of 30 years. For this study, data from all SNOTEL sites in CONUS and Alaska and northeast USA SCAN sites (Figure 2b) were obtained with the exception of sites whose period of record data were unavailable online. Only stations with both SWE and *h* data were retained.

2.1.1.2 Canada (British Columbia) Snow Survey Data

Goodison et al. (1987) note that Canada has no national digital archive of snow observations from the many independent agencies that collect snow data and that snow data are instead managed provincially. The quantity and availability of the data vary considerably among the provinces. The Water Management Branch of the British Columbia (BC) Ministry of the Environment manages a comparatively dense network of Automated Snow Weather Stations (ASWS) that measure SWE, *h*, accumulated precipitation, and other weather variables (Figure 2a). For this study, data from all British Columbia ASWS sites were initially obtained. As with the NRCS stations, only ASWS stations with both SWE and *h* data were retained.

2.1.1.3 USA NRCS Snow Course / Aerial Marker Data

The snow survey program (USDA, 2008) dates to the 1930s and includes a large number of snow course and aerial marker sites (Figure 2c) in western North America. While the measurement frequency is variable, it is most commonly monthly. To generate a dataset for this study, data were extracted using the National Water and Climate

Center Report Generator 2.0. This allows filtering by time period, elevation band, and other elements. All sites with data between 1980-2018 were included (Figure 2c).

2.1.1.4 Northeast USA Data

In addition to the data from the SCAN sites, snow data for this project from the northeast US come from two networks and three research sites (Figure 2b). The Maine Cooperative Snow Survey (MCSS, 2018) network includes h and SWE data collected by the Maine Geological Survey, the United States Geological Survey, and numerous private contributors and contractors. MCSS snow data are collected using the Standard Federal or Adirondack snow sampling tubes typically on a weekly to bi-weekly schedule throughout the winter and spring, 1951-present. The New York Snow Survey network data were obtained from the National Oceanic and Atmospheric Administration's Northeast Regional Climate Center at Cornell University (NYSS, 2018). Similar to the MCSS, NYSS data are collected using Standard Federal or Adirondack snow sampling tubes on weekly to bi-weekly schedules, 1938-present.

The Sleepers River, Vermont Research Watershed in Danville, Vermont (Shanley and Chalmers, 1999) is a USGS site that includes 15 stations with long-term weekly records of h and SWE collected using Adirondack snow tubes. Most of the periods of record are 1981-present, with a few stations going back to the 1960s. The sites include topographically flat openings in conifer stands, old fields with shrub and grass, a hayfield, a pasture, and openings in mixed softwood-hardwood forests. The Hubbard Brook Experiment Forest (Campbell et al., 2010) has collected weekly snow observations at the Station 2 rain gauge site, 1959-present. Measurement protocol collects ten samples 2 m apart along a 20 m transect in a hardwood forest opening about ¼ hectare in size. At each sample location along the transect, h and SWE are measured using a Mt. Rose snow tube and the ten samples are averaged for each transect. Finally, the Thompson Farm Research site includes a mixed hardwood forest site and an open pasture site (Burakowski et al. 2013; Burakowski et al. 2015). Daily (from 2011-2018), at each site, a snow core is extracted with an aluminum tube and weighed (tube + snow) using a digital hanging scale. The net weight of the snow is combined with the depth and the tube diameter to determine ρ_b, similar to a Federal or Adirondack sampler.

2.1.1.5 Chugach Mountains (Alaska) Data

In the spring of 2018, we conducted three weeks of fieldwork in the Chugach mountains in coastal Alaska, near the city of Valdez (Figure 2d-e). We measured h using an avalanche probe at 71 sites along elevational transects during March, April, and May. The elevational transects ranged between 250 and 1100 m (net change along transect) and were accessible by ski and snowshoe travel. At each site, we measured h in 8 locations within the surrounding 10 m^2, resulting in a total of 550+ snow depth measurements. These 71 sites were scattered across 8 regions in order to capture spatial gradients that exist in the Chugach mountains as the wetter, more-dense maritime snow near the coast gradually changes to drier, less dense snow on the interior side.

2.1.1.5 Data Pre-Processing

Figure 3 demonstrates that it is not uncommon for automated snow pillow measurements to become noisy or non-physical, at times reporting large depths when there is no SWE reported. This is different from instances when physically plausible, but very low densities might be reported; say in response to early season dry, light snowfalls. It was therefore desirable to apply some objective, uniform procedure to each station's dataset in order to remove clear outlier points, while minimizing the removal of valid data points. We recognize that there is no accepted standardized method for cleaning bivariate SWE-*h* data sets. While Serreze et al. (1999) offer a procedure for SNOTEL data in their appendix, it is relevant only for precipitation and SWE values, not *h*. Given the strong correlation between *h* and SWE, we instead choose to use common outlier detection techniques for bivariate data.

The Mahalanobis distance (MD; Maesschalck et al., 2000) quantifies how far a point lies from the mean of a bivariate distribution. The distances are in terms of the number of standard deviations along the respective principal component axes of the distribution. For highly correlated bivariate data, the MD can be qualitatively thought of as a measure of how far a given point deviates from an ellipse enclosing the bulk of the data. One problem is that the MD is based on the statistical properties of the bivariate data (mean, covariance) and these properties can be adversely affected by outlier values. Therefore, it has been suggested (e.g., Leys et al., 2018) that a 'robust' MD (RMD) be calculated. The RMD is essentially the MD calculated based on statistical properties of the distribution unaffected by the outliers. This can be done using the Minimum Covariance Determinant (MCD) method as first introduced by Rousseeuw (1984).

Once RMDs have been calculated for a bivariate data set, there is the question of how large an RMD must be in order for the data point to be considered an outlier. For bivariate normal data, the distribution of the square of the RMD is χ^2 (Gnanadesikan and Kettenring, 1972), with p (the dimension of the dataset) degrees of freedom. So, a rule for identifying outliers could be implemented by selecting as a threshold some arbitrary quantile (say 0.99) of χ_p^2. For the current study, a threshold quantile of 0.999 was determined to be an appropriate compromise in terms of removing obviously outlier points, yet retaining physically plausible results.

A scatter plot of SWE vs. *h* for the SNOTEL dataset from CONUS and AK reveals many non-physical points, mostly when a very large *h* is reported for a very low SWE (Figure 4a). Approximately 0.7% of the original data points were removed in the pre-processing described above, creating a more physically plausible scatter plot (Figure 4b). Note that the outlier detection process was applied to each station individually. The distribution of 'day of year' (*DOY*) values of removed data points was broad, with a mean of 160 and a standard deviation of 65. Note that the *DOY* origin is 1 October. The same procedure was applied to the BC snow pillow, NRCS snow course, and northeast USA data sets as well (not shown). Table 1 summarizes useful information about the numerous data sets described above and indicates the final number of data points retained for each. We acknowledge that our process inevitably removes some valid data points, but, as a small percentage of an already small 0.7% removal rate, we judged this to be acceptable.

Table 1: Summary of information about the datasets used in this study. Datasets in bold font were used to construct the regression model. The numbers of stations and data points reflect the post-processed data.

Dataset Name	Dataset Type	Number of retained stations	Number and percentage of retained data points	Precision (*h* / SWE)
NRCS SNOTEL	**Snow pillow (SWE), ultrasonic (*h*)**	**791**	**1,900,000 (99.3%)**	**(0.5 in / 0.1 in)**
NRCS SCAN	Snow pillow (SWE), ultrasonic (*h*)	5	7094 (97.8%)	(0.5 in / 0.1 in)
British Columbia Snow Survey	**Snow pillow (SWE), ultrasonic (*h*)**	**31**	**61,000 (97.5%)**	**(1 cm / 1 mm)**
NRCS Snow Survey	Federal sampler / Aerial marker	1085	116,000 (99.6%)	(0.5 in / 0.1 in) for manual sampler (2 in / n/a) for aerial marker
Maine Geological Survey	Adirondack or Federal sampler (SWE and *h*)	431	28,000 (99.3%)	(0.5 in / 0.5 in)
Hubbard Brook (Station 2), NH	Mount Rose sampler (SWE and *h*)	1	704 (99.4%)	(0.1 in / 0.1 in)
Thompson Farm, NH	Snow core (SWE and *h*)	2	988 (99.4%)	0.5 in / 0.5 in)
Sleepers River, VT	Adirondack sampler	14	7214 (99.4%)	(0.5 in / 0.5 in)
New York Snow Survey	Adirondack or Federal sampler (SWE and *h*)	523	44,614 (98.2%)	(0.5 in / 0.5 in)
Chugach Mountains, AK	Avalanche probe (*h*)	71	71 (100%)	(1 cm)

2.1.2 Climatological Variables

30-year climate normals at 1 km resolution for North America were obtained from the ClimateNA project (Wang et al., 2016). This project provides grids for minimum, maximum, and mean temperature, and total precipitation for a given month. These grids are based on the PRISM normals (Daly et al., 1994) and are available for the periods 1961-1990 and 1981-2010. For this study, the more recent climatology was used. The ClimateNA project also provides a wide array of derived bioclimatic variables, such as precipitation as snow (*PAS*), frost-free-period (*FFP*), mean annual relative humidity (*RH*) and others. Wang et al. (2012) summarize these additional variables and how they are derived. Figure 5 shows gridded maps of winter (sum of December, January, February) precipitation (*PPTWT*) and the temperature difference (*TD*) between the mean temperature of the warmest month and the mean temperature of the coldest month. The latter variable (*TD*) is a measure of continentality.

2.2 Regression Model

In order to demonstrate the varying degrees of influence of explanatory variables, several regression models were constructed. In each case, the model was built by randomly selecting 50% of the paired SWE-*h* measurements from the aggregated CONUS, AK, and BC snow pillow datasets. The model was then validated by applying it to the remaining 50% of the dataset and comparing the modeled SWE to the observed SWE for those points. We constructed a second version of the regression models by randomly selecting 50% of the snow pillow stations and using all of the data from those stations. The model was then validated by applying it to the data from the remaining 50% of the stations. These two methods provided identical results, likely due to the very large sample size (N) of our dataset. In all cases, the p values from the linear regression were 0, again due to the large sample size. Additional validation was done with the northeast USA datasets (SCAN snow pillow and various snow coring datasets) and the NRCS snow course dataset, which were completely left out of the model building process.

2.2.1 One-Equation Model

The simplest equation, and one that is supported by the strong correlation seen in the portions of Figure 3 when SWE is present, is one that expresses SWE as a function of *h*. A linear model is attractive in terms of simplicity, but this limits the snowpack to a constant density. An alternative is to express SWE as a power law, i.e.,

(1) $$SWE = Ah^{a_1}.$$

This equation can be log-transformed into

(2) $$log_{10}(SWE) = log_{10}(A) + a_1 log_{10}(h)$$

which immediately allows for simple linear regression methods to be applied. With both *h* and SWE expressed in units of mm, the obtained coefficients are $(A, a_1) = (0.146, 1.102)$. Information on the performance of the model will be deferred until the results section.

2.2.2 Two-Equation Model

Recall from Figures 1 and 4 that there is a hysteresis loop in the SWE-*h* relationship. During the accumulation phase, snow densities are relatively low. During the ablation phase, the densities are relatively high. So, the same snowpack depth is associated with two different SWEs, depending upon the time of year. The regression equation given above does not resolve this difference. This can be addressed by developing two separate regression equations, one for the accumulation (*acc*) and one for the ablation (*abl*) phase. This approach takes the form

(3) $$SWE_{acc} = Ah^{a_1}; \quad DOY < DOY^*$$

(4) $$SWE_{abl} = Bh^{b_1}; \quad DOY \geq DOY^*$$

where *DOY* is the number of days from the start of the water-year, and DOY^* is the critical or dividing day-of-water-year separating the two phases. Put another way, DOY^* is the day of peak SWE. Interannual variability results in a range of DOY^* for a given site. Additionally, some sites, particularly the SCAN sites in the northeast USA, demonstrate multi-peak SWE profiles in some years. To reduce model complexity, however, we investigated the use of a simple climatological (long term average) value of DOY^* at each site. For each snow pillow station, the average DOY^* was computed over the period of record of that station. Analysis of all of the stations revealed that this average DOY^* was relatively well correlated with the climatological mean April maximum temperature (the average of the daily maximums recorded in April; $R^2 = 0.7$). However, subsequent regression analysis demonstrated that the SWE estimates were relatively insensitive to DOY^* and the best results were actually obtained when DOY^* was uniformly set to 180 for all stations. Again, with both SWE and h in units of mm, the regression coefficients turn out to be $(A, a_1) = (0.150, 1.082)$ and $(B, b_1) = (0.239, 1.069)$.

As these two equations are discontinuous at DOY^*, they are blended smoothly together to produce the final two-equation model

(5) $$SWE = SWE_{acc}\frac{1}{2}(1 - tanh[0.01\{DOY - DOY^*\}]) + SWE_{abl}\frac{1}{2}(1 + tanh[0.01\{DOY - DOY^*\}])$$

The coefficient 0.01 in the tanh function controls the width of the blending window and was selected to minimize the root mean square error of the model estimates.

2.2.3 Two-Equation Model with Climate Parameters

A final model was constructed by incorporating climatological variables. Again, the emphasis in this study is on methods that can be implemented at locations lacking the time series of weather variables that might be available at a weather or SNOTEL station. Climatological normals are unable to account for interannual variability, but they do preserve the high spatial gradients in climate that can lead to spatial gradients in snowpack characteristics. Stepwise linear regression was used to determine which variables to include in the regression. The initial list of potential variables included was

(6) $$SWE = f(h, z, PPTWT, PAS, TWT, TD, DOY, RH)$$

where z is the elevation (m), *PPTWT* is the winter (sum of December, January, February) precipitation (mm), *PAS* is mean annual precipitation as snow (mm), *TWT* is the winter (December, January, February) mean temperature (°C), *TD* is the difference between the mean temperature of the warmest month and the mean temperature of the coldest month (°C), *DOY* is the day of water year, and *RH* is the relative humidity (%). In the stepwise regression, explanatory variables were accepted only if they improved the adjusted R^2 value by 0.001. The result of the regression yielded

(7) $$SWE_{acc} = Ah^{a_1}PPTWT^{a_2}TD^{a_3}DOY^{a_4}; \quad DOY < DOY^*$$

(8) $$SWE_{abl} = Bh^{b_1}PPTWT^{b_2}TD^{b_3}DOY^{b_4}; \quad DOY \geq DOY^*$$

or, in log-transformed format,

(9) $$log_{10}(SWE_{acc}) = log_{10}(A) + a_1 log_{10}(h) + a_2 log_{10}(PPTWT) + a_3 log_{10}(TD) + a_4 log_{10}(DOY); \quad DOY < DOY^*$$

(10) $$log_{10}(SWE_{abl}) = log_{10}(B) + b_1 log_{10}(h) + b_2 log_{10}(PPTWT) + b_3 log_{10}(TD) + b_4 log_{10}(DOY); \quad DOY \geq DOY^*$$

indicating that only snow depth, winter precipitation, temperature difference, and day of water year were relevant. Manual tests of model construction with other variables included confirmed that Eqns. (7-8) yielded the best results. These two SWE estimates for the individual (*acc* and *abl*) phases of the snowpack were then blended with Eqn. (5) to produce a single equation for SWE spanning the entire water year. The obtained regression coefficients were $(A, a_1, a_2, a_3, a_4) = (0.0533,\ 0.9480,\ 0.1701, -0.1314, 0.2922)$ and $(B, b_1, b_2, b_3, b_4) = (0.0481, 1.0395, 0.1699, -0.0461, 0.1804)$. The physical interpretation of these coefficients is straightforward. For example, both a_2 and b_2 are greater than zero. So, for two locations with equal *h*, *DOY*, and *TD*, the location with greater *PPTWT* will have a greater SWE and therefore density. These locations are typically maritime climates with wetter, denser snow. In contrast, both a_3 and b_3 are less than zero. Therefore, for two locations with equal *h*, *DOY*, and *PPTWT*, the location with greater TD (a more continental climate) will have a lower density, which is again an expected result. These trends are similar in concept to Sturm et al. (2010), whose discrete snow classes (based on climate classes) indicate which snow will densify more rapidly.

3 Results

A comparison of the three regression models (one-equation model, Eq. (2); two-equation model, Eqs. (3-5); multi-variable two-equation model, Eqs. (5, 7-8)) is provided in Figure 6. The left column shows scatter plots of modeled SWE to observed SWE for the validation data set with the 1:1 line shown in black. The right column shows distributions of the model residuals. The vertical lines in the right column show the mean error, or model bias. Visually, it is clear that the one-equation model performs relatively poorly with a large negative bias. This large negative bias is partially overcome by the two-equation model (middle row, Figure 6). The cloud of points is closer to the 1:1 line and the vertical black line indicating the mean error is closer to zero. In the final row of Figure 6, we see that the multi-variable two-equation model yields the best result by far. The residuals are now evenly distributed with a small bias. Several metrics of performance for the three models, including R^2 (Pearson coefficient), bias, and root-mean-square-error (RMSE), are provided in Table 2. Figure 7 shows the distribution of model residuals for the multi-variable two-equation model as a function of DOY.

Table 2: Summary of performance metrics for the three regression models presented in Section 2.2.

Model	R^2	Bias (mm)	RMSE (mm)
One-equation	0.946	-19.5	102
Two-equation	0.962	-5.1	81
Multi-variable two-equation	0.978	-1.2	59

It is useful to also consider the model errors in a non-dimensional way. Therefore, an RMSE was computed at each station location and normalized by the winter precipitation ($PPTWT$) at that location. Figure 8 shows the probability density function of these normalized errors. The average RMSE is approximately 15% of $PPTWT$ with most values falling into the range of 5-30%. The spatial distribution of these normalized errors is shown in Figure 9. For the SNOTEL stations, it appears there is a slight regional trend, in terms of stations in continental climates (Rockies) having larger relative errors than stations in maritime climates (Cascades). The British Columbia stations also show higher relative errors.

3.1 Results for Snow Classes

A key objective of this study is to regress climatological information in a continuous rather than a discrete way. The work by Sturm et al. (2010) therefore provides a valuable point of comparison. In that study, the authors developed the following equation for density ρ_b

(11) $$\rho_b = (\rho_{max} - \rho_0)\left[1 - e^{(-k_1 h - k_2 DOY)}\right] + \rho_0$$

where ρ_0 is the initial density, ρ_{max} is the maximum or 'final' density (end of water year), k_1 and k_2 are coefficients, and *DOY* in this case begins on January 1. This means that their *DOY* for October 1 is -92. The coefficients vary with snow class and the values determined by Sturm et al. (2010) are shown in Table 3.

Table 3: Model parameters by snow class for Sturm et al. (2010).

Snow Class	ρ_{max}	ρ_0	k_1	k_2
Alpine	0.5975	0.2237	0.0012	0.0038
Maritime	0.5979	0.2578	0.0010	0.0038
Prairie	0.5941	0.2332	0.0016	0.0031
Tundra	0.3630	0.2425	0.0029	0.0049
Taiga	0.2170	0.2170	0.0000	0.0000

To make a comparison, the snow class for each SNOTEL and British Columbia snow survey (Rows 1 and 3 of Table 1) site was determined using a 1-km snow class grid (Sturm et al., 2010). The aggregated dataset from these stations was made up of 27% Alpine, 14% Maritime, 10% Prairie, 11% Tundra, and 38% Taiga data points. Equation (11) was then used to estimate snow density (and then SWE) for every point in the validation dataset described in Section 2.2. Figure 10 compares the SWE estimates from the Sturm model and from the current multi-variable, two-equation model (Equations 5, 7-8). The upper left panel of Figure 10 shows all of the data, and the remaining panels show the results for each snow class. In all cases, the current model provides better estimates (narrow cloud of points; closer to the 1:1 line). Plots of the residuals by snow class are provided in Figure 11, giving an indication of the bias of each model for each snow class. Summaries of the model performance, broken out by snow class, are given in Table 4. The current model has smaller biases and RMSEs for each snow class.

Table 4: Comparison of model performance by Sturm et al. (2010) and the current study.

Model	Sturm et al. (2010)			Multi-variable two-equation model		
Snow Class	R^2	Bias (mm)	RMSE (mm)	R^2	Bias (mm)	RMSE (mm)
All Data	0.928	-29.2	111	0.978	-1.2	59
Alpine	0.973	10.1	55	0.978	-2.7	48
Maritime	0.968	-16.8	109	0.975	-7.8	95
Prairie	0.967	18.7	56	0.971	-0.7	45
Tundra	0.956	-10.5	82	0.974	-2.9	59
Taiga	0.943	-80.0	151	0.978	2.6	54

3.2 Comparison to Pistocchi (2016)

In order to provide an additional comparison, the simple model of Pistocchi (2016) was also applied to the validation dataset. His model calculates the bulk density as

(12) $$\rho_b = \rho_0 + K(DOY + 61),$$

where ρ_0 has a value of 200 kg m^{-3} and K has a value of 1 kg m^{-3}. The DOY for this model has its origin at November 1. Application of this model to the validation dataset yields a bias of 55 mm and an RMSE of 94 mm. These results are comparable to the Sturm et al. (2010) model, with a larger bias but smaller RMSE.

3.3 Comparison to Jonas et al. (2009)

A final point of comparison can be provided by the model of Jonas et al. (2009). The full version of that model contains region-specific offset parameters that are not relevant to North America, so the following partial version of the model is used (their Eq. 4):

(13) $$\rho_b = ah + b,$$

where the parameters (a, b) vary with elevation and month, as given by Table 5. Note that coefficients are not given for every month. Application of the Jonas et al. (2009) model to the snow pillow dataset yields a bias of -5 mm and an RMSE of 69 mm. These results are not directly comparable to those of the current model (Table 2, row 3) since the Jonas et al. (2009) model is unable to compute results for several months of the year. To make a direct comparison to the current model, it is necessary to first remove those data points (about 5%). When this is done, the current model yields a bias of -0.3 mm and an RMSE of 59 mm.

Table 5: Model coefficients (a, b) for the Jonas et al. (2009) model.

Month	$z > 2000\ m$	$2000\ m > z > 1400\ m$	$1400\ m > z$
January	$(206, 52)$	$(208, 47)$	$(235, 31)$

February	(217, 46)	(218, 52)	(279, 9)
March	(272, 26)	(281, 31)	(333, 3)
April	(331, 9)	(354, 15)	(347, 25)
May	(378, 21)	(409, 29)	(413, 19)
June	(452, 8)	n/a	n/a
July	(470, 15)	n/a	n/a
August	n/a	n/a	n/a
September	n/a	n/a	n/a
October	n/a	n/a	n/a
November	(206, 47)	(183, 35)	(149, 37)
December	(203, 52)	(190, 47)	(201, 26)

3.4 Results for Northeast USA

The regression equations in this study were developed using a large collection of snow pillow sites in CONUS, AK, and BC. The snow pillow sites are limited to locations west of approximately W 105° (Figure 2a). By design, the data sets from the northeastern USA (Section 2.1.1.3) were left as an entirely independent validation set. These northeastern sites are geographically distant from the training data sets, subject to a very different climate, largely use different methods (snow coring, with the exception of the SCAN network) and are generally at much lower elevations than the western sites, providing an interesting opportunity to test how robust the current model is.

Figure 12 graphically summarizes the datasets and the performance of the multi-variable two-equation model of the current study. The RMSE values are comparable to those found for the western stations, but, given the comparatively thinner snowpacks in the northeast, represent a larger relative error (Table 5). The bias of the model is consistently positive, in contrast to the western stations where the bias was negligible. Note that Table 5 also includes results from the application of the other three models discussed. Sturm et al. (2010) cannot be applied to several of the datasets since their available 1 km snowclass dataset cuts off at -71.6° longitude. The current model and the Jonas et al. (2009) model perform better than the other two models, with the current model generally outperforming the Jonas et al. (2009) model. The two datasets where the Jonas et al. (2009) model has a slightly better performance are the two smallest datasets (less than 1000 measurements; see Table 1).

Table 5: Performance metrics for various models applied to the northeastern USA datasets. Bold font is used to highlight the model with the best performance for each dataset.

	Multi-variable, two-equation model		Sturm et al. (2010)		Jonas et al. (2009)		Pistocchi (2015)	
Dataset Name	Bias (mm)	RMSE (mm)	Bias (mm)	RMSE (mm)	Bias (mm)	RMSE (mm)	Bias (mm)	RMSE (mm)
Maine Geological Survey, ME	**13.1**	**34.0**	n/a	n/a	25.1	46.0	59.2	77.1
Hubbard Brook (Station 2), NH	21.8	66.6	34.2	76.9	**19.4**	**65.4**	52.0	90.8
Thompson Farm, NH	7.1	20.2	n/a	n/a	**5.6**	**19.9**	20.4	32.3
NRCS SCAN	**-1.2**	**39.2**	8.4	45.0	-2.8	40.6	23.4	56.9
Sleepers River, VT	**14.4**	**28.2**	36.5	48.9	20.4	33.5	55.8	67.1
New York Snow Survey	**14.8**	**31.2**	21.0	49.3	16.3	33.0	41.3	56.1

3.5 Results for NRCS Snow Course / Aerial Marker Data

The NRCS snow course and aerial marker data were also left out of the model building process so they provide an additional and completely independent comparison of the various models considered. Recall that these data come from snow course (coring measurements) and aerial surveys, which are different measurement methods than the snow pillows which provided the data for construction of the current regression model. Table 6 shows the results and demonstrates that the current model has the best performance.

Table 6: Performance metrics for various models applied to the NRCS snow course and aerial marker dataset. Bold font is used to highlight the model with the best performance.

	Multi-variable, two-equation model		Sturm et al. (2010)		Jonas et al. (2009)		Pistocchi (2015)	
Dataset Name	Bias (mm)	RMSE (mm)	Bias (mm)	RMSE (mm)	Bias (mm)	RMSE (mm)	Bias (mm)	RMSE (mm)
NRCS Snow Course / Aerial Marker	**0**	**59**	-24	123	24	72	71	99

4 Discussion

The results presented in this study show that the regression equation described by equations (5, 7-8) is an improvement (lower bias and RMSE) over other widely used bulk density equations. The key advantage is that the current method regresses in relevant parameters directly, rather than using discrete bins (for snow class, elevation, month of year, etc.), each with its own set of model coefficients. The comparison (Figs. 10-11; Table 4) to the model of Sturm et al. (2010) reveals a peculiar behavior of that model for the Taiga snow class, with a large negative bias in the Sturm estimates. Inspection of the coefficients provided for that class (Table 3) shows that the model simply predicts that $\rho_b = \rho_{max} = 0.217$ for all conditions.

When our multi-variable two-equation model, developed solely from western North American data, is applied to northeast USA locations, it produces SWE estimates with smaller RSME values and larger biases than the western stations. When comparing the SWE-h scatter plots of the SNOTEL data (Figure 4b) to those of the east coast data sets (left column; Figure 12), it is clear that the northeast data generally have more scatter. This is confirmed by computing the correlation coefficients between SWE and h for each dataset. It is unclear if this disparity in correlation is related to measurement methodology or is instead a ‘signal to noise’ issue. Comparing Figures 4 and 12 shows the considerable difference in snowpack depth between the western and northeastern data sets. When the western dataset is filtered to include only measurement pairs where $h < 1.5$ m, the correlation coefficient is reduced to a value consistent with the northeast datasets. This suggests that the performance of the current (or other) regression model is not as good at shallow snowpack depths. This is also suggested upon examination of the time series of observed $\rho_b = SWE/h$ for a given season at a snow pillow site. Very early in the season, when the depths are small, the density curve has a lot of variability. Later in the season, when depths are greater, the density curve becomes much smoother. Very late in the season, when depths are low again, the density curve becomes highly variable again.

Measurement precision and accuracy affect the construction and use of a regression model. Upon inspection of the snow pillow data, it was observed that the precision of the depth measurements was approximately 25 mm and that of the SWE measurements was approximately 2.5 mm. To test the sensitivity of the model coefficients to the measurement precision, the depth values in the training dataset were randomly perturbed by +/- 25 mm and the SWE values were randomly perturbed by +/- 2.5 mm and the regression coefficients were recomputed. This process was repeated numerous times and the mean values of the perturbed coefficients were obtained. These adjusted coefficients were then used to recompute the SWE values for the validation data set and the bias and RMSE were found to be -10.5 mm and 72.7 mm. This represents a roughly 10% increase in RMSE, but a considerable increase in bias magnitude (see Table 4 for the original values). This sensitivity of the regression analysis to measurement precision underscores the need to have high-precision measurements for the training data set. Regarding accuracy, random and systematic errors in the paired SWE - h data used to construct the regression model will lead to uncertainties in SWE values predicted by the model. As noted in the introduction, snow pillow errors in SWE estimates do not follow a simple pattern. Additionally, they are complicated by the fact that the errors are often computed by comparing snow pillow data to coring data, which itself is subject to error. Lacking quantitative information on the distribution of snow pillow errors, we are unable to quantify the uncertainty in the SWE estimates.

Another important consideration has to do with the uncertainty of depth measurements that the model is applied to. For context, one application of this study is to crowd-sourced, opportunistic snow depth measurements from programs like the Community Snow Observations (CSO; Hill et al., 2018) project. In the CSO program, backcountry recreational users submit depth measurements, typically taken with an avalanche probe, using a smartphone in the field. The measurements are then converted to SWE estimates which are assimilated into snowpack models. These depth measurements are ‘any time, any place’ in contrast to repeated measurements from the same location, like snow pillows or snow courses. Most avalanche probes have cm-scale graduated markings, so measurement precision is not a major issue. A larger problem is the considerable variability in snowpack depth that can exist over short (meter scale) distances. The variability of the Chugach avalanche probe measurements was assessed by taking the standard deviation of 8 h measurements per site. The average of this standard deviation over the sites was 22 cm and the average coefficient of variation (standard deviation normalized by the mean) over the sites was 15%. This variability is a function of the surface roughness of the underlying terrain, and also a function of wind redistribution of snow. Propagating this uncertainty through the regression equations yields a slightly higher (16%) uncertainty in the SWE estimates. CSO participants can do three things to ensure that their recorded depth measurements are as representative as possible. First, avoid measurements in areas of significant wind scour or deposition. Second, avoid measurements in terrain likely to have significant surface roughness (rocks, fallen logs, etc.). Third, take several measurements and average them.

Expansion of CSO measurements in areas lacking SWE measurements can increase our understanding of the extreme spatial variability in snow distribution and the inherent uncertainties associated with modeling SWE in these regions. It could also prove useful for estimating watershed-scale SWE in regions like the northeastern USA, which is currently limited to five automated SCAN sites with historical SWE measurements for only the past two decades. Additionally, historical snow depth measurements are more widely available in the Global Historical Climatology Network (GHCN-Daily; Menne et al. 2012), with several records extending back to the late 1800s. While many of the GHCN stations are confined to lower elevations with shallower snow depths, the broader network of quality-controlled snow depth data paired with daily GHCN temperature and precipitation measurements could potentially be used to reconstruct SWE in the eastern US given additional model development and refinement.

5 Conclusions

We have developed a new, easy to use method for converting snow depth measurements to snow water equivalent estimates. The key difference between our approach and previous approaches is that we directly regress in climatological variables in a continuous fashion, rather than a discrete one. Given the abundance of freely available climatological norms, a depth measurement tagged with coordinates (latitude and longitude) and a time stamp is easily and immediately converted into SWE.

We developed this model with data from paired SWE-*h* measurements from the western United States and British Columbia. The model was tested against entirely independent data (primarily snow course; some snow pillow) from the northeastern United States and was found to perform well, albeit with larger biases and root-mean-squared-errors. The model was tested against other well-known regression equations and was found to perform better. The model was also tested against a large dataset of independent snow course and aerial marker measurements from western North America. For this second independent test, the current model outperformed the other models considered.

This model is not a replacement for more sophisticated snow models that evolve the snowpack based on high frequency (e.g., daily or sub-daily) weather data inputs. The intended purpose of this model is to constrain SWE estimates in circumstances where snow depth is known, but weather variables are not, a common issue in sparsely instrumented areas in North America.

6 Acknowledgements

Support for this project was provided by NASA (NNX17AG67A). R. Crumley acknowledges support from the CUAHSI Pathfinder Fellowship. E. Burakowski acknowledges support from NSF (MSB-ECA #1802726). We thank M. Sturm, A. Winstral and a third anonymous referee for their careful and thoughtful reviews of this manuscript.

7 Data Access

Numerous online datasets were used for this project and were obtained from the following locations:

1. NRCS Snow Telemetry: https://www.wcc.nrcs.usda.gov/snow/SNOTEL-wedata.html
2. NRCS Soil Climate Analysis Network: https://www.wcc.nrcs.usda.gov/scan/

3. British Columbia Automated Snow Weather Stations: https://www2.gov.bc.ca/gov/content/environment/air-land-water/water/water-science-data/water-data-tools/snow-survey-data/automated-snow-weather-station-data
4. Maine Cooperative Snow Survey: https://mgs-maine.opendata.arcgis.com/datasets/maine-snow-survey-data
5. New York Snow Survey: http://www.nrcc.cornell.edu/regional/snowsurvey/snowsurvey.html
6. Sleepers River Research Watershed. Snow data not available online; request data from contact at: https://nh.water.usgs.gov/project/sleepers/index.htm
7. Hubbard Brook Experimental Forest: https://hubbardbrook.org/d/hubbard-brook-data-catalog
8. Climatological Data: https://adaptwest.databasin.org/pages/adaptwest-climatena
9. NRCS Snow Course / Aerial Marker Data: https://wcc.sc.egov.usda.gov/reportGenerator/

A Matlab function for calculating SWE based on the results is this paper has been made publicly available at Github (https://github.com/communitysnowobs/snowdensity).

References

Alford, D.: Density variations in alpine snow, J. Glaciol., 6(46), 495-503, https://doi.org/10.3189/S0022143000019717, 1967.

Avanzi, F., De Michele, C., and Ghezzi, A.: On the performances of empirical regressions for the estimation of bulk snow density, Geogr. Fis. Dinam. Quat., 38, 105-112, doi:10.4461/GFDQ.2015.38.10, 2015.

Beaumont, R.: Mt. Hood pressure pillow snow gage, J. Appl. Meteorol., 4, 626-631, https://doi.org/10.1175/1520-0450(1965)004<0626:MHPPSG>2.0.CO;2 1965.

Beaumont, R., and Work, R.: Snow sampling results from three samplers, Hydrol. Sci. J., 8(4), 74-78, https://doi.org/10.1080/02626666309493359, 1963.

Burakowski, E.A., Ollinger, S., Lepine, L., Schaaf, C.B., Wang, Z., Dibb, J.E., Hollinger, D.Y., Kim, J.-H., Erb, A., and Martin, M.E.: Spatial scaling of reflectance and surface albedo over a mixed-use, temperate forest landscape during snow-covered periods, Remote Sens. Environ., 158, 465-477, https://doi.org/10.1016/j.rse.2014.11.023, 2015.

Burakowski, E.A., Wake, C.P., Stampone, M., and Dibb, J.: Putting the Capital 'A' in CoCoRAHS: An Experimental Program to Measure Albedo using the Community Collaborative Rain Hail and Snow (CoCoRaHS) Network, Hydrol. Process., 27(21), 3024-3034, https://doi.org/10.1002/hyp.9825, 2013.

Campbell, J., Ollinger, S., Flerchinger, G., Wicklein, H., Hayhoe, K., and Bailey, A.: Past and projected future changes in snowpack and soil frost at the Hubbard Brook Experimental Forest, New Hampshire, USA, Hydrol. Process., 24, 2465-2480, https://doi.org/10.1002/hyp.7666, 2010.

Church, J.E.: Snow surveying: its principles and possibilities, Geogr. Rev., 23(4), 529-563, DOI: 10.2307/209242, 1933.

Church, J.E., and Marr, J.C.: Further improvement of snow-survey apparatus, Transactions of the American Geophysical Union, 18(2), 607-617, 10.1029/TR018i002p00607, 1937.

Daly, C., Neilson, R., and Phillips, D.: A statistical-topographic model for mapping climatological precipitation over mountainous terrain, J. Appl. Meteorol., 33, 140-158, https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2, 1994.

Dixon, D., and Boon, S.: Comparison of the SnowHydro sampler with existing snow tube designs, Hydrol. Process., 26(17), 2555-2562, https://doi.org/10.1002/hyp.9317, 2012.

Dressler, K., Fassnacht, S., and Bales, R.: A comparison of snow telemetry and snow course measurements in the Colorado River basin, J. Hydrometeorol., 7, 705-712, https://doi.org/10.1175/JHM506.1, 2006.

Goodison, B.: Accuracy of snow samplers for measuring shallow snowpacks: An update, Proceedings of the 35th Annual Eastern Snow Conference, Hanover, NH, 36-49, 1978.

Goodison, B., Ferguson, H., and McKay, G.: Measurement and data analysis. The Handbook of Snow: Principles, Processes, Management, and Use, D. Gray and D. Male, Eds., Pergamon Press, 191-274, 1981.

Goodison, B., Wilson, B., Wu., K, and Metcalfe, J.: An inexpensive remote snow-depth gauge: An assessment, Proceedings of the 52nd Annual Western Snow Conference, Sun Valley, ID, 188-191, 1984.

Goodison, B., Glynn, J., Harvey, K., and Slater, J.: Snow Surveying in Canada: A Perspective, Can. Water Resour. J., 12(2), 27-42, https://doi.org/10.4296/cwrj1202027, 1987.

Gnanadesikan, R., and Kettenring, J.: Robust estimates, residuals, and outlier detection with multiresponse data, Biometrics, 28, 81-124, DOI: 10.2307/2528963, 1972.

Hill, D.F., Wolken, G. J., Wikstrom Jones, K., Crumley, R., and Arendt, A.: Crowdsourcing snow depth data with citizen scientists, Eos, 99, https://doi.org/10.1029/2018EO108991, 2018.

Johnson, J., and Marks, D.: The detection and correction of snow water equivalent pressure sensor errors, Hydrol. Proc., https://doi.org/10.1002/hyp.5795, 2004.

Johnson, J., Gelvin, A., Duvoy, P., Schaefer G., Poole, G., and Horton, G.: Performance characteristics of a new electronic snow water equivalent sensor in different climates, Hydrol. Proc., DOI: 10.1002/hyp.10211, 2015.

Jonas, T., Marty, C., and Magnusson, M.: Estimating the snow water equivalent from snow depth measurements, J. Hydrol., 378, 161-167, https://doi.org/10.1016/j.jhydrol.2009.09.021, 2009.

Leys, C., Klein, O., Dominicy, Y., and Ley, C.: Detecting multivariate outliers: use a robust variant of the Mahalanobis distance, J. Exp. Soc. Psychol., 74, 150-156, https://doi.org/10.1016/j.jesp.2017.09.011, 2018.

Liang, X., Lettermaier, D., Wood, E., and Burges, S.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, J. Geophys. Res. Atmos., 99(D7), 14,415-14,428, https://doi.org/10.1029/94JD00483, 1994.

Liston, G., and Elder, K.: A distributed snow evolution modeling system (SnowModel), J. Hydrometerol., 7, 1259-1276, https://doi.org/10.1175/JHM548.1, 2006.

Lundberg, A., Richardson-Naslund, C., and Andersson, C.: Snow density variations: consequences for ground penetrating radar, Hydrol. Process., 20, 1483-1495, https://doi.org/10.1002/hyp.5944, 2006.

Maine Geological Survey: Maine Cooperative Snow Survey Dataset, https://www.maine.gov/dacf/mgs/hazards/snow_survey/, 2018.

McKay, G., and Findlay, B., 1971: Variation of snow resources with climate and vegetation in Canada, Proceedings of the 39th Western Snow Conference, Billings, MT, 17-26, 1971.

De Maesschalck, R., Jouan-Rimbaud, D., and Massart, D.: The Mahalanobis distance, Chemometr. Intell. Lab. Syst., 50(1), 1-18, https://doi.org/10.1016/S0169-7439(99)00047-7, 2000.

McCreight, J., and Small, E.: Modeling bulk density and snow water equivalent using daily snow depth observations, The Cryosphere, 8, 521-536, https://doi.org/10.5194/tc-8-521-2014, 2014.

Meloysund, V., Leira, B., Hoiseth, K., and Liso, K.: Predicting snow density using meterological data, Meteorol. Appl., 14, 413-423, https://doi.org/10.1002/met.40, 2007.

Menne, M.J., I. Durre, R.S. Vose, B.E. Gleason, and Houston, T.G.: An overview of the Global Historical Climatology Network-Daily Database, J. Atmos. Ocean. Technol., 29, 97-910, doi:10.1175/JTECH-D-11-00103.1, 2012.

Molotch, N.P., and Bales, R.C.: SNOTEL representativeness in the Rio Grande headwaters on the basis of physiographics and remotely sensed snow cover persistence, Hydrol. Process., 20(4), 723-739, https://doi.org/10.1002/hyp.6128, 2006.

Mote, P., Li, S., Letternaier, D., Xiao, M., and Engel, R.: Dramatic declines in snowpack in the western US,” npj Clim. Atmos. Sci., 1(2), 1-6, doi:10.1038/s41612-018-0012-1, 2018.

Mizukami, N., and Perica, S.: Spatiotemporal characteristics of snowpack density in the mountainous regions of the western United States, J. Hydrometeorol., 9, 1416-1426, https://doi.org/10.1175/2008JHM981.1, 2008.

New York Snow Survey, NOAA, Northeast Regional Climate Center at Cornell University, 2018.

Pagano, T., Garen, D., Perkins, T., and Pasteris, P.: Daily updating of operational statistical seasonal water supply forecasts for the western U.S., J. Am. Water Resour. Assoc., 45(3), 767-778, https://doi.org/10.1111/j.1752-1688.2009.00321.x, 2009.

Painter, T., Berisford, D., Boardman, J., Bormann, K., Deems, J., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw, R., Marks, D., Mattmann, C., Mcgurk, B., Ramirez, P., Richardson, M., Skiles, S., Seidel, F., and Winstral, A.: The Airborne Snow Observatory: fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo, Remote Sens. Environ., 184, 139-152, doi:10.1016/j.rse.2016.06.018, 2016.

Pistocchi, A.: Simple estimation of snow density in an Alpine region, J. Hydrol. Reg. Stud., 6, 82-89, http://dx.doi.org/10.1016/j.ejrh.2016.03.004, 2016.

Rousseeuw, P.: Least Median of Squares Regression, J. Am. Stat. Assoc., 79, 871-880, DOI: 10.1080/01621459.1984.10477105, 1984.

Ryan, W., Doesken, N., and Fassnacht, S.: Evaluation of Ultrasonic Snow Depth Sensors for U.S. Snow Measurements, J. Atmos. Ocean. Technol., 25, 667-684, https://doi.org/10.1175/2007JTECHA947.1, 2008.

Schaefer, G., Cosh, M., and Jackson, T.: The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN), J. Atmos. Ocean. Technol., 24, 2073-2077, https://doi.org/10.1175/2007JTECHA930.1, 2007.

Serreze, M., Clark, M., Armstrong, R., McGinnis, D., and Pulwarty, R.: Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data, Water Resour. Res., 35(7), 2145-2160, https://doi.org/10.1029/1999WR900090, 1999.

Shanley, J., and Chalmers, A.: The effect of frozen soil on snowmelt runoff at Sleepers River, Vermont, Hydrol. Process., 13(12-13), 1843-1857, https://doi.org/10.1002/(SICI)1099-1085(199909)13:12/13<1843::AID-HYP879>3.0.CO;2-G, 1999.

Sturm, M., Holmgren, J., and Liston, G.: A seasonal snow cover classification system for local to global applications, J. Clim., 8, 1261-1283, https://doi.org/10.1175/1520-0442(1995)008<1261:ASSCCS>2.0.CO;2, 1995.

Sturm, M., Taras, B., Liston, G.E., Derksen, C., Jonas, T., and Lea, J.: Estimating snow water equivalent using snow depth data and climate classes, J. Hydrometeorol., 11, 1380-1394, https://doi.org/10.1175/2010JHM1202.1, 2010.

U.S. Army Corps of Engineers: Snow hydrology: Summary report of the snow investigations of the North Pacific Division, 437pp., 1956.

U.S. Department of Agriculture: The History of Snow Survey and Water Supply Forecasting. Interviews With U.S. Department of Agriculture Pioneers, D. Helms, S. Phillips and P. Reich (eds.), Natural Resources Conservation Service, 2008.

U.S. Department of Agriculture: Snow Survey and Water Supply Forecasting. National Engineering Handbook Part 622, Water and Climate Center, Natural Resources Conservation Service, 2011.

Wang, T., Hamann, A., Spittlehouse, D.L., and Murdock, T.: ClimateWNA - High-Resolution Spatial Climate Data for Western North America, J. Appl. Meteorol. Climatol., 51, 16-29, https://doi.org/10.1175/JAMC-D-11-043.1, 2012.

Wang, T., Hamann, A., Spittlehouse, D.L., and Carroll, C: Locally downscaled and spatially customizable climate data for historical and future periods for North America, PLoS One, 11, DOI:10.1371/journal.pone.0156720.

Wigmosta, M.S., Vail, L., and Lettenmaier, D.: A distributed hydrology-vegetation model for complex terrain, Water Resour. Res., 30, 1665-1679, https://doi.org/10.1029/94WR00436, 1994

Figure 1: Conceptual sketch of the evolution of snow water equivalent (SWE) over the course of a water year (black line). Also shown is the evolution of SWE with snowpack depth over a water year (red line). Note the hysteresis loop due to the densification of the snowpack.

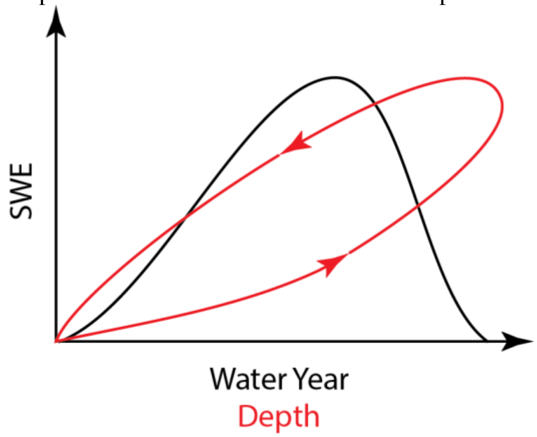


Figure 2: Distribution of measurement locations used in this study. (a) Western USA and Canada snow pillow locations, with colors indicating station elevation in meters. (b) Northeast USA snow pillow and snow course locations, with stations colored according to data source. (c) Western North America snow course and aerial marker locations, with colors indicating station elevation in meters. (d, e) Measurement sites in the Chugach Mountains, southcentral Alaska.

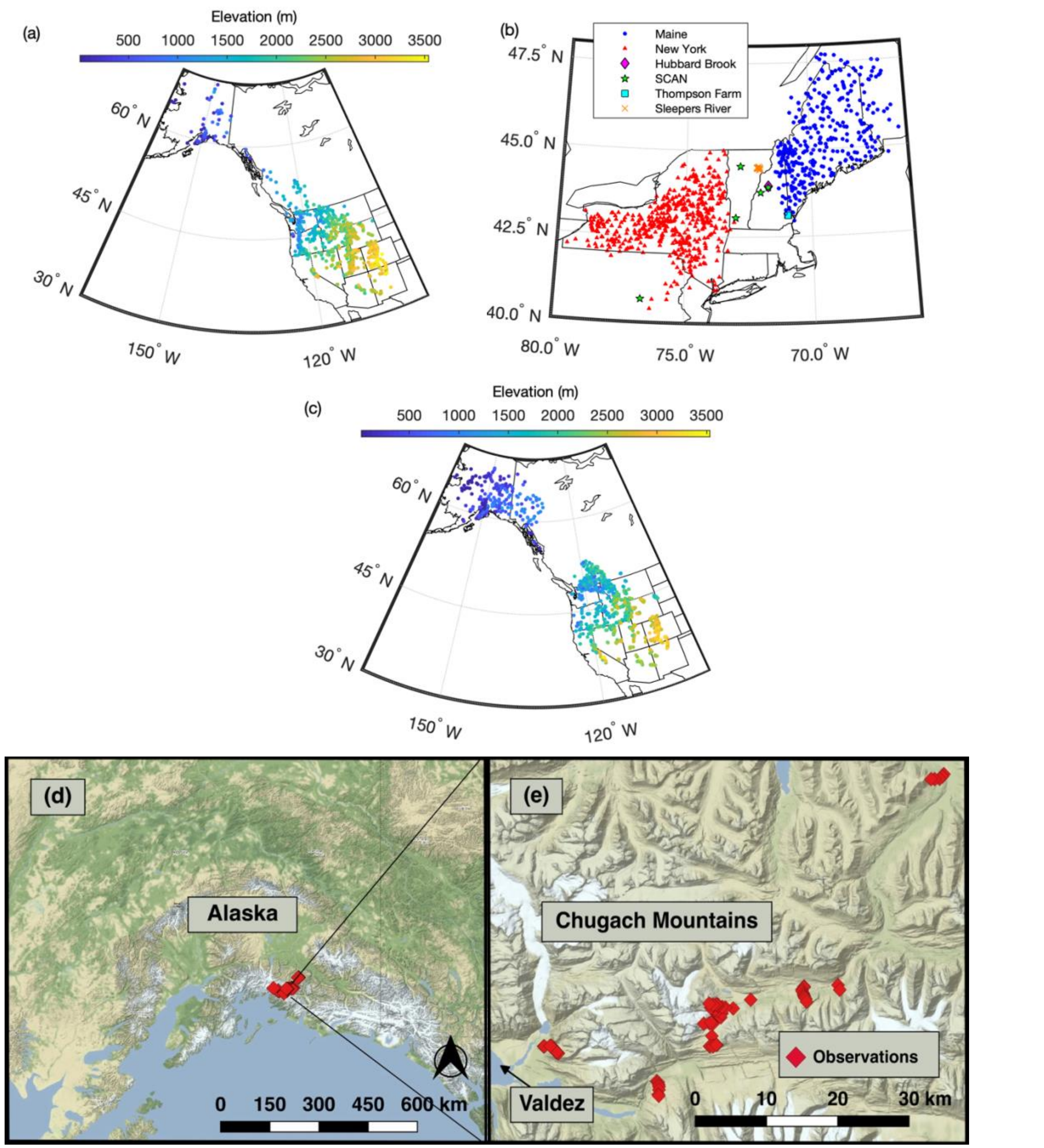


Figure 3: Sample time series of SWE and *h* from the Rex River (WA) SNOTEL station. Observations of *h* at times when SWE is zero are likely spurious.

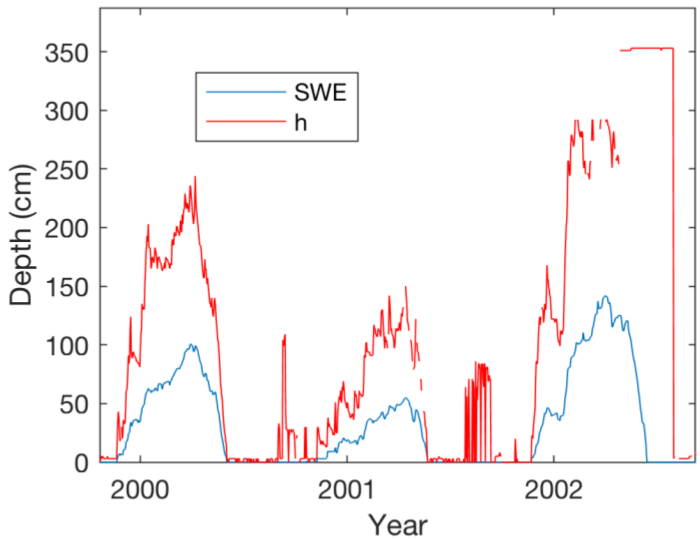


Figure 4: Scatter plot of SWE vs. *h* for the complete SNOTEL dataset before (a) and after (b) removing data points, following the method described in Section 2.1.1.5. Symbols are colored by 'day of water year' (*DOY*; October 1 is the origin).

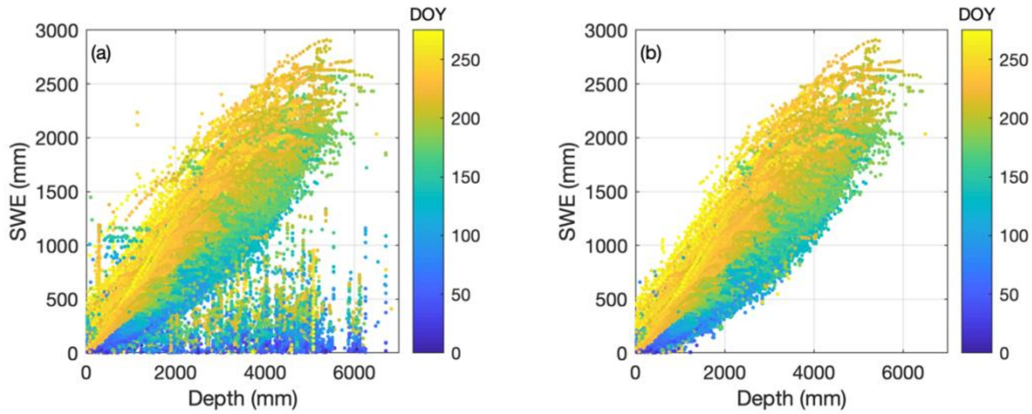


Figure 5: Gridded maps of winter (December, January, February) precipitation (PPTWT) and temperature difference (TD) between mean of warmest month and mean of coldest month) for North America. Maps are for the 1981-2010 climatological period.

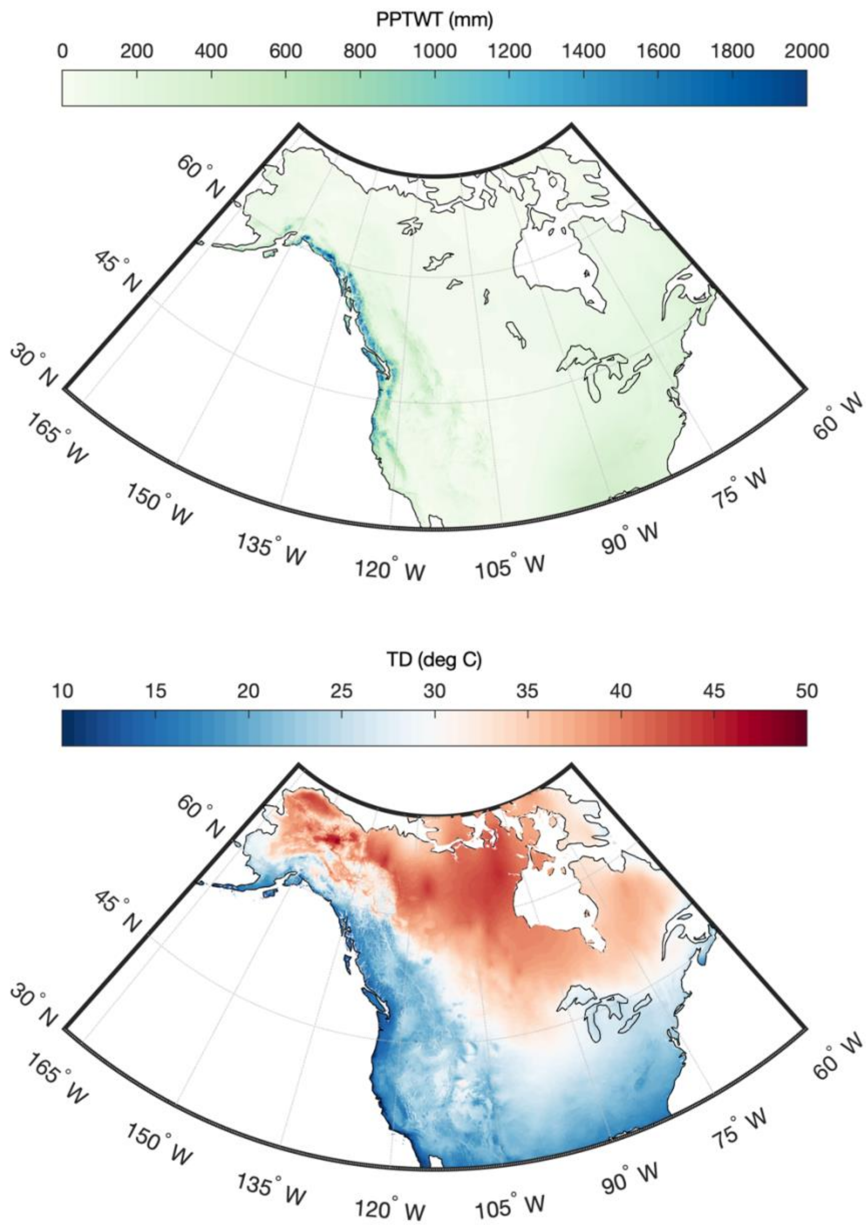


Figure 6: Two-dimensional histograms (heat maps; left column) of modeled vs. observed SWE and probability density functions (right column) of the residuals for three simple models applied to the CONUS, AK, and BC snow pillow data. Warmer colors in the heat maps indicate greater density. The vertical lines in the right column indicate the location of the mean residual, or bias. Top row (a-b): One-equation model (Section 2.2.1). Middle row (c-d): Two-equation model (Section 2.2.2). Bottom row (e-f): Multi-variable two-equation model (Section 2.2.3).

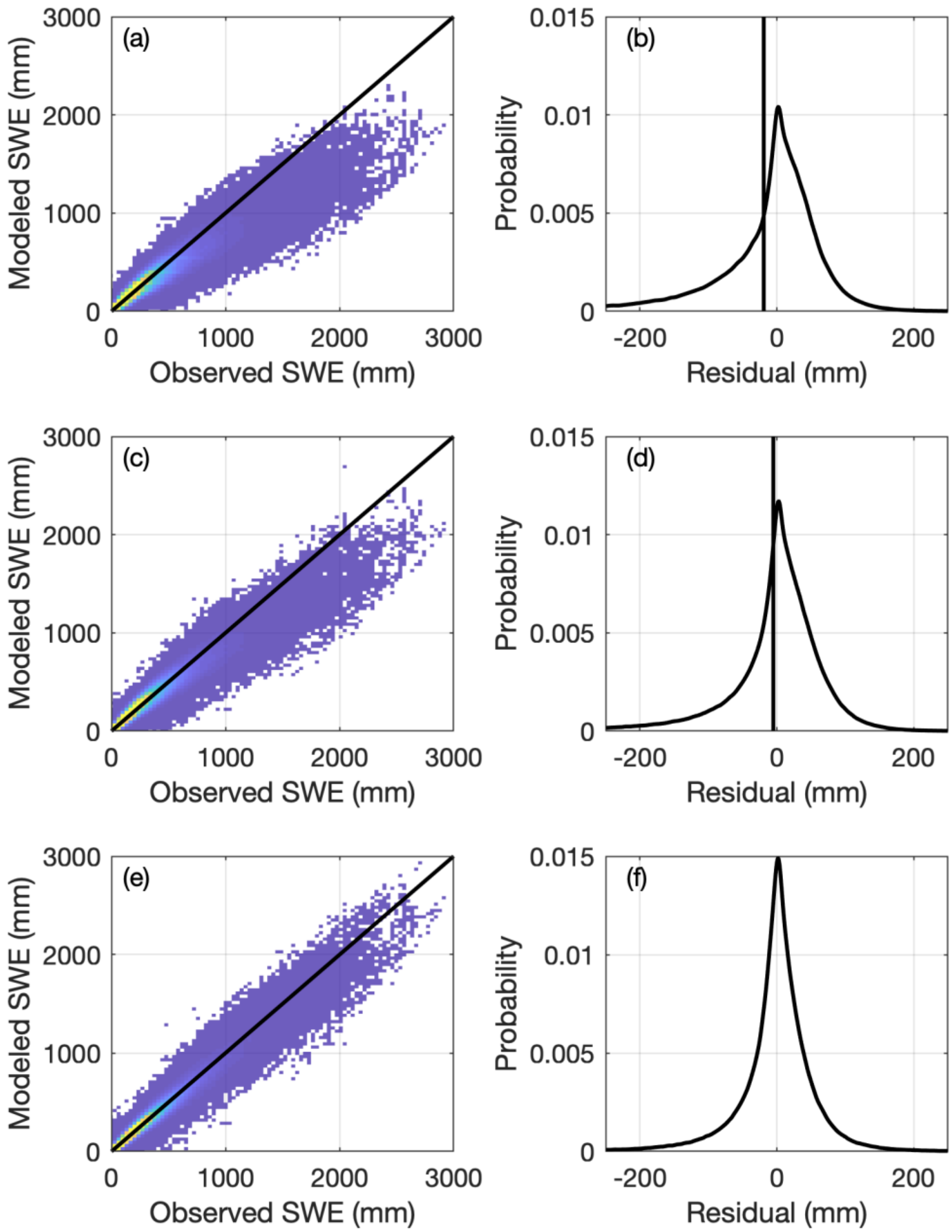


Figure 7: Heat map of SWE residuals as a function of *DOY*.

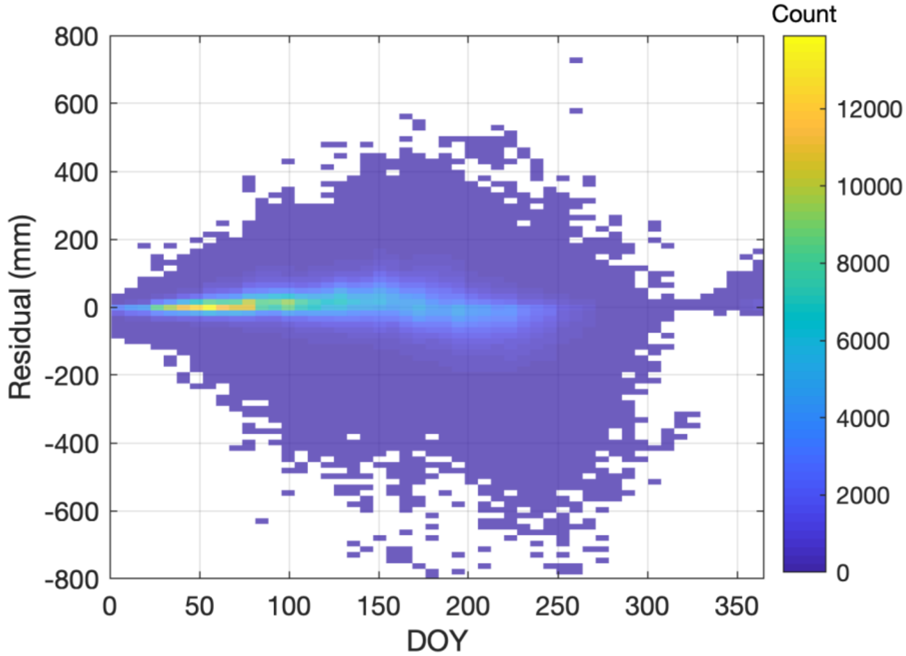


Figure 8: Probability density function of snow pillow station root-mean-square error (RMSE) normalized by station winter precipitation (*PPTWT*).

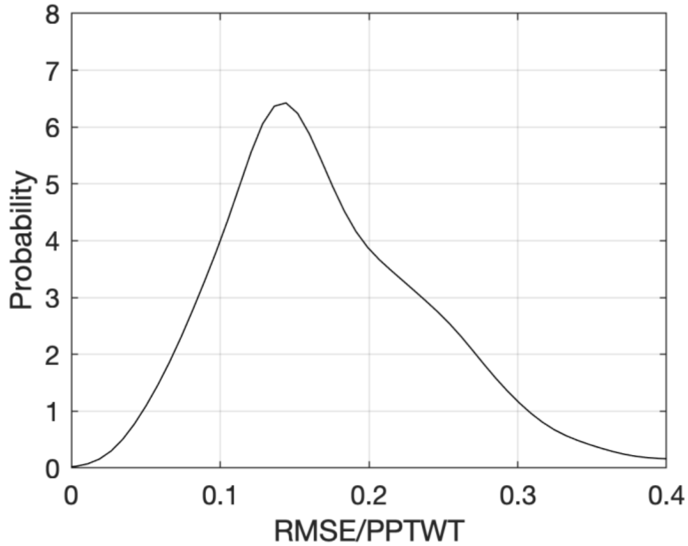


Figure 9: Spatial distribution of snow pillow station root-mean-square error (RMSE) normalized by station winter precipitation (PPTWT).

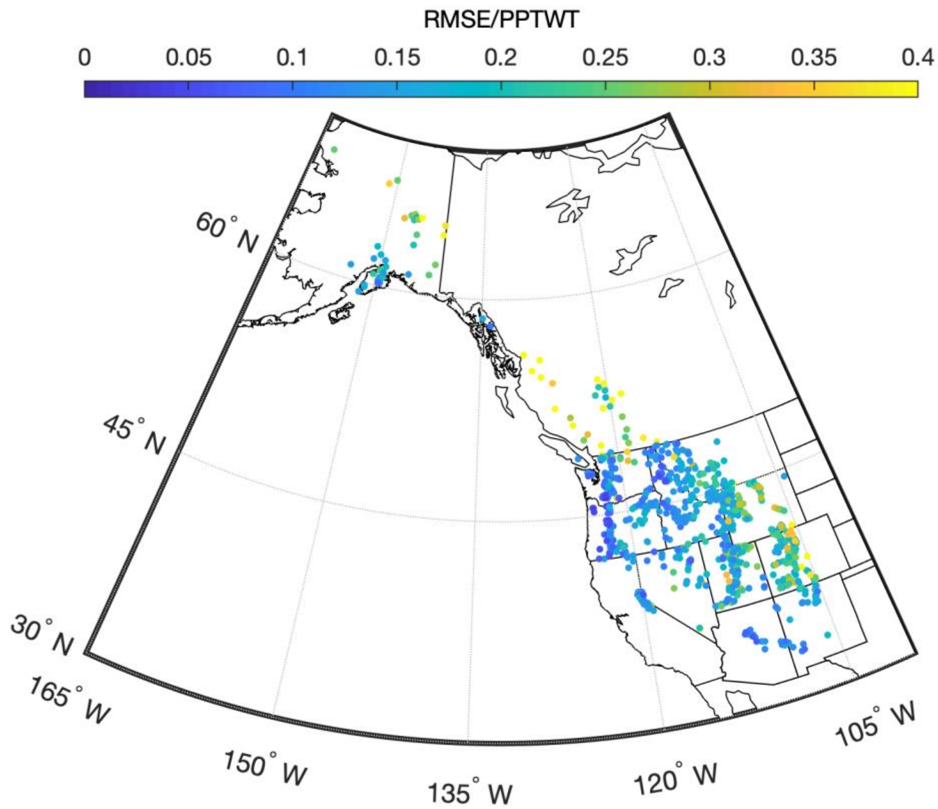


Figure 10: Comparison of the multi-variable, two-equation model of the current study with the model of Sturm et al. (2010). The subpanels show modeled SWE vs. observed SWE for all of the data binned together, as well as for the data broken out by the snow classes identified by Sturm et al. (1995). The gray symbols show the Sturm result and the transparent heat maps (warmer colors indicate greater density) show the current result. The models are being applied to the validation data set (50% of the aggregated snow pillow data for CONUS, AK, and BC).

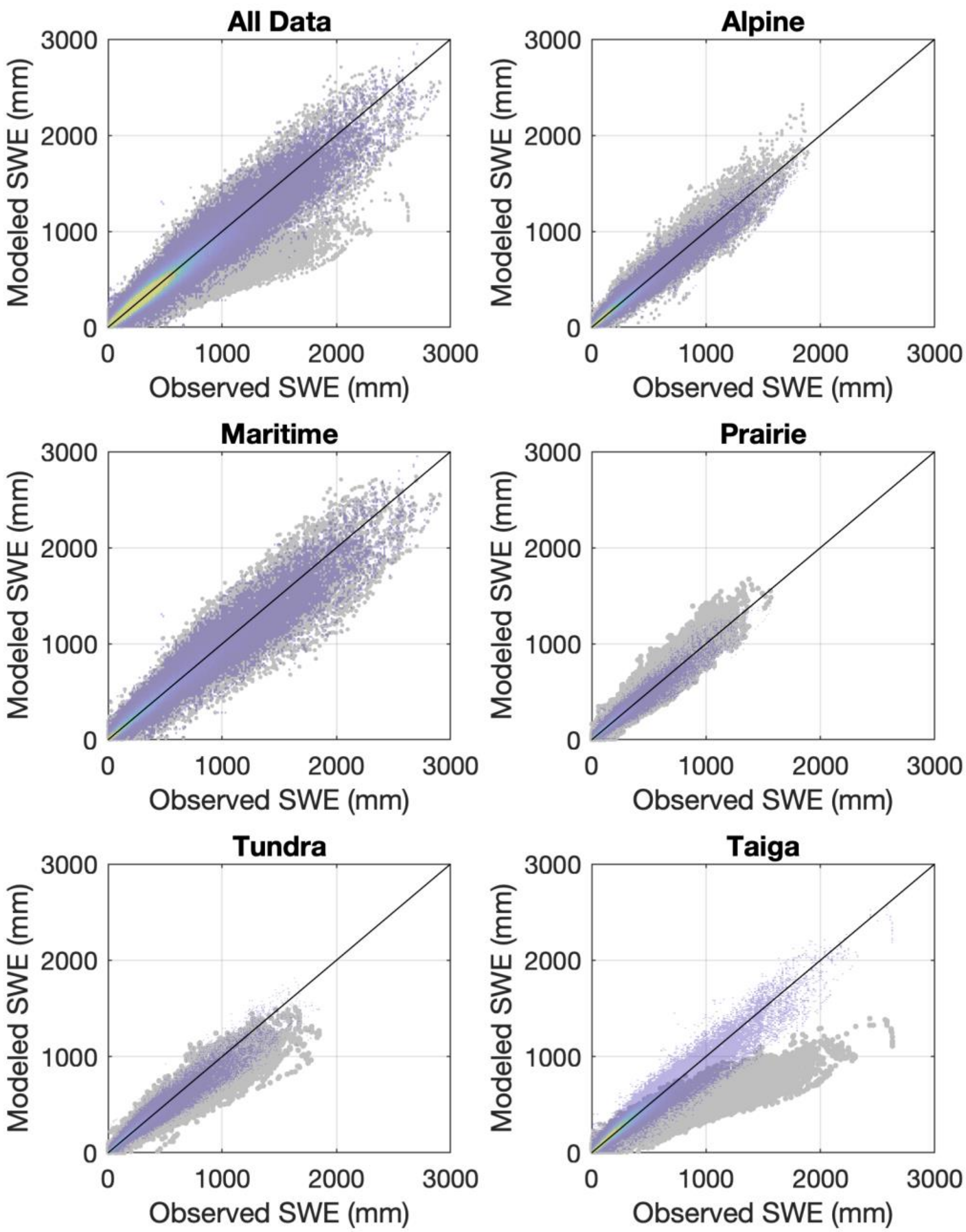


Figure 11: Comparison of the multi-variable, two-equation model of the current study with the model of Sturm et al. (2010). The subpanels show probability density functions of the residuals of the model fits for all of the data binned together, as well as for the data broken out by the snow classes identified by Sturm et al. (1995). The gray lines show the Sturm result and the colored lines show the current result. The vertical lines show the mean error, or the model bias, for both the Sturm and the current result. The models are being applied to the validation data set (50% of the aggregated snow pillow data for CONUS, AK, and BC).

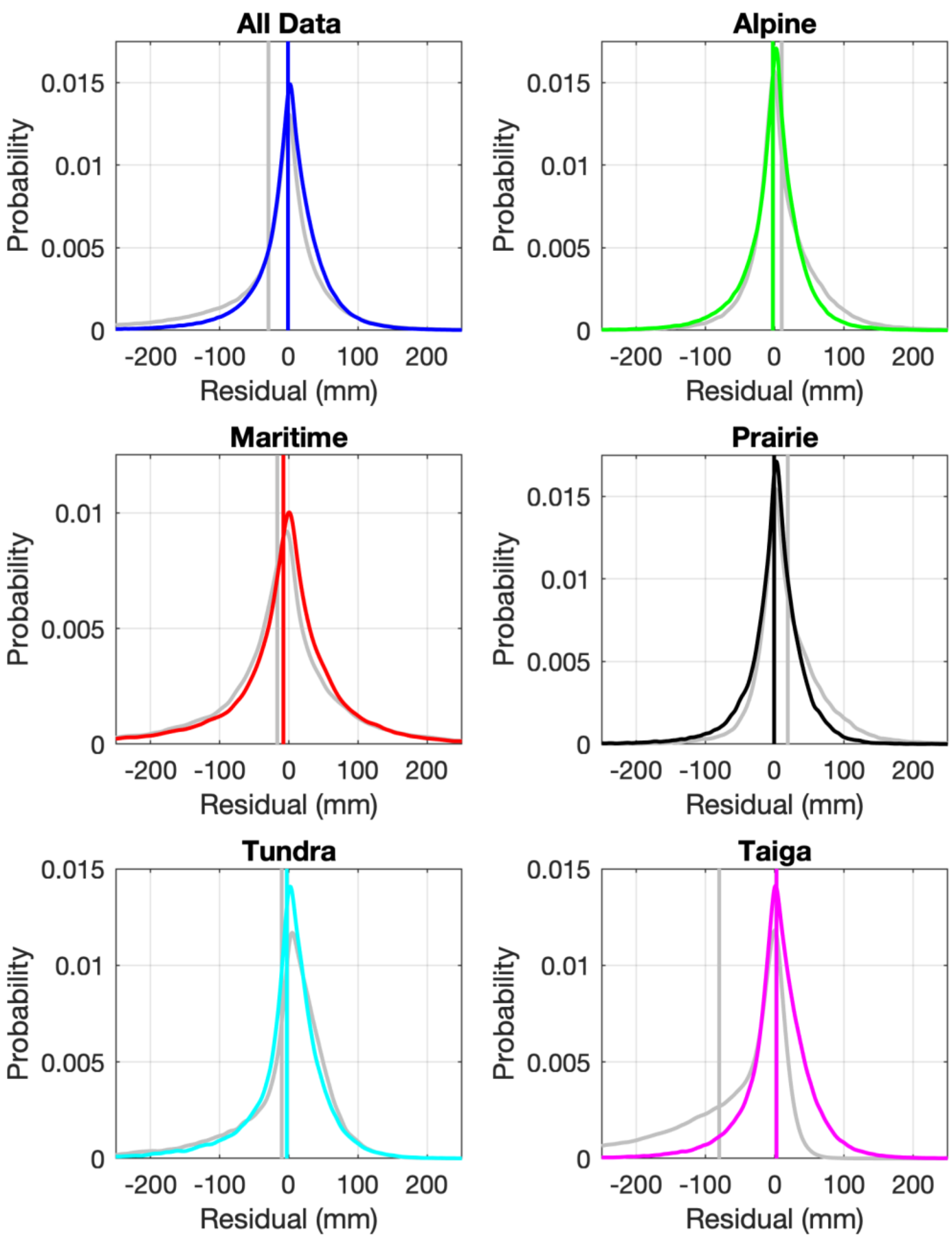
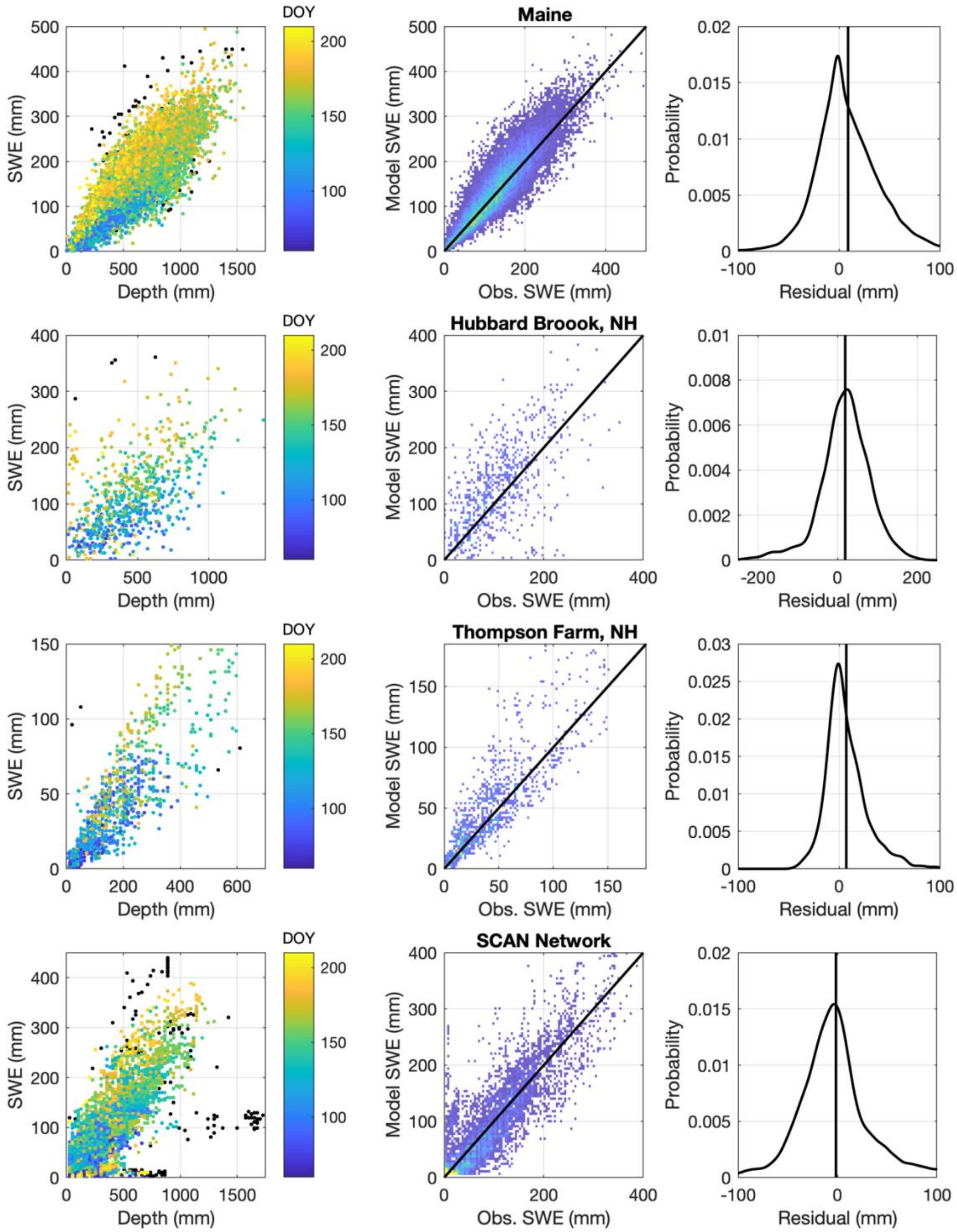
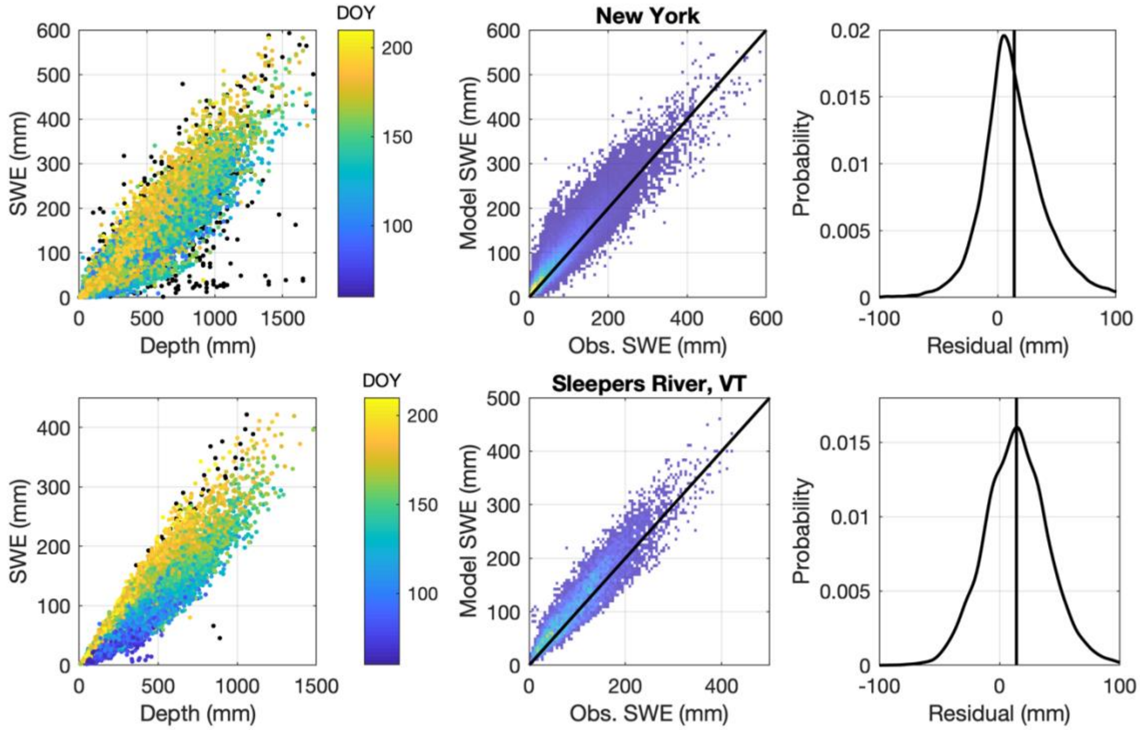


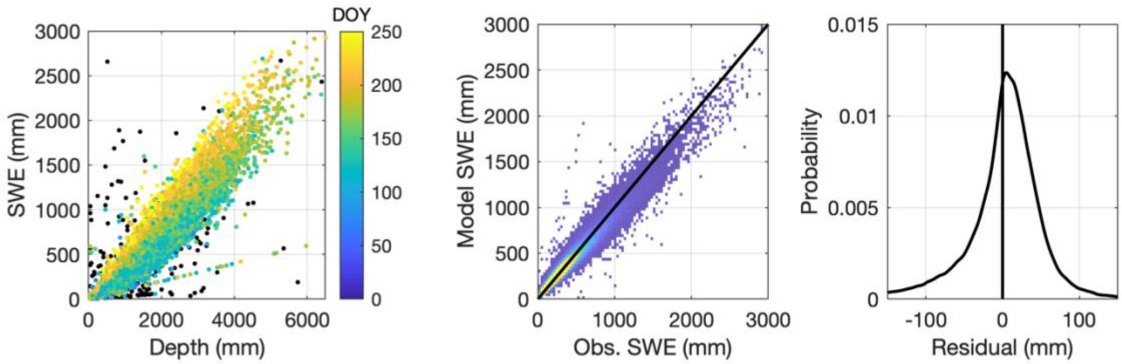
Figure 12: Results from application of the multi-variable, two-equation model to numerous east coast datasets. The left column shows the SWE-*h* data for each dataset. Note that the black symbols are points removed by the outlier detection procedure discussed in section 2.1.1.4. The remaining symbols are colored by *DOY*. The middle panel plots heat maps of the model estimates of SWE against the observations of SWE with the 1:1 line included. Warmer colors indicate higher densities. The right panel shows probability density functions of the model residuals, with the vertical line indicating the mean error, or bias. Individual rows correspond to individual data sets and are labeled.





DOY
New York
SWE (mm)
Depth (mm)
Model SWE (mm)
Obs. SWE (mm)
Probability
Residual (mm)
Sleepers River, VT

Figure 13: Results from application of the multi-variable, two-equation model to the NRCS snow course / aerial marker dataset. The left column shows the SWE-*h* data for each dataset. Note that the black symbols are points removed by the outlier detection procedure discussed in section 2.1.1.5. The remaining symbols are colored by *DOY*. The middle panel plots heat maps of the model estimates of SWE against the observations of SWE with the 1:1 line included. Warmer colors indicate higher densities. The right panel shows probability the density function of the model residuals, with the vertical line indicating the mean error, or bias.



Converting Snow Depth to Snow Water Equivalent Using Climatological Variables

David F. Hill[1], Elizabeth A. Burakowski[2], Ryan L. Crumley[3], Julia Keon[4], J. Michelle Hu[5], Anthony A. Arendt[6], Katreen Wikstrom Jones[7], Gabriel J. Wolken[8]

[1]Civil and Construction Engineering, Oregon State University, OR, USA
[2]Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, NH, USA
[3]Water Resources Graduate Program, Oregon State University, OR, USA
[4]Civil and Construction Engineering, Oregon State University, OR, USA
[5]Civil and Environmental Engineering, University of Washington
[6]Applied Physics Laboratory, University of Washington
[7]Alaska Division of Geological & Geophysical Surveys, Fairbanks, AK, USA
[8]Alaska Division of Geological & Geophysical Surveys, Fairbanks, AK, USA; International Arctic Research Center, University of Alaska Fairbanks, Fairbanks, AK, USA

Correspondence to: David F. Hill (david.hill@oregonstate.edu)

Abstract. We present a simple method that allows snow depth measurements to be converted to snow water equivalent (SWE) estimates. These estimates are useful to individuals interested in water resources, ecological function, and avalanche forecasting. They can also be assimilated into models to help improve predictions of total water volumes over large regions. The conversion of depth to SWE is particularly valuable since snow depth measurements are far more numerous than costlier and more complex SWE measurements. Our model regresses SWE against snow depth (h), day of water year (DOY) and climatological (30-year normal) values for winter (December, January, February) precipitation ($PPTWT$) and the difference (TD) between mean temperature of the warmest month and mean temperature of the coldest month, producing a power-law relationship. Relying on climatological normals rather than weather data for a given year allows our model to be applied at measurement sites lacking a weather station. Separate equations are obtained for the accumulation and the ablation phases of the snowpack. The model is validated against a large database of snow pillow measurements and yields a bias in SWE of less than 2 mm and a root-mean-squared-error (RMSE) in SWE of less than 60 mm. The model is additionally validated against two completely independent sets of data; one from western North America, and one from the northeast United States. Finally, the results are compared with three other models for bulk density that have varying degrees of complexity and that were built in multiple geographic regions. The results show that the model described in this paper has the best performance for the validation data sets.

Formatted: Font: Italic
Deleted: mean annual
Formatted: Font: Italic
Deleted: *MAP*
Formatted: Font: Italic
Deleted: mean February temperature ($\bar{T}_{F_{mean}}$),
Deleted: , which introduces 'day of water year' (*DOY*) as an additional variable
Deleted: 0.5
Deleted: approximately 65
Deleted: When the errors are investigated on a station-by-station basis, the average RMSE is about 5% of the *MAP* at each station.
Deleted: a

1 Introduction

In many parts of the world, snow plays a leading-order role in the hydrological cycle (USACE, 1956; Mote et al., 2018). Accurate information about the spatial and temporal distribution of snow water equivalent (SWE) is useful to many stakeholders (water resource planners, avalanche forecasters, aquatic ecologists, etc.), but can be time consuming and expensive to obtain.

Snow pillows (Beaumont, 1965) are a well-established tool for measuring SWE at fixed locations. Figure 1 provides a conceptual sketch of the variation of SWE with time over a typical water year. A comparatively long accumulation phase is followed by a short ablation phase. While simple in operation, snow pillows are relatively large in size and they need to be installed prior to the onset of the season's snowfall. This limits their ability to be rapidly or opportunistically deployed. Additionally, snow pillow installations tend to require vehicular access, limiting their locations to relatively simple topography. Finally, snow pillow sites are not representative of the lowest or highest elevation bands within mountainous regions (Molotch and Bales, 2005). In the western United States (USA), the Natural Resources Conservation Service (NRCS) operates a large network of Snow Telemetry (SNOTEL) sites, featuring snow pillows. The NRCS also operates the smaller Soil Climate Analysis Network (SCAN) which provides the only, and very limited, snow pillow SWE measurements in the eastern USA.

SWE can also be measured manually, using a snow coring device that measures the weight of a known volume of snow to determine snow density (Church, 1933). These measurements are often one-off measurements, or in the case of 'snow courses' they are repeated weekly or monthly as a transect of measurements at a given location. The simplicity and portability of coring devices expand the range over which measurements can be collected, but it can be challenging to apply these methods to deep snowpacks due to the limited length of standard coring devices. Note that there are numerous different styles of coring devices, including the Adirondack sampler and the Mt. Rose / Federal sampler (Church and Marr, 1937). The NRCS operates a large network of snow course sites (USDA, 2011) in the western United States.

There are a number of issues that affect the accuracy of both snow pillow and snow coring measurements. With coring measurements, if the coring device is not carefully extracted, a portion of the core may fall out of the device. Or, snow may become compressed in the coring device during insertion. These effects have led to varying conclusions, with some studies (e.g., Sturm et al., 2010) showing a low SWE bias and other studies (e.g., Goodison, 1978) showing a high SWE bias. As noted by Johnson et al. (2015) a good rule of thumb is that coring devices are accurate to around ± 10%. Also, studies comparing different styles of snow samplers report statistically different results, suggesting that SWE measurements are sensitive to the design of the specific coring device, such as the presence of holes or slots, the device material, etc. (Beaumont and Work, 1963; Dixon and Boon, 2012). With snow pillows, some studies (e.g., Goodison et al., 1981) note that ice bridging can lead to low biases in measured SWE, with the snow surrounding the pillow partly supporting the snow over the pillow. Other studies (Johnson and Marks, 2004; Dressler et al., 2006; Johnson et al., 2015) note a more complex situation with SWE under-reported at times, but over-reported at other times. Note that when snow pillow data are evaluated, they are most commonly compared to coring measurements at the same location.

All methods of measuring SWE are challenged by the fact that SWE is a depth-integrated property of a snowpack. This is why the snowpack must be weighed, in the case of a snow pillow, or a core must be extracted from the surface to the ground. This measurement complexity makes it difficult to obtain SWE information with the spatial and temporal resolution desired for watershed-scale studies. Other snowpack properties, such as the depth h, are much easier to measure. For example, using a graduated device such as a meterstick or an avalanche probe to measure the depth takes only seconds. Automating depth measurements at a fixed location can easily be done using low-cost ultrasonic devices (Goodison et al., 1984; Ryan et al., 2008). High-spatial-resolution measurements of snowpack depth are commonly made with Light Detection and Ranging (LIDAR). One example of this is the Airborne Snow Observatory program (ASO; Painter et al., 2016). The comparatively high expense of airborne LIDAR surveys typical limits measurements geographically (to a few basins) and temporally (weekly to monthly interval).

Given the relative ease in obtaining depth measurements, it is common to use h as a proxy for SWE. Figure 1 shows a conceptual sketch of the variation of SWE with h over a typical water year. Noting the arrows on the curve, we see that SWE is multi-valued for each h. This is due to the fact that the snowpack increases in density throughout the water year, producing a hysteresis loop in the curve. A large body of literature exists on the topic of how to convert h to SWE. It is beyond the scope of this paper to provide a full review of these 'bulk density equations,' where the density is given by $\rho_b = \mathrm{SWE}/h$. Instead, we refer readers to the useful comparative review by Avanzi et al. (2015). Here, we prefer to discuss a limited number of previous studies that illustrate the spectrum of methodologies and complexities that can be used to determine ρ_b or SWE.

Many studies express ρ_b as an increasing function (often linear) of h. In some cases (e.g., Lundberg et al., 2006) a second equation is added where ρ_b attains a constant value when a threshold h is exceeded. A single linear equation captures the process of densification of the snowpack during the accumulation phase, but performs poorly during the ablation phase, where depths are decreasing but densities continue to increase or approach a constant value. Other approaches choose to parameterize ρ_b in terms of time, rather than h. Pistocchi (2016) provides a single equation while Mizukami and Perica (2008) provide two sets of equations, one set each for early and late season. Each set contains four equations, each of which is applicable to a particular 'cluster' of stations. This clustering was driven by observed densification characteristics and the resulting clusters are relatively spatially discontinuous. Jonas et al. (2009) take the idea of region- (or cluster-) specific equations and extend it further to provide coefficients that depend on time and elevation as well. They use a simple linear equation for ρ_b in terms of h and the slope and intercept of the equation are given as monthly values, with three elevation bins for each month (36 pairs of coefficients). There is an additional contribution to the intercept (or 'offset') which is region-specific (one of 7 regions).

These classifications, whether based on region, elevation, or season, are valuable since they acknowledge that all snow is not equal. McKay and Findlay (1971) discuss the controls that climate and vegetation exert on snow density, and Sturm et al. (2010) address this directly by developing a snow density equation where the coefficients depend upon the 'snow class' (5 classes). Sturm et al. (1995) explain the decision tree, based on temperature, precipitation, and wind speed, that leads to the classification. The temperature metric is the 'cooling degree month' calculated during winter months only. Similarly, only precipitation falling during winter months was used in the classification. Finally, given the challenges in obtaining high quality, high-spatial-resolution wind information, vegetation classification was used as a proxy. Using climatological values (rather than values for a given year), Sturm et al. (1995) were able to develop a global map of snow classification.

There are many other formulations for snow density that increase in complexity and data requirements. Meloysund et al. (2007) express ρ_b in terms of sub-daily measurements of relative humidity, wind characteristics, air pressure, and rainfall, as well as h and estimates of solar exposure ('sun hours'). McCreight and Small (2014) use daily snow depth measurements to develop their regression equation. They demonstrate improved performance over both Sturm et al. (2010) and Jonas et al. (2009). However, a key difference between the McCreight and Small (2014) model and the others listed above is that the former cannot be applied to a single snow depth measurement. Instead, it requires a continuous time series of depth measurements at a fixed location. Further increases in complexity are found in energy-balance snowpack models (SnowModel, Liston and Elder, 2006; VIC, Liang et al., 1994, DHSVM, Wigmosta et al., 1994, others), many of which use multi-layer models to capture the vertical structure of the snowpack. While the particular details vary, these models generally require high temporal-resolution time series of many meteorological variables as input.

Despite the development of multi-layer energy-balance snow models, there is still a demonstrated need for bulk density formulations and for vertically integrated data products like SWE. Pagano et al. (2009) review the advantages and disadvantages of energy-balance models and statistical models and describe how the NRCS uses SWE (from SNOTEL stations) and accumulated precipitation in their statistical models to make daily water supply forecasts. If SWE information is desired at a location that does not have a SNOTEL station, and is not part of a modeling effort, then bulk density equations and depth measurements are an excellent choice.

The present paper seeks to generalize the ideas of Mizukami and Perica (2008), Jonas et al. (2009), and Sturm et al., (2010). Specifically, our goal is to regress physical and environmental variables directly into the equations. In this way, environmental variability is handled in a continuous fashion rather than in a discrete way (model coefficients based on classes). The main motivation for this comes from evidence (e.g., Fig. 3 of Alford, 1967) that density can vary significantly over short distances on a given day. Bulk density equations that rely solely on time completely miss this variability and equations that have coarse (model coefficients varying over either vertical bins or horizontal grids) spatial resolution may not fully capture it either.

Our approach is most similar to Mizukami and Perica (2008), Jonas et al. (2009), and Sturm et al., (2010) in that a minimum of information is needed for the calculations; we intentionally avoid approaches like Meloysund et al. (2007) and McCreight and Small (2014). This is because our interests are in converting *h* measurements to SWE estimates in areas lacking weather instrumentation. The following sections introduce the numerous data sets that were used in this study, outline the regression model adopted, and assess the performance of the model.

2 Methods

2.1 Data

2.1.1 Snow Depth and Snow Water Equivalent

In this section, we list sources of 1970-present snow data utilized for this study (Table 1). With regards to snow coring devices, we refer to them using the terminology preferred in the references describing the datasets.

2.1.1.1 USA NRCS Snow Telemetry and Soil Climate Analysis Networks

SNOTEL (Serreze et al., 1999; Dressler et al., 2006) and SCAN (Schaefer et al. 2007) stations in the contiguous United States (CONUS) and Alaska typically record sub-daily observations of *h*, SWE, and a variety of weather variables (Figure 2a). The periods of record are variable, but the vast majority of stations have a period of record in excess of 30 years. For this study, data from all SNOTEL sites in CONUS and Alaska and northeast USA SCAN sites (Figure 2b) were obtained with the exception of sites whose period of record data were unavailable online. Only stations with both SWE and *h* data were retained.

Deleted: -b

2.1.1.2 Canada (British Columbia) Snow Survey Data

Goodison et al. (1987) note that Canada has no national digital archive of snow observations from the many independent agencies that collect snow data and that snow data are instead managed provincially. The quantity and availability of the data vary considerably among the provinces. The Water Management Branch of the British Columbia (BC) Ministry of the Environment manages a comparatively dense network of Automated Snow Weather Stations (ASWS) that measure SWE, *h*, accumulated precipitation, and other weather variables (Figure 2a). For this study, data from all British Columbia ASWS sites were initially obtained. As with the NRCS stations, only ASWS stations with both SWE and *h* data were retained.

2.1.1.3 USA NRCS Snow Course / Aerial Marker Data

Formatted: Font: Bold

The snow survey program (USDA, 2008) dates to the 1930s and includes a large number of snow course and aerial marker sites (Figure 2c) in western North America. While the measurement frequency is variable, it is most commonly monthly. To generate a dataset for this study, data were extracted using the National Water and Climate

Center Report Generator 2.0. This allows filtering by time period, elevation band, and other elements. All sites with data between 1980-2018 were included (Figure 2c).

2.1.1.4 Northeast USA Data

Deleted: 3

In addition to the data from the SCAN sites, snow data for this project from the northeast US come from two networks and three research sites (Figure 2b). The Maine Cooperative Snow Survey (MCSS, 2018) network includes *h* and SWE data collected by the Maine Geological Survey, the United States Geological Survey, and numerous private contributors and contractors. MCSS snow data are collected using the Standard Federal or Adirondack snow sampling tubes typically on a weekly to bi-weekly schedule throughout the winter and spring, 1951-present. The New York Snow Survey network data were obtained from the National Oceanic and Atmospheric Administration's Northeast Regional Climate Center at Cornell University (NYSS, 2018). Similar to the MCSS, NYSS data are collected using Standard Federal or Adirondack snow sampling tubes on weekly to bi-weekly schedules, 1938-present.

The Sleepers River, Vermont Research Watershed in Danville, Vermont (Shanley and Chalmers, 1999) is a USGS site that includes 15 stations with long-term weekly records of *h* and SWE collected using Adirondack snow tubes. Most of the periods of record are 1981-present, with a few stations going back to the 1960s. The sites include topographically flat openings in conifer stands, old fields with shrub and grass, a hayfield, a pasture, and openings in mixed softwood-hardwood forests. The Hubbard Brook Experiment Forest (Campbell et al., 2010) has collected weekly snow observations at the Station 2 rain gauge site, 1959-present. Measurement protocol collects ten samples 2 m apart along a 20 m transect in a hardwood forest opening about ¼ hectare in size. At each sample location along the transect, *h* and SWE are measured using a Mt. Rose snow tube and the ten samples are averaged for each transect. Finally, the Thompson Farm Research site includes a mixed hardwood forest site and an open pasture site (Burakowski et al. 2013; Burakowski et al. 2015). Daily (from 2011-2018), at each site, a snow core is extracted with an aluminum tube and weighed (tube + snow) using a digital hanging scale. The net weight of the snow is combined with the depth and the tube diameter to determine ρ_b, similar to a Federal or Adirondack sampler.

2.1.1.5 Chugach Mountains (Alaska) Data

Deleted: 4

In the spring of 2018, we conducted three weeks of fieldwork in the Chugach mountains in coastal Alaska, near the city of Valdez (Figure 2d-e). We measured *h* using an avalanche probe at 71 sites along elevational transects during March, April, and May. The elevational transects ranged between 250 and 1100 m (net change along transect) and were accessible by ski and snowshoe travel. At each site, we measured *h* in 8 locations within the surrounding 10 m^2, resulting in a total of 550+ snow depth measurements. These 71 sites were scattered across 8 regions in order to capture spatial gradients that exist in the Chugach mountains as the wetter, more-dense maritime snow near the coast gradually changes to drier, less dense snow on the interior side.

Deleted: c-d

2.1.1.5 Data Pre-Processing

Figure 3 demonstrates that it is not uncommon for automated snow pillow measurements to become noisy or non-physical, at times reporting large depths when there is no SWE reported. This is different from instances when physically plausible, but very low densities might be reported; say in response to early season dry, light snowfalls. It was therefore desirable to apply some objective, uniform procedure to each station's dataset in order to remove clear outlier points, while minimizing the removal of valid data points. We recognize that there is no accepted standardized method for cleaning bivariate SWE-*h* data sets. While Serreze et al. (1999) offer a procedure for SNOTEL data in their appendix, it is relevant only for precipitation and SWE values, not *h*. Given the strong correlation between *h* and SWE, we instead choose to use common outlier detection techniques for bivariate data.

Deleted: depth

The Mahalanobis distance (MD; Maesschalck et al., 2000) quantifies how far a point lies from the mean of a bivariate distribution. The distances are in terms of the number of standard deviations along the respective principal component axes of the distribution. For highly correlated bivariate data, the MD can be qualitatively thought of as a measure of how far a given point deviates from an ellipse enclosing the bulk of the data. One problem is that the MD is based on the statistical properties of the bivariate data (mean, covariance) and these properties can be adversely affected by outlier values. Therefore, it has been suggested (e.g., Leys et al., 2018) that a 'robust' MD (RMD) be calculated. The RMD is essentially the MD calculated based on statistical properties of the distribution unaffected by the outliers. This can be done using the Minimum Covariance Determinant (MCD) method as first introduced by Rousseeuw (1984).

Once RMDs have been calculated for a bivariate data set, there is the question of how large an RMD must be in order for the data point to be considered an outlier. For bivariate normal data, the distribution of the square of the RMD is χ^2 (Gnanadesikan and Kettenring, 1972), with p (the dimension of the dataset) degrees of freedom. So, a rule for identifying outliers could be implemented by selecting as a threshold some arbitrary quantile (say 0.99) of χ_p^2. For the current study, a threshold quantile of 0.999 was determined to be an appropriate compromise in terms of removing obviously outlier points, yet retaining physically plausible results.

A scatter plot of SWE vs. *h* for the SNOTEL dataset from CONUS and AK reveals many non-physical points, mostly when a very large *h* is reported for a very low SWE (Figure 4a). Approximately 0.7% of the original data points were removed in the pre-processing described above, creating a more physically plausible scatter plot (Figure 4b). Note that the outlier detection process was applied to each station individually. The distribution of 'day of year' (*DOY*) values of removed data points was broad, with a mean of 160 and a standard deviation of 65. Note that the *DOY* origin is 1 October. The same procedure was applied to the BC snow pillow, NRCS snow course, and northeast USA data sets as well (not shown). Table 1 summarizes useful information about the numerous data sets described above and indicates the final number of data points retained for each. We acknowledge that our process inevitably removes some valid data points, but, as a small percentage of an already small 0.7% removal rate, we judged this to be acceptable.

Deleted: source

Deleted:

Table 1: Summary of information about the datasets used in this study. Datasets in bold font were used to construct the regression model. The numbers of stations and data points reflect the post-processed data.

Dataset Name	Dataset Type	Number of retained stations	Number and percentage of retained data points	Precision (h / SWE)
NRCS SNOTEL	**Snow pillow (SWE), ultrasonic (*h*)**	**791**	**1,900,000 (99.3%)**	**(0.5 in / 0.1 in)**
NRCS SCAN	Snow pillow (SWE), ultrasonic (*h*)	5	7094 (97.8%)	(0.5 in / 0.1 in)
British Columbia Snow Survey	**Snow pillow (SWE), ultrasonic (*h*)**	**31**	**61,000 (97.5%)**	**(1 cm / 1 mm)**
NRCS Snow Survey	Federal sampler / Aerial marker	1085	116,000 (99.6%)	(0.5 in / 0.1 in) for manual sampler (2 in / n/a) for aerial marker
Maine Geological Survey	Adirondack or Federal sampler (SWE and *h*)	431	28,000 (99.3%)	(0.5 in / 0.5 in)
Hubbard Brook (Station 2), NH	Mount Rose sampler (SWE and *h*)	1	704 (99.4%)	(0.1 in / 0.1 in)
Thompson Farm, NH	Snow core (SWE and *h*)	2	988 (99.4%)	0.5 in / 0.5 in)
Sleepers River, VT	Adirondack sampler	14	7214 (99.4%)	(0.5 in / 0.5 in)
New York Snow Survey	Adirondack or Federal sampler (SWE and *h*)	523	44,614 (98.2%)	(0.5 in / 0.5 in)
Chugach Mountains, AK	Avalanche probe (*h*)	71	71 (100%)	(1 cm)

2.1.2 Climatological Variables

30-year climate normals at 1 km resolution for North America were obtained from the ClimateNA project (Wang et al., 2016). This project provides grids for minimum, maximum, and mean temperature, and total precipitation for a given month. These grids are based on the PRISM normals (Daly et al., 1994) and are available for the periods 1961-1990 and 1981-2010. For this study, the more recent climatology was used. The ClimateNA project also provides a wide array of derived bioclimatic variables, such as precipitation as snow (*PAS*), frost-free-period (*FFP*), mean annual relative humidity (*RH*) and others. Wang et al. (2012) summarize these additional variables and how they are derived. Figure 5 shows gridded maps of winter (sum of December, January, February) precipitation (*PPTWT*) and the temperature difference (*TD*) between the mean temperature of the warmest month and the mean temperature of the coldest month. The latter variable (*TD*) is a measure of continentality.

2.2 Regression Model

In order to demonstrate the varying degrees of influence of explanatory variables, several regression models were constructed. In each case, the model was built by randomly selecting 50% of the paired SWE-*h* measurements from the aggregated CONUS, AK, and BC snow pillow datasets. The model was then validated by applying it to the

Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Not Highlight
Formatted: Font: Not Bold
Formatted: Not Highlight
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Deleted: 800
Deleted: (nominal)
Deleted: CONUS
Deleted: and for the period 1981-2010
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Deleted: PRISM website (Daly et al., 1994). PRISM normals for British Columbia (BC), Canada, were obtained from the ClimateBC project (Wang et al., 2012), also for the 1981-2010 period. Finally, PRISM normals for Alaska (AK) were obtained from the Integrated Resource Management Applications (IRMA) Portal run by the National Park Service. The AK normals are for the 1971-2000 period and have a slightly coarser resolution (approximately 1.5 km). Figure 5 shows gridded maps of mean annual precipitation (*MAP*) and mean February Temperature ($\bar{T}_F$) for these three climate products, plotted together. Other temperature products (max and min temperatures; other months) were obtained as well, but are not shown.

remaining 50% of the dataset and comparing the modeled SWE to the observed SWE for those points. We constructed a second version of the regression models by randomly selecting 50% of the snow pillow stations and using all of the data from those stations. The model was then validated by applying it to the data from the remaining 50% of the stations. These two methods provided identical results, likely due to the very large sample size (N) of our dataset. In all cases, the p values from the linear regression were 0, again due to the large sample size. Additional validation was done with the northeast USA datasets (SCAN snow pillow and various snow coring datasets) and the NRCS snow course dataset, which were completely left out of the model building process.

2.2.1 One-Equation Model

The simplest equation, and one that is supported by the strong correlation seen in the portions of Figure 3 when SWE is present, is one that expresses SWE as a function of *h*. A linear model is attractive in terms of simplicity, but this limits the snowpack to a constant density. An alternative is to express SWE as a power law, i.e.,

(1) $SWE = Ah^{a_1}.$

This equation can be log-transformed into

(2) $log_{10}(SWE) = log_{10}(A) + a_1 log_{10}(h)$

which immediately allows for simple linear regression methods to be applied. With both *h* and SWE expressed in units of mm, the obtained coefficients are $(A, a_1) = (0.146, 1.102)$. Information on the performance of the model will be deferred until the results section.

2.2.2 Two-Equation Model

Recall from Figures 1 and 4 that there is a hysteresis loop in the SWE-*h* relationship. During the accumulation phase, snow densities are relatively low. During the ablation phase, the densities are relatively high. So, the same snowpack depth is associated with two different SWEs, depending upon the time of year. The regression equation given above does not resolve this difference. This can be addressed by developing two separate regression equations, one for the accumulation (*acc*) and one for the ablation (*abl*) phase. This approach takes the form

(3) $SWE_{acc} = Ah^{a_1}; \quad DOY < DOY^*$

(4) $SWE_{abl} = Bh^{b_1}; \quad DOY \geq DOY^*$

where *DOY* is the number of days from the start of the water-year, and DOY^* is the critical or dividing day-of-water-year separating the two phases. Put another way, DOY^* is the day of peak SWE. Interannual variability results in a range of DOY^* for a given site. Additionally, some sites, particularly the SCAN sites in the northeast USA,

Deleted: (October 1 is the origin),

demonstrate multi-peak SWE profiles in some years. To reduce model complexity, however, we investigated the use of a simple climatological (long term average) value of DOY^* at each site. For each snow pillow station, the average DOY^* was computed over the period of record of that station. Analysis of all of the stations revealed that this average DOY^* was relatively well correlated with the climatological mean April maximum temperature (the average of the daily maximums recorded in April; $R^2 = 0.7$). However, subsequent regression analysis demonstrated that the SWE estimates were relatively insensitive to DOY^* and the best results were actually obtained when DOY^* was uniformly set to 180 for all stations. Again, with both SWE and h in units of mm, the regression coefficients turn out to be $(A, a_1) = (0.150, 1.082)$ and $(B, b_1) = (0.239, 1.069)$.

As these two equations are discontinuous at DOY^*, they are blended smoothly together to produce the final two-equation model

(5) $$SWE = SWE_{acc}\frac{1}{2}(1 - tanh[0.01\{DOY - DOY^*\}]) + SWE_{abl}\frac{1}{2}(1 + tanh[0.01\{DOY - DOY^*\}])$$

The coefficient 0.01 in the tanh function controls the width of the blending window and was selected to minimize the root mean square error of the model estimates.

2.2.3 Two-Equation Model with Climate Parameters

A final model was constructed by incorporating climatological variables. Again, the emphasis in this study is on methods that can be implemented at locations lacking the time series of weather variables that might be available at a weather or SNOTEL station. Climatological normals are unable to account for interannual variability, but they do preserve the high spatial gradients in climate that can lead to spatial gradients in snowpack characteristics. Stepwise linear regression was used to determine which variables to include in the regression. The initial list of potential variables included was

(6) $$SWE = f(h, z, PPTWT, PAS, TWT, TD, DOY, RH)$$

where z is the elevation (m), $PPTWT$ is the winter (sum of December, January, February) precipitation (mm), PAS is mean annual precipitation as snow (mm), TWT is the winter (December, January, February) mean temperature ($°C$), TD is the difference between the mean temperature of the warmest month and the mean temperature of the coldest month ($°C$), DOY is the day of water year, and RH is the relative humidity (%). In the stepwise regression, explanatory variables were accepted only if they improved the adjusted R^2 value by 0.001. The result of the regression yielded

(7) $$SWE_{acc} = Ah^{a_1}PPTWT^{a_2}TD^{a_3}DOY^{a_4}; \quad DOY < DOY^*$$

Deleted: is

Deleted: , MAP,

Deleted: $\bar{T}_{Jmin}, \bar{T}_{Jmean}, \bar{T}_{Jmax}, \bar{T}_{Fmin}, \bar{T}_{Fmean}, \bar{T}_{Fmax}, \bar{T}_{Mmin}, \bar{T}_{Mmean}$

Deleted: , *MAP* is the mean annual precipitation (mm)

Formatted: Font: Italic

Formatted: Font: Italic

Deleted: and the

Deleted: s

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Deleted:

Deleted: represent the mean of minimum, mean, and maximum daily values for the months January through April (J, F, M, A). For example, $\bar{T}_{J_{min}}$ is the climatological normal of the average of the daily minimum temperatures observed in January.

(8) $$SWE_{abl} = Bh^{b_1}PPTWT^{b_2}TD^{b_3}DOY^{b_4}; \quad DOY \geq DOY^*$$

or, in log-transformed format,

(9) $$log_{10}(SWE_{acc}) = log_{10}(A) + a_1 log_{10}(h) + a_2 log_{10}(PPTWT) + a_3 log_{10}(TD) + a_4 log_{10}(DOY); \quad DOY < DOY^*$$

(10) $$log_{10}(SWE_{abl}) = log_{10}(B) + b_1 log_{10}(h) + b_2 log_{10}(PPTWT) + b_3 log_{10}(TD) + b_4 log_{10}(DOY); \quad DOY \geq DOY^*$$

indicating that only snow depth, winter precipitation, temperature difference, and day of water year were relevant. Manual tests of model construction with other variables included confirmed that Eqns. (7-8) yielded the best results. These two SWE estimates for the individual (*acc* and *abl*) phases of the snowpack were then blended with Eqn. (5) to produce a single equation for SWE spanning the entire water year. The obtained regression coefficients were $(A, a_1, a_2, a_3, a_4) = (0.0533, 0.9480, 0.1701, -0.1314, 0.2922)$ and $(B, b_1, b_2, b_3, b_4) = (0.0481, 1.0395, 0.1699, -0.0461, 0.1804)$. The physical interpretation of these coefficients is straightforward. For example, both a_2 and b_2 are greater than zero. So, for two locations with equal *h*, *DOY*, and *TD*, the location with greater *PPTWT* will have a greater SWE and therefore density. These locations are typically maritime climates with wetter, denser snow. In contrast, both a_3 and b_3 are less than zero. Therefore, for two locations with equal *h*, *DOY*, and *PPTWT*, the location with greater TD (a more continental climate) will have a lower density, which is again an expected result. These trends are similar in concept to Sturm et al. (2010), whose discrete snow classes (based on climate classes) indicate which snow will densify more rapidly.

3 Results

A comparison of the three regression models (one-equation model, Eq. (2); two-equation model, Eqs. (3-5); multi-variable two-equation model, Eqs. (5, 7-8)) is provided in Figure 6. The left column shows scatter plots of modeled SWE to observed SWE for the validation data set with the 1:1 line shown in black. The right column shows distributions of the model residuals. The vertical lines in the right column show the mean error, or model bias. Visually, it is clear that the one-equation model performs relatively poorly with a large negative bias. This large negative bias is partially overcome by the two-equation model (middle row, Figure 6). The cloud of points is closer to the 1:1 line and the vertical black line indicating the mean error is closer to zero. In the final row of Figure 6, we see that the multi-variable two-equation model yields the best result by far. The residuals are now evenly distributed with a small bias. Several metrics of performance for the three models, including R^2 (Pearson coefficient), bias, and root-mean-square-error (RMSE), are provided in Table 2. Figure 7 shows the distribution of model residuals for the multi-variable two-equation model as a function of DOY.

Deleted: mean annual precipitation

Deleted: and mean February temperature

Deleted: In the above equations, note that an offset is added to the temperature in order to avoid taking the log of a negative number.

Deleted: are

Deleted: 128

Deleted: 1.070

Deleted: 0.132

Deleted: 0.506

Deleted: 271

Deleted: 8

Deleted: 201

Deleted: 310

Deleted: If a_1 and b_1 were equal to unity, then the density, given by (SWE/*h*), would be a constant at a given location. Since they are greater than unity, they capture the effect that snow density increases as depth increases. Turning to the coefficients on the climate variables,

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Deleted: equal depth, equal temperature characteristics, but different precipitation characteristics, the regression model predicts that the wetter location (larger *MAP*) will have a greater density

Deleted: Finally

Deleted: , regarding temperature

Deleted: greater

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Deleted: equal depth, equal precipitation characteristics, but different temperature characteristics, the regression model predicts that the warmer location (larger $\bar{T}_{F_{mean}}$) will have a greater density.

Deleted: histograms

Deleted: negligible

Table 2: Summary of performance metrics for the three regression models presented in Section 2.2.

Model	R^2	Bias (mm)	RMSE (mm)
One-equation	0.946	-19.5	102
Two-equation	0.962	-5.1	81
Multi-variable two-equation	0.978	-1.2	59

It is useful to also consider the model errors in a non-dimensional way. Therefore, an RMSE was computed at each station location and normalized by the winter precipitation (*PPTWT*) at that location. Figure 8 shows the probability density function of these normalized errors. The average RMSE is approximately 15% of *PPTWT* with most values falling into the range of 5-30%. The spatial distribution of these normalized errors is shown in Figure 9. For the SNOTEL stations, it appears there is a slight regional trend, in terms of stations in continental climates (Rockies) having larger relative errors than stations in maritime climates (Cascades). The British Columbia stations also show higher relative errors.

3.1 Results for Snow Classes

A key objective of this study is to regress climatological information in a continuous rather than a discrete way. The work by Sturm et al. (2010) therefore provides a valuable point of comparison. In that study, the authors developed the following equation for density ρ_b

(11) $$\rho_b = (\rho_{max} - \rho_0)\left[1 - e^{(-k_1 h - k_2 DOY)}\right] + \rho_0$$

where ρ_0 is the initial density, ρ_{max} is the maximum or 'final' density (end of water year), k_1 and k_2 are coefficients, and *DOY* in this case begins on January 1. This means that their *DOY* for October 1 is -92. The coefficients vary with snow class and the values determined by Sturm et al. (2010) are shown in Table 3.

Table 3: Model parameters by snow class for Sturm et al. (2010).

Snow Class	ρ_{max}	ρ_0	k_1	k_2
Alpine	0.5975	0.2237	0.0012	0.0038
Maritime	0.5979	0.2578	0.0010	0.0038
Prairie	0.5941	0.2332	0.0016	0.0031
Tundra	0.3630	0.2425	0.0029	0.0049
Taiga	0.2170	0.2170	0.0000	0.0000

To make a comparison, the snow class for each SNOTEL and British Columbia snow survey (Rows 1 and 3 of Table 1) site was determined using a 1-km snow class grid (Sturm et al., 2010). The aggregated dataset from these stations was made up of 27% Alpine, 14% Maritime, 10% Prairie, 11% Tundra, and 38% Taiga data points. Equation (11) was then used to estimate snow density (and then SWE) for every point in the validation dataset described in Section 2.2. Figure 10 compares the SWE estimates from the Sturm model and from the current multi-variable, two-equation model (Equations 5, 7-8). The upper left panel of Figure 10 shows all of the data, and the remaining panels show the results for each snow class. In all cases, the current model provides better estimates (narrow cloud of points; closer

Deleted: 2
Deleted: 0.5
Deleted: 67
Deleted: mean annual maximum SWE (SWE_{max})
Formatted: Font: Italic
Deleted: 1
Deleted: SWE_{max},
Deleted: 25
Deleted: northern
Deleted: smaller
Formatted: Font: Italic
Formatted: Font: Italic
Deleted: present

to the 1:1 line). Plots of the residuals by snow class are provided in Figure 11, giving an indication of the bias of each model for each snow class. Summaries of the model performance, broken out by snow class, are given in Table 4. The current model has smaller biases and RMSEs for each snow class.

Table 4: Comparison of model performance by Sturm et al. (2010) and the current study.

Model	Sturm et al. (2010)			Multi-variable two-equation model		
Snow Class	R^2	Bias (mm)	RMSE (mm)	R^2	Bias (mm)	RMSE (mm)
All Data	0.928	-29.2	111	0.978	-1.2	59
Alpine	0.973	10.1	55	0.978	-2.7	48
Maritime	0.968	-16.8	109	0.975	-7.8	95
Prairie	0.967	18.7	56	0.971	-0.7	45
Tundra	0.956	-10.5	82	0.974	-2.9	59
Taiga	0.943	-80.0	151	0.978	2.6	54

3.2 Comparison to Pistocchi (2016)

In order to provide an additional comparison, the simple model of Pistocchi (2016) was also applied to the validation dataset. His model calculates the bulk density as

(12) $$\rho_b = \rho_0 + K(DOY + 61),$$

where ρ_0 has a value of 200 kg m^{-3} and K has a value of 1 kg m^{-3}. The DOY for this model has its origin at November 1. Application of this model to the validation dataset yields a bias of 55 mm and an RMSE of 94 mm. These results are comparable to the Sturm et al. (2010) model, with a larger bias but smaller RMSE.

3.3 Comparison to Jonas et al. (2009)

A final point of comparison can be provided by the model of Jonas et al. (2009). The full version of that model contains region-specific offset parameters that are not relevant to North America, so the following partial version of the model is used (their Eq. 4):

(13) $$\rho_b = ah + b,$$

where the parameters (a, b) vary with elevation and month, as given by Table 5. Note that coefficients are not given for every month. Application of the Jonas et al. (2009) model to the snow pillow dataset yields a bias of -5 mm and an RMSE of 69 mm. These results are not directly comparable to those of the current model (Table 2, row 3) since the Jonas et al. (2009) model is unable to compute results for several months of the year. To make a direct comparison to the current model, it is necessary to first remove those data points (about 5%). When this is done, the current model yields a bias of -0.3 mm and an RMSE of 59 mm.

Table 5: Model coefficients (a, b) for the Jonas et al. (2009) model.

Deleted: present
Deleted: 2
Deleted: -0.5
Deleted: 67
Deleted: 1
Deleted: 0.3
Deleted: 55
Deleted: 0
Deleted: 4.5
Deleted: 105
Deleted: 65
Deleted: 2
Deleted: 51
Deleted: 69
Deleted: 6.1
Deleted: 67
Formatted Table
Deleted: 1
Deleted: 4
Deleted: 62
Deleted: one
Formatted: Font: Italic
Formatted: Font: 12 pt, Font color: Auto
Formatted: Normal
Formatted: Font: 10 pt, Bold
Formatted: Not Highlight
Formatted: Normal, Line spacing: single

Month	$z > 2000\ m$	$2000\ m > z > 1400\ m$	$1400\ m > z$
January	$(206, 52)$	$(208, 47)$	$(235, 31)$
February	$(217, 46)$	$(218, 52)$	$(279, 9)$
March	$(272, 26)$	$(281, 31)$	$(333, 3)$
April	$(331, 9)$	$(354, 15)$	$(347, 25)$
May	$(378, 21)$	$(409, 29)$	$(413, 19)$
June	$(452, 8)$	n/a	n/a
July	$(470, 15)$	n/a	n/a
August	n/a	n/a	n/a
September	n/a	n/a	n/a
October	n/a	n/a	n/a
November	$(206, 47)$	$(183, 35)$	$(149, 37)$
December	$(203, 52)$	$(190, 47)$	$(201, 26)$

3.4 Results for Northeast USA

The regression equations in this study were developed using a large collection of snow pillow sites in CONUS, AK, and BC. The snow pillow sites are limited to locations west of approximately W 105° (Figure 2a). By design, the data sets from the northeastern USA (Section 2.1.1.3) were left as an entirely independent validation set. These northeastern sites are geographically distant from the training data sets, subject to a very different climate, largely use different methods (snow coring, with the exception of the SCAN network) and are generally at much lower elevations than the western sites, providing an interesting opportunity to test how robust the current model is.

Figure 12 graphically summarizes the datasets and the performance of the multi-variable two-equation model of the current study. The RMSE values are comparable to those found for the western stations, but, given the comparatively thinner snowpacks in the northeast, represent a larger relative error (Table 5). The bias of the model is consistently positive, in contrast to the western stations where the bias was negligible. Note that Table 5 also includes results from the application of the other three models discussed. Sturm et al. (2010) cannot be applied to several of the datasets since their available 1 km snowclass dataset cuts off at -71.6° longitude. The current model and the Jonas et al. (2009) model perform better than the other two models, with the current model generally outperforming the Jonas et al. (2009) model. The two datasets where the Jonas et al. (2009) model has a slightly better performance are the two smallest datasets (less than 1000 measurements; see Table 1).

Table 5: Performance metrics for various models applied to the northeastern USA datasets. Bold font is used to highlight the model with the best performance for each dataset.

	Multi-variable, two-equation model		Sturm et al. (2010)		Jonas et al. (2009)		Pistocchi (2015)	
Dataset Name	Bias (mm)	RMSE (mm)	Bias (mm)	RMSE (mm)	Bias (mm)	RMSE (mm)	Bias (mm)	RMSE (mm)

Formatted Table

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Deleted: ¶

Formatted: Font color: Auto

Formatted: Line spacing: 1.5 lines

Deleted: 3

Deleted: SNOTEL

Deleted: are

Deleted: ,

Deleted: present

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Deleted: the multi-variable two-equation model

Deleted: various

Maine Geological Survey, ME	**13.1**	**34.0**	n/a	n/a	25.1	46.0	59.2	77.1
Hubbard Brook (Station 2), NH	21.8	66.6	34.2	76.9	**19.4**	**65.4**	52.0	90.8
Thompson Farm, NH	7.1	20.2	n/a	n/a	**5.6**	**19.9**	20.4	32.3
NRCS SCAN	**-1.2**	**39.2**	8.4	45.0	-2.8	40.6	23.4	56.9
Sleepers River, VT	**14.4**	**28.2**	36.5	48.9	20.4	33.5	55.8	67.1
New York Snow Survey	**14.8**	**31.2**	21.0	49.3	16.3	33.0	41.3	56.1

3.5 Results for NRCS Snow Course / Aerial Marker Data

The NRCS snow course and aerial marker data were also left out of the model building process so they provide an additional and completely independent comparison of the various models considered. Recall that these data come from snow course (coring measurements) and aerial surveys, which are different measurement methods than the snow pillows which provided the data for construction of the current regression model. Table 6 shows the results and demonstrates that the current model has the best performance.

Table 6: Performance metrics for various models applied to the NRCS snow course and aerial marker dataset. Bold font is used to highlight the model with the best performance.

	Multi-variable, two-equation model		Sturm et al. (2010)		Jonas et al. (2009)		Pistocchi (2015)	
Dataset Name	Bias (mm)	RMSE (mm)	Bias (mm)	RMSE (mm)	Bias (mm)	RMSE (mm)	Bias (mm)	RMSE (mm)
NRCS Snow Course / Aerial Marker	**0**	**59**	-24	123	24	72	71	99

4 Discussion

The results presented in this study show that the regression equation described by equations (5, 7-8) is an improvement (lower bias and RMSE) over other widely used bulk density equations. The key advantage is that the current method regresses in relevant parameters directly, rather than using discrete bins (for snow class, elevation, month of year, etc.), each with its own set of model coefficients. The comparison (Figs. 10-11; Table 4) to the model of Sturm et al. (2010) reveals a peculiar behavior of that model for the Taiga snow class, with a large negative bias in the Sturm estimates. Inspection of the coefficients provided for that class (Table 3) shows that the model simply predicts that $\rho_b = \rho_{max} = 0.217$ for all conditions.

When our multi-variable two-equation model, developed solely from western North American data, is applied to northeast USA locations, it produces SWE estimates with smaller RSME values and larger biases than the western stations. When comparing the SWE-h scatter plots of the SNOTEL data (Figure 4b) to those of the east coast data sets (left column; Figure 12), it is clear that the northeast data generally have more scatter. This is confirmed by computing the correlation coefficients between SWE and h for each dataset. It is unclear if this disparity in correlation is related to measurement methodology or is instead a 'signal to noise' issue. Comparing Figures 4 and 12 shows the considerable difference in snowpack depth between the western and northeastern data sets. When the western dataset is filtered to include only measurement pairs where $h < 1.5$ m, the correlation coefficient is reduced to a value consistent with the northeast datasets. This suggests that the performance of the current (or other) regression model is not as good at shallow snowpack depths. This is also suggested upon examination of the time

Deleted: 8.9
Formatted: Font: Bold
Deleted: 3.3
Deleted: 18.9
Deleted: 64.2
Formatted: Font: Bold
Deleted: 21.6
Formatted: Font: Bold
Formatted: Font: Bold
Deleted: 1.8
Deleted: 8.7
Deleted: 14
Formatted: Font: Bold
Deleted: .0
Deleted: 29.7
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Font: Bold
Deleted: 3
Formatted: Font: (Default) Times New Roman, Not Italic, Font color: Text 1
Formatted Table
Deleted: ¶
Deleted: present
Deleted: physical
Deleted: curves

series of observed $\rho_b = SWE/h$ for a given season at a snow pillow site. Very early in the season, when the depths are small, the density curve has a lot of variability. Later in the season, when depths are greater, the density curve becomes much smoother. Very late in the season, when depths are low again, the density curve becomes highly variable again.

Measurement precision and accuracy affect the construction and use of a regression model. Upon inspection of the snow pillow data, it was observed that the precision of the depth measurements was approximately 25 mm and that of the SWE measurements was approximately 2.5 mm. To test the sensitivity of the model coefficients to the measurement precision, the depth values in the training dataset were randomly perturbed by +/- 25 mm and the SWE values were randomly perturbed by +/- 2.5 mm and the regression coefficients were recomputed. This process was repeated numerous times and the mean values of the perturbed coefficients were obtained. These adjusted coefficients were then used to recompute the SWE values for the validation data set and the bias and RMSE were found to be -10.5 mm and 72.7 mm. This represents a roughly 10% increase in RMSE, but a considerable increase in bias magnitude (see Table 4 for the original values). This sensitivity of the regression analysis to measurement precision underscores the need to have high-precision measurements for the training data set. Regarding accuracy, random and systematic errors in the paired SWE - h data used to construct the regression model will lead to uncertainties in SWE values predicted by the model. As noted in the introduction, snow pillow errors in SWE estimates do not follow a simple pattern. Additionally, they are complicated by the fact that the errors are often computed by comparing snow pillow data to coring data, which itself is subject to error. Lacking quantitative information on the distribution of snow pillow errors, we are unable to quantify the uncertainty in the SWE estimates.

Another important consideration has to do with the uncertainty of depth measurements that the model is applied to. For context, one application of this study is to crowd-sourced, opportunistic snow depth measurements from programs like the Community Snow Observations (CSO; Hill et al., 2018) project. In the CSO program, backcountry recreational users submit depth measurements, typically taken with an avalanche probe, using a smartphone in the field. The measurements are then converted to SWE estimates which are assimilated into snowpack models. These depth measurements are 'any time, any place' in contrast to repeated measurements from the same location, like snow pillows or snow courses. Most avalanche probes have cm-scale graduated markings, so measurement precision is not a major issue. A larger problem is the considerable variability in snowpack depth that can exist over short (meter scale) distances. The variability of the Chugach avalanche probe measurements was assessed by taking the standard deviation of 8 h measurements per site. The average of this standard deviation over the sites was 22 cm and the average coefficient of variation (standard deviation normalized by the mean) over the sites was 15%. This variability is a function of the surface roughness of the underlying terrain, and also a function of wind redistribution of snow. Propagating this uncertainty through the regression equations yields a slightly higher (16%) uncertainty in the SWE estimates. CSO participants can do three things to ensure that their recorded depth measurements are as representative as possible. First, avoid measurements in areas of significant wind scour or deposition. Second, avoid measurements in terrain likely to have significant surface roughness (rocks, fallen logs, etc.). Third, take several measurements and average them.

Expansion of CSO measurements in areas lacking SWE measurements can increase our understanding of the extreme spatial variability in snow distribution and the inherent uncertainties associated with modeling SWE in these regions. It could also prove useful for estimating watershed-scale SWE in regions like the northeastern USA, which is currently limited to five automated SCAN sites with historical SWE measurements for only the past two decades. Additionally, historical snow depth measurements are more widely available in the Global Historical Climatology Network (GHCN-Daily; Menne et al. 2012), with several records extending back to the late 1800s. While many of the GHCN stations are confined to lower elevations with shallower snow depths, the broader network of quality-controlled snow depth data paired with daily GHCN temperature and precipitation measurements could potentially be used to reconstruct SWE in the eastern US given additional model development and refinement.

5 Conclusions

We have developed a new, easy to use method for converting snow depth measurements to snow water equivalent estimates. The key difference between our approach and previous approaches is that we directly regress in climatological variables in a continuous fashion, rather than a discrete one. Given the abundance of freely available climatological norms, a depth measurement tagged with coordinates (latitude and longitude) and a time stamp is easily and immediately converted into SWE.

We developed this model with data from paired SWE-*h* measurements from the western United States and British Columbia. The model was tested against entirely independent data (primarily snow course; some snow pillow) from the northeastern United States and was found to perform well, albeit with larger biases and root-mean-squared-errors. The model was tested against other well-known regression equations and was found to perform better. The model was also tested against a large dataset of independent snow course and aerial marker measurements from western North America. For this second independent test, the current model outperformed the other models considered.

This model is not a replacement for more sophisticated snow models that evolve the snowpack based on high frequency (e.g., daily or sub-daily) weather data inputs. The intended purpose of this model is to constrain SWE estimates in circumstances where snow depth is known, but weather variables are not, a common issue in sparsely instrumented areas in North America.

6 Acknowledgements

Support for this project was provided by NASA (NNX17AG67A). R. Crumley acknowledges support from the CUAHSI Pathfinder Fellowship. E. Burakowski acknowledges support from NSF (MSB-ECA #1802726). We thank M. Sturm, A. Winstral and a third anonymous referee for their careful and thoughtful reviews of this manuscript.

7 Data Access

Numerous online datasets were used for this project and were obtained from the following locations:

1. NRCS Snow Telemetry: https://www.wcc.nrcs.usda.gov/snow/SNOTEL-wedata.html
2. NRCS Soil Climate Analysis Network: https://www.wcc.nrcs.usda.gov/scan/
3. British Columbia Automated Snow Weather Stations: https://www2.gov.bc.ca/gov/content/environment/air-land-water/water/water-science-data/water-data-tools/snow-survey-data/automated-snow-weather-station-data
4. Maine Cooperative Snow Survey: https://mgs-maine.opendata.arcgis.com/datasets/maine-snow-survey-data
5. New York Snow Survey: http://www.nrcc.cornell.edu/regional/snowsurvey/snowsurvey.html
6. Sleepers River Research Watershed. Snow data not available online; request data from contact at: https://nh.water.usgs.gov/project/sleepers/index.htm
7. Hubbard Brook Experimental Forest: https://hubbardbrook.org/d/hubbard-brook-data-catalog
8. Climatological Data: https://adaptwest.databasin.org/pages/adaptwest-climatena
9. NRCS Snow Course / Aerial Marker Data: https://wcc.sc.egov.usda.gov/reportGenerator/

A Matlab function for calculating SWE based on the results is this paper has been made publicly available at Github (https://github.com/communitysnowobs/snowdensity).

Formatted: Default Paragraph Font, Font: (Default) +Body (Calibri), 12 pt

Formatted: Default Paragraph Font, Font: (Default) +Body (Calibri), 12 pt

Formatted: Default Paragraph Font, Font: (Default) +Body (Calibri), 12 pt

Formatted: No underline, Font color: Auto

Formatted: Default Paragraph Font, Font: (Default) +Body (Calibri), 12 pt

Formatted: No underline, Font color: Auto

Formatted: Default Paragraph Font, Font: (Default) +Body (Calibri), 12 pt

Formatted: Default Paragraph Font, Font: (Default) +Body (Calibri), 12 pt

Formatted: No underline, Font color: Auto

Formatted: Default Paragraph Font, Font: 12 pt

Deleted: ¶

Formatted: Font: (Default) Times New Roman, No underline, Font color: Auto

Deleted: CONUS PRISM Data: http://www.prism.oregonstate.edu/

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Deleted: ¶

Formatted: Font: 10 pt

Formatted: Default Paragraph Font, Font: 12 pt

Formatted: Font:

Formatted: No underline, Font color: Auto

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: 10 pt

Formatted: List Paragraph, Line spacing: single

Deleted: British Columbia PRISM Data: http://climatebcdata.climatewna.com/¶
Alaska PRISM Data: https://irma.nps.gov/Portal/¶

Deleted: has been

Deleted: (https://github.com/communitysnowobs/snowdensity).

Formatted: Default Paragraph Font, Font: 12 pt

References

Alford, D.: Density variations in alpine snow, J. Glaciol., 6(46), 495-503, https://doi.org/10.3189/S0022143000019717, 1967.

Avanzi, F., De Michele, C., and Ghezzi, A.: On the performances of empirical regressions for the estimation of bulk snow density, Geogr. Fis. Dinam. Quat., 38, 105-112, doi:10.4461/GFDQ.2015.38.10, 2015.

Beaumont, R.: Mt. Hood pressure pillow snow gage, J. Appl. Meteorol., 4, 626-631, https://doi.org/10.1175/1520-0450(1965)004<0626:MHPPSG>2.0.CO;2 1965.

Beaumont, R., and Work, R.: Snow sampling results from three samplers, Hydrol. Sci. J., 8(4), 74-78, https://doi.org/10.1080/02626666309493359, 1963.

Burakowski, E.A., Ollinger, S., Lepine, L., Schaaf, C.B., Wang, Z., Dibb, J.E., Hollinger, D.Y., Kim, J.-H., Erb, A., and Martin, M.E.: Spatial scaling of reflectance and surface albedo over a mixed-use, temperate forest landscape during snow-covered periods, Remote Sens. Environ., 158, 465-477, https://doi.org/10.1016/j.rse.2014.11.023, 2015.

Burakowski, E.A., Wake, C.P., Stampone, M., and Dibb, J.: Putting the Capital ‘A’ in CoCoRAHS: An Experimental Program to Measure Albedo using the Community Collaborative Rain Hail and Snow (CoCoRaHS) Network, Hydrol. Process., 27(21), 3024-3034, https://doi.org/10.1002/hyp.9825, 2013.

Campbell, J., Ollinger, S., Flerchinger, G., Wicklein, H., Hayhoe, K., and Bailey, A.: Past and projected future changes in snowpack and soil frost at the Hubbard Brook Experimental Forest, New Hampshire, USA, Hydrol. Process., 24, 2465-2480, https://doi.org/10.1002/hyp.7666, 2010.

Church, J.E.: Snow surveying: its principles and possibilities, Geogr. Rev., 23(4), 529-563, DOI: 10.2307/209242, 1933.

Church, J.E., and Marr, J.C.: Further improvement of snow-survey apparatus, Transactions of the American Geophysical Union, 18(2), 607-617, 10.1029/TR018i002p00607, 1937.

Daly, C., Neilson, R., and Phillips, D.: A statistical-topographic model for mapping climatological precipitation over mountainous terrain, J. Appl. Meteorol., 33, 140-158, https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2, 1994.

Formatted: Default Paragraph Font, Font: 12 pt

Deleted:
Deeb, E., Marshall, H.-P., Forster, R., Jones, C., Hiemstra, C., and Siqueira, P.: Supporting NASA SnowEx remote sensing strategies and requirements for L-band interferometric snow depth and snow water equivalent estimation, Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium, Fort Worth, TX, 1395-1396, DOI: 10.1109/IGARSS.2017.8127224, 2017.

Dixon, D., and Boon, S.: Comparison of the SnowHydro sampler with existing snow tube designs, Hydrol. Process., 26(17), 2555-2562, https://doi.org/10.1002/hyp.9317, 2012.

Dressler, K., Fassnacht, S., and Bales, R.: A comparison of snow telemetry and snow course measurements in the Colorado River basin, J. Hydrometeorol., 7, 705-712, https://doi.org/10.1175/JHM506.1, 2006.

Deleted: ¶
Fick, S., and Hijmans, R.: WorldClim2: new 1-km spatial resolution climate surfaces for global land areas, Int. J. Climatol., 37, 4302-4315, https://doi.org/10.1002/joc.5086, 2017.¶

Goodison, B.: Accuracy of snow samplers for measuring shallow snowpacks: An update, Proceedings of the 35th Annual Eastern Snow Conference, Hanover, NH, 36-49, 1978.

Goodison, B., Ferguson, H., and McKay, G.: Measurement and data analysis. The Handbook of Snow: Principles, Processes, Management, and Use, D. Gray and D. Male, Eds., Pergamon Press, 191-274, 1981.

Goodison, B., Wilson, B., Wu., K, and Metcalfe, J.: An inexpensive remote snow-depth gauge: An assessment, Proceedings of the 52nd Annual Western Snow Conference, Sun Valley, ID, 188-191, 1984.

Goodison, B., Glynn, J., Harvey, K., and Slater, J.: Snow Surveying in Canada: A Perspective, Can. Water Resour. J., 12(2), 27-42, https://doi.org/10.4296/cwrj1202027, 1987.

Gnanadesikan, R., and Kettenring, J.: Robust estimates, residuals, and outlier detection with multiresponse data, Biometrics, 28, 81-124, DOI: 10.2307/2528963, 1972.

Hill, D.F., Wolken, G. J., Wikstrom Jones, K., Crumley, R., and Arendt, A.: Crowdsourcing snow depth data with citizen scientists, Eos*, 99,* https://doi.org/10.1029/2018EO108991, 2018.

Johnson, J., and Marks, D.: The detection and correction of snow water equivalent pressure sensor errors, Hydrol. Proc., https://doi.org/10.1002/hyp.5795, 2004.

Johnson, J., Gelvin, A., Duvoy, P., Schaefer G., Poole, G., and Horton, G.: Performance characteristics of a new electronic snow water equivalent sensor in different climates, Hydrol. Proc., DOI: 10.1002/hyp.10211, 2015.

Jonas, T., Marty, C., and Magnusson, M.: Estimating the snow water equivalent from snow depth measurements, J. Hydrol., 378, 161-167, https://doi.org/10.1016/j.jhydrol.2009.09.021, 2009.

Deleted: ¶
Kang, D.H., and Barros, A.P.: Full-system testing in laboratory conditions of an L-band snow sensor system for in situ monitoring of snow-water content, IEEE Trans. Geosci. Remote Sens., 49(3), 908-919, 10.1109/TGRS.2010.2072786, 2011.¶
¶
Leinss, S., Wiesmann, A., Lemmetyinen, J., and Hajnsek, I.: Snow water equivalent measured by differential interferometry, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 8(8), 3773-3790, 10.1109/JSTARS.2015.2432031, 2015.¶

Leys, C., Klein, O., Dominicy, Y., and Ley, C.: Detecting multivariate outliers: use a robust variant of the Mahalanobis distance, J. Exp. Soc. Psychol., 74, 150-156, https://doi.org/10.1016/j.jesp.2017.09.011, 2018.

Liang, X., Lettermaier, D., Wood, E., and Burges, S.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, J. Geophys. Res. Atmos., 99(D7), 14,415-14,428, https://doi.org/10.1029/94JD00483, 1994.

Liston, G., and Elder, K.: A distributed snow evolution modeling system (SnowModel), J. Hydrometerol., 7, 1259-1276, https://doi.org/10.1175/JHM548.1, 2006.

Lundberg, A., Richardson-Naslund, C., and Andersson, C.: Snow density variations: consequences for ground penetrating radar, Hydrol. Process., 20, 1483-1495, https://doi.org/10.1002/hyp.5944, 2006.

Maine Geological Survey: Maine Cooperative Snow Survey Dataset, https://www.maine.gov/dacf/mgs/hazards/snow_survey/, 2018.

McKay, G., and Findlay, B., 1971: Variation of snow resources with climate and vegetation in Canada, Proceedings of the 39th Western Snow Conference, Billings, MT, 17-26, 1971.

De Maesschalck, R., Jouan-Rimbaud, D., and Massart, D.: The Mahalanobis distance, Chemometr. Intell. Lab. Syst., 50(1), 1-18, https://doi.org/10.1016/S0169-7439(99)00047-7, 2000.

McCreight, J., and Small, E.: Modeling bulk density and snow water equivalent using daily snow depth observations, The Cryosphere, 8, 521-536, https://doi.org/10.5194/tc-8-521-2014, 2014.

Meloysund, V., Leira, B., Hoiseth, K., and Liso, K.: Predicting snow density using meterological data, Meteorol. Appl., 14, 413-423, https://doi.org/10.1002/met.40, 2007.

Menne, M.J., I. Durre, R.S. Vose, B.E. Gleason, and Houston, T.G.: An overview of the Global Historical Climatology Network-Daily Database, J. Atmos. Ocean. Technol., 29, 97-910, doi:10.1175/JTECH-D-11-00103.1, 2012.

Molotch, N.P., and Bales, R.C.: SNOTEL representativeness in the Rio Grande headwaters on the basis of physiographics and remotely sensed snow cover persistence, Hydrol. Process., 20(4), 723-739, https://doi.org/10.1002/hyp.6128, 2006.

Mote, P., Li, S., Letternaier, D., Xiao, M., and Engel, R.: Dramatic declines in snowpack in the western US," npj Clim. Atmos. Sci., 1(2), 1-6, doi:10.1038/s41612-018-0012-1, 2018.

Formatted: No underline, Font color: Auto

Mizukami, N., and Perica, S.: Spatiotemporal characteristics of snowpack density in the mountainous regions of the western United States, J. Hydrometeorol., 9, 1416-1426, https://doi.org/10.1175/2008JHM981.1, 2008.

New York Snow Survey, NOAA, Northeast Regional Climate Center at Cornell University, 2018.

Pagano, T., Garen, D., Perkins, T., and Pasteris, P.: Daily updating of operational statistical seasonal water supply forecasts for the western U.S., J. Am. Water Resour. Assoc., 45(3), 767-778, https://doi.org/10.1111/j.1752-1688.2009.00321.x, 2009.

Painter, T., Berisford, D., Boardman, J., Bormann, K., Deems, J., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw, R., Marks, D., Mattmann, C., Mcgurk, B., Ramirez, P., Richardson, M., Skiles, S., Seidel, F., and Winstral, A.: The Airborne Snow Observatory: fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo, Remote Sens. Environ., 184, 139-152, doi:10.1016/j.rse.2016.06.018, 2016.

Pistocchi, A.: Simple estimation of snow density in an Alpine region, J. Hydrol. Reg. Stud., 6, 82-89, http://dx.doi.org/10.1016/j.ejrh.2016.03.004, 2016.

Rousseeuw, P.: Least Median of Squares Regression, J. Am. Stat. Assoc., 79, 871-880, DOI: 10.1080/01621459.1984.10477105, 1984.

Ryan, W., Doesken, N., and Fassnacht, S.: Evaluation of Ultrasonic Snow Depth Sensors for U.S. Snow Measurements, J. Atmos. Ocean. Technol., 25, 667-684, https://doi.org/10.1175/2007JTECHA947.1, 2008.

Schaefer, G., Cosh, M., and Jackson, T.: The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN), J. Atmos. Ocean. Technol., 24, 2073-2077, https://doi.org/10.1175/2007JTECHA930.1, 2007.

Serreze, M., Clark, M., Armstrong, R., McGinnis, D., and Pulwarty, R.: Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data, Water Resour. Res., 35(7), 2145-2160, https://doi.org/10.1029/1999WR900090, 1999.

Shanley, J., and Chalmers, A.: The effect of frozen soil on snowmelt runoff at Sleepers River, Vermont, Hydrol. Process., 13(12-13), 1843-1857, https://doi.org/10.1002/(SICI)1099-1085(199909)13:12/13<1843::AID-HYP879>3.0.CO;2-G, 1999.

Sturm, M., Holmgren, J., and Liston, G.: A seasonal snow cover classification system for local to global applications, J. Clim., 8, 1261-1283, https://doi.org/10.1175/1520-0442(1995)008<1261:ASSCCS>2.0.CO;2, 1995.

Deleted: ¶
Sokol, J., Pultz, T., and Walker, A.: Passive and active airborne microwave remote sensing of snow cover," Int. J. Remote. Sens., 24, 5327-5344, https://doi.org/10.1080/0143116031000115076, 2003.¶

Sturm, M., Taras, B., Liston, G.E., Derksen, C., Jonas, T., and Lea, J.: Estimating snow water equivalent using snow depth data and climate classes, J. Hydrometeorol., 11, 1380-1394, https://doi.org/10.1175/2010JHM1202.1, 2010.

U.S. Army Corps of Engineers: Snow hydrology: Summary report of the snow investigations of the North Pacific Division, 437pp., 1956.

U.S. Department of Agriculture: The History of Snow Survey and Water Supply Forecasting. Interviews With U.S. Department of Agriculture Pioneers, D. Helms, S. Phillips and P. Reich (eds.), Natural Resources Conservation Service, 2008.

U.S. Department of Agriculture: Snow Survey and Water Supply Forecasting. National Engineering Handbook Part 622, Water and Climate Center, Natural Resources Conservation Service, 2011.

Wang, T., Hamann, A., Spittlehouse, D.L., and Murdock, T.: ClimateWNA - High-Resolution Spatial Climate Data for Western North America, J. Appl. Meteorol. Climatol., 51, 16-29, https://doi.org/10.1175/JAMC-D-11-043.1, 2012.

Wang, T., Hamann, A., Spittlehouse, D.L., and Carroll, C: Locally downscaled and spatially customizable climate data for historical and future periods for North America, PLoS One, 11, DOI:10.1371/journal.pone.0156720.

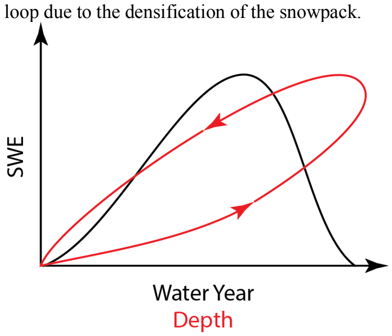
Wigmosta, M.S., Vail, L., and Lettenmaier, D.: A distributed hydrology-vegetation model for complex terrain, Water Resour. Res., 30, 1665-1679, https://doi.org/10.1029/94WR00436, 1994

Deleted: ¶
Vuyovich, C., Jacobs, J., and Daly, S.: Comparison of passive microwave and modeled estimates of total watershed SWE in the continental United States, Water Resour. Res., 50(11), 9088-9012, https://doi.org/10.1002/2013WR014734, 2014.¶

Formatted: Space After: 5 pt

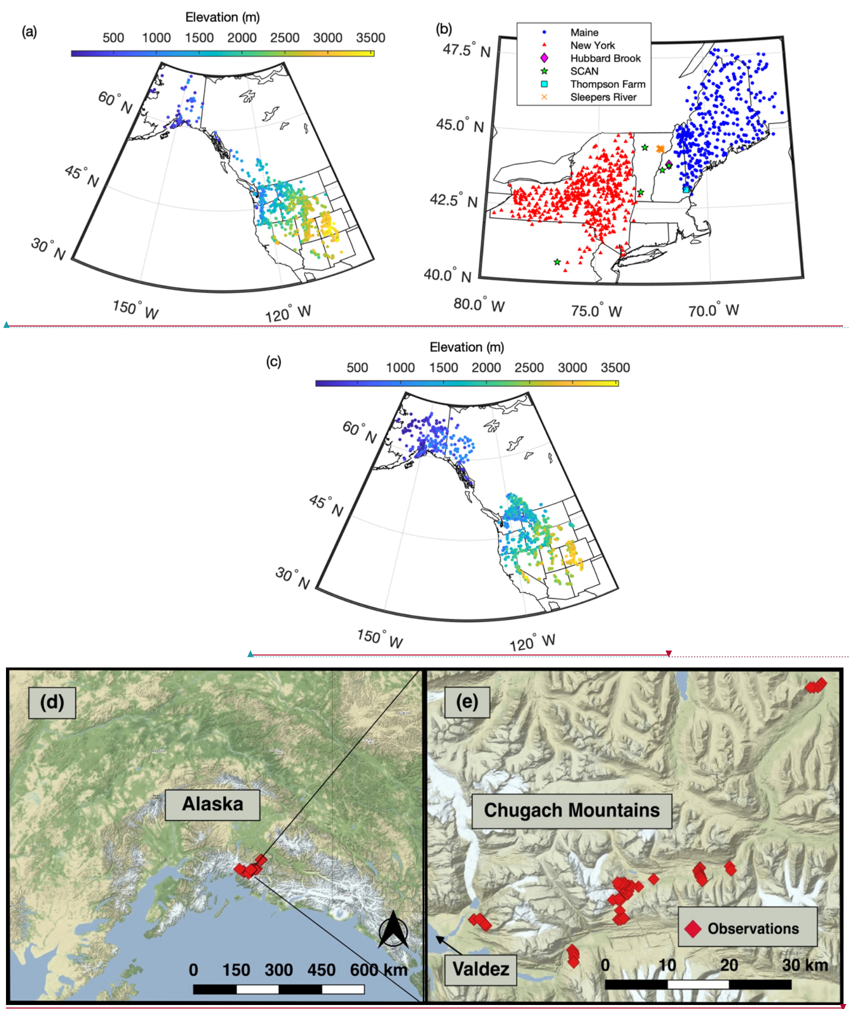
Deleted: ¶

Figure 1: Conceptual sketch of the evolution of snow water equivalent (SWE) over the course of a water year (black line). Also shown is the evolution of SWE with snowpack depth over a water year (red line). Note the hysteresis loop due to the densification of the snowpack.



Formatted: Line spacing: single

Figure 2: Distribution of measurement locations used in this study. (a) Western USA and Canada snow pillow locations, with colors indicating station elevation in meters. (b) Northeast USA snow pillow and snow course locations, with stations colored according to data source. (c) Western North America snow course and aerial marker locations, with colors indicating station elevation in meters. (d, e) Measurement sites in the Chugach Mountains, southcentral Alaska.



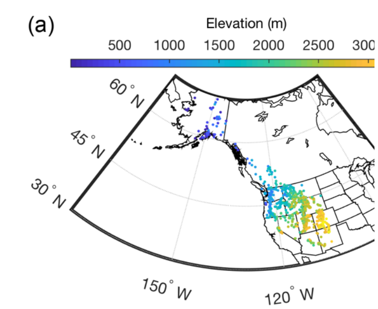
Formatted: Line spacing: single

Deleted: station

Deleted: c

Deleted: d

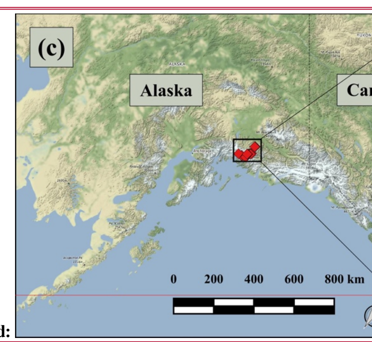
Formatted: Font: 10 pt



Deleted: ... [1]

Formatted: Font: 10 pt

Formatted: Centered



Deleted:

Deleted:

Figure 3: Sample time series of SWE and *h* from the Rex River (WA) SNOTEL station. Observations of *h* at times when SWE is zero are likely spurious.

Formatted: Line spacing: single

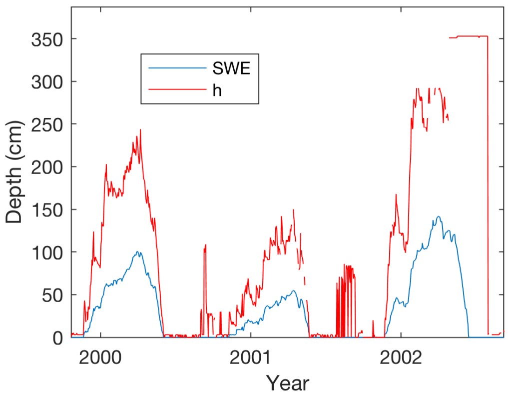
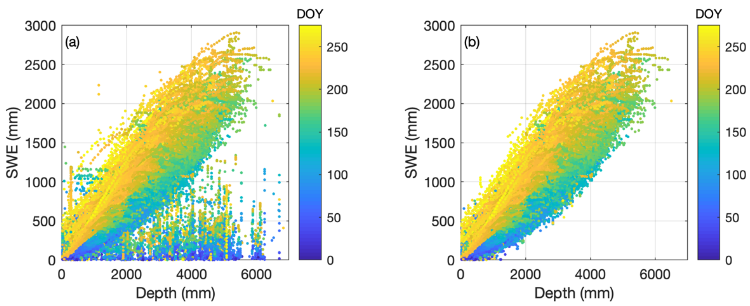


Figure 4: Scatter plot of SWE vs. *h* for the complete SNOTEL dataset before (a) and after (b) removing data points, following the method described in Section 2.1.1.5. Symbols are colored by 'day of water year' (*DOY*; October 1 is the origin).



Formatted: Line spacing: single

Formatted: Font: 10 pt

Deleted:

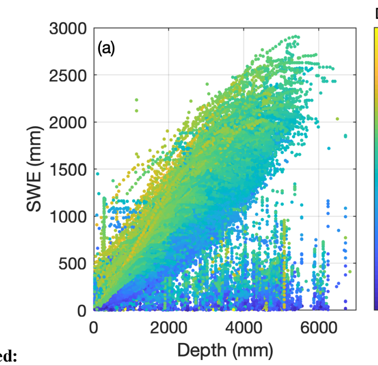
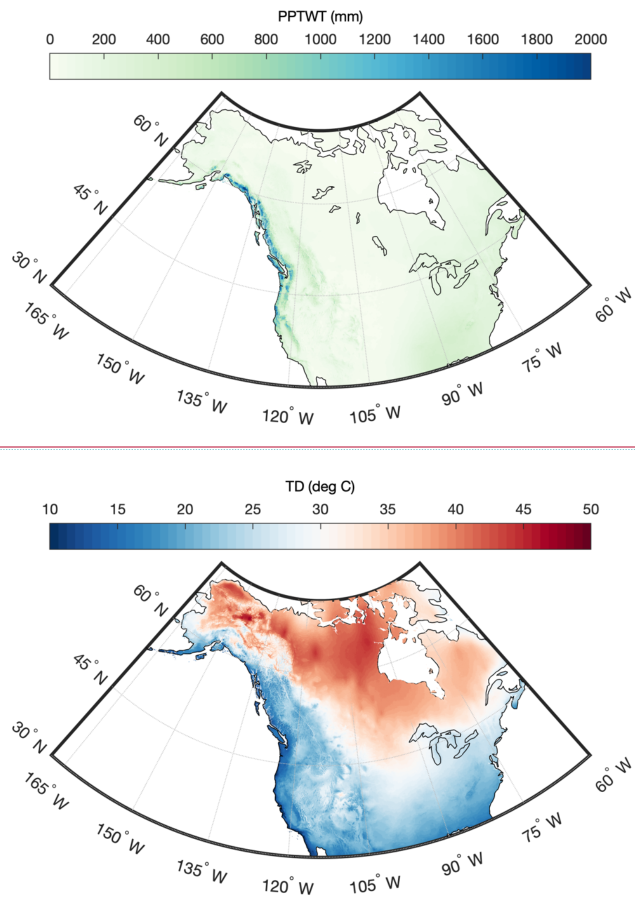


Figure 5: Gridded maps of winter (December, January, February) precipitation (PPTWT) and temperature difference (TD) between mean of warmest month and mean of coldest month) for North America. Maps are for the 1981-2010 climatological period.



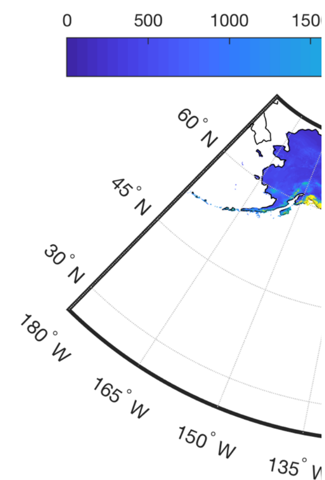
Formatted: Line spacing: single

Deleted: mean annual precipitation…inter (December, January, February) (*MAP*)…recipitation (PPTWT) and temperature difference (TD) between mean of warmest month and mean of coldest month) and mean February temperature ($\bar{T}_{F_{mean}}$) for the study regions ... [2]

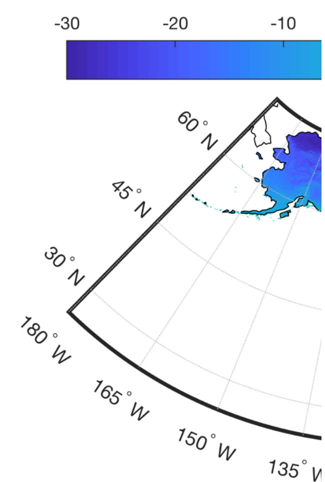
Deleted: Climate normals are from the PRISM data set (1981-2010 for CONUS and British Columbia; 1971-2000 for Alaska).

Formatted: Font:

Formatted: Font:



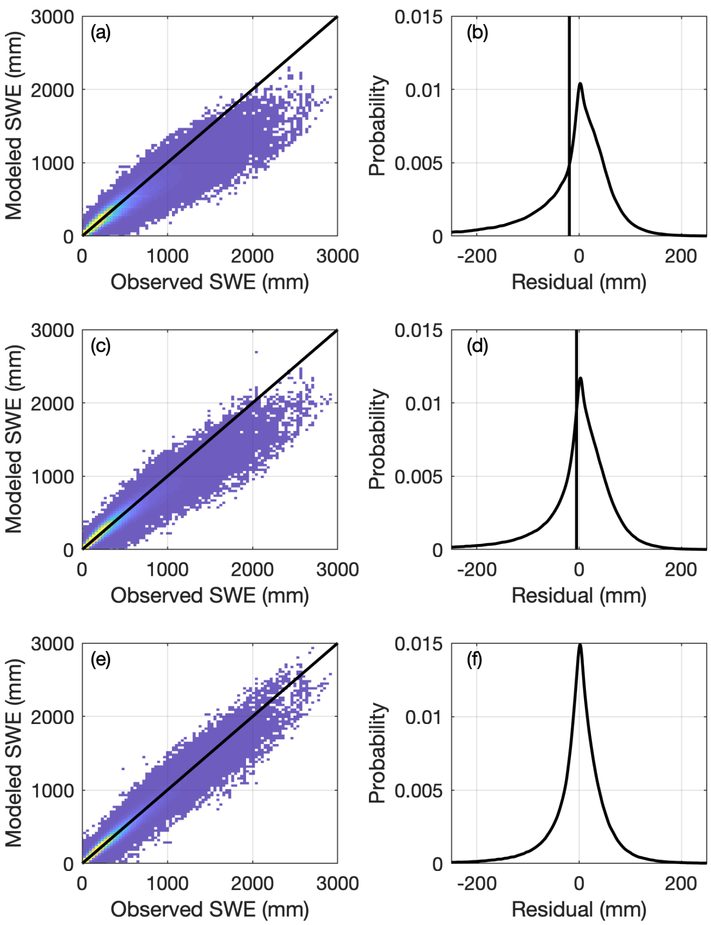
Deleted:



Deleted:

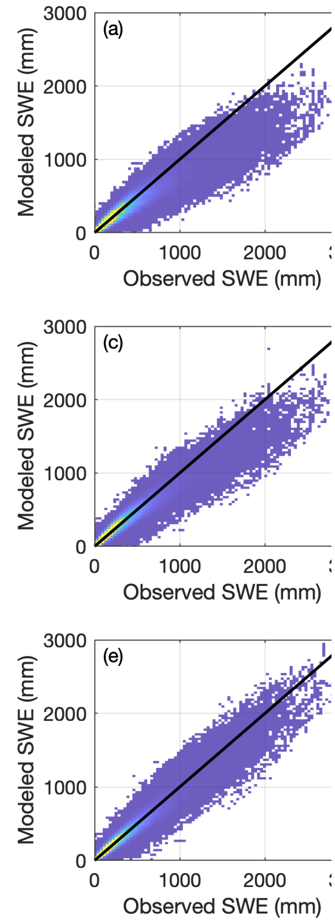
Formatted: Font:

Figure 6: Two-dimensional histograms (heat maps; left column) of modeled vs. observed SWE and probability density functions (right column) of the residuals for three simple models applied to the CONUS, AK, and BC snow pillow data. Warmer colors in the heat maps indicate greater density. The vertical lines in the right column indicate the location of the mean residual, or bias. Top row (a-b): One-equation model (Section 2.2.1). Middle row (c-d): Two-equation model (Section 2.2.2). Bottom row (e-f): Multi-variable two-equation model (Section 2.2.3).



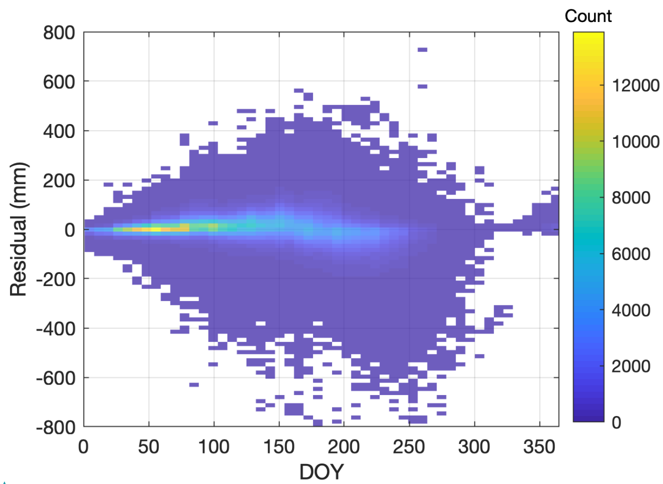
Commented [Office2]: FIX (F)

Formatted: Line spacing: single

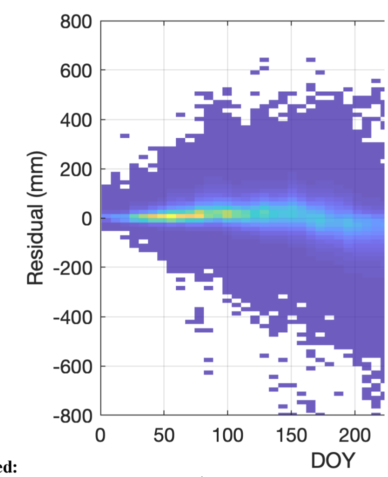


Deleted:

Figure 7: Heat map of SWE residuals as a function of *DOY*.



Formatted: Font: 10 pt

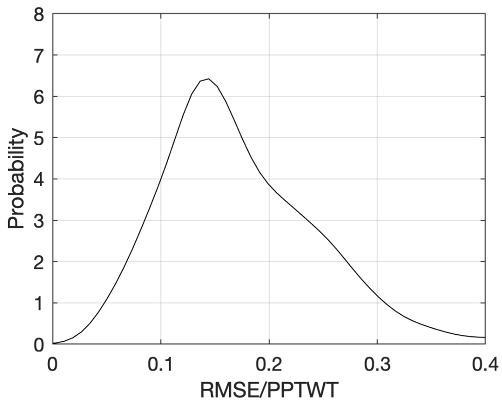


Deleted:

Page Break

Formatted: Font: 10 pt

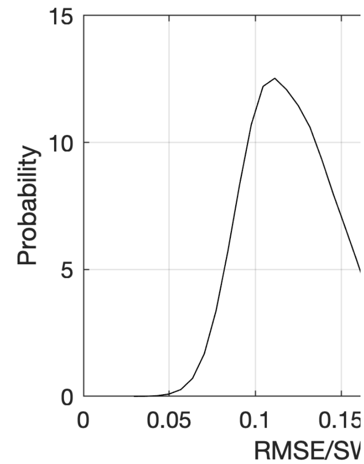
Figure 8: Probability density function of snow pillow station root-mean-square error (RMSE) normalized by station winter precipitation (*PPTWT*).



Formatted: Line spacing: single

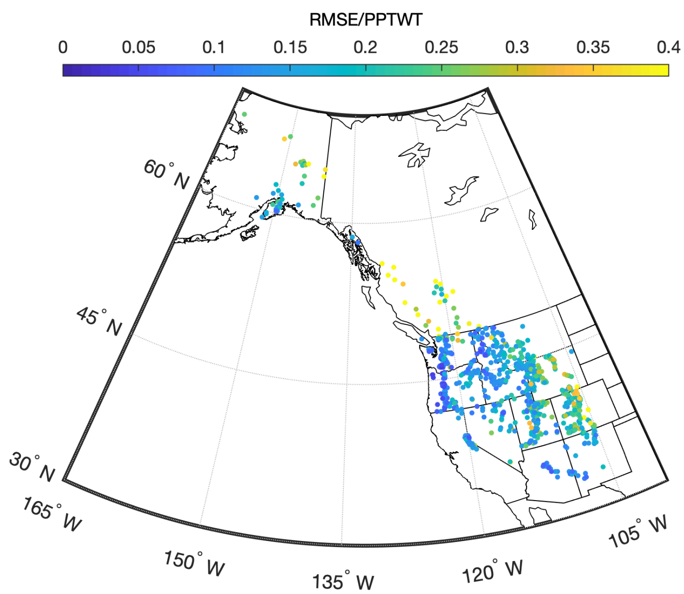
Deleted: mean annual maximum SWE

Formatted: Font: Italic



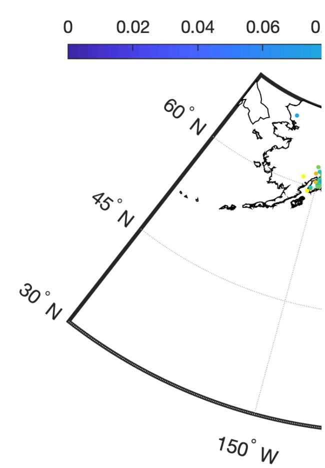
Deleted:

Figure 9: Spatial distribution of snow pillow station root-mean-square error (RMSE) normalized by station winter precipitation (PPTWT).



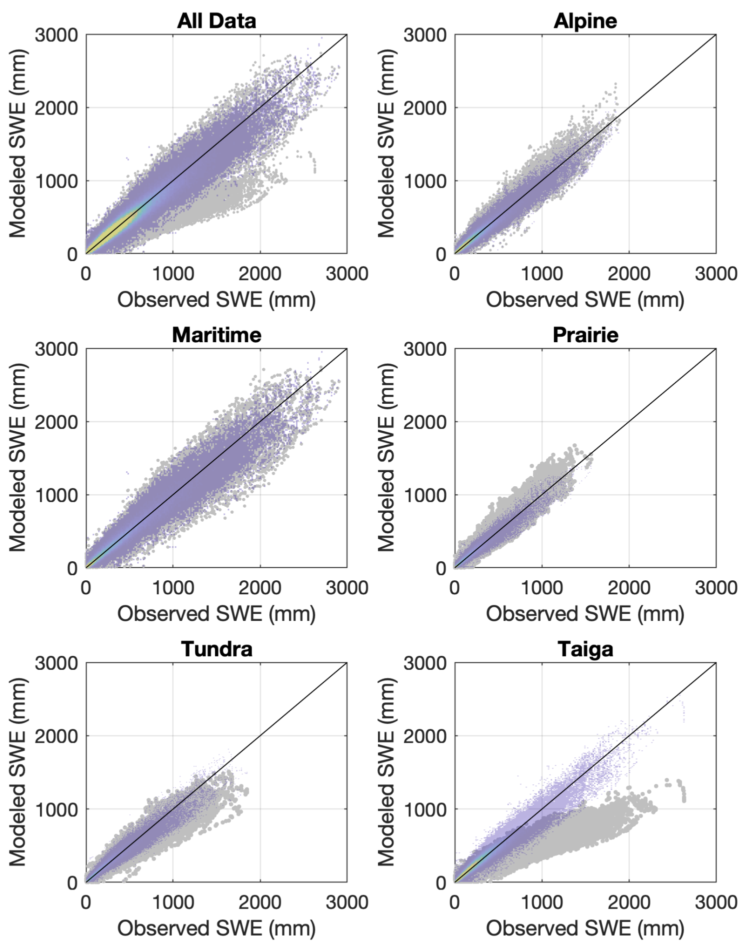
Formatted: Line spacing: single

Deleted: mean annual maximum SWE



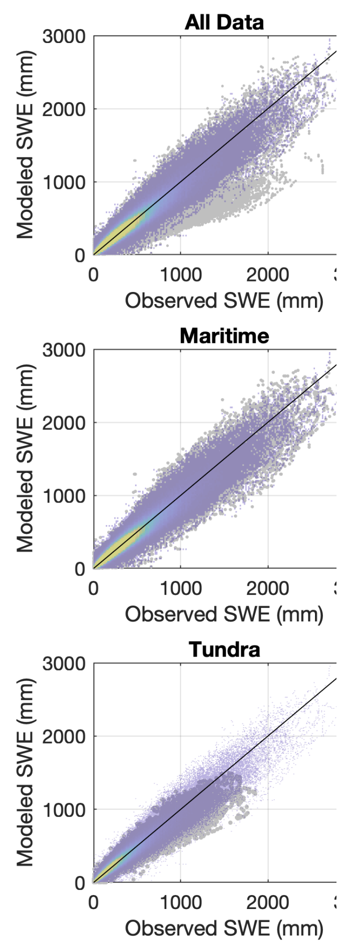
Deleted:

Figure 10: Comparison of the multi-variable, two-equation model of the current study with the model of Sturm et al. (2010). The subpanels show modeled SWE vs. observed SWE for all of the data binned together, as well as for the data broken out by the snow classes identified by Sturm et al. (1995). The gray symbols show the Sturm result and the transparent heat maps (warmer colors indicate greater density) show the current result. The models are being applied to the validation data set (50% of the aggregated snow pillow data for CONUS, AK, and BC).



Formatted: Line spacing: single

Deleted: present



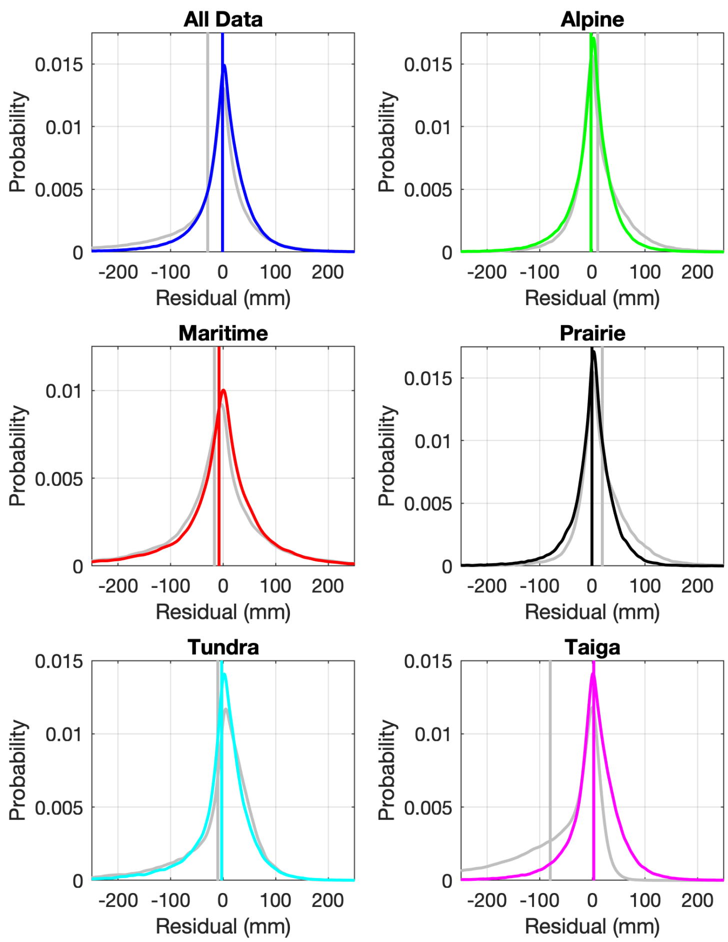
Deleted:

Formatted: Font: 10 pt

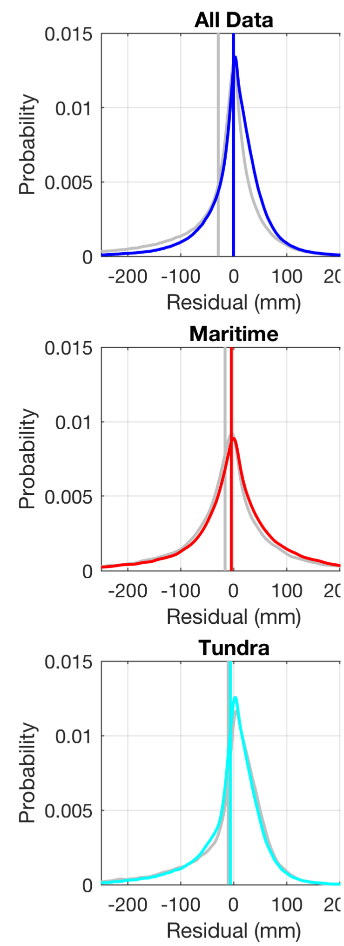
Figure 11: Comparison of the multi-variable, two-equation model of the current study with the model of Sturm et al. (2010). The subpanels show probability density functions of the residuals of the model fits for all of the data binned together, as well as for the data broken out by the snow classes identified by Sturm et al. (1995). The gray lines show the Sturm result and the colored lines show the current result. The vertical lines show the mean error, or the model bias, for both the Sturm and the current result. The models are being applied to the validation data set (50% of the aggregated snow pillow data for CONUS, AK, and BC).

Formatted: Line spacing: single

Deleted: present

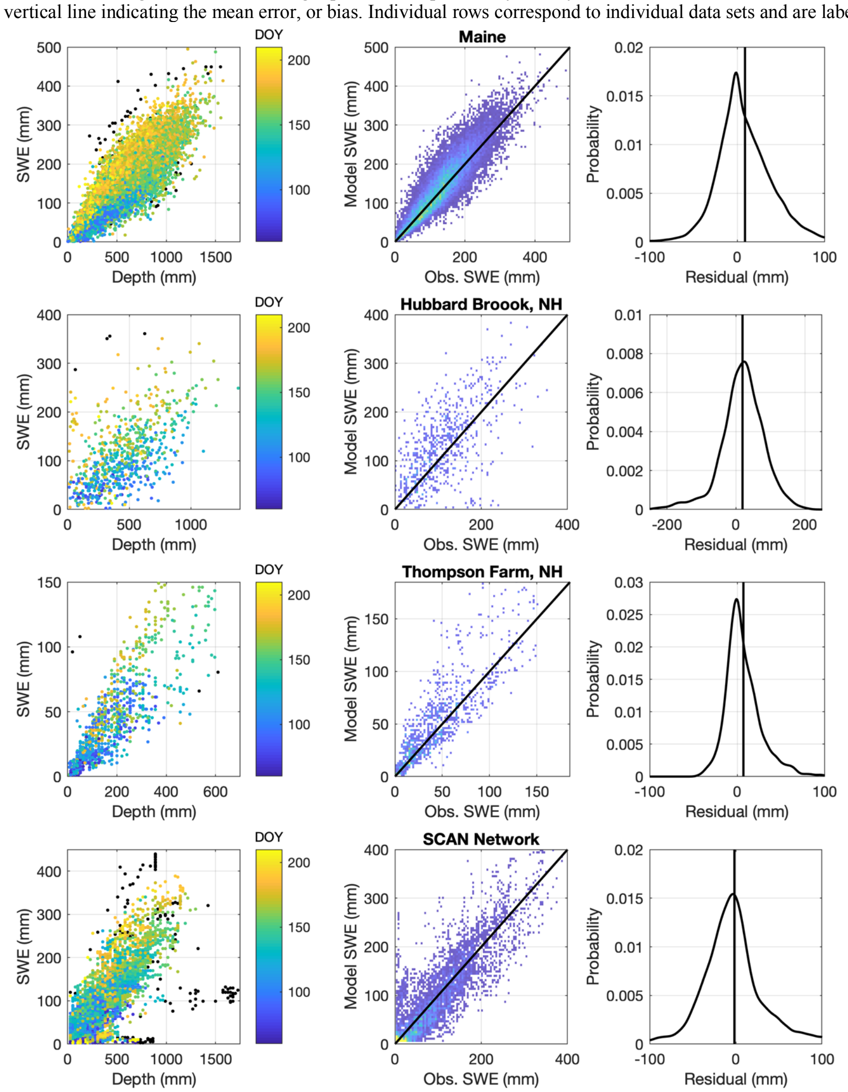


Deleted:



Formatted: Font: 10 pt

Figure 12: Results from application of the multi-variable, two-equation model to numerous east coast datasets. The left column shows the SWE-*h* data for each dataset. Note that the black symbols are points removed by the outlier detection procedure discussed in section 2.1.1.4. The remaining symbols are colored by *DOY*. The middle panel plots heat maps of the model estimates of SWE against the observations of SWE with the 1:1 line included. Warmer colors indicate higher densities. The right panel shows probability density functions of the model residuals, with the vertical line indicating the mean error, or bias. Individual rows correspond to individual data sets and are labeled.



Formatted: Line spacing: single

Formatted: Font: Italic

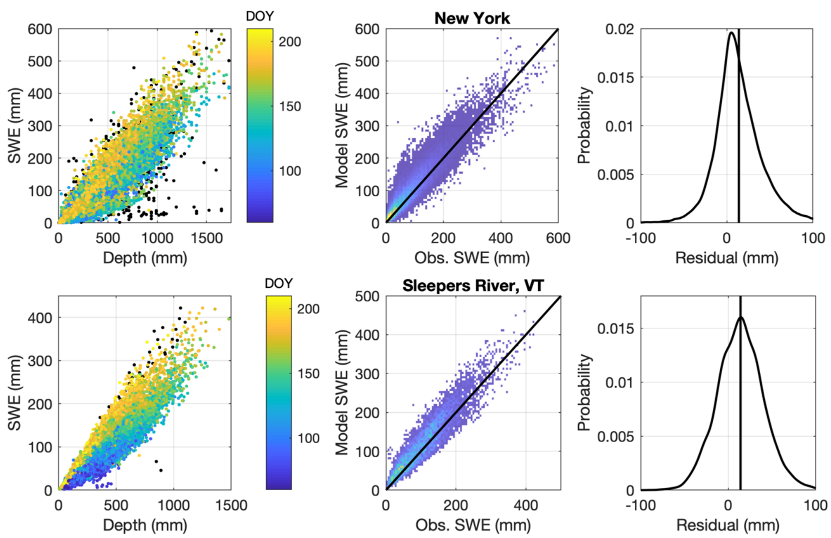
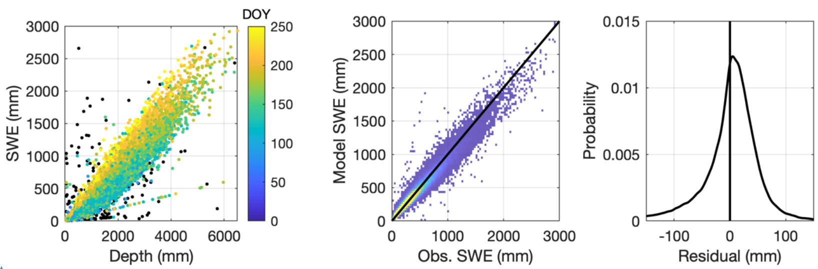

New York
Sleepers River, VT
SWE (mm)
Depth (mm)
DOY
Model SWE (mm)
Obs. SWE (mm)
Probability
Residual (mm)

Figure 13: Results from application of the multi-variable, two-equation model to the NRCS snow course / aerial marker dataset. The left column shows the SWE-*h* data for each dataset. Note that the black symbols are points removed by the outlier detection procedure discussed in section 2.1.1.5. The remaining symbols are colored by *DOY*. The middle panel plots heat maps of the model estimates of SWE against the observations of SWE with the 1:1 line included. Warmer colors indicate higher densities. The right panel shows probability the density function of the model residuals, with the vertical line indicating the mean error, or bias.



Formatted: Font: Italic

Formatted: Font: 10 pt

Page 25: [1] Deleted	Microsoft Office User	5/28/19 11:05:00 AM

Page 28: [2] Deleted	Microsoft Office User	5/21/19 1:48:00 PM

Page 28: [2] Deleted	Microsoft Office User	5/21/19 1:48:00 PM

Page 28: [2] Deleted	Microsoft Office User	5/21/19 1:48:00 PM